The contribution of tax statistics for analysing regional income disparities in Italy*

by

PAOLO DI CARO

aDepartment of Finance; Ministry of Economy and Finance; Italy;
bPortsmouth Business School, University of Portsmouth; U.K.

Abstract

This paper provides new insights on the study of regional income disparities in Italy by using administrative microdata on individual tax returns. The main results of the paper can be listed as follows. It is documented that inequality in this country is important when looking at the distribution of income both between and within regions. High inequality levels are recorded particularly when comparing the distribution of income of women and young people across regions. From the decomposition of the redistributive effects of the Italian personal income tax, we find that tax credits act for redistribution mostly in the South and tax schedules in the Centre-North. The strengths and weaknesses of using tax files for analysing regional income disparities are also discussed.

Keywords: Tax statistics; Regional income disparities; Inequality measures.


Paolo Di Caro – Department of Finance, Ministry of Economy and Finance, Italy.
Via dei Normanni 5 – 00184 Rome – Italy;
tel.: +39-06-93836167; e-mail paolo.dicaro@mef.gov.it
(CORRESPONDING AUTHOR)


* I am grateful to the Editor-in-chief (Eva Sierminska), the Editors and three anonymous referees for comments and suggestions. I wish to thank Fabrizia Lapecorella for the insightful support. I am grateful to Massimo Bordignon, Andrea Brandolini, Michele Costa, Fernando Di Nicola, Ernesto Longobardi, Marco Manzo, Maria Teresa Monteduro, Jorge Onrubia, Ivica Urban, Achille Vernizzi and the participants to the 58th SIE Conference for helpful comments. I wish to thank SOGEI s.p.a. for the support with the data. The views expressed in the paper are those of the author and do not reflect those of the institutions of affiliation. The usual disclaimers apply.
1. Introduction

The Great Recession produced uneven economic and social consequences across places, by challenging the capacity of modern tax-benefit systems to achieve redistributive objectives. Recent comparative studies among OECD nations highlighted the presence of significant differences in income distribution in the regions of the same country (OECD, 2016). In the European Union, from 2007 to 2014, the Gini coefficient of disposable income increased by about 6.1% and 2.5% between and within Member States, respectively. At the end of the crisis, in the Thessaly region in Greece the individual disposable income in purchasing power parity was 35% lower than that registered at the beginning of the recession; whereas in the Dolnoslaskie region in Poland the individual disposable income increased by about 34% over the same period. In Italy, during the years of the Great Recession, individual disposable income was less equal, in terms of standard deviation, within (+0.87%) than between (-0.17%) regions. Spatial inequalities have continued to play a relevant role during the first years of the recovery (OECD, 2015).

In the past decade, the study of inequality issues has benefited from the use of tax administrative microdata for different reasons (Bourguignon, 2016). Administrative data are very valuable for performing public policy evaluations given that they cover larger sample sizes and present fewer problems with attrition, measurement error, and non-response than survey data (Card et al., 2010). The availability of detailed tax statistics at a regional level is important for analysing income disparities between and within regions, as well (Atkinson et al., 2017). Tax statistics, however, are influenced by some limitations such as the lack of information on family conditions and individuals with income below the tax threshold, and the influence of underreporting (Obersky et al., 2017). The latter aspect can be particularly relevant in Italy, where tax evasion registers one of the highest levels among developed countries and substantial regional differences exist (Schneider et al., 2015).

This paper uses administrative microdata on individual tax returns for the fiscal year 2014 elaborated at the Italian Ministry of Economy and Finance (MEF) in order to study regional income disparities in the Italian NUTS-1 macro-regions. Three are the main contributions of the work. First, we throw further light into regional inequalities in Italy by updating and integrating the findings of previous studies that mostly adopted survey data for different years (Morelli, 2016). Specifically, a set of inequality indexes is employed for describing the recent dynamics of regional income distribution between and within the Italian macro-regions. Second, we investigate particular profiles of inequality such as age
and gender (Jenkins and Van Kerm, 2014), which play a crucial role for understanding spatial income disparities in Italy (Checchi and Peragine, 2010). Moreover, we conduct an exploratory analysis on some factors that can contribute to explaining the distribution of inequality among the Italian macro-regions, with a particular interest on specific aspects that are currently debated in the empirical literature. Third, we decompose the effects of the Italian personal income tax (PIT) or IRPEF on redistribution and progressivity by assessing the impact of tax schedules and tax credits at a regional level. The latter research objective is motivated by the increasing importance of studying the redistributive capacity of the tax system from a regional perspective (Biswas et al., 2017).

The study of the Italian case is interesting for several reasons. In Italy, the historical North-South divide still remains a problem with deep economic and social differences registered at a regional level. The spatial dimension of income inequality in this country is of particular importance when looking at distributional aspects between and within regions (Jesuit, 2008). In Italy, moreover, the analysis of regional income disparities has been mostly based on survey data due to the limited availability of tax files (Ceriani et al., 2013). The redistributive capacity of the Italian PIT has been questioned on the basis of different elements (Verbist and Figari, 2014): for instance, the high number of tax expenditures in the Italian tax system has been criticised from a distributional perspective (Tyson, 2014). Reforming the IRPEF structure is an open issue for the Italian government and a primary topic in the public debate (Baldini et al., 2017).

The paper is structured as follows. Section 2 reviews the main research lines that are directly related to our work. Section 3 describes the data and provides an overview of the Italian context. Section 4 contains the empirical analysis. Section 5 presents the summary of the results and discusses some future research developments.

