



INEQUALITY IN SOUTH AFRICA AND BRAZIL Can we trust the numbers?

Murray Leibbrandt is a Professor in the School of Economics at the University of Cape Town and the Director of the Southern Africa Labour and Development Research Unit. He holds the DSD/NRF National Research Chair of Poverty and Inequality Research. His research focuses on South African poverty, inequality and labour market dynamics. He is currently one of the Principal Investigators on the National Income Dynamics Study. He has been a Visiting fellow at Cornell University and was Visiting Professor at the University of Michigan. He holds a PhD from the University of Notre Dame in the USA.

Arden Finn has a Master's degree in Economic Development from Oxford University. Following this he returned to UCT in order to work within SALDRU, first within the National Income Dynamics Study's survey operation and now as a researcher focusing on the measurement and analysis of poverty and inequality, income mobility and microeconometrics using panel data.

Introduction

Much academic and policy-related research over the past decade has focused on one of two stylised facts: Brazilian inequality has decreased sharply over the last 10 years and South Africa's level of inequality has remained stubbornly high. Nieri (2010) concludes the Brazilian Gini coefficient fell from 0.604 in 1993 to 0.55 in 2008. Barros *et al* (2010) back this up by concluding that the Gini fell from 0.60 in 1995 to 0.552 in 2007. Both of these studies use household income per capita in their calculations. On the South African side Leibbrandt *et al* (2010) calculate Gini coefficients using per capita income to show that South Africa's income inequality increased from 0.66 in 1993 to 0.70 in 2008.¹ Bhorat and van der Westhuizen (2011), use expenditures rather than income as their measure but show a remarkably similar increase in inequality from 0.64 in 1995 to 0.69 in 2005. Thus the internal picture within both countries is pretty consistent. This is important because, irrespective of the comparability of the measures of inequality in Brazil and South Africa, the consistency of the

¹ Household per capita income gives a much more accurate idea of individual welfare within the household, compared to an overall measure of aggregated household income. Not taking household size effects into account can lead to very misleading conclusions.

results suggests that the general argument is correct: Brazil has become less unequal than it was 15 years ago, while South Africa has not. But is Brazil more or less unequal than South Africa? To answer this question, we must establish how comparable are the methods used in calculating the Gini coefficients of each country.

The purpose of this note is to address this issue – are these regularly cited Gini coefficients of Brazil and South Africa generally comparable? The note begins with an explanation of how the Gini coefficient is calculated. Next, it discusses how the collection and measurement of income data may influence inequality measures. The key question of whether there are significant methodological differences in the calculation of inequality in South Africa and Brazil is tackled, before various adjusted measures are provided.

In order to make a fair comparison, we make use of nationally representative datasets from 2008 for both countries – the National Income Dynamics Study (NIDS) for South Africa, and the Pesquisa Nacional por Amostra de Domicílios (PNAD) for Brazil. We focus on the calculation of the Gini coefficient from income, rather than consumption, data. Whereas recent studies of inequality in South Africa use both income and expenditure data in order to check for internal consistency, as illustrated by the two studies cited above, studies of inequality in Brazil tend to use income data. Indeed, the PNAD has income data but not expenditure data and therefore, for a comparison of South Africa and Brazil we are limited to inequality as measured by household per capita income.

How is the Gini coefficient calculated?

