Working paper 16
15 July 2017

UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

Expert meeting on measuring poverty and equality
26-27 September 2017, Budva, Montenegro
Session D: Methodological issues in measuring economic inequalities

Towards a more realistic estimate of the income distribution in Mexico

Prepared by INEGI [1]

Abstract

As in many other instances, sound decision-making in fighting poverty based on high quality information leads to efficiencies. More accurate measures of the proportion of households in poverty conditions, or at risk, will result in appropriately sized budgetary allocations. In contrast, when faced with larger uncertainties, excessive appropriations may be made “to be on the safe side”, distracting badly needed funds from attention of the really poor or other pressing needs.

The distribution of income is an important input, if not the most important, when measuring poverty. This is usually computed from data collected through household surveys on income. However, the declaration of income in these statistical exercises is flawed to a certain (but unknown) extent. Additionally, random sample selection procedures may leave out very small sub-populations which accrue a disproportionately large part of household income. Oversampling richer households does not completely eliminate this shortcoming. Furthermore, even when included in the sample, there is no guarantee they will volunteer their information in any accurate manner. In general, this results in an underestimated total household income for a country or a region.

There are however other sources, albeit limited, which reflect one or more aspects of an income distribution. We will consider two in particular. In the first place, the institutional sector accounts from recent versions of SNA provide a single figure for the aggregation of all income households receive in one calendar year, either directly or indirectly. Note that we cannot conclude how such figure is distributed among households. Instead, we can get from this figure an idea of the above mentioned underestimation resulting from household surveys from any country in the world. Latin America seems to have lead the way with many

[1] Ms. Alfredo Bustos and Mr. Gerardo Leyva.

imaginative approaches to reconcile both sources. Many countries in the region, and even ECLAC, are considering however to phase out such attempts since a consensus criterion regarding which approach is best in any given situation is missing. Oddly, this happens when other regions are wondering whether to correct or not income values from surveys in order to get so-called distributional national accounts.

The second complementary source we will briefly look into is given by tax records or some summary measures derived from them. In Mexico these give a reasonably good approximation as to the income of high earning households which may be missing from the survey, as was mentioned above. For our approach we use only two summary measures which tax authorities may be more willing to share. Firstly, we use the relative size of the population whose income is above a given threshold, for a number of alternative thresholds. Secondly, we require the average income of households or individuals in the sub-populations so defined.

Our approach seeks to reconcile all three information sources to which we have referred above. This is achieved by fitting at least one parametric family of distributions to the survey data. To this end we developed a new criterion we call Maximum Constrained Pseudo Likelihood (MCPL) to guide us in the search for optimal solutions. As every likelihood method, it is based on consideration of parametric families of distributions. It is a pseudo likelihood because only part of the information provided by the sampling design is considered; namely, the expansion factors. And is constrained since this is a convenient way to introduce the information provided by the two complementary sources.

Since we search for values of the parameters which optimize the MCPL criterion for each family, and since we impose two constraints, one for each complementary data source, we have considered families with at least three parameters. In this paper, we tried two families often reported in the income literature: the 3-parameter Generalized Gamma (GG), and the 4-parameter Type II Generalized Beta (GB2). The two constraints require that (i) the value of the mean of the distribution expressed as a function of the parameters equals the average household income according to SNA, and (ii) that the average income of the 0.001% and 0.0001% higher income households in the model equals the same average from tax records.

In all four cases (two distributions, two thresholds) the fitted models happen to be remarkably similar. In fact, the difference between criterion values is relatively small. However, the *maximum maximorum* corresponds to the combination GG and the 0.0001% threshold. While the survey reported a value for the Gini coefficient of 0.454 for Mexico in 2012, the corresponding values from our approach lie between 0.61 and 0.64. This highlights the underestimation of inequality that results from income household surveys where upper incomes are not adequately represented in the sample. It must be emphasized that no attempt should be made at comparing these figures with underestimated ones from other countries or regions. On the other hand, the proportion of households whose income lies below an income threshold (termed in Mexico "Welfare line") goes down from 51.6% according to CONEVAL, Mexican institution in charge of measuring poverty, to 28.0% for the best fitting model. The corresponding values for the "Minimum welfare line" are down from 20% to 8.1%. This is so once income under-reporting in the survey is allowed for.

Once the same approach is used, comparisons over time are allowed even when successive rounds of the survey are not deemed comparable. In Mexico, minor operation and training changes were introduced in the 2015 survey but their consequences in terms of total or average current income caught everyone by surprise. Preliminary results from our approach indicate that the Gini coefficient went from a low 0.445 to 0.64 again. The proportion of households below the welfare line reached 30.4% and below the minimum welfare line went to 11.7%, in accordance with trends apparent in all approaches.

Introduction

Mexico is a country that exhibits important deficiencies and inequalities of all types, but the numerical expressions of these problems have not always been sufficiently precise. In fact, the basic statistical information available has led us with certainty to underestimate inequality and possibly also to overestimate poverty. For example, data from the National Household Income and Expenditure Survey (ENIGH, for its acronym in Spanish) indicate that by 2012 about 44% of households received incomes below the line established by the National Council for the Assessment of Social Development Policy (CONEVAL, for its acronym in Spanish)[2]. At the same time, as published by the National Institute of Statistics and Geography (INEGI), 10% of households with higher incomes perceived "only" 19 times what the 10% with the lowest incomes did. However, the study of the distribution of income by ENIGH is limited mainly by two causes:

1. Incomes of households in the survey seem to be higher than what they report (what we shall refer to as "underreporting"), so that income poverty seems higher since households whose actual income is higher than the threshold are considered in poverty, and
2. There are households, not included in the ENIGH sample, with incomes much higher than any of those reported (what we call "truncation"). Therefore, inequality is underestimated when ENIGH alone is used since the difference between large and small incomes becomes smaller than it would otherwise be.