2. Related literature

This work is linked to the following research areas: the analysis of spatial inequality and the evaluation of the place-specific redistributive effects of personal income taxes by using tax statistics; and the study of regional income differences in Italy based on microdata. As for the first research area, tax records have been progressively used for investigating income disparities between and within countries and assessing the redistributive capacity of particular tax systems (Kopczuk and Saez, 2004). Using income data from tax returns for Canadian provinces, Alasia (2003) documented that in Canada the
rise in spatial inequality during the 1990s was mostly driven by within-province differences rather than between-province variations. Using tax files for Spain, Bonhomme and Hospido (2013) provided evidence on the role of the regional and gender dimensions for understanding inequality patterns in that country and the redistributive ability of the Spanish tax system. Hortas-Rico et al. (2014) applied a set of inequality measures to individual data on personal income tax returns in order to understand the spatial distribution of income among Spanish municipalities. Moser and Schnetzer (2017) studied the income-inequality nexus in Austria, that is, the relation between inequality and average income, by relying on individual wage tax data at the municipal level.

The usage of tax microdata for analysing income inequality issues presents both strengths and weaknesses (Atkinson and Brandolini, 2001). On the one side, tax statistics provide higher accuracy and coverage of regional observations with respect to survey data by increasing sample sizes (Perugini, 2014). In addition, data derived from tax records do not suffer from problems of sampling and designing that are present in surveys (Iacovou et al., 2012). They are also helpful for understanding the redistributive impact of the complex set of tax instruments (Chetty and Hendren, 2013). On the other hand, tax microdata show the following main limitations. They provide an incomplete description of individuals with incomes below the tax threshold. Information obtained from tax data can be influenced by the specific set of rules present in a given year (e.g. inclusion/exclusion of particular income categories from the taxable income); tax declarations are affected by underreporting that can originate from avoidance and/or evasion. Despite the presence of these shortcomings, tax statistics represent an additional tool for understanding income differences at a regional level (Koijen et al., 2014). Comparing the results obtained from tax records with those derived from different sources is crucial for improving the reliability of tax statistics for policy purposes (Di Nicola et al., 2015).

This paper also contributes to the literature studying regional income disparities in Italy by using microdata (Consolini and Donatiello, 2015). Table 1 reports a selected list of works that are related to our analysis. The chosen contributions differ with respect to the level of analysis, the period under observation, the measurement of inequality. Most of the works in table 1 are based on survey data that derive from different sources (EU-SILC, LIS, SHIW); the only exception is the contribution of Acciari and Mocetti (2012) that used

---

1 A complete discussion of the pros and cons of using tax statistics in distributional studies is a task outside the boundaries of the present paper. For a more detailed treatment, see Bakker (2012).
data on individual tax records for provinces. Some of the common findings in the existing literature are: the confirmation of North-South disparities, the relevance of regional patterns for explaining inequality in Italy, the importance of economic and social determinants of income disparities. Previous studies did not look at two aspects that we consider in this paper, namely the place-specific redistributive effects of the Italian personal income tax and the relationship between spatial inequality and factors such as preferences about inequality and quality of institutions.

Insert about here.

Table 1. Regional inequality in Italy, selected contributions.

3. Data description and the Italian context

3.1 The tax files 2014

To describe regional income disparities in the four Italian macro-regions (North-West, North-East, Centre, and South) individual observations obtained from a sample of tax administrative microdata for the fiscal year 2014 are used. The dataset provides information on 80,000 anonymised taxpayers that are equal to 0.2% of the total Italian taxpayers filling personal income tax returns. The dataset contains 56 variables that include information on: taxpayers (gender, macro-region, marital status, number of children); income (gross income, taxable income, total income divided by income source); gross and net tax liabilities; tax deductions and tax credits. In the empirical analysis, we take into consideration two tax deductions (main residence; pension contributions of self-employed and additional pension contributions) and five tax credits (dependent family members; employment, retirement, similar to employment; mortgage interests; home restructuring; interventions for energy savings). These are the main tax deductions and tax credits in the Italian PIT structure in terms of number of beneficiaries and share of income. The presentation of the tax files and summary statistics can be found in Di Caro (2017a).3

In addition, the tax records provide information on the ‘80 Euro bonus’, the main countercyclical tax policy adopted by the Italian government during the Great Recession (Baldini et al., 2015). The ‘80 Euro bonus’ is a refundable tax credit of 80 euro per month that is given to dependent workers with a gross income between €8,145 and €26,000; the

---

2 The focus on NUTS-1 regions is justified by data availability and the fact that macro-regions represent an appropriate level of territorial disaggregation in terms of sample size (Longford et al., 2012).
3 More information on the sample design and the availability of tax files for research can be found in Acciari (2016).
tax credit is due only if the gross tax liability is higher than tax credit for employment, retirement, and similar to employment. For taxpayers with gross income between €24,000 and €26,000, the amount of the tax credit is calculated on the basis of the following formula $EUR \{ \text{Total bonus} \times \left(26,000 - \frac{\text{taxable income}}{2,000}\right) \}$. In 2014, the total attainable amount of the ‘80 Euro bonus’ was equal to 640 euro given that it was introduced in May. The bonus has been confirmed in subsequent years.

*Insert about here.*

**Table 2. Comparison of average gross income, macro-regions.**

Table 2 compares the average gross income in the four Italian macro-regions as resulting from the EU-SILC data, the tax-benefit microsimulation dataset from the Ministry of Economy and Finance, and the tax files. The tax-benefit microsimulation dataset provides a better description of gross income at a regional level than EU-SILC and tax files by taking into account property income and income from financial activities (Di Nicola et al., 2017). Although the presence of some differences in the three datasets, T-test results do not reject the null hypothesis of the equality of standard deviations across regions, at 5% level of statistical significance. Tax records potentially covers all the Italian municipalities (about 8,000), whereas the EU-SILC allows for the coverage of about 800 municipalities. Moreover, if we look at the tax files, the ratio of the sampling standard error divided by the estimated income at a regional level is on average 0.5%. In the EU-SILC, this ratio is on average 4% (Ceriani et al., 2013); in the Bank of Italy’s SHIW database the ratio is 7% (Cannari and D’Alessio, 2003).

