CAPP
Centro di Analisi delle Politiche Pubbliche

From SHIW to IT-SILC: construction and representativeness of the new CAPP_DYN first-year population

Emanuele Ciani, Donatella Fresu

CAPP*aper* n. 92
giugno 2011

Università di Modena e Reggio Emilia Facoltà di Economia Marco Biagi

Università di Bologna Dipartimento di Scienze Economiche

CAPP - Centro di Analisi delle Politiche Pubbliche
Dipartimento di Economia Politica - Università di Modena e Reggio Emilia
Ufficio 54 - Ala Ovest
Viale Berengario, 51 41100 Modena - ITALY
phone: +39 059 2056854 fax: +39 059 2056947
email capp@unimo.it

EUROPEAN COMMISSION
Employment, Social Affairs and Equal Opportunities

Social Protection and Integration
Social Protection, Social Services

COMMUNITY PROGRAMME FOR EMPLOYMENT AND SOCIAL SOLIDARITY (PROGRESS)

Action related to the development of administrative datasets and models for labour market and pensions analysis

Project ASSESSING ADEQUACY AND LONG TERM DISTRIBUTIVE EFFECTS OF THE ITALIAN PENSION SYSTEM. A MICROSIMULATION APPROACH VS/2009/0498

This publication is supported under the European Community Programme for Employment and Social Solidarity – PROGRESS (2007-2013).

This programme is managed by the Directorate-General for Employment, social affairs and equal opportunities of the European Commission. It was established to financially support the implementation of the objectives of the European Union in the employment and social affairs area, as set out in the Social Agenda, and thereby contribute to the achievement of the Lisbon Strategy goals in these fields.

The seven years Programme targets all stakeholders who can help shape the development of appropriate and effective employment and social legislation and policies across the EU-27, EFTA- EEA and EU candidate and pre-candidate countries.

PROGRESS mission is to strengthen the EU contribution in support of Member States commitment. PROGRESS will be instrumental in:

- *providing analysis and policy advice on PROGRESS policy areas;*
- *monitoring and reporting on the implementation of EU legislation and policies in PROGRESS policy areas;*
- *promoting policy transfer, learning and support among Member States on EU objectives and priorities; and*
- *relaying the views of the stakeholders and society at large*

For more information see: http://ec.europa.eu/progress

From SHIW to IT-SILC: construction and representativeness of the new CAPP_DYN first-year population

Emanuele Ciani, Donatella Fresu

Abstract

This paper describes the construction of the initial population in CAPP_DYN and illustrates the degree of representativeness of the Italian population in 2006. While the previous version of the model was constructed using the Bank of Italy Survey on Household Income and Wealth, the current version is based on the new ISTAT Survey on Income and Living Conditions. The first part of the paper discusses the reasons that led us to this switch. It also provides full details on the operations carried out on the original dataset in order to obtain a sample that can be used as the first year population of the Dynamic Simulation Model. In the second part of the paper, the demographic and socio-economic characteristics of the initial population sample are compared with information coming from other sources, such as administrative archives, national accounts and Labour Force Surveys. This exercise is crucial in assessing the capability of CAPP_DYN to represent the population's characteristics at the starting point.

JEL Classification: C63, C81, D31, H55

Keywords: Dynamic micro-simulation, model validation, income distribution, pensions and social security

Summary

1. Introduction

The choice of the database to use in setting up the initial population of a dynamic micro-simulation model is extremely important. Its representativeness — i.e. the capability of the data source selected to represent at the starting year the reference population's socio-economic characteristics, can decisively influence the evolution of the medium and long term outputs generated by the dynamic model. This is the main reason why many dynamic micro-simulation models use census or administrative data in building the population at the base year[1]. Aside from benefits and limits deriving from using random draws from census/administrative records with the purpose of building the population at the base year, in the Italian context making use of sample surveys turns out to be in many ways an unavoidable choice. Setting up a population drawn from census records would encounter administrative difficulties not compatible with the time horizon of this research project.

That being so, it's then a matter of defining the criterion of choice to recognize which, among those available, can be the "best" source of microdata.

A previous CAPP report (Ministry of Social Affairs, 2008) evaluated the degree of representativeness of the two primary sample surveys in Italy giving detailed information on socio-economic characteristics of the population: the Bank of Italy's *Survey on Household Income and Wealth* (Indagine sui Bilanci delle Famiglie Italiane, SHIW) and the Italian component of the European survey *Statistics on Income and Living Conditions* (IT-SILC). That study, referring to the period 2004-2005, revealed a greater reliability of SHIW with respect to the civil status and of IT-SILC with respect to the occupational and income condition, whereas no significant difference was found between the two sources with respect to age and educational qualification.

In this part of the report the comparison has been improved and extended in light of the new data available (2007) and of a series of adjustments in processing the IT-SILC data. The objective of this work is to (re-)evaluate pros and cons of building the CAPP_DYN initial sample by using IT-SILC.

---

[1] Models which make use of data coming from census or administrative records are for instance: DYNAMOD (Australia), DYNACAN (Canada), NEDYMAS (Netherlands), MOSART3 (Norway), PENSIM2 (England), CORSIM (USA). But the limited informative content of these archives requires the employment of appropriate *statistical matching* procedures with other statistical sources (of cross-section or panel data) to supplement the set of information available; on the other hand, the more widespread *matching* techniques, based on *propensity score* or on the *mahalanobis distance*, are founded on some hardly testable assumptions (see for instance Blundell et al. (2005); Caliendo and Kopeinig (2005)), which, if not fulfilled, invalidate the results produced.

The former CAPP_DYN version did indeed use the 2002 SHIW cross-section as main informative basis for setting up the initial population. In the period in which the model frame was built, the Bank of Italy's survey represented the richest and most analytical informative source available in our country for empirical analysis of the economic behaviour of individuals and households. It collected (with biennial frequency) detailed information referring to income, savings and property, as well as on socio-demographic characteristics (Banca d'Italia, 2004).

Starting from 2004, a new sample survey is available in Italy. It collects very comprehensive socio-demographic and income information of individuals and households. Conducted annually, the Italian component of the European survey *Statistics on Income and Living Conditions* (IT-SILC) is a rich source of data which offers many advantages with respect to any other available sample survey, first and foremost, the numerousness of observations in the sample and the integration of income information provided by interviewees with administrative records. During the updating phase of the CAPP_DYN basis, our research unit therefore felt it appropriate to reconsider meticulously the use of IT-SILC in place of, or together with, SHIW. The feasibility study, in addition to exploiting the knowledge acquired with former projects (see Ministry of Social Affairs, 2008; Savegnago, 2008), required investing further skills in understanding the extensive informative content of the survey and its distinctive features. The research unit then performed some comparisons with the socio-economic information made available by the official sources, enriching those previously published, in the version with and without use of sampling weights. Having confirmed the superiority of SILC with respect to SHIW, the group then proceeded to select which, among the different IT-SILC cross-sections available, should be used in order to build the initial base. Ultimately the choice fell on the last survey available to date and referring to 2007 (income 2006). This is the fourth IT-SILC survey which:

- closes the first complete rotational panel of four-year length[2];
- has probably overcome the classic problems generated in the initial survey adjustment phase (especially a smaller incidence of the selection bias caused by a large and selective attrition of the survey in the early stages);
- allows a comparison with the SHIW 2006 sample.

The next step involved selection of the set of information necessary for the dynamic simulation and its validation. As described hereafter, some variables underwent careful tests

and, in some cases, some sort of calibration or other devices were required to improve the statistical representativeness of the sample.

Paragraph 2 presents the main survey characteristics and describes in detail pros and cons related to using IT-SILC. Paragraph 3 describes the procedures followed to set up the initial population. Some variables, not directly included in IT-SILC, were derived on the basis of some assumption, better illustrated hereafter. Paragraph 4 reports the results of the representativeness analysis with respect to socio-demographic variables on the initial CAPP_DYN population. In this analysis the initial CAPP_DYN population was compared to the information coming from administrative records and other sample surveys. Paragraph 5 analyzes the degree of representativeness of income variables, relative both to earned income and to pension transfers. Paragraph 6 compares the IT-SILC and SHIW income distributions.

2. The choice of IT-SILC

2.1 Description of the SILC survey

As from 2004, the European survey *Statistics on Income and Living Conditions* (hereinafter EU-SILC) is conducted yearly in 25 member States of the European Union, plus Norway and Iceland. The survey has the purpose of collecting information on income and living conditions of households in different countries of the European Union. The main parameters of interest for each year of the survey are the percentage of individuals in poor circumstances and the average household income (Istat, 2008: 23). The survey enables the variation in time in the level of the economic well-being or the persistence of interviewees in poverty circumstances to be detected. More specifically, EU-SIL acquires information on the main income categories, inclusive or not of tax burden, social security contributions and positive transfers by the tax-benefit system, i.e. family allowances, unemployment benefits and any other welfare allowance. Unlike the previous *European Community Household Panel* (ECHP), EU-SILC has been modelled to be *output-harmonized*, and not *input-harmonized*; in other words, the sample and the questionnaire are built up in different ways among the participating countries, then arranged to construct a set of common variables on the basis of

[2] The 2004-2007 longitudinal component could then be used to generate the transition conditional probabilities required by the model for the medium/long-term simulation.

established definitions[3]. With respect to living conditions, EU-SILC contains both quantitative information concerning income received by family members and qualitative indicators. Among the latter, in addition to the traditional questions referring to house quality and occupational status, other questions have been included to detect conditions of financial stress.

For instance, households are asked to report whether they had any difficulties in bearing essential expenses (like going to medical specialists or buying teaching material) or paying the instalments on loans. A possible problem is linked to the different time horizon to which qualitative and quantitative indicators refer. Part of the information concerning occupational status refers to the year of the survey (hereinafter defined as period *t*), while income variables refer to the previous calendar year (period *t-1*)[4]. This survey characteristic may cause some inconsistencies in the model base population. In the attempt to reduce these problems, we chose to prefer variables referring to the period *t-1*, with respect both to occupational status and to socio-demographic information. This decision is in line with what has been done in the context of the principal static model of *Tax-Benefit* at European level, EUROMOD, which also makes use of the EU-SILC data base. A detailed description of the procedure followed is available in paragraph 3.

The Italian survey (hereinafter IT-SILC), conducted by Istat, is made available to researchers in the form of microdata. The Italian component is of particular interest since, besides containing all the common variables at European level, it includes specific variables, which are available only in the Italian questionnaire. In our case, the analytic information regarding social security benefits and earned income received by interviewees is especially useful. In Italy, income information is integrated with data coming from administrative records, with the purpose of reducing measurement errors due to the quality of answers provided by the interviewees (Istat, 2009a). This operation is based on *exact matching* among individual sampling record and records coming from fiscal and pension archives ("Casellario centrale dei pensionati"), using the tax code as an identification key. Istat then proceeds to reconcile the two sources on the basis of some operational assumptions, described briefly in the paragraph 2.2.

---

[3] This aspect entails some difficulties in using the survey for comparisons among different countries, also since some member states opted for the use of administrative records alone. However, for our purposes the problem is not significant, since the simulation refers only to the Italian component.

[4] To be precise, the individual questionnaires were administered only to individuals who were at least 15 years old at the end of period *t-1*.

The survey design selected by Istat has an integrated nature (Istat, 2008: 24 and following), in which four panels of four year length overlap in time enabling both longitudinal and cross-sectional estimates to be performed. The sampling scheme for each panel is characterized by two stages: the first concerns municipalities, stratified according to their demographic dimension; the second concerns households[5]. To take into consideration the possible distortions deriving from selective missing responses to the survey, Istat provides a vector of sampling weight, describing "*how many*" individuals each observation represents in the real population, which are built at the household level as the inverse of the probability to be included in the sample. The weights are developed conditionally to a set of observable characteristics deemed to be significant in explaining the possible selection effects in failing to answer the interview[6]. The sampling total of the fourth survey (2007) comprises 52,772 individuals and 20,982 households, a number roughly double the SHIW sample.

2.2 Pros and Cons of using IT-SILC

There are at least three main reasons supporting the choice of shifting from SHIW to the IT-SILC survey: the integration procedure of Istat sampling data with information from administrative records; the presence in IT-SILC of more detailed and accurate items regarding pensions; the greater amount of sampling observations in the Istat survey.

In particular, the main advantage consists in Istat's choice to implement an innovative integration process between sample IT-SILC information and data from administrative

---

[5] Reading the reference manual provided by Istat (2008: 30), one can infer that the four rotational groups are independent of each other: "The sample of each Ar municipality (*self-representative*, or "*autorappresentativo*", *N.d.R.*) has been divided in four sub-samples of equal dimension, each of them randomly assigned to a rotational group. For Nar municipalities (*non self-representative*, or " *non autorappresentativo*", *N.d.R.*) instead, the assignment to groups concerned the whole sample of households of each sample municipality; therefore, every municipality drawn in each Nar stratum is randomly linked to a single rotational group."

[6] The IT-SILC survey provides a set of weights, defined as "coefficienti di riporto all'universo" , that enable more precise estimates of the overall value of the surveyed variable. Weights are developed in such a way as to take into consideration the different probability of inclusion given by the sampling design and by the totally missing response to the interview. Moreover, the reference manual (Istat, 2008: 38) specifies that the weights of the cross-sectional sample are linked to some known values:

- Resident population by region, sex and age class at 31st December of period *t-1;*
- Resident population by area, sex and age class at 31st December of period *t*;
- Foreign resident population over eighteen years old, by area and sex at 31st December of period *t-1*;
- Foreign resident population over eighteen years old, by area and origin at 31st December of period *t-1*;
- Resident population by area and demographic size of the municipality in period *t-1*.

records. This characteristic ensures not only a greater reliability of income data, but also the opportunity to dispose of some variables that are not included in the sampling survey alone: the amount of social contributions and taxes paid during the period *t-1*. To sum up, and referring to the specific manual for further details (Istat, 2009a), Istat traced each individual tax code, then matched exactly sample information with records contained in the administrative sources:

- The CUD form, for earned income, pensions and severance pay;
- The 730 form, which, in addition to the previous form, contains information on income from land and buildings and on income from self-employment not arising from professional activities;
- The "*Unico Persone Fisiche*" form, for further income from self-employment and employer-coordinated freelance jobs (the so-called "co.co.co".);
- The Central Pension Register ("Casellario Centrale dei Pensionati"), managed by Inps[7].

The early phase of integration consists in piecing together, on the basis of the aforementioned records, income variables compatible with those emerging from the IT-SILC survey questionnaires (Istat, 2009a: 37-77). Istat then breaks down each item contained in fiscal records, with the purpose of identifying, for each sort of income (from employment, self-employment, transfers or grants), the net amount, obtained as the difference between the gross amount, on the one hand, and social contributions (or solidarity contributions) plus tax deductions on the other. It is interesting to observe that IT-SILC, like SHIW, detects solely income after tax. The availability of information with merely administrative origin on gross income, social contributions and taxes paid represents an important novelty with respect to the former model version based on the Bank of Italy sample.

Once net income variables have been pieced together in the administrative records, Istat proceeds to "reconcile" (Istat, 2009a: 79) the income items of different sources and to resolve any possible inconsistency with sample data. With respect to the integration of the various administrative sources, the main problem consists in harmonizing the items coming from the Central Pension Register and from fiscal archives, about which the Istat reference handbook could be consulted (2009a: 79-91). The inconsistencies between administrative and sample data can instead be of two types, and have been resolved with different procedures

[7] Istat tries to find a match for every individual belonging to the theoretical sample. In 2006 it was possible to identify the tax code in 97.4% of the cases (Istat, 2009a: 21).

according to the income type[8]. In the former, one of the two sources depicts the individual as being a receiver of a given type of income, while the other does not report such receipts. With reference to employment income, Istat takes into account the whole income structure and other signals, especially the occupational condition during the various months of the year. If the income item results in sampling records but not in fiscal ones, the best is done to connect it with items certificated by the fiscal records. If the income type results only in fiscal records, this fiscal entry is included in the absence of other revenues, otherwise Istat tries to link it and assign it to income from self-employment, if the latter is reported in the questionnaire. Income from self-employment also goes through an integration procedure which takes into account the whole income structure and information about the occupational condition. The procedure is, moreover, aimed at resolving the inconsistencies in splitting up the item between self-employment income in the strict sense and income from "para-subordinate" employment[9]. Lastly, those who emerge as earners only in fiscal or administrative records are considered as pensioners (Istat, 2009a: 98-99)[10].

The second kind of inconsistency pertains to the level of income reported by individuals in the two different sources. With reference to employment income, the "true" value is assumed to be the one reported in fiscal records, except where the latter turns out to be lower than the sample value[11].

Referring to self-employment income, Istat assumes, as is likely to be the case, that both the sample and the administrative records underestimate the true value, because of interviewees *under-reporting* on the one hand and, on the other, because of tax evasion and tax avoidance. Consequently, the rule followed consists in considering the maximum between the two sources.

---

[8] For some income types no procedure was adopted. Severance pay originates only from CUD/770: this is considered to be a more reliable source for the sample individuals for which it was possible to find a match, while the sample value remained unchanged for the others. The same logic is applied to family allowances, since part of recipients are not detected by the Cud/770 source (Istat, 2009a: 79).

[9] As we shall call the working category in Italy which lies on the borderline between employees and self-employed.

[10] In fact, as can be noticed in tab. 1, for some pension types the number of earners decreases as a consequence of the integration process. Where the receipt of the pension results only from sample records, the Institute adopted both standard procedures of data control and adjustment and some compatibility checks with maximum and minimum values provided by law and with the minimum retirement age (Istat, 2009a: 99). For instance, if an individual declares receipt of welfare benefits, but is below age 65 (minimum age for welfare benefits), Istat proceeds with the correction.

[11] In this case Istat resorts to different procedures, which consider other information, such as the evidence of tax-free revenues and the reliability of the answers provided, on the basis of the interviewer's judgment (Istat, 2009a: 99).

Finally, with reference to pension, the value given by the administrative records is always taken to be more reliable.

For the purposes of this study, it is important to observe how the integration process with administrative data is particularly useful to "correct" the information about pensions (especially disability pensions), concerning which the individual's inclination not to report their receipt and/or to report the type erroneously during the interview is well known (see Table 1, taken from Istat (2008: 99))[12].

Table 1
Results of pension correction procedure

Reference database	Recipients resident in Italy (absolute values in thousands)				
	Old age and retirement	Survival	Disability	Welfare benefits	Total
Data EU-SILC 2004	10,672	3,375	2,210	1,021	14,668
Data EU-SILC 2004 integrated with fiscal data and the Pension Register	10,573	4,194	4,148	749	15,861
Pension Register (b)	10,471	4,602	4,523	743	15,726

Source: table 4.9 in Istat (2008: 99).
Note: recipients of the so-called "attendance allowances" only (benefits provided to handicapped individuals with 100% and needing continuous care in daily living and walking) have been excluded from the category of disability benefits. In line with the IT-SILC sample, individuals aged 14 years or less have been excluded from the Pension Register.

The process of integration with administrative data enables microdata inclusive of some information of exclusively administrative origin to be released, particularly useful in performing simulations on the tax/benefit system. This kind of information is, for instance, the total amount of personal income taxes, employee's (also for self-employed workers) and employer's contributions. Such information allows the exact gross income to be reconstructed.

An additional advantage of IT-SILC is the presence of analytical entries referring to pensions, reported with greater detail than in SHIW[13]. In particular, in the Istat survey it is

[12] In addition to this correction process, Istat uses more standard procedures to apportion missing data and to check for the eventual presence of outliers. The procedure is made transparent by the presence of flag variables for European income items, which report the imputation factor for each observation that has been subjected to correction.
[13] The description of integrative variables is contained in a document attached to the microdata released.

possible to distinguish between occupational pensions, survival pensions, welfare benefits, supplementary and disability pensions[14].

For each entry the monthly amount and the number of monthly payments received are available. For the disability category it is also possible to distinguish if the interviewee receives disability support benefits, contributory disability benefits (paid by Inps or Inpdap), revenues from accident insurance (paid by Inail or Ipsema) or war pensions. The monthly amount of such revenues and the number of monthly payments received are, however, provided in the form of aggregate value. The special disability benefit, the so-called "attendance allowance", is included in a comprehensive entry called "disability allowances", but it is possible to disentangle from it the number of recipients and the monthly amount of attendance allowances, since they are paid in a fixed amount that does not depend on the interviewee's economic condition.

The third reason in support of IT-SILC is its sampling dimension. The cross-section used in this study (2007) covers 20,982 households and 52,772 individuals (of which 45,133 are at least 15 years old at the end of period t-1). The sampling dimension is such as to enable precise estimates also at regional level (Istat, 2008: 32), which instead are not feasible with a SHIW cross-section that, in 2006, contains information regarding 7,768 households (19,551 individuals). In spite of its countless advantages, IT-SILC is not free from limitations. In the first place, not much detailed information is available about the level of wealth owned by interviewed households. The poor data collected relating to personal estate are gathered much less systematically than in SHIW[15].

Secondly, though it has been conducted for four years, IT-SILC does not have the SHIW tradition. The empirical research pursued for decades on SHIW has indeed enabled its limits and distortions to be perceived. On this matter see, for instance, Cannari and D'Alessio (1992) and Brandolini (1999).

---

[14] At European level, making use of the ESSPROS classification renders it very difficult to work on the Italian situation. For instance, welfare benefits and the special disability benefits, called "attendance allowances" in the Italian system, paid to persons aged over 65 years end up in the entry py110n, presented as "old-age benefits". The integrative Italian variables enable this difficulty to be overcome.

[15] Note that individuals are required to report if they have some savings, how they invest them and their total amount. Record is also kept of the individual payments to supplementary pension funds or if the family obtained a loan (specifying the initial amount, the total duration, the interest rate and the installment paid in t-1). Finally, the amount of ICI paid enables the property cadastral income to be reconstructed, making some general assumptions on the nature of the property starting from available information. The research unit maintains feasible a *statistical matching* with the SHIW survey in order to compensate for the lack of some entries regarding households' wealth in IT-SILC.

IT-SILC microdata are, moreover, subject to top-coding of interviewees at the age of 80 years. The preservation of statistical privacy has a particularly restrictive effect on the sort of analysis that we intend to perform, as it does not enable the socio-economic conditions of individuals aged over 80 years to be reliably focused.

The imputation of the age for individuals aged 80 and over was carried out by exploiting the information contained in demo.istat and using the procedure described in the next paragraph.

Finally, sampling data are collected through interviews to individuals residing in private dwellings located on the Italian territory. Like SHIW, IT-SILC fails to sample some groups of individuals in the population, such as those residing in nursing homes ("case di cura"). The result is that the analysis performed by means of sample data refers to the population living in private dwellings, completely omitting the socio-economic condition of elderly people dwelling in residential facilities (defined "strutture residenziali" in Italy). These individuals, apart from representing the greatest part of the population not residing with their family, constitute an important share of recipients of pensions and disability benefits. The next paragraph explains in detail the procedure used to circumvent this limit, together with the set of criteria used to build the initial population in the dynamic model.

3. The procedure followed to build the initial CAPP_DYN population

This paragraph describes the procedure adopted to build the initial CAPP_DYN population using the IT-SILC 2007 cross-section. Many of the variables contained in IT-SILC required recoding, while others were constructed on the basis of some assumptions. In this paragraph we present some of the most important assumptions, among which those relating to:

1. The choice of the reference period between t and t-1 and the reconstruction of the professional condition in period t-1;
2. The top-coding on the interviewee's age;
3. The creation of a sample representative of individuals residing in nursing homes;
4. The definition of the variable "years of contribution" and the rule used to detect the contribution spell;
5. The procedure followed to identify the amount of each pension and disability

benefit received, and the relative checks with respect to administrative records.

Many of the socio-economic variables included in the survey did not require intervention by the research unit and were directly used. The incidence of missing values for these variables is absolutely negligible in IT-SILC[16].