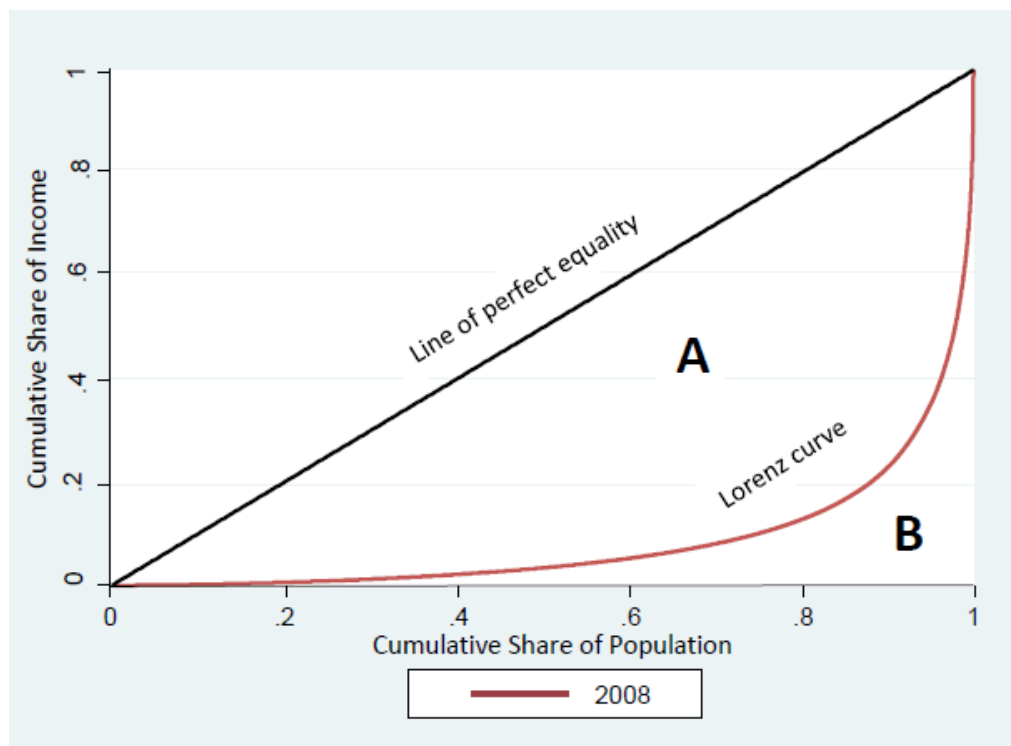
There are a number of methods that can be used to calculate the Gini coefficient. Perhaps the most popular of these involves deriving the coefficient from the Lorenz curve.

The Lorenz curve is a graphical representation of how the cumulative share of income (on the y-axis) is related to the cumulative share of the population (on the x-axis). A 45-degree line represents perfect equality in society – the bottom 20% of people have 20% of total income, the bottom 65% of people have 65% of total income, and so forth. The further away the Lorenz curve is from the line of perfect equality, the greater the level of inequality. Figure 1, below, presents the Lorenz curve for South Africa in 2008.

The Gini coefficient is calculated by taking the ratio of the area between the line of perfect equality and the Lorenz curve (A in Figure 1) to the total area below the line of perfect equality (A + B in Figure 1).² The measure always lies in the interval [0, 1], although it is sometimes multiplied by 100 when it is reported.

²For a Lorenz curve $Y=L(X)$, the Gini coefficient is given by

Figure 1: Lorenz Curve for South Africa



Source: NIDS 2008. Own Calculations

Measuring income and the potential ramifications for the Gini coefficient

Income, whether at the individual or household level, is notoriously difficult to measure accurately. Individuals answering questions about income may intentionally provide erroneous information if, for example, they are suspicious that records of their income will not be kept confidential, or if other people are within earshot of the interview.

Unintentional errors may also bias measured income. Many surveys ask respondents how much income was received from source *x* in the last 30 days or in the last 12 months, and precise amounts are often difficult to recall. It is also often the case that total household income is measured by asking a single person about every household member's income, which he or she may not be able to report accurately.

Methodological differences between surveys are usually the main stumbling block when it comes to directly comparing inequality measures. Labour market income may be gross or net of taxes, transfers from government may be included or excluded, income data may be at the household or individual level, may be monthly or annual (or both), and household income may be measured by a single (one-shot) question or as the aggregation of many different sources.

All of these factors may have a significant effect on the Gini coefficient. Under the mild assumption of a progressive income tax system, if income data are gross rather than net of taxes, the Gini coefficient will be higher. This is because the wealthy forgo a higher proportion of their salaries than the poor, thus compressing the distribution of income and making the net-of-taxes Gini coefficient smaller than its gross counterpart. Where tax systems rely heavily on more regressive taxes – such as sales taxes or VAT – gross income may be more equally distributed than net income.

In addition, the question of whether income is measured pre or post government welfare transfers is important. Using a measure of household income that excludes government transfers will generally lead to an overestimation of inequality, as many government programmes target the poor. In fact, the difference in household income before and after government transfers has been used to evaluate the redistributive effectiveness of the state (for examples see Goni *et al* (2008) and Leibbrandt *et al* (2011)).

There are no *a priori* predictions as to how the Gini coefficient differs according to whether income is measured as a one-shot question (how much income did this household receive from all sources in the last month) or as an aggregation of different sources from different individuals (wages, government transfers, remittances, capital gains).

