Draft
April, 2004

Poverty and Inequality in Brazil: New Estimates from Combined PPV-PNAD Data[1]

Chris Elbers (*Vrije* Universiteit, Amsterdam)
Jean Olson Lanjouw (University of California, Berkeley, and Brookings Institution)
Peter Lanjouw (World Bank)
Phillippe George Leite (DELTA, Paris)

1. Introduction

Inequality and poverty occupy a prominent place in debates surrounding the recent development experience of Brazil, its future prospects and available policy options. There is an extensive literature on the distribution of well-being in Brazil - describing levels and dynamics of poverty and inequality outcomes; scrutinizing regional and sectoral disparities; studying the links to labour markets, human capital outcomes, public spending patterns; and so on.[2] An important stylized fact that emerges from this body of research is that, compared to other countries, Brazil is a clear outlier in terms of inequality and also accounts for a dominant share of the total number of poor in Latin America.

[1] We are grateful to Francois Bourguignon, Francisco Ferreira, Pedro Luis do Nascimento Silva, Ricardo Paes de Barrros, and Martin Ravallion for useful discussions. Financial support was gratefully received from to the Bank-Netherlands Partnership program and the World Bank PREM Inequality Thematic Group. The views presented in this paper are those of the authors only and should not be taken to reflect the views of the World Bank or any affiliated institution.

Conclusions regarding measured poverty and inequality levels, and trends, depend crucially on the underlying empirical foundations that support such analysis. Almost all of what is known about the distribution of economic welfare in Brazil, at the level of the country as a whole, comes from the well-known PNAD (*Pesquisa Nacional por Amostra de Domicílios*) household surveys. These are large surveys, fielded on an annual basis since the late 1960s, covering virtually all of Brazil (except the sparsely populated north of the country). The PNAD survey permits the construction of a measure of household income, and this indicator of economic welfare underpins much of the subsequent analysis of well-being that has drawn on PNAD data.

Figure 1 compares inequality in Brazil, measured on the basis of the Gini coefficient of inequality and the per capita measure of income from the PNAD, against that in a selection of other countries for which Gini coefficients have been calculated using an income measure of welfare.[3] The stylized fact of a particularly high level of inequality in Brazil can be readily observed. The Gini for Brazil in 1996-7 is calculated to take a value of about 0.6. If we confine our attention to only the Northeast and Southeast of Brazil (see below) the Gini takes the value of 0.61. The only other countries in our sample that record a level of inequality as high as Brazil are Bolivia and Swaziland. Inequality in Latin America tends to be fairly high, relative to other countries in the sample. In the United States, for example, the Gini coefficient takes a value of 0.41.

A recent study by Ferreira, Lanjouw and Neri (2000) suggests that there are at least some reasons for concern regarding the welfare indicator available in the PNAD surveys. Because the survey is essentially an earnings survey, it is oriented towards formal sector

---

[2] The literature is very large. Useful recent contributions include Camargo and Ferreira (1999), Ferreira and Litchfield (1996), Ferreira and Paes de Barros (1999) and World Bank (2001a, 2001b**, 2002?**).

employment. As a result income data from households engaged in self-employment activities are only cursorily collected. These problems may result in inaccurate measures of income from two groups of particular importance in distributional analyses: self-employed informal sector and cultivating households. The question thus arises whether the limitations of the PNAD income figures are driving some of the conclusions about welfare outcomes in Brazil – with respect to both levels and patterns across population subgroups.

In 1996 a pilot household survey was fielded in Brazil's northeast and southeast regions. This survey, known as the *Pesquisa sobre Padrões de Vida* (PPV), is modelled after the World Bank Living Standard Measurement Survey (LSMS). The PPV is a multi-module integrated survey which collects, in addition to information on incomes, data on household consumption. Fairly detailed information on consumption expenditures are collected and it is also possible to impute values of consumption streams from items such as housing and home-produced food products.

