WIDER Working Paper 2018/165

Top income adjustments and tax reforms in Ecuador

H. Xavier Jara[1] and Nicolás Oliva[2]

December 2018

United Nations University World Institute for Development Economics Research

wider.unu.edu

Abstract: Top income under-coverage in developing countries not only leads to downward biased inequality indicators but might also affect the ex-ante evaluation of progressive tax reforms. We propose a simple adjustment to top incomes for formal employees (e.g. employees affiliated with social security) in survey data from Ecuador based on information from administrative tax records. ECUAMOD, the tax-benefit microsimulation model for Ecuador, is then used to compare income inequality, total tax revenue, and work incentives with and without top income adjustment under the baseline tax-benefit system and for a hypothetical income tax reform. Our study shows that adjusting top incomes in survey data allows us to capture 98 per cent of income tax revenue in our simulations. Moreover, income inequality measured by the Gini coefficient increases by 3 points following the adjustments of top incomes. The evaluation of our progressive tax reform shows that our baseline results with unadjusted top income data capture well the percentage change in tax revenue but underestimate the additional tax revenue amount by 22 per cent. We conclude with a discussion of the challenges to combine survey data and tax records in developing countries.

Keywords: top incomes, survey under-coverage, tax return data, tax reforms
JEL classification: D31, C81

[1] ISER, University of Essex, Colchester, UK, corresponding author, email: hxjara@essex.ac.uk; [2] Centro de Estudios Fiscales – SRI, Quito, Ecuador.

1   Introduction

Latin America remains one of the regions with the highest income inequality and the redistributive role of tax-benefit systems in the region is very modest. Household survey data is the main source of information used to produce income inequality indicators in Latin America. Recently, the growing literature on top incomes based on administrative tax records has stressed the problems with household survey data in capturing the top of the income distribution (Atkinson et al 2011, Burkhauser et al. 2012). The use of income tax return data represents in this sense a valuable source of information to provide a more precise picture of top incomes.

Over the last years, the study of top incomes based on tax records has gained increasing interest. A large number of studies have focused on calculating top income-corrected inequality estimates. The majority of this work has proposed calculating top income-corrected inequality measures by combining inequality estimates calculated separately for the top and non-top groups of the population, where inequality for the top group has been calculated either based on survey data (Alfons et al. 2013, Burkhauser et al. 2012, Cowell and Flachaire 2007) or income tax returns (Atkinson 2007, Atkinson et al 2011, Alvaredo 2011). Instead of combining inequality estimates, one could directly combine data sources (household surveys and tax records) and then obtain inequality indicators based on the modified data (Bach et al. 2009, Burkhauser et al. 2016, DWP 2015). The latter approach has the additional advantage of improving the ex-ante evaluation of policy reforms targeting the top of the income distribution.

The aim of this paper is twofold. First, we analyse the effect on income inequality of correcting top incomes in survey data based on information from income tax return data from Ecuador. For this, we propose a simple adjustment to top incomes in household survey data, taking into consideration that tax records capture only a fraction of the working population in developing countries, namely formal workers. Second, we make use of tax-benefit microsimulation techniques to assess the effect of a hypothetical reform aiming to make personal income tax more progressive, before and after correcting top incomes in the survey.

Our results show that income equality, measured by the Gini coefficient, would increase by 3 percentage points in Ecuador when household survey data is enriched with information of top incomes from tax records. Moreover, the simulation of personal income tax improves significantly following the adjustment of top incomes in survey data. Simulations based on our unadjusted microdata capture only 80 per cent of income tax revenue, whereas correcting for top incomes allows us to capture around 98 per cent of income tax revenue from tax records. We simulate a hypothetical reform whereby personal income tax is made more progressive and show that ECUAMOD baseline results with unadjusted top income data capture well the percentage change in tax revenue but underestimate the additional tax revenue amount by 22 per cent compared to results based on the adjusted data. The reform has only very minor effects on income inequality because only a small fraction of the population is affected by the increase in income tax due to the high non-taxable threshold and the presence of deductions from personal expenditures characterizing the design of the Ecuadorian personal income tax. Work incentives are also only minimally affected by the reform, with mean Marginal Effective Tax Rates (METR) for the top earnings decile group increasing by 1.6 percentage points with the unadjusted data, and by 2.4 points based on the top income corrected data.

Our analysis contributes to the literature in two main respects. First, to the best of our knowledge this is the first study that attempts to directly combine household survey data with income tax records in Latin America for the study of income inequality. We show that this approach provides

results in line with previous studies for Ecuador, which have calculated top-corrected Gini coefficients by combining inequality estimates rather than combining data sources. Second, we propose a simple approach to enrich household survey data with information from tax records in developing countries for the purpose of ex-ante policy evaluation and show that such adjustments improve the simulations of personal income tax.

The remainder of the paper is organized as follows. Section 2 briefly summarizes the literature on top incomes and income inequality. Section 3 describes the data sources and the approach employed to adjust top income information from survey data based on tax records. Section 4 reports the results of the analysis on the effect of top income corrections on income inequality and ex-ante evaluation of tax reforms. Section 5 concludes with a discussion of the challenges faced in combining survey data and tax records in developing countries.

2   Top incomes and income inequality

Over the last decades there has been an increasing interest in the study of top incomes based on information from income tax returns and national accounts. The majority of studies on top incomes have followed one of two main strands of research. The first strand has focused on analyzing the evolution of top income shares over time and across countries. The second strand has focused on using income tax return data to estimate top income-corrected inequality indicators.

The first strand of the literature was inspired by the work by Piketty (2001) and Piketty and Saez (2003) on the evolution of top income shares in France and the United States. Since then, a large number of studies at the national, cross national and global level has emerged (Atkinson 2005; Atkinson and Piketty 2007, 2010). The World Inequality Database (WID) initiative provides access to historical series on top income shares for a number of developed and developing countries.[1] WID currently provides information on top income shares for five Latin American countries: Argentina, Brazil, Chile, Colombia, and Uruguay. For Ecuador, a study by Cano (2014) analyses the evolution of top income shares over the period 2004–11 based on income tax return data, together with National Accounts and population projections for control totals for income and population. Using gross income as the concept of interest, their results show that the income share accruing to the top 1 per cent increased from 13.1 per cent in 2004 to 20.8 per cent in 2009, then decreased to 16.2 per cent in 2011.