In this paper, we exemplify the application of a method for the statistical adjustment of models to survey data, which also uses anonymized tax information and national accounts results, and, recognizing the simultaneous presence of the above limitations, improves on the estimation of an income distribution (see Bustos(2015a, b)). The still preliminary results of the optimal fit show an estimate of nearly 30%[3] of households whose incomes lie below CONEVAL welfare line. This estimate is close to two thirds of what CONEVAL estimated using the uncorrected income reported by the survey. Similarly, a significantly greater quantification of inequality is obtained, with a ratio of revenues of the upper income decile to the first of almost 53 times.

The discussion of economic inequality between people and households, both in wealth and in income, has recently returned to the center of the attention of politicians and of academics. The book *Capital in the Twenty-First Century* (Piketty(2015)) appears in this context and exemplifies the interest in the topic, by becoming a bestseller in Europe and the United States. Other prominent economists such as Joseph Stiglitz (2013, 2015a, 2015b), Paul Krugman (2009) and Tony Atkinson

---

[2] For the sake of consistency with the rest of the paper and for illustrative purposes only, this figure was obtained by calculating the poverty incidence from CONEVAL "welfare line", using information from ENIGH and adjusting it to quarters and households. This can lead to differences with the poverty incidence obtained using the Socioeconomic Conditions Module, as CONEVAL does for calculation of multidimensional poverty.

[3] This is the value that results from calculating the intersection between the “welfare line” and the fitted model. Given that the “welfare line” is built based on food and non food expenditures, rather than on income, we don´t need to modify or recalculate its value when we correct the income vector shifting upward the entire distribution. The “welfare line” is calculated by CONEVAL using a version of the Orchansky coefficient in which total expenditure is used instead of total income; which makes the level of the “welfare line” independent of the absolute level of income of the percentile of reference (the one whose observed food expenditure is just enough to comply with the minimum nutritional requirements). Since we have no reason to believe that expenditures, specifically at the level observed at the group of reference, are under-reported, we don´t find any justification to modify the “welfare line”.

(2015), have argued that uncontrolled inequality is not only a reflection of an unjust economic system, but ultimately is an obstacle to efficiency and economic growth. Authors such as Wilkinson and Pickett go further to point out that the most healthy, happy and functional societies are those where economic inequality is less acute. In our country, the note *Extreme inequality in Mexico: Concentration of economic and political power* (Esquivel (2015)), and the work *High income, optimal taxation and possible tax collection* (Campos, et al. (2014)), set off a wave of reflections and a renewed focus on inequality in Mexico.

Moreover, attention to the challenges that high inequality imposes on the progress of mankind has led to incorporate the subject into the Sustainable Development Goals (SDGs[4]) with which the United Nations proposes to lead the development efforts around the world towards 2030. Thus, SDG number 10 refers to "reducing inequality between countries and within them". Moreover, during the Fifth Global Forum of the OECD on “Statistics, knowledge and policy: Transforming policies, changing lives", held in Guadalajara, Mexico, in 2015[5], the discussion about different forms of inequality played a central role. Among other things, it was mentioned that inequality in Latin America remains among the highest in the world, even though traditional statistical tools fail to cover adequately the "mega-rich". In this regard, it is noted that the High Level Expert Group on the Measurement of Economic Performance and Social Progress (which continues the Stiglitz-Sen-Fitoussi Commission work) has proposed, among other things, to work on the integration of microeconomic and macroeconomic sources so that statistics shed a more realistic image of what happens to inequality in countries.[6]

Thus, the renewed interest in inequality necessarily comes hand in hand with major efforts to measure it in the best way possible. However, proper measurement is far less trivial than it seems. The analysis of inequality often resorts to household income surveys or tax records, but both sources have limitations that, to a greater or lesser extent, yield inaccurate figures on the subject. Surveys do not capture well enough the top of an income distribution and tax records can present an incomplete picture, especially for the lower income groups or those that escape tax control actions.

If we put the spotlight on household income surveys, we note that -pursuant to what was proposed originally in Cortés (2001) and in Leyva (2004)- they are often affected by two problems: underreporting and truncation. Consequently, total income estimated by a household survey normally shows a more or less large deficit when compared with other figures of total household income considered more reliable, such as household income reported by the System of National Accounts (SNA)[7].