If not otherwise specified, data processing in the rest of the chapter is performed without resorting to sampling weights (released together with microdata). After a careful examination, the research unit decided not to use sampling weights in the construction of the initial CAPP_DYN population. While sampling weights improve the representativeness of the sample as per age range, gender and residence region (Istat, 2008: 38), they do involve some critical issues. Firstly, as argued elsewhere (see the previous report, Ministry of Social Affairs, 2008), sampling weights cannot be used *tout court* in a dynamic micro-simulation model. The procedure used in previous versions of the model (based on SHIW data) consisted in building the initial population by multiplying each individual by a number equal to their sampling weight. This procedure is unsuitable from the computational point of view with IT-SILC, as the application of sampling weights would reproduce the entire Italian population. This procedure would imply expanding the sample up to a number of observations exactly equal to the Italian resident population in 2007. Even if we decide to divide sampling weights by the smallest observed value in such a way as to reduce to a minimum the scale by standardization, 10% of individuals are characterized by a standardized weight of 27 and with a maximum equal to 156. This procedure would require building a dataset of about 750,000 records for the base year, rendering future scenarios extremely difficult to process with the computers at our disposal. In addition, some interviewees (especially the elderly living on their own), owing to the objective difficulties in getting information on their socio-economic situation from them, are assigned a particularly high weight, which is a function of known totals relating to demographic variables, such as gender, age or residence area. Such a procedure does not guarantee the representativeness of the sample with respect to other variables of interest, which are not covered in this post-stratification process. In consequence,

[16] It is important to specify that in IT-SILC microdata variables can only be missing when no imputation process has been possible (source: document "Eu-Silc User Database Description - Version 2007-1 from 01-02-2009" (page 40) attached to microdata released).For some variables, mainly income ones, are provided *flag* variables which report the imputation factor in case of correction. A flag assumes a positive value when a variable is *filled*, -1 if the variable instead is missing. Much of the information used by our research unity is *non missing*, and this is explained by the fact that Istat extended the imputation procedure also to qualitative variables (Istat, 2008: 15). For some variables, like the state of health, it is furthermore included the category "I don't know" or "he/she refuses to answer".

assigning a large weight to a small share of the sampling population may lead to significant distortions when other interviewees' characteristics, aside from age, gender or residence area, are considered.

3.1 Compatibility between professional condition and reported income

As already mentioned, the IT-SILC survey, carried out at the end of the reference year for living conditions (period t), measures the level of income in the previous year (period t-1). Hence the time rift between socio-demographic and income variables represents a problem.

In line with the main Tax-Benefit model built at European level on the basis of EU-SILC (EUROMOD), we decided to pick period t-1 (2006) as reference period. The primary reason is the presence in IT-SILC of detailed information on the professional condition in each month of period t-1, whereas it would be necessary to introduce very strict assumptions to enable income variables to be consistent with the condition in period t[17].

With respect to the main demographic variables (age and sex) no important problems arise, whereas educational qualification, civil status, residence area and composition of the family unit could change in the period under consideration between t-1 and t. Regarding educational qualification and civil status, IT-SILC reports the year the individual achieved the current status and which was the previous one. Consequently, we merely modify the status of those who obtained a new qualification or changed their marital status in 2007. Referring to the residence area and the composition of the family unit, we use, instead, data referring to the time of the interview, in the absence of further information. Table 2 reports time definitions.

The professional condition, instead, shows greater complexity. For period *t*-1 we have available the professional condition in each month of the year, broken down into employed *full-time*, employed *part-time*, self-employed *full-time*, self-employed *part-time*, unemployed, retired from working activity, student, outside the labour force, in military service. To decide the professional condition we make use of a simple rule: the individuals are defined as

[17] In a early stage of the research, we also evaluated the possibility of favoring the information collected through the question relating to the professional condition at time *t*, trying to solve the time rift only for individuals in transition from a state of inactivity to a state of employment, and *vice versa*, from period *t-1* to period *t*. however this choice requires a greater number of assumptions and thus turns out more complex and less transparent. Moreover, in the comparison the annual mean resulting from the quarterly Labor Force Survey, the current strategy emerges as superior.

"employed" if they worked at least two months during the year. Although the rule does not seem very strict, it has been chosen for two reasons. First of all, not many individuals have worked for less than six months. Secondly, as can be observed from Table 20 and from Table 24, this choice involves a good degree of fit with respect to the employment rate estimated according to the quarterly Labour Force Survey (averages for 2006; see Istat, 2007).

Table 2

Definition of the main demographic variables on the basis of period t-1 (2006)

Variable	Information used
Age	Age at the end of the income reference period (2006)
Residence area	Residence area in 2007.
Educational qualification	Highest qualification if obtained before 2007, previous qualification if obtained in 2007.
Occupational status	Individuals are defined "employed" if they worked at least 2 months in 2006. The other variables are defined according to the foremost condition in the months of 2006, controlling for the receipt of the relevant revenues.
Occupational sector (Public, Private)	Variable defined only if the individual is employed in 2006.It is equal to the occupation sector in 2007 if the individual is employed also in 2007, while in the other cases the sector of the last company for which he/she worked is used.
Part-time job	If employed, the individual has a part-time job if in the majority of months of employment he/she worked part-time.

Where the individual is defined as "employed", the choice between employee *full* or *part-time*, or between self-employed worker *full* or *part-time*, depends upon the prevalent condition during the period in which he/she worked. To assure consistency with respect to reported income, an individual is not considered to be an employee if he/she does not receive the respective income, nor to be a self-employed worker if the respective gains or losses do not appear. Workers with employer-coordinated freelance contracts (co.co.co) are identified as workers whose self-employment income comes predominantly from such type of contracts. Finally, the sector (public or private) is defined according to the condition in 2007 for those who were employed in both periods. For the rest of the individuals, we use the NACE code of the last company for which they worked. In this case, the only way to identify public employees is to define as such those whose code indicates "*Public administration and defence, compulsory social security*", and every individual who has been employed in the "*Education*" sector.

For those who are defined as *non-employed*, the professional condition is broken down into unemployed, student, pensioner and outside the labour force, according to the prevailing condition in the months the individual did not work. Pensioners are further split up into retirement/contributory pensioners and non-contributory pensioners, where the former are identified as those who report receiving an old age pension.

The presence of individuals who report getting both employment income and old age pensions could be a problem[18] in building the dynamic model. For 1,927 individuals report receiving (in period *t*-1) both earned income and a contributory pension. This group represents 4,32% of the sample population over age 15. From Table 3 it can be observed that 87% of these individuals received 12 or more monthly pensions in period *t*-1; 434 individuals reported receiving (in addition to contributory pensions) at least a monthly salary as employee, while 484 individuals worked at least one month as self-employed workers. It is therefore necessary to know the number of individuals in transition toward retirement (less than 12 months from pensions) and the number of those who continue, even if retired, to work as self-employed or employees. The table suggests, moreover, the presence of a group of individuals with positive earned income even though declaring not to have worked in any month of period *t*-1.

Table 3

Monthly pensions received, monthly pays for employees and months of self-employment in period *t*-1 for individuals receiving both earned income and old age/retirement pensions

Number of monthly contributory pensions			Number of monthly pays for employees			Number of months as self-employed worker		
	Frequency	Percentage		Frequency	Percentage		Frequency	Percentage
1	14	0.7%	0	1,493	77.5%	0	1,440	74.7%
2	13	0.7%	1	58	3.0%	1	2	0.1%
3	33	1.7%	2	23	1.2%	2	3	0.2%
4	47	2.4%	3	39	2.0%	3	1	0.1%
5	10	0.5%	4	22	1.1%	4	2	0.1%
6	13	0.7%	5	12	0.6%	5	1	0.1%
7	41	2.1%	6	27	1.4%	6	10	0.5%
8	11	0.6%	7	17	0.9%	7	4	0.2%
9	11	0.6%	8	44	2.3%	8	2	0.1%
10	45	2.3%	9	19	1.0%	9	3	0.2%

[18] It is important to specify which are the original variables used to this purpose. The general definition of "pension income" is not always clear, since the structure defined by the ESSPROSS classification for European variables and the structure of the Italian pension system overlap. In this case, by "earned income" we refer to the variable "ylav" in the Italian dataset, while by "contributory pensioner" we refer to those who answered positively to question 13.11 of the individual questionnaire: "In 2006, have you received one or more contributory pensions, i.e. old age or retirement pension?" (variable "plav" equal to 1).

11	7	0.4%	10	10	0.5%	10	4	0.2%
12	19	1.0%	11	12	0.6%	11	2	0.1%
13	1,663	86.3%	12	151	7.8%	12	453	23.5%
Total	1,927	100.0%	Total	1.927	100.0%	Total	1,927	100.0%

Sources: IT-SILC data processing.

Table 4

Age of individuals receiving both earned income and old age/retirement pensions

Age range	Frequency	Percentage	Cumulative
Less than 50	8	0.4%	0.4%
50-54	28	1.5%	1.9%
55-59	405	21.0%	22.9%
60-64	586	30.4%	53.3%
65-69	475	24.7%	77.9%
70 or more	425	22.1%	100.0%
Total	1,927	100.0%	

Sources: IT-SILC data processing.

On the basis of information available in IT-SILC we can build seven groups of individuals, listed in

Table ***5***, making clearer the economic and occupational condition of individuals considered.

Table 5

Breakdown of the group of recipients of earned income and old age/retirement pensions

Description	Number	Percentage
He/she received 12 or 13 monthly pensions and at least one monthly pay as employee, but did not receive income from self-employment	233	12.1%
He/she received 12 or 13 monthly pensions and worked at least one month as self-employed worker, but did not receive income from employment	426	22.1%
He/she received 12 or 13 monthly pensions and self-employment income, but did not work one month as self-employed worker, nor received income from employment	470	24.4%
He/she received 12 or 13 monthly pensions and received both income from employment and self-employment	29	1.5%
He/she received 12 or 13 monthly pensions, did not receive income from employment or self-employment, but got arrears	524	27.2%
He/she received less than 12 monthly pensions and received at least one monthly pay as employee or self-employment income	240	12.4%
He/she received less than 12 monthly pensions and received arrears or severance pays	5	0.3%
Total	1,927	100%

Note: all characteristics refer to period *t*-1. The third category includes those who receive business income or royalties but did not actually work.

Given the CAPP_DYN characteristics, it was necessary to manoeuvre individuals in transition into a well-defined condition. In other words, since the minimum time unit of the dynamic simulation is the year, we need to move every retirement choice into this time horizon. Moreover, it was decided not to admit the pensioner/worker condition. Even though this condition is not only admitted by the pension system discipline but actually encouraged by our legislator, it is well-known that the effectiveness of the incentive measures turns out to be very poor. For simplicity's sake we therefore decided to follow the following rules (see also

Table ***6***):

1. For individuals whose earned income exceeds the amount of the old age/retirement pension, we re-code the latter to zero. The operation is carried out only if the person concerned declares having paid less than 40 years of contributions, and if under age 65 if male, or under 60 if female;

2. For the remaining individuals whose earned income is lower than their old age/retirement pension, we re-code earned income to zero. In this case we modify the revenue of 1,687 individuals, of which 52% below age 66 years (mean age: 66 years).

Table 6
Characteristics of "treated" individuals

Description	Elimination of pension income		Elimination of earned income	
	Number	Distribution	Number	Distribution
Received 12 or 13 monthly pensions and at least one monthly pay as employee, but did not receive income from self-employment	35	14.6%	198	11.7%
Received 12 or 13 monthly pensions and worked at least a month as self-employed worker, but did not receive income from employment	84	35.0%	342	20.3%
Received 12 or 13 monthly pensions and self-employment income, but did not work one month as self-employed worker, nor received income from employment	20	8.3%	450	26.7%
Received 12 or 13 monthly pensions and received both income from employment and self-employment	4	1.7%	25	1.5%
Received 12 or 13 monthly pensions, did not receive income from employment or self-employment, but got arrears	0	0.0%	524	31.1%
Received less than 12 monthly pensions and received at least one monthly pay as employee or self-employment income	97	40.4%	143	8.5%

Received less than 12 monthly pensions and received arrears or severance pays	0	0.0%	5	0.3%
Total	240	100%	1.687	100%

Note: total number of observations 1,927.
Sources: IT-SILC data processing.

3.2 The "top-coding" of age

Standard SILC microdata are subject to *top-coding* at the age of 80 years old, in order to preserve the statistical secret. Given the importance of this variable in the dynamic simulation, the research unit chose to assign the age according to the frequency distributions by age and gender observed in the official statistics (Table 7, from *demo.istat* website).

Table 7
Resident population (80 years old and over) on 1st January 2007 by age and sex-ITALY

			Men +	*Relative frequency*	
Age	Total men	Total women	Women	*Men*	*Women*
80	158,738	253,987	412,725	*0.1514*	*0.1215*
81	144,940	241,889	386,829	*0.1382*	*0.1157*
82	130,849	228,340	359,189	*0.1248*	*0.1092*
83	118,343	216,167	334,510	*0.1128*	*0.1034*
84	102,541	200,200	302,741	*0.0978*	*0.0957*
85	90,171	183,556	273,727	*0.0860*	*0.0878*
86	77,868	165,748	243,616	*0.0742*	*0.0793*
87	44,137	98,275	142,412	*0.0421*	*0.0470*
88	28,551	67,088	95,639	*0.0272*	*0.0321*
89	25,385	61,667	87,052	*0.0242*	*0.0295*
90	26,140	66,175	92,315	*0.0249*	*0.0316*
91	26,607	70,913	97,520	*0.0254*	*0.0339*
92	21,668	61,301	82,969	*0.0207*	*0.0293*
93	16,861	49,752	66,613	*0.0161*	*0.0238*
94	12,442	38,816	51,258	*0.0119*	*0.0186*
95	8,273	27,604	35,877	*0.0079*	*0.0132*
96	5,822	20,663	26,485	*0.0056*	*0.0099*
97	3,796	13,859	17,655	*0.0036*	*0.0066*
98	2,253	9,433	11,686	*0.0021*	*0.0045*
99	1,321	6,038	7,359	*0.0013*	*0.0029*
100 and over	2,025	9,472	11,497	*0.0019*	*0.0045*
TOTAL	**1,048,731**	**2,090,943**	**3,139,674**	**1.0000**	**1.0000**

Sources: demo.istat.it *(access: November 2010)*.

In particular, the relative frequencies shown in the last two columns of the following table were used to assign to each IT-SILC interviewee aged 80 years and over the probability of being 80, 81, 82 and so on, according to the interviewee gender. The probability thus computed was then compared to a random number drawn for each individual from a uniform distribution in the interval [0,1]: if the random value is lower than the aforementioned probability (conditional and cumulative), the respective age is assigned to the individual.

3.3 Individuals living in residential facilities

Like any other sample survey, IT-SILC reveals information on households and individuals living in the household at the time of the interview and for a certain time span before. Individuals living in nursing institutions and other facilities remain outside the survey sampling scheme; but, for our purposes, it is necessary that the initial CAPP_DYN population be representative of the whole population residing on the national territory. Residential facilities in Italy identify those institutions (public or private, operating or not within the national health service) offering medical, recovery and welfare services towards persons affected by certain health problems (even if not serious) and elderly individuals, together with all other services connected with the individual stay in the structure, like refectory or recreational services.

There follows an analytical description of the procedure adopted to include in the initial CAPP_DYN population individuals representative of guests in residential facilities. The latter, besides representing the most important slice of the population residing outside the household, constitute a not negligible quota of recipients of pensions and disability benefits.

It is well-known that guests of residential facilities have special characteristics in terms of age, health conditions, gender, residence area and economic conditions. From the Istat survey "L'Assistenza residenziale e socio-assistenziale in Italia" (Table 8) it emerges that, at December 31, 2006, the elderly individuals residing in such facilities numbered 230,468 (54,262 men; 176,205 women) and represented about 3% of the population aged 65 and over. Of these, 73% resided in the North (the 15.33% in the Centre). 70.39% were non-self-sufficient (66.11% of males; 71.71% of females). Finally, nearly 70% of them were over age 80.

Table 8

Elderly persons in residential facilities by gender and age range

AGE RANGE	Men	Women	Total
	ABSOLUTE VALUES		
65-74	13,746	17,755	31,501
75-79	12,083	27,078	39,161
80 and over	28,434	131,372	159,805
TOTAL	**54,263**	**176,205**	**230,467**
	PERCENTAGE VALUES		
65-74	25.33	10.08	13.67
75-79	22.27	15.37	16.99
80 and over	52.40	74.56	69.34
TOTAL	**100.00**	**100.00**	**100.00**

Sources: Table 4.4 – Survey "L'Assistenza residenziale e socio-assistenziale in Italia", 31/12/2006. Available at http://www.istat.it/dati/dataset/20100211_00/ *(last access: 26/11/2010)*

The recent "Report on non-self-sufficiency" ("Rapporto sulla non autosufficienza") released on July 2010 and published by the Ministry of Welfare reports that:

- 3% of elderly individuals aged 65 and over are guests of residential facilities (table 9, p.24);
- 8.7% of non-self-sufficient elderly individuals aged 65 and over are guests of residential facilities (table 11, p.34);
- In Italy 265,326 sleeping accommodations are available in residential facilities for elderly persons (table 15, p.47);
- The elderly individuals who live in residential facilities total 345,093, including also non-self-sufficiency facilities (table 16, p.48)[19].

The above-mentioned information was used to generate artificially a sample of representative individuals living in residential facilities[20]. We assume a number of 265,000 guests in residential facilities, a number consistent with the 230,467 individuals mentioned in the Istat report (about 3% of the population aged 65 and over). The sample totals 335 individuals, which represents exactly 3% of the 65+ population in IT-SILC. The information available does not, however, enable the economic characteristics of the population of study to be inferred. To compensate for the unavailability of such data, we make use of the information coming from the last survey "L'Assistenza residenziale e socio-assistenziale in

[19] http://www.lavoro.gov.it/NR/rdonlyres/9B939247-1A95-468A-9A54-6E58BE0DD85C/0/RapportosullanonautosufficienzainItalia27072010.pdf

Italia" (2006), implementing a two-stage procedure. In a first stage we select a group of IT-SILC interviewees older than 65 years, non-immigrant, living alone. We calibrate the sample in order for it to contain 70.39% of non-self-sufficient elderly individuals (with a serious disability level and receiving the so-called "attendance allowances") and to comply with the frequency distributions by gender and age range, with 73% of the sample living in the North of the country. The imputation of income components is carried out at the second stage, where we assume that the economic status of elderly individuals in residential facilities be identical with those of a similar individual (with respect to gender, age range, residence area and disability/non-self-sufficiency level) randomly drawn from all the IT-SILC interviewees.

3.4 Years of contribution

The Italian dataset contains the variable "*acontrib*", reporting the years of contribution to the social security system until the date of the interview (period *t*)[21]. This information is collected for all employed individuals or those who have been employed in the past. This variable enables us not only to compute the future pension, but also to determine the belonging pension system and the respective computation formula for future pensions (earning-related, contribution-related or mixed)

To this end, we must trace the years of contribution up to 1995[22]. Ideally, it would require knowing the slice of contributions paid before that date, back to rebuilding the entire working career of individuals. The breakdown of contributory spells during the individual's life, not available in the original microdata, was built according to the interviewee's (adjusted) professional condition and exploiting the (self-declared) information on the age when the individual started working regularly (variable *pl190*):

1. For non-employed, neither at the time of the interview nor in period *t-1*, we assume that the contribution spell is concentrated in the initial period of working life. This enables us to avoid that individuals with few contributions paid and who started working at a

[20] In the absence of elementary microdata, the available information on the characteristics of guests of residential facilities does not enable us to build multivariate tables (jointly) according to gender, age class, residence area and disability/self-sufficiency level.

[21] The variable *acontrib* refers to question 8.3 of the individual questionnaire: "How many years of contributions have been paid for your pension?". Some interviewees may answer "I don't know" (acontrib=99), but in the microdata released to the public this does not occur for any individual.

[22] The choice of the year is not casual: starting from 1995 public pensions for newly employed individuals will be computed through the "contributions-based" system.

date sufficiently far from 1995 end up automatically in the "Notional Defined Contribution" system introduced in Italy after 1995.

2. For individuals who are employed at the time of the interview we assume that the contribution period is concentrated at the end of the working life[23].

For both groups we assume that there is no interruption in the contribution spell, since we have no information about possible unemployment/inactivity periods during the individual's working life. Moreover, we implicitly assume the absence of measurement errors with respect to variables "*acontrib*" and "*pl190*".

Table ***9*** and Table 10 highlight the relationship between number of years of contribution, age and the year the individual started working regularly. In the first table, note a non-negligible presence of individuals older than 65 years with a relatively limited contribution period, shorter than 20 years. In addition, Table 10 highlights how a significant slice of individuals, although they started working before 1955, are characterized by a very narrow contribution period with respect to the time horizon existing between this date and the time of the interview (2007).

Table 9

Distribution of the number of years of contribution by age

Age range	Years of contribution						N. of observations
	Less than 10	10-19	20-29	30-39	40 or more	Total	
Less than 30	93%	7%	0%	0%	0%	100%	4,787
30-49	27%	46%	24%	3%	0%	100%	14,367
50-64	4%	13%	29%	47%	7%	100%	9,328
65-74	2%	11%	22%	48%	17%	100%	4,986
75 and over	2%	10%	21%	41%	27%	100%	3,837
Total	24%	24%	22%	23%	7%	100%	37,215

Sources: data processing on the initial sample. Sampling weights are not used (see the beginning of par. 4).

Table 10

Distribution of the number of years of contribution by the year individuals started working regularly

Starting year of working activity	Years of contribution						N. of observations
	Less than 10	10-19	20-29	30-39	40 or more	Total	
Prima del 1955	10%	20%	43%	26%	100%	10%	5,735
1955-1974	12%	27%	49%	9%	100%	12%	11,591

[23] The relevant variable is surveyed for all individuals who work or have worked at least once in their life. Any missing value is observed for sample individuals belonging to this category.

1975-1984	36%	47%	8%	0%	100%	36%	6,831
1985-1994	57%	9%	0%	0%	100%	57%	6,810
1995-2004	13%	0%	0%	0%	100%	13%	5,589
From 2005 on	2%	0%	0%	0%	100%	2%	663
Total	24%	22%	23%	7%	100%	24%	37,129

Sources: data processing on the initial sample. Sampling weights are not used (see the beginning of par. 4).

The same consideration holds for those who started working prior to 1975 and, in a different and limited way, also for the subsequent age groups. In line with our operating proposal, these observations set some limits to the common assumption among dynamic micro simulation models that the contribution spells are concentrated in the final part of the individual's working life.

3.5 Disability benefits

A detailed reconstruction of disability benefits can be an important prerequisite for simulating expenditures relating to money allowances toward population of non-self-sufficient individuals and investigating the relationship between such expenditure and socio-economic conditions of recipients. In this paragraph we concentrate on the distribution of these benefits in the IT-SILC sample and on the procedures used to solve some inconsistencies. The analysis here proposed refers to the non-institutionalized population. The same procedures apply to the representative sample of inmates of residential facilities.

The individual Italian questionnaire of the IT-SILC survey (question 13.4) asks different questions about the perception of the four most important benefits:

1. Disability pension or allowance paid to public or private employees or self-employed workers;
2. Pension for accident at work or occupational disease (INAIL, IPSEMA);
3. Benefits to civil invalid, civil unsighted and civil hearing-impaired individuals (called *disability support benefits* in this report, to distinguish them from disability contributory pensions reported in point 1);
4. War pensions (excluding those paid to orphans and widows/widowers)[24]

The number of monthly payments received and the mean monthly amount relating to the four categories is collected in a single variable, even if the individual receives more than

one type of benefit. The next question (13.7 in the individual questionnaire) asks if the amount declared by the interviewee includes possible attendance allowances or other disability related benefits (like taxi vouchers). In the Italian cross-section all answers to this question have been re-classified at the time of integration with administrative data, reporting a negative answer for all individuals and keeping the attendance allowances separate[25]. The monthly amount of attendance allowances is reported through the variable "*pacc_e*", the number of monthly payments through the variable "*pmacc*". According to the structure of the relevant question, these two variables also concern other general benefits related to disability, precisely like taxi vouchers. Analyzing the amount reported in Table 11 it is possible, however, to indirectly retrace the nature of the reported benefits, thus justifying the choice to identify as recipients of attendance allowances every individual with a variable "*pacc_e*" greater than zero[26]. Given the complexity of the system of monetary disability allowances, it is necessary to specify that in the administrative statistics, contained in the yearly Istat-Inps survey "Statistiche della previdenza e assistenza sociale" (hereinafter Istat-Inps survey), attendance allowances include only the allowances paid to civil invalids, while those paid out in relation to (contributions-related) disability pensions, to accidents at work or war pensions are included in the amount of the relevant pensions (since it is impossible to receive them separately)[27]. The allowances relating to civil disability, to which people frequently but improperly refer by the term "attendance allowances", also include allowances toward minors, communication allowances, allowances to totally blind persons, allowances to 20% blind persons and allowances to workers affected by sickle cell anaemia and thalassemia major.

Table 11

Amount of the variable "pacc_e" in the sample

Amount (euro)	Type of allowance	Number of

[24] It can be observed how, in table 1 of this chapter, the correction is particularly relevant for disability benefits. In fact, it is likely that individuals mix up different benefits or do not report their collection.

[25] Such information has been provided by Istat, in reply to our request for clarification.

[26] With respect to the amount referring to different annual payments, we refer to values reported on the internet website: http://www.handylex.org/gun/importi2006.shtml (last access 29th may, 2010), by the Italian Union Lotta alla Distrofia Muscolare (Fight against Muscolar Distrophy). Of 51 non-identified cases, 49 are characterized by a monthly amount greater than the special allowance in favor of 20% hearing-impairment, and 24 receive a monthly payment greater than the attendance allowances to totally invalid individuals. It is thus likely that the majority of these individuals combine an allowance like those reported in the table with other monetary allowances paid at local authority level.

