J Econ Inequal (2015) 13:557–578
DOI 10.1007/s10888-015-9304-4

Appraising income inequality databases in Latin America

François Bourguignon[1]

Received: 9 March 2015 / Accepted: 28 May 2015 / Published online: 20 August 2015

Abstract This paper provides an evaluation of the main income distribution databases in Latin America: the CEPALSTAT and the SEDLAC databases. Alhough they rely on the same household surveys conducted by national statistical offices in the region, the indicators reported in the two databases differ substantially in a number of cases. Those differences come from distinct adjustments made to the original data, in particular an adjustment to National Accounts aggregate income data in the CEPALSTAT database. Based on this comparison, the paper then provides a general discussion of the adjustment of household survey data to National Accounts as well as other issues which may be responsible for biases in the way inequality is estimated and reported.

Keywords Inequality in Latin America · Income distribution databases · Household survey data and national accounts

1 Introduction

Income inequality in Latin America is an important topic in view of the comparatively high level of inequality in the region. Besides national statistical offices, it is monitored internationally through two databases. The first one is produced by CEPALSTAT, the statistical department of the Economic Commission for Latin America and the Caribbean (ECLAC); the second one, SEDLAC (Socioeconomic Database for Latin America and the Caribbean),

✉ François Bourguignon
francois.bourguignon@psemail.eu

[1] Paris School of Economics, Paris, France

is a joint venture of the Center for Distributive, Labor and Social Studies (CEDLAS), at the National University of la Plata in Argentina, and the World Bank's Poverty and Gender group for Latin America and the Caribbean Region. Both provide time series of inequality and poverty indicators for a majority of countries in the region – although few Caribbean countries are covered. These indicators are based on the microdata collected in the household surveys undertaken by national statistical offices. Yet, the inequality indicators reported in the two databases differ substantially in a number of cases. Even their evolution over time is, at times, not fully comparable. As both databases do produce their inequality indicators on the basis of the same data sources it is important to understand where the difference is coming from and the best use to be made of these two data sets.

It turns out that the main methodological difference between the two sources lies mainly in the fact that the SEDLAC database makes minimal adjustments to the raw data from household surveys (HS) whereas CEPALSTAT adjusts those data so that they fit the household income account of National Accounts (NA). It is not clear, a priori, that this methodological difference necessarily leads to huge discrepancies in the distribution of income and in inequality measures. Yet it potentially can. Also, it is very likely to produce differences in poverty measurement.

There are other possible weaknesses in the methodology used for constructing the two databases, especially when they are compared to distributional data sources in other regions of the world – in particular the OECD Income Distribution Database and the Luxembourg Income Study (LIS) – although they too are imperfect. It thus appears that progress can be made in Latin America in monitoring inequality and poverty, not only in the way the basic data are collected by national statistical bureaus, but also in the way they are processed and then published by CEPALSTAT on the one hand, and CEDLAS and World Bank's LAC poverty and gender group, on the other.

Statistical departments in other international organizations also access primary household survey data to produce summary measures of income inequality in developing countries. This is the case, in particular, of the World Bank and its POVCAL database. When relevant, it will thus be interesting to evaluate how the two databases analyzed in this paper compare with this international source and methodological choices made there. It must be kept in mind however, that POVCAL uses microdata which have been previously processed by SEDLAC, in particular to harmonize income concepts across countries as much as possible. Even though POVCAL makes further adjustments, it cannot be considered as a completely independent source on income inequality and poverty indicators.

The present paper is organized as follows. The first section is devoted to a comparison of the two databases in terms of the income Gini coefficient and the poverty measures they report. A first set of methodological differences in the way the microdata from household surveys are processed by the two sources is scrutinized in Section 2. The main issue taken up in the following section is the major problem of whether HS data must be adjusted to match NA figures, as in CEPALSTAT, or left as much unchanged as possible, as done in SEDLAC. This issue has been discussed abundantly in the literature on poverty measurement based on consumption expenditures, in particular by Deaton (2003, 2005) and Ravallion (2003), but much less in the context of distributional statistics based on income data. Section 3 then focuses on several other issues which may introduce biases in the way inequality is estimated, and which might require more attention as the experience accumulates in dealing with inequality and poverty measurement in Latin America. The final section draws some

general conclusions about ways to improve the monitoring of inequality and poverty across Latin American countries.[1]

2 Income inequality and poverty in Latin America as reported by CEPALSTAT and SEDLAC

Both the CEPALSTAT and SEDLAC databases deliver the message that inequality is high in Latin America, in comparison with the rest of the world. According to the CEPALSTAT, the Gini coefficient of the distribution of disposable household income per capita was above 0.5 average over the last few years, a level comparable to what is observed in sub-Saharan Africa but much above other regions.[2]

The average Gini coefficient is substantially lower, although still high in absolute terms, within the SEDLAC database. As can be seen in Fig. 1, which compares Gini coefficients reported in the two databases, CEPALSTAT indeed tends to report higher levels of inequality than SEDLAC in a majority of countries. In some cases, the difference may be quite substantial: almost 5 % points in the case of Brazil or Guatemala, and more for the Dominican Republic. Yet, the difference is not systematic. Gini Coefficients from the two sources converge in the cases of Panama, Chile and El Salvador. SEDLAC's estimates are even slightly higher than CEPALSTAT's in the case of Costa-Rica, Peru and Uruguay. This heterogeneity implies that the ranking of countries according to inequality may be quite different in the two sources. Brazil is the most inegalitarian Latin American country in CEPALSTAT, but it is overtaken by Guatemala or Colombia in SEDLAC.

Those differences in the level of inequality would be less of a problem if they were approximately constant over time. Were that the case, the two databases would give consistent information about the way in which inequality changes over time across countries. Yet, although this is not the rule, the gap between inequality measures in the two databases is far from being constant.

These gaps are shown for a few selected countries in Fig. 2. For the four countries shown there, trends and trend reversals over the last 15 or 20 years roughly coincide. Yet one can observe several noticeable short-term deviations – circled in Fig. 2. For instance, there has been no change in inequality between 2002 and 2004 in Argentina according to CEPALSTAT, but a 3 percentage points drop according to SEDLAC.[3] Distributional changes at a time of a major crisis and then recovery are an issue of importance, so that such a discrepancy is troublesome. In Brazil, SEDLAC shows the beginning of the fall in inequality as early as in the mid-1990s whereas CEPALSTAT shows a rise.[4] In Mexico, the CEPALSTAT and SEDLAC series are very close to each other. Yet, a discrepancy of close to 2 percentage points in 1996 makes the variation of the Gini between 1994 and 1999 go in opposite direc-

[1] It turns out that no Caribbean country is present in either the CEPALSTAT or the SEDLAC database for more than one or two years. This is the reason this paper focuses exclusively on Latin American countries.

[2] See Alvaredo and Gasparini (2015) for such a comparison based on the World Bank's POVCALdatabase.

[3] Yet, the drop in inequality is comparable when considering the 2002–2006 period. The 2002–04 fall is confirmed in other studies with their own treatment of the microdata (see for instance Lustig et al. (2013)).

[4] SEDLAC is in agreement here with the independent study by Ferreira et al. (2008), based on the same household PNAD primary data used by both SEDLAC and CEPALSTAT.