Tax records have the following main limitations. First, in 2014, taxpayers filling IRPEF returns counted for about two-thirds of the total Italian population, that is, individuals at the bottom of the income distribution, which normally do not fill tax returns because their income is below the tax threshold, are not completely covered in our study. When poor individuals not filling tax returns are concentrated in particular areas, like in the South of Italy where poverty levels are twice those observed in the Centre-North (D’Agostino and Scarlato, 2015), the regional patterns of inequality emerging from tax files provide a limited representation of true income disparities. Second, information on households that are relevant when dealing with redistributive issues is not taken into consideration (Coulter et al., 1992a). In the absence of information for reconstructing
households’ composition in the tax files, in the next section, we describe the distribution of income in the four Italian macro-regions by classifying taxpayers with respect to the number of dependent children. This choice, which does not overcome the issue of having individuals as unit of analysis, is motivated by the fact that information on children can be important when studying regional income disparities in Italy (D’Ambrosio, 2001). Third, when tax evasion is not uniformly distributed across regions, like in Italy where regions located in the South show levels of tax evasion higher than those observed in the rest of the country and evasion activities are particularly spread among small businesses and professionals (Carfora et al., 2017), true income disparities in the population can be different than those deriving from tax files where underreporting is not considered. In the subsection 4.3, we discuss the potential implications of these issues for our analysis and some possible ways forward.

### 3.2 Italian regional disparities

In Italy, one of the most relevant dimensions of income inequality is represented by geography. Southern regions have been characterised by poor economic performance, weak labour markets, and high incidence of poverty and inequality. Interestingly, income differences across Italy are relevant between and within regions, as well (Cerqueti and Ausloos, 2015). Figure 1 reports the regional distribution of the Gini index of pre-tax gross income obtained from the tax files for the year 2014. Observe that, Southern regions (dark grey) report higher values of the Gini index (0.48) with respect to Central and Northern counterparts (light grey) where the index is on average 0.45. Regional differences in the Gini index are significant, at 1% statistical level, after performing the ANOVA Test that rejects the null of equality. This implies that: income inequality is more relevant in the South than in the Centre-North; the distribution of income is less equal within the South than in the rest of the country.

**Figure 1. Regional distribution of the Gini index, tax files 2014.**

The Great Recession contributed to amplifying the differences in the distribution of individual and household income between and within regions. From 2007 to 2014, data from the Italian National Institute of Statistics (ISTAT), the Gini index registered different variations: North-West (-1%), North-East (-3%), Centre (0%), and South (+2.5%). The
recent crisis, moreover, produced asymmetric consequences on particular categories of individuals. Table 3 reports the variations in the employment rate for specific age and gender profiles observed during the Great Recession. In 2014, in the South, the employment rate of persons aged below 35 years was about 27% less than that recorded at the beginning of the recession. If we consider the employment rate of women in the total population, in the Southern regions, the drop in the crisis period was about four times higher than in the rest of the country. Regional differences are also observed when looking at total population and the age class 36-65.

*Insert about here.*

**Table 3. Regional variations (%) in the employment rate during the Great Recession.**

4. Inequality and the effects of IRPEF: regional results

4.1 Spatial patterns of inequality in Italy

This section investigates the spatial distribution of regional income disparities in Italy by combining graphical analysis with the application of a set of inequality indexes in the spirit of Jenkins (1999). The graphs in figure 2 report the frequency density function of gross income in the four Italian macro-regions. The variable gross income includes all taxable income subject to the Italian PIT plus the property income subject to the flat rate Cedolare Secca. The latter is a proportional tax with two rates (21% and 10%) that is applied to some income deriving from rented properties having specific features (Di Caro, 2017a). This variable represents the gross income in the tax files and it is used for determining the eligibility of most of tax credits within the IRPEF structure. Observe that, this definition of gross income does not include most of the income deriving from properties and that generating from financial activities and, consequently, the results hereafter presented have to be interpreted with caution given that the after-tax income considered here represents a subset of true disposable income (Figari et al., 2015). The uneven distribution regarding the initial conditions of Italian taxpayers living in the four macro-regions is confirmed after applying the ANOVA F Test on equality of the mean level, which rejects the null hypothesis of equality at 5% level of statistical significance.

*Insert about here.*

**Figure 2. Gross Income, macro-regional distribution.**
We first apply the Theil index to the tax data in our sample. The Theil index is a Generalized Entropy Measure that describes the degree of dispersion of a subject in a population with respect to a given variable (Cowell, 2003). This index is useful for studying spatial inequalities because it allows for the decomposition of regional income dispersion into two parts: the within component, the weighted sum of the within-region inequality index; and, the between component, the inter-regional difference in average income (Jenkins, 2006). In short, decomposing the Theil index has the merit of showing how intra- and inter-regional income disparities contribute to determining income inequality in Italy. In 2014, we calculated that the Theil index for the variables pre-tax gross income and net income was equal to 0.4143 and 0.3185: in both cases, the within-region component counted for about 95%.\(^4\) These findings, which can appear \textit{prima facie} surprising in a country like Italy where disparities across regions are relevant, support the view that the distribution of income within regions has a relevant influence on overall inequality, particularly in recent years (Bönke and Schröder, 2015). For instance, Hoffmeister (2009) found that since the second half of 1990s regional differences among NUTS-1 regions in the European Union were due to within-region variations for about 90%. As for the Italian case, using the EU-SILC data for the year 2014, Mussida and Parisi (2016) estimated that the within-region component of the Theil index counted for more than 92% of total inequality.