The literature on top income shares has also highlighted the problems of household survey data in capturing top-income groups (Atkinson et al. 2011, Burkhauser et al. 2012) and prompted the second strand of the literature which focuses on estimating inequality indicators accounting for top income corrections. Jenkins (2017) provides a clear summary of the three different approaches that have been used in the literature to estimate top income corrected inequality indicators. The first approach is based purely on survey data and consists in calculating inequality estimates separately for the non-top income group and for the top-income group, with inequality estimates for the top group usually calculated by fitting a Pareto Type I distribution (Alfons et al. 2013; Burkhauser et al. 2012; Cowell and Flachaire 2007). The estimate of total inequality is then obtained as the sum of inequality among the non-top group, inequality among the top group and inequality between the top and non-top groups. The second approach is identical to the first except that it calculates the inequality estimate for the top group using tax return data (Atkinson 2007; Atkinson

---

[1] For more information about WID see: https://wid.world/.

et al 2011; Alvaredo 2011). The third approach differs from the previous two and consists in directly replacing the income information from the top group in the survey data with the income information from the top group in tax return data (Bach et al. 2009; Burkhauser et al. 2016; DWP 2015). Jenkins (2017) shows that the second and third approaches provide similar results in terms of the evolution of income inequality in the UK from 1995 to 2010. The second approach described above has been used by Rossignolo et al. (2016) to calculate top-corrected Gini coefficients from gross labour income in Ecuador. Their results show that the Gini coefficient from gross labour income would increase by 3.1 points in 2012 after correcting top income inequality estimates based on administrative tax records.

An important advantage of the third approach discussed above has been overlooked by the literature and refers to the fact that directly improving the income information in survey data would allow improving ex-ante evaluation of policy reforms affecting the top of the income distribution. This is particularly important in developing countries, where survey data captures poorly the richest population and governments might be interested in assessing the effect of progressive income tax reforms to increase tax revenue. We thus follow this approach and show how correcting top income information in survey data based on tax returns in Ecuador would not only provide a better picture of income inequality but could help assess the effect of potential income tax reforms in the country, as discussed in the following sections.

3   Data and methodology

Our analysis makes use of income tax return data to adjust top income information in household survey data from the National Survey of Income and Expenditures of Urban and Rural Households in Ecuador (Encuesta Nacional de Ingresos y Gastos de Hogares Urbanos y Rurales, ENIGHUR). The adjusted data is then used to simulate a hypothetical income tax reform using ECUAMOD, the tax-benefit microsimulation model for Ecuador. This section starts describing the different data sources and the methodology to adjust top incomes in survey data. Then, the hypothetical reforms to be analysed are discussed.

3.1   Income tax return data

Income tax return data has been provided by the Ecuadorian tax administration (Servicio de Rentas Internas, SRI) and contains information on all individuals who have filed tax returns. Personal income tax in Ecuador is assessed at the individual level and levied jointly on labour and capital income according to a progressive tax schedule with rates from 0 to 35 per cent. Income tax is automatically withheld and reported by employers of employees with no other sources of income than employment income. Employees with other sources of income and the self-employed are required to file a tax return consolidating all their sources of income (employment and self-employment income and capital income).[2] In order to ensure comparability in the concepts used from tax records and household survey data, we focus here only on employment income and use

---

[2] Income tax return data comprises information from three different tax forms. Tax form 107 reports income information for employees formal employment. Tax forms 102 and 102A contain information on business, self-employment income, capital returns, salaries and other income sources, respectively, from individuals who are (form 102) and who aren't (form 102A) required to keep accounting books.

2011 information to match the year of the survey.[3] Formal employees represent 90 per cent of the formal working population covered by tax records in Ecuador.

Income tax return data for 2011 contains information for around 2.7 million employees. Despite the rich income information available in tax records in Ecuador (as in many developing countries), the data is not suitable to perform distributional analysis at the country level for several reasons. First, there is a large informal sector in the economy, which is not captured by tax records. Informality defined in terms of non-affiliation to social security amounted to around 60 per cent of the working population in 2011 (Canelas 2014). Second, income tax return data contains information only for individuals in work who file taxes and therefore does not capture the inactive or unemployed population; neither does the data contain information on benefits receipt. Therefore, the bottom of the income distribution is not captured. Finally, there is no socio-demographic information for individuals and no household information in the data. For this reason, most distributional analysis in Ecuador is based on household survey data, which we discuss in the following section.

3.2 Household survey data and tax-benefit microsimulation

Household survey data. Our study uses data from the latest wave of the National Survey of Income and Expenditures of Urban and Rural Households in Ecuador (Encuesta Nacional de Ingresos y Gastos de Hogares Urbanos y Rurales, ENIGHUR). ENIGHUR 2011/2012 is a nationally representative cross-sectional survey, conducted approximately every eight years and contains information for 39,617 households and 153,341 individuals.[4] ENIGHUR was chosen over other household survey data sources in Ecuador due to its very detailed information on income and expenditures of households. Gross employment and self-employment income from principal and secondary occupations are reported, including extra pay, bonuses, in-kind income and self-consumption from self-employment activities. Moreover, detailed information about affiliation to social security is reported, allowing to classify individuals as formal or informal and to categorize them to specific social security regimes. Cash transfers from social benefits and public pensions are also available. Other sources of income such as income from capital and property, private transfers, remittances, other rents, are also reported in the data. Finally, detailed data on household expenditures is also available in the survey, as well as information on demographic and socio-economic characteristics of each member of the household. Because of the rich information available in the data, ENIGHUR is also the underlying microdata used with our microsimulation model, ECUAMOD.

Tax-benefit simulations. Our analysis makes use of ECUAMOD, the tax-benefit microsimulation model for Ecuador, which has been developed as part of UNU-WIDER's project on 'SOUTHMOD-simulating tax and benefit policies for development' in which tax-benefit microsimulation models have been built for selected developing countries. ECUAMOD and other

---

[3] The tax return data provided contains self-employment income net of social insurance contributions. It is not possible to differentiate between own social insurance contribution payments and employer social insurance payments for their workers. Moreover, social insurance contributions are part of the same variable containing salary payments to workers. For this reason, it was not possible to recreate a variable of self-employment income gross of social insurance payments to compare it to the concept used in the survey and simulations.