A common problem faced by surveys everywhere is what to do with missing values. For example, a respondent may claim to earn a wage, but refuse to tell the interviewer what that wage is. This person's wage needs to be imputed in order to limit the amount of bias in the data. Imputation strategies may be very simple or relatively complex, and may add another source of non-comparability to income inequality comparisons. Discussion of imputation methods used in survey analysis is beyond the scope of this note. However, if a large percentage of the income data is missing, it becomes important to ascertain what was done about this in the public release data and how this was done.

In general, surveys try to estimate a measure of the welfare of a household and, if carefully measured, income is a good proxy of this welfare measure. To this end, some aggregated household income measures include imputed rental income to capture the stream of benefits that that households derive from occupying dwellings that they own. While this is not income per se (nor is it easily measured), it is important to ascertain whether or not imputed rental income is included when comparing different surveys.

Finally, some kinds of in-kind income – particularly the receipt of free public services – can affect the welfare of a household even if the contribution the household's welfare is not recorded as income. Neither the NIDS nor the PNAD includes a measure of this kind. There is, therefore, no bias in the two measures of inequality. If, however, there are meaningful differences in the extent to which the Brazilian and South African states redistribute income in this manner, comparisons of the Gini coefficients for income may misrepresent the extent to which household welfare is unequally distributed. In this regard, South Africa's state is generally thought to be more effective in delivering non-cash services to the poor than states in Latin America, and some fiscal incidence studies in Brazil

suggest that relatively little income is redistributed in this fashion. A comprehensive assessment of this is beyond the scope of this paper.

This list of factors represents areas that could confound cross-national income inequality comparisons. They provide useful context for a discussion of how income inequality calculations are actually undertaken in Brazil and South Africa. Despite the regularity with which the inequality calculations are undertaken, are cross country comparisons actually comparable? If not, can they be made comparable or, at least, can we ascertain whether measured differences are too high or too low?

Comparing the Gini coefficients in South Africa and Brazil

As mentioned in the introduction, this note makes use of two datasets that have been cited widely and have formed the basis for academic analysis and policy-making. The NIDS and PNAD datasets are broadly comparable, as both are nationally representative and comprehensively record all major sources of household income. The sample size of the surveys is 7 302 and 118 138 households (28 250 and 391 867 individuals) respectively.

As far as defining a household is concerned, in NIDS 2008 an individual is considered to be a resident household member (and therefore eligible for interview) if he or she usually spends four or more night per week in the dwelling unit, and shares food and resources from a common pool with other household members. In PNAD 2008 a household is defined as the person or people related either by kinship, dependence or norms of common living that share a dwelling. Different families can occupy the same dwelling unit, but each family is counted as a separate household.

Table 1, below, presents 11 areas where these surveys may diverge. Each point is discussed in further detail below.

Table 1: Potentially Confounding Factors in NIDS and PNAD

	South Africa	Brazil
Wage data	Net	Gross
Government transfers	Post	Post
Sporadic income	No	No
Non-cash income	Yes	Yes
Time frame	Monthly	Monthly
One-shot or aggregation	Aggregation	Aggregation
Imputed rent	Yes	No
Weights	Yes	Yes
Representivity	National	National
Imputation	Yes	No
Survey type	Panel	Cross-section

Wage data

Perhaps the most important difference between NIDS and PNAD 2008 is the fact that the former includes a post-tax measure of labour market income while the latter includes a pre-tax measure. This means that a naïve comparison between the two would result in Brazil's Gini being upwardly biased compared to South Africa's. A shortcoming of the Brazilian data is the fact that there are no questions in the survey about taxes, and this makes deriving a precise measure of net labour market income very difficult. However, the South African data contains questions about gross and net wages, which means that a like-for-like comparison is possible. Users should note that the household income variable in the NIDS data contains net wages as one of its components, and this should be replaced by gross wages before any comparison is made with Brazil. Replacing the net wage with its gross counterpart increases the South African Gini coefficient by about 1.75%, from 0.686 to 0.698.³

Government transfers

The household income variables of both surveys contain comprehensive measures of government (cash) transfers. For South Africa this includes the state pension, unemployment insurance, workmen's compensation, disability grant, war veteran's pension, child support grant, foster care grant and care dependency grant. For Brazil, all government transfers (conditional and unconditional) are included as well.

Sporadic income

Neither survey includes income that is received as a one-off, as the aim is to record income that is "usually" received by the household in the last month. Examples of sporadic incomes include inheritances and bridal dowries. The South African data includes income received from a 13th cheque and annual work bonuses, and these amounts are divided by 12 in order to derive a more appropriate monthly amount.