\textit{Insert about here.}

\textit{Figure 3. Concentration curves for gross income and tax, by macro-regions.}

The two graphs in figure 3 report the concentration curves for gross income (left-hand side) and net tax liabilities (right-hand side), gross tax liabilities minus tax credits, for Italy and the NUTS-1 regions. Concentration curves display the share of a variable of interest accounted for by cumulative proportions of individuals in the population ranked from poorest to richest (Bishop \textit{et al.}, 1992). In particular, the graphs are useful for mapping the presence of regional differences in the concentration/dispersion of pre-tax gross income and tax liabilities. The net tax liabilities, graph on the right, are more concentrated in the Southern regions than in the rest of the country. This result can derive

\(^4\) The results of the Theil index for pre-tax gross income are: Estimate (0.4143), Std. Err. (0.0149), z (27.77), P>|z| (0.000), 95% Conf. Interval [0.3850, 0.4435], within-group inequality GE(1) (0.3935), between-group inequality GE(1) (0.0208). The results of the Theil index for post-tax net income are: Estimate (0.3185), Std. Err. (0.0116), z (27.44), P>|z| (0.000), 95% Conf. Interval [0.2958, 0.3413], within-group inequality GE(1) (0.3025), between-group inequality GE(1) (0.0160).
from different reasons. Income differences among individuals living in the same area are more pronounced in the South than in the rest of the country as confirmed by the high value of the Gini index in Southern regions (figure 1). In the South, moreover, where the average income is lower than in the rest of Italy, the high concentration of tax credits that substantially influence redistribution in this area in relative terms (see section 4.5) produces effects on the concentration of net tax liabilities.

4.2 Specific profiles of regional disparities

This section provides evidence on some aspects of regional disparities in Italy. Table 4 reports the Gini coefficients of pre-tax gross income and net income for the observations contained in the tax files grouped on the basis of different criteria: region, gender, age, and number of children. The last column contains the difference (in %) between the two Gini coefficients by providing some initial insights on the redistributive effects of the Italian PIT. More precisely, remembering that the Gini coefficient varies from 0 (perfect income equality) to 1 (perfect income inequality), a reduction of the Gini coefficient after the application of the personal income tax can be interpreted as a decrease of income inequality among individuals. From the observation of the upper part of table 4, it can be noted that the Gini coefficient of pre-tax gross income and post-tax net income vary in the four Italian macro-regions. Interestingly, regional differences in the Gini coefficients remain almost unchanged after the application of the PIT by motivating further exploration on the place-specific consequences of IRPEF, as documented in section 4.5.

Insert about here.

Table 4. Gini Index for gross and net income, different breakdowns.

Tax statistics have been progressively adopted for analysing the place- and time-specific features of gender inequalities (Atkinson et al., 2016). Also, the availability of tax records is important for evaluating the distributional impact of gender-based tax reform proposals (Alesina et al., 2011). Despite the fact that there are large labour market differences for men and women in Italy (Colonna and Marcassa, 2015), from table 4, it can be observed that the Gini coefficients of gross and net income do not differ so much when looking at the national distribution of income from a gender perspective. The Gini coefficient of gross income is higher for women than men; the difference between the Gini coefficient of pre- and post-tax income is higher for men than women by suggesting that
the Italian PIT produces more redistributive consequences among the former category of individuals. However, if we look at the distribution of income among women from a regional perspective, and by adopting percentile ratios, some additional features emerge.

The share of women in the top 1% of the distribution of gross income is quite homogenous across places: North-West (10.30%), North-East (10.21%), Centre (11.72%), and South (9.70%). Yet, the share of women in the bottom 10% shows relevant spatial variations: 58.84% (North-West), 60.12% (North-East), 45.88% (Centre), and 41.13% (South). The latter finding is explained by two main reasons. First, in the South of Italy, women are at the margin of the labour market and they are mostly employed in irregular activities (Clementi and Giammatteo, 2014); consequently, they do not fill tax returns. Second, most of women living in the Southern regions, where the weight of family ties continues to be more pronounced than in the rest of the country (Alesina and Giuliano, 2010), are occupied in home production activities that are not subject to taxes and, again, they do not fill tax returns. Therefore, our evidence suggests that the study of the spatial patterns of gender inequality in Italy requests the combination of multiple indicators (Bozzano, 2012).

Investigating the age-specific patterns of income inequality is relevant when comparing inequality between countries (Sierminska and Takhtamanova, 2012). The age profile is viewed as a crucial element also for studying spatial income differences in the Italian regions (Fiorio, 2011). The two graphs in figure 4 report the concentration curves for the variable gross income in the four Italian macro-regions for the age class 0-35 years (left-hand side) and the age class 36-65 (right-hand side). The selection of the two specific age classes is due to the following reasons: the sample of tax statistics presents a breakdown in three age classes; the age class 65+ years is not considered because of it is mostly made up of retired taxpayers and it shows low spatial dispersion.

Income differences across geographical areas are less pronounced at the beginning of the life cycle: the standard deviation of gross income registered in the four regions is about €1,872.87 and €3,872.04 for the age class 0-35 and 36-65 years, respectively. One of the reasons can be the common trend observed in the Italian regions for which youths show difficulties in entering the labour market (Scarpetta et al., 2012). From the comparison of the Gini coefficients of gross and net income (table 4), we can note that the Italian PIT produces more redistributive consequences among persons aged above 35 years. This is a
direct consequence of the combination of the difficulties of young people of finding full-time dependent occupations and the IRPEF’s structure.

Insert about here.

**Figure 4. Concentration curves for gross income, by macro-regions and age classes.**

Differences in the households’ composition of individual taxpayers provide further insights into the asymmetric distribution of inequality (Koulovatianos et al., 2005). In particular, the results of the Gini coefficients reported in table 4 suggest that income is more equal among taxpayers with one child and/or two children than among those having no child and/or 3+ children; the redistributive impact of the Italian PIT increases with the number of children. The first aspect derives from the fact that most of Italian taxpayers filling tax returns have one child/two children; the second aspect is explained by the importance of the tax credits for dependent members within the IRPEF structure. The graphs in figure 5 report the density functions for the (log of) gross income in the four Italian macro-regions, where the individual taxpayers have been grouped according to the number of children (0, 1, 2, and 3+). Note that, the presence of fat tails in regional income distributions is in line with the findings in Mussida and Parisi (2016).