[4] Individuals recorded as domestic employees in a household have been dropped, as information about their own household (e.g. number of children, expenditures) is not available. In total, 103 individuals (0.07 per cent of the sample) were dropped from the original sample. No households were dropped and no adjustments to the weights were made as a result of dropping individual observations. For more information on data adjustment made as part of ECUAMOD work, see Jara et al. (2017).

country models from the SOUTHMOD project have been implemented on the EUROMOD platform, which provides a harmonized setting for cross country comparative analysis on the redistributive effect of tax-benefit policies (see Sutherland and Figari 2013). ECUAMOD combines detailed country-specific coded policy rules with microdata from ENIGHUR to simulate direct and indirect taxes, social insurance contributions, as well as cash transfers for the household population of Ecuador. ECUAMOD is a static model in the sense that tax-benefit simulations abstract from behavioural reactions of individuals and no adjustments are made for changes in the population composition over time. Simulation results for ECUAMOD have been validated against external statistics (see Jara et al., 2017).

ECUAMOD version 1.4 covers policy years 2011–18 for the purpose of tax-benefit simulations, based on ENIGHUR 2011–12. To account for time inconsistencies between the input data year and the policy year in the simulations, market incomes and non-simulated tax-benefit variables in the data are adjusted using source-specific updating factors (see Jara et al. 2017). In this study, 2011 policies (as on June 30th) in Ecuador are used for the analysis.

The baseline simulations of personal income tax in Ecuador assume full compliance, meaning that personal income tax is simulated for all earners in ENIGHUR data, irrespectively of whether they are defined as formal or informal workers (i.e. whether they report being affiliated to social security or not). As part of our analysis and in order to be consistent with the information available in tax records, once top income information has been adjusted in the survey, we relax the assumption on full compliance and simulate personal income tax only for earners defined as formal workers.

3.3 Adjusting top income information in survey data

The adjustment to improve top income coverage in survey data proposed in this paper draws from the methodology used by the Department for Work and Pensions in the UK (DWP 2015) and from the revisions to this methodology proposed by Burkhauser et al. (2016). The two approaches consist of replacing incomes from the top of the distribution in the survey with cellmean values of the top of the distribution in the tax return data. Here, we propose to adjust the top of the distribution in the survey based on information from cellmean values obtained from both data sources.

Our proposed top income adjustment works as follows. First, in both datasets we create a variable of gross employment income with a harmonized definition, which is the sum of salaries, extra pay, bonuses, utilities participation, 13th and 14th month pay and reserve funds. Second, we select a sample of employees in the survey data that is the most comparable to the sample of employees in the income tax return data, i.e. formal employees. For this, we select in the survey only employees who report being affiliated to social security. In Ecuador, affiliation to social security is mandatory for employees. Those not affiliated to social security are therefore most likely employed in the informal sector. Third, we rank individuals by their gross employment income in the two datasets and allocate them to income percentile groups. Fourth, we calculate the average income for each income percentile in the two datasets. Fifth, for each percentile group we calculate the ratio between the average income in the tax records and the average income in the survey data. Sixth, we identify the point at the top of the income distribution when the gap between the average income in the survey and the average income in the tax records starts widening. For this, we select as threshold the point where average income in the tax records is more than 10 per cent larger than the average income in the survey (a ratio above 1.10). Finally, for employees in the top percentiles of the survey data where the ratio between the average income in the tax record and the average income in the survey is above 1.10, we adjust their gross employment income by multiplying it to the ratio of their corresponding percentile.

Note that differences in incomes between survey data and income tax return data could also be observed at the bottom of the distribution. Previous studies on measurement error of earnings in survey data have found that low income individuals would tend to over-report earnings, whereas high income individuals would tend to under-report earnings (Bound and Krueger 1991; Bollinger 1998; Kapteyn and Ypma 2007). Whereas there seems to be large agreement on the fact that differences between survey and tax return data at the top of the distribution are mainly the result of top income under-coverage in survey data, some studies have opposed the view that administrative data is error-free (Kapteyn and Ypma 2007; Meijer et al. 2012; Abowd and Stinson 2013). In this paper, we concentrate on top income under-coverage in household survey data and do not discuss data quality issues at the bottom of the distribution.

4 Empirical results

This section presents the results of our assessment of top income under-coverage in survey data from Ecuador and the effect of adjusting top incomes based on tax records. We first compare income information from survey and tax records data to analyse the extent to which top incomes of formal workers are under-covered in ENIGHUR data. Then, we assess the effects of adjusting top incomes in the survey on simulated tax revenue and income inequality. Finally, we illustrate the difference made by top-income adjustments on the evaluation of the effect of a hypothetical progressive tax reform using ECUAMOD.

4.1 Top income under-coverage by household survey data

Tables 1 and 2 present some basic descriptive statistics to provide a first impression of the degree of top income under-coverage in the household survey data in Ecuador. Hereafter we refer to the survey data as "ECUAMOD data" as we use the harmonized version of ENIGHUR 2011/2012 data that runs with the tax-benefit microsimulation model ECUAMOD.

Table 1. Population totals by income thresholds – formal employees (2011)

	ECUAMOD		Tax records	Ratio
	(unweighted)	(weighted)	(unweighted)	
All	20,548	2,062,475	2,716,664	1.3
Above 1st tax threshold	6,224	524,752	757,029	1.4
Above 2nd tax threshold	4,164	349,150	556,032	1.6
Above 3rd tax threshold	2,642	227,007	383,465	1.7
Above 4th tax threshold	1,659	151,595	278,554	1.8
Above 5th tax threshold	214	26,658	72,414	2.7
Above 6th tax threshold	53	8,125	29,682	3.7
Above 7th tax threshold	13	3,087	14,300	4.6
Above 8th tax threshold	3	749	6,704	8.9

Source: Authors elaboration based on ECUAMOD version v1.4 and income tax return data from SRI.

Table 1 shows the population totals for formal employees in the whole sample of each dataset, as well as the totals above the different income tax thresholds in Ecuador. Table A1 in the annex provides detailed information on the structure of the tax schedule in Ecuador for year 2011. It is worth noting that in Ecuador, as in many Latin American countries, the exemption threshold for income tax payment (the 1st tax threshold) is rather high and equalled USD 9,210 in 2011, which was equivalent to 2.9 times the annualized minimum wage (USD 3,168). The results show that top

income under-coverage in survey data is an important issue in Ecuador. The under-representation of formal employees increases significantly at the highest income tax thresholds in the survey compare to the tax records. ECUAMOD data identifies only 3 observations with gross employment income above the 8th income tax threshold, which corresponds to 749 employees using the survey weights. On the contrary, tax records identify 6,704 employees above this threshold, which is 8.9 times larger than the weighted number of employees in the survey. This first set of results makes it evident that income tax simulations based on survey data would underestimate the number of tax payers and total tax revenue in the country. Potential counterfactual simulations of progressive tax reforms (e.g. increase the tax rate of the top tax band) based on household survey data would most likely underestimate the effect of such policies.