Non-cash income

Both surveys incorporate a variety of non-cash income sources in the composition of aggregate household income. NIDS and PNAD 2008 data capture non-cash income from subsistence farming and from income received "in-kind" (for example, an in-kind transfer from a family member living in

³It is possible to see the effect on the Gini coefficient of grossing up wage data as being relatively small given the progressivity of the South African tax code. The main reason for this is that the share of overall net-of-tax income is dominated by the top two deciles to such an extent, that "grossing up" wage data doesn't bloat the Gini by as much as we expect. That is to say, the top two deciles contain about 75% of overall income anyway, so adding tax payments to these deciles does not dramatically alter the already skew distribution.

another household). As noted, neither the NIDS nor the PNAD data attempts to monetise the value of free government services.

Time frame

Both surveys aim to capture income received in the last month, rather than in the last year. This suggests that recall bias is unlikely to be a major problem or relatively more important in either dataset.

One-shot versus aggregation

Both surveys are consistent when it comes to totalling up household income. In both, the default procedure is to sum each individual's reported income in order to arrive at a more detailed household-level aggregate.

Imputed rent

The household income variable in the NIDS dataset includes imputed rental income for owner-occupied housing. The Brazilian data does not include this measure. Therefore, it is recommended that imputed rent is removed from the NIDS measure of household income before a comparison of Gini coefficients is made. Removing imputed rental income is not unprecedented in a South African context, as there is no standardised way of calculating it.⁴ In fact, a recent comparison of poverty and inequality in South African between 1993 and 2008 (Leibbrandt *et al* 2010) removes imputed rent from all calculations because of this problem. Removing imputed rent from total household income, results in an increase in South Africa's Gini coefficient from 0.686 to 0.700.⁵

Weights

Both surveys contain a set of weights that should be used when calculating inequality measures as they ensure that the sample is reflective of the population of the relevant country.

Imputation

The NIDS dataset contains imputed values for all components of income where there are missing values (Argent 2009). The extent to which the data exhibit missing values varies across variables,

⁴ For example, the PSLSD 1993 survey took an over-arching definition of 7% of the value of the property, while NIDS 2008 used regression-based techniques that controlled for a wide variety of household characteristics.

⁵ The reason why inequality increases when imputed rent is removed from incomes in the South African data is that imputed rent received from RDP housing makes up a relatively large proportion of the incomes of poor households.

although the imputation procedure is consistent. However, no imputations were carried out on the PNAD dataset. It is not known whether the lack of imputation biases Brazil's Gini coefficient and, even if there is some bias, we have no prior assumption as to its direction. If we assume that missing data was not a pervasive problem in the Brazilian data, it should be acceptable to calculate the Gini coefficient on the data as they stand.

The South African data has a far greater missing data problem than Brazil's, where the response rate was over 97%. Imputing for these missing values has the effect of decreasing the Gini coefficient from 0.724 to 0.686. Despite this evidence of missing data biasing South Africa's Gini coefficient upwards, it is not possible to say that the direction of the bias will be the same in Brazil. However, given that accuracy of the South African Gini coefficient is increased by imputing missing data, our calculations for South Africa are based on the imputed data.

Survey type

Both NIDS and PNAD are nationally representative surveys. NIDS is a panel (longitudinal) survey that tracks and re-interviews individuals over time. PNAD is a cross-sectional survey that interviews a sample of the population at one point in time only. This does not affect the comparability of the surveys in any way as the 2008 NIDS Wave is the base wave of a panel study and, as such, was designed to be a nationally representative.

Towards a more comparable measure of inequality

Although there are some significant differences between the South African and Brazilian datasets, it is possible to adapt NIDS 2008 in order to closely align comparisons between the two countries. Table 2, below, summarises some of the findings previously discussed, and provides a Gini coefficient comparison that takes these into account. It should be noted that all changes are made to make the South African data fit better with the Brazilian, and not the other way round. In an ideal state, the PNAD 2008 data would include net wages, imputed rent and would also impute for missing values. However, as this is not the case, the South African data is adjusted to take this into account.

Table 2: Comparing Gini Coefficients

	South Africa	Brazil	Difference
Naïve	0.686	0.550	0.136
Gross vs. Gross	0.698	0.550	0.148
Without Imputed Rent	0.700	0.550	0.150
Final	0.712	0.550	0.162

Source: NIDS 2008 for South Africa (own calculations) and Nieri (2010) for Brazil.