Some comments are worth pointing out. Spatial inequalities are relevant for taxpayers with a large number of dependent children: in the four macro-regions, the standard deviation of the gross income for taxpayers with 3+ dependent children is more than two-thirds higher than that registered for taxpayers with zero child. From the comparison of graphs in figure 5, we can note that the different number of dependent children produces more consequences on the distribution of income in Central and Northern regions rather than in the South. These descriptive findings seem to confirm the idea that understanding households’ composition and individual characteristics is crucial for analysing spatial income disparities in Italy (Papagni, 2006).

Insert about here.

**Figure 5. Kernel densities of gross income, by macro-areas and n. of children.**

4.3 **Interpretation of the results**

The set of evidence presented in the previous pages suggests that, in Italy, the distribution of income among individuals shows specific patterns both between and within
regions. In short, the discussion on inequality is important when comparing individuals living in different regions, by confirming the historical North-South disparities; but, it is also relevant when looking at the situation of individuals living in the same region. In addition, regional inequality patterns in this country are characterized by some featuring elements such as the gender and age profiles, and the influence of the different number of children. It is important to note that, however, the findings presented in this paper can be influenced by the specific limitations of tax files discussed beforehand (section 3.1). Consequently, the robustness of our results has to be checked with respect to these issues in order to provide a correct understanding of regional income disparities in Italy (Di Caro, 2017b).

The fact that we do not consider individuals with income levels below the tax threshold that are mostly located in Southern regions means that spatial differences described here can represent a lower bound for true regional income disparities present in Italy. Therefore, our results have to be read in combination with the findings of the contributions that used data with information on non-fillers (D’Alessio and Neri, 2015). Although the household dimension is not covered in our study (Coulter et al., 1992b), it is worth observing that: similar regional differences to those presented here have been documented when using data based on households (Pesole and Raitano, 2012); the breakdown for the number of children is useful for understanding the redistributive effects of the Italian PIT (Blangiardo and Vernizzi, 2013).

The lack of updated information on the individual distribution of tax evasion in the Italian regions supports the need of conducting further investigations on a micro level before identifying the precise effects of tax evasion on regional inequalities (Marino and Zizza, 2012). As for the spatial consequences of IRPEF (section 4.5), however, it is worth noting that our results show commonalities with those presented in the work of Albarea et al. (2015) where a microsimulation model taking into account tax evasion has been applied for studying the Italian PIT. To check for the influence of tax evasion on spatial income disparities in Italy, we are working for developing micro estimations of tax evasion by merging tax files with survey data; a research area that had positive results in other European countries (Paulus, 2015).
4.4 Mapping spatial differences that can influence regional inequality

Identifying the reasons behind the uneven distribution of income between and within countries and regions is one of the most important research themes in the inequality literature (Corak, 2013). In this section, our objective is much more limited, that is, we explore how specific factors, which have been progressively adopted for addressing inequality issues (Scheve and Stasavage, 2009), are distributed across Italy. In particular, we look at the spatial distribution of two variables: individual preferences about inequality and the quality of regional institutions. The interest for these specific variables is motivated by two main reasons. Institutional and social factors are gaining relevance in the empirical inequality literature given the need of finding explanations to income differences across places that are additional to economic determinants (Atkinson and Bourguignon, 2014). To our knowledge, these particular aspects received limited application to the study of regional income disparities in Italy when using microdata (Ballarino et al., 2012).

In a cross-country study, Kerr (2014) analysed the relationships between inequality and individual preferences regarding the support for redistribution and the degree of inequality acceptance. The author documented that high inequality in a given context is able to deteriorate individual concerns for redistribution, which can lead to a greater acceptance of inequalities, and, in turn, to the formation of a vicious cycle: the reduced social concern can amplify inequality. Although Kerr (2014)’s hypotheses need further empirical scrutiny from a regional perspective, we are interested in showing how preferences regarding inequality are distributed across Italy. As in Kerr (2014), we use information deriving from the International Social Survey Programme (ISSP) - the latest version refers to the year 2009 - which contains a breakdown for the Italian macro-regions.

Specifically, we use the categorical variable describing inequality acceptance at a regional level that derives from the answers to the ISSP’s question V28 ‘Large differences in income are necessary for (your country) prosperity?’ The variable takes values from 1 (strong agreement) to 5 (strong disagreement); high (low) values indicate low (high) degree of inequality acceptance. From the graph on the left in figure 6, we can note that Southern regions (dark values) report higher degree of inequality acceptance than the rest of the country. The significance of regional differences is confirmed after applying the ANOVA F Test on equality of the mean level, which rejects the null hypothesis of equality at 5% level of statistical significance.
As for the link between income disparities and the quality of institutions at a regional level, the presence of low levels of institutional quality has been associated to high inequality in a given place (Tabellini, 2010). This relationship can derive from the fact that low levels of institutional quality contribute to increasing transaction costs and rent seeking, and reduce mutual trust and reciprocal cooperation among individuals (Rodriguez-Pose, 2013). Consequently, places where institutions are of low quality can show reduced support for redistributive policies and high inequality (Kyriacou and Roca-Sagalés, 2014). To map the quality of institutions in the Italian regions we use the Index of Institutional Quality (IQI) elaborated for Italy by Nifo and Vecchione (2014). The IQI, which takes values from 0 (low quality of institutions) to 1 (high quality of institutions), contains information on five institutional dimensions that include variables obtained from different sources: voice and accountability; government effectiveness; regulatory quality; rule of law; corruption. From the graph on the right in figure 6, we can note that Southern regions (dark values) report higher degree of inequality acceptance than the rest of the country. The significance of spatial differences is confirmed after applying the ANOVA F Test.

From the comparison of figure 1 and graphs in figure 6, it can be noted that the geographical distribution of inequality acceptance and quality of institutions suggests that both variables can be relevant for the understanding of regional inequalities in Italy. A careful interpretation, however, is requested in this case, given that our mapping is far from representing a causal relation between inequality and these specific factors. More simply, the graphical representation presented here supports the need of scrutinising the relevance of preferences about inequality and quality of institutions as additional factors that can explain regional income disparities in Italy.