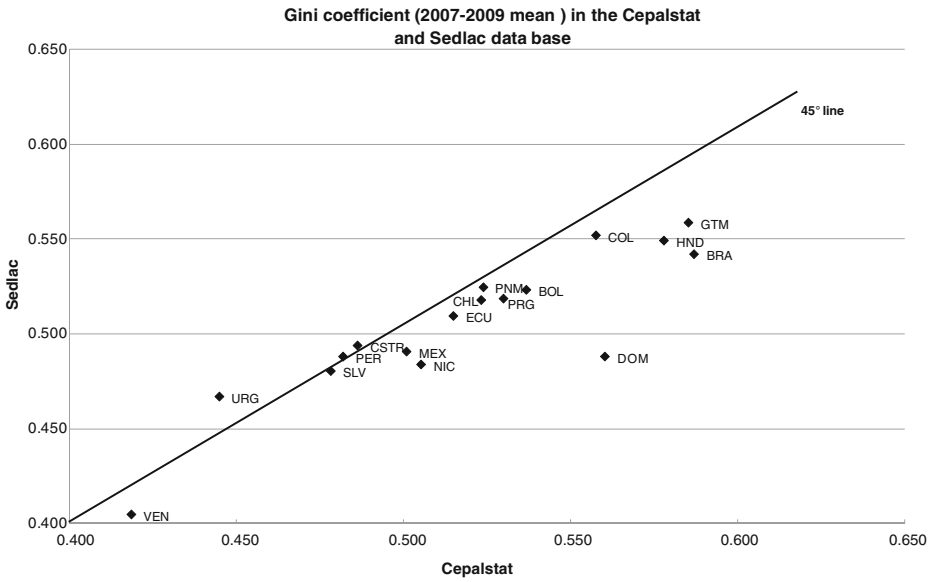


Fig. 1 Gini coefficient (2007–2009 mean) in the Cepalstat and Sedlac data base

tions across the two sources. Finally, Bolivia gives the example of large, but time consistent (except over 2004–2007) discrepancies between the two sources. This is a case of a widely fluctuating series and it is rather natural that the two sources coincide in such an instance, even in showing a huge, and rather suspicious, drop of inequality over the recent years.[5]

Things seem to go better in the recent years. These were rather quiet years in Latin America in general, whereas the discrepancies mentioned above often corresponded to times of crisis. Note also that data are more frequent and are more systematically present in both data sources in the recent period. This was not the case earlier, making the evaluation of the consistency of the two databases more difficult.

Additional data sources are mentioned in the charts appearing in Fig. 2. The most frequent one is the World Bank's POVCAL database, based on the same household survey data as CEPALSTAT and SEDLAC. Ignoring two obvious errors of reporting (in Brazil and Bolivia), the Gini estimates from that source are in general extremely close to SEDLAC's. This is not unexpected since POVCAL applies its own treatment to the harmonized micro-data provided as primary data by SEDLAC. Yet, some substantial departures are readily apparent in the case of Bolivia.

Other additional sources are available in the case of Mexico, which also directly rely on the original data in household surveys. The OECD estimates are actually produced by the countries themselves following OECD guidelines, whereas the LIS estimates result from an independent treatment of the original data by the LIS research team. Both sources report much lower Gini coefficients, the main reason being that they refer to disposable household income *per consumption unit* rather than per capita as an individual living standard. Introducing equivalence scales to account for economies of scale in consumption and the fact

[5]In this respect, it is quite abnormal that no explanation is given of such an anomalous change in either source, and no consistency check of the original household survey data is reported.

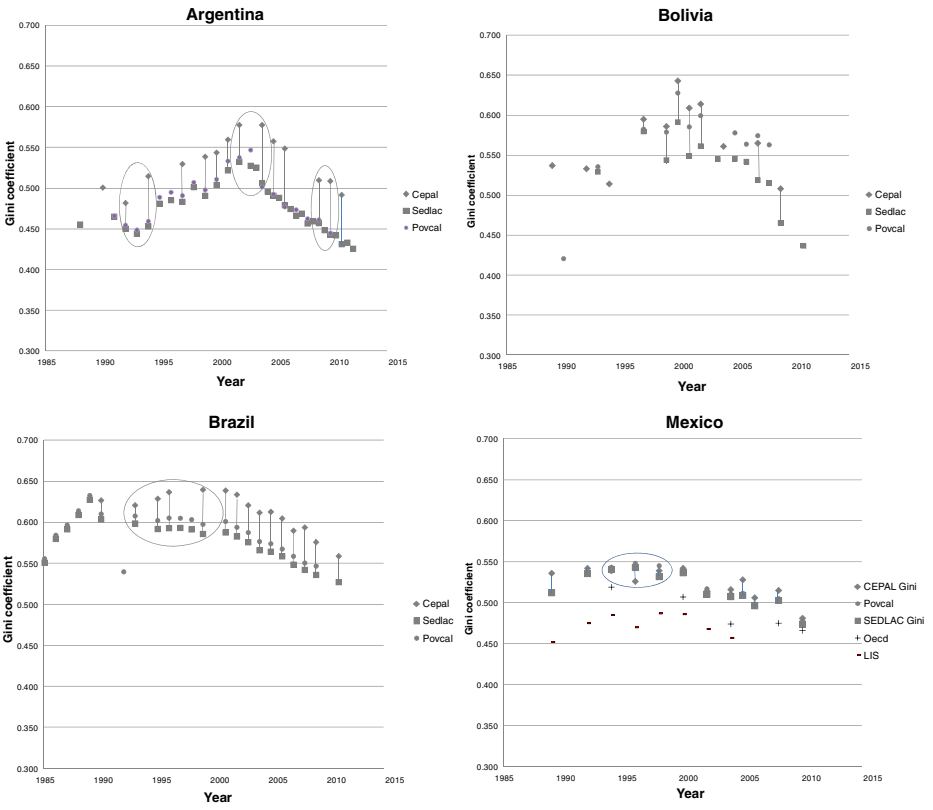


Fig. 2 ComparingGini time series from various sources: selected countries

that children have less consumption needs than adults increases the standard of living of large poor families relative to the case where household income is divided by family size. This logically leads to lower inequality summary measures. As a matter of fact, the SEDLAC site also reports Gini coefficients for equivalized household incomes which are close to OECD estimates.[6]

The three main data sources for Latin America (CEPALSTAT, SEDLAC and POVCAL) also report decile shares and other summary measures of inequality. A comparison of the decile shares across the first two sources does not show any systematic pattern of difference. Yet, it is often, but certainly not always, the case that the main difference lies in the share of the top decile, all the other decile shares being approximately proportional. This is consistent with differences between the two sources being mostly located in high incomes.

A rigorous comparison across sources could not be undertaken for poverty measures. The difficulty here lies in the fact that the two databases use and report poverty as measured with

[6]The issue of equivalence scales is taken up below in Section 3.

different poverty lines. On top of this, each database uses several of them. CEPALSTAT reports poverty measures that rely on specific poverty lines based on the updated value of some basic food basket defined in the 1980s, as described in ECLAC (2013, p. 54). But they also report the poverty measures published by the World Bank out of the POVCAL database with the Millennium Development Goals: the 1.25 PPP 2005 USD a day extreme poverty line and the 2 PPP 2005 USD a day poverty line.[7] CEPALSTAT poverty lines are generally set at a higher level but their level in PPP USD as well as the way they are updated is not known with precision, which is very problematic for interpretation and comparison.

On their side, SEDLAC show poverty measures based on the same data as those used in POVCAL but with international poverty lines of respectively 2.5 and 4 PPP 2005 USD, the former being more or less the median of the official national poverty lines among Latin American countries. They also show relative poverty measures defined on national poverty lines set to 50 % of the median living standard. Finally, they also report national poverty measures as evaluated and published by National Statistical Offices in Latin America, which do not necessarily coincide either with CEPALSTAT's own poverty measures or with the World Bank's POVCAL.

Although no direct comparison is possible across measures based on different poverty lines, they look roughly consistent with each other in the sense that they generally show the same directional change in the poverty headcount. Yet, as with income Gini coefficients, sizable discrepancies are not infrequent. Figure 3 compares CEPALSTAT's estimates of the extreme poverty headcount in selected countries with the POVCAL 1.25 PPP 2005 USD a day estimates – both series are available on the CEPALSAT site. In the four cases shown there, the trends are clearly parallel, with the CEPALSTAT poverty figures generally above POVCAL, thus suggesting that CEPALSTAT's own poverty line is above the $1.25 international poverty line. Brazil and Costa-Rica are good illustrations of that consistency. On the contrary, the Colombian series show a huge discrepancy between the two databases between 2005 and 2008. A change in the survey methodology took place in Colombia at that time and it possible that one of the two databases has not fully taken into account its implications.[8] Another example of inconsistency is provided by Mexico in Fig. 3d. According to CEPALSTAT, the poverty headcount increased between 1992 and 1994, yet it decreased according to POVCAL. The discrepancy is even more serious between 2004 and 2010.