In addition to being more detailed and comprehensive in construction, a consumption measure is generally perceived to provide a more reliable indicator of economic well-being than income, even when there are no clear biases in the income measure. This is particularly so when the purpose of the analysis is to study poverty. Essentially it is argued that it is easier to collect reliable and reasonably complete information on consumption than on income. In addition, there are theoretical arguments in favor of

[3] The data for Brazil refer to the years 1996-7, while those for the other countries refer to different reference years in the 1990s. These statistics for other countries have been taken from the World Bank's World Development Report, 2000.

using consumption as a measure of welfare because consumption is thought to better proxy long term living standards than current income.[4]

Despite this surface appeal, the PPV also has clear limitations. Compared to the PNAD survey, the PPV sample is tiny. In addition, the PPV is not designed to be representative of the country as a whole – its coverage extends only to the northeast and southeast of Brazil (covering roughly three quarters of the national population).

The purpose of this paper is to report on the results of an attempt to exploit the best features of the two datasets described above, so as to produce consumption-based estimates of poverty and inequality, but in the large PNAD sample. We employ a recently developed methodology which permits the analyst to impute a welfare indicator from one survey, the PPV in our case, into another survey, the PNAD.

Imputing consumption into the much larger PNAD survey allows us then to estimate summary measures of poverty and inequality at levels of regional disaggregation significantly lower than what would have been possible in the PPV survey. An additional feature of the methodology is that we are able to assess the statistical precision of the welfare measures we estimate. One of our concerns in this paper will be to determine whether the imputation methodology comes at an unacceptably high price in terms of statistical significance of the estimates.

Aside from demonstrating the feasibility of imputing consumption from the PPV into the PNAD survey, we also aim in this paper to ask whether the picture of poverty and inequality that derives via this approach differs significantly from that which obtains from standard analysis based on the conventional PNAD income measure. In that sense

---

[4] See Ravallion (1994) for further discussion.

we are interested to use our approach as a means to gauge the robustness of the conventional picture of poverty and inequality in Brazil.

It is clear that we cannot directly assess the reliability of the PNAD income measure by simply comparing the conventional results against those we obtain following our approach: the welfare *concepts* of income and consumption are different and could not be expected to yield identical quantitative estimates of poverty and inequality. However it may perhaps be arguable that if the two welfare measures were sound, their *qualitative* implications, in terms of the profile of poverty they yield, would be broadly similar. To that end we compare the spatial profile of poverty and inequality across these two approaches on the basis of simple cross tabulations and decompositions.

Figure 2 foreshadows one of the main findings from this analysis. When inequality is measured on the basis of the Gini coefficient and imputed consumption in the PNAD, Brazil is less obviously an outlier as compared to other countries for which inequality has been measured on the basis of a consumption definition of welfare. Inequality in Brazil as a whole is measured at around 0.45. If we confine our attention to the Northeast and Southeast only, inequality is a bit higher – 0.46. Inequality in the Northeast and Southeast, measured directly with the PPV household survey data, is a bit higher still – 0.49. All three of these measures are significantly below those observed in a sizeable number of other countries.

The structure of the remainder of this paper is as follows. In the next section we describe in greater detail the two data sources we draw on in this analysis. Section III turns to an overview of the methodology we employ. Section IV describes how we implement the methodology in this particular setting. Section V presents results at the level of the PPV's representative region. We ask whether our estimates of poverty and inequality accord with those of the PPV, we assess the statistical precision of our estimates, and we compare our consumption based estimates against the estimates that would obtain had we used the income measure that is conventionally analyzed using the PNAD survey. Section VI produces further consumption-based estimates, at levels of disaggregation which the PPV survey could not support. Once again we compare findings against the PNAD income concept in order to develop a further sense of whether, and where, the two approaches part company. We next compare the occupational profile of poverty based on the PNAD income concept against that from the PNAD consumption-based estimates. In Section VII we report some basic inequality decomposition results, again in turn based on the PNAD income measure and the PNAD imputed consumption measure. We ask whether qualitative conclusions regarding the relative contribution to overall inequality from certain population subgroups are robust to the welfare measure that is employed. Section VIII summarizes our findings and discusses directions for future enquiry.

II. Data

As described above, the analysis in this paper draws on two sources of household survey of data: the combined 1996 and the 1997 rounds of the *Pesquisa Nacional por Amostra de Domicílios* (PNAD); and the *Pesquisa sobre Padrões de Vida* (PPV) of 1996.