4.5 The place-specific redistributive effects of the Italian PIT

The Reynolds-Smolensky (RS) index is useful for studying the redistributive effect of taxes with tax microdata (Jenkins, 1988). More precisely, the decomposition of the RS index suggested by Pfähler (Lambert, 2001) allows for the evaluation of the different effects of each component of the structure of an income tax in terms of progressivity and redistribution. In this case, we are interested in providing evidence on the place-specific
redistributive effects of the Italian PIT in order to show how the Italian IRPEF operates at a regional level. In particular, we apply the decomposition of the RS index proposed by Onrubia et al. (2014) to the tax files, which has three main merits: to reduce some methodological problems present in the Pfähler-Lambert method, such as the relevance of the sequential order adopted and the need of including reranking terms; to use the same benchmark; to provide a flexible formula for the redistributive effects of the PIT’s components.

Specifically, and using the Onrubia et al. (2014)’s original notation, for each NUTS-1 region we use the following decomposition of the RS index:

$$\Pi^{RS} = \frac{\bar{B}}{\bar{S} - \bar{B}} S \Pi^{K}_{B,B-S} - \frac{\bar{Y}}{\bar{T} - \bar{T}} \sum_{i=1}^{m} \frac{C_i}{\bar{Y}} \Pi^{K}_{Y-S,Y-S+C_i} - \frac{\bar{Y} \bar{S}}{B(\bar{Y} - \bar{S})} \sum_{i=1}^{n} \frac{D_i}{\bar{Y}} \Pi^{K}_{Y,Y-D_i} - R \quad (1)$$

where \(Y\) and \(B\) are the gross income and the taxable income, respectively; \(S\) and \(T\) are the gross and the net tax liabilities. \(C\) denotes the sum of the \(m = 5\) tax credits, and \(D\) the sum of the \(n = 2\) deductions. The presence of an upper bar indicates the average of a variable; \(\Pi^{K}_{X,Z}\) is the Kakwani index of progressivity of the generic variables \(X\) and \(Z\). The reranking term \(R\) captures the effects on horizontal equity (Urban, 2014).5

Insert about here.

Table 5. Reynolds-Smolensky decomposition, Italian macro-areas.

Table 5 reports the results of the decomposition of the RS index applied to the tax files for the year 2014. Note that, the decomposition of the RS index has been calculated for each macro-region separately and it measures the effects of the Italian PIT at a regional level. Tax credits represent the most relevant IRPEF’s elements to achieve redistribution. This is in line with the results obtained by a recent study conducted for the Italian case at a national level (Barbetta et al., 2016), which was based on a sample of more than one million individual tax returns for the 2011 fiscal year. The contribution of Barbetta and co-authors relied on a sample that is not fully representative of the Italian population, as recognised by the authors, due to the high representation of dependent workers and pensioners. The tax

---

5 A more detailed discussion of the decomposition of the RS index can be found in Di Caro (2017b), where an extended version of the formula in (1) has been applied to the Italian case.
files used here are representative of the Italian population of taxpayers: our sample contains information on dependent workers, self-employed, pensioners, and other categories.

In addition, spatial differences are observed when comparing the redistributive consequences of the IRPEF’s elements. In the Central and Northern regions, redistributive objectives are mostly reached by means of tax schedules, while in the South tax credits represent the main redistributive instrument. This result derives from the combination of regional differences in the initial economic conditions and the IRPEF structure. In the Centre-North, where the average reported income is relatively high, tax schedules operate for smoothing the initial distribution of income among taxpayers. In the South, where average reported income is relatively low, tax credits, particularly those decreasing with respect to income such as tax credits for families and working conditions, play a crucial role for achieving redistribution. The role of deductions is of limited importance at a national as well as at a regional level.

As for the effects on horizontal equity, the reranking term is higher in the regions located in the South than in the rest of the country. This means that IRPEF creates more distortions on the redistribution of income within regions located in the South than in the Centre-North; a possible justification can be related to the design of most of tax credits that are tailored for particular taxpayers (e.g. dependent workers). The reranking term partially describes horizontal equity; measures based on social welfare functions and disaggregate dimensions are needed for having a more complete picture (Dardanoni and Lambert, 2002). The tax credit ‘80 Euro bonus’ made a positive contribution to reducing income inequality in Italy: the share of the tax credits is high when the bonus is accounted for. This confirms the view that the ‘80 Euro bonus’ operated for smoothing the adverse effects of the Great Recession in Italy (Baldini et al., 2015). Moreover, it can be noted that the bonus increased the reranking term, that is, it produced negative effects on horizontal equity. This result derives from the specific design of the bonus, which do not take into consideration poor households below the tax threshold and it is valid only for dependent workers (Morini and Pellegrino, 2016).

5. Concluding remarks

Income inequalities between and within regions have increasingly attracted the interest of researchers by posing novel theoretical and practical questions (Kastrop, 2016). This paper made a contribution to the study of regional income disparities in Italy by
relying on tax statistics. The main results of the work can be summarised as follows. By updating and integrating the findings of previous analyses conducted for the Italian case with different datasets, we have documented that regional income disparities are relevant between and within the four Italian NUTS-1 regions. In addition, supporting evidence has been provided on the idea that gender and age profiles play an important role for understanding inequality differences across Italy. An exploratory analysis, moreover, suggests that factors like individual preferences about inequality and the quality of regional institutions can be useful for explaining spatial differences in income across Italian regions. Exploiting the informative contents of the tax records in our sample, we have decomposed the redistributive consequences of the Italian PIT at a regional level by finding that tax credits and tax schedules register place-specific patterns. The effects of the ‘80 Euro bonus’ on vertical and horizontal equity have been also studied from a regional perspective.