It is not impossible, but somewhat unlikely, that such discrepancies are due to the difference of poverty lines. In the case of Mexico, the contradictory evolution shown in Fig. 3d could be explained by fewer people falling below the $1.25 a day line between 2005 and 2010 and more people with an income per capita between that line and the line used by CEPALSTAT. This is indeed conceivable, but, practically, this is a rather infrequent outcome. Indeed, the POVCAL and SEDLAC poverty headcount estimates defined by the $1.25 a day, $2 a day, and $2.5 a day poverty lines evolved more or less in a parallel way in all Latin American countries. This suggests that there is something else in the difference between CEPALSTAT and other data sources. Given the increasing importance of the poverty measures and their time evolution, it is certainly crucial that some agreement be reached on the poverty line to use – and the way to define living standards. In particular, it might be important for each data source to use and to report on multiple poverty lines below

[7]The PPP notation means that exchange rates to convert local currencies into USD have been adjusted for cost of living differences across countries.

[8]SEDLAC's own estimate using a 2.5 PPP 2005 USD poverty line is consistent with the POVCAL series in Fig. 3.

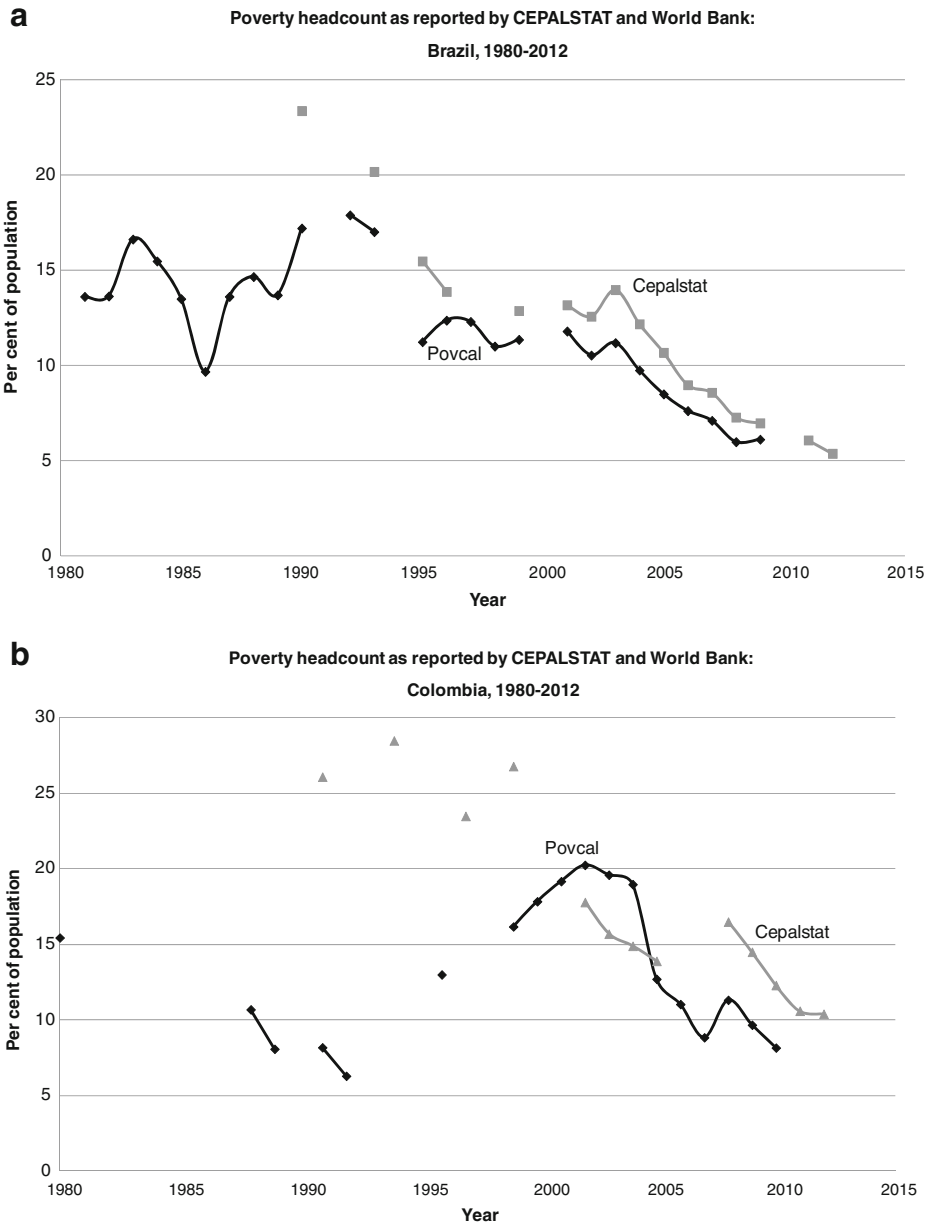
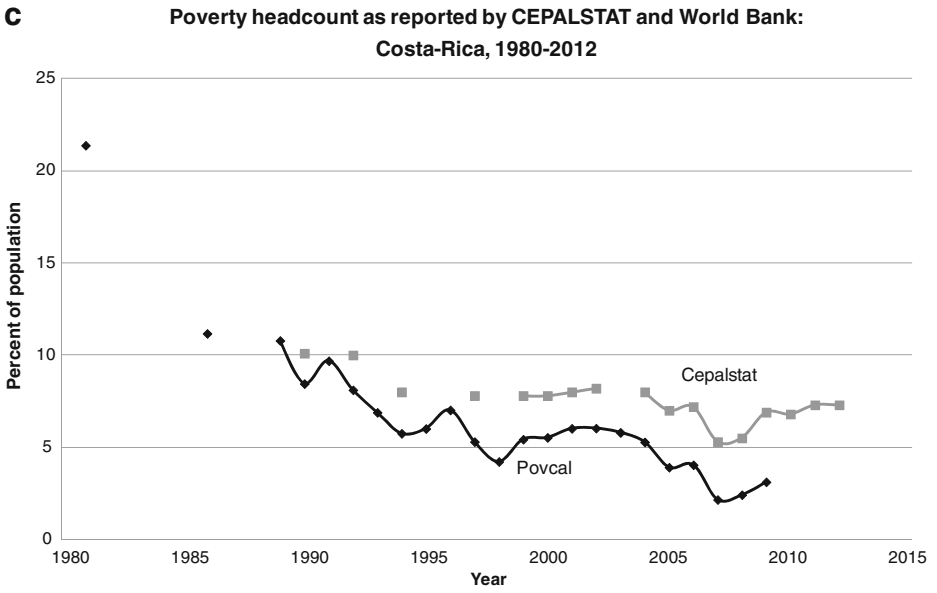


Fig. 3 **a** Poverty headcount as reported by CEPALSTAT and World Bank: Brazil, 1980–2012. **b** Poverty headcount as reported by CEPALSTAT and World Bank: Colombia, 1980–2012. **c** Poverty headcount as reported by CEPALSTAT and World Bank: Costa-Rica, 1980–2012. **d** Poverty headcount as reported by CEPALSTAT and World Bank: Mexico, 1980–2012

some arbitrary maximum. This would both help to make the databases more comparable and allow for a more detailed treatment of the distribution of income among the poor than is the case with standard poverty indicators.