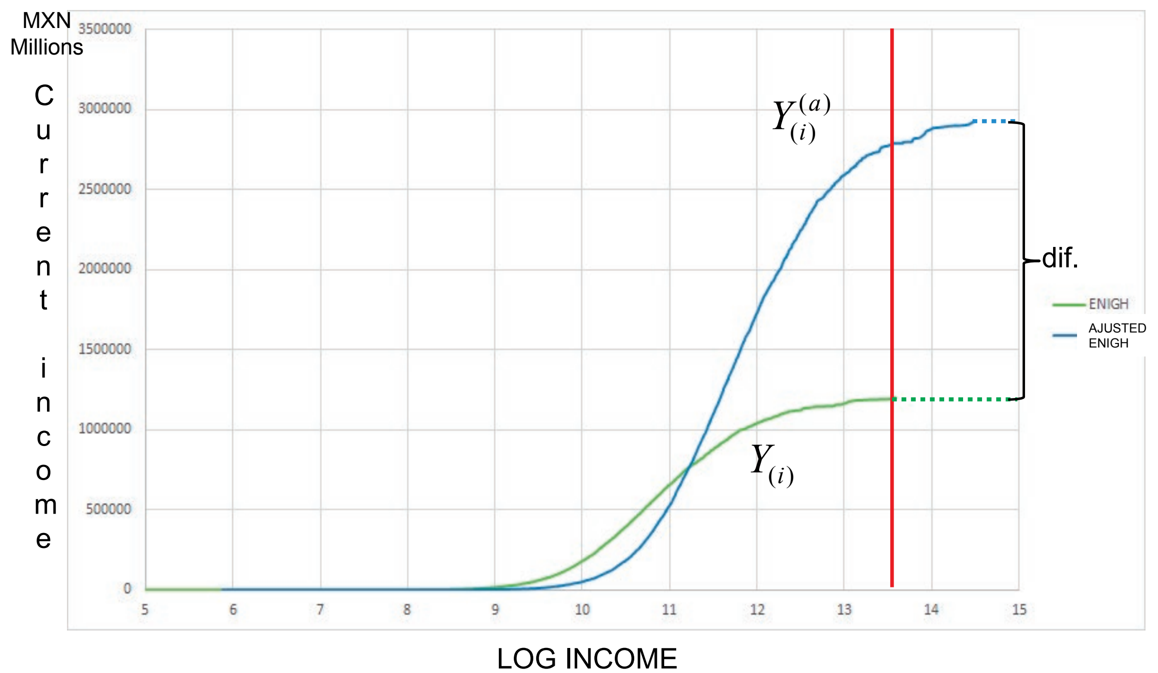
[4] <https://sustainabledevelopment.un.org/index.php?menu=1300>

[5] See especially the intervention of Nora Lustig in session 4.a. "Plenary: High Level Dialogue on 'What are the implications of increasing inequality?'", <http://www.oecd-5wf.mx/>.

[6] <http://www.oecd.org/statistics/measuring-economic-social-progress/Main%20conclusions%20HLEG%20meeting%20Jan%202014.pdf#2>

[7] The "Adjustment to National Accounts" is an extended practice in the measurement of poverty, at least in Latin America, where Altimir's (1987) work, adopted by ECLAC, has perhaps been most influential. In Mexico, we can trace the first efforts in this matter with what was done by Ifigenia Martínez (1970), Enrique de Alba (1967), Enrique Hernández Laos (1991) and Julio Boltvinik (1999). OECD is currently developing a project to reconcile household income records from surveys and national accounts. It should be noted that in this context the case of Mexico stands out as the one that shows the most extreme differences (Fesseau, 2013).

Figure 1. Accumulated expanded income before and after adjusting to System of National Accounts figures, 2012.



Source: Own from the ENIGH 2012 database. Calculations in this case are the result of multiplying each survey income value by 2.43; i.e., a proportional adjustment to the sample values.

In order to correct such deficit, amending or adjusting incomes declared in the survey has become common practice. Mostly, this is done so that their expanded sum matches the total household income from SNA. That is, since the total household income from SNA is usually higher than the one derived from surveys, the adjustment to accounts involves distributing the difference between households in the sample (see Fig. 1). Many procedures have been suggested in the literature for carrying out this task. All of them, however, are based on more or less arbitrary assumptions. In the Mexican case, the deficit has always been important, so the choice of one method over another becomes of central importance, since it will yield not only a different version of the magnitude of inequality but also of poverty.[8]

In the absence of solid and convincing methods for determining which part of the deficit is due to income underreporting, and which to truncation, it has been usual to act as if only one of the causes of the deficit is present. Thus, when it is assumed that truncation does not the cause the deficit, the part of the income difference corresponding to the super-rich is distributed among households included in the sample of the survey, artificially increasing the income of this subpopulation. Poverty is easily underestimated when income from super-rich households is transferred (but only

[8] It has already been pointed out that, when it is decided to exclude from the adjustment or reconstruction of the distribution of income to the population with lower incomes, poverty is being overestimated.

on paper!) along the whole distribution reported by the survey, including those who would otherwise be considered in poverty.

On the other hand, when truncation is assumed to be the sole cause of the oft-cited deficit, this figure will be distributed only among high-income households (e.g., higher-income 10 or 20%), using some arbitrary allocation rule[9]. Consequently, inequality becomes even more extreme (again, on paper) as the solution to truncation ends bulging income of households at the top of the distribution. This, under the dubious assumption that households at the lower 80 to 90% part accurately declared their income in the survey, and should not be subjected to any correction.

Our proposal

In view of the above, the research area at INEGI initiated a project, still in progress, to propose alternatives for the study of the income distribution in Mexico, using all available information on the subject (surveys, national accounts and tax records). Although our results are still preliminary, we want to open up our proposal for discussion by the community interested in these issues. Publication will help identify shortcomings to overcome, as well as avenues unexplored by us, always with the aim of achieving the most realistic possible statistical representation of the distribution of income in Mexico.

