Research questions
The World Income Distribution

Michel Lubrano

2019–2020
Master in economics (2nd year)

Contents

1 Introduction

With the rise of international agencies starting during the second half of the XXth century, more concern was expressed for world poverty and world inequality. We can point out two major landmarks. One concerns a United Nation program and the second a book published by François Bourguignon, a former Chief Economist at the World Bank.

1. The United Nations issued the project *Objectives for the Millennium*: the aim was to cut extreme poverty by half for 2015. We can find details on the web at:

   $$http://www.un.org/millenniumgoals$$

2. Bourguignon (2016), *The Globalization of Inequality* and Milanovic (2016) *Global Inequality*. Globalization has induced a drop in the between countries inequality while inequality has increased within countries. This topic is related to the decomposition of an inequality index. If we take the case of the Generalized Entropy index I^{GE}, it can be decomposed between k subgroups, using the formula:

   $$I^{GE} = \sum_i w_i I_i(x, n_i) + \sum_i B_i(\mu_i, \mu, w_i),$$

   where $I_i(x, n_i)$ represents inequality within group i and $\sum_i B_i(\mu_i, \mu, w_i)$ represents inequality between the k groups. The w_i are the weights to be used in the decomposition.

These two books are sufficiently important to have motivated a paper by Ravallion (2018) published in the *Journal of Economic Literature*. The purpose of Ravallion (2018) was to appraise the role of globalization for explaining world income inequality.

When discussing income inequality and also poverty, a key concept is the income distribution. At the level of a country, this is relatively easy to define, if not to measure. For discussing poverty and inequality at the world level, we should need a world income distribution. It is not easy to define an income distribution at the World level and a lot of specific questions are involved. They will be the main topics of this lecture.

1.1 The importance of the topic

The World Income Distribution (WID) is an important topic for a variety of arguments.

We can first think of problem around **distributive justice** as *Large disparities can be considered as unjust.* The notions of justice at the world level are not evident. In a way Rawls (1971) is avoiding the question because there are no institution able to enforce a social contract at the world level. Brandolini and Carta (2016) contribute to the debate.

There are however international institutions where these questions can be discussed. They are concerned by inequality and bargaining power. Emerging countries are willing to insist on their economic weight in order to gain more bargaining power and a WID is a tool for that purpose.

Migrations are becoming a problem as they can be determined by unequal economic conditions. It becomes important to localize poverty over the world. In many papers that we shall examine, there is a section about the geography of poverty.

From an academic point of view, neoclassical growth theory is concerned about **convergence** between countries. It becomes important to be able to measure this convergence, or its absence.

The last question is **Dependency theory and divergence**, see for instance Stiglitz (2017) on the Euro problem and European convergence.

1.2 What is world inequality

World inequality can be defined and measured in **three** different way, if we follow Milanovic (2005) and Lakner and Milanovic (2013), also quoted in Anand and Segal (2008). However Anand and Segal (2015) introduce a fourth level, they call it concept 0. It is interesting to try to relate these concepts to the notion of global justice.

Concept 0 Observed at the country level, the total national income of that country, measured in dollars or PPP. Useful to measure the weight of a country for international trade. Canada and India has the same GDP in 2012 (but not the same population!).

Concept 1 Inter-country inequality: each country is treated as a point equipped with GDP per capita. This is useful to study economic convergence between countries. Measured at PPP. But there is no possibility to measure within country inequality. So we skip one part of the question. However, this type of inequality is important for leading discussions inside international organizations.

Concept 2 Population-weighted intercountry inequality, the same as before but with population weights. This is inequality among individuals in the world with each individual assigned the average per capita income of his or her country of residence, measured with PPP. If we keep the level of international institutions discussion, this would deny the right to small countries of having a voice in these institutions. This would be a lack of procedural justice.

Concept3 Global inequality between world citizen. The key income concept is the household income, so each country income distribution is needed. In term of redistributive justice, this is certainly a valid point of view. Which also allows for studying between countries inequality if the information concerning borders is not lost. Certainly the most complex notion of inequality.

The WID is concerned with the last definition where the world is considered as a unique country. The notion of redistributive justice is easily discussed at this level, this is the cosmopolitan view.

1.3 Tools

We need tools for measuring world poverty and inequality:

1. An international poverty line: Xun and Lubrano (2018), following several papers of the World Bank around Chen and Ravallion, but also the very important approach of Atkinson and Bourguignon (2001) and Deaton (2010). What is an international poverty line and how should it be defined. We cannot have the same line for all countries. An absolute IPL is valid for the poorest countries. But after a certain level, it has to become a relative poverty line. We cannot measure poverty in the US using the IPL of one dollar a day.

2. Data and a method to estimate the world income distribution. The question of data is of prime importance. But also what to do with the data, in other words which kind of distribution can we adjust to the data.

3. The use of dynamic quantiles: Ravallion and Chen (2003) in order to obtain the famous *Elephant Curve*. The aim is to compare the evolution between two income distributions. The tool of stochastic dominance gives a global answer. The GIC is related to stochastic dominance but it also provides a picture of how the income distribution was modified.