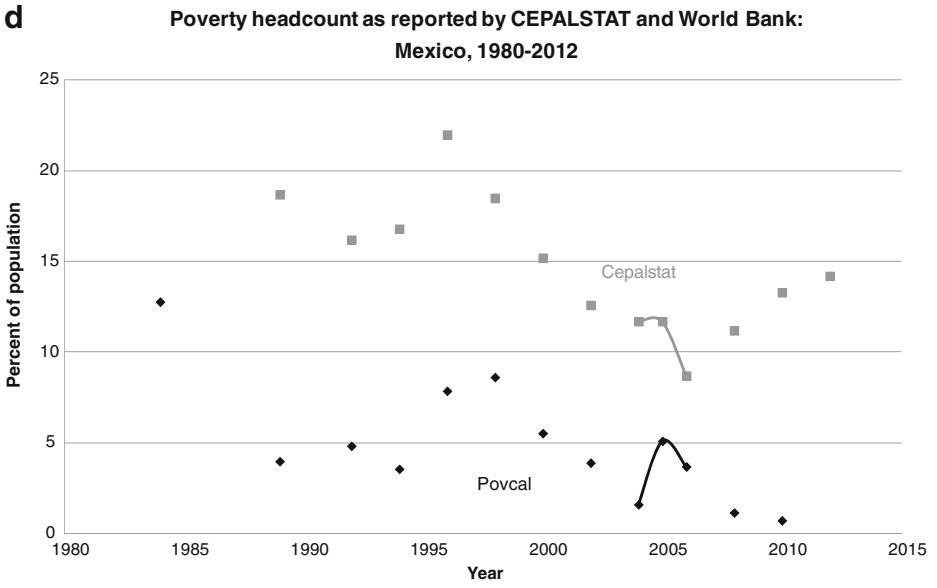


Fig. 3 (continued)

3 Missing data, under-reporting and the National Account-Household Survey gap

CEPALSTAT and SEDLAC rely on identical primary household survey data and have to deal with whatever information is available in those surveys, e.g. income rather than

consumption, imputed values for in-kind payments or benefits and for self-consumption, the presence or the absence of imputed rents for owner-occupiers, a first set of adjustments for missing data, etc.... It follows that the differences in reported distributional indicators can only come from different methodologies in dealing with available data. The methodology used by SEDLAC is precisely described in CEDLAS (2012). CEPAL is about to revise its own methodology but, as of the time this was written, it is unfortunate that no single document seemed to give the full detail of the way the primary data are handled.[9] The basic principles were set in an interesting 1987 paper by O. Altimir about household surveys in Latin America and the quality of the distributional data that could be drawn from them. The way these principles were operationalized is described in some detail in Feres (2004). Several documents also discuss the adjustment procedure as applied to the Chilean HS data – see Beccaria (2007), Bravo and Valderrama Torres (2011), Feres (1997), CEPAL (2012a). The methodological difference between the two sources essentially lies in the way missing income data and under-reporting are being handled, in taking into account urban/rural differences in cost of living and in how imputed rents for owner-occupiers are handled when data are missing or unreliable. We deal with the issue of missing data and under-reporting in the present section and with the other issues in the next.

Few corrections of the original data are made by SEDLAC, whose main task really consists of harmonizing the definition of household income in inequality measurement across countries and over time. As far as missing incomes are concerned, they are handled as follows. If the information for income source s of individual i is missing, then it is taken to be zero provided s is not the main income source of the individual – essentially labor income but possibly pensions. If s is the main income source, individual i's income is simply disregarded when computing the total income of the household (s)he belongs to,[10] except if i is the head of the household, in which case the household observation is dropped. The same rule applies to zero incomes when there are good reasons to expect that a particular income should not be zero – e.g. a salaried worker reporting zero salary income. On the other hand, non-zero incomes are left uncorrected, even when there are reasons to suspect they are under-reported.

CEPALSTAT's attitude is more interventionist. A first set of adjustments is made for missing, and presumably zero income data when there are good reasons to believe that this income source should not be zero – people who report being gainfully employed not reporting their salary income, pensioners not reporting pensions or home owner-occupiers not reporting imputed rents. In these cases, a value is imputed that is obtained by a matching technique – i.e. equal to the average value of the same variable among people or households with complete information and the same socio-demographic characteristics – or by 'hot deck' – drawing randomly an observation with complete data and the same socio-demographic characteristics.

A second adjustment is then made for under-reporting as measured by the discrepancy between the aggregate income reported by income source in the HS in comparison with entries in the household income account of the NA. For each income source, all individual

[9] At this stage, CEPALSTAT's modus operandi in treating household survey data can only be checked in a piecemeal way in a few working papers in the CEPALSTAT series. Yet, the view that one gets from such an investigation is only partial.

[10] Also, the income of domestic servants or people renting a room in the dwelling are not considered as household members and their own income is not included in the household income. Although not seen explicitly in the available documents, it may be assumed that CEPALSTAT does the same.

Table 1 Chile: household survey to national account adjustment factors by income source

	NA/HS income ratio (all households)							Structure of total income in 2011 (%)	
Year	1996	1998	2000	2003	2006	2009	2011	NA	HS
Wage and salaries	0.99	1.00	0.96	1.00	1.01	1.10	1.09	55.3	60.3
Self-employment	2.04	1.96	1.83	1.98	1.98	2.05	2.05	27.4	15.9
Pensions and benefits	1.40	1.35	1.47	1.15	1.13	0.98	0.98	7.4	9.0
Property	2.74	2.75	2.43	1.84	2.14	1.94	3.51	4.8	1.7
(Top quintile adjustment)[a]	*1.06*	*1.07*	*1.05*	*1.03*	*1.04*	*1.03*	*1.06*		
Imputed rents	0.45	0.44	0.45	0.44	0.44	0.42	0.48	5.1	13.1
Total	1.21	1.19	1.15	1.14	1.15	1.19	1.19	100.0	100.0

Source: Bravo and Valderrama Torres (2011) for 1996–2000, CEPAL (2012a, b) for 2003–2011

Coefficients used to adjust household survey income data so that they fit the household account figures in National Accounts

[a]Adjustment factor when the whole discrepancy between NA ad HS is imputed to the top quintile group proportionally to total market income

data in the survey are thus scaled up so as for the household survey total income figures to match the entries in the National Accounts. Thus, salary income is multiplied by a factor generally slightly above unity whereas self-employment income is expanded by a factor that may be close to two in some cases. An exception to this proportional correction applies to the 'property income', for which the adjustment bears exclusively on the incomes reported in the top quintile group of the distribution, so as to account for the fact that this type of income predominantly accrues to well-off households. In principle, another exception occurs in the case where survey data over-estimate National Accounts, in which case no correction is performed. This is seldom the case, except for self-reported imputed rents which, on average, owner-occupiers tend to over-estimate. In the case of Chile, for instance, this over-estimation is close to 100 % – see below – and survey data are scaled down in that particular instance.[11]

Unfortunately, the information on these coefficients of adjustment of HS to NA are not available on a regular basis for all countries in the region, so that it is difficult to evaluate systematically the actual change such a procedure causes on the estimated distribution of income. Yet, judging from the coefficients publicly released in the case of Chile, the change is potentially substantial. When the National Socioeconomic Characterization Survey (CASEN) is compared to National Accounts – after multiplying reported monthly incomes by 12 and taking into account the effect of inflation during the year – it can be seen in Table 1 that self-employment income, pensions and property income are quite off the mark. The reverse is true of imputed rents, and the gap is moderate in the case of salaries.

Presumably scaling self-employment incomes by 2 or property income by almost 3 is not likely to leave the distribution of household income per capita unchanged. A simple calculation based on quintile group shares and the relative size of the NA-HS gap in property income gives some idea of the direction and the size of this adjustment on inequality

[11]Table 1 also suggests that HS wage and salary incomes were marginally scaled down in 1997 and 2000.

estimates. The results of the calculation are shown in Table 2 in the case of Chile in 2009 and 2011, and in the case of Brazil for 2005. It can be seen there that adjusting to NA logically tends to increase inequality since the adjustment is equivalent to increasing the income of the richest quintile.[12] As the property income gap represents a rather small proportion of the total household income in 2009's Chile, the correction is of moderate size. It is much more pronounced in 2011 because the property income gap is larger. Then, this simple correction amounts to a 1.2 percentage points in the Gini coefficient. The same adjustment is much more sizable with Brazilian data – almost 2 percentage points – because the property income gap represents a much larger share of the total household income and the share of the top quintile is also bigger than in Chile.