Table 2 presents similar information but now compares total gross employment income at different income thresholds between survey and tax records data. The results reflect the under-coverage of top incomes observed in Table 1. The total income of formal employees in tax records is 1.4 times higher than that of formal employees in survey data. The underestimation of total employment income in survey data compared to tax records becomes increasingly severe at higher income tax thresholds. Total employment income of individuals with income above the 8th income tax threshold is 15 times larger in tax records compared to survey data.

Table 2. Total income by income thresholds – formal employees (in millions USD for 2011)

|	ECUAMOD	Tax records	Ratio
All	16,368	23,441	1.4
Above 1st tax threshold	8,834	15,395	1.7
Above 2nd tax threshold	7,009	13,306	1.9
Above 3rd tax threshold	5,413	11,042	2.0
Above 4th tax threshold	4,209	9,362	2.2
Above 5th tax threshold	1,340	4,492	3.4
Above 6th tax threshold	568	2,683	4.7
Above 7th tax threshold	258	1,756	6.8
Above 8th tax threshold	76	1,146	15.0

Source: Authors elaboration based on ECUAMOD version v1.4 and income tax return data from SRI.

Another way to assess the degree of top income under-coverage in survey data, which is the basis for our adjustment methodology, is to compare average employment income by income groups in each data source. Figure 1 presents average yearly employment income by income percentiles from each data source. The results show that the gap between average employment income between survey data and tax records starts widening after the 80th percentile, where average employment income from tax records is 1.11 times higher than that of survey data, representing around half of an annualized minimum wage. The ratio between average employment income in tax records and survey data amounts to 1.73 in the top income percentile.

Note that for low income percentile groups average employment income is observed to be higher in survey data than in tax records, which is consistent with previous research (Bound and Krueger 1991; Bollinger 1998; Kapteyn and Ypma 2007). In addition to the problem of over-reporting of low earners discussed in the literature, in the case of Ecuador, additional factors could drive these results. First, employees in ENIGHUR data report their gross employment income over the last month. In order to compare this information with yearly income recorded in tax return data, we multiply gross employment income by 12 months in the survey. Annualizing employment income from survey data might create some noise for individuals with low earnings who could potentially be working less than 12 months over a year, as low earnings might be correlated with more insecure

jobs. This noise is likely to be lower for higher income groups as individuals with high earnings are likely to be in more secure jobs. Second, there could be underreporting of earnings in tax records. For instance, some employers might register their employees as earning the minimum wage in order to pay lower social insurance contributions, whereas they are being paid more than the minimum.[5] This issue might be mitigated in household survey data, where individuals report directly how much they were paid by their employers. As previously mentioned, in this study we focus on adjustments at the top of the income distribution because it is less clear to know what might be driving differences between survey data and tax records at the bottom end of the distribution and given that the difference is small in absolute terms.

Figure 1. Average income by income percentiles – formal employees (in 2011 USD)
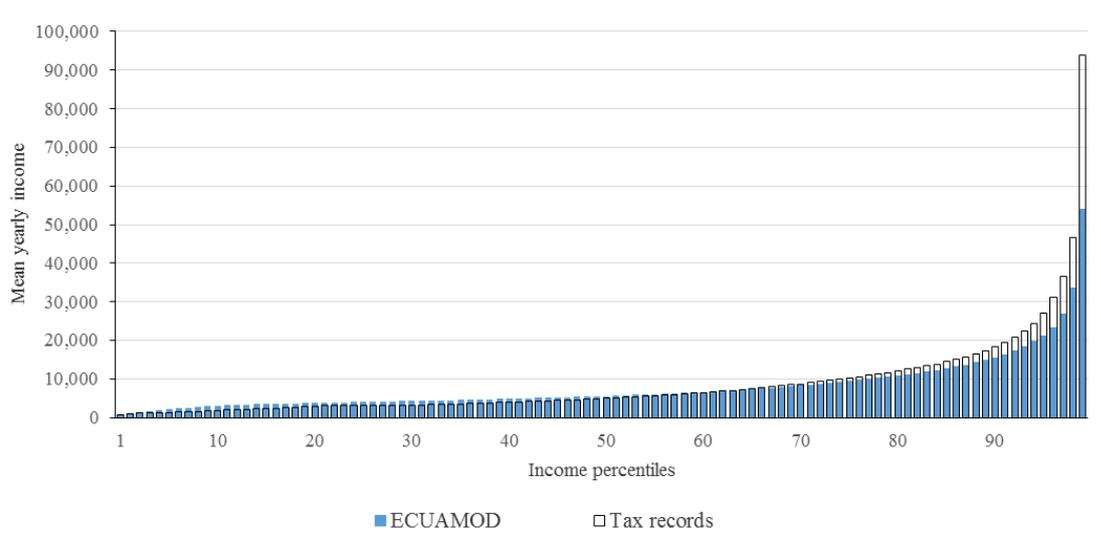
Source: Authors elaboration based on ECUAMOD version v1.4 and income tax return data from SRI.

4.2 Distributional and budgetary impacts of top income adjustments

This section presents the effects of adjusting household survey data to better capture incomes at the top of the distribution in Ecuador. We focus on the effect on two particular outcomes: simulated tax revenue and income inequality.

Table 3 shows the total number of income tax payers and the total tax revenue obtained with ECUAMOD simulations before (column A) and after (column B) adjusting top incomes in the survey data. The table also presents the number of tax payers and the tax revenue from tax records (column C). Our results show that without adjusting top incomes in the survey, ECUAMOD underestimates the number of tax payers by around 30 per cent and income tax revenue by around 20 per cent. Adjusting top incomes in the survey increases only marginally the number of tax payers, however, simulated tax revenue captures around 98 per cent of income tax revenue from tax records.

---

[5] This practice has been referred to in the literature as "envelope wages" and seems to be particularly prevalent in Central and Eastern European countries (Meriküll and Staehr 2010; Woolfson 2007; Williams and Padmore 2013).

Table 3. Top income adjustments, tax payers and tax revenue (2011)

	ECUAMOD	ECUAMOD adjusted	Tax records	Ratios	
	(A)	(B)	(C)	(A)/(C)	(B)/(C)
Number of tax payers (in thousands)	334	336	476	0.70	0.71
Tax revenue (in million US dollars)	639	765	784	0.82	0.98

Source: Authors elaboration based on ECUAMOD version v1.4 and income tax return data from SRI.