[27] These indications were provided by Istat on 3rd April 2009, in reply to our request for clarification.

		observations
165	Special allowance to 20% blind persons, year 2006	52 (4.2%)
690	Attendance allowances to totally blind civil persons, year 2006	33 (2.7%)
710	Attendance allowances to totally blind persons, year 2007	1 (0.0%)
227	Communication allowances to deaf-mute individuals, year 2006	27 (2.2%)
234	Attendance allowances to minors, year 2005	4 (0.3%)
238	Attendance allowances to minors, year 2006	16 (1.3%)
451	Attendance allowances to totally disabled individuals, year 2006	1,027 (82.8%)
444	Attendance allowances to totally disabled individuals, year 2005	7 (0.6%)
450	Attendance allowances, year 2006 (with error rounding)	17 (1.4%)
1140	Combination between attendance allowances for totally disabled and totally blind individuals	6 (0.5%)
-	Unidentified amount	51 (4.1%)
	Total	*1,241 (100%)*

Note: amounts relating to different years may be justified by the presence of arrears, where they are prior to 2006, or by the presence of mistakes that it has not been possible to correct through integration with the administrative data.

However, the analysis of disability support benefits, reported in Table 12, highlights a possible mistake in re-classifying as negative all answers to the question "*does the amount you have just reported also include attendance allowances or other disability-related benefits (like taxi vouchers)?*". In particular, it seems that Istat re-classifies as negative answers also the variable that reports the receipt of disability support benefits for those who receive only attendance allowances [28]. From Table 12 it can be observed that the 65 and over age band contains a proportion of recipients of "pension+benefit" much higher than that reported in Istat (2009b), referring to the population aged 15 and over at 31/12/2006[29]. This difference may be the result of a mistake, since disability support benefits change into non-contributory pensions at age 65 years, excluding blind individuals[30]. It must be pointed out that errors may emerge only where the individuals declare receipt of more than one type of disability benefit,

[28] Recall that Istat separates attendance allowances where interviewees declare having included them in the monthly amount of disability benefits.

[29] In all comparisons between the Istat-Inps survey statistics and the sample statistics we refer to the population aged 15 and over at the end of period *t-1*. In IT-SILC income variables are indeed surveyed only for individuals at least 15 at the end of period *t-1*. We used tables 6.1, 6.2 and 6.3 of the volume on pensions' recipients from the 2006 Istat-Inps survey, available on the website http://www.istat.it/dati/catalogo/20090618_01/ (last access 30th May, 2010).

[30] Only to report the totality of recipients according to IT-SILC, sampling weights are used. To obtain the totality of recipients, sampling weights built by Istat must be summed up for all individuals who come out to be recipients of disability support benefits in the sample (variable "pinv3" equal to 5). Remember that in the sample, variable "paccomp" ("*does the amount you have just reported also*

in addition to disability support allowances. If indeed they reported only the latter, they are necessarily in receipt of disability support benefits, not only of attendance allowances; otherwise the monthly payment would be equal to zero due to the re-classification process[31]. The last column of Table 12 reports the significance level of the difference between the two proportions, using the null hypothesis test that the estimated proportion in the sample is identical to the one computed according to the Istat-Inps survey[32]. The absence of asterisks means that the difference referring to the cell considered is not significant at 10% level, whereas if three asterisks are reported the difference is significant at 1% level. It can be observed that differences are not negligible, especially for upper age ranges.

Table 12

Share of recipients of disability support benefits by age range, before controls

AGE RANGE	Istat-Inps survey 31/12/2006			IT-SILC original sample, before controls			Significance level for the difference between the two proportions		
	Pension with att. allow.	Pension without att. allow.	Att. allow. only	Pension with att. allow.	Pension without att. allow.	Att. allow. only	Pension with att. allow.	Pension without att. allow.	Att. allow. only
15-34	0.64%	0.48%	0.15%	0.68%	0.61%	0.15%		**	
35-49	0.73%	1.16%	0.15%	0.65%	1.26%	0.14%			
50-64	1.10%	2.41%	0.39%	0.80%	2.16%	0.33%	***		
65-69	0.23%	0.00%	2.58%	0.86%	0.06%	1.47%	***	***	***
70-74	0.37%	0.00%	4.54%	1.16%	0.07%	3.10%	***	***	***
75-79	0.56%	0.01%	8.64%	2.61%	0.05%	4.81%	***	**	***
80 or more	1.35%	0.01%	26.63%	11.59%	0.31%	10.82%	***	***	***
Total	0.76%	0.97%	2.65%	1.46%	1.03%	1.29%	***		***

Sources: the share of recipients by age range in the Istat-Inps survey is obtained dividing the total number of recipients at the 31/12/2006 resulting from Istat (2009b) and the respective population at the 1/1/2007 resulting from demo.istat.it.

Note: the significance level: *** 1%; ** 5%; * 10%. The absence of asterisks means that the null hypothesis is rejected at the 10% level of significance. Sampling weights are not used (see the beginning of par. 4).

include attendance allowances or other disability-related benefits (like taxi vouchers)?") is equal to 2 ("No") for all individuals to whom the question was submitted.

[31] Moreover, there is no individual with pinv3=5 and no other disability pension with a value equal to the monthly amount of attendance allowances.

[32] In the z-score test we assume that the value obtained from Istat-Inps and demo.istat.it is the true value p_0 in the population (and not a random variable). The choice is justified by the non-sampling nature of the Istat-Inps survey, which makes direct use of administrative records. Defining as p the estimated value in the sample, the z statistic (Lindgren, 1993) is obtained as $(p-p_0)/ [(p_0 \text{ X } (1-p_0) / n)^{1/2}]$, where n represents the number of observations belonging to a cell. The statistic is asymptotically distributed as a standard normal variable. Every age cell has a number of observations greater than 2,100. The null hypothesis is that the two proportions are different, and a two tails test is used. Consequently, the p-value is equal to 2 X $(1 - \Phi(|z|)$, where Φ is the distribution function of a

To correct this inconsistency we adopt some working hypotheses that we also use to distinguish the total amount among different types of benefits in cases of individuals receiving more than one pension. We assume that the number of monthly payments (variable "*pminv*") is equal for all types of disability pensions received and we take into account that disability support benefits are paid in fixed amounts (equal to 238.07 euro per month in 2006)[33]. Cases and assumptions are listed in Table 13. Moreover, in building the initial population we added a sample of individuals living in residential facilities, as described in paragraph 3.3, that might offset the absence of a share of individuals with high probability of receiving such benefits.

Table 13
Assumptions made and operations carried out in different cases

Characteristics	Justification	N	Treatment
• 65years old and over • Two disability pensions, including disability support benefits • Attendance allowances (but not the benefit to totally blind or 20% blind civil individuals)	Disability support benefits change into non-contributory pensions for age 65 years and over, with the exception of totally blind civil individuals	325	Zeroing the disability support benefit
• 65years old and over • Two disability pensions, including disability support benefits • He/she does not receive attendance allowances, or he/she receives them but because totally blind or 20% blind civil individual • The monthly amount is at least 238 euro	Blind civil individuals, totally or partially, might continue to receive the disability benefit after age 65. They also receive the corresponding attendance allowances (independently of income)	26	The amount of the disability support benefit is equal to pminv*238 (257 if totally blind). The other pension is equal to the remaining part
• 65years old and over • Two disability pensions, including disability support benefits • He/she does not receive attendance allowances, or he/she receives them but because totally blind or 20% blind civil individual • The monthly amount is less 238 euro	Same, but on the basis of the monthly payment, they do not seem to receive pensions	3	Zeroing the disability support benefit
• 65years old and over	In the absence of information, we assume they	64	The amount of the disability support benefit is

standard normal variable, while |z| is the absolute value of the test statistic. If the p-value is lower than the significance level, the test leads to rejection of the null hypothesis.

[33] Remember that attendance allowances are kept apart from pensions.

• Two disability pensions, including disability support benefits • The monthly amount is at least 238 euro	receive the disability benefit		equal to pminv*238 (257 if totally blind). The other pension is equal to the remaining part
• 65years old and over • Two disability pensions, including disability support benefits • The monthly amount is less 238 euro	Same, but on the basis of the monthly payment, they do not seem to receive disability benefits	2	Zeroing the disability support benefit
• He/she receives both a contributory disability pension (or accident pension) and the war pension	We do not have no information to split the amount	17	Each pension is half of the whole amount
• He/she receives both a contributory disability pension and accident payment	From 1995 the two benefits are incompatible, but those which started earlier continue to be paid.	108	Each pension is half of the whole amount
• 65years old and over • Three disability pensions, including disability support benefits • Attendance allowances (but not the allowances in favour of totally blind or 20% blind civil individuals)	Disability support benefits change into non-contributory pensions for age 65 and over, with the exception of totally blind civil individuals.	27	Zeroing the disability support benefit. The other pensions are of equal amounts (half of the total).
• 65years old and over • Three disability pensions, including disability support benefits • He/she does not receive the attendance allowances, or he/she receives them but as totally blind or 20% blind civil individual • The monthly amount is at least 238 euro	Blind civil individuals, totally or partially, might continue to receive the disability benefits after age 65. They also receive the corresponding attendance allowances (independently of income).	3	Amount of the disability support benefit equal to pminv*238 (257 if totally blind civil individual). The other two pensions amount to the remaining part divided by two.
• Younger than 65 years • Three disability pensions, including disability support benefits • The monthly amount is at least 238 euro	In the absence of information, we assume they receive the disability benefit	4	Amount of the disability support benefit equal to pminv*238. The other two pensions amount to the remaining part divided by two.
• Three or more disability pensions and particular characteristics.		3	If the monthly amount is sufficient, we set the disability support benefit amount equal to pminv*238, while the amount of other pensions is the residual divided by the number. Otherwise, we reset the disability support benefit to zero.
Total cases		582	

Note: to identify blind civil persons, we exploited the fact that they receive an allowance with a specific value (see tab. 17). The monthly amount for totally blind civil persons is slightly higher (257.47 euro in 2006), but we ignored this difference.

The correction made significantly reduces the distortion. As can be noted in Table 14, the initial model population retains a small (but significant) share of elderly individuals receiving only the disability support benefit[34]. Given the lack of further information, it was thought more appropriate to consider the amount provided by Istat as correct, instead of proceeding with questionable calibration operations. The share of individuals receiving only attendance allowances remains, however, substantially underestimated. This can be explained by two reasons. The first is that the cloning procedure of individuals living in residential facilities is not able to reproduce exactly the number of individuals in residential facilities receiving the benefit. This result is unavoidable, given the lack of information on these individuals. We preferred, however, not to attribute ad-hoc benefits to avoid introducing further arbitrariness in building this subgroup of individuals. The second reason explaining the residual underestimate is that, in surveying in 2007 income received in 2006, the IT-SILC survey necessarily underestimates those individuals who, in serious health conditions, receive attendance allowances in the last stage of their life. Basically, part of the 2006 recipients are no longer alive at the time the survey is carried out (outflow), while individuals starting to receive the benefit in 2007 are not considered (inflow).

Table 14

Share of recipients of disability support benefits by age range, after controls

AGE RANGE	Istat-Inps survey 31/12/2006			Initial CAPP_DYN population			Significance level for the difference between the two proportions		
	Pension with att. allow.	Pension without att. allow.	Att. allow. only	Pension with att. allow.	Pension without att. allow.	Att. allow. only	Pension with att. allow. att. allow.	Pension without att. allow.	Att. allow. only
15-34	0.64%	0.48%	0.15%	0.70%	0.62%	0.15%		**	
35-49	0.73%	1.16%	0.15%	0.67%	1.28%	0.16%			
50-64	1.10%	2.41%	0.39%	0.90%	2.25%	0.33%	**		
65-69	0.23%	0.00%	2.58%	0.10%	0.03%	2.26%		***	
70-74	0.37%	0.00%	4.54%	0.30%	0.11%	4.16%		***	
75-79	0.56%	0.01%	8.64%	0.74%	0.05%	6.82%		**	***
80 or more	1.35%	0.01%	26.63%	1.55%	0.30%	21.52%		***	***
Total	*0.76%*	*0.97%*	*2.65%*	0.72%	1.05%	2.14%		*	***

[34] These are individuals declaring receipt of only one type of disability pensions and any receipt of attendance allowances.

Sources: the share of recipients by age range in the Istat-Inps survey is obtained by dividing the total number of recipients at 31/12/2006 resulting from Istat (2009b) and the respective population at 1/1/2007 resulting from demo.istat.it.
Note: the significance level: *** 1%; ** 5%; * 10%. The absence of asterisks means that the null hypothesis is rejected at the 10% level of significance. Sampling weights are not used (see the beginning of par. 4).

The two major disability benefits emerge as sufficiently balanced. Refer to paragraph 5.2 for more detailed analysis.

4. Socio-demographic characteristics of the initial CAPP_DYN population

4.1 Indices used

The following comparisons aim to analyze the possible distortions in CAPP_DYN data (as processed to implement the dynamic micro-simulation procedure) with respect to the information published by official sources, mainly Istat, constituting the benchmark. This analysis concentrates on the relative frequency distribution regarding the phenomenon of interest, of which we investigate the level of similarity or dissimilarity with respect to the benchmark. In particular, two distributions are defined "similar" when characterized by the same relative frequencies, as well as by the same mean value, the same mean square deviation or mode. If this does not occur, to evaluate how far two distributions diverge from each other, one can resort to the "dissimilarity indices among simple distributions according to the same characteristics"[35]. The simple dissimilarity indices are a symmetrical (the order of the difference does not matter) and increasing function of the differences among frequencies belonging to the two distributions. These indices may be absolute or relative, of order equal to *r* (for each *r* positive integer) or of order equal to 1. For absolute indices the ceiling depends on the number of modalities that the variable assumes, while relative indices range between 0 and 1. For instance the first of the two indices that follow, (number 1), represents an absolute index of dissimilarity, the second a relative index of order 2 (or quadratic index).

$$z_1 = \sum_{i=1}^{k} |f_{Ai} - f_{Bi}|$$

$$z_2 = \sqrt{\frac{1}{2}\sum_{i=1}^{k} |f_{Ai} - f_{Bi}|^2}$$

[35] Which will be variously age, civil status, qualification, occupation, etc., qualitative, like civil status, or quantitative, like age.

f_{Ai} represents the relative frequency density associated to the modality *i* of distribution *A*, and f_{Bi} the relative frequency density associated with modality *i* of distribution *B*, the benchmark in this case. We will use relative indices of order 1, in particular:

The average difference:

$$\Delta = \frac{1}{k}\sum_{i=1}^{k}|f_{Ai} - f_{Bi}|$$

Substantially this is the average of the differences between relative frequency densities. This index is greater than or equal to zero, where the equality holds in case of similar distributions, i.e. in case of equality among relative frequencies characterizing the two distributions. The ceiling depends on the number of modalities of the variable. For a given number of modalities, the maximum average difference is achieved when one of the two distributions is entirely concentrated on the *h* modality ($f_{Ah} = 1$), while the other is entirely concentrated on the *j* modality ($f_{Bj} = 1$), with $h \neq j$;

The simple relative dissimilarity index:

$$z_1 = \frac{1}{2}\sum_{i=1}^{k}|f_{Ai} - f_{Bi}|$$

This index ranges in the interval [0,1]. Like the previous one, it assumes 0 value in case of similar distributions, value equal to 1 in case of maximum dissimilarity, which occurs where the distributions are entirely concentrated in two different modalities;

The simple relative dissimilarity index by cumulative frequencies:

$$Z_1 = \frac{1}{k-1}\sum_{i=1}^{k-1}|F_{Ai} - F_{Bi}|$$

F represents the cumulative frequency density[36]. This index also ranges in the interval [0,1]. The ceiling is reached in case of a distribution with all observations concentrated on the first modality ($F_{A1} = 1$), while in the other distribution all observations are concentrated on the latter ($F_{B,k-1} = 0$).

We proceed with the comparisons with respect to demographic and occupational characteristics between the initial CAPP_DYN population and statistics coming from official

[36] The consideration of cumulative frequencies makes sense in case of an ordered rectilinear (qualitative or quantitative) characteristic. Variables with values that are non-ordered and that cannot be put on a ordinal scale are, for instance, gender, civil status, professional activity. Even for nominal data the simple dissimilarity index by cumulative frequencies will be computed, according to an order of categories chosen arbitrarily.

sources. The initial sample of the dynamic model built according to the procedures described in paragraph 4 is called "CAPP_DYN 2006".

The aspects analyzed are: the frequency distribution by age, by educational qualification, by civil status, by occupational status, the distribution of employed individuals by professional activity, the characteristics of foreign individuals residing in our country, individual income and the distribution of recipients of pensions and other benefits.

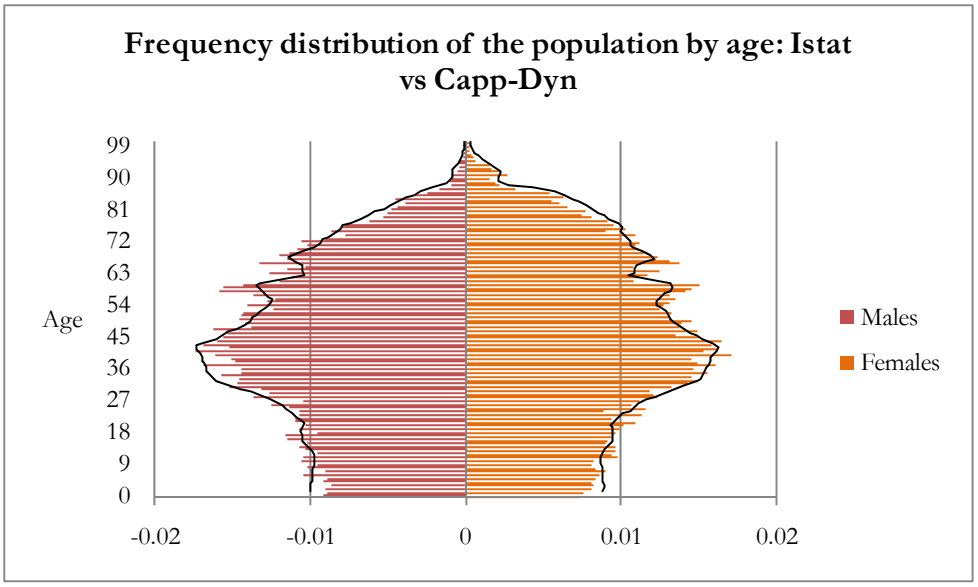
4.2 Age distribution

First of all we examine the demographic characteristics. We analyzed the structure by age and gender according to the information published by Istat[37]. We refer to 1st January 2007 information, as surveyed by the Municipal Registry Office and published by the National Institute of Statistics (Istat).

Figure 1 reports the frequency distribution by age, distinguishing by sex, of the CAPP_DYN sample. Negative frequencies represent female population, while the dark line represents the frequency distribution derived from official data. Concerning the male population, the CAPP_DYN sample tends to underestimate to a greater extent the population aged 0-45, with the largest differences being concentrated among the individuals aged 0-10 and the 27-40 years. Sampling data tend, instead, to overestimate the population of individuals aged 45 and over, even if the differences are not as extensive, with the exception of the 58-61 and 68-70 age ranges. Concerning women, as against men, the frequencies are closer to the official data. There is some correspondence between genders in the age groups for which the sample tends to over- or under-represent the effective population. Indeed, also for women the survey tends to under-represent the population aged 0-45 years, with the largest gaps concentrating among the youngest individuals. Above this age, sampling data in some cases overestimate the effective frequencies, especially around the ages of 60 and 70.

Figure 1
Frequency distributions by age. Istat vs. CAPP_DYN

[37] The webpage http://demo.istat.it reports the official data from the Municipal Registry Offices.



The age distribution figure is re-presented after grouping the population into 5 years ranges, with the exception of individuals aged 80 years and over, who are included in a single cluster.

Dark bars represent official statistics. Comparing them with CAPP_DYN data, it can again be observed that the largest gaps for males refer to the 25-44 groups. For them the gap ranges from about 0.4 to 0.9 percentage points. From the age of 45 generally the opposite is generally the case, i.e. CAPP_DYN data generate larger frequencies than the Registry Office statistics, especially for the 55-59 (0.5%) and 60-64 (0.6%) age ranges. The female population, instead, presents smaller gaps. The largest discrepancies for them are represented by the 0-4, 30-34 and 80 and over age ranges, with a CAPP_DYN frequency 0.5 percentage points lower than the official one, and by the 60-64 age range, with an overestimate of 0.5% with respect to the Registry records.

Figure 2
Frequency distributions by age range. Istat vs. CAPP_DYN

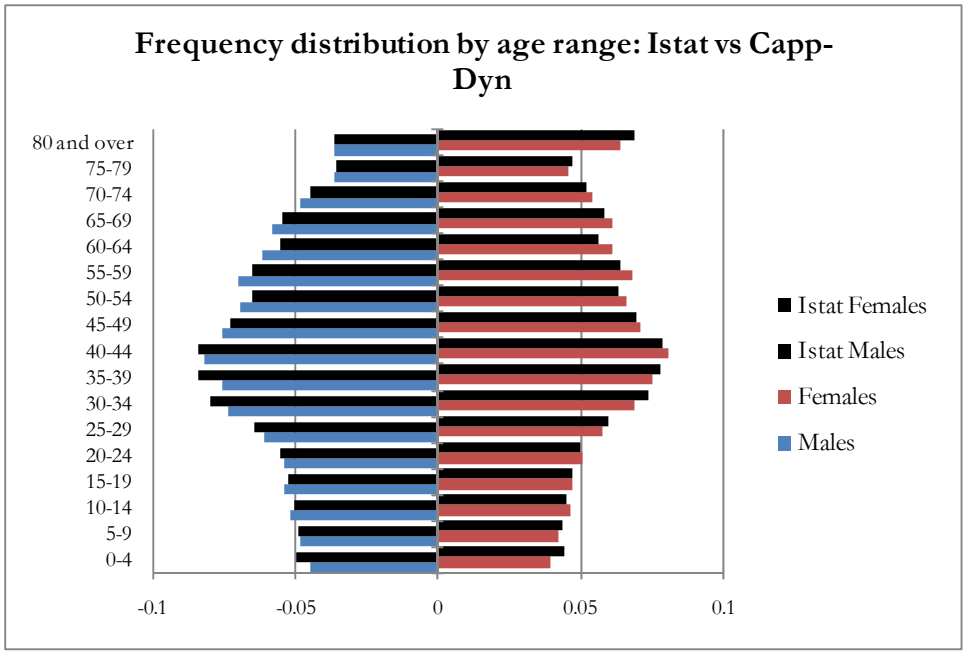


Table ***15*** reports the dissimilarity indices by age. As stated previously, the indices reported are: the average difference, the simple relative index of dissimilarity and the dissimilarity index by cumulative frequencies.

Table 15
Dissimilarity indices by age, divided by gender

	Females	Males	Total
Average difference	0.00062	0.00074	0.00058
Simple dissimilarity	0.03105	0.03726	0.02951
Dissimilarity by cum. frequencies	0.00562	0.00831	0.00678

The table reports larger differences with respect to official data for the CAPP_DYN male population. For both genders, however, the dissimilarity is fairly limited: on the whole, the simple dissimilarity index is about 3%, while the dissimilarity index by cumulative frequencies is lower than 0.7%, values quite near to 0 (no dissimilarity).

The dissimilarity indices are re-presented in Table 16, where the population is grouped into 5-year ranges. Again individuals aged 80 and over are included in a single cluster.

Table 16
Dissimilarity indices by age ranges, divided by gender

	Females	Males	Total
Average difference	0.00266	0.00339	0.00286
Simple dissimilarity	0.02258	0.02884	0.02429
Dissimilarity by cum. frequencies	0.00695	0.01027	0.00833

A further comparison may be made through the demographic structure indicators, like the total dependency ratio, the age dependency ratio, the ageing index and the average age. The indicators obtained from the CAPP_DYN sample are compared, in Table 17, to the indicators computed using Istat values for 2007. The comparison is made for each geographic area and for Italy overall. The total dependency ratio represents the ratio between the "not of working age" population (individuals aged 0-14 and 65 years) and the "working age" population (individuals aged15-64 years), multiplied by 100. CAPP_DYN reports a number of about 51 "not of working age" individuals per 100 individuals aged 15-64 in Italy, very close to the official data (52 individuals). For the North, CAPP_DYN overestimates the "not of working age" population, while for the Centre and the South the opposite occurs, where the latter registers the highest deviation among all areas with respect to the official data (2 individuals). Differences are, however, not very great (1-2 units), and offset each other in such a way as to be small for Italy. The age dependency ratio is constructed as the ratio between persons aged 65 years old and over and the working age population, multiplied by 100. Again, CAPP_DYN and official data do not deviate greatly. The first overestimate the second for all areas except the Centre, with a moderate difference between 0 and 2 elderly per 100 active individuals. For Italy, CAPP_DYN exceeds the official statistics by 1 unit. The ageing index represents the ratio between individuals aged 65 and over and those aged 0-14 years, multiplied by 100. This is the index which shows the largest differences, in excess with respect to Istat for two of the three areas (Centre and South), while for the North the two sources approximately coincide. The largest deviation corresponds to the South (12 individuals), with 125 elderly per 100 children, compared to 113 resulting from the Istat statistics. The average age is weighted by the relative abundance of individuals of each age. This indicator shows agreement between the two sources, taking into consideration the approximation of the official statistics.