Implicitly, the calculation reported in Table 2 assumes that the adjustment of the other income sources is strictly proportional to reported HS incomes. This is far from being the case and other effects may come on top of the property income correction. As imputed rents in Chile are being more concentrated at the top of the distribution than other incomes, scaling down this income to make it consistent with NA data necessarily reduces inequality. On the other hand, scaling up self-employment income is likely to have an ambiguous effect since this type of income tends to concentrate at the two ends of the income distribution.

Thanks to their access to the original HS data, Bravo and Valderrama Torres (2011) were able to simulate the full effect of the NA/HS adjustment on the estimation of inequality in Chile between 1996 and 2006. Overall, they found that the adjustment tended to increase inequality, especially with inequality measures that give very much weight to the top of the distribution. The adjustment to NA turned out to be responsible for approximately a 3 % point increase in the Gini coefficient, substantially more than suggested by the illustrative calculation above. A similar illustrative comparison is reported by Feres (2004, p.398) for a few countries during the 1980–1999 period. The effect of the adjustment is found to vary across countries. It leads to an increase in the Gini coefficient roughly equal to 4 percentage points in the case of Brazil and Chile, no change in the case of Mexico or Panama and a 3 to 5 percentage point drop in the case of Costa-Rica. No explanation is given about why the NA adjustment leads to such different results. It is also interesting that some of these variations do not fit the SEDLAC-CEPALSTAT difference reported in Fig. 1, despite the fact that SEDLAC data should presumably be quite close to the unadjusted inequality indicators in Feres (2004). For instance, according to the preceding result and assuming no big change in the direction and size of the adjustment, Costa-Rica should be at some distance on the left hand side of the diagonal in Fig. 1 instead of next to it.

This last remark raises the issue of the stability of the NA adjustment over time. The example of the calculation made for Chile in 2009 and 2011 suggests some variability of the adjustment due to the changing size of the gaps between NA and HS. In view of this simple example, even the idea that the adjustment modifies cross-country comparisons but not really inequality trends is far from certain. It can be seen in Table 2 that the drop in the NA adjusted Ginis between 2009 and 2011 is much smaller than the drop in the unadjusted HS Gini.[13] In Feres (2004), it turns out that the NA adjustment modifies the trend

[12] If v_i the original share of quintile group i in the raw data, the adjustment leads to a new share v'_I given by: $v'_I = v_i.100/(100+P)$ for $i \neq 5$ and $v'_5 = (v_5 + P).100/(100+P)$ where P is the share of the property income gap in the total household income.

[13] Bravo and Valderrama Torres (2011) find the effect of the adjustment on the Gini coefficient – but not for the coefficient of variation or the Top 10 %/Bottom 10 % mean income ratio – is rather stable between 1996 and 2006. Yet, it can be seen in Table 1 that the property income gap was more stable over that period than it has been more recently.

Table 2 Inequality effect of adjusting the NA/HS property income gap on the top quintile group: rough calculation on Chile and Brazil

	Aggregate income by source (%)				Quintile group shares[a] (%)	
	Household survey (HS)[b]	NA-HS gap as % of HS total income	National Accounts (NA)		Household survey	NA-Adjusted
Chile (2009)						
Labor income	75.7	22.2	84.4	0–20%	4.5	4.4
Property income	2.5	1.9	3.9	20–40 %	8.2	8.0
Transfers	8.5	0.0	7.0	40–60 %	11.9	11.7
Imputed rents	13.3	−6.3	4.6	60–80 %	18.7	18.3
Total	100	17.8	100	80–100 %	56.8	57.6
				Gini[c]	46.0	46.7
Chile (2011)						
Labor income	76.3	19.9	82.7	0–20 %	4.8	4.6
Property income	1.7	3.4	4.8	20–40 %	8.5	8.2
Transfers	9.0	0.0	7.4	40–60 %	12.2	11.8
Imputed rents	13.1	−5.7	5.1	60–80 %	19.1	18.4
Total	100	17.6	100	80–100 %	55.5	57.0
				Gini[c]	44.8	46.0
Brazil (2005)						
Labor income	76.2	−4.1	62.6	0–20 %	3.0	2.8
Property income	3.6	10.1	11.9	20–40 %	6.5	6.1
Transfers	20.2	9.2	25.5	40–60 %	11.0	10.3
				60–80 %	18.6	17.4
Total	100.0	15.2	100.0	80–100 %	60.9	63.4
				Gini[c]	51.2	53.0

[a]For Brazil, the household survey quintile group share are from Sedlac. For Chile the adjustment goes in the opposite direction. As Sedlac gives NA-adjusted quintile group shares, the correction procedure estimates the HS quintile share which would have led to the Sedlac shares with the procedure described in Adjustment consists of allocating the NA-HS property income gap to top quintile group

[b]In Chile, HS aggregate income by source is given in CEPAL (2012a, b). For Brazil, this information is taken from Medina and Galvan (2008)

[c]Calculated on quintile group shares

Source: Author's calculation

of the Gini coefficient in Brazil during the 1990s. It is falling without the adjustment and rising after adjustment, thus providing an explanation to the discrepancy noted in Fig. 2. Unfortunately, no data are regularly published on the decomposition of aggregate household income by source in household surveys, so that it is difficult to get an intuition for the cause of such diverging trends. This would be very valuable information and it should be recom-

mended that both CEPALSTAT and SEDLAC publish such information on a regular basis. On the other hand, it must be pointed out that some countries do not report a full Household Account as part of their NA system. In such a case, CEPALSTAT relies on own estimates based on scattered information. These estimates should be reported too.

An issue that is beyond the scope of the present paper despite its relevance for the comparison between NA and HS is that of the definitions being used and the data collection methodology in NA. Some household income components like salaries, dividends, pensions, and social transfers are readily estimated because they are most often provided on a compulsory basis by formal units of production and government agencies. Things are not as simple for the informal sector of the economy. A variety of sources are being used for the latter, which include informal sector surveys – like the ENANIM in Mexico[14] or the 'Economia Informal Urbana' survey in Brazil – as well as labor force and household surveys themselves. The latter provide year by year information on formal and informal employment, which may be combined with other less regular sources to produce estimates of total informal value added and income.[15] More generally, the consistency of NA and HS is better when the former makes direct use of the latter and the coefficients that permit expanding the information in the survey to cover all operations of the informal sector are not modified too often. Could this be the cause of the congruence between SEDLAC and CEPALSTAT observed for several countries in Fig. 1 and the disparity observed for others? This is difficult to say without a detailed knowledge of the National Account methodology in use in the various countries appearing in this figure. It should be expected from National Statistical Offices that they publish regularly a full account of the way in which they estimate the size and activity of the informal sector. Yet, the general information in the methodological volumes they release is most often insufficient to understand the possible origin of the discrepancies with household surveys.

Now that we understand the nature of the possible discrepancies between NA adjusted and plain HS distributional data and the variability of those across countries and possibly over time, what conclusion should we draw? Is there any reason to prefer one to the other?

There is a large literature dealing with the issue of whether HS should be adjusted so that the mean income or consumption expenditure fit NA. As a matter of fact, the agreement is rather wide that such a procedure should be avoided when measuring poverty.[16] As the under-reporting in household surveys is likely to come from the top part of the distribution, applying a proportional correction to all survey incomes or consumption expenditures per capita to fit the National Accounts would introduce a rather severe bias in the estimation of poverty. Restricting the porportional adjustment to all salaries and self-employment earnings also runs the risk of introducing a downward bias in poverty measurement if, as is generally expected, actual under-reporting is more severe at the top of the distribution. Even

[14]Encuesta nacional de micro negocios see INEGI, Mexico: http://www.inegi.org.mx/est/contenidos/proyectos/encuestas/hogares/modulos/enamin/enamin2012/

[15]A detailed methodological account of how to measure the informal sector in developing economies is provided by ILO (2013). See also specific methodological issues in CEPAL (2012b).