Table 4 compares gross employment income inequality among formal employees in the survey and in tax records using a battery of indicators. As expected, the top income adjustment in the survey produces inequality indicators closer to those obtained from the tax records. The Gini coefficient from gross employment income increases by around 5 percentage points after the adjustments. The largest improvements are observed when top income shares from adjusted and unadjusted survey data are compared to tax records, with the adjusted data matching particularly well the share of gross employment income held by the top 1 per cent of the population.

Table 4. Top income adjustments and labour income inequality - formal employees (2011)

	ECUAMOD	ECUAMOD adjusted	Tax records	Ratios	
	(A)	(B)	(C)	(C)/(A)	(C)/(B)
Gini	43.3	48.2	51.0	1.2	1.1
Atkinson 0.5	15.7	19.9	21.4	1.4	1.1
Atkinson 1	28.5	33.8	36.4	1.3	1.1
P90/P10	7.3	8.2	9.8	1.3	1.2
Top 10% share	31.2	37.6	39.4	1.3	1.0
Top 5% share	20.1	25.9	27.2	1.4	1.1
Top 1% share	6.8	10.3	10.9	1.6	1.1

Source: Authors elaboration based on ECUAMOD version v1.4 and income tax return data from SRI.

The advantage of using household survey data compared to tax records is that it allows us to calculate inequality estimates at the population level. As previously mentioned, income tax return data in developing countries might capture only a small fraction of the population due to the presence of a larger informal sector. Moreover, combining household survey data with tax-benefit microsimulation allows us to go beyond the study of labour or market income, and look at household disposable income, defined as the sum of market income plus cash transfers plus public pensions net of social insurance contributions and direct taxes.[6]

---

[6] Market income is the sum of employment and self-employment income, bonuses, in-kind income, self-consumption from self-employment activities, capital and property income, inter-household payments, private transfers, minus alimony payments. Imputed rent is not included as part of market income.

Table 5 presents inequality indicators for the whole population in the survey obtained before and after the top income adjustment. The results show the effect of our top-incomes adjustment on a battery of inequality indicators based on per capita household disposable. Income inequality measured by the Gini coefficient from disposable income increases by 3 percentage points following the adjustment of top incomes in the survey. This result is in line with the 3.1 point increase obtained by Rossignolo et al. (2016), who calculated top-corrected Gini coefficients from gross labour income in Ecuador. The difference between their approach and ours is that they calculate top-corrected Gini coefficients by calculating separately the Gini for the non-top income group based on survey data, and the Gini for the top-income group based on tax records. Our approach relies, on the contrary, on directly modifying top income information in the survey and then calculates inequality indicators based on the adjusted data. Other inequality indicators presented in Table 5 also increase by around 10 per cent following the adjustment to top incomes.

Table 5. Top income adjustments and disposable income inequality (2011)

	ECUAMOD	ECUAMOD adjusted	Ratio
	(A)	(B)	(B)/(A)
Gini coefficient	46.1	49.1	1.1
Atkinson 0.5	17.8	20.3	1.1
Atkinson 1	30.8	34.0	1.1
P90/P10	7.4	7.9	1.1

Source: Authors elaboration based on ECUAMOD version v1.4 and income tax return data from SRI.

Table A.2 in the appendix shows that the increase in the Gini coefficient from disposable income after the adjustment is mainly driven by the increase in the Gini from market income (and therefore from labour income). The last row of Table A.2 shows that the redistributive effect of the tax-benefit system would only marginally increase following the adjustment (from 4.1 to 4.3 points in the Gini coefficient).

4.3   Top income adjustments and ex-ante evaluation of tax reforms

Adjusting top-incomes information in household survey data has the additional advantage of allowing to improve the ex-ante evaluation of policy reforms targeting the top of the income distribution. Assessing the effect of progressive income tax reforms might be particularly important in developing countries aiming to create more stables sources of government revenue to finance sustainable social programs.

In order to illustrate the effect of top income adjustments on tax reforms, we analyse a simple hypothetical reform scenario whereby the tax schedule in Ecuador is made more progressive. More precisely, we increase the tax rate of the fourth tax bracket from 12 to 15 per cent and increase all other rates from the fifth to the ninth tax brackets by 5 percentage points. The reform is purely illustrative and aims to provide some insights into the effect of our top income correction on ex-ante policy evaluation. The baseline and modified tax schedules are presented in Table 6.

We start by assessing the effects of our hypothetical reform in terms of the total number of tax payers and aggregate tax revenue with and without the adjustment of top incomes in ECUAMOD data. Table 7 shows that, as expected, the number of tax payers does not change between the baseline and the reform scenario because the hypothetical reform affects tax rates from tax brackets 4 and above and would therefore have an effect only on individuals who are already paying taxes.

Table 6. Personal income tax schedule for baseline and reform scenarios (2011)

Tax band	Lower limit	Upper limit	Marginal tax rate (%) baseline	Marginal tax rate (%) reform
1	0	9,210	0	0
2	9,210	11,730	5	5
3	11,730	14,670	10	10
4	14,670	17,610	12	15
5	17,610	35,210	15	20
6	35,210	52,810	20	25
7	52,810	70,420	25	30
8	70,420	93,890	30	35
9	93,890	-	35	40

Source: Authors elaboration and SRI (2010).

In terms of total tax revenue, our results show that without adjusting top incomes, tax revenue would increase from 639 to 757 million US dollars (118 million US dollars), which represents an 18 per cent increase. Using our adjusted survey data, tax revenue would increase from 765 to 914 million USD dollars (150 million US dollars), representing a 20 per cent increase. As such, our baseline results seem to capture well the relative increase in tax revenue following the reform. However, in absolute terms not adjusting top incomes would result in a 22 per cent underestimation of the increase of tax revenue following the reform (comparing 118 to 150 million US dollars).