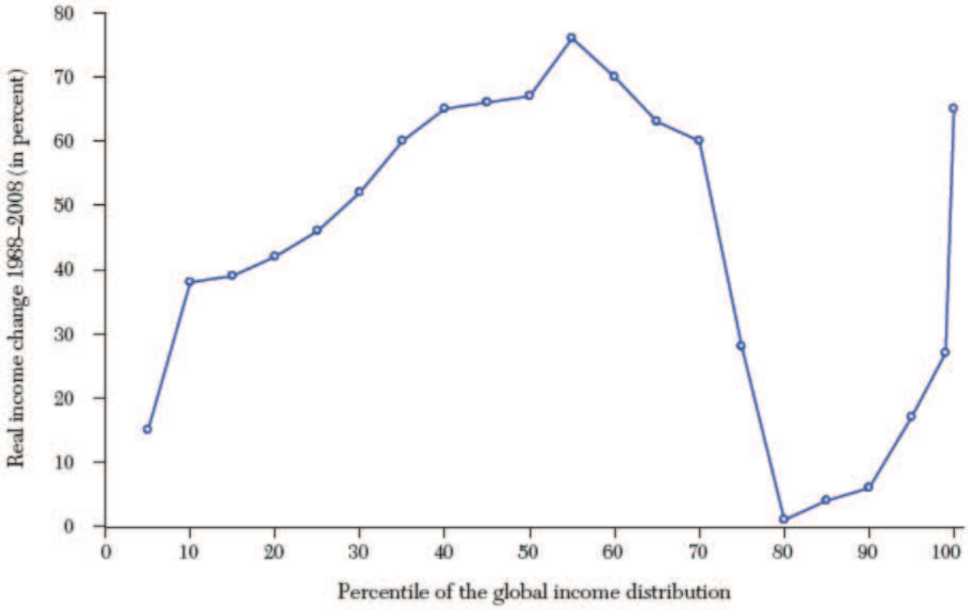

Figure 1: The elephant graph of Lakner and Milanovic

1.4 Main references in the Literature

There is a vast literature around the World Income Distribution. Basis for this introduction to the topic. These references can be ordered according to the topic they treat.

- The entrance to the topic can be a survey published in JEL, Anand and Segal (2008): What Do We Know about Global Income Inequality? Due to measurement uncertainty, it is difficult to conclude about the influence of globalization on global inequality, despite the numerous published papers. All papers indicate a high level of inequality, but they do not agree on the direction of change.

  A chapter in a handbook Anand and Segal (2015): The Global Distribution of Income. This chapter covers in fact the main topics of this lecture. It is an upgrade of the previous paper, introducing for instance the question of top incomes.

- An important, but controversial paper: Sala-i-Martin (2006). It is concerned by measuring both poverty and inequality using non-parametric density inference applied to very sparse data sets based on quantiles.

- Parametric approach: Chotikapanich et al. (2012) WID and inequality (mixture of Beta 2), Pinkovskiy and Sala-i-Martin (2009) (mixture of lognormals) with the same pitfalls as Sala-i-Martin (2006).

- Using quantiles with a uniform assumption: Bourguignon and Morrisson (2002), Milanovic (2002). In fact this is not uniform but the identical income assumption. All the households in the same quantile have the same income. So there is no inequality inside a quantile, which is a simplifying assumption that lead to an under-evaluation of inequality.

- Darvas (2016) compares three methods: non-parametric, direct quantiles and parametric.

- Correction for top incomes: Imputation methods, Jenkins et al. (2011); Mixing surveys and tax data, Atkinson et al. (2011).

- The elephant curve: Lakner and Milanovic (2016), Alvaredo et al. (2018)

1.5 General research questions

- The determination of an international poverty line, using the fixed point method of Deaton (2010). See also Xun and Lubrano (2018).

- The geography of poverty

- The questions around data

  - Survey data, National accounts, PPP and money conversions
  - The measurement of top incomes, tax files or statistical imputation?
  - The mixing of macro and micro data, divergence and price indices

- World income dynamics and the elephant curve: what is explaining this shape. Totaly unexpected at a national level. Ravallion (2018) minimizes the role of globalization.

- Questions around measurement of growth dynamics: anonymous versus non-anonymous GIC. Which parametric form?

2 Mixing macro and micro data

2.1 The literature

- Atkinson and Bourguignon (2001): mixture of lognormal densities, population weights, the mean comes from NA and the use of an hypothetical Gini, PPP correction for GDP per capita:

  $$\mathrm{E}(y) = \exp(\mu + \sigma^2/2), \qquad Gini = 2\Phi(\sigma/\sqrt{2}) - 1$$

  and then

  $$f(x) = \sum_i w_i f_\Lambda(x_i|\mu_i, \sigma_i^2)$$

- Using reported Gini and auxiliary information from PovCalnet: Xun and Lubrano (2018). 2005 PPP.

- Sala-i-Martin (2006) mixing macro and micro and using a non-parametric approach to estimate the WID, based on quantiles coming from the Deninger and Squire data base.

2.2 An attempt for developping countries

With this type of assumption, Xun and Lubrano (2018) estimated a WID based on a lognormal, GDP per capita and Gini index coming from the World Bank. The WID is reported in Figure 2. The resulting density is dominated by the shape of the lognormal distribution, despite the fact that it is a mixture of 74 different countries. But the weights of India and China are dominant. And very rich countries are excluded from this sample which contains mainly developing countries because the main objective of the paper was to derive an updated International Poverty Line (IPL).

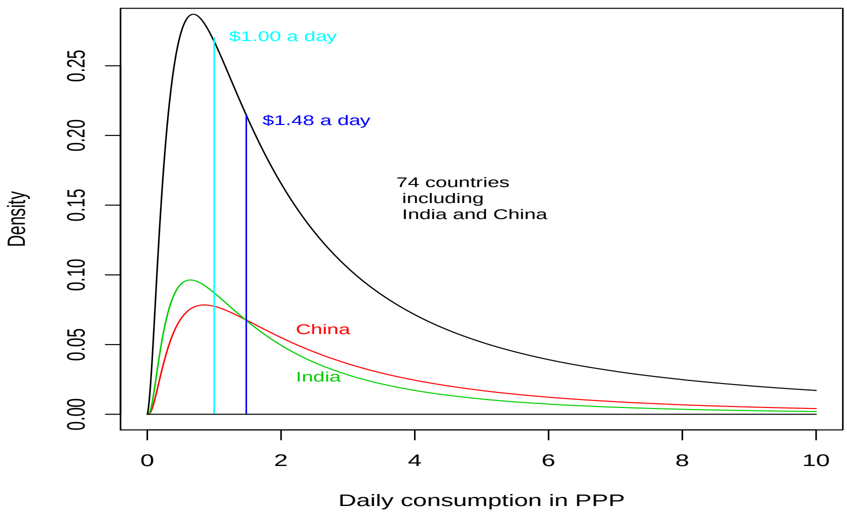

Figure 2: World Income Distribution around 2001

2.3 The geography of poverty

The geography of poverty aims at the locating the poor over the world. The World Bank has divided the world into six homogeneous regions, with for instance Middle-East-North-Africa (MENA). This geography depends on the definition of an IPL and of a WID. Table 1, drawn from Xun and Lubrano (2018), investigates the influence of various weighting assumptions on the value of the IPL and its consequences in term of million of poor. The first