Method

We decided to follow an alternative path, particularly when focus is placed only on the income distribution. In principle, we decided not to modify declared income values in the survey but rather to use them as one among other sources to produce a likelier distribution of income.

Figure 2. Outline of the MCPLL criterion and examples of data sources considered.

$$\text{Model}: f(y;\theta) \Rightarrow \begin{cases} \ell(\underline{\theta};\underline{Y}_{(i)}) = \ln\left(f(\underline{Y}_{(i)};\underline{\theta})\right) \\ h(\underline{\theta}) \end{cases};$$

$$\text{Criterion}: \underset{\underline{\theta},\underline{\lambda}}{Max}\left\{ \sum_{i=1}^{n} \frac{1}{\pi_{(i)}} \ell(\underline{\theta};\underline{Y}_{(i)}) - \underline{\lambda}'\left(\underline{h}(\underline{\theta}) - \underline{c}\right) \right\}$$

$$\text{ENIGH}: \begin{cases} Y_{(i)} \\ \pi_{(i)} \end{cases}, i = 1, \ldots, n;$$

$$\text{SCN}: c_1 = \text{Total}(\hat{Y}_{SNC});$$

$$\text{SAT}: c_2 = \text{Average}(Y_{Max-k}, \ldots, Y_{Max});$$

Since one of our main concerns is to reduce the arbitrariness of other approaches, we require a criterion to guide us when comparing between alternative models fitted to the data, so the best among various tested can be chosen. This criterion includes three conditions: (1) is based on unmodified survey values, (2) takes into account the sampling design as far as this is possible, and (3) makes compatible survey results with figures from the SNA and from the Mexican tax authority (SAT, for its Spanish acronym), in order to bring the results of this exercise closer to "reality".

[9] Here, it is important to take into account that only "income" is adjusted, the relationship between this and other economic variables (consumption, wealth, etc.) will be distorted.

Figure 2 provides an outline of the criterion and exemplifies the sources that can be considered (See Bustos (2015a)).

On top of three information sources, we included alternative families of parametric distributional models, thus avoiding the arbitrariness of sticking to one preferred model. These families have been widely used in the income literature (see Kleiber et al. (2003)). The family being tried determines not only the form assumed by the likelihood but also that of a number of constraints imposed on the parameter values. These constraints allow us to consider the effect of non-survey data sources on the fitted distribution.

The proposed criterion consists of assembling procedures already available but scattered in the literature: (1) maximum log-likelihood, but (2) pseudo-, since it is not possible in general to include all the consequences of the sample design, and (3) parameter values are constrained in order to account for contributions from other sources. Hence the name: Maximum Constrained Pseudo Log-Likelihood (MCPLL).

We knew we faced inhomogeneity challenges in terms of both observation units (households vs. people), and of income concepts between SAT declared income and those covered by the ENIGH and the SNA. However, if we could pull this off, we would be able to produce reconstructions of the entire Mexican income distribution, compatible with all three information sources, for which we know of no precedent. This was achieved by considering known income values of SAT records and SNA as conditions to be met by the parameter values of the MCPLL fitted models.

Specifically, SNA information was included by requiring that the value of the mean of each fitted distribution, expressed as a function of the parameters, equals that of the average household income according to SNA (see first row of Table 1).

Table 1. Summary of constraints considered.

Concept (Source)	Constraint	Interpretation
Average household income (SNA)	$h_1(\underline{\theta}) = E[Y \mid \underline{\theta}] = c_1$	Mean income for fitted model equals average household income, according to SNA.
Household Integral (SAT)	$h_2(\underline{\theta}) = \int_{\varphi_\alpha}^{\infty} f_Y(y\|\underline{\theta})dy = \alpha = c_2$	Proportion of households whose income is greater than threshold φ_α is, according to the model, equal to similar number from SAT.
Income Integral (SAT)	$h_3(\underline{\theta}) = \frac{1}{\alpha}\int_{\varphi_\alpha}^{\infty} y f_Y(y\|\underline{\theta})dy = M = c_3$	Mean income for households whose income is greater than threshold φ_α is, according to the model, equal to average household income from SAT.

Regarding the SAT data, we considered a number of options. Since individual tax returns were made available to us, we thought of merely adding them to the survey records with an expansion weight equal to one. However, the above mentioned inconsistencies would be exacerbated specially in the region where both data sets join, which made it impossible to follow this course of action. Given that the main contribution of the SAT data would be to describe the presence of extreme situations at the upper end of the distribution, we decided to impose instead additional constraints derived from tax records and which are to be satisfied by the parameter values. Note that at very high income levels, the difference between individuals and households becomes blurred, which acts to our benefit. In addition, regarding extraordinarily high incomes, differences between SAT records and INEGI sources in terms of consideration of other sources of income (e.g., imputed housing rent or in-kind transfers) would make little or no difference in our proof of concept.

These considerations led to two types of constraint, summarized in rows two and three, in table 1. To begin with, we identified an income threshold, denoted by φ_α, from the SAT records. This value is such that only a small proportion (α, say) of "households" declares an income greater than the threshold . In the first instance, referred to as the "household integral constraint", we required that the fraction of households in the upper tail of the fitted model (i.e., beyond income threshold φ_α) also equals α. In the second case, referred to as the "income integral constraint", the (conditional) average income in the upper tail of the fitted model is forced to attain the same value as that of the group of households declaring an income greater than φ_α. Two things are to be stressed in the latter case. First, the proportion of households in the upper tail may be different from α since it is required only that one average equals the other. Finally, we now know not only that the incomes of a fraction of households lie to the right our threshold but also that the average of those incomes is more or less distant from the threshold. Therefore, we consider this constraint to be more informative than the first. For this reason, our numerical examples in the following section will use this version.