[16]See in particular Deaton (2001, 2003, 2005) and Ravallion (2001, 2003). Such a view seems ideed in full contradiction with the position of Cepal as stated in various documents. For instance, Beccaria (2007) ends the summary of a paper dealing with the measurement of poverty in Latin America with the following statement:" (The paper thus).. provides an additional reason for adjusting the income data in household surveys through the comparison of the main aggregates obtained from them with their counterparts in household incomes recorded by the National Accounts system".

though there may also be some under-reporting among poor people, the effect of under-reporting on poverty measurement is most likely to be smaller than the effect of the gap between NA and HS.[17]

The case against adjusting HS data may be slightly less strong when dealing with inequality rather than poverty and with income rather than consumption. If the under-reporting in household surveys comes from the richest households, unadjusted data from HS will underestimate inequality. Unlike in the case of consumption, incomes are observed from various sources, the misreporting of which have differing effects on the overall distribution of income. Property income is known to be concentrated in the top percentiles of the distribution, whereas, in developing countries, salaries probably accrue to the middle of it. Thus, knowing the sources of the overall gap between NA and HS certainly help determining the direction and the order of magnitude of the bias in inequality due to HS under-reporting. The calculations reported in Table 2 are a good illustration of this approach.

Is it possible to go beyond this kind of adjustment without adding noise rather than correcting biases? This is not clear. As other incomes than property are more uniformly distributed in the population there is a real danger that, by trying to correct the inequality bias that comes from the under-reporting of specific income sources, one would in fact add imprecision to the estimates obtained from raw survey data. For this to be the case, it would be sufficient that the rate of under-reporting of that income source is correlated with total household income per capita. Even for property income, imputing the gap to the top quintile may lead to some severe under-estimation of actual inequality if, as suggested by the recent work on 'top incomes' – see Atkinson and Piketty (2010) – the under-reporting emanates mostly from very rich or 'super-rich' households, which are most likely absent from household survey samples.[18]

When one adds to all of this that there also is some imprecision in the way NA themselves evaluate the activity and the aggregate income of the informal sector and some particular non-market income sources like imputed rents, the case for a full adjustment of HS data to NA aggregates weakens.

Should distribution analysis thus focus essentially on surveys ignoring the information coming from National Accounts? The Chilean and Brazilian examples discussed above suggest that this would be disregarding some valuable information. It is of crucial importance to know that some evolution of the observed HS income distribution comes together with substantial changes in the apparent gaps with respect to NA, whether overall or for specific income sources like property or self-employment incomes. This should encourage the analyst to investigate the possible causes for such changes, wondering whether the methodology is faulty either on the side of the HS or that of NA, or whether substantial modifications have taken place in the economy that could explain the change in the size of the gap. If on the contrary, no noticeable alteration in the relative sizes of the gaps were detected, changes in the survey distribution of income would acquire more significance.

Table 3 shows the evolution of the ratio between the mean income per capita in HS and the consumption expenditure per capita in NA over the last 10 to 15 years for selected countries regularly covered by SEDLAC. Note that the comparison is between income on the one hand and consumption expenditure on the other. This is because not all countries report

[17] For a general discussion of biases in the measurement of poverty in Latin America, especially those arising from the inconsistency between NA and HS, see Szekely et al. (2004).

[18] For a discussion of that point and an example of combining survey data with top income tax data see Alvaredo and Londoño-Velez (2013) on Colombia.

Table 3 Ratio of the mean income in household survey to the mean household final consumption expenditure per capita in National Accounts

Country	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Bolivia	1.26		1.07	1.01	1.08	1.08			1.17	1.19	1.17	1.21	1.27		1.26	1.27
Brazil				0.84	0.84	0.85	0.82	0.82	0.84	0.86	0.83	0.84	0.82		0.78	0.81
Colombia					0.50	0.67	0.60	0.63	0.66			0.65	0.68	0.71	0.71	0.67
Costa-Rica					0.80	0.79	0.80	0.75	0.76	0.75	0.80	0.80	0.90			
Dominican Republic				0.92	0.88	0.82	0.69	0.58	0.60	0.57	0.59	0.48	0.55	0.49	0.47	0.50
Ecuador				0.47			0.66	0.87	0.70	0.75	0.75	0.66	0.66	0.70	0.69	0.71
El Salvador								0.57	0.56	0.53	0.54	0.49	0.55	0.52	0.50	0.51
Honduras					1.13	0.93	0.95	0.96	0.91	0.92	0.98	1.03	1.03	1.01	0.98	
Mexico		0.44		0.49		0.47		0.43	0.43	0.43		0.43		0.42		0.44
Paraguay	1.44		1.34		1.32	1.22	1.26	1.15	1.18	1.08	1.07	0.98	1.10	1.06	1.06	0.00
Peru	0.74	0.81	0.81	0.70	0.67	0.76	0.71	0.72	0.67	0.72	0.75	0.73	0.77	0.78	0.76	0.77
Uruguay	0.76	0.80		0.83	0.82	0.76	0.71	0.70	0.70	0.69	0.69	0.71	0.82	0.74	0.73	0.70

Source: Sedlac and WDI, author's calculation

household disposable income in their NA and consumption expenditure is known to evolve in a more or less parallel way with disposable income. The figures in that table are at the same time worrying and reassuring. They are worrying because they show a huge heterogeneity of countries with respect to this particular HS/NA gap, which varies from 50 % to more than 100 % across countries. If National Accounts are in any sense right, the cross-country structure of HS income per capita is quite different from that of NA consumption, and most likely NA household income per capita. But figures in Table 3 are somewhat reassuring because, in general, they appear to be rather stable over time for the various countries in the table. Of course, there are countries like the Dominican Republic, Paraguay or even Uruguay where the ratio shows a very clear downward trend. There also are temporary changes like those in Bolivia starting in 2008, and in Uruguay or in Colombia from 2008 to 2011. These should be the warning signals mentioned above, alerting the analysts about the need to investigate the cause for such changes, including possible modifications in the household survey or the National Account methodology and the worsening of biases in estimates of the income distribution.

4 Other issues in dealing with household surveys

Until now, this paper has concentrated on the major cause for differences between reported distributional statistics by SEDLAC and CEPALSTAT, which is undoubtedly the adjustment of HS data to NA performed by the latter. Yet, the rather complete document that describes the methodology used by SEDLAC stresses various other issues worth some discussion.

Some issues have to do with the surveys themselves. Maybe the most crucial one is whether regular Latin American surveys should keep focusing exclusively on income only and not ask questions about consumption expenditures, which would give a better idea of the distribution of welfare in the population. At the same time, income is the concept used in most developed countries to measure income inequality, and the one a database like LIS focuses on. A few years ago, another major issue would also have been the partial coverage of the surveys as in many countries the rural sector was left aside. Today, Argentina may be the only country where HS data do not cover the rural areas. There also are various improvements that could be made in the way surveys are taken or in the questionnaires themselves. But, these are issues to be taken up at the level of national statistical offices which go beyond the present paper and, as a matter of fact, also beyond SEDLAC and CEPALSTAT responsibilities. What follows focuses essentially on the way the two databases deal with specific data or conceptual issues in elaborating distributional indicators.

Non-response

This issue has been alluded to earlier in this paper when stressing that CEPALSTAT was using matching techniques in case of major non-response – i.e. where the non-response was concerned with a major component of household income – whereas SEDLAC was simply dropping the corresponding observations. Both approaches have weaknesses. Matching techniques, which may be considered as particular cases of regression techniques, replace the missing variable with the predicted value of a regression run on observations with complete data and where regressors are the observed socio-economic characteristics of households or individual household members. Typically, a missing earning for an active individual would be replaced either by the average earning of people with the same characteristics – age, education, occupation, etc. – or by the predicted value of a regression of earnings on the same characteristics. Issues of concern in both cases are: a) the need to

introduce some heterogeneity in the procedure so as not to artificially reduce inequality by imputing the same earnings to sub-populations where this variable was missing and with identical socio-economic characterisics; b) the neglect of possible selection bias behind non-reporting, which means that non-reporting people are not a random sample of the reporting population.