Table 1: Poverty count in the developing world around 2001 (millions)

Group	Reference	World	China	India
Poverty line	IPL	$\max(IPL, z_i)$	IPL	IPL
Unweighted	1 448	1 698	409	498
Pop weighted	1 505	1 846	459	547
Poor weighted	1 584	1 833	455	543
Official figures	Reference	World	China	India
	1 195	1 599	360	416

Official figures were computed using the official poverty rate at the national poverty line. No figures exist for 43 countries in the World Bank data set. So we determined which of the normalized poverty lines of the World Bank ($1.25, $2.00, $2.50, $4.00 and $5.00) was closest to the national poverty line and took the corresponding poverty rates.

column corresponds to the case where the IPL is defined by reference to a low-income group of countries. Depending on the weighting, the IPL is equal to $1.48, $1.65 or $1.63, using 2005 PPP. The second column defines a variable IPL depending on the country to which it is applied. It draws on the ideas of Atkinson and Bourguignon (2001). The last two columns apply the IPL to China and India separately at a time where these two countries were still considered as poor countries.

Table 2 details the location of the poor for different poverty line calculations. The last column gives the results published in Sala-i-Martin (2006) for comparison. Most of the poor are located in East Asia (China) and South Asia (India). These figures are much higher than in Sala-i-Martin (2006), who finds around 400 million of poor people in 2000 using a common poverty line of $1.50. Using our IPL, we find 1 698 million. Where does it come from?

The figures for Africa found in Xun and Lubrano (2018) are similar to those of Sala-i-Martin (2006). Where we find hugely different figures is for East and South Asia, essentially China, India and Indonesia. It is worth comparing the data themselves. We collected the official poverty rates on the web site of the World Bank. When we multiply the official poverty

Table 2: The location of poor people in the developing world around 2001

Region	Unweighted	Pop weighted	Poor weighted	Sala-i-Martin
Africa	245	263	262	297
South Asia	639	702	697	33
East Asia	576	639	634	41
East. Europe	36	36	36	4
Latin America	177	177	177	21
MENA	26	29	28	1
Total	1,698	1,840	1,834	397

Figures are in millions. The last column comes from Table 2 of Sala-i-Martin (2006) based on a $1.50 a day poverty line.

rates by the population and sum up the countries, we find a total of 1 599 million poor people. Our unweighted evaluation of 1 698 million appears to be consistent with the information contained in our data base. This raises question about the way Sala-i-Martin (2006) estimated his WID.

2.4 What is wrong with Sala-i-Martin (2006)?

Sala-i-Martin (2006) is using GDP per capita for scaling. He makes use of the Deininger and Squire data set.

1. There are problems in this data set, underlined in Atkinson and Brandolini (2001): some incomes are for individuals, some for households. Changes in definitions over time.

2. Data are available in the form of quintiles. 5 observations to adjust a density using Kernel Density Estimation. Usual asymptotic theory cannot apply.

3. Use of the same window size for all years and countries. So the differences in inequality are ignored as the window is defined as $h = 1.06 \times \hat{\sigma} \times n^{-1/5}$.

4. Because of a constant window size, the spread of the distribution is modified, and so inequality is badly measured.

5. A Gaussian kernel is not adapted for positive random variables. The left tail of the distribution is thus badly estimated and no poverty count can be validly inferred.

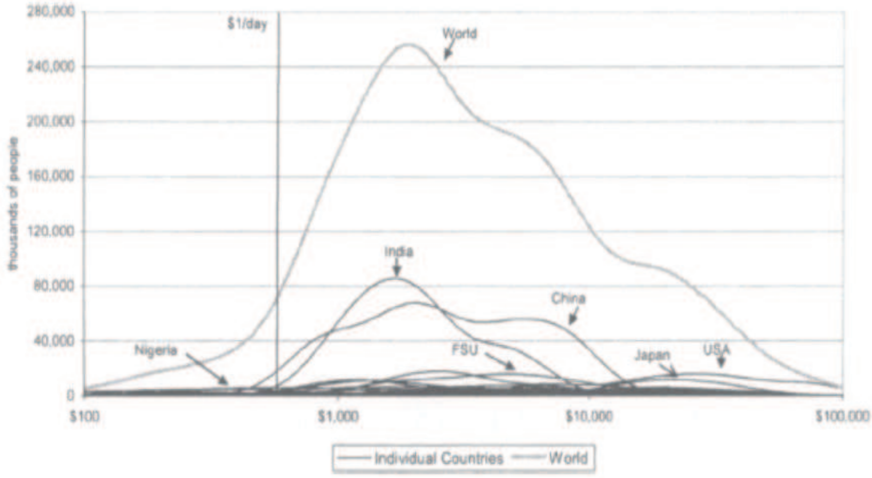

FIGURE IIIb
The WDI and Individual Country Distributions in 2000

Figure 3: Estimated WID from Sala-i-Martin

Figure 3 is reproduced from Sala-i-Martin (2006). Compared to Figure 2, rich countries are added. But we can already note that the income distribution of China is very strange as it presents several modes. For instance, we could show the Chinese income distribution in 2006 as estimated from survey data using the CGSS in Figure 4. This Figure was obtained by kernel smoothing with an asymmetric lognormal kernel and a sample determined window size. They have nothing in common and the Chinese income distribution cannot have changed so much in six years. Let us comment about the influence of the bandwidth.