Table 17
Demographic structure indicators

	CAPP_DYN	Istat

	North	Centre	South	Italy	North	Centre	South	Italy
Total dependency	53.35	52.11	47.96	51.32	52	53	50	52
Age dependency	32.73	32.53	26.65	30.72	32	33	26	30
Ageing index	158.7	166.1	125.0	149.1	159	162	113	142
Average age	43.73	43.88	41.15	42.95	44	44	41	43

4.3 Qualification distribution

Regarding qualification, comparisons are made with respect to the information coming from the Labour Force Survey (LFS), conducted weekly by Istat and whose findings are published quarterly and yearly. This is a sample survey which represents the main statistical source on the Italian labour market and that, each quarter, concerns about 175,000 individuals. The alternative would be to refer to the last population census, but given the six years of difference with respect to CAPP_DYN data, the first source, although of sampling character, seems more appropriate. We refer to the 2006 survey, since, to eliminate the temporal imbalance between income and other characteristics, variables originally referring to 2007 have been retraced to 2006. The variable regarding the qualification has been encoded according to 4 values, which represent: title from compulsory education (primary and secondary school), diploma (received after 2-3 years or 5 years school attendance), three-year degree and higher qualifications. The Labour Force Survey, however, does not distinguish postgraduate degrees. It was therefore necessary to merge the last two classes of individuals in a single one, reducing the number of values assumed by the variable to three. Table 18 shows the dissimilarity indices by qualification, divided by area and sex, for the population aged at least 15 years. For all areas, the dissimilarity indices assume relatively low values. For Italy the simple dissimilarity index ranges from 3.2% (for females) to 4.5% (for males), while the dissimilarity index by cumulative frequencies is around 2%. The highest values correspond to the North, though the values for the South do not differ much. In all areas females report smaller indices than males, almost always close to zero in the Centre.

Table 18

Dissimilarity indices by qualification, divided by area and gender

	Females	Males	Total	Females	Males	Total
	North			Centre		
Average difference	0.0256	0.0307	0.0271	0.0078	0.0213	0.0128
Simple dissimilarity	0.0384	0.0461	0.0406	0.0116	0.0319	0.0191
Dissimilarity by cum. frequencies	0.0192	0.0265	0.0213	0.0077	0.0160	0.0096
	South			Italy		
Average difference	0.0241	0.0300	0.0268	0.0213	0.0300	0.0246
Simple dissimilarity	0.0361	0.0451	0.0402	0.0320	0.0450	0.0369
Dissimilarity by cum. frequencies	0.0186	0.0276	0.0228	0.0160	0.0259	0.0195

Figure 3, Figure 4 and Figure 5 illustrate the relative frequency distributions by age, respectively for holders of compulsory school completion, diploma and tertiary education.

Figure 3

Frequency distribution by age, divided by gender, for those completing compulsory education

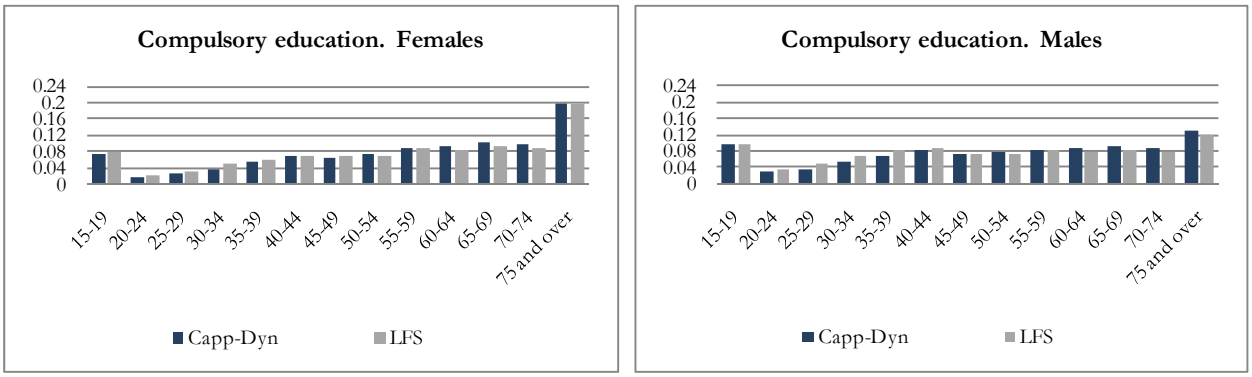


Figure 4

Frequency distribution by age, divided by gender, for holders of diploma

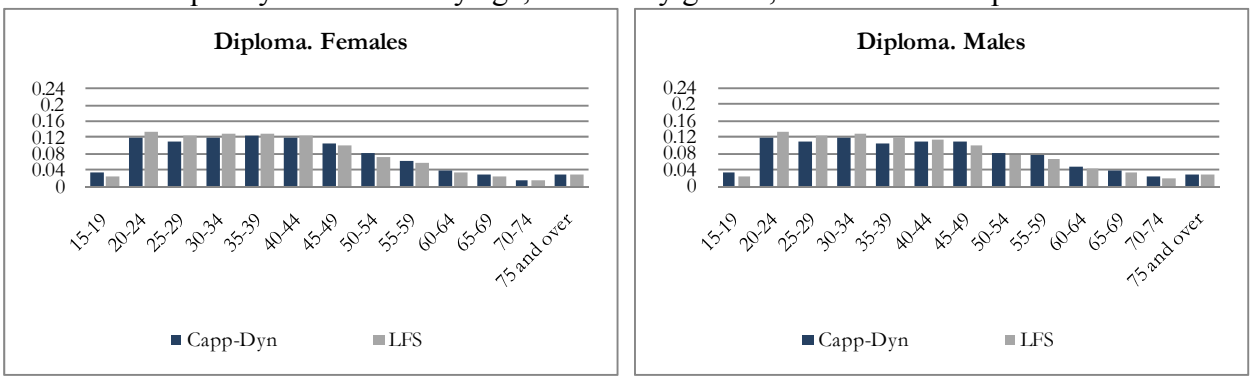
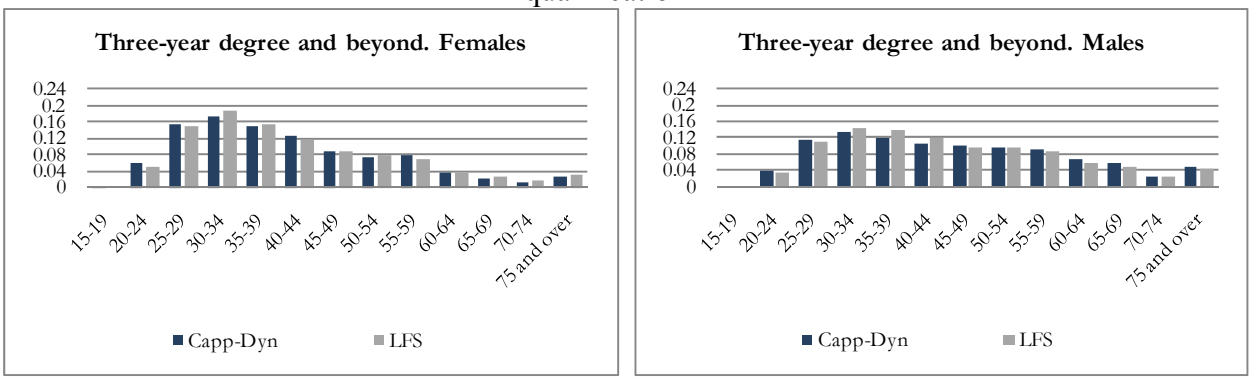


Figure 5

Frequency distribution by age, divided by gender, for holders of three-year degree or higher qualification



For all education levels, CAPP_DYN data tend to report lower frequencies with respect to the Labour Force Survey for the first age ranges, the opposite from age 45. The figures do not reveal significant differences between the two sources, for both genders.

4.4 Marital status

The civil status variable was encoded so as to assume 4 values: married, single, divorced and widow/er individuals. The comparison is made with respect to the official statistics referring to the population on 1st January 2007 published by Istat. Table 19 reports the dissimilarity indices between frequency distributions (CAPP_DYN and Istat).

Table 19

Dissimilarity indices by marital status, divided by gender

	Females	Males	Total

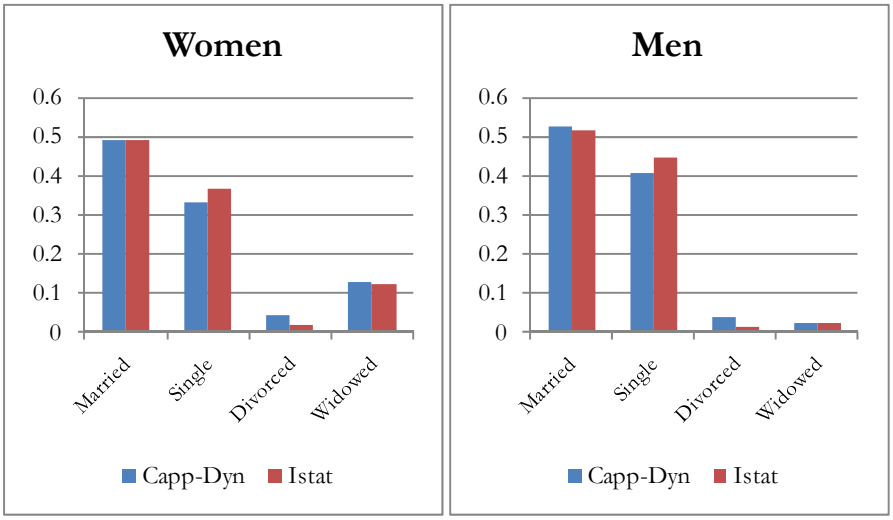
Average difference	0.01574	0.01977	0.01786
Simple dissimilarity	0.03148	0.03953	0.03571
Dissimilarity by cum. frequencies	0.01201	0.01370	0.01309

Index values are fairly limited and appear to be larger for males than females. However, there are sizeable differences for some groups of individuals that can be better detected by showing the frequency distribution. The differences between the two frequency distributions are therefore presented in

Figure ***6***. For married and widowed individuals values are sufficiently aligned between the two sources, whereas the major differences can be found for single individuals (3.2% for women, 4% for men) and, to a lesser extent, for divorcees (2.7% for women, 2.6% for men).

Figure 6

Frequency distribution by marital status



4.5 Employment status

The employment status variable was encoded according to 4 values: full-time employed, part-time employed, unemployed and "outside the labour force" individuals. Comparisons are made for the population aged at least 15 years with respect to data coming from the Labour Force Survey. Table 20 illustrates the dissimilarity indices, both by gender and on the whole population.

Table 20

Dissimilarity indices by employment status, divided by gender. Population aged at least 15 years

	Females	Males	Total
Average difference	0.0181	0.0069	0.0094
Simple dissimilarity	0.0363	0.0137	0.0188
Dissimilarity by cum. frequencies	0.0166	0.0083	0.0063

With respect to this characteristic, the indices assume different values for men and women. The simple dissimilarity index assumes a value of about 1.4% for men, as against 3.6% for women. The dissimilarity index by cumulative frequencies is pretty close to zero for men, equal to 1.7% for women. The differences between the two sources are thus quite limited.

Figure 7, Figure 8, Figure 9 and Figure 10 compare the CAPP_DYN and the LFS distributions for Italy as a whole and for each geographic area. The largest differences found are the following: in the North, for men outside the labour force (2.2 percentage points) and for employed women, both full-time and part-time (1.8-1.9%); in the Centre, for full-time employed women (2.7%) and for part-time employed women (2.9%); in the South, for unemployed, both women and men (3.2 and 3.7% respectively) and for individuals outside the labour force, both women and men (4.1 and 2.8% respectively).

Figure 7

Frequency distribution by employment status, divided by gender. Italy

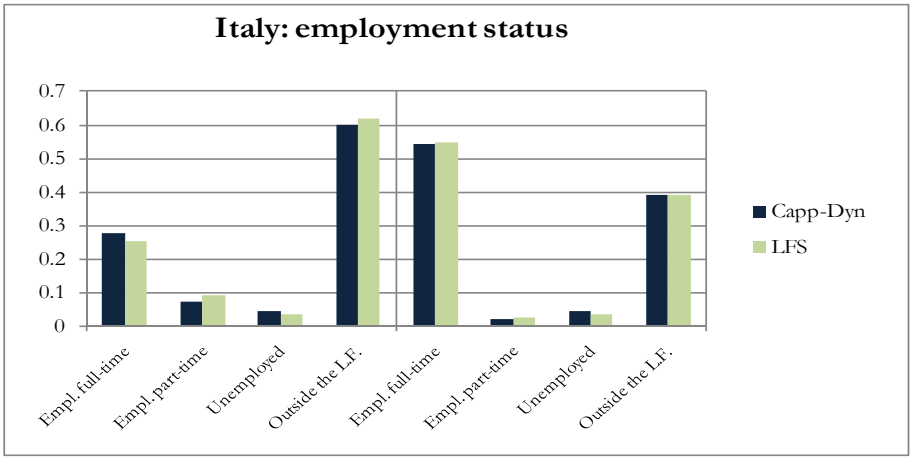


Figure 8
Frequency distribution by employment status, divided by gender. North

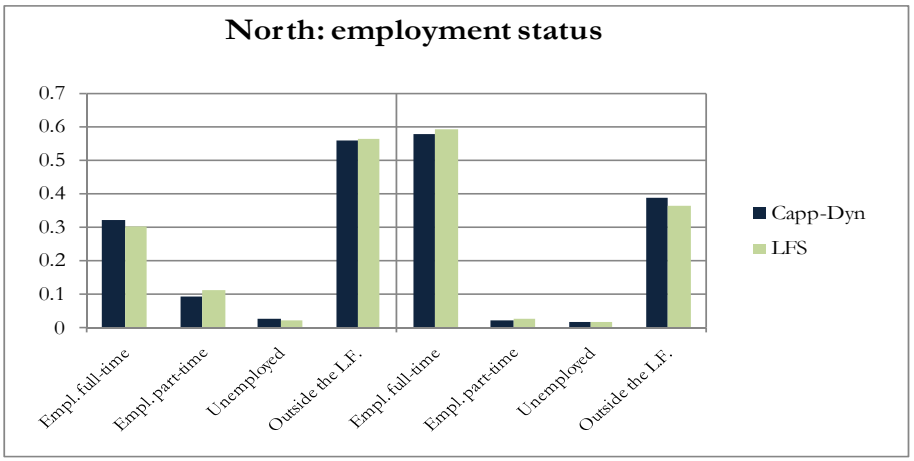


Figure 9
Frequency distribution by employment status, divided by gender. Centre

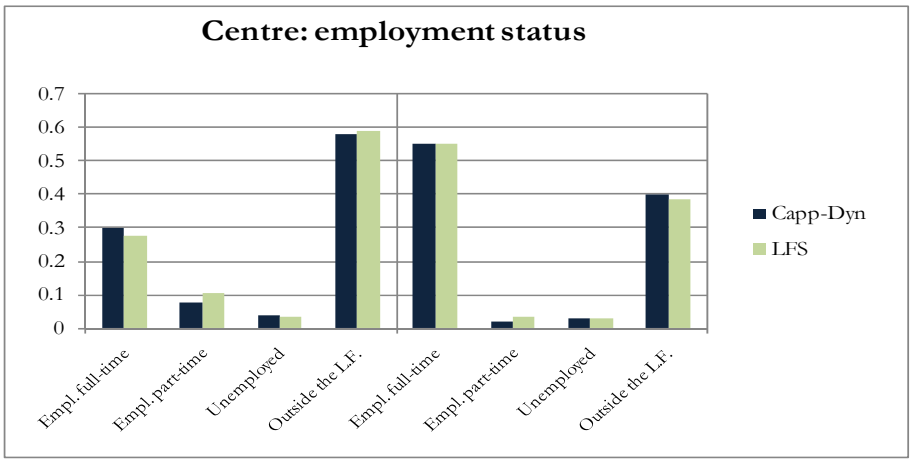
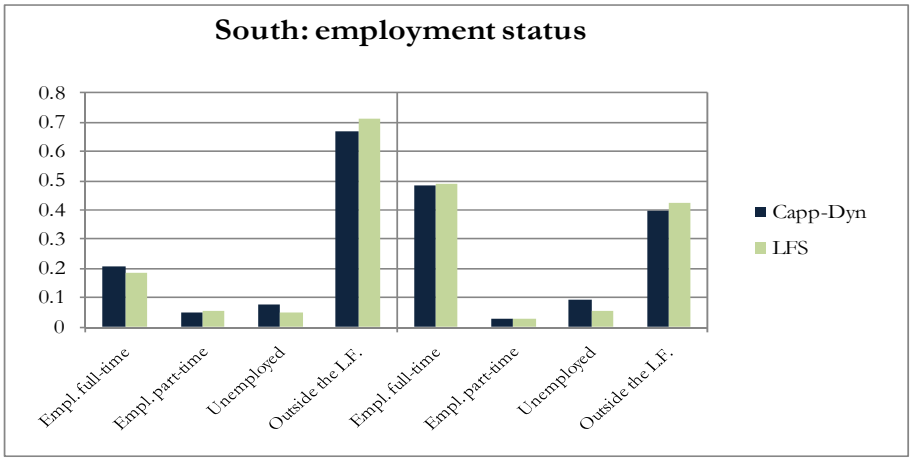


Figure 10
Frequency distribution by employment status, divided by gender. South



Note that part-time jobs are more widespread among women than men. In Italy, according to the Labour Force Survey, 9.2% of women are characterized by this occupational status (7.6% according to CAPP_DYN). Only 2.7% of men are part-time workers according to the Labour Force Survey (2.2% according to CAPP_DYN). However for the South the differences between sexes tend to fade. In this area about 5% of women are part-time workers (5.6% according to LFS, 4.7% according to CAPP_DYN), against 2.6-2.7% for men. Data also show an higher level of employment of men, both full-time and part-time. In Italy 56% of the male population above 15 years old are employed according to CAPP_DYN, against 35% of females according to CAPP_DYN. The following tables show the comparison between the two frequency distributions (CAPP_DYN and LFS) by occupational status, for each gender and age range.

Table 21
Frequency distribution by employment status for individuals aged 15-24

	Women		Men	
	CAPP_DYN	LFS	CAPP_DYN	LFS
Empl. full-time	16.30%	14.28%	26.65%	27.50%
Empl. part-time	4.65%	5.86%	2.07%	3.10%
Unemployed	9.23%	6.81%	11.16%	7.22%
Outside the L.F.	69.82%	73.05%	60.11%	62.18%
Total	100.00%	100.00%	100.00%	100.00%

Table 22
Frequency distribution by employment status for individuals aged 25-34

	Women		Men	
	CAPP_DYN	LFS	CAPP_DYN	LFS
Empl. full-time	47.59%	43.96%	76.95%	76.98%
Empl. part-time	12.84%	15.21%	4.29%	3.90%
Unemployed	11.91%	7.62%	8.37%	6.55%
Outside the L.F.	27.66%	33.22%	10.38%	12.57%
Total	100.00%	100.00%	100.00%	100.00%

Table 23

Frequency distribution by employment status for individuals aged 35-54

	Women		Men	
	CAPP_DYN	LFS	CAPP_DYN	LFS
Empl. full-time	47.02%	43.05%	87.90%	87.51%
Empl. part-time	13.72%	16.25%	2.64%	2.78%
Unemployed	4.93%	3.82%	3.82%	2.95%
Outside the L.F.	34.33%	36.89%	5.64%	6.76%
Total	100.00%	100.00%	100.00%	100.00%

Table 24

Frequency distribution by employment status for individuals aged 55 and over

	Women		Men	
	CAPP_DYN	LFS	CAPP_DYN	LFS
Empl. full-time	8.27%	6.68%	21.00%	19.76%
Empl. part-time	1.35%	1.79%	0.86%	1.94%
Unemployed	0.58%	0.11%	0.92%	0.56%
Outside the L.F.	89.80%	91.42%	77.23%	77.73%
Total	100.00%	100.00%	100.00%	100.00%

Large differences between the two sources can be found for women of all ages, and are much more concentrated in the categories of full-time employed and "outside the labour force" women. Again for women, the discrepancies are larger for those aged 25-34; the differences range from 2.37 percentage points, in absolute value, for part-time employed to 5.56% for outside the labour force women. Women aged 55 years and over show, instead, the lowest disagreement: the difference ranges from 0.44% for part-time employed to 1.59% for full-time employed. Important differences can also be found for full-time employed women of 25-34 and 35-54 years (3.63% and 4.15% respectively) and for women outside the labour force in the first age range (3.23%). As stated previously, men show greater convergence between the two sources. Differences stay below or around 1 percentage point. The largest relate to unemployed and to outside the labour force individuals in the 15-24 age range (3.94% and 2. 07% respectively) and to outside the labour force individuals in the 25-34 age

range (2.19%). With the exception of the last age range, unlike women, the larger discrepancies refer to unemployed and outside the labour force individuals.

4.6 Working activity

This subsection analyzes in greater detail the category of workers. Comparisons will focus on the character of the occupation and on income. In this case the benchmark is the information released by INPS referring to 2006 for employees in the private sector, self-employed and para-subordinates.[38] By employees, we refer to individuals who have had at least one contributory payment during the year; by self-employed, the number of those who have been enrolled in the register of this category of workers during the year or even for a fraction of the year; by para-subordinates, we refer to the number of individuals with cooperation agreements enrolled in the separate register who have had at least one contributory payment during the year.[39] Regarding the latter, information is rather limited: we can observe the total number, the breakdown by sex but, unlike other types of workers, we do not know the breakdown by age ranges. There is no available information on public employees, nor from national social security institutes. Referring to them, we will initially use the data published by the State General Accounting Department (in Italian "Dipartimento della Ragioneria Generale dello Stato", or for brevity RGS), coming from the survey "Conto Annuale", containing the results of the census survey aiming to control the costs of public work. Unfortunately we can proceed to a detailed breakdown of public employment (as for age ranges and by area) only for employees with open-ended jobs. For workers with temporary jobs we know instead the total number of employees divided by sex. Table 25 reports the dissimilarity indices by working category, for women, men and overall, where INPS data were used to identify the employees of the private sector, while RGS information was used to identify public employees (both with open-ended and temporary jobs). The dissimilarity indices reach higher values for women with respect to men and turn out to be quite large: the simple dissimilarity index assumes a value of about 9% for women, 8% for men.

Table 25

[38] See the information published on the INPS website, in the section titled "Osservatori statistici".
[39] As established by clause 2, paragraph 26 of law no. 335/1995.

Dissimilarity indices by working category

	Females	Males	Total
Average difference	0.04570	0.03985	0.04224
Simple dissimilarity	0.09141	0.07970	0.08448
Dissimilarity by cum. frequencies	0.03323	0.02657	0.02816

However, because of lack of RGS information, it is impossible to make further comparisons between survey and official data. For this reason, from now on, only private sector employees will be dealt with. The only benchmark will therefore be the INPS data. Later on, the Department of Finance fiscal statistics will also be considered. The latter can be a useful complementary source, since it considers the universe of employees, not only those belonging to the private sector. Furthermore, this source contains additional information, such as the breakdown by area and age. Table 26 contains the dissimilarity indices by working category and gender, for all working individuals, with the exception of public employees. Since this variable is a nominal one, the simple dissimilarity index by cumulative frequencies makes sense only for the category order chosen.[40]

Table 26

Dissimilarity indices by working category, divided by sex (excluding public employees)

	Females	Males	Total
Average difference	0.03182	0.03382	0.03344
Simple dissimilarity	0.04774	0.05073	0.05017
Dissimilarity by cum. frequencies	0.02387	0.02537	0.02508

All indices are higher for men with respect to women (even if the differences are quite limited), but for both they remain at fairly low levels; although not very close to zero, they are well below 10%. Moreover, it may be noted how the discrepancies are smaller when only private sector employees are considered. In the comparison with the information from the Department of Finance, we will try to investigate the extent to which the differences emerging from Table 25 still persist, considering this alternative source for the purpose of including the entire universe of employees. It is interesting to note the diversity of the two distributions for each area. To this end, Table 27 re-presents the dissimilarity indices, but for each area. We omit the indices computed on the entire population, already included in Table 26.

[40] Note that a nominal variable is characterized by categories that are non-ordered and that cannot be put on an ordinal scale.