Fig 3. Optimal fitted models					
Generalized Gamma			Type II Generalized Beta		
Fig. 3 GG here			Fig. 3 GB2 here		
	0.0010%	0.0001%		0.0010%	0.0001%
OPTIMAL MCPLL VALUE	-365,349,780	-364,628,437	OPTIMAL MCPLL VALUE	-365,541,743	-364,913,147
$E(X\|\theta)$	✓	✓	$E(X\|\theta)$	✓	✓
$E(X\|X > \varphi_\alpha, \theta)$	✓	✓	$E(X\|X > \varphi_\alpha, \theta)$	✓	✓
GINI COEFFICIENT	0.62	0.64	GINI COEFFICIENT	0.61	0.63
RATIO X/I	52.09	54.16	RATIO X/I	51.66	53.20
Source: Own from ENIGH 2012 Data base.					

Originally, using general numerical optimization routines, five alternative families of distributional models were fitted to survey data according to the above criterion, when only one SNA constraint was considered (Bustos(2015a)). Once tax data was made available to us, the number of constraints grew, which lead us to drop all two-parameter families. Therefore, in what follows we shall only consider the three-parameter Generalized Gamma (GG) and the four-parameter Type II Generalized Beta (GB2) families.

In addition, since the relevant statistical theory to discriminate between models fitted to sample data is still under development, we use other measures on adjustments (as quantiles, ratios and Gini coefficients) to help us assess which model to choose as the "one closest to reality" from among those tested. In other words, although we have not completely done away with arbitrariness, we have achieved a method to overcome limitations of other available methods, to contribute to the discussion and one that yields reasonable results.

Results

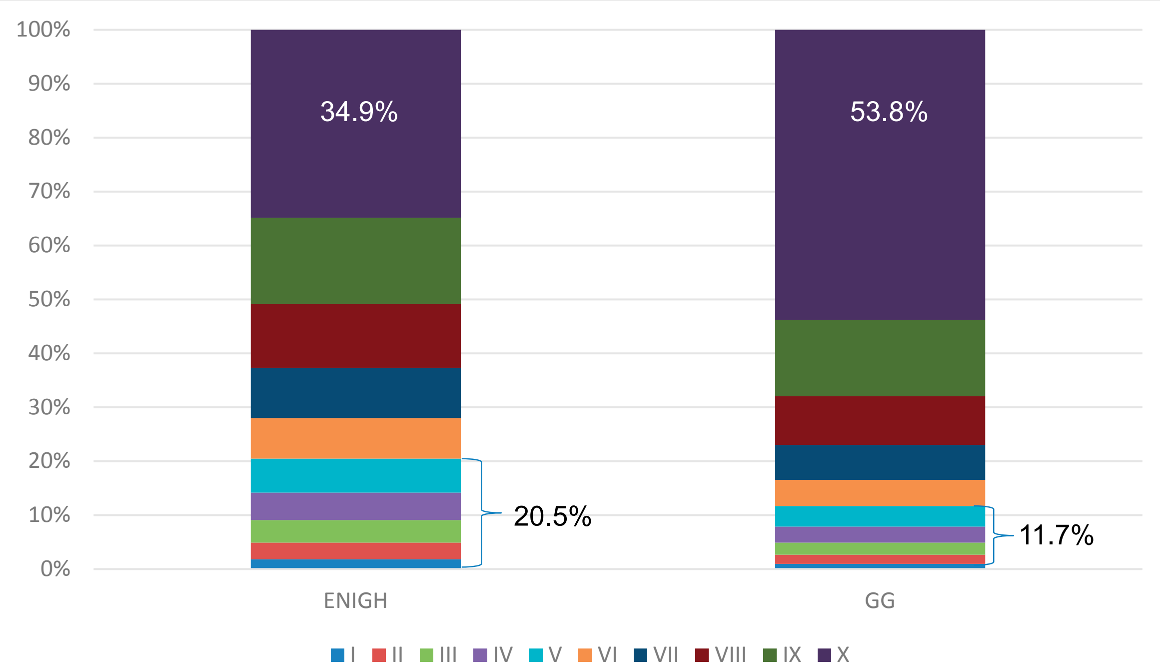
As has already been mentioned, in addition to data from ENIGH and the SNA for 2012, we worked with anonymized information on almost 3 million individual tax returns for the same year. All fitted models consider also the threshold established based on SAT information, and which is such that only 1 in 1’000,000 households has higher quarterly incomes. The determination of the threshold assumes that each higher-income taxpayer counts as a household; it takes into consideration that the total number of households in the country is about 31.5 million. It should be noted that, after a number of trials, this threshold resulted in optimal fitted models, shown in Fig. 3. It is to be stressed that all four fitted models appear to be very similar. For instance, the optimal values of the proposed criterion are close to one another and their graphical representations show little difference, except maybe at the bottom and at the top of the income range.