Table 7. Number of tax payers and income tax revenue under the baseline and reform scenarios (2011)

	ECUAMOD		ECUAMOD adjusted		Ratios	
	Baseline	Reform	Baseline	Reform		
	(A)	(B)	(C)	(D)	(B)/(A)	(D)/(C)
Number of tax payers (in thousands)	334	334	336	336	1.00	1.00
Tax revenue (in million US dollars)	639	757	765	914	1.18	1.20

Source: Authors elaboration based on ECUAMOD version v1.4

We now turn to the effects of our illustrative reform in terms of income inequality. Table 8 compares the effects of our reform using both the unadjusted and adjusted survey data. In both cases, our hypothetical reform has only minor effects in terms of income inequality. For instance, the Gini coefficient from disposable income would decrease from 46.1 to 45.9 per cent based on the unadjusted data, and from 49.1 to 48.9 when top incomes are adjusted in the survey. In both cases, the effect of the reform represents only around 0.2 percentage points decrease in the Gini coefficient. This minor effect is also observed for other inequality indicators presented in Table 8. The minor effect of the reform in terms of income inequality might be related to two main factors. First, our illustrative reform increases only marginally the progressivity of the tax schedule, with an increase in the tax rate from 12 to 15 per cent for the fourth tax bracket and a 5 percentage points increase for higher tax brackets. Second, as previously discussed, personal income tax in Ecuador is characterized by a high exemption threshold and by the presence of deductions from personal expenditures. Despite the 20 per cent increase in tax revenue following the reform, the counterfactual scenario is effectively affecting only a small fraction of high earners in the population.

Table 8. Income inequality under the baseline and reform scenarios (2011)

	ECUAMOD		ECUAMOD adjusted		Ratios	
	Baseline	Reform	Baseline	Reform		
	(A)	(B)	(C)	(D)	(B)/(A)	(D)/(C)
Gini coefficient	46.1	45.9	49.1	48.9	1.00	1.00
Atkinson 0.5	17.8	17.6	20.3	20.0	0.99	0.99
Atkinson 1	30.8	30.6	34.0	33.8	0.99	0.99
P90/P10	7.4	7.4	7.9	7.9	1.00	1.00

Source: Authors elaboration based on ECUAMOD version v1.4

Finally, ex-ante policy evaluation should also take into consideration potential effects of counterfactual reforms on work incentives. Here, we compare Marginal Effective Tax Rates (METRs) under the baseline and reform scenarios based on unadjusted and adjusted survey data. METRs measure the proportion of a marginal increase in earnings that would be taxed away because of increased taxes, social insurance contributions and benefit withdrawal. METRs provide thus an indicator of incentives at the intensive margin of labour supply, i.e. incentives to work more or earn more.

We follow Jara and Tumino (2013) and calculate METR as follows. First, household disposable income is calculated for each household in the survey. Then, for each earner in the household, separately, individual earnings are increased by 3 per cent and the corresponding household disposable income following the increase is computed. More formally, the METR of individual i in household h can then be represented by:

$$METR_i = 1 - \frac{Y_h^1 - Y_h^0}{E_i^1 - E_i^0} ,$$

where the numerator measures the change in household disposable income before (Y_h^0) and after (Y_h^1) the increase in individual earnings (E_i) and the denominator is equal to the increase in earnings itself.

Table 9 compares METRs under the baseline and reform scenarios based on unadjusted and adjusted survey data. Results are presented for formal workers only because METRs are equal to zero for informal workers, as following an increase in earnings they would pay no additional SICs or taxes, and the main social assistance benefit is not means-tested but proxy means-tested, meaning the increase in earnings would be fully reflected by the same increase in disposable income. In the baseline scenario, mean METRs are only slightly higher based on the top-income adjusted data (8.8 per cent compared to 8.4 per cent). The median, 25th and 75th percentiles of METR is the same in both cases. The change in mean METRs following our hypothetical reform is also similar in magnitude with the unadjusted and adjusted data, representing only around a 2 per cent increase. Therefore, on average, our hypothetical reform does not seem to affect formal workers incentives to earn more.

Table 9. Marginal effective tax rates – formal workers (2011)

| | ECUAMOD | | ECUAMOD adjusted | | Ratios | |
| | Baseline | Reform | Baseline | Reform | | |
	(A)	(B)	(C)	(D)	(A)/(B)	(C)/(D)
Mean	8.4	8.6	8.8	9.0	1.02	1.03
Median	9.1	9.1	9.1	9.1	1.00	1.00
P25	4.4	4.4	4.4	4.4	1.00	1.00
P75	11.1	11.1	11.2	11.2	1.00	1.00

Source: Authors elaboration based on ECUAMOD version v1.4

Due to the particularities in the design on personal income tax in Ecuador (high exemption threshold and presence of deductions from personal expenditures), it seems important to look at the effect of our reform on METRs across the earnings distribution of formal workers. Figure 2 presents Mean METRs by earning deciles based on the unadjusted and adjusted survey data (blue part of the bars), as well as the effect of our hypothetical reform under both cases (white part of the bars).

Figure 2. Mean METR by earnings deciles – formal workers (2011)
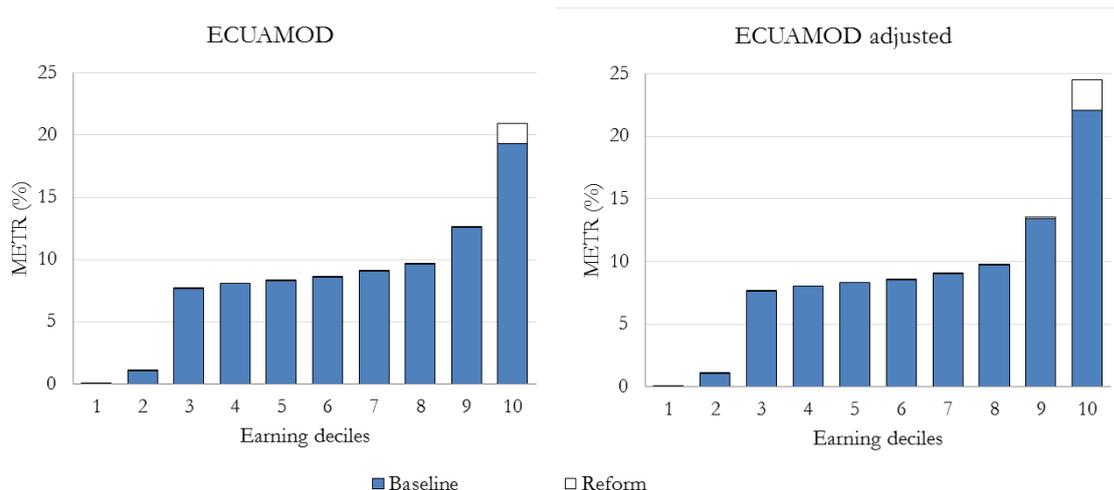
Source: Authors elaboration based on ECUAMOD version v1.4.