Table 27

Dissimilarity indices by working category, divided by area (excluding public employees)

	North	Centre	South
Average difference	0.02769	0.03472	0.04589
Simple dissimilarity	0.04154	0.05208	0.06884
Dissimilarity by cum. frequencies	0.02077	0.02604	0.03442

As shown by this table, considering a narrower territorial level, the dissimilarity turns out to be larger in some cases, especially for the South and for the second indicator, which in any case remains below the 10% level. The dissimilarity index by cumulative frequencies is conditioned by the fact that, if for employees and self-employed the discrepancies in the South are the widest among all areas, for para-subordinates, the characteristic belonging to the right tail of the distribution according to the selected order, they are smaller. As already said, indeed, this index is very sensitive to a greater concentration on the extreme tails of the distribution. The differences between the two distributions illustrated in Table 26 and in Table 27 are highlighted through Figure 11 and Figure 12.

Figure 11

Frequency distribution by working category, by gender (excluding public workers)

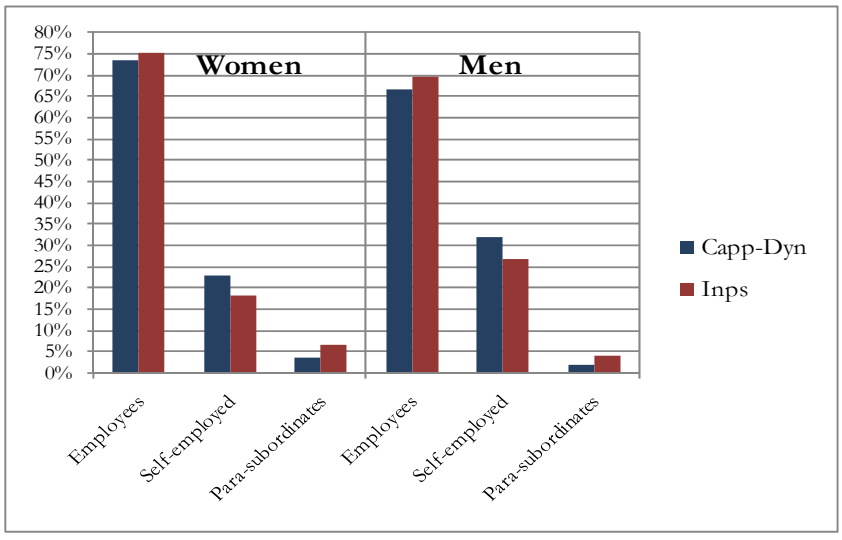
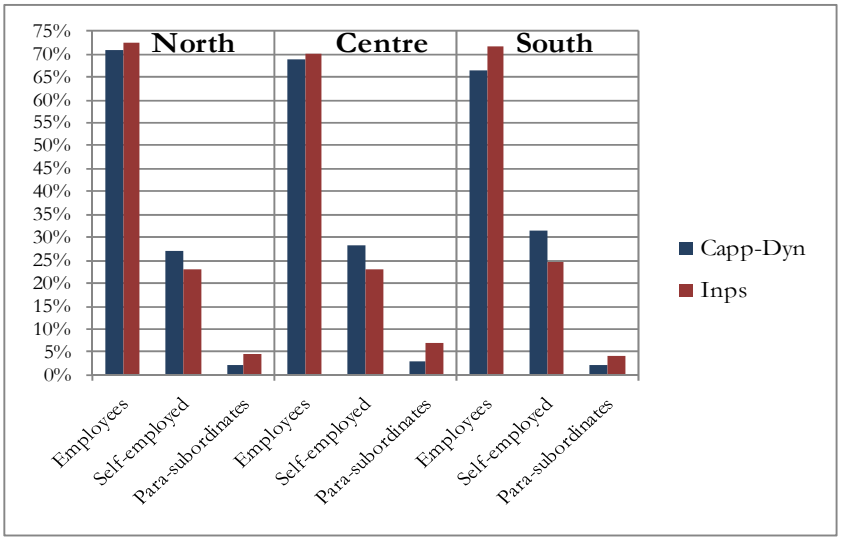


Figure 12

Frequency distribution by working category, by area (excluding public workers)



The figures illustrate that for employees and self-employed workers the largest differences concentrate in the South; for para-subordinate workers the most important discrepancies emerge in the Centre. With respect to the comparison between genders, the deviations between the two distributions are similar, slightly larger for men.

However the comparison by working activity cannot go any further, given the insufficient articulation of the INPS data for the para-subordinate workers category. It is noteworthy that if in CAPP_DYN the activity has an exclusive nature, in the official statistics this is not the case.[41] Hence it is possible that the INPS data contain workers that belong to more than one category. Further comparisons with respect to data published by INPS can be made for private sector employees and for self-employed workers. **Errore. L'origine riferimento non è stata trovata.** and Figure 14 shows the comparison between distributions for the categories of workers above, divided by age.

With reference to employees, for women the whole distribution shows differences, with the exception of individuals over sixty years old, and is concentrated in the first three age ranges, with a peak for the 40-49 age band (of about 3 percentage points). For men the dissimilarities concentrate mainly in the first and the fourth age range (up to 29 and 50-59 years old), amounting respectively to 3.5% and 2.5% in absolute value.

For self-employed, instead, the two distributions seem to converge. For women higher differences emerge, especially for the 22-49 age band (3.7%) and 65+ (1.9%). For men, the largest discrepancies (2.1%) emerge for the 50-59 age band.

[41] For instance, if an individual is party to an employment agreement and a collaboration contract, he/she is considered in only one of the two positions, according to which one prevails in his/her working life.

Figure 13

Frequency distribution of private sector employees by age

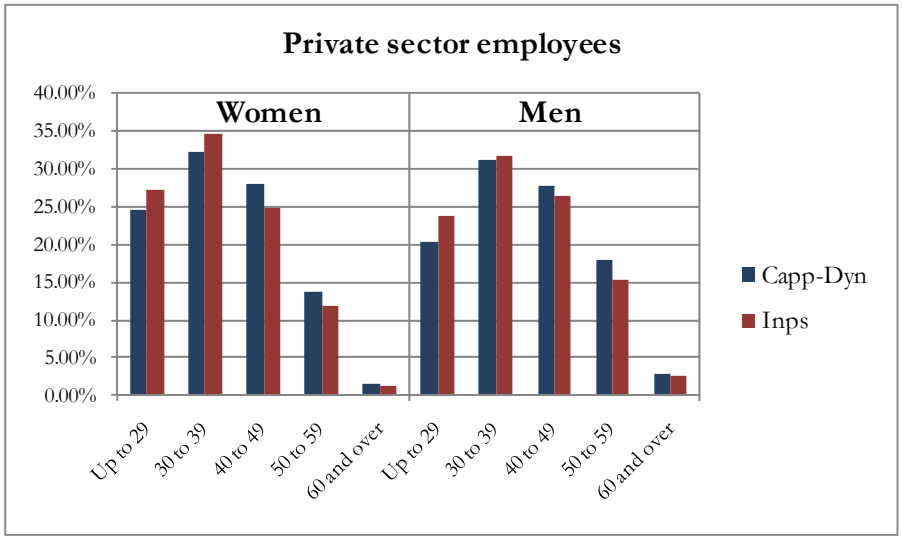
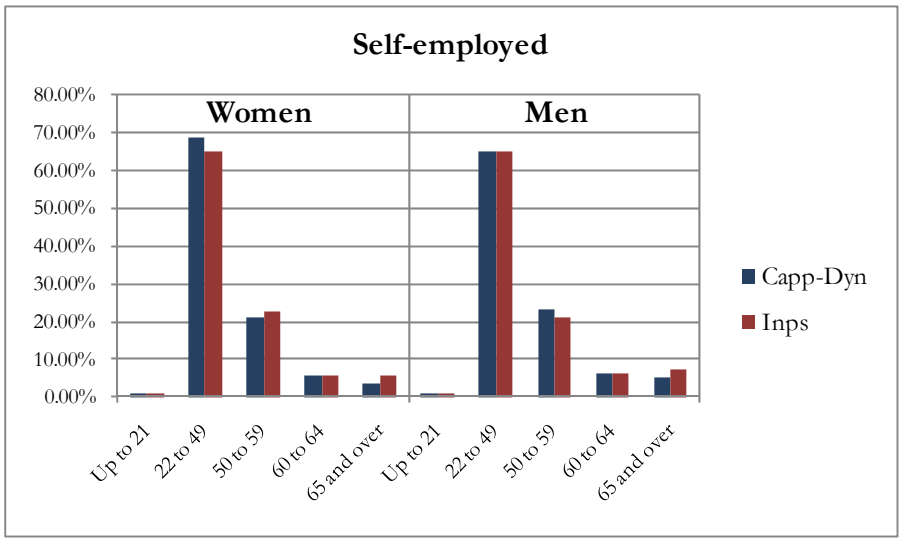


Figure 14

Frequency distribution of self-employed workers by age



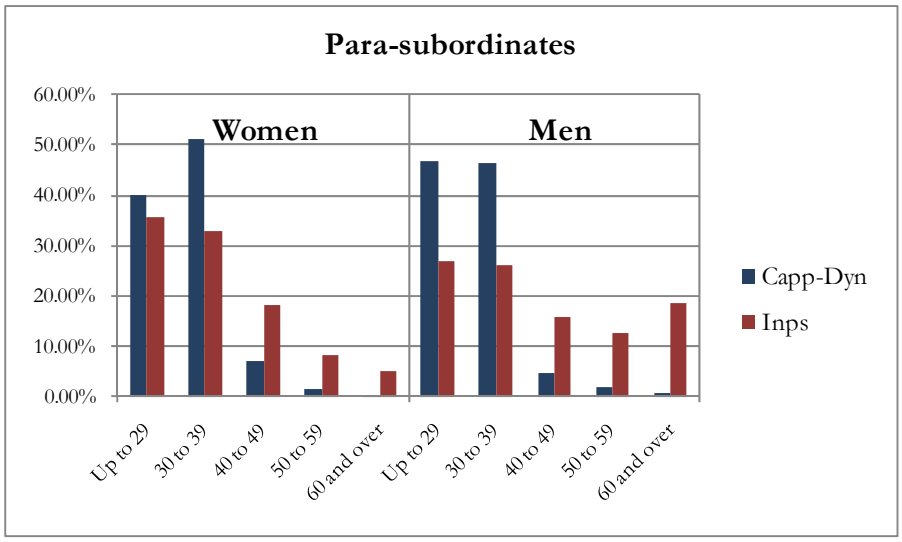
As already said, there is no information available on the distribution of para-subordinates by age range for 2006. INPS published such information only for 2008. However it may be safe to assume that there have been no significant disruptions in the age distribution of such workers over two years (from 2006 to 2008). Therefore, with the aim of obtaining some minimum indication about the closeness of the sampling population of para-subordinates to the real one, the CAPP_DYN distribution by age (which, remember, refers to

2006) is compared to the effective 2008 distribution, as obtained from INPS information. The two distributions are illustrated in

Figure *15*. Assuming that the distribution in 2008 follows the same trend as in 2006, para-subordinates seem to be too concentrated in the group of individuals younger than 40 years, whereas for workers above than this age too few observations are reported by CAPP_DYN. Although we can reasonably assume an excessive concentration in the relatively younger age bands, it would be necessary to have adequate information to quantify the difference between the image outlined by CAPP_DYN and the actual situation.

Figure 15

Frequency distribution of "para-subordinates" holders of cooperation agreements by age



As previously stated, a further comparison, with reference to the category of employees, can be made on the basis of data published the Department of Finance, a branch of the Ministry of the Economy and Finance (MEF). These data report some statistics, such as the number or the relative frequency, about employees and their distribution by region, age range or gender. The category of employees reported by the ministerial source includes not only private and public sector employees, but also recipients of income assimilated to employees' wages (according to tax legislation), and therefore para-subordinates are also included. The comparison will thus be made with respect to the CAPP_DYN total number of employees (public and private), of employer-coordinated freelance workers ("titolari di rapporti di collaborazione coordinata e continuative") and of project workers ("collaboratori a progetto"). Table 28, Table 29 and Table 30 report the two relative frequency distributions, respectively by sex, age range and area. With respect to the division by sex and age range, the

differences between the two frequency distributions turn out to be rather small, of one percentage point at most. Greater differences can be found dividing by area, especially in the Centre (with a difference of 4.5 percentage points with respect to the ministerial data) and in the South (3.5%). Such evidence reflects to some extent the findings emerging in the comparison between CAPP_DYN and INPS data (Table 27, Figure 11 and Figure 12). In that case the greater differences for private sector employees were found in the Centre and in the South, while for para-subordinates in the Centre. Regarding the distribution by age range, the similarities between CAPP_DYN and ministerial data are very obvious and contradict the findings emerging from the previous comparisons. This may be due to the existence of differences between the two benchmarks, respectively INPS and the Department of Finance (also because of different criteria in computing the number of individuals), but also in part to the existence of some compensation through the aggregation of employees and para-subordinates.

Table 28

Frequency distribution of employees and para-subordinates by gender (CAPP_DYN against MEF)

	Females	Males
CAPP_DYN	43.91%	56.09%
Dept. Finance	42.83%	57.17%

Table 29

Frequency distribution of employees and para-subordinates by age range (CAPP_DYN against MEF)

	15-24	25-44	45-64	65 and over
CAPP_DYN	7.61%	55.78%	35.83%	0.79%
Dept. Finance	8.06%	55.09%	35.08%	1.78%

Table 30

Frequency distribution of employees and para-subordinates by area (CAPP_DYN vs. MEF)

	North	Centre	South
CAPP_DYN	49.73%	24.62%	25.64%
Dept. Finance	50.81%	20.10%	29.10%

4.7 Immigrants

In the same way as we studied the total population, we can now study the population of foreigners residing in Italy. First, **Errore. L'origine riferimento non è stata trovata.** illustrates the CAPP_DYN distribution by age against the distribution obtained from Istat data referring to foreign residents on 1st January 2007. The dark line represents official data, the horizontal bars the CAPP_DYN population. There are significant differences across the whole distribution between the two sources, but the largest can be found for males up to the age of 40 (with a maximum difference of 1.1% for age 32) and from ages 47 to 56 (with a maximum difference of 1-2% for individuals aged 54); for females, the largest discrepancies can be found up to age 46 (with a maximum difference of 1.1% for ages 27 and 46 years).

Again, we can search if there is some compensation dividing the population by age bands. The two populations are therefore divided into 5-year range groups, with the exception of individuals aged 80 years over, included in a single group. The respective frequency distributions are illustrated in Figure 17. However, grouping by age range does not seem to mitigate the differences, especially for individuals up to the age of 59, also because of the equality of sign of the differences between contiguous ages (**Errore. L'origine riferimento non è stata trovata.**). For females, the largest discrepancies emerge for the 5-9 and 25-29 age bands, with a difference, in absolute value, of 1.4% and 2.2% respectively; for males, instead, they can be found in the 30-34 and 50-54 age bands, with a difference of 1.9 and 2.2% respectively. With the exception of very few cases, the CAPP_DYN population generally under-represents the official one for ages below 50 years. For ages above, exactly the opposite occurs.

Figure 16

Frequency distribution by age of immigrants

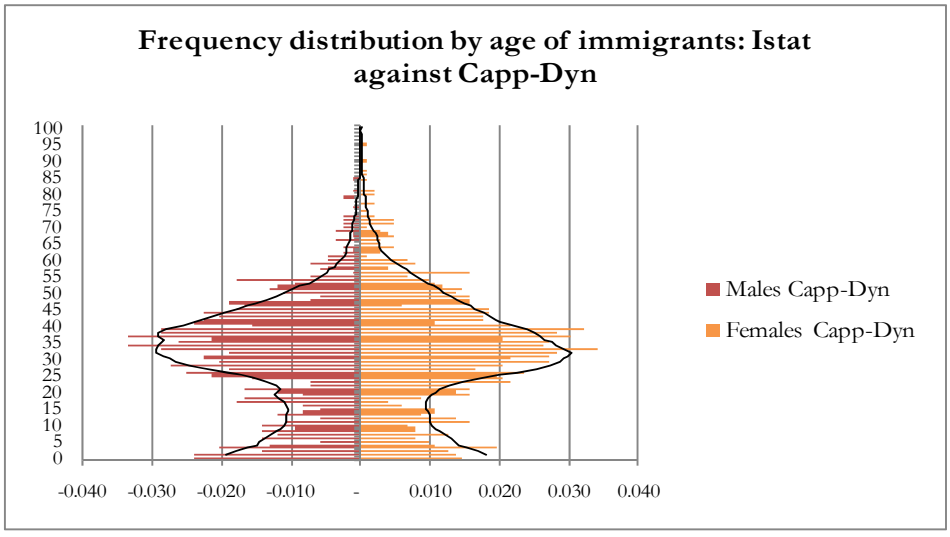


Figure 17
Frequency distribution by age range of immigrants

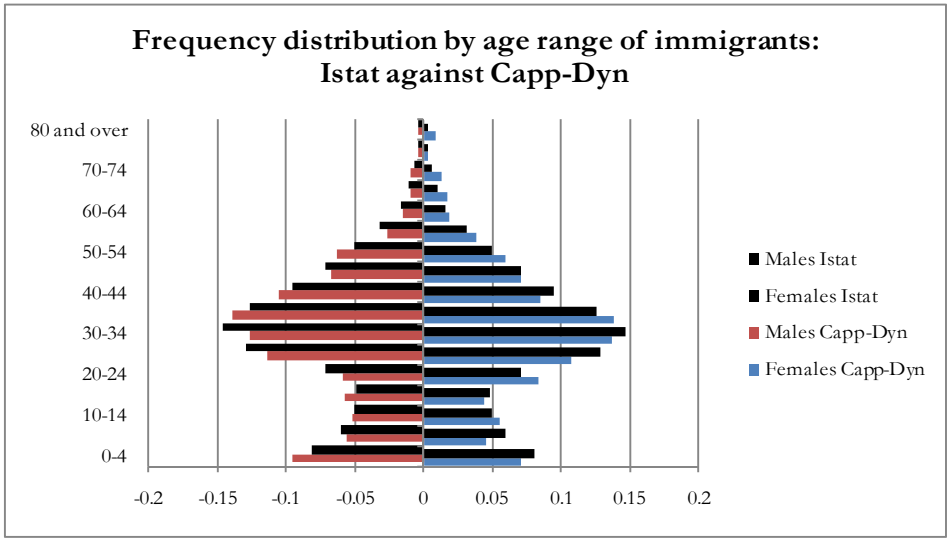


Table 31 reports the dissimilarity indices by age for immigrants, divided by gender and as a whole. The evidence emerging from the first two figures is confirmed in that table. The simple dissimilarity index shows high values for both sexes, higher than 12%, although it seems that the differences decrease when the total is considered. The dissimilarity index by cumulative frequencies remains fairly low, since the most important differences regard central ages in the frequency distribution. Table 32 shows how the division by classes affects the dissimilarity indices. The fact that the average difference is greater than in Table 31 is due to the lower number of values assumed by the variable and, as illustrated in the previous pages, the range of variation of this index depends on the number of values assumed by the variable. The simple dissimilarity index reduces considerably, while the index of dissimilarity by cumulative frequencies increases slightly.

Table 31
Dissimilarity indices by age of the foreign population, divided by gender

	Females	Males	Total
Average difference	0.00241	0.00242	0.00174
Simple dissimilarity	0.12157	0.12214	0.08777
Dissimilarity by cum. frequencies	0.01737	0.01166	0.01467

Table 32
Dissimilarity indices by age range of the foreign population, divided by gender

	Females	Males	Total
Average difference	0.00799	0.00595	0.00593
Simple dissimilarity	0.06790	0.05061	0.05041
Dissimilarity by cum. frequencies	0.02055	0.01364	0.01747

We now illustrate how the foreign population affects the overall number of residents. **Errore. L'origine riferimento non è stata trovata.**,

Figure *19* and Figure 20 show, by age ranges, the percentage ratio between foreign and overall population. Significant differences emerge also in this case. The CAPP_DYN percentage of foreigners under-represents the one arising from official data by 1.4% (1% for females 1.8% for males). Looking more in depth at the structure by age, large differences can be found up to age 50, especially for males. If, for instance, CAPP_DYN reports a number of 5.7 foreigners for 100 residents in the 30-34 age band, according to Istat they number 9.3. In this case the major discrepancies are concentrated in the 0-10 and 25-49 age ranges for females, while they remain quite high and fairly stable for males up to age 49.

Figure 18

Foreigners in relation to the total population by age ranges (percentage values). Females

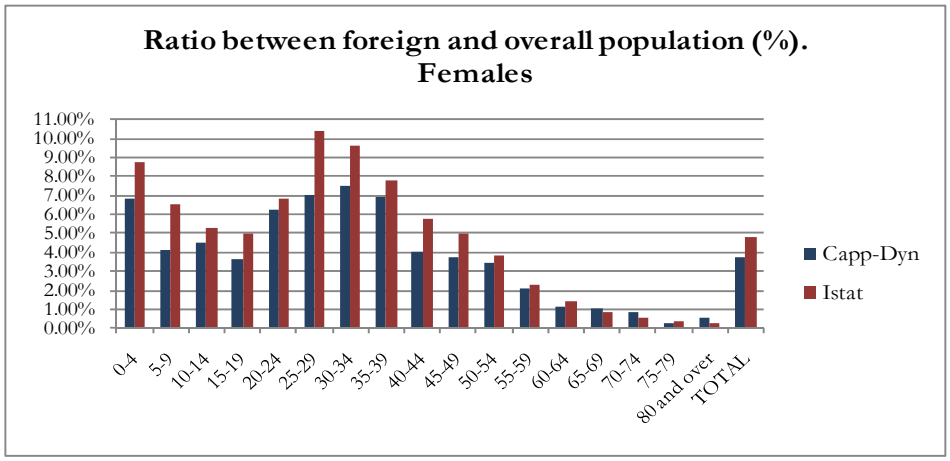


Figure 19

Foreigners in relation to the total population by age ranges (percentage values). Males

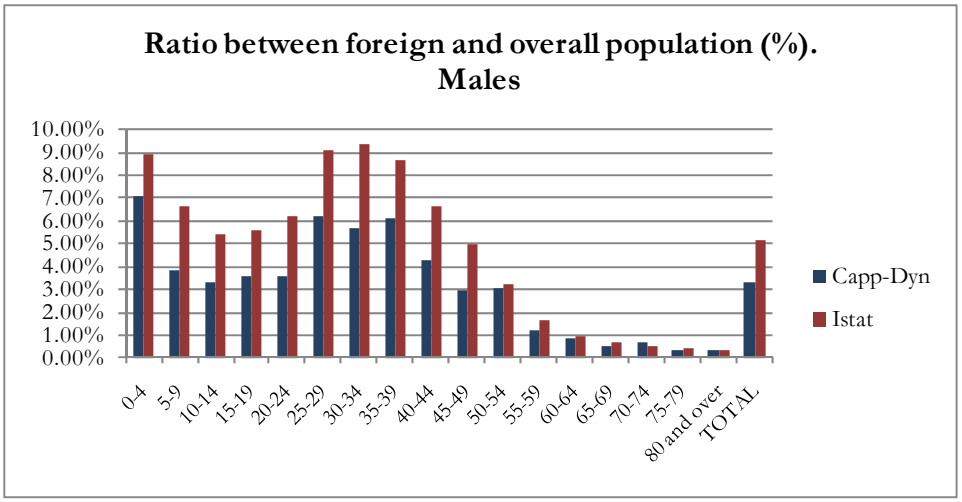
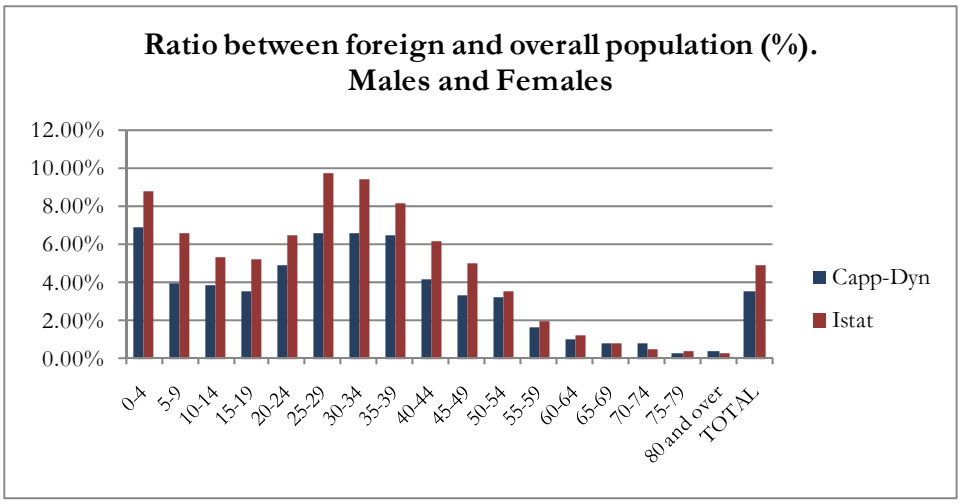


Figure 20

Foreigners in relation to the total population by age ranges (percentage values). Males and Females



We now proceed to the comparison between frequency distributions by residence area and employment status of the foreign population.