Our analysis can be improved along three main directions, which represent future avenues of research. First, the availability of longitudinal tax statistics can help the evaluation of the distributive implications of the tax instruments used in the Italian tax system by adopting panel techniques. This opportunity, however, will pose additional challenges on the use of tax records: time-specific patterns have to be correctly specified for studying income disparities (Schröder et al., 2014). Second, tax statistics with a regional breakdown like those used in this paper can be used as inputs for the definition and application of spatial microsimulation models (Tanton, 2014). In Italy, for instance, they can be useful for enhancing the regional tax-benefit microsimulation model that uses EU-SILC data (Matino et al., 2017). Third, the study of the spatial dimension is important when assessing the distributional implications of tax credits and PIT’s elements in order to have a more accurate picture on the overall effects of IRPEF on progressivity and redistribution (Diamond and Saez, 2011). These questions are left for future work.
References


FIGURES AND TABLES

Figures

Figure 1. Regional distribution of the Gini index, tax files 2014

Note: The graph reports the Gini Index of the average gross income. The variable is classified on the basis of the equal intervals method.

Figure 2. Gross Income, macro-regional distribution

Note: The graphs above plot the frequency density function (y-axis) and the gross income between 0 and 100,000 EUR (x-axis) in the four Italian macro-regions.
Figure 3. Concentration curves for gross income and tax, by macro-areas

GROSS INCOME

TAX LIABILITY

Note: The graphs plot the concentration curves for gross income (left) and net tax (right) in the four Italian macro-regions; the percentage of population (x-axis) is ordered by the gross income. CE: Centre; NE: North-East; NW: North-West; SO: South; ITA: Italy.

Figure 4. Concentration curves for gross income, by macro-regions and age classes

0-35

36-65

Note: The graphs plot the concentration curves for gross income for the age classes 0-35 years (left) and 36-65 (right) in the four Italian macro-regions; the percentage of population (x-axis) is ordered by the gross income. CE: Centre; NE: North-East; NW: North-West; SO: South; ITA: Italy.
Figure 5. Kernel densities of gross income, by macro-areas and n. of children

Note: The graphs plot the kernel densities (y-axis) for the (log of) gross income (x-axis) in the four Italian macro-regions according to the number of children: 0, 1, 2, 3+.

Figure 6. Regional distribution of inequality acceptance and quality of institutions

Note: The graphs report the regional distribution of the inequality acceptance variable (left) and the variable describing the quality of institutions (right). In both cases, variables are classified on the basis of the equal intervals method.
### Tables

#### Table 1. Regional inequality in Italy, selected contributions

<table>
<thead>
<tr>
<th>Contribution</th>
<th>Level of analysis</th>
<th>Data source</th>
<th>Period</th>
<th>Inequality measure</th>
<th>Main results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cannari &amp; D’Alessio (2003)</td>
<td>Region (NUTS-2)</td>
<td>SHIW</td>
<td>1995-2000</td>
<td>Gini Index, MLD</td>
<td>North-South disparities; High concentration of income in the South; Relevance of socio-demographic factors</td>
</tr>
<tr>
<td>Jesuit et al. (2003)</td>
<td>Region (NUTS-2)</td>
<td>LIS</td>
<td>1991; 1995</td>
<td>Median Income</td>
<td>North-South disparities; Region-specific patterns</td>
</tr>
<tr>
<td>Acciari &amp; Mocetti (2012)</td>
<td>Province (NUTS-3)</td>
<td>MEF</td>
<td>2000-2011</td>
<td>Gini Index</td>
<td>North-South disparities; Within province differences; Relevance of economic factors</td>
</tr>
<tr>
<td>Longford et al. (2012)</td>
<td>Region (NUTS-2)</td>
<td>EU-SILC</td>
<td>2006</td>
<td>Gini Index, Lorenz curve</td>
<td>North-South disparities; Region-specific patterns</td>
</tr>
<tr>
<td>Mussida &amp; Parisi (2016)</td>
<td>Macro-region (NUTS-1)</td>
<td>EU-SILC</td>
<td>2009; 2014</td>
<td>Gini Index, Atkinson Index, Palma Index, Percentile ratio</td>
<td>North-South disparities; Role of the crisis; Differences between nationals and foreigners</td>
</tr>
</tbody>
</table>

Note: EU-SILC (European Statistics on Income and Living Conditions Survey); LIS (Luxembourg Income Study); MEF (Ministry of Economy and Finance); MLD (Mean Logarithmic Deviation); SHIW (Survey of Household Income and Wealth).

#### Table 2. Comparison of average gross income, macro-regions

<table>
<thead>
<tr>
<th>Region</th>
<th>EU-SILC</th>
<th>MEF</th>
<th>Tax Files</th>
</tr>
</thead>
<tbody>
<tr>
<td>North-West</td>
<td>22,047.90</td>
<td>28,858.23</td>
<td>22,083.90</td>
</tr>
<tr>
<td>North-East</td>
<td>21,313.41</td>
<td>28,206.13</td>
<td>21,280.88</td>
</tr>
<tr>
<td>Centre</td>
<td>20,117.17</td>
<td>25,191.71</td>
<td>20,608.71</td>
</tr>
<tr>
<td>South</td>
<td>15,836.57</td>
<td>18,761.41</td>
<td>15,819.32</td>
</tr>
<tr>
<td>St.dev.</td>
<td>2,777.86</td>
<td>4,613.86</td>
<td>2,786.76</td>
</tr>
<tr>
<td>Italy</td>
<td>20,072.02</td>
<td>23,955.78</td>
<td>19,002.32</td>
</tr>
</tbody>
</table>

Note: EU-SILC (European Statistics on Income and Living Conditions Survey); MEF (Ministry of Economy and Finance).
Table 3. Regional variations (%) in the employment rate during the Great Recession