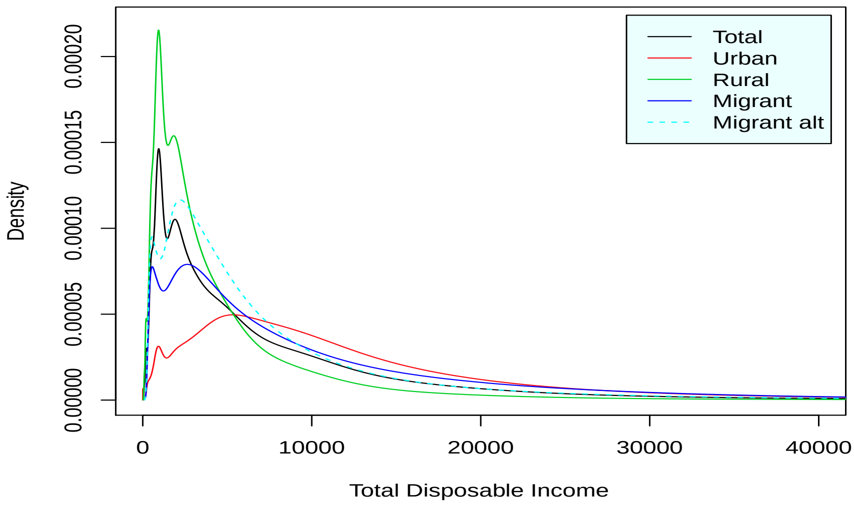

Figure 4: Chinese Income distribution per capita in 2006

We have generated a sample according to a lognormal distribution so as to reproduce the main characteristics of the WID appearing in Figure 2. The

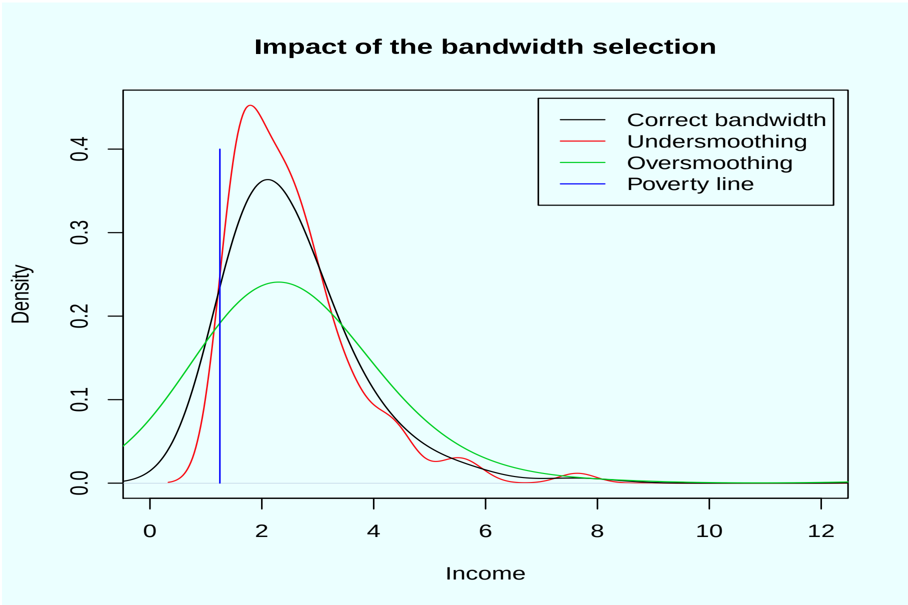

Figure 5: The influence of the bandwidth

bandwidth has no influence on the mean (with a symmetric kernel), but a serious influence on the variance and the shape of the tails. So the proportion of poor is going to be greatly affected.

3 The need of a correct econometric approach

We have seen that there was something wrong in the paper of Sala-i-Martin (2006) and that this was probably due to the econometric method, at least partially. Let us now review the other possible econometric models that can be applied.

3.1 A distribution free approach

The pioneering papers of Milanovic (2002) and Bourguignon and Morrisson (2002) were relying on a common econometric assumption. The type of data that are available are income shares sw_i from household surveys. These papers assume that inside each quantile, all households have the same income. This is the *identical quantile income* method. This is a simulation methods which generates income values according to the population size of each country. It works as follows. For each country, using a `R` code convention:

1. $sw_i = p_i \times \mu_i/\mu$. If deciles, $p_i = 0.10$. Population size $= N$.
2. $\mu_i = (sw_i/p_i)\times$ mean income in PPP
3. For each decile i, we take $x_i = \texttt{rep}(\mu_i, N/10)$, all people within each decile have the same income.
4. The distribution is build as $x = \texttt{c}(x_1, x_2, \cdots, x_{10})$
5. Lorenz and dominance curves are built in the same way. Derivation of the Gini.

For the World income distribution, all these artificial samples are pooled. The global incomes shares and Lorenz curves are computed from this sample.

3.2 A full parametric approach

Chotikapanich et al. (2012): modelling country-specific income distributions using Beta 2 and incorporating them into the global distribution. Relax the constant-income-within-subgroups assumption made by Sala-i-Martin (2006), Bourguignon and Morrisson (2002), Milanovic (2002). Same data set as Sala-i-Martin (2006), but with totaly different econometric methods. Investigate global inequality and globalization.

- Population and income shares or as population shares and class mean incomes in local currency units.
- Country mean incomes, in common currency units (Penn World Tables), to scale the local currency data and compute the required class mean incomes.
- Deininger and Squire updated database from the World Bank (WIID2b)
- Country real per capita (mean) income drawn from national income sources

In Figure 6, we have the income distribution of two selections of countries, using a Beta 2 parametric density. Of course, they are very regular. And the Chinese and Indian densities look very much the same as in our Figure 2: more inequality in China, but a larger mean income, compared to India.