The PNAD, implemented by the national statistical organization IBGE (*Instituto Brasileiro de Geografia e Estatística*), has been the main staple of country-wide (as opposed to metropolitan) distributional analysis in Brazil since the mid-1970s. It is an annual survey covering both urban and rural areas (except in the Northern region), and is representative at the level of the state and all metropolitan areas. Its sample size, currently around 105,000 dwellings per survey-round, is generally viewed as ample to produce reliable estimates of poverty or inequality at the regional, state, or possibly even lower, level. However, for such a large survey, and one which is fielded so often, some of the PNAD questionnaire shortcomings are remarkable. The questionnaire has evolved a great deal between the mid-1970s and 1996, generally much for the better. Nevertheless, there is one aspect, crucial for poverty and income distribution analysis, which has remained rather problematic: the income questions for any income source other than wage employment are insufficiently disaggregated and detailed.[5]

In principle, the nonsampling errors likely to arise from the absence of these more detailed questions could bias income measurement in either direction. Too few questions about in-kind benefits or the values of different types of production for own consumption are likely to lead to an underestimate of welfare, through forgetfulness. On the other hand, the absence of detailed questions about expenditure on inputs is likely to lead to an overestimate of net incomes from home production. In practice, the international evidence suggests that the first effect often predominates, and the absence of such detailed questions can lead to income under-reporting by categories of workers which, as it happens, are quite likely to be poor. Ferreira, Lanjouw and Neri (2000) examine these issues for the case of Brazil in some detail and suggest not only that under-reporting of

[5] For example, careful scrutiny of the PNAD survey instrument reveals that if a respondent happens to be self-employed in the informal sector (a small farmer in rural areas, for example), then only one *single* question regarding the respondent's income is posed in the PNAD questionnaire – namely what was his *net income in the past month*. The data issues addressed in this section are more thoroughly discussed in Ferreira, Lanjouw and Neri (2000).

income in the PNAD may well be significant, but that the degree is likely to vary considerably significantly across population subgroups.

As mentioned earlier, our second data source, the PPV, is a household survey modeled on the Living Standard Measurement Survey. It was fielded in 1996-97 by IBGE to assess the poverty targeting of Government social spending in Brazil. The aim of the PPV was to supplement the information already available through the PNAD, in order to improve the data available for poverty monitoring and policy analysis in Brazil.

The PPV was designed to fill some of the data gaps left by the PNAD. It provides a much more detailed picture of household expenditures and consumption, as well as utilization of various publicly subsidized services, particularly education, health, and transportation. The questionnaire is much longer, and requires multiple visits to each household. This richer information comes at a price. To keep survey expenses within reason, the sample size is much smaller (just under 5000 households in total) and the survey only covers the two most populous of Brazil's five regions, the Northeast and Southeast. These two regions together account for 73% of Brazil's population. The PPV is representative for ten spatial units (the metropolitan areas of São Paulo, Rio de Janeiro, Belo Horizonte, Salvador, Recife, and Fortaleza; the non-metropolitan urban Northeast; the non-metropolitan urban Southeast; the rural Northeast; and the rural Southeast). However, as we shall see below, even at the representative region level, estimates of poverty and other welfare indicators may be rather imprecisely estimated.

The purpose of this paper is to report on the application of a technique to combine the PPV and the PNAD datasets, seeking to complement their respective strengths and to compensate for their weaknesses. Because a maintained hypothesis of the imputation is that the consumption models estimated on the PPV data apply to PNAD households, it is most tenable to implement our method with reference to the Northeast and Southeast of Brazil only (that part of the country which the PPV is representative of). Results presented in this paper, including those based on the PNAD, thus pertain only to these two regions. We comment briefly, in Section VIII, on the feasibility of extending the analysis to other regions of Brazil.