A simple way of reintroducing heterogeneity in the matching or regression methodology consists of introducing random deviations from the matched or predicted value drawn from the distribution estimated from reporting individuals. This does not solve the selection bias, though. In the case of missing earnings data, this bias can be corrected through the standard two-stage procedure a la Heckman by which one first estimates the probability that an individual reports his/her earnings and then runs the earning regression, adding to the list of regressors a term that depends on the probability of earnings being reported. In the case of matching, the mean earnings among individuals with the same socio-economic characteristics should be evaluated by weighting the corresponding observations by the probability that earnings data are not missing.[19]

An alternative that does not need to rely on a specific model or a specific set of observable characteristics for imputing a value to the missing data consists of eliminating the observations with missing data and reweighting all households within the sample with complete information using the ex-ante estimated probability for a household to report the relevant information. The selection model would then be based on a set of variables that are readily available throughout the entire sample.

Of course, such a procedure, or the more cumbersome matching and regression procedures should be applied only insofar as: a) some characteristics indeed prove to be significant in the selection probability; and b) the overall impact on distributional statistics has any statistical significance. Some tests should be run on a reduced set of surveys to see whether it is worth embarking in such a procedure or whether one can be satisfied with the SEDLAC procedure to mostly ignore non-response observations.[20]

Equivalence scales

That the distribution of income depends on the kind of equivalence scale that is being used to define the welfare of the average member of a household is pretty obvious. The relative welfare of large families, which often are poor, is much higher when economies of scale or the lesser needs of younger children are taken into account. This is readily seen on the online results shown by SEDLAC with alternative equivalence scales. Inequality is substantially reduced when moving from the per capita to more general equivalence scales.

The dominance of the income (or consumption expenditure) per capita as a welfare norm in international databases owes very much to the international definition of poverty being set on per capita terms – i.e. the 1.25 PPP 2005 USD per day and per person poverty line in the first objective of the Millennium Development Goals. In advanced countries, however, poverty is systematically measured on the basis of 'equivalized' income, as in Europe, or using poverty lines that differ across family sizes, as in the US.

[19] See Little and Rubin (1987).

[20] Yet, this procedure might be preferable to the way SEDLAC presently deals with zero incomes, discarding the corresponding observation when estimating inequality but keeping them for poverty measurement. It is not unlikely that such a treatment leads to an underestimation of inequality and an over-estimation of poverty, if indeed zero income cases are more frequent among households with relatively low income potential.

To the extent that the composition of families changes with the development process, working exclusively with per capita income or expenditures may miss an important part of distributional changes. It is true that equivalence scales are likely to differ across countries and there would be something arbitrary in choosing one particular scale. Yet, there is the same level of arbitrariness in using per capita income to measure poverty or inequality. Moreover, in the case of Latin America the countries may not be too different to establish some uniform scale to be used in estimating the distribution of income. The OECD scale, for instance, is of rather wide use and could be a common norm for evaluating poverty and inequality among Latin American and Caribbean countries. Of course, this would not prevent SEDLAC or CEPALSTAT to simultaneously publish inequality and poverty figures defined in per capita terms.[21]

Imputed rents

The implicit income of home owner-occupiers should certainly be part of their total income when they are compared to households who rent their flat or their house. But the information on the market value of this implicit rent is very often missing or, when reported by households less aware of the rental market around them, quite unreliable. SEDLAC handles that issue by granting a 10 % income premium to owner-occupiers in cases where imputed rents reported by households are missing or clearly inconsistent, whereas CEPALSTAT uses data matching. The 10 per cent imputation used by SEDLAC corresponds to the average size of reported imputed rents by owner-occupiers relative to their income in Latin America (Beccaria and Guzmann, 2013, p. 42).

It may be asked whether a simple x percent imputation for all owner-occupiers would not be the best way of dealing with this difficult misreporting issue, the distributional impact of which is far from negligible. Some support for such an approach is to be sought in the literature on the demand for housing services, where the income elasticity of housing expenditures appear to be close to unity.[22] Yet, another difficulty with both using reported imputed rents and making an x percent allowance for it has to do with mortgages. At a given point of time, an owner-occupier household reporting an imputed rent exactly equal to its mortgage is strictly equivalent to a renter. This means that mortgage information should be collected in addition to owner-occupier status and/or imputed rent if this matter is to be handled in a rigorous way. Some partial information is sometimes available in household surveys that permit to distinguish owners and 'acquirers'.

Geographical differences in the cost of living

Geographical differences in the cost of living within a country can substantially affect inequality and poverty estimates at the national level. In the absence of precise data, SEDLAC researchers propose to simply scale up rural household incomes by 15 % in all Latin American countries. On the other side, CEPALSTAT relies on food prices drawn from household budget survey information to estimate the cost of living differential between

[21] Of course, poverty lines to be used with equivalized income should be made consistent with poverty lines presently defined in per capita terms.

[22] An early attempt at synthesizing what was known for developing countries is Malpezzi and Mayo (1987), A survey of the more recent literature and possibly new studies would have to be undertaken to check this important point.

urban and rural areas. Yet, such a procedure has clear limits. Inferring prices from reported unit costs in household consumption data is notoriously imprecise. Moreover, even in rural areas, food accounts for less than half of a households' budget. Note however how important this correction for cost of living differences across geographical areas is for poverty measurement in view of the fact that in most countries poverty tends to concentrate in rural areas and small cities.

Without price data of strictly comparable goods – food products but also fuel, transport, housing, etc. – systematically surveyed in different parts of a country, there may be little else to do other than relying on some rule of thumb of the type used by SEDLAC. Yet, it should be checked whether the International Comparison Program that recently collected price data on a very extensive basis, often at the sub-national level, to establish international purchasing power parities, could not help to partly solve that problem (see World Bank 2013).

Poverty measurement and poverty lines

The estimation of the extent of poverty in a world where countries or statistical entities use different poverty lines and different definitions of household income necessarily leads to inconsistencies. It was shown above that it was most likely that such an inconsistency exists between SEDLAC and CEPALSTAT, even though they do not report poverty measures based on the same poverty lines. This is too important an issue both at the national and the international levels to leave things in such disorder.

A few rules should be sufficient to reestablish consistency and comparability. Following the previous discussion of the consequences of adjusting survey data to National Accounts, a first rule would be to rely exclusively on the raw household survey data to measure poverty. As seen above, the risk would be that NA adjustment rules introduce a severe bias in poverty measurement. The second rule should precisely be to systematically provide poverty measures defined on several poverty lines. As a matter of fact, such a rule would simply put into practice the notion of 'poverty dominance', introduced more than two decades ago in the poverty measurement literature (Foster and Shorrocks 1988). According to that principle, there is less poverty in distribution A than in B for all social welfare consistent poverty measures only if the poverty headcount is lower in A than in B for all possible poverty lines below the one actually chosen. In Latin America, the MDG $1.25 per person and per day poverty line, in 2005 purchasing power parity dollars, does not mean very much as only a small proportion of people are below it. It is nevertheless important to know how many they are. The same may apply to $2 or $2.5 a day, as official poverty lines often are above these thresholds. Reporting poverty headcounts for this multiplicity of lines might in some cases blur the message that poverty has fallen or increased, but it would also better inform policy makers and the public about the actual state of affairs. At the same time, there would be some cause for concern if the poverty headcount did not go down simultaneously across all those poverty lines.

Practically, both SEDLAC and CEPALSTAT report poverty measures defined on several poverty lines but they do it in some scattered way and, in the case of CEPALSTAT, without even mentioning the actual level of the poverty line in international purchasing power parity USD. What is needed is to expand the notion of 'extreme' and 'less extreme' poverty used in CEPALSTAT to a set of poverty lines covering a broader interval, including of course the $1.25 and $2 a day. If all statistical entities elaborating and publishing poverty measures followed that rule, they would also be constrained to be more mutually consistent than they are today. SEDLAC and CEPALSTAT could be the obvious pioneers in this area.