Figure 7 collects the different country income distributions and assemble them into a mixture to find the WID and the distribution for five groups

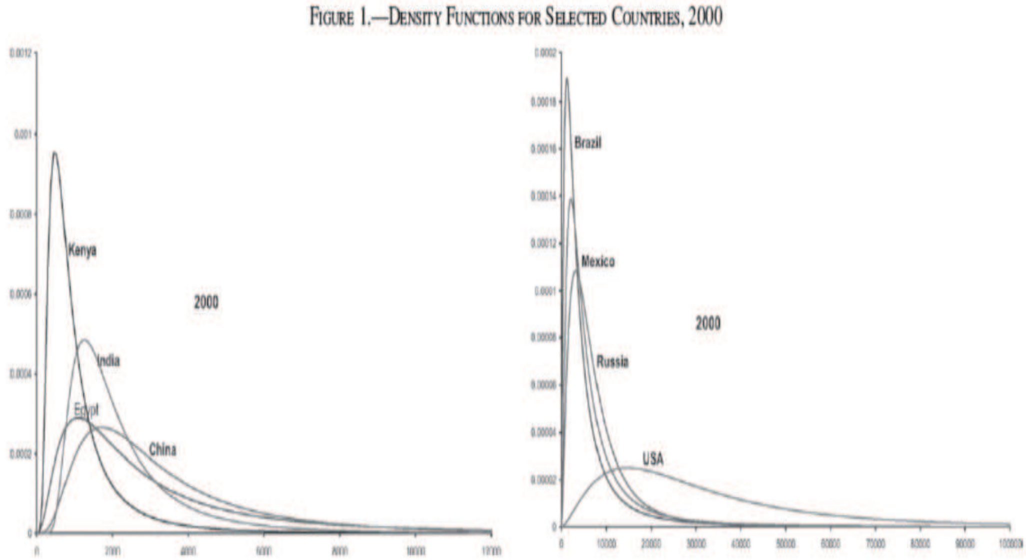

Figure 6: Selected countries income distribution

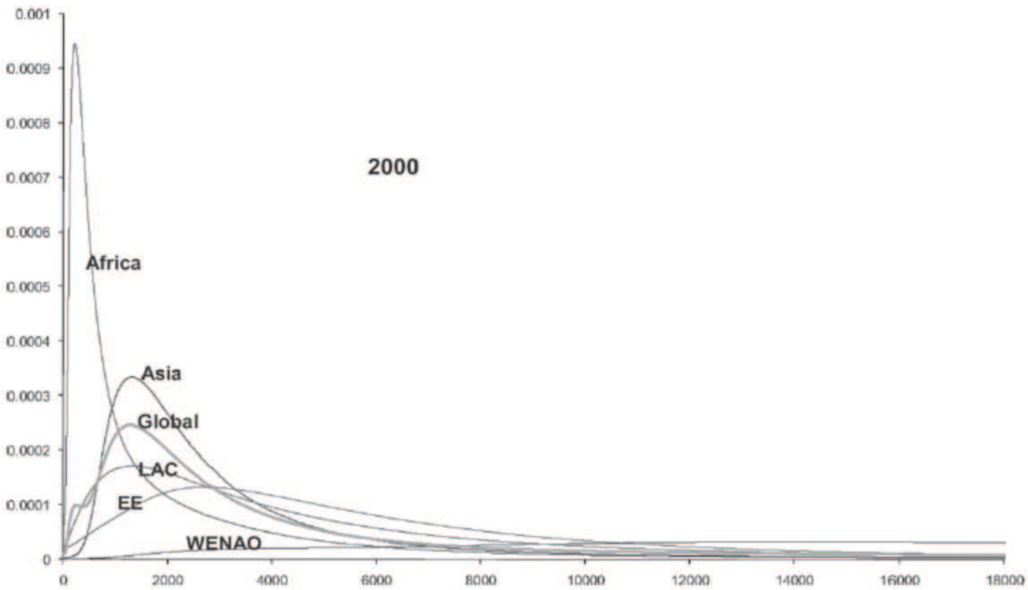

Figure 7: WID in 2000 and its group decomposition

of countries. Because the Theil index is decomposable, Chotikapanich et al. (2012) present the evolution of world income inequality between 1993 and 2000 and its decomposition in within country inequality and between country inequality. Here, we have a decrease in World total inequality, but and an

Table 3: Global Inequality Decomposition

Theil	1993	2000
Global	0.813	0.795
Within	0.288	0.302
Between	0.525	0.493

increase of inequality within each country, while it has decreased between the countries. This is coherent with the idea reported in Bourguignon (2012, 2016).

Note that Milanovic (2002) concluded to an increase of global inequality between 1988 and 1993. We have here a decrease after 1993. Sala-i-Martin (2006) concluded also to a decrease of global inequality after 1993, but fluctuations in the seventies.

3.3 Comparing the methods

Darvas (2016) compares different methods used to adjust a WID. The performance is judged on the ability to reproduce Gini indices when starting from quantiles data.

- The identical quantile income method gives satisfactory results
- The NP method of Sala-i-Martin (2006) is the worse method
- The best method consist in adjusting a parametric distribution
- The regression method of Kakwani (1980) gives good results

  $$\log[p - L(p)] = \beta_0 + \beta_1 \log(p) + \beta_2 \log(1-p)$$

  Adjusted on quantiles which in fact correspond to a Lorenz curve. Prediction using:

  $$L(p) = p - \exp[\beta_0 + \beta_1 \log(p) + \beta_2 \log(1-p)]$$

  Using integral calculus, the Gini is found to be:

$$G = \exp(\beta_0)\frac{\Gamma(\beta_2+1)\Gamma(\beta_1+1)}{\Gamma(\beta_1+\beta_2+2)}.$$

So inequality is measured directly from the shape of the Lorenz curve, estimated parimetrically.

We can try to go a bit further on with this method. The quantile function corresponds to $\mu L'(p) = x$ and has a long expression:

$$
\begin{aligned}
Q(p) &= \mu \frac{p^2 - p}{p(p-1)} - \\
&\quad \mu \exp(\beta_0) \frac{p^{\beta_1+1}(1-p)^{\beta_2}\beta_1 + p^{\beta_1+1}(1-p)^{\beta_2}\beta_2 - p_1^{\beta}(1-p)_2^{\beta}\beta_1}{p(p-1)}
\end{aligned}
$$

For finding the density function, one has to first invert $x = Q(p)$ and then differentiate the result in x.