Table 33 reports the dissimilarity indices by residence area for foreign residents. There is only a slight disagreement between distributions for both genders. If the frequency of foreigners living in the South coincides almost perfectly between the two sources, some disagreement exists in the North (CAPP_DYN frequencies under-represent the official statistics by 5 percentage points, both for females and males) and in the Centre (CAPP_DYN frequencies exceed the Istat frequencies by 5%).

Table 33

Dissimilarity indices by residence area for foreigners, by gender

	Females	Males	Total
Average difference	0.03553	0.03152	0.03494
Simple dissimilarity	0.05330	0.04728	0.04728
Dissimilarity by cum. frequencies	0.02665	0.02600	0.02761

Regarding the employment status, comparisons are made with respect to the information coming from the Labour Force Survey.

First, we report in Table 34 the dissimilarity indices by employment status.[42] On the whole, the differences are quite limited: the simple dissimilarity index amounts to 1.8%, quite near to a dissimilarity equal to zero, like the dissimilarity index by cumulative frequencies. However for women, values significantly higher than for men are observed.

Table 34

Dissimilarity indices by employment status for foreigners in the 15-64 age range

	Females	Males	Total
Average difference	0.03120	0.00615	0.00854
Simple dissimilarity	0.06239	0.01229	0.01798
Dissimilarity by cum. frequencies	0.02080	0.00410	0.00954

The following figures (Figure 21 and Figure 22) report the economic activity, the employment and the unemployment rate for foreigners, divided by residence area and gender. Some important differences emerge between the two sources. The greatest discrepancies regard women: CAPP_DYN rates exceed the others with respect to the first and third indicator, while for the employment rate the difference is not of the same sign among areas. Then, CAPP_DYN data report a greater participation rate of foreign women in the labour market, but also a much higher unemployment. For men the main differences relate to unemployment rates and generally to the South (the greatest difference can be observed for the activity rate in the South, of about 4.4%). In sum, unemployment rates and the South are characterized by the highest discrepancies, especially for women, for which the maximum difference corresponds to 14 percentage points for unemployed females in the South.

[42] Similarly to what has been done for the total population, foreigners are grouped by full-time workers, part-time workers, unemployed and those outside the labour force, this time in the age class 15-64.

Figure 21
Economic activity rate for foreigners in the 15-64 age range

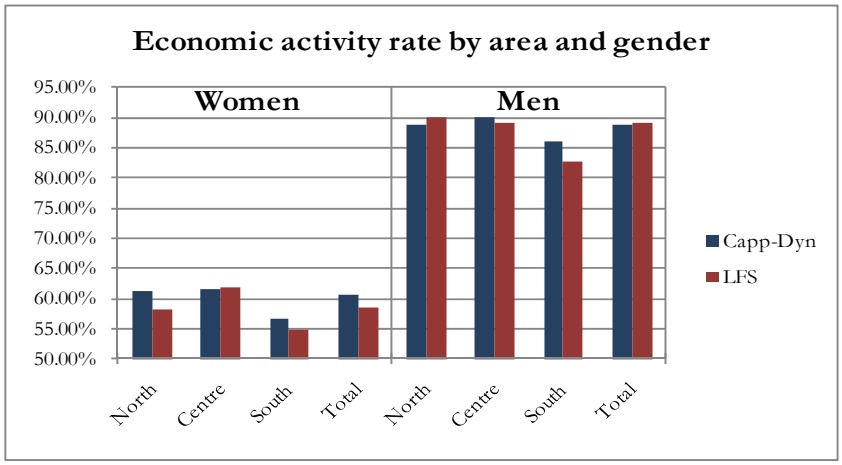


Figure 22
Employment rate for foreigners in the 15-64 age range

Figure 23
Unemployment rate for foreigners in the 15-64 age range

5. Income distribution in the initial CAPP_DYN population

5.1 Labour income

In this subsection we compare self-employment and employment income reported by CAPP_DYN with the information coming from 2006 National Accounts, released by Istat as time series data by institutional sector (financial and non-financial corporations, public administrations, family businesses and consumers, etc.). The relevant entries in National Accounts data are: *gross earnings* for employees (net of social contributions paid by employers); for self-employed workers income is the sum of different items. The first element is the *share of mixed income transferred by family businesses* to families of consumers, i.e. the share of profits from production that families allocated to consumption and savings, assuming that the entrepreneur transfers to his family the remaining profits after paying taxes and land rents, once the cost of debt and the amortization of plants and equipments have been covered. Family businesses are considered the units classified as partnerships and non-financial one-man businesses who occupy up to 5 employees. A second element consists of income received by members of *quasi-corporations*, that is, partnerships or cooperative firms with more than 5 employees. A third element consists in *profits distributed by companies*, including the remuneration of directors and auditors of limited companies. Table 35 illustrates the aggregate employment and self-employment income and the disposable income

after tax (the individual net income, in CAPP_DYN data) coming from the two sources.[43] CAPP_DYN and National Accounts values are made comparable by multiplying the CAPP_DYN amount by the ratio between the effective and the sample population. Income entries are reported gross of taxes and social contributions to be paid by workers, in line with National Accounts data.

Table 35

Aggregate income (in millions of euro at current prices). CAPP_DYN against National Accounts data (CN)

	CAPP_DYN	CN	CAPP_DYN/CN
Employment income	419,858	444,802	94.39%
Self-employment income	190,138	293,300	64.83%

Note: CAPP_DYN data are made comparable to CN data.

Employment income does not disagree much between the two sources, while for self-employed workers a higher difference can be observed.

A second possibility consists in making the two sources comparable by the number of workers for each of the two categories considered (employees or self-employed workers), not by the total number of individuals, as shown in Table 36. In this case the weight is computed as the ratio between the actual number of workers (employees or self-employed workers) according to the National Accounts and the corresponding CAPP_DYN information. This calculation could reduce the differences highlighted in Table 35, since some of them could, in effect, be caused by a different composition by category of workers with respect to the official figures.

Table 36

Aggregate income (in millions of euro at current prices). CAPP_DYN against National Accounts data (CN)

	CAPP_DYN	CN	CAPP_DYN/CN
Employment income	433,874	444,802	97.54%
Self-employment income	249,174	293,300	84.96%

Note: CAPP_DYN data are made comparable to CN data.

[43] With regard to self-employment income from National Accounts, the *share of mixed income transferred by family businesses* amounts to 185.65 billion euro, *income received by members of quasi-corporations* to 56.38 billion euro and *the other profits distributed by companies* to 51.27 billion euro.

In effect, for the two categories of workers considered the differences tend to diminish, especially for self-employed workers. The percentage values for the latter go from 65 to 85% (with a difference in absolute value that decreases from 103 billion euro to 44). In both cases employment income is close to National Accounts values (CAPP_DYN data constitute 94 or 98% of the CN aggregates, respectively in the first and in the second case), with a difference that ranges from 11 to 25 billion euro.

Some discrepancies probably emerge since in this case it is as if individuals with different amounts of income have the same weight in building the aggregate amount. With regard to this aspect, a distribution of weights could be generated (workers, for instance, are characterized by larger frequencies for low or medium levels of pre-tax income, between 7 and 30,000€), able to reflect the relationship between observed frequencies in the population and sample frequencies, for instance by income ranges.

Another part of the existing discrepancies may be due to a different classification of income. It is possible that some National Account income entries attributed to self-employed workers should actually be considered as unearned income. Another criterion can then be adopted to get the share of the operating profit produced by self-employed workers, namely to multiply the average income for employees by the number of self-employed workers. According to National Accounts data the average income for employees amounts to 35,000€ (precisely 34,529€, comprehensive of social contributions paid by employers). Multiplying this value by just over 7 million self-employed, the value reported in Table 37 can be obtained. In the first row the CAPP_DYN income is made comparable to the total population (the weight is computed as the ratio between the effective and the CAPP_DYN population), while in the second row the CAPP_DYN income is made comparable to the effective number of self-employed (the weight is computed as the ratio between the effective number and the CAPP_DYN number of self-employed workers).

Table 37

Aggregate income (in millions of euro at current prices). CAPP_DYN vs. National Accounts data (CN). Self-employment income

	CAPP_DYN	CN	CAPP_DYN/CN
Self-employment (relative to the population)	190,138	247,065	76.96%
Self-employment (relative to the number of workers)	249,174	247,065	100.85%

Note: self-employment income comparable to CN data through a weight based on the population or on the total number of self-employed workers.

The values obtained in this way are closer than before to National Accounts data, especially when the different number of self-employed workers is considered.

5.2 Pensions

Regarding the population pension recipients, we distinguish the following payments: disability support benefits ("pensione di invalidità civile"), social pension/allowances ("pensione/assegno sociale"), attendance allowances ("assegno di accompagnamento"), disability pensions and ordinary disability benefits ("pensioni di inabilità e assegni ordinari di invalidità"), pensions for accident at work ("rendite per infortunio sul lavoro"), war pensions ("pensioni di guerra"), survivors' pensions ("pensioni ai superstiti") and old age or retirement pensions ("pensioni di vecchiaia o anzianità"). For the following comparisons the benchmark is constituted by the information produced by Istat in cooperation with Inps about pensions paid and their respective recipients for the year 2006. First, we illustrate the distribution of the overall number of recipients by geographic area and sex (

Figure *24* and Figure 25).

Figure 24

Distribution of recipients of pensions and other security and welfare benefits by area

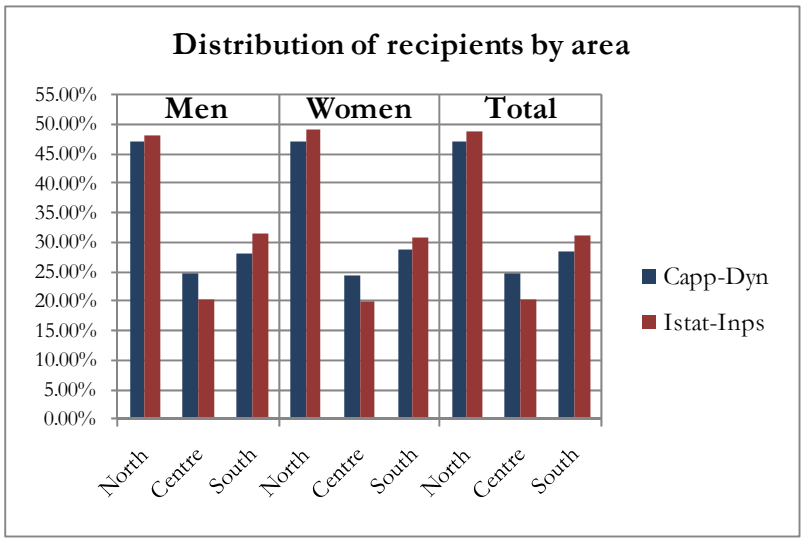
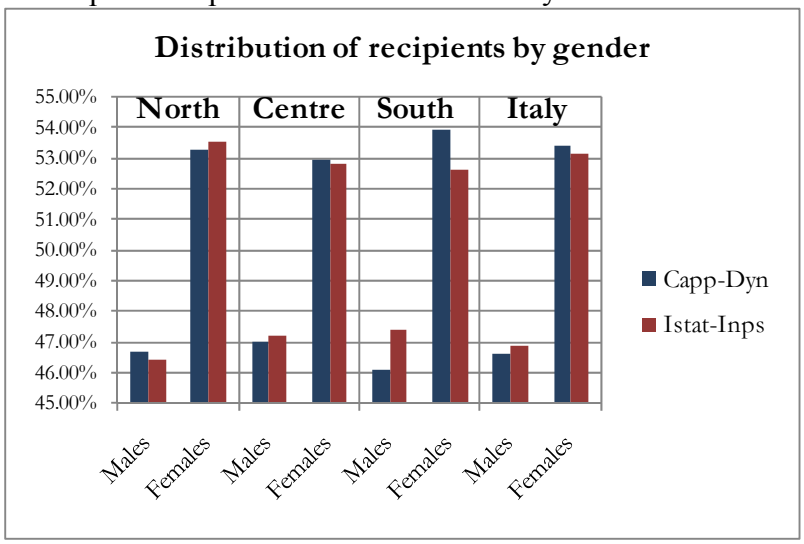


Figure 25

Distribution of recipients of pensions and other security and welfare benefits by gender



With respect to the breakdown by area (Figure 24), CAPP_DYN data are always below the frequencies derived from official statistics in the North and South (to a greater extent in the South, by about 3.5 percentage points for men and 2 for women), while the former over-represent the latter in the Centre (by about 4.5 percentage points for men, 4 for women). With respect, instead, to the distribution by gender, while in the North and the Centre the official statistics coincide almost perfectly with CAPP_DYN data, in the South some difference emerges, by about 1.3 percentage points for both sexes. We now try to understand whether the composition by type of payment of pension recipients obtained from CAPP_DYN data reflects the one derived from official sources. Table 38 shows the number of recipients of pensions and the percentage composition by type of payment.[44]

Table 38

Composition by type of recipients of pensions and other security and welfare benefits

	Number		Percentage	
	CAPP_DYN	Istat-Inps	CAPP_DYN	Istat-Inps
Old-age/retirement	9,778	10,789,819	51.37%	49.00%
Survivors	3,787	4,771,202	19.89%	21.67%
Social pension/allowance	637	775,197	3.35%	3.52%
Disability pension	1,764	1,911,168	9.27%	8.68%

[44] Regarding war pensions, the official figures on the number of recipients include survivors' benefits as well (received by family members of the original deceased recipient). As the number of recipients and the number of benefits do not differ significantly (i.e. every recipient usually receives a single war pension), information on the number of benefits has been used to separate indirect treatments and to merge them instead with survivors' pensions.

(contributory)				
Pension for accident at work	865	842,917	4.54%	3.83%
Disability support benefits	805	949,728	4.23%	4.31%
War	106	133,641	0.56%	0.61%
Attendance allowances	1,293	1,848,122	6.79%	8.39%
Total	**19,035**	**22,021,794**	**100.00%**	**100.00%**

Sources: Istat-Inps data.

The last two columns show that the disagreement between the two sources is minimal, around 1 to 2% in all cases. The maximum difference is observed for survivors' pensions: Istat-Inps frequencies exceed the CAPP_DYN frequency by about 2 percentage points. For social pensions/allowances, disability support benefits and war pensions, the agreement between the two sources is almost perfect. Note that that the number and the percentage of recipients of survivors' pensions, of disability support benefits and of attendance allowances suffer from the fact that sample data do not report the number of recipients under age 16, because not subject to interview. This then leads to distortions in the frequency of recipients of this kind of benefits and, in consequence, also of the others. If the total number of recipients resulting from CAPP_DYN data is compared to the effective number (obtained from Istat-Inps data) scaling the former by a factor equal to the ratio between the effective and the sample population, a total of 21,503 million recipients in 2006 can be obtained, compared to 22,022 resulting from Istat-Inps data. Therefore there is an underestimate of 500,000 recipients.[45]

We now analyze the distribution by gender, area and age groups of recipients of each kind of security and welfare benefits.

First we consider the old-age and retirement pensions. Table 39 and Table 40 report the CAPP_DYN frequency distributions by area and gender compared to the benchmark.

Table 39
Distribution of recipients of old-age pensions by area

	CAPP_DYN	Istat-Inps
North	53.58%	55.76%
Centre	24.59%	19.89%

[45] Remember that an individual could receive more than one benefit or pension. Given that the total in Table 38 has been obtained as the sum of recipients of each type of benefit or pension, the value reported here is necessarily higher than the actual value, equal instead to 16,162 million individuals.

South	21.83%	24.35%
Total	100.00%	100.00%

Table 40
Distribution of recipients of old-age/retirement pensions by gender

	CAPP_DYN			Istat-Inps		
	Men	Women	Total	Men	Women	Total
North	52.01%	47.99%	100.00%	51.96%	48.04%	100.00%
Centre	54.87%	45.13%	100.00%	56.32%	43.68%	100.00%
South	59.42%	40.58%	100.00%	59.49%	40.51%	100.00%
Total	54.33%	45.67%	100.00%	54.66%	45.34%	100.00%

Regarding the distribution by area, the largest differences can be found in the Centre (especially) and in the South. Also with respect to the distribution by gender, the major discrepancies can be observed in the Centre, while in the other areas the CAPP_DYN frequencies coincide almost perfectly with the benchmark. On the whole, however, the differences compensate each other, so that the total differs minimally between the two sources. Figure 26 reports the two distributions by age for both genders. Five-year range groups are used, although the first age group with non-zero frequency is the 30-34 range for men belonging to the CAPP_DYN sample (men belonging to this group account for 0.04% of males receiving old-age/retirement pension), the 40-44 age group in all other cases (even for women belonging to the CAPP_DYN sample).

Figure 26
Distribution of recipients of old-age pension by age

Regarding men, the CAPP_DYN distribution is very similar to the benchmark. Some differences can be observed for women and especially for ages ranging from 60 to 84 years. The CAPP_DYN sample generally tends to under-represent the recipients from ages 75 to 84 and tends to over-represent those from 60 to 74.

We move now to survivors' pensions (both contributory and war pensions).

Table ***41*** illustrates the comparison between frequency distributions by geographic area. Also in this case the Centre displays the highest differences, amounting to 4.4 percentage points. As for old-age pensions, the CAPP_DYN distribution reports lower frequencies in the North and South, higher frequencies in the Centre with respect to the benchmark.

Table 41
Distribution of survivors' pensions recipients by area

	CAPP_DYN	Istat-Inps
North	45.77%	48.21%
Centre	24.80%	20.39%
South	29.42%	31.39%
Total	100.00%	100.00%

In the Centre are found also the largest discrepancies (2.4%) between distributions by gender (Table 42).

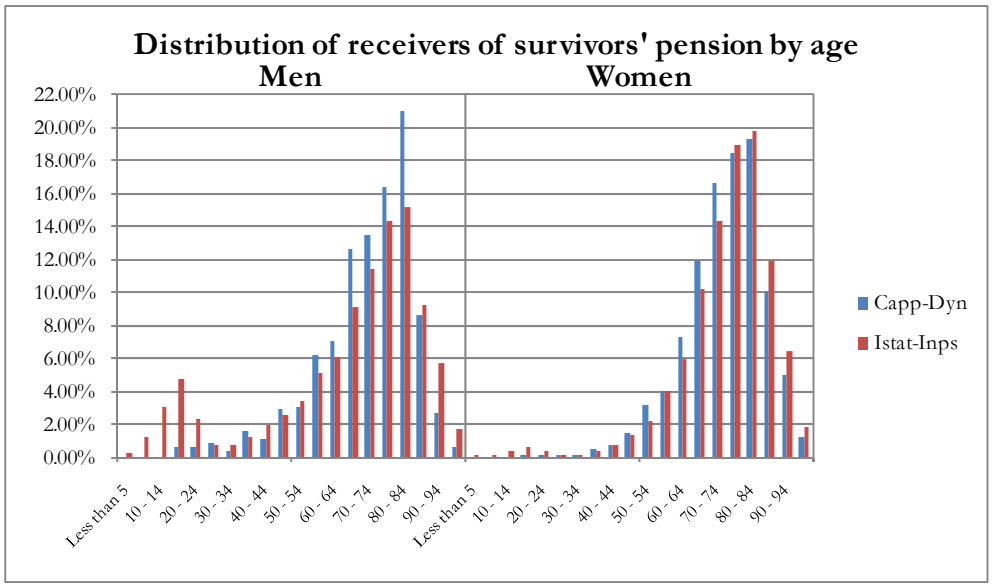
Table 42
Distribution of survivors' pension recipients by gender

	CAPP_DYN			Istat-Inps		
	Men	Women	Total	Men	Women	Total
North	11.77%	88.23%	100.00%	12.32%	87.68%	100.00%
Centre	9.80%	90.20%	100.00%	12.09%	87.91%	100.00%
South	14.00%	86.00%	100.00%	13.43%	86.57%	100.00%
Total	11.94%	88.06%	100.00%	12.62%	87.38%	100.00%

Regarding the distribution by age (Figure 27), here men show the largest discrepancies. Particularly, for early ages up to 24 years, CAPP_DYN data under-represent the effective frequency of recipients of survivors' pensions (with a maximum difference of 4 percentage points for the 15-19 age group). For ages up to 15 years, as already stated at the beginning of this paragraph, this is due to the exclusion from the population of recipients of this kind of payments of individuals younger than age 16. From 25 years onwards and up to

84 usually the contrary occurs (with some minor exceptions). The biggest differences are found in the 65-69 and 80-84 age ranges, with differences of 3.5 and 5.8 percentage points respectively. Again from 85 years onwards benchmark frequencies lie beyond the CAPP_DYN ones, especially for the 90-94 age group. For women the trend is similar to what has been seen for men, although the disagreement is limited: the largest differences occur for the 65-69 (1.8%), 70-74 (2.3%) and 85-89 (1.9%) age ranges.

Figure 27
Distribution of recipients of survivors' pension by age



With respect to the holders of social pension/allowances and to the distribution by area (Table 43), the differences between the two distributions are fairly limited. In the Centre, this time, an almost perfect coincidence can be found between CAPP_DYN frequencies and the benchmark.

Table 43
Distribution of recipients of social pension/allowances by area

	CAPP_DYN	Istat-Inps
North	29.04%	27.90%
Centre	20.88%	20.68%
South	50.08%	51.42%
Total	100.00%	100.00%

With respect to the distribution by gender (

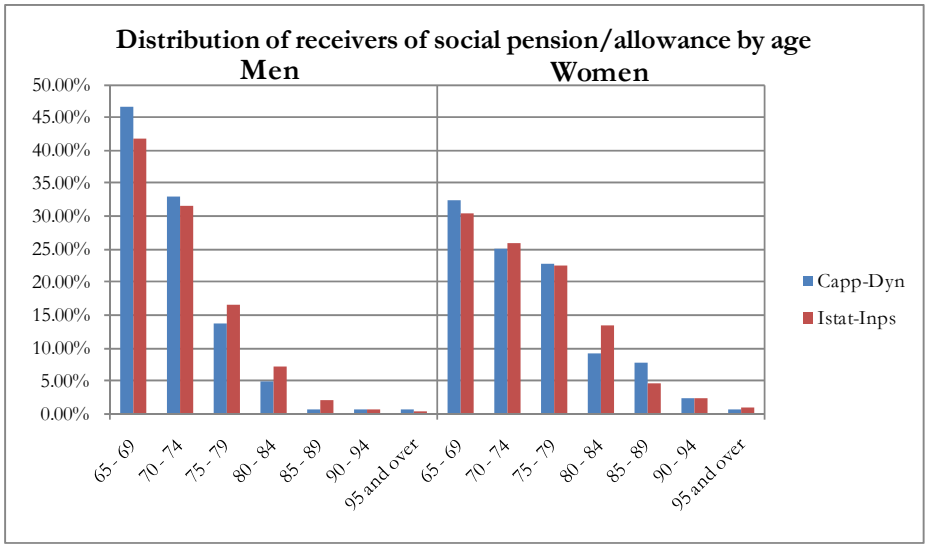
Table ***44***), men are under-represented in CAPP_DYN data: in the South the difference amounts to 5.7 percentage points, in the Centre to 1.3 points and in the North to 3.3 points.

Table 44
Distribution of recipients of social pension/allowances by gender

	CAPP_DYN			Istat-Inps		
	Men	Women	Total	Men	Women	Total
North	22.16%	77.84%	100.00%	25.42%	74.58%	100.00%
Centre	24.81%	75.19%	100.00%	26.15%	73.85%	100.00%
South	27.27%	72.73%	100.00%	33.01%	66.99%	100.00%
Total	25.27%	74.73%	100.00%	29.47%	70.53%	100.00%

Figure 28 compares the two frequency distributions by age groups for recipients of social pension or allowance. Obviously, the analysis includes only individuals aged 65 and over, since falling into this age group is one of the essential requirements in order to receive this kind of benefits. The major differences are observed for men, especially those belonging to the 65-69 (with a difference of 4,6%), 75-79 (with a difference of 3%) and 80-84 age groups (with a difference of 2,1%). For women, the 80-84 and 85-89 age groups present the major differences (respectively amounting to 4,1 and 3 percentage points).

Figure 28
Distribution of recipients of social pension/allowance by age



The following tables and figure refer to the population of recipients of disability pensions and of ordinary disability benefits. Regarding the distribution by gender (Table 45), in the Centre and in the South CAPP_DYN frequencies differ significantly from the

benchmark, respectively by 6.4 percentage points in excess and 5.2 points in default. Computing the difference, in the North sample frequencies are below the benchmark by 1.2 percentage points.

Table 45

Distribution of recipients of disability pension and ordinary disability benefits by area

	CAPP_DYN	Istat-Inps
North	29.71%	30.91%
Centre	28.40%	21.96%
South	41.89%	47.13%
Total	100.00%	100.00%

Referring to the distribution by gender (Table 46), if in the North and in the Centre the discrepancies are rather small, in the South they are higher, though not excessive (by 3.5%).