A final word on the data concerns the comparability of the PPV and the PNAD surveys. It is imperative for the successful implementation of our methodology, that the two data sources we draw on be closely comparable. The methodologies underlying sampling, data collection methods, questionnaire design, etc., across these two datasets are quite different. Nonetheless, Soares de Freitas, *et al* (1997) find little evidence, in a comparison of the PPV with the 1995 PNAD survey, that these basic methodological differences introduce major discrepancies across the two data sources in terms of population characteristics.[6] The PPV survey was fielded during a period of one year spanning 1996 and 1997. The annual PNAD surveys are fielded on or around a given date, usually in September, in their respective survey years. In a further attempt to ensure that the two data sources we work with are as comparable as possible, and also in order to maximize the sample size of the database into which consumption is imputed, we have merged the 1996 and 1997 rounds of the PNAD on the grounds that these two neatly bridge the period covered by the PPV survey. Although the geographic coverage of this combined dataset is confined to the Northeast and Southeast of Brazil only, the size of the sample is sufficiently large (around 111,000 households) to permit considerable disaggregation.

III. Methodology

The methodology we implement here has been described in detail in Elbers, Lanjouw and Lanjouw (2002, 2003). The basic idea is straightforward. We estimate poverty and inequality based on a household per-capita measure of consumption expenditure, y_h. A model of y_h is estimated using the PPV survey data, restricting explanatory variables to those that can be linked to households in both sets of data.[7] Then, letting W represent an indicator of poverty or inequality, we estimate the expected level of W given the PNAD-based observable characteristics of the area of interest using parameter estimates from the 'first-stage' model of y. The same approach could be used with other household measures of well-being, such as per-capita expenditure adjusted by equivalence scales, or to estimate inequalities in the distribution of household characteristics other than expenditures, such as assets or income.

Definitions

The basis of the approach is that per-capita household expenditure, y_h, is related to a set of observable characteristics, $\mathbf{x}_h$, that can be linked to households in both the PPV and PNAD sample surveys:[8]

(1) $$\ln y_h = E[\ln y_h \mid \mathbf{x}_h] + u_h.$$

Using a linear approximation to the conditional expectation, we model the observed log per-capita expenditure for household *h* as:

---

[6] Bianchini and Albieri (1998) provide further details on the survey designs of all of Brazil's major household surveys.
[7] Elbers et al (2002, 2003) describe the case when we impute expenditure from a household survey into the population census.
[8] The explanatory variables are observed values and thus need to have the same degree of accuracy in addition to the same definitions across data sources. From the point of view of our methodology it does not matter whether these variables are exogeneous.

(2) $\ln y_h = \mathbf{x}_h b + u_h$,

where b is a vector of k parameters and u_h is a disturbance term satisfying $E[u_h|x_h] = 0$. The vector of disturbances in the population is distributed $\mathbf{u} \sim Á(0,\Sigma)$.

The model in (2) is estimated using the PPV data. We are interested in using these estimates to calculate the welfare of an area or group for which we do not have any, or insufficient, expenditure information. Although the disaggregation may be along any dimension – not necessarily geographic – for convenience we will refer to our target population as a 'UF' (union federação). There are M_v households in UF v, M_v^s households in the PNAD sample from UF v, and household h has m_h family members.

While the unit of observation for expenditure in these data is the household, we are more often interested in poverty and inequality measures based on individuals. Thus we write $W(\mathbf{m}, \mathbf{X}, b, \mathbf{u}_v)$, where $\mathbf{m}$ is a vector of household sizes, $\mathbf{X}$ is a matrix of observable characteristics and $\mathbf{u}$ is a vector of disturbances.

Because the disturbances for households in the target population are always unknown, we consider estimating the expected value of the indicator given the PNAD households' observable characteristics and the model of expenditure in (2).[9] We denote this expectation as

(3) $m_v^s = E[W \mid \mathbf{m}_v^s, \mathbf{X}_v^s, x]$,

where x is the vector of model parameters, including those which describe the distribution of the disturbances, and the superscript 's' indicates that the expectation is conditional on the sample of PNAD households from UF v rather than a census of households.

In constructing an estimator of m_v^s we replace the unknown vector x with consistent estimators, $\hat{x}$, from the first-stage expenditure regression. This yields $\hat{m}_v^s$ = E[W | $\mathbf{m}_v^s$, $\mathbf{X}_v^s$, $\hat{x}$]. This expectation is generally analytically intractable so we use simulation to obtain our estimator, $\tilde{m}_n^s$.