4 The bias in micro and macro data

It is easy to understand that the main question for estimating a WID is data. The sources are diverse and hardly comparable. And for some countries, they are very scarce. The main questions are the relation between national accounts (NA) and survey data, the use of PPP to convert currency units and the under-reporting of high incomes.

4.1 National accounts and surveys

Anand and Segal (2008, 2015) discuss the concepts of income involved in NA compared to survey data and the problems induced by PPP conversion. Nolan et al. (2016) try to find the reason why GDP per capita can evolve over time differently from household income. Cynamon and Fazzari (2017) try to analyse the differences between the income concept from NA and household income in surveys.

- Household consumption per capita from NA exceeds average consumption from surveys. Deaton (2005) discrepancy has increased over time.
- National Accounts: GDP per capita, C+I+X-M
- National accounts: household final consumption expenditure (HFCE)
- Survey: disposable income. Under-reporting of very poor and very rich. Imputed rents not included
- Anchoring survey means on GDP per capita would mean a proportional under-evaluation
- NA are not more reliable than surveys, depending on the countries.

Anand and Segal (2008) document the divergence between these three possible scalings of an income distribution. HFCE is 72% of GDP in the US in 2006 and disposable income 90% of HFCE. NA estimates of HFCE are subject to error as they come from a difference. Under-reporting for surveys.

4.2 PPP conversion

If the law of one price held and there were no non-tradables, we could simply use market exchange rates. Use of PPP exchange rates to account for differences in the cost of living across countries.

- Using exchange rate would under-estimate the income of poorer countries and increase inequality.
- International Comparison Program for collecting CPIs. China included in 2005.
- Deaton (2010) underlines that the regular revision of PPP changes the geography of poverty. Consequences of 2005PPP to 2011PPP.

For a quantity expressed in 2005PPP denoted $x_{2005PPP}$, a conversion would mean

$$x_{2011PPP} = x_{2005PPP} \times \frac{CPI_{2011}}{CPI_{2005}} \times \frac{PPP_{2005}}{PPP2011}$$

4.3 Interpreting discrepancy NA Surveys

Lakner and Milanovic (2016) global interpersonal inequality 1988 and 2008. Their data set is a panel of country-deciles, without GDP anchoring. 2005PPP. PovcalNet, WYD, LIS, BHPS, SILC, 565 surveys 1988, 1993, 1998, 2003, 2008.

- Lakner and Milanovic (2016) The under-reporting of top incomes in household surveys and their discrepancy with national accounts are closely connected issues.
- Surveys give a good account up to the ninth decile.
- Pareto imputation to the last decile of the excess consumption recorded in NA. The mean is increased.
- To take into account imputed rents, consumption of public goods

5 Correction methods

Because data set are not enough informative for high incomes, some correction methods were proposed. A first type of method can be qualified as statistical methods. In the case of top coding for instance, method simulates missing observations. This is for instance multiple imputation. A second type of method proposes to combine survey data which are informative for income up to say the 90% quantile or the 99% quantile with other sources of data, typically tax data. The question is of course to have a convincing method of combination.

5.1 Multiple imputation for survey data

One main reference is Jenkins et al. (2011). This paper addresses the question of top coding. In official surveys, like the US March Current Population Survey (CPS), but also many other sources, high income or high wages are top coded which means that the only indication is that the related wage or income is greater then a given bound for confidentiality reasons. The correcting method works as follows:

1. Estimate a truncated GB2 over the sample up to the 9th decile
2. Generate random numbers in the top tail of the GB2
3. Estimate Gini using the initial observations plus the random draws from the tail

There is a vast literature on multiple imputation, mainly in medical statistics. Jenkins et al. (2011) is one example applied to income data. The purpose of Atkinson (2017) is slightly different. He is concerned with investigating the importance of high incomes, using tax data and the Pareto coefficient. He investigates historical UK data. A good introduction to the next subsection.

5.2 Tax records and top incomes

The use of tax record for completing survey data is becoming more and more popular. Atkinson (2017) is investigating tax data with the Pareto distribution in order to study the evolution of high incomes. Jenkins (2017) analyses inequality still in the UK, but he is interested in the whole income distribution. So he want to combine survey data and tax data. There are two ways of doing this. Jenkins (2017) fits a Pareto II to the tax data and derives a Gini coefficient for the rich group. Then he applies a decomposition formula for the Gini index due to Atkinson (2007) to combine the Gini computed over the survey data and the Gini derived from the Pareto estimation. The formula for the Gini decomposition is:

$$G = p_r s_r G_r + p_n s_n G_n + G_B.$$

The indices r and n refer to rich and non-rich. p_r is the proportion of rich, s_r the income share of the rich, computed as $s_r = p_r \mu_r / \mu$. The overall mean income is $\mu = p_r \mu_r + p_n \mu_n$. The last term is the between group inequality $G_B = s_r - p_r$. p_r has to be fixed by the statistician, with values ranging from 0.90 to 0.99.

The other way of combining tax and survey data does not rely on a parametric model for high incomes. The highest incomes in the survey data are replaced by with cell-mean imputations based on the corresponding observations in the tax return data. The method can be illustrated in Atkinson et al. (2011), Blanchet et al. (2017). Here again one has to fix which proportion of the population the tax data represents for top incomes. Quantiles are then reevaluated using an interpolation method.

The account which is given in Anand and Segal (2015) is interesting, because it shed light on the method and its limitations. "*These estimates present the incomes of the top 0.1%, top 1%, and top 10% as a share of control income, where control income is an estimate of total personal income in the economy (not just taxable income)*". There are two important limitations. Taxes can change over time, and there is also tax avoidance and tax evasion. And the measured inequality can vary a lot with the definition of the reference income which is used to compute the top income shares. So, even if these data help a lot for measuring inequality (they correct the under-evaluation resulting from the use of survey data), they also introduce a lot of variation and add uncertainty to the final estimates.