Table 46

Distribution of recipients of disability pension and ordinary disability benefits by gender

	CAPP_DYN			Istat-Inps		
	Men	Women	Total	Men	Women	Total
North	39.89%	60.11%	100.00%	38.92%	61.08%	100.00%
Centre	40.12%	59.88%	100.00%	40.95%	59.05%	100.00%
South	43.57%	56.43%	100.00%	47.02%	52.98%	100.00%
Total	41.50%	58.50%	100.00%	43.18%	56.82%	100.00%

With regard to the distribution by age (Figure 29), only individuals belonging to the age groups from 20 years upward are considered, since this kind of benefit is subject to some contribution and minimum insurance requirements (5 years of contribution are necessary). CAPP_DYN data tend generally to over-represent the frequency of recipients belonging to age groups up to 69 years (with some exceptions, especially for the 55-59 age range) and from 85 years onwards. The most significant differences can be found for men, particularly in the 55-59, 80-84 and 90-94 age ranges, with discrepancies amounting respectively to 2%, 3. 2% and 1.6%. For women, instead, the greatest differences occur for the 65-69, 80-84 and 90-94 age groups, with discrepancies amounting respectively to 1.3%, 1.5% and 2%.

Figure 29
Distribution of recipients of disability pension and ordinary disability benefits by age

We now consider the holders of pension for accidents at work (not reversionary). As shown in Table 47, there exist wide differences in the CAPP_DYN distribution by geographic area of holders of pension for accidents at work: while in the North and in the South CAPP_DYN data under-represent the effective frequency (respectively by 5.3 and 3.2 percentage points), in the Centre CAPP_DYN over-estimates the effective frequency by 8.5%.

Table 47
Distribution of recipients of pensions for accidents at work by area

	CAPP_DYN	Istat-Inps
North	38.61%	43.92%
Centre	33.29%	24.80%
South	28.09%	31.28%
Total	100.00%	100.00%

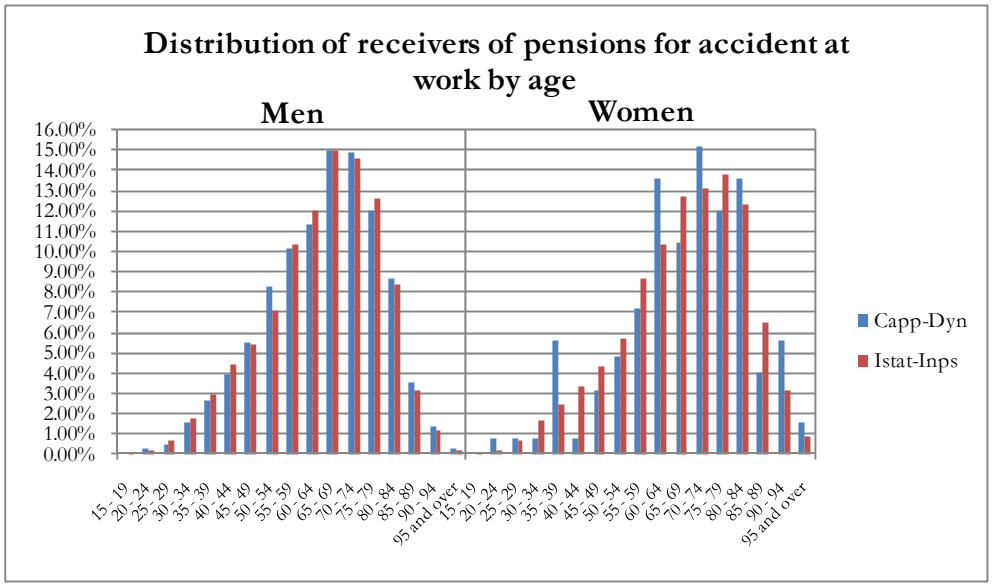
With reference to the gender distribution (Table 48), CAPP_DYN over-estimates in all geographic areas the effective number of men receiving pensions for accident at work, with a maximum difference in the South of about 12%. So far, then, this is the category of benefits that experiences the greatest mismatch with respect to official information.

Table 48
Distribution of recipients of pensions for accidents at work by gender

	CAPP_DYN			Istat-Inps		
	Men	Women	Total	Men	Women	Total
North	86.09%	13.91%	100.00%	75.03%	24.97%	100.00%
Centre	84.15%	15.85%	100.00%	75.70%	24.30%	100.00%
South	87.28%	12.72%	100.00%	75.65%	24.35%	100.00%
Total	85.98%	14.02%	100.00%	75.39%	24.61%	100.00%

Figure 30 illustrates the frequency distribution of recipients of pensions for accidents at work by age groups for the two sources. Only the recipients of direct pensions are considered (survivors' payments are included in the separate category illustrated above) and for this reason the first age group represented is the one ranging from ages 15 to 19. CAPP_DYN and official statistics agree well for men: the maximum difference of 1.2% is found in the 50-54 age range. Different evidence emerges for women: with the exception of the tails, the whole CAPP_DYN distribution is more irregular and differs significantly from the benchmark. An explanation for this lies in the limited number of observations, even in the central age groups of the distribution. The major deviations from the benchmark are observed for the 35-39 (with a difference of 3%), 40-44 (2,5%) and 60-64 age groups (3.2%).

Figure 30
Distribution of recipients of pensions for accidents at work by age



With reference to recipients of disability support benefits, Table 49 and Table 50 report the frequency distributions respectively by area and by gender. Remember that the difference between this kind of benefits and disability pensions is that the latter are paid out in favour of individuals with minimum contribution requirements, while the benefits

considered here have a purely welfare function. Table 49 shows that CAPP_DYN frequencies exceed the benchmark by 4.4 percentage points for the Centre, while the opposite occurs for the North (by 1.8%) and the South (by 2.7%). Once again, the most significant discrepancies are found for the Centre.

Table 49
Distribution of recipients of disability support benefits by area

	CAPP_DYN	Istat-Inps
North	28.94%	30.70%
Centre	22.11%	17.69%
South	48.94%	51.62%
Total	100.00%	100.00%

With regard to the distribution by gender, the most significant discrepancies are found in the South, with a CAPP_DYN frequency 5% below the benchmark for men. The second area with respect to the order of deviations is the North, with a difference of 2.8 percentage points.

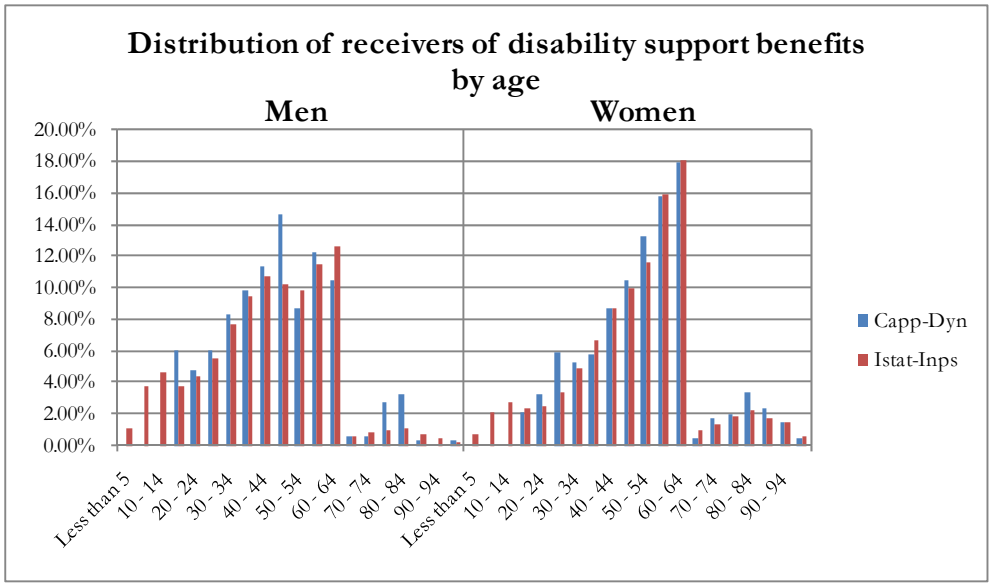
Table 50
Distribution of recipients of disability support benefits by gender

	CAPP_DYN			Istat-Inps		
	Men	Women	Total	Men	Women	Total
North	42.06%	57.94%	100.00%	44.90%	55.10%	100.00%
Centre	43.82%	56.18%	100.00%	42.32%	57.68%	100.00%
South	40.36%	59.64%	100.00%	45.30%	54.70%	100.00%
Total	41.61%	58.39%	100.00%	44.65%	55.35%	100.00%

With regard to the distribution by age (Figure 31), it must be pointed out that at age 65 the disability support benefit for handicapped and deaf-mutes changes into social allowances, while the benefits paid to blind persons do not change even after age 65. For this reason the figure shows a significant jump in the frequency distribution for the 65-69 age range. CAPP_DYN, instead, does not report recipients younger than 16, since they are not subject to interview. The younger ages, then, suffer from this problem, which necessarily affects also the other age groups. The frequency distributions differ more for males: the greatest discrepancies are found in the 5-9 (with a difference of 3.8%), the 10-14 (4.6%) and the 45-49 (4,5%) age groups, and concentrate mainly in the early age ranges. For women as

well, the first age bands show larger deviations, especially the 5-9 (2.1%), 10-14 (2,7%) and 25-29 (2.6%) age ranges.

Figure 31
Distribution of recipients of disability support benefits by age



With respect to war pensions, we refer only to direct pensions (the others, as before, are included in survivors' pensions). As can be noted in Table 51, the CAPP_DYN frequencies exceed the benchmark in the North by about 3.2 percentage points, while the opposite occurs in the Centre and in the South.

Table 51
Distribution of recipients of war pensions by area

	CAPP_DYN	Istat-Inps
North	40.57%	37.35%
Centre	27.36%	29.20%
South	32.08%	33.44%
Total	100.00%	100.00%

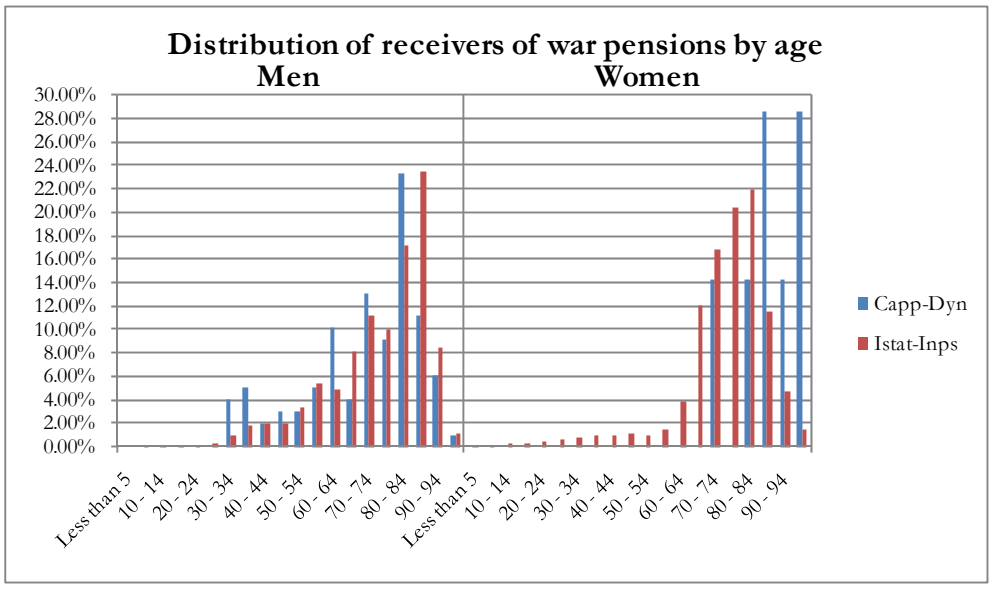
Regarding the gender distribution, in the North and in the Centre the CAPP_DYN data over-represent the effective percentage frequency for men by 5.7 and 5.3 percentage points respectively, while in the South CAPP_DYN under-represents the effective frequency by 2.4 points.

Table 52
Distribution of recipients of war pensions by gender

	CAPP_DYN			Istat-Inps		
	Men	Women	Total	Men	Women	Total
North	95.35%	4.65%	100.00%	89.69%	10.31%	100.00%
Centre	93.10%	6.90%	100.00%	87.82%	12.18%	100.00%
South	91.18%	8.82%	100.00%	93.57%	6.43%	100.00%
Total	93.40%	6.60%	100.00%	90.44%	9.56%	100.00%

Figure *32* illustrates that for this kind of benefits large differences exist between the two distributions, both for males and females. Starting from the left of the figure, for men the largest differences correspond to the age groups 60-64 (5.3%), 80-84 (6%) and 85-89 (12.3%). For women the age groups characterized by the greatest differences between CAPP_DYN frequencies and the benchmark are the 65-69 (12%), the 75-79 (20.3%), the 85-89 (17%) and the 95 and over (25.2%) age ranges. For them the CAPP_DYN distribution shows a bimodal pattern, compared to the unimodal pattern of the benchmark. Basically, CAPP_DYN reports war pensions only for women aged 70 and over, unlike the official statistics. But in this case too, both for men and for women, an important role in determining the differences is played by the small number of recipients.

Figure 32
Distribution of recipients of war pensions by age



Analyzing now the universe of recipients of attendance allowances, with regard to the distribution by geographic area (Table 53), the major differences are found in the North (about 8.5 percentage points), followed by the Centre (6.5%) and the South (2%).

Table 53
Distribution of recipients of attendance allowances by area

	CAPP_DYN	Istat-Inps
North	30.86%	39.40%
Centre	27.22%	20.66%
South	41.92%	39.94%
Total	100.00%	100.00%

Regarding the distribution by gender (

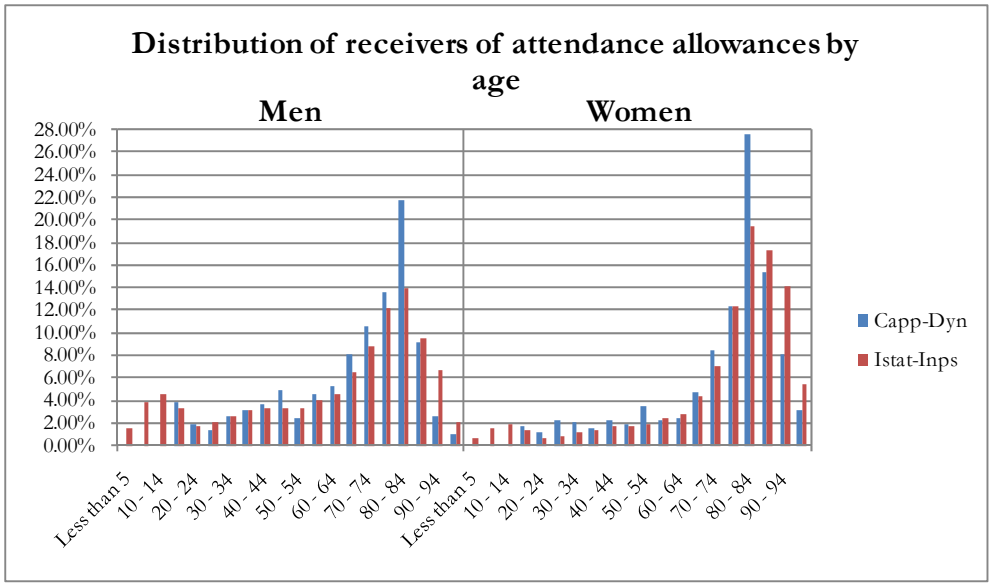
Table ***54***), in the South there is an almost perfect agreement between the CAPP_DYN distribution and the benchmark; in the other two geographic areas the discrepancies are around 2-3%.

Table 54
Distribution of recipients of attendance allowances by gender

	CAPP_DYN			Istat-Inps		
	Men	Women	Total	Men	Women	Total
North	35.34%	64.66%	100.00%	33.51%	66.49%	100.00%
Centre	32.67%	67.33%	100.00%	35.61%	64.39%	100.00%
South	39.67%	60.33%	100.00%	39.32%	60.68%	100.00%
Total	36.43%	63.57%	100.00%	36.27%	63.73%	100.00%

Regarding the frequency distribution by age (Figure 33), it must be reiterated that CAPP_DYN data do not include recipients aged less than 16, since they are not subject to interview. From ages 16 to 85 the benchmark generally exceed the CAPP_DYN frequencies, both for men and for women. From the age of 85 onwards, instead, the opposite holds. For men, the greatest differences are observed for the age groups 10-14 (with a difference of 4.5%), 80-84 (7.8%) and 90-94 (4%); for women, the ranges 80-84 (8.1%), 90-94 (6.1%) and 95 and over (2.4%).

Figure 33
Distribution of recipients of attendance allowances by age



We now analyze the total and average amounts of pensions and other security and welfare benefits paid. The benchmark is still represented by the Istat statistics prepared in collaboration with the National Institute of social security.

In the first two columns Table 55 reports the total amount for each category of payments. CAPP_DYN values are obtained by multiplying the total CAPP_DYN amount by the ratio between the effective number and the CAPP_DYN number of recipients. In the third and fourth columns the amount of each category of benefit is expressed as a percentage of the total. The last column shows the percentage ratio between the CAPP_DYN amount and the official figures.

Table 55

Total amount of pensions and other security and welfare benefits (millions of euro)

	CAPP_DYN	Istat	CAPP_DYN: % of total	Istat: % of total	CAPP_DYN/Istat
Old-age (or retirement) pensions	155,245	153,277	69.5%	69.0%	101.3%
Survivors' pensions	36,948	35,559	16.5%	16.0%	103.9%
Social pensions/allowances	3,221	3,505	1.4%	1.6%	91.9%
Disability pensions and ordinary disability benefits	11,650	13,203	5.2%	5.9%	88.2%
Pensions for accidents at work	3,137	3,087	1.4%	1.4%	101.6%
Disability support benefits	2,929	2,805	1.3%	1.3%	104.4%
War pensions	1,026	890	0.5%	0.4%	115.3%

Attendance allowances	9,356	9,766	4.2%	4.4%	95.8%
Total	**223,511**	**222,092**	**100.00%**	**100.00%**	**100.00%**

For each kind of pension or benefit, CAPP_DYN data report a total amount very close to the corresponding official statistic. In some cases, as can be seen from the first two or the last column, the differences are slightly larger: this is the case for instance for old-age (and retirement) pensions, for survivors' pensions and for disability pensions. Also in these cases, however, the deviation remains quite limited, maximum 2 billion euro. For old-age pensions, moreover, the difference does not weigh very much, given the high total amount of payments. Better performance is featured by social pensions and allowances, pensions for accidents at work, disability support benefits and war pensions. For the latter the ratio between the two sources is heavily influenced by the low amount of benefits. On the whole, the CAPP_DYN overall value of payments constitutes nearly 100% of the corresponding Istat-Inps amount.

Table 56 reports the average amount paid for each pension or benefit category. The behaviour of CAPP_DYN values with respect to official statistics is more or less the same as appears in the previous table (Table 55).

Table 56

Average amount of pensions and other security and welfare benefits

	CAPP_DYN	Istat	CAPP_DYN/Istat
Old-age (or retirement) pensions	13,741	14,206	96.7%
Survivors' pensions	7,756	7,453	104.1%
Social pensions/allowances	4,235	4,522	93.6%
Disability pensions and ordinary disability benefits	6,105	6,908	88.4%
Pensions for accidents at work	3,743	3,663	102.2%
Disability support benefits	3,084	2,953	104.4%
War pensions	7,677	6,659	115.3%
Attendance allowances	5,051	5,284	95.6%

The following tables illustrate the composition of each type of security and welfare benefit by area and gender. The values corresponding to each cell are obtained by multiplying the CAPP_DYN value by the ratio between the effective and the CAPP_DYN number of recipients.

Table 57

Total amount of old-age and retirement pensions (millions of euro)

	CAPP_DYN Males	Istat Males	CAPP_DYN Females	Istat Females	CAPP_DYN Total	Istat Total
North	60,066	57,181	28,297	27,261	88,363	84,443
Centre	22,891	23,067	9,351	9,801	32,242	32,868
South	24,110	25,154	10,242	10,812	34,352	35,966
Total	107,067	105,402	47,889	47,875	154,956	153,277

With respect to old-age and retirement benefits (Table 57), the most significant differences relate to men, especially those residing in the North of Italy. They are indeed responsible for almost all of the 1.7 billion euro of difference highlighted in Table 55. Also for women the largest discrepancies concentrate in the North.

Table 58

Total amount of survivors' pensions (millions of euro)

	CAPP_DYN Males	Istat Males	CAPP_DYN Females	Istat Females	CAPP_DYN Total	Istat Total
North	1,284	1,320	16,583	16,225	17,867	17,545
Centre	553	573	7,011	7,035	7,564	7,608
South	1,032	933	10,357	9,473	11,389	10,406
Total	2,869	2,827	33,952	32,732	36,820	35,559

In relation to survivors' pensions, women are mainly responsible for the overall discrepancy (in particular, for them the difference amounts to 1,219 out of the total difference of 1,261 billion). For a large part this affects the South (for about 890 million) in great part. For men, too, the South reports the highest difference, although limited.

Table 59

Total amount of social pensions/benefits (millions of euro)

	CAPP_DYN Males	Istat Males	CAPP_DYN Females	Istat Females	CAPP_DYN Total	Istat Total
North	253	270	655	723	908	993
Centre	164	202	461	532	626	734
South	513	576	1,168	1,202	1,681	1,778
Total	931	1,048	2,284	2,457	3,215	3,505

For social pensions and allowances (Table 59), the low total amount implies that the disagreement is quite limited compared to the previously illustrated benefits. Women are

characterized by major differences, although the evidence for men is not much dissimilar (the discrepancies for women and men amount respectively to 173 and 117 million euro).

Table 60

Total amount of disability pensions and ordinary disability benefits (millions of euro)

	CAPP_DYN Males	Istat Males	CAPP_DYN Females	Istat Females	CAPP_DYN Total	Istat Total
North	1,806	2,226	2,081	2,174	3,887	4,400
Centre	1,204	1,526	1,247	1,389	2,451	2,915
South	2,867	3,179	2,514	2,709	5,381	5,888
Total	5,877	6,930	5,842	6,273	11,719	13,203

With regard to disability pensions and ordinary disability allowances (Table 60), the overall difference of 1.4 billion euro concentrates to a greater extent among men, for 1 billion euro. The size of the discrepancy is not remarkable for any geographic area.

Table 61

Total amount of pensions for accidents at work (millions of euro)

	CAPP_DYN Males	Istat Males	CAPP_DYN Females	Istat Females	CAPP_DYN Total	Istat Total
North	1,225	1,191	150	153	1,374	1,344
Centre	623	652	84	91	707	743
South	1,008	907	82	93	1,090	1,000
Total	2,856	2,750	316	337	3,172	3,087

With regard to pensions for accidents at work (Table 61), men report the most significant deviations, especially in the South, with an overall difference of 106 million euro versus 22 million for women.

Table 62

Total amount of disability support benefits (millions of euro)

	CAPP_DYN Males	Istat Males	CAPP_DYN Females	Istat Females	CAPP_DYN Total	Istat Total
North	408	380	487	486	894	866
Centre	202	206	301	291	502	497
South	694	637	844	804	1,538	1,442
Total	1,303	1,223	1,632	1,582	2,935	2,805

Regarding disability support benefits (Table 62) no significant difference is observed between genders and the extent of the deviation is quite similar (80 vs. 50 million euro respectively for men and women). Again, given the limited total amount of this kind of benefits, also the overall deviation turns out to be small and as a whole amounts to 130 million euro.

Table 63
Total amount of war pensions (millions of euro)

	CAPP_DYN Males	Istat Males	CAPP_DYN Females	Istat Females	CAPP_DYN Total	Istat Total
North	317	290	16	28	333	318
Centre	335	219	16	25	351	244
South	319	309	10	18	329	328
Total	970	818	42	72	1,013	890

What we remarked above about the relationship between the magnitude of the discrepancy and the total amount paid is particularly relevant to war pensions (Table 63). The CAPP_DYN value exceeds the official value by 124 million euro, although for women the opposite occurs. Men, and particularly the Centre, are responsible for most of the difference.

Table 64
Total amount of attendance allowances (millions of euro)

	CAPP_DYN Males	Istat Males	CAPP_DYN Females	Istat Females	CAPP_DYN Total	Istat Total
North	870	1,274	1,940	2,581	2,809	3,855
Centre	513	718	1,005	1,318	1,518	2,035
South	944	1,501	1,729	2,375	2,673	3,876
Total	2,327	3,493	4,673	6,273	7,000	9,766

With regard to attendance allowances (Table 64), the major differences are observed for women, for about 1.6 billion euro compared to a 2.77 billion overall difference. For both sexes they are most in evidence in the North and the South.