Now, if we were to accept the model that optimizes the criterion value (see Figure 3), we find that:

- There is evidence of significant and growing underreporting over the entire income range in the survey, as expected. There is also an important truncation at the top. Thus, we see underreporting even for lower income households, although the proportion of underreported income to survey income would increase as higher-income households are considered. Consequently, this more realistic estimate of income distribution in Mexico results in greater inequality but also in a lower monetary poverty.[10]
- The gap between the haves and the have-nots is larger than that estimated with the traditional source, so that the X-th decile would have received in 2012 almost 57 times that obtained by decile I, against 19 times according to ENIGH reported figures. That figure, however, is below the more than 83 times the estimate by Esquivel (2015) for Oxfam.

Figure 4. Distribution of income over population deciles according to ENIGH and to the optimal model (Generalized Gamma, GG) compatible with national accounts and fiscal data, Mexico, 2012.

[10] Monetary poverty is mentioned to indicate that this is the relationship between household income and the monetary value of the Coneval welfare line, in order to make clear the contrast with the official measurement of poverty in Mexico that is multidimensional, In respect of which we do not report any exercise in this article.

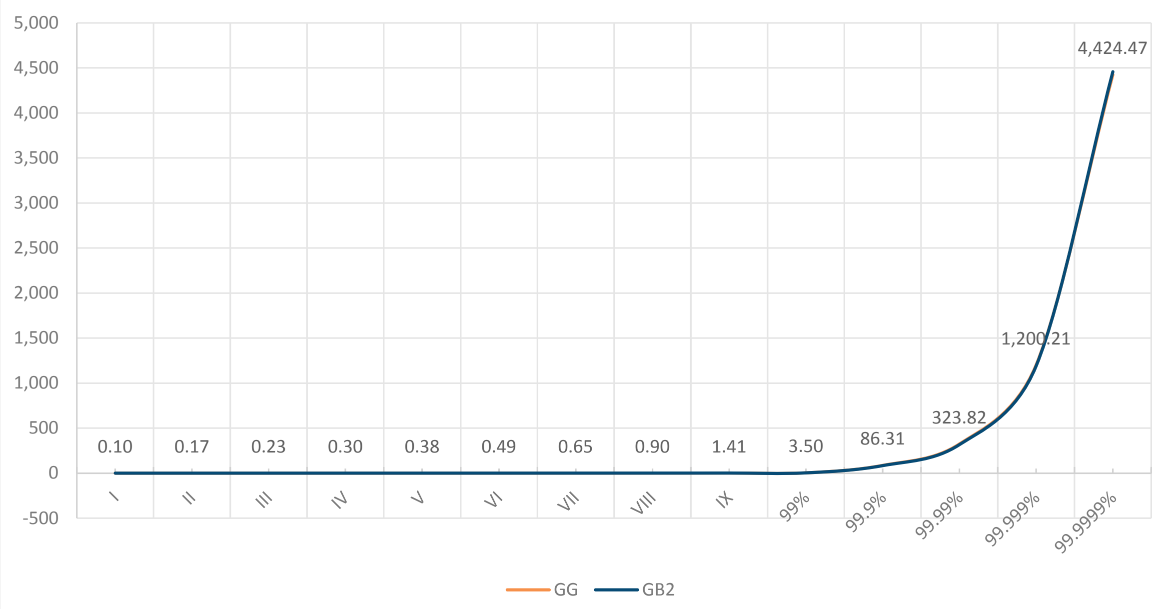


Source: Own calculations from the ENIGH 2012 database.

- The richest 1% of households concentrates almost as much income as the bottom 60%. Indeed, deciles I to VI receive 17.2% of total income, while the X-th decile would receive just over 50%, and the higher 1% would receive 17.3%.
- Even within the top 1% there are important differences. The top 0.1% receives 8.6% of total income; i.e., 86 times its relative size among households. The factor grows to 324 times for the 0.01%; to 1,200 times for the 0.001%, and to 4,424 times for the top 0.0001%, reflecting a significant inequality even within the higher income group of households (see fig. 5)[11].
- This inequality is reflected in a Gini coefficient (indicating less inequality as it approaches zero and greater inequality as it gets closer to one) with a value of 0.630, which contrasts with the 0.453 from the original figures of the ENIGH. It should be noted that the former figure would only be internationally comparable if data for other countries were adjusted using the same methodology (MCPLL).

Figure 5. Ratio of relative income to relative size of households. Selected quantiles.

[11] For the sake of clarity, shorter intervals at the top are shown with equal length.



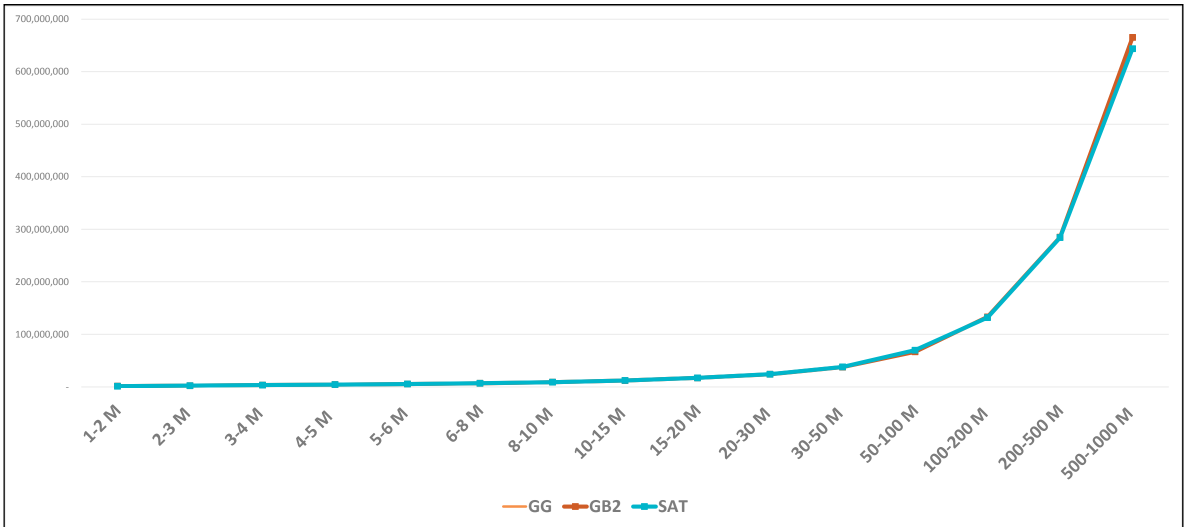
Source: own computations.