<table>
<thead>
<tr>
<th>Area</th>
<th>North-West</th>
<th>North-East</th>
<th>Centre</th>
<th>South</th>
<th>Italy</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;=35 years</td>
<td>-22.5</td>
<td>-22.2</td>
<td>-20.8</td>
<td>-26.6</td>
<td>-23.0</td>
</tr>
<tr>
<td>Men</td>
<td>-23.5</td>
<td>-22.1</td>
<td>-22.3</td>
<td>-31.0</td>
<td>-25.2</td>
</tr>
<tr>
<td>Women</td>
<td>-21.2</td>
<td>-22.2</td>
<td>-19.0</td>
<td>-18.7</td>
<td>-20.0</td>
</tr>
<tr>
<td>36-65 years</td>
<td>-3.14</td>
<td>-3.68</td>
<td>-2.07</td>
<td>-10.2</td>
<td>-4.90</td>
</tr>
<tr>
<td>Men</td>
<td>-5.94</td>
<td>-5.50</td>
<td>-6.05</td>
<td>-14.1</td>
<td>-8.37</td>
</tr>
<tr>
<td>Women</td>
<td>0.69</td>
<td>-0.98</td>
<td>-3.39</td>
<td>-2.63</td>
<td>-0.39</td>
</tr>
<tr>
<td>Tot. Population</td>
<td>-5.04</td>
<td>-5.01</td>
<td>-2.99</td>
<td>-11.9</td>
<td>-6.45</td>
</tr>
<tr>
<td>Men</td>
<td>-8.11</td>
<td>-7.11</td>
<td>-7.12</td>
<td>-15.7</td>
<td>-10.0</td>
</tr>
<tr>
<td>Women</td>
<td>-0.92</td>
<td>-2.02</td>
<td>-2.69</td>
<td>-4.68</td>
<td>-1.10</td>
</tr>
</tbody>
</table>

Note: our elaborations, data from ISTAT.

Table 4. Gini Index for gross and net income, different breakdowns

<table>
<thead>
<tr>
<th>Region</th>
<th>Gini gross income</th>
<th>Gini net income</th>
<th>Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>North-West</td>
<td>0.4468</td>
<td>0.3946</td>
<td>-11.7</td>
</tr>
<tr>
<td>North-East</td>
<td>0.4306</td>
<td>0.3828</td>
<td>-11.1</td>
</tr>
<tr>
<td>Centre</td>
<td>0.4643</td>
<td>0.4128</td>
<td>-11.1</td>
</tr>
<tr>
<td>South</td>
<td>0.4754</td>
<td>0.4276</td>
<td>-10.1</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.4507</td>
<td>0.4061</td>
<td>-9.9</td>
</tr>
<tr>
<td>Male</td>
<td>0.4477</td>
<td>0.3946</td>
<td>-11.9</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-35</td>
<td>0.4951</td>
<td>0.4624</td>
<td>-6.6</td>
</tr>
<tr>
<td>36-65</td>
<td>0.4490</td>
<td>0.3962</td>
<td>-11.8</td>
</tr>
<tr>
<td>65+</td>
<td>0.4210</td>
<td>0.3658</td>
<td>-13.1</td>
</tr>
<tr>
<td>N. of child.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.4727</td>
<td>0.4251</td>
<td>-10.1</td>
</tr>
<tr>
<td>1</td>
<td>0.3816</td>
<td>0.3218</td>
<td>-15.7</td>
</tr>
<tr>
<td>2</td>
<td>0.4102</td>
<td>0.3437</td>
<td>-16.2</td>
</tr>
<tr>
<td>3+</td>
<td>0.4465</td>
<td>0.3688</td>
<td>-17.4</td>
</tr>
</tbody>
</table>

Note: Data refer to tax files for the year 2014.

Table 5. Reynolds-Smolensky decomposition, Italian macro-areas

<table>
<thead>
<tr>
<th>Region</th>
<th>Tax Schedules</th>
<th>Tax Credits</th>
<th>Deductions</th>
<th>Re-ranking</th>
<th>R-S Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>North-West</td>
<td>0.4225</td>
<td>0.5842</td>
<td>0.0077</td>
<td>-0.0144</td>
<td>1.0000</td>
</tr>
<tr>
<td>North-East</td>
<td>0.3855</td>
<td>0.6204</td>
<td>0.0102</td>
<td>-0.0161</td>
<td>1.0000</td>
</tr>
<tr>
<td>Centre</td>
<td>0.3845</td>
<td>0.6140</td>
<td>0.0173</td>
<td>-0.0158</td>
<td>1.0000</td>
</tr>
<tr>
<td>South</td>
<td>0.2691</td>
<td>0.7342</td>
<td>0.0164</td>
<td>-0.0197</td>
<td>1.0000</td>
</tr>
<tr>
<td>Italy</td>
<td>0.3592</td>
<td>0.6523</td>
<td>0.0116</td>
<td>-0.0231</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Region</th>
<th>Tax Schedules</th>
<th>Tax Credits</th>
<th>Deductions</th>
<th>Re-ranking</th>
<th>R-S Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>North-West</td>
<td>0.4501</td>
<td>0.5553</td>
<td>0.0082</td>
<td>-0.0136</td>
<td>1.0000</td>
</tr>
<tr>
<td>North-East</td>
<td>0.4114</td>
<td>0.5931</td>
<td>0.0109</td>
<td>-0.0154</td>
<td>1.0000</td>
</tr>
<tr>
<td>Centre</td>
<td>0.4088</td>
<td>0.5880</td>
<td>0.0184</td>
<td>-0.0152</td>
<td>1.0000</td>
</tr>
<tr>
<td>South</td>
<td>0.2816</td>
<td>0.7190</td>
<td>0.0171</td>
<td>-0.0177</td>
<td>1.0000</td>
</tr>
<tr>
<td>Italy</td>
<td>0.3806</td>
<td>0.6284</td>
<td>0.0123</td>
<td>-0.0213</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Note: Values expressed as percentage of the total Reynolds-Smolensky Index.