Anand and Segal (2015) have combined tax data coming from the WID data base to the data of Milanovic, assuming that the Milanovic data set covers only the bottom 99% of the population in each country. They multiply the population in each income group in the surveys by 0.99 and then append the top percentile with its income share from the tax data. The countries for which this operation is possible are: China, India, Indonesia, Argentina, South Africa and the G7 countries. They use a regression method for the other countries.

6 Growth Incidence Curve

6.1 The algebra of the growth incidence curve

Ravallion and Chen (2003): measuring pro-poor growth. Trickle down is not necessarily true.

Quantile functions $Q(p)$ are useful for defining the Lorenz curve

$$L(p) = \frac{1}{\mu}\int_0^p Q(t)\,dt, \Rightarrow Q(p) = \mu L'(p).$$

When we have two points in time, we can draw the Growth Incidence Curve:

$$g_t(p) = \frac{Q_t(p)}{Q_{t-1}(p)} - 1 = \frac{L'_t(p)}{L'_{t-1}(p)}(\gamma_t + 1) - 1 \simeq \log Q_t(p) - \log Q_{t-1}(p).$$

Can be related to the Watts poverty index as $W = \int_0^q (\log z - \log Q(t))dt$.

6.2 Comparing elephant curves

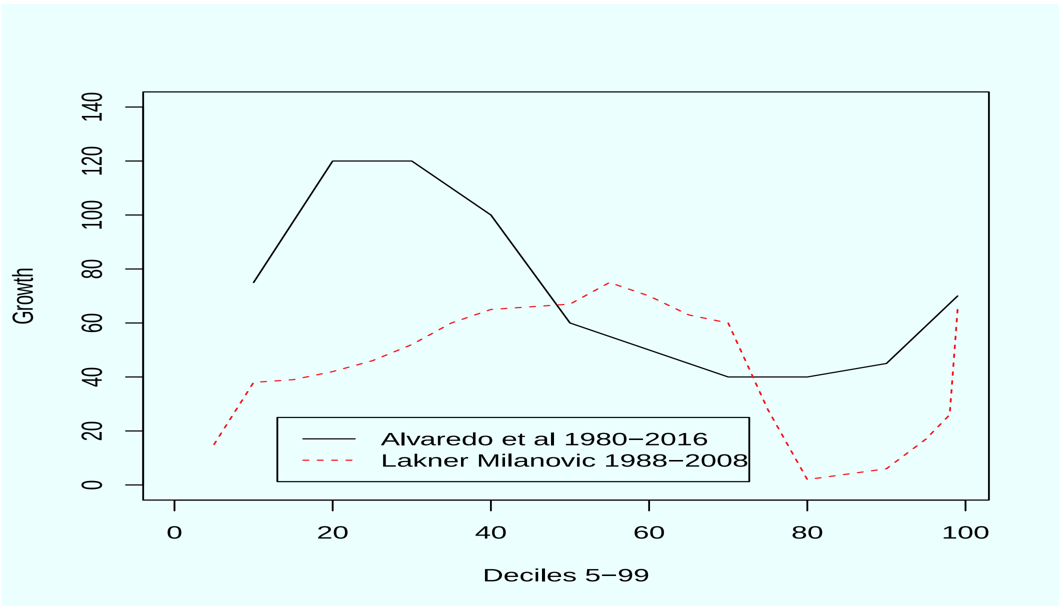

Figure 8: Two evaluations of the Elephant curve

With Figure 8 we compare the two published elephant curves that are coming from Lakner and Milanovic (2016) and from Alvaredo et al. (2017, 2018). The general features of these two curves are similar. But when reported on the same graph, there are strong differences. For instance in Alvaredo et al. (2017, 2018), the poor have gained much more than in Lakner and Milanovic (2016).

It is interesting to try to explain the origin of these differences. In Lakner and Milanovic (2016) the data report income per capita when in Alvaredo et al. (2017, 2018), income is per adult. This makes a large difference for families with a large number of children. Both approaches try to correct for high incomes. But they use two different imputation methods. Finally, the two data sets are different. Alvaredo et al. (2017, 2018) have a much greater concern for high incomes. When they speak about poor people they refer to those below the median. A closer look at they data sets shows that they data are not representative of poor people and are unsuitable for analysing poverty.

7 Data sources

Various data basis are available for deriving a World Income Distribution. Essentially we need income data, most of the time from surveys, but we have seen also that National Accounts played a significant role.

The first source, but which is partial, is formed by the main national surveys. They are available at the national level for large countries like the PSID for the US, the BHPS for the UK, the GSOEP for Germany. There is the EU-SILC project at the level of the European Union which started to be implemented in 2003. All the figures are converted in euros. But comparisons outside Europe are difficult to make, because of different concepts and different monetary units.

A large effort has been made for providing compatible data at the world level. Essentially, we have three sources: the World Bank, the project around Branko Milanovic and the project around Thomas Piketty.

7.1 The World Bank

This the tool `Povcalnet` which is available at the World Bank:

$$http://iresearch.worldbank.org/PovcalNetPPP2005/index.htm$$

We should note that the same division is in charge of defining PPP conversions so that countries are made comparable.

7.2 The World Inequality Data Base

This is a project gathering a team of more than 100 researchers. It was initiated by Tony Atkinson and Thomas Piketty. The data are available at:

$$https://wid.world$$

together with R programs to access them. However, not all the countries are available. Global macro data are available for a large list of countries (39 in Europe, 56 in Africa, but only 2 for Americas and 4 for Asia). When we look for more detailed data sets, is available a much more reduced list of countries.

If the data are very good for high income, they are not feasible for lower incomes and totally unadapted for analysing poverty.

7.3 The Stone Center on Socio-Economic Inequality

The *Stone Center on Socio-Economic Inequality* is a US research center hosted by the City University of New York, the working place of Branko Milanovic. The data base is available at

$$https://www.gc.cuny.edu$$

and is maintained by Branko Milanovic. It is available for a large number of countries, summarizing surveys, but providing of course only quantiles. Data are running over 1988, 1993, 1998, 2002, and 2005 with over 100 different countries. However, there are only 67 countries for which we have data for the five years.