- Using the unmodified CONEVAL welfare lines, the methodology suggests that the incidence of income poverty would reach only 30% of households, smaller than the 44% figure reported by CONEVAL. However, even if indeed poverty incidence measurement is lower than would be estimated with uncorrected income figures from ENIGH, which apparently understate household income throughout the distribution, the size of the challenge poverty poses remains overwhelming and focus on it continues to be of paramount importance. Moreover, we recognize that we still have to incorporate additional criteria for a more accurate reconstruction of the lower part of the income distribution; moreover, we have not yet developed a way to link the new distribution with the data gaps (education, food, health, social security, housing and housing services) per household necessary for measuring multidimensional poverty.

Comparison with fiscal data

In order to assess the accuracy of the model at the top of the distribution, two sets of values were obtained for 15 yearly high-income brackets. In the first place, we computed the average household income in each of the income brackets from both the best two fitted models, and from the tax records. Results are summarized in Fig. 6. Over 14 out of 15 brackets all three lines are nearly the same. Only at the very top, for incomes between 500 million and 1000 million Mexican pesos a year, a minor discrepancy is perceptible. We can safely conclude that as far as this measure is concerned there is agreement between models and data.

Figure 6. Income average over selected income brackets.

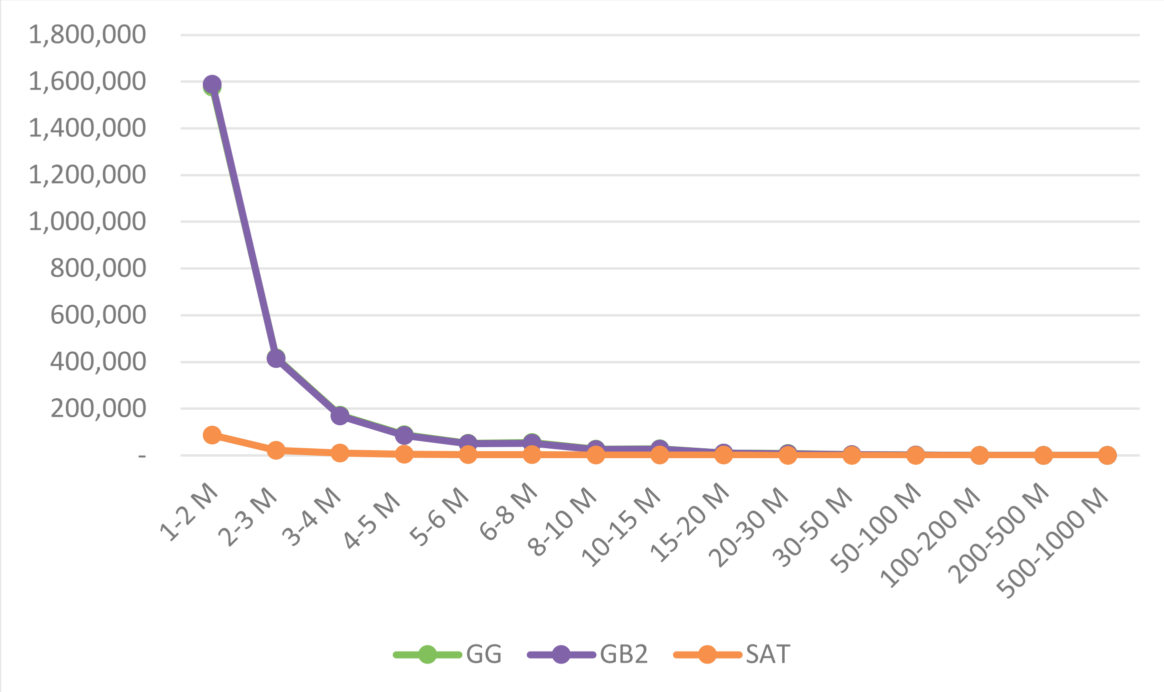


Source: own using ENIGH and SAT 2012 data basis.

Secondly, we computed the number of households within each income bracket, again according to the same three sources. In this instance, an important difference became quickly apparent (see fig. 7). For the first bracket, with yearly incomes between 1 million and 2 million, the number of households according to the model is close to twenty times that of the tax records. Since tax and model numbers decrease rapidly, the difference between them does the same, as income increases. However, in relative terms the discrepancy is more persistent.

Great care has to be exercised when looking at the above results, since they may raise concerns about, say, tax elution. Conceptual and observation units differences may be at play, particularly among lower incomes where the largest differences appear.