5 Conclusion

CEDLAS and ECLAC, the organizations behind SEDLAC and CEPALSTAT respectively, have to be commended for the work they do and their efforts at providing income distribution estimates in a way that is consistent both across countries and over time. At a time in which inequality issues are getting more traction around the world, and still more so in Latin America, this is a most important endeavor. This is the reason that inconsistencies that do exist between the databases created and maintained by these two entities should be reduced so as to give them more credibility, more robustness, and more complementarity. The methodologies they use are different. The SEDLAC database relies more directly on raw household survey data, whereas the CEPALSTAT database adjusts them so as to fit National Account data. Both approaches are relevant. Income distribution and poverty statistics based on minimally adjusted survey data are of course essential. But analyzing the actual coverage of household survey data in the light of National Accounts is crucial too.

Yet, it must be kept in mind that the NA adjustment actually adds a layer of measurement errors in the estimation of income distribution indicators. There is some uncertainty on what National Accounts actually measure under the headings of disposable household income and its various components, as emphasized by Deaton (2005).There is also uncertainty about the effect of applying uniform adjustment ratios to the whole population - or the upper quintile for property income. Under these conditions, it is absolutely essential that distributional indicators based on unadjusted data be made available in a consistent manner over time, an enterprise which is quite successfully and remarkably managed by the entities responsible for SEDLAC. NA adjusted indicators are needed too as they remedy some structural weaknesses of survey data, like the under-reporting of some specific income sources. But it is important that more detail be reported about the nature and size of the adjustments being performed, year by year, so that users are able to understand the possible causes of the discrepancy with unadjusted indicators.

The idea would thus be to move to a kind of dual apprehension of distributional changes within an economy that would consider simultaneously the distributional statistics drawn from the raw survey data and the nature and size of the gaps with NA. The appraisal of the evolution of the income distribution would thus be typically multidimensional. For instance, one would follow the Gini coefficient defined by HS data, but also the standard household aggregates in the NA and simple ratios as the one shown above comparing mean survey incomes and aggregate private consumptions expenditure as a proxy of disposable income per capita. That household property income increases faster than other income sources or that survey incomes grow systematically at a slower rate than NA aggregates may be the sign that changes may be taking place in the distribution of income that are imperfectly recorded in surveys. One could thus draw conclusions of the following type: "Yes, income inequality has gone down in the household survey but the coverage of the survey has fallen relative to NA or property income has increased in comparison with other income sources, which may mean more inequality at the top". Of course, estimates of the evolution of the income distribution when survey data are adjusted to NA would permit us to go beyond such simple statements, provided the way they have been elaborated is fully documented. Things are different for poverty measurement, however, as it was seen that any adjustment in the raw data for potential under-coverage, except possibly for missing data, is likely to introduce serious biases.

This paper has also listed several technical points which would be worth investigating further for improving the appraisal of distributional matters in Latin America. However,

the role of national Statistical Offices in improving the quality of the indicators provided by international databases should not be under-estimated. In this respect, better recording individual incomes, checking their consistency with consumption expenditure information, and possibly combining surveys with administrative data when possible should be seriously considered.

Acknowledgments I thank Francisco Ferreira, Nora Lustig, Leonardo Gasparini, Xavier Mancero and Daniel Teles for most helpful comments. Remaining errors or omissions are my responsibility.

References

Altimir, O.: Income distribution statistics in Latin America and their reliability. Rev. Income Wealth, 111–54 (1987)
Alvaredo, F., Gasparini, L.: Recent trends in inequality and poverty in developing countries. In: Atkinson, A., Bourguignon, F. (eds.) *Handbook of income distribution*, vol. 2. Elsevier (2015)
Alvaredo, F., Londoño-Velez, J.: High incomes and personal taxation in a developing economy: Colombia 1993–2013. Commitment to Equity-CEQ Working Paper n. 12 (2013)
Atkinson, A.B., Piketty, T.: Top incomes: a global perspective. Oxford University Press (2010)
Beccaria, L.: La medición del ingreso para los estudios de pobreza en América Latina: aspectos conceptuales y empíricos, Estudios estadisticos y prospectivos, N° 60. Cepal, Santiago (2007)
Beccaria, L., Guzmann: Medición de los Ingresos y la Pobreza Oficial en América Latina y el Caribe, Documento de Trabajo 148, CEDLAS, Universidad Nacional de la Plata (2013)
Bravo, D., Valderrama Torres, J.A.: The impact of income adjustments in the Casen Survey on the measurement of inequality in Chile. Estudios de Economia **38**(1), 43–65 (2011)
CEDLAS: A guide to SEDLAC, socio-economic database for Latin America and the Caribbean, http://SEDLAC.econo.unlp.edu.ar/eng/methodology.php (2012)
CEPAL: La medicion de los ingresos en la encuesta CASEN 2011-R2, http://observatorio.ministeriodesarrollosocial.gob.cl/layout/doc/casen/La_Medicion_de_los_Ingresos_CASEN_2011.pdf (2012a)
CEPAL: Elementos para una metodología de medición del sector informal en las cuentasNacionales, Estudios estadisticos y prospectivos, N° 77, Santiago, Chile (2012b)
Deaton, A.: Counting the world's poor; problems and possible solutions. World Bank Res. Obs. **16**(2), 125–147 (2001)
Deaton, A.: How to monitor poverty for the Millenium Development Goals. J. Hum. Dev. **4**(3), 353–378 (2003)
Deaton, A.: Measuring poverty in a growing world (or measuring growth in a poor world). Rev. Econ. Stat. **87**(1), 15 (2005)
Feres, J.C.: La medicion de los ingresos e la encuesta Casen –1996. Cepal, Santiago (1997)
Feres, J.C.: Confiabilidad de la medicion del ingreso en las encuestas de hogares, in MECOVI, Imputacion de datos en las encuestas de hogares: los procedimientos metodologicos y sus implicaciones, Taller 14, Buenos Aires, 385–402 (2004)
Ferreira, F., Leite, P., Litchfield, J.: The rise and fall of Brazilian inequality: 1981–2004, Macroeconomic Dynamics, 12(S2), pp. 199–230 (2008)
Foster, J., Shorrocks, A.: Poverty orderings and welfare dominance. Soc. Choice Welf. **5**, 179–98 (1988)
ILO: Measuring informality: a statistical manual on the informal sector and informal employment. International Labor Office, Geneva (2013)
Little, R.J.A., Rubin, D.B.: *Statistical analysis with missing data*. John Wiley, New York (1987)
Lustig, N., Lopez-Calva, L.F., Ortiz-Juarez, E.: Declining inequality in Latin America in the 2000s: the cases of Argentina, Brazil, and Mexico. World Dev. **44**(C), 129–141 (2013)
Malpezzi, S., Mayo, K.: The demand for housing in developing countries: empirical estimates from household data. Econ. Dev. Cult. Chang. **35**(4), 687–721 (1987)
Medina, F., Galvan, C.: Descomposición del coeficiente de Gini por fuentes de ingreso: Evidencia empírica para América Latina 1999-2005, Estudios estadisticos y prospectivos, N° 63. Cepal, Santiago (2008)
Ravallion, M.: Measuring aggregate welfare in developing countries: how well do National Accounts and surveys agree? Rev. Econ. Stat. **85**(3), 645–52 (2003)

Ravallion, M.: The debate on globalization, poverty and inequality: why measurement matters. Int. Aff. **79**(4), 739–753 (2001)
Szekely, M., Lustig, N., Cumpa, M., Mejia, J.-A.: Do we know how much poverty there is? Oxf. Dev. Stud. **32**(4), 523–558 (2004)
World Bank: Measuring the real size of the world economy: the framework, methodology, and Results of the International Comparison Program (ICP). Washington (2013)