6. Comparison between the IT-SILC and the SHIW income distribution

Since the previous version of the model was based on the Bank of Italy's Survey on the Household Income and Wealth (SHIW), we proceed in this section to a comparison

between the two sample surveys, IT-SILC and SHIW. A comparative analysis between IT-SILC 2005 and SHIW 2004 with respect to the administrative data and, when possible, other sources of survey information is contained in the previous report (Ministry of Social Affairs, 2008; Savegnago, 2008). In sum, it has been observed that family incomes were significantly different in the two surveys. In IT-SILC the average income for households was higher and more concentrated. Self-employment income was on average much lower in SHIW, in particular due to a much lower number of earners. Compared to the National Accounts aggregates, the 2005 IT-SILC survey was generally characterized by a greater agreement.

Similar results comparing income information were reported as well in Istat (2008: 162 et seq.). This document contains a comparative analysis between the pilot survey on living conditions Icv03 of 2003 (referring to income information in 2002) and SHIW 2002. In particular, it can be observed how the household incomes net of the imputed rents were on average higher in the survey on living conditions, but more concentrated. The Gini index was equal to 0.396 in Icv03, versus 0.373 in SHIW. The values up to the fifth percentile of the income distribution were lower in Icv03, while they exceeded the SHIW values for the subsequent percentiles.

Compared to the previous analysis, we re-propose some assessments, focusing on two major aspects: the distribution of pension transfers and the distribution of earnings.[46] Unlike in the previous paragraphs, the comparison with respect to SHIW is carried out using the original IT-SILC sample. The use of the initial CAPP_DYN population would require the construction of an initial population also on the basis of the SHIW 2006 sample, a task which goes beyond the purposes of this report. Consequently, the analysis is performed solely on the non-institutionalized population.

A more detailed discussion about pensions and other security and welfare benefits is justified by their importance in the dynamic model. It should also be noted that, unlike the previous report, the estimates here reported (updated now to 2006) do not use sampling weights and therefore highlight the possible problems of representativeness that could be present in the initial CAPP_DYN population.

Table 65 and Table 66 report the comparison between averages for the population aged at least 16 for the main types of pensions and benefits. For disability benefits, it may be

[46] For the evaluation of economic welfare, it should be appropriate to consider the household equivalent income. In this report we use individual income instead, since the unit of analysis of the dynamic model is primarily the individual and the purpose is to validate the individual variables. Differences in the household variables may also reflect a different household composition that we do not control for at this stage.

observed how SHIW reports values significantly lower than the IT-SILC estimates.[47] In particular, SHIW shows lower values for the first three types of pensions, while it reports a higher value for disability support benefits, especially for the higher age groups. A possible explanation may be linked to errors in classifying the correct type of benefit (misreporting).[48]

For contributory pensions, the values seem to be quite similar, even if the average SHIW value exceeds the IT-SILC average by about 76€. This difference is offset by the survivors' pensions, whose average is higher in IT-SILC. Finally, the column reporting social pensions illustrates an inclination of the SHIW respondents to report (probably) incorrectly the nature of the benefit, that is not found instead in IT-SILC: for social pensions/allowances cannot be received by individuals below age 65.

Table 65

Average annual income from pensions and other security and welfare benefits for all individuals aged 16 and over, by age group

Age group	Contributory disability benefits		Pensions from accident		War pensions		Disability support benefits	
	SHIW	IT-SILC	SHIW	IT-SILC	SHIW	IT-SILC	SHIW	IT-SILC
16-50	34	26	14	16	1	5	30	48
50-54	96	144	24	60	0	8	31	73
55-59	182	195	45	71	8	12	18	93
60-64	141	246	29	103	31	24	77	113
65-69	124	277	42	141	7	7	78	4
70-74	223	385	70	180	4	60	100	11
75-79	264	793	72	205	10	78	111	22
80 and over	345	1479	52	255	38	72	302	61
Total	110	226	30	72	7	18	61	53

Sources: SHIW 2006 and IT-SILC 2007 data processing.
Note: in this table sample weights have not been used. The IT-SILC sample has been used after the correction reported in paragraph 3.5.

[47] In reconstructing the amounts of pensions and benefits in SHIW, we used the file "allb5.dta", taking into account both the information on the pension type and that on the Institution paying the pension. The benefits reported in the category "inabilità/invalidità (previdenziale)" (that refer to disability benefits paid to individuals with minimum working and contributory requirements) are identified as contributory disability pensioners if the social security institution is INPS, INPDAP or the State, while they are identified as pensioners for accidents at work if the paying institution is INAIL. War pensions are already specified in the questionnaire, as well as the disability support benefits. However we reclassified the latter as pensions for accidents at work in cases where the paying institution is INAIL.

[48] For both samples, we excluded attendance allowances from disability support benefits. In SHIW they should be reported separately, in the entry "economic assistance" (Annex B6). However, in the absence of any reference to these allowances in the questions about pensions and benefits, it is possible that interviewees reported them within the "disability support benefits", contributing to the over-estimation observed in the tables.

Table 66

Average annual income from pensions and other security and welfare benefits for all individuals aged 16 and over, by age group

Age group	Old-age/retirement pensions		Survivors' pensions		Social pensions/allowances	
	SHIW	IT-SILC	SHIW	IT-SILC	SHIW	IT-SILC
16-50	13	10	23	45	4	0
50-54	424	437	182	252	7	0
55-59	3,884	3,928	206	307	4	0
60-64	8,364	8,153	353	584	54	0
65-69	9,048	9,024	663	1,031	225	281
70-74	8,246	7,971	1,349	1,608	285	276
75-79	7,608	7,129	1,787	2,127	266	249
80 and over	6,369	5,197	2,855	3,323	344	173
Total	3,064	2,699	477	575	78	58

Sources: SHIW 2006 and IT-SILC 2007 non-weighted data processing. The IT-SILC sample used takes into account the correction reported in paragraph 3.5.

Table 67 and Table 68 illustrate the distribution of some security and welfare benefits in the IT-SILC and SHIW sample population and the picture provided by the official ISTAT-INPS statistics (Istat, 2009b).

Table 67

Percentage of individuals aged 16 and over receiving pensions and other security and welfare benefits, by age group

Age	Contributory disability benefits			Pensions from accident			War pensions			Disability support benefits		
	IST.	SILC	SHIW	IST.	SILC	SHIW	IST.	SILC	SHIW	IST.	SILC	SHIW
16-50	0.3%	0.4%	0.6%	0.5%	0.5%	0.2%	0.0%	0.1%	0.0%	1.5%	1.6%	0.7%
50-54	2.0%	2.3%	1.5%	1.7%	1.9%	0.6%	0.1%	0.1%	0.0%	2.7%	2.4%	0.7%
55-59	3.3%	2.6%	2.2%	2.5%	2.3%	1.0%	0.2%	0.1%	0.1%	3.4%	3.0%	0.4%
60-64	3.8%	3.6%	2.5%	3.3%	3.2%	0.7%	0.3%	0.3%	0.1%	4.5%	3.6%	1.9%
65-69	4.1%	4.1%	2.3%	4.1%	3.9%	0.8%	0.6%	0.1%	0.1%	0.2%	0.1%	1.5%
70-74	7.2%	6.4%	3.7%	4.9%	4.8%	1.3%	1.2%	0.5%	0.1%	0.4%	0.3%	1.5%
75-79	13.7%	13.1%	4.6%	5.2%	4.7%	1.5%	2.2%	0.4%	0.3%	0.6%	0.7%	1.9%
80+	26.8%	27.7%	5.9%	4.9%	5.0%	1.0%	7.2%	1.8%	1.0%	1.4%	1.7%	4.9%
Tot	3.8%	3.8%	1.8%	1.9%	1.9%	0.6%	0.7%	0.2%	0.1%	1.7%	1.7%	1.2%

Sources: SHIW 2006 and IT-SILC 2007 data processing.

Note: IST. Indicates the administrative Istat-Inps statistics. In this table sample weights have not been used. The IT-SILC sample has been used after the correction reported in paragraph 3.5.

Comparing the percentage of recipients of disability pensions in Table 24, it can be observed how the 2007 IT-SILC sample, as adjusted through the procedure described in the previous paragraphs, is more consistent with the Istat-Inps statistics on pension recipients

than SHIW. Savegnago (2008: 34) reaches the same conclusion for different waves and taking into account also declared income. The unique category that is significantly underestimated in both surveys is that of recipients of war pensions. Regarding this aspect, it should be noted that the Istat-Inps percentages also take into account war pensions paid to survivors of the original recipients, i.e. indirect pensions (Istat, 2009b: 82), not considered in the IT-SILC and SHIW war pension variables and that, according to official information, amount to 63.8% of the total. If we exclude these benefits, the share of recipients in the total population aged 16 or over amounts to 0.26%, a value significantly close to the 0.23% estimated in IT-SILC.

Even with regard to occupational and survivors' pensions (Table 68) the agreement of IT-SILC with Istat data seems to be greater than the SHIW survey. The fact that SHIW tends to overestimate the number of occupational pensioners and to underestimate the number of recipients of survivors' pensions may be due to *misreporting* of respondents, who may tend to confuse the two types of pensions. The same problem may explain the differences emerging for social pensions/allowances for individuals younger than 65.

Table 68

Percentage of individuals aged 16 and over receiving pensions and other security and welfare benefits, by age group

Age	Old-age/retirement pensions			Survivors' pensions			Social pensions/allowances		
	ISTAT	IT-SILC	SHIW	ISTAT	IT-SILC	SHIW	ISTAT	IT-SILC	SHIW
16-50	0.1%	0.1%	0.1%	0.8%	0.8%	0.3%	0.0%	0.0%	0.1%
50-54	2.8%	2.9%	3.0%	2.9%	3.3%	2.1%	0.0%	0.0%	0.1%
55-59	24.0%	28.3%	27.2%	5.1%	4.5%	2.8%	0.0%	0.0%	0.1%
60-64	61.1%	66.4%	62.2%	8.6%	8.3%	5.5%	0.0%	0.0%	0.8%
65-69	75.8%	76.7%	71.3%	14.2%	14.3%	9.5%	7.9%	7.1%	4.0%
70-74	74.2%	74.1%	70.3%	22.9%	22.5%	16.9%	7.4%	6.1%	5.2%
75-79	68.3%	68.4%	68.9%	34.4%	31.8%	23.3%	6.5%	5.9%	5.0%
80+	55.1%	52.2%	62.4%	55.0%	50.5%	36.8%	4.4%	3.9%	5.5%
Total	21.9%	23.5%	25.1%	8.9%	8.5%	6.3%	1.5%	1.4%	1.4%

Sources: SHIW 2006 and IT-SILC 2007 data processing.
Note: ISTAT indicates the administrative Istat-Inps statistics. In this table sample weights have not been used. The IT-SILC sample has been used after the correction reported in paragraph 3.5.

The next analysis refers to the population aged at least 16 at the end of the reference period for income variables (2006 for both surveys), since the income information contained

in IT-SILC about children is rather limited.[49] By "earners" we always mean individuals with a positive income, not taking into account that a self-employed person could be an earner although receiving a negative or zero income.

In Table 69 and Table 70 we compare the annual income from pensions and related benefits in the two samples for individuals at least 16 years of age.[50] The estimated averages for the entire reference population are quite close for the two samples, although the use of sampling weights changes the sign of the difference: for, without using these, the average income coming from the SHIW survey is higher. The share of recipients does not differ significantly between the two surveys. Considering only recipients instead, it can be observed that income, in the columns referring to Italy as a whole, tends to be more dispersed in the IT-SILC sample. The first percentiles tend to be lower than the respective SHIW values, while the 99th percentile in IT-SILC becomes higher. It should be noted that for pensions differences seem to be more pronounced when using weights.

Table 69

Distribution of annual income from pension and related benefits in SHIW and IT-SILC; only individuals aged at least 16 years

	SHIW, WITHOUT WEIGHTS				IT-SILC, WITHOUT WEIGHTS			
	North	Centre	South and Islands	Italy	North	Centre	South and Islands	Italy
Mean	4,361	4,437	2,930	3,878	4,184	4,002	2,883	3,736
Standard Deviation	7,135	6,904	5,535	6,608	7,112	7,450	5,641	6,801
Share of recipients	34.5%	38.6%	28.8%	33.4%	35.3%	34.4%	29.7%	33.4%
Percentile	Income distribution by percentiles only for recipients							
1st	2,328	2,064	2,400	2,314	1,105	988	754	936
25th	7,982	6,500	6,045	6,760	7,072	6,604	5,564	6,370
50th	11,700	10,400	9,000	10,400	10,784	9,841	7,970.5	9,750
75th	15,600	14,400	13,000	14,690	15,015	14,560	12,369.5	14,366
99th	35,000	35,750	28,600	33,900	37,492	41,509	31,450	36,374

Sources: SHIW 2006 and IT-SILC 2007 data processing taking into account the correction reported in paragraph 4.

[49] In reality the IT-SILC individual questionnaire is completed for all individuals with minimum age of 15 years at the end of the income reference period. We chose to raise the threshold to 16 to minimize the differences between the two surveys.

[50] Also in this case the age of the individual in IT-SILC is reduced by one year. See paragraph 3.5 for the explanation.

Table 70

Comparison between IT-SILC and SHIW annual income from pensions and related benefits; only individuals aged at least 16 years

	DIFFERENCE IT-SILC - SHIW, WITH WEIGHTS				DIFFERENCE IT-SILC - SHIW, WITHOUT WEIGHTS			
	North	Centre	South and Islands	Italy	North	Centre	South and Islands	Italy
Mean	305	160	189	242	-176	-434	-46	-141
Standard Deviation	239	1,060	489	500	-23	546	106	193
Share of recipients	4.2%	-0.1%	1.5%	2.4%	0.8%	-4.2%	0.9%	0.0%
Percentile	Income distribution by percentiles only for recipients							
1st	-526	-1,109	-1,308	-806	-1,223	-1,076	-1,646	-1,378
25th	-702	156	-286	-91	-910	104	-481	-390
50th	-884	-338	-434	-624	-917	-559	-1,030	-650
75th	-676	840	70	13	-585	160	-631	-324
99th	2,141	4,264	5,564	3,497	2,492	5,759	2,850	2,474

Sources: SHIW 2006 and IT-SILC 2007 data processing taking into account the correction reported in paragraph 4.

The analysis of individual annual earnings reveals two main differences between IT-SILC and SHIW:

- Average incomes tend to be higher in IT-SILC, as well as the number of earners. This is particularly true for self-employment;
- SHIW seems to underestimate the number of individuals receiving income from both employment and self-employment.[51]

The average income for employees (Table 71 and Table 72) appears significantly higher in IT-SILC, as well as the share of recipients. Again, there is a greater dispersion of incomes within the population of earners. Weights seem to affect less the difference between the two surveys. In particular, without them the difference in the share of wage earners in the North, in the Centre and Italy overall increases.

[51] In SHIW individual incomes are obtained from the file "rper06", from variables "tp1" (income from pension without arrears), "yl" (employment income) and "ym" (net income from self-employment). In IT-SILC the individual income entries included in the file "cit07p" have been used. To take into account the diversity in the construction of variables, to earnings (variables "yaut" and "ydip") and pensions we added family allowances ("dass_e*numass", "aass_e*meassa", "assfam_e*massfam", "pass_e *npass"), since reading the SHIW questionnaire reveals that they are included in the respective income entries (see Annex B1 of the SHIW questionnaire, question 7.4). In order to

Table 71

Distribution of annual income from employment in SHIW and IT-SILC; only individuals aged at least 16 years

	SHIW, WITHOUT WEIGHTS				IT-SILC, WITHOUT WEIGHTS			
	North	Centre	South and Islands	Italy	North	Centre	South and Islands	Italy
Mean	6,511	5,519	4,205	5,505	7,345	6,711	5,236	6,538
Standard Deviation	9,522	9,639	7,676	9,005	11,207	10,666	9,264	10,547
Share of recipients	39.5%	33.5%	29.1%	34.6%	43.4%	41.1%	35.9%	40.5%
Percentile	Income distribution by percentiles only for recipients							
1st	1,500	1,200	1,500	1,400	130	117	130	129
25th	12,600	11,900	10,000	12,000	10,788	9,940	6,600	9,376
50th	15,500	15,000	14,500	15,000	16,116	15,544	14,303	15,575
75th	19,500	19,500	18,000	19,000	21,583	21,059	20,343	21,138
99th	45,000	43,800	40,000	45,000	55,521	55,070	48,483	53,898

Sources: SHIW 2006 and IT-SILC 2007 data processing taking into account the correction reported in paragraph 4.

Table 72

Comparison between IT-SILC and SHIW annual income from employment; only individuals aged at least 16 years

	DIFFERENCE IT-SILC - SHIW, WITH WEIGHTS				DIFFERENCE IT-SILC - SHIW, WITHOUT WEIGHTS			
	North	Centre	South and Islands	Italy	North	Centre	South and Islands	Italy
Mean	464	669	1,225	781	834	1,193	1,031	1,033
Standard Deviation	1,785	264	1,850	1,416	1,686	1,027	1,589	1,542
Share of recipients	0.7%	5.2%	7.3%	3.9%	3.9%	7.5%	6.7%	5.8%
Percentile	Income distribution by percentiles only for recipients							
1st	-1.367	-1,862	-1,366	-1,366	-1,370	-1,083	-1,370	-1,271
25th	-1,328	-1,742	-2,926	-2,400	-1,812	-1,960	-3,400	-2,625
50th	870	76	-207	640	616	544	-197	575
75th	2,254	1,711	2,580	2,299	2,083	1,559	2,343	2,138
99th	13,984	9,160	11,269	9,922	10,521	11,270	8,483	8,898

Sources: SHIW 2006 and IT-SILC 2007 data processing taking into account the correction reported in paragraph 4.

reconstruct a pension variable consistent with SHIW, we summed all pensions and related benefits reported in tables 24 and 25 plus supplementary pensions ("pminte*pinte_e+pintun_e").

The most significant difference is observed for self-employed income (Table 73 and

Table ***74***). In particular, the difference between averages for the entire population is considerable and, expressed as a percentage, amounts to 45.9%. The main reason is the large difference in the percentage of receivers of self-employed income, which is equal to 16.5% in IT-SILC vs. the 9.2% estimated using SHIW.

Table 73

Distribution of annual income from self-employment in SHIW and IT-SILC; only individuals aged at least 16 years

	SHIW, WITH WEIGHTS				IT-SILC, WITHOUT WEIGHTS			
	North	Centre	South and Islands	Italy	North	Centre	South and Islands	Italy
Mean	2,491	2,161	1,205	1,976	3,475	3,085	1,868	2,882
Standard Deviation	12,680	16,556	8,228	12,323	12,119	11,225	8,236	10,854
Share of recipients	10.6%	9.4%	7.3%	9.2%	18.2%	17.7%	13.2%	16.5%
Percentile	Income distribution by percentiles only for recipients							
1°	600	400	700	560	210	262	376	275
25°	10,000	9,000	7,000	9,000	6,000	6,000	4,659	5,538
50°	17,500	15,000	12,000	15,000	14,198	13,000	10,000	12,533
75°	27,000	25,000	20,000	24,500	24,075	22,199	17,479	22,001
99°	130,000	150,000	90,000	130,000	105,084	86,470	83,403	96,790

Sources: SHIW 2006 and IT-SILC 2007 data processing taking into account the correction reported in paragraph 4.

Table 74

Comparison between IT-SILC and SHIW annual income from self-employment; only individuals aged at least 16 years

	DIFFERENCE IT-SILC - SHIW, WITH WEIGHTS				DIFFERENCE IT-SILC - SHIW, WITHOUT WEIGHTS			
	North	Centre	South and Islands	Italy	North	Centre	South and Islands	Italy
Mean	460	461	385	441	984	924	663	907
Standard Deviation	-3,458	-6,511	-1,460	-3,637	-561	-5,331	8	-1,469
Share of recipients	6.2%	7.3%	4.8%	5.9%	7.6%	8.3%	5.9%	7.3%
Percentile	Income distribution by percentiles only for recipients							
1°	-327	-80	-774	-170	-390	-138	-324	-285
25°	-4,000	-4,000	-2,538	-3,602	-4,000	-3,000	-2,341	-3,462
50°	-4,523	-5,000	-2,759	-2,868	-3,303	-2,000	-2,000	-2,468
75°	-3,000	-8,391	-3,000	-4,000	-2,925	-2,801	-2,521	-2,500
99°	-60,000	-42,598	-16,272	-25,000	-24,916	-63,530	-6,597	-33,210

Sources: SHIW 2006 and IT-SILC 2007 data processing taking into account the correction reported in paragraph 4.

The large difference in self-employment income between the two surveys is partly justified by a reduced presence of individuals declaring a positive self-employed income in SHIW, and seems partly to be a consequence of the reduced presence of individuals receiving income from both employment and self-employment in the survey carried out by the Bank of Italy. Table 75 illustrates this situation.

Table 75

Recipients of employment and self-employment income in IT-SILC and SHIW; only individuals aged at least 16 years

SHIW, WITH WEIGHTS				IT-SILC, WITHOUT WEIGHTS			
	Self-employment income		Total		Self-employment income		Total
Employment income	Non-recipient	Recipient		Employment income	Non-recipient	Recipient	
Non-recip.	56.81%	8.55%	65.35%	Non-recip.	46.94%	12.57%	59.51%
Recipient	33.98%	0.66%	34.65%	Recipient	36.56%	3.93%	40.49%
Total	90.79%	9.21%	100.00%	Total	83.50%	16.50%	100.00%

Sources: SHIW 2006 and IT-SILC 2007 data processing taking into account the correction reported in paragraph 4.

Like the analysis of pensions, the comparison of incomes between the two surveys does not enable definite conclusions about the superiority of either survey to be drawn. We therefore refer to the conclusions reported in the previous ministerial report. Looking at the aggregate income, that report (Savegnago, 2008: 31) highlighted a negative bias for employment income equal to 10.2% in SHIW 2004 and to 0.6% in IT-SILC 2005.[52] However the underestimate of income from self-employment, which is very high and negative for both surveys, was equal to 53.6% in IT-SILC vs. 66.9% in SHIW.

This difference was justified by a lower presence of people who declared receiving this type of income in SHIW, as also shown for 2006 in tables 31 and 32. The same conclusion held for transfers, which total for 2004 in SHIW (Savegnago, 2008: 31) was only 71.0% of the aggregate value reported in the National Accounts, against a 89.9% estimated using IT-SILC. Only for unearned income the distortion was larger in the survey carried out by Istat, with a total value equal to the 33.8% of the aggregate National Accounts value for SHIW, compared to a smaller 23.4% for IT -SILC.

In sum, the IT-SILC survey seems to be more reliable in describing the distribution of pensions, other security or welfare benefits and earned income, especially for income from self-employment. These variables play a key role in the dynamic model. While from the point of view of the demographic structure no significant differences seem to exist between the two surveys regarding their correspondence to the National Accounts data, the analysis of the economic well-being of respondents suggests that the IT –SILC survey is more accurate. This advantage, most probably due to the greater sample size and to the integration procedures with administrative data, justifies the choice of the research team to build the initial CAPP_DYN population by using the IT-SILC sample.

[52] By aggregate values we mean here the total amount in the whole sample population, shifted to the overall Italian population by directly using the sampling weights (in IT-SILC) or multiplying the sample values by a factor equal to the ratio between the Italian population and the number of observations (in SHIW). See Savegnago (2008: 20).

REFERENCES

Blundell, R., Dearden, L. and Sianesi, B. (2005) Evaluating the impact of education on earnings in the UK: models, methods and results from the NCDS. *Journal of the Royal Statistical Society, Series A* 168(3): 473–512.

Brandolini A., (1999) The distribution of personal income in post-war Italy: source description, data quality, and the time pattern of income inequality, Temi di discussione, n. 350, Banca d'Italia, Roma.

Caliendo M. and Kopeinig S. (2005), Some Pratical Guidance for the Implementation of Propensity Score Matching, *IZA DP N.1588*.

Cannari L., D'Alessio G., (1992), Mancate interviste e distorsione degli stimatori, Temi di Discussione, No. 172, Servizio Studi, Banca d'Italia, Roma.

European Commission. 2009. Employment in Europe 2009, ISSN 1016-5444.

Istat. 2007. Forze di lavoro – Media 2006. *Annuario*, n. 12

Istat. 2008. L'indagine europea dei redditi e le condizioni di vita delle famiglie (EU-SILC), *Metodi e Norme*, n. 37.

Istat. 2009a. Integrazione di dati campionari EU-SILC con dati di fonte amministrativa, *Metodi e Norme*, n. 38.

Istat. 2009b. Statistiche della previdenza e dell'assistenza sociale, Anno 2006, Volume II - I beneficiari delle prestazioni pensionistiche, *Annuari*, n.12, ISBN 978-88-458-1616-1.

Lindgren B. W. 1993. Statistical theory, 4th edition, Chapman & Hall, London.

Ministero della Solidarietà Sociale, (2008), Studio prospettico della povertà in Italia, Rapporto finale, Roma.

Savegnago M. 2008. Rappresentatività campionaria delle indagini Banca d'Italia e Silc, *CAPPaper,* n. 46.