Figure 7. Average number of households in selected income brackets.



Source: own using ENIGH and SAT 2012 data basis.

Future work

To help us understand more fully the implications of these results and give us a better idea of how reasonable they are, it is convenient to make international comparisons using the same methodology. To do this, we have approached international agencies to make appropriate comparisons and to get feedback and exchange ideas in order to join efforts in the common interest of better represent reality through statistics. You also need to make comparisons over time for the case of Mexico, which will start soon as we have information of anonymized tax records that this requires. It is also essential to open the debate with national and foreign experts to create synergies that help us better understand income distribution in Mexico, with all its consequences. After all, the role of statistics in society is to describe reality of the most likely way possible.

Final remarks

The statistical description of income distribution coming out from sources like household surveys is subject to limitations such as under-reporting and truncation, which may impair our ability to make realistic assessments about socially relevant phenomena such as inequality and poverty. Most approaches to overcome these limitations normally depend on the assumption that either truncation or under reporting can be more or less neglected. In contrast, our approach doesn't require this kind of assumption, given that it allows data to express itself in such a way that the amounts of truncation and underreporting are a result rather than an assumption. Instead of "adjusting" data from the income survey we offer a method to generate a new and more realistic version of income distribution by using CMPL to combine different data sources while using an optimality criterion which allows us to select the functional form with the best fit. We think this a step forward in the way towards a better statistical representation of the income distribution, even if we also understand there is still some way to go.

References

Alba Guerra, Enrique de, *Información estadística que se requiere para conocer el comportamiento económico de las familias*, vol. II, Banco de México, Departamento de Estudios Económicos, México, 1967.

Altimir, Oscar, "Income Distribution Statistics in Latin America and Their Reliability", Review of Income and Wealth, vol. 33, no. 2, June 1987, pp. 111-155.

Atkinson, Anthony B., "Inequality: What Can Be Done?", Harvard University Press, 2015.

Boltvinik Kalinka, Julio, y Enrique Hernández Laos, *Pobreza y distribución del ingreso en México*, Siglo XXI Editores, México, 1999.

Bustos, Alfredo, "Estimation of the Distribution of Income from Survey Data, Adjusting for Compatibility with Other Sources", Statistical Journal of the IAOS, vol. 31, no. 4, 2015, pp. 565-577.

----, "Estimation of the Distribution of Income from Survey Data, Adjusting for Compatibility with Other Sources", Workshop on Measuring Inequalities of Income and Wealth, Berlin, Germany 15-16 September 2015, http://fr.slideshare.net/StatsCommunications/hleg-thematic-workshop-on-measuring-inequalities-of-income-and-wealth-alfredo-bustos

Campos Vázquez, Raymundo Miguel, Emmanuel Salvador Chávez Jiménez, y Gerardo Esquivel Hernández, *Los ingresos altos, la tributación óptima y la recaudación posible*, Premio Nacional de Finanzas Públicas, 2014. http://www.cefp.gob.mx/portal_archivos/convocatoria/pnfp2014/primerlugarpnfp2014.pdf

Cortés-Cáceres, F., "El cálculo de la pobreza y la desigualdad a partir de la Encuesta Nacional de Ingresos y Gastos de los Hogares", *Comercio Exterior*, vol. 51, núm. 10, 2001, pp. 879-884.

Esquivel, Gerardo, *Desigualdad extrema en México: Concentración del poder económico y político*, Oxfam, México, 2015.

Fesseau, Maryse and Maria Liviana Mattonetti, *Distributional Measures Across Household Groups in a National Accounts Framework: Results from an Experimental Cross-Country Exercise on Household Income, Consumption and Saving*, OECD Statistics Working Papers, no. 2013/04, November 2013 http://dx.doi.org/10.1787/5k3wdjqr775f-en.

Hernández-Laos, Enrique, "Tendencias recientes de la distribución del ingreso en México, 1977-1984", en *La economía mexicana actual: Pobreza y desarrollo incierto*, Universidad Autónoma Metropolitana-Iztapalapa, México, 1991.

Kleiber, Christian, and Samuel Kotz. *Statistical size distributions in economics and actuarial sciences*. Vol. 470. John Wiley & Sons, 2003.

Krugman, Paul, The Conscience of a Liberal, W. W. Norton & Co. Inc., New York, 2009.

Leyva-Parra, Gerardo, *El ajuste del ingreso de la ENIGH con la contabilidad nacional y la medición de la pobreza en México*, Sedesol, Serie Documentos de Investigación, México, 2004.

Martínez, Ifigenia, *Distribución del ingreso en México: Tendencias y proyección a 1980*, vol. 1, Siglo XXI, México, 1970.

Piketty, Thomas, Capital in the Twenty-First Century, Fondo de Cultura Economica, Mexico, 2015.

Stiglitz, Joseph E., The Price of Inequality: How Today's Divided Society Endangers Our Future, W. W. Norton & Co. Inc., New York, 2013.

----, The Great Divide: Unequal Societies and What We Can Do About Them, W. W. Norton & Co. Inc., New York, 2015.

----, Rewriting the Rules of the American Economy: An Agenda for Growth and Shared Prosperity, W. W. Norton & Co. Inc., New York, 2015.

Wilkinson, Richard, and Kate Pickett, The Spirit Level: Why Equality is Better for Everyone, Tantor Media Inc., Connecticut, 2011.