Our analysis provides a number of interesting findings. First METR are very low for formal workers in the bottom two deciles of the earnings distribution. This is most likely related to the fact that formal part-time employees pay SICs proportionally with respect to their working time. Moreover, the majority of agricultural workers are affiliated to a special social security regime with a very low contribution rate is equal to 2.5 per cent of 22.5 per cent of the minimum wage. Second, METR of formal workers in deciles three to eight range between 7 to 10 per cent and are driven by the contribution of SICs to METR. Third, mean METR are much larger in the top two earning deciles due to the contribution of income tax and the adjustment of top incomes has a clear effect of METR for these groups. For the ninth decile, mean METR equals 12.5 per cent based on the unadjusted data and 13.5 per cent when top incomes are adjusted. The difference is larger for the tenth decile, where based on the unadjusted data mean METR equals 19.3 per cent compared to 22.1 per cent with the top income adjusted data. Finally, our illustrative reform affects mostly work incentives of the top earnings decile group. Based on the unadjusted data, mean METR would

increase by 1.6 percentage points, whereas a 2.4 points increase would be observed at the top of the distribution based on the adjusted data.

5   Conclusion

Over the last years, the growing literature on top incomes has stressed the importance of going beyond household survey data for the analysis of income inequality. The use of administrative tax records has the advantage of helping to overcome the problem of top income under-coverage characterizing survey data. From a policy perspective, combining survey and tax return data would not only contribute to provide a better picture of income inequality but would also improve the ex-ante evaluation of tax reforms, which could be of particular interest in developing countries.

The present paper makes use of income tax return data from Ecuador to adjust top incomes of formal workers in household survey data, which is then used for ex-ante evaluation of policy reforms. Our study contributes to the literature on income inequality and top incomes in developing economies in two main respects. First, we combine household survey data with income tax records to provide a picture of the effect of correcting top incomes on income inequality. Second, we exploit tax-benefit microsimulation techniques to assess the effect of a counterfactual reform scenario aiming to make personal income tax more progressive, before and after correcting top incomes in survey data.

Our study provides a number of interesting findings. First, inequality measured by the Gini coefficient from disposable income would increase by 3 points following the adjustment of top incomes in household survey data. Second, correcting top incomes information in the survey improves the fit of our simulations of personal income tax, allowing us to capture around 98 per cent of income tax revenue in Ecuador in 2011. Finally, ECUAMOD simulations of a counterfactual policy reform which makes personal income tax more progressive shows that based on the unadjusted data, the model would capture well the percentage change in tax revenue; however, it would underestimate the additional tax revenue amount by 22 per cent compared to results based on the adjusted data. Our hypothetical reform would have very minor effects on income inequality and would decrease work incentives only slightly. In particular, work incentives would mostly be affected at the top of the earnings distribution with the top-incomes adjusted data showing a larger increase in Marginal Effective Tax Rates (2.4. points compared to 1.6 points with the unadjusted data).

Adjusting top incomes in household survey data based on tax records in developing countries presents however a number of challenges. First and foremost, contrary to the recent trends observed in developed countries where access to administrative data for research has become less stringent, access to this type of data is extremely restrictive in developing countries. Second, although income tax return data has the advantage of providing more precise income information, it does so only for a fraction of the working population in developing countries, namely workers in the formal sector who are filing taxes. The larger the informal sector in the economy, the smaller the fraction of the population that would be captured by tax records. Third, combining household survey data with income tax return data in developing countries requires, therefore, identifying the population in the survey that matches the best the population captured by tax records. This requires detailed information in the survey about affiliation to social security or other indicators allowing to identify workers in the informal sector to compare them with the population captured by tax records. Finally, it is essential to construct comparable income concepts in both survey data and tax records in order to be able to combine both data sources. This might not always be possible

depending on unit of analysis (e.g. the tax unit might not be the individual in all countries) and the level of disaggregation in income tax return data (e.g. whether net or gross incomes are available).

Due to the difficulties in combining household survey data and tax records, a number of caveats should be considered in our analysis. First, our approach consisted in adjusting only top incomes of formal employees in the survey. The decision to limit the adjustment to employees was driven by the fact that it was not possible to construct a harmonized concept of self-employment income gross of social insurance contributions in the income tax return data. Further work is needed to analyse the effect of adjusting incomes for the self-employed. In particular, preliminary analysis based on self-employment income net of social insurance contributions points to a problem of underreporting throughout the income distribution rather than only under-coverage of top incomes. Second, our study has not considered the effect of another important source of information for the simulation of personal income tax in Ecuador, namely deductions from personal expenditures. In ECUAMOD, we use information from household expenditures reported in the survey to calculate deductions from personal expenditures as part of the simulation of personal income tax. Our simulations of personal income tax could be affected by two factors. First, there can be misreporting of household expenditures in the survey which could bias our simulations of personal income tax. Second, our simulations attribute all household expenditures to the person with the highest market income in the household for the purpose of the calculation of deduction from personal expenditures in the simulation of personal income tax. This simplifying assumption could not necessarily reflect the behaviour of household members in terms of declaring deductions from personal expenditures for tax purposes. Income tax return data from Ecuador contains information on deductions from personal expenditures, which could, in some way, be used to validate our assumption.

Despite these shortcomings, the analysis presented in this paper provides encouraging results on the possibility of enriching household survey data with tax records. Future research should focus on the possibility of adjusting self-employment income in the survey based on information from tax records, and on the effect of information from expenditures and the assumptions made to calculate deductions for the purpose of income tax simulations. Finally, the approach proposed in this paper to adjust top incomes is rather simple and aims at exploiting limited information available from tax records. Other, more sophisticated statistical techniques should be used for sensitivity tests. In particular, in countries where income tax returns data contains information on socio-demographic characteristics, these should be exploited to combine household survey data and tax records.

References

Abowd, J. M., and Stinson, M. H. (2013). 'Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data'. *Review of Economics and Statistics*, Vol XCV, no. 5, pp. 1451–67.

Alfons, A., Templ, M. and Filzmoser, P. (2013). 'Robust estimation of economic indicators from survey samples based on Pareto tail modelling'. *Journal of the Royal Statistical Society*, Series C, 62, 271–86.

Alvaredo, F. (2011). 'A note on the relationship between top income shares and the Gini coefficient'. *Economic Letters*, 110, 274–7.