8 Conclusion

An important literature centered on two main points:

1. The impact of globalization on poverty and inequality

2. How to correct surveys for high incomes

A side, but important methodological question: the choice of a correct econometric method for treating those data in order to avoid the pitfalls of Sala-i-Martin (2006).

References

Alvaredo, F., Chancel, L., Piketty, T., Saez, E., and Zucman, G. (2017). The elephant curve of global inequality and growth. Technical Report 2017/20, World Wealth and Income Databse.

Alvaredo, F., Chancel, L., Piketty, T., Saez, E., and Zucman, G. (2018). The elephant curve of global inequality and growth. *AEA Papers and Proceedings*, 108:103–08.

Anand, S. and Segal, P. (2008). What do we know about global income inequality? *Journal of Economic Literature*, 46(1):57–94.

Anand, S. and Segal, P. (2015). Chapter 11 - the global distribution of income. In Atkinson, A. B. and Bourguignon, F., editors, *Handbook of Income Distribution*, volume 2 of *Handbook of Income Distribution*, pages 937 – 979. Elsevier.

Atkinson, A. B. (2007). Measuring top incomes: Methodological issues. In Atkinson, A. and Piketty, T., editors, *Top Incomes over the 20th Century*, pages 18–42. Oxford University Press.

Atkinson, A. B. (2017). Pareto and the upper tail of the income distribution in the uk: 1799 to the present. *Economica*, 84(334):129–156.

Atkinson, A. B. and Bourguignon, F. (2001). Poverty and inclusion from a world perspective. In Stiglitz, J. E. and Muet, P.-A., editors, *Governance, Equity and Global Markets*. Oxford University Press.

Atkinson, A. B. and Brandolini, A. (2001). Promise and pitfalls in the use of 'secondary' data-sets: Income inequality in OECD countries as a case study. *Journal of Economic Literature*, 39(3):771–799.

Atkinson, A. B., Piketty, T., and Saez, E. (2011). Top incomes in the long run of history. *Journal of Economic Literature*, 49(1):3–71.

Blanchet, T., Fournier, J., and Piketty, T. (2017). Generalized Pareto curves: Theory and applications. Working Paper Series 3, World Wealth and Income Databse.

Bourguignon, F. (2012). *La mondialisation de l'inégalité*. Seuil, Paris.

Bourguignon, F. (2016). *The Globalisation of Inequality*. Princeton University Press.

Bourguignon, F. and Morrisson, C. (2002). Inequality among world citizens: 1820-1992. *The American Economic Review*, 92(4):727–744.

Brandolini, A. and Carta, F. (2016). Some reflections on the social welfare bases of the measurement of global income inequality. *Journal of Globalization and Development*, 7(1):1–15.

Chotikapanich, D., Griffiths, W. E., Rao, D. S. P., and Valencia, V. (2012). Global income distributions and inequality, 1993 and 2000: Incorporating country-level inequality modeled with beta distributions. *The Review of Economics and Statistics*, 94(1):52–73.

Cynamon, B. Z. and Fazzari, S. M. (2017). Household income, demand, and saving: Deriving macro data with micro data concepts. *Review of Income and Wealth*, 63(1):53–69.

Darvas, Z. (2016). Some are more equal than others: new estimates of global and regional inequality. Discussion Paper 8, Bruegel, Brussels, Belgium.

Deaton, A. (2005). Measuring poverty in a growing world (or measuring growth in a poor world). *The Review of Economics and Statistics*, 87(1):1–19.

Deaton, A. (2010). Price indexes, inequality, and the measurement of world poverty. *American Economic Review*, 100(1):5–34.

Jenkins, S. P. (2017). Pareto models, top incomes and recent trends in uk income inequality. *Economica*, 84(334):261–289.

Jenkins, S. P., Burkhauser, R. V., Feng, S., and Larrimore, J. (2011). Measuring inequality using censored data: a multiple-imputation approach to estimation and inference. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 174(1):63–81.

Lakner, C. and Milanovic, B. (2013). Global income distribution: From the fall of the Berlinwall to the great recession. Policy Research Working Paper 6719, World Bank.

Lakner, C. and Milanovic, B. (2016). Global income distribution: From the fall of the Berlinwall to the great recession. *World Bank Economic Review*, 30(2):203–232.

Milanovic, B. (2002). True world income distribution, 1988 and 1993: First calculation based on household surveys alone. *The Economic Journal*, 112(476):51–92.

Milanovic, B. (2005). *Worlds Apart: Measuring International and Global Inequality*. Princeton University Press.

Milanovic, B. (2016). *Global Inequality: A New Approach for the Age of Globalization*. Harvard University Press, Cambridge.

Nolan, B., Roser, M., and Thewissen, S. (2016). Gdp per capita versus median household income: What gives rise to divergence over time? Discussion Paper 3, Institute for New Economic Thinking, Oxford University.

Pinkovskiy, M. and Sala-i-Martin, X. (2009). Parametric estimations of the world distribution of income. Working Paper 15433, National Bureau of Economic Research, 1050 Massachusetts Avenue, Cambridge, MA 02138.

Ravallion, M. (2018). Inequality and globalization: A review essay. *Journal of Economic Literature*, 56(2):620–642.

Ravallion, M. and Chen, S. (2003). Measuring pro-poor growth. *Economics Letters*, 78:93–99.

Rawls, J. (1971). *A Theory of Justice*. Oxford University Press, London.

Sala-i-Martin, X. (2006). The world distribution of income: Falling poverty and... convergence, period. *Quarterly Journal of Economics*, 121:351–397.

Stiglitz, J. (2017). The fundamental flaws in the euro zone framework. In da Costa Cabral, N., Gonçalves, J., and Cunha Rodrigues, N., editors, *The Euro and the Crisis. Financial and Monetary Policy Studies*, volume 43. Springer, Cham.

Xun, Z. and Lubrano, M. (2018). A Bayesian measure of poverty in the developing world. *Review of Income and Wealth*, 64(3):649–678.