As changes are made to the South African data, so the Gini coefficient increases, as does the difference in inequality between the two countries. A naïve comparison of the "raw" data, yields a difference of

0.136. This increases to 0.162 when gross wages are added and imputed rent is removed from the South African data.

Conclusion

The South African NIDS and the Brazilian PNAD of 2008 have both been widely used to quantify the level of inequality present in each society. Because the approaches to calculating the two countries' Gini coefficients are reliable and have been used consistently, the statistical evidence suggests that levels of inequality in Brazil have been falling but has been broadly unchanged in South Africa. Comparison of levels of inequality in the two countries, however, can only be done if the datasets used in calculating the Gini coefficients are comparable. Broadly speaking, they are, and, while some differences exist, enough data are available to adjust for these.

Both datasets are nationally representative and include comprehensive measure of total household monthly income. Areas covered include income from the labour market, government transfers, remittances and income of a capital nature. A naïve comparison of the Gini coefficients based on this measure of household income ignores a number of potentially confounding factors.

By far the biggest difference between the two datasets is that the South African data is reported net of taxes, while the Brazilian data is gross of taxes. As noted above, however, this obstacle may be overcome by "grossing up" the South African data before any comparison is made. Other differences are the facts that imputation for missing values is not carried out in PNAD and imputed rental income for owner-occupied housing is probably not included either. It is unlikely that the former is of major concern, while the latter may be dealt with by simply adding the relevant amount from each household in the South African data. Taking note of these facts will enable the researcher to perform a like-with-like comparison between the Gini coefficients of both countries, while preserving the integrity of the data.

One outstanding question relates to the extent to which public policy redistributes income through the delivery of free public services, such as health and education, and whether there are significant differences between the extent to which the South African and Brazilian states achieve this. While fiscal incidence studies suggest that the South African state does use this technique to redistribute income more effectively than Brazil, calculating the extent of the difference is beyond the scope of this paper.

References

Argent, J., (2009). "Household Income: Report on NIDS Wave 1". NIDS Technical Paper. Southern

Africa Labour and Development Research Unit, University of Cape Town.

Barros, R., De Carvalho, M., Franco, S. and Mendonca, R., (2010). "Markets the State and the Dynamics of Inequality in Brazil". Chapter 6 in Lopez-Calva, L. and Lustig, N. Eds. *Declining Inequality in Latin America: A Decade in Progress?* United Nations Development Programme, New York.

Bhorat, H. and van der Westhuizen, C., (2011). "Pro-Poor Growth and Social Protection in South Africa: Exploring the Interactions". Unpublished paper, DPRU, School of Economics, University of Cape Town. May.

Brazilian Geographical and Statistical Institute (IBGE), (2008) "Pesquisa Nacional por Amostra de Domicilios". Vol. 29.

Goñi, E., Lopez, J.H., and Serven, L., (2008). "Fiscal Redistribution and Income Inequality in Latin America." World Bank Policy Research Working Paper No. 4487.

Leibbrandt, M., Woolard, I., Finn, A. and Argent, J., (2010). "Trends in South African income distribution and poverty since the fall of Apartheid" OECD Social, Employment and Migration Working Papers No. 101, January.

Leibbrandt, M., Wegner, E and Finn, A., (2011). "Policies for Reducing Inequality and Poverty in South Africa". Unpublished paper, SALDRU, School of Economics, University of Cape Town. May.

Neri, M., (2010). "The decade of falling income inequality and formal employment generation in Brazil". Chapter 2 in OECD 2010. *Tackling Inequalities in Brazil, India, China and South Africa: the Role of Labour Market and Social Policies*. OECD Publishing.

Published in July 2012 by The Centre for Development and Enterprise
5 Eton Road, Parktown, Johannesburg 2193, South Africa
P O Box 1936, Johannesburg 2000, South Africa
Tel +27 11 482 5140 • Fax +27 11 482 5089 • info@cde.org.za • www.cde.org.za

BOARD

L Dippenaar (chairman), A Bernstein (executive director), A Ball, E Bradley, C Coovadia, M Cutifani, B Figaji, F Hoosain,
M Le Roux, S Maseko, I Mkhabela, M Msimang, W Nkuhlu, S Pityana, S Ridley, A Sangqu, E van As

INTERNATIONAL ASSOCIATE

Professor Peter L Berger

© The Centre for Development and Enterprise

All rights reserved. This publication may not be reproduced, stored, or transmitted without the express permission of the publisher. It may be quoted, and short extracts used, provided the source is fully acknowledged.

This CDE Insight was funded by Business Leadership South Africa.
The research was conducted, and the note was written, by Professor Murray Leibbrandt and Arden Finn of the Department of Economics at UCT.