Alvaredo, F., Atkinson, A. B., Piketty, T. and Saez, E. (2013). 'The top 1 percent in international and historical perspective'. *Journal of Economic Perspectives*, 27, 3–20.

Alvaredo, F. and Londoño Velez, J. (2013). 'High incomes and personal taxation in a developing economy: Colombia 1993–2010'. CEQ Working Paper no. 12. New Orleans, LA: CEQ Institute, Tulane University.

Atkinson, A. B. (2005). 'Top incomes in the UK over the twentieth century'. *Journal of the Royal Statistical Society*, Series A, 168, 325–43.

Atkinson, A., and T. Piketty (eds) (2007). *Top Incomes over the Twentieth Century: A Contrast between Continental European and English-Speaking Countries*. Oxford: Oxford University Press.

Atkinson, A., and T. Piketty (eds) (2010). *Top Incomes: A Global Perspective*. Oxford University Press.

Atkinson, A. B., Piketty, T. and Saez, E. (2011). 'Top incomes in the long run of history'. *Journal of Economic Literature*, 49, 3–71.

Bach, S., Corneo, G. and Steiner, V. (2009). 'From bottom to top. The entire income distribution in Germany, 1992–2003'. *Review of Income and Wealth*, 55, 303–30.

Bargain, O., Jara, H. X. and Rodriguez, D. (2017). 'Learning from your neighbour: swapping tax-benefit systems in Latin America'. *Journal of Economic Inequality*, 15(4), pp.369–92.

Bollinger, C. R. (1998). 'Measurement Error in the CPS: A non-parametric Look'. *Journal of Labor Economics*, Vol. 16, no. 3, pp 576–94.

Bound, J., and Kreuger, A. B. (1991). 'The Extent of Measurment Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?' *Journal of Labor Economics*, Vol 9, pp. 1–24.

Burkhauser, R. V., Feng, S., Jenkins, S. P. and Larrimore, J. (2012). 'Recent trends in top income shares in the USA: reconciling estimates from March CPS and IRS tax return data'. *Review of Economics and Statistics*, vol. 94, pp. 371–88.

Burkhauser, R. V., Herault, N., Jenkins, S. P. and Wilkins, R. (2016). 'What has been happening to UK income inequality since the mid-1990s? Answers from reconciled and combined household survey and tax return data'. ISER Working Paper no. 2016-03. Colchester: Institute for Social and Economic Research.

Canelas, C. (2014). 'Minimum wage and informality in Ecuador'. WIDER Working Paper 2014/006. Helsinki: UNU-WIDER.

Cano, L. (2014). 'Top Income Shares in a Growing South American Economy: Ecuador 2004–2011'. Working Paper University Toulouse 1 Capitole 1(2):1-60. Toulouse: University of Toulouse.

Cowell, F. and Flachaire, E. (2007). 'Income distribution and inequality measurement: the problem of extreme values'. *Journal of Econometrics*, 141, 1044–72.

Department for Work and Pensions (DWP) (2015). *Households Below Average Income An Analysis of the Income Distribution 1994/95–2013/14*. Analysis. London: Department for Work and Pensions.

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

Jara, H.X., Varela, C., Cuesta, M., and Amores, C. (2017). *SOUTHMOD Country Report Ecuador. ECUAMOD v1.0. 2011-2016*. Report. Helsinki: UNU-WIDER.

Jara, H.X., and Varela, C. (2018). 'Tax-benefit microsimulation and income redistribution in Ecuador'. *International Journal of Microsimulation* (forthcoming).

Jenkins, S. P. (2017), 'Pareto Models, Top Incomes and Recent Trends in UK Income Inequality', *Economica*, vol. 84, pp. 261–89.

Kapteyn, A., and Ypma, J. Y. (2007). 'Measurement Error and Misclassi.cation: A Comparison of Survey and Administrative Data'. *Journal of Labor Economics*, Vol 25, pp. 513–50.

Meijer, E., Rohwedder, S., and Wansbeek, T. (2012). 'Measurement Error in Earnings Data: Using a Mixture Model Approach to Combine Survey and Register Data'. *Journal of Business and Economic Statistics*, Vol 30, pp. 191–201.

Meriküll, J., and Staehr, K. (2010). 'Unreported employment and envelope wages in mid-transition: Comparing developments and causes in the Baltic countries'. *Comparative Economic Studies* 52, pp. 637–70.

Piketty, T. (ed.) (2001). *Les Hauts Revenus en France au XXe Siècle: Inégalités et Redistributions, 1901–1998*. Paris: Grasset.

Piketty, T., and E. Saez (2003). 'Income Inequality in the United States, 1993–1998'. *Quarterly Journal of Economics*, 118 (1) (1): 1–39.

Rossignolo, D., Oliva, N., and Villacreses, N. (2016). 'Cálculo de la concentración de los altos ingresos sobre la base de los datos impositivos: un análisis para el Ecuador', serie Macroeconomía del Desarrollo, Nº 184 (LC/L.4278). Santiago: Comisión Económica para América Latina y el Caribe (CEPAL).

SRI (2010). Resolución No. NAC- DGERCGC10-00733. Registro Oficial No 352 – Suplemento. Quito: Servicio de Rentas Internas.

Sutherland, H., and F. Figari (2013). 'EUROMOD: the European Union Tax-benefit Microsimulation Model'. *International Journal of Microsimulation*, 6(1): 4–26.

Williams, C. C., and Padmore, J. (2013). 'Envelope wages in the European Union'. *International Labour Review 152*, No. 3−4, pp. 411-430.

Woolfson, C. (2007). 'Pushing the envelope: The 'informalization' of labour in post-communist new EU member states'. *Work, Employment and Society* 21 (3), pp. 551–64.

Appendix

Table A.1. Personal income tax schedule in 2011

Tax band	Lower limit	Upper limit	Marginal tax rate (%) baseline
1	0	9,210	0
2	9,210	11,730	5
3	11,730	14,670	10
4	14,670	17,610	12
5	17,610	35,210	15
6	35,210	52,810	20
7	52,810	70,420	25
8	70,420	93,890	30
9	93,890	-	35

Source: SRI (2010).

Table A.2. Top income adjustments and the redistributive effect of the tax-benefit system (2011)

	Market income	Disposable income (DPI)	Difference
Gini coefficient			
ECUAMOD	50.2	46.1	4.1
ECUAMOD adjusted	53.4	49.1	4.3

Source: Authors' elaboration based on ECUAMOD version v1.4 and income tax return data from SRI.