Properties

The difference between $\tilde{m}_n^s$, our estimator of the expected value of W for the UF, and the actual level of welfare for the UF may be written (suppressing the index v):

(4) $$W - \tilde{m}^s = (W - m) + (m - m^s) + (m^s - \hat{m}^s) + (\hat{m}^s - \tilde{m}^s).$$

Thus the prediction error has four components: the first due to the presence of a disturbance term in the first stage model which implies that households' actual expenditures deviate from their expected values (*idiosyncratic error*); the second due to the fact that we are imputing into a sample rather than a census of households (*sampling error*); the third due to variance in the first-stage estimates of the parameters of the expenditure model (*model error*); and the forth due to using an inexact method to compute $\hat{m}^s$ (*computation error*). Elbers, Lanjouw and Lanjouw (2002) provide a detailed description of the properties of the first and last two components of the prediction error.

To summarize, the variance in our estimator due to idiosyncratic error falls approximately proportionately in M_v, the size of the actual population of households in

[9] If the target population includes PPV households then some information is known. As a practical matter we do not use these few pieces of direct information on y.

the UF. In other words, the smaller the target population, the greater is this component of the prediction error, and there is thus a practical limit to the degree of disaggregation possible. At what population size this error becomes unacceptably large depends on the explanatory power of the **x** variables in the expenditure model and, correspondingly, the importance of the remaining idiosyncratic component of the expenditure.

We calculate *sampling errors* on our poverty estimates taking into account the fact that the PNAD surveys are complex samples which involve stratification and multi-stage clustering (see Howes and Lanjouw, 1998, Deaton, 1997) .

We employ the delta method to calculate the variance due to *model error*:

$V_M \approx \nabla^T V(\hat{x})\nabla$, where $\nabla = [\partial\tilde{\mu}^s / \partial x]|_{\hat{x}}$ and $V(\hat{x})$ is the asymptotic variance covariance matrix of the first-stage parameter estimators. Because this component of the prediction error is determined by the properties of the first-stage estimators, it does not increase or fall systematically as the size of the target population changes. Its magnitude depends, in general, only on the precision of the first-stage coefficients and the sensitivity of the indicator to deviations in household expenditure. For a given UF its magnitude will also depend on the distance of the explanatory variables for households in that UF from the levels of those variables in the sample data.

The variance in our estimator due to *computation error* depends on the method of computation used. As our calculations of the idiosyncratic and models errors are based on simulations, we can make the computation error become as small as desired by choosing a large enough number of simulation draws (at the cost of computational resources and time).

We use Monte Carlo simulations to calculate: $\hat{m}^s$, the expected value of the poverty or inequality measure conditional on the first stage model of expenditure; V_I, the variance in *W* due to the idiosyncratic component of household expenditures; and, for use in determining the model variance, the gradient vector $\nabla = [\partial\tilde{\mu}^s / \partial x]|_{\hat{x}}$.

Let the vector $\hat{\mathbf{u}}^r$ be the *r*th draw from our estimated disturbance distribution – a random draw from an M_v-variate standard normal or *t* distribution, pre-multiplied by a matrix $\boldsymbol{T}$, defined such that $\boldsymbol{TT}^T = \hat{\Sigma}_n$, where $\hat{\Sigma}_n$ is the estimated disturbance covariance matrix for the population of households in UF *v*. With each vector of simulated disturbances we construct a value for the indicator, $\hat{W}_r = W(\mathbf{m}, \mathbf{X}, \hat{x}, \hat{\mathbf{u}}^r)$, where **m** and **X** represent numbers of households and observable characteristics of PNAD households, respectively, each repeated in accordance with its expansion factor so as to have rows equal to the census number of households, M_v. The simulated expected value for the indicator is the mean over *R* replications:

(5) $$\tilde{m}^s = \frac{1}{R}\sum_{r=1}^{R}\hat{W}_r .$$

Having estimated *m* using the population number of households, the variance of *W* around its expected value due to the idiosyncratic component of expenditures can be estimated in a straightforward manner using the same simulated values:

(6) $$\hat{V}_I = \frac{1}{R}\sum_{r=1}^{R}(\hat{W}_r - \tilde{m}^s)^2 .$$

Simulated numerical gradient estimators are constructed as follows: We make a positive perturbation to a parameter estimate, say $\hat{b}_k$, by adding $d\,|\,\hat{b}_k\,|$, and then calculate $\tilde{m}^{s+}$. A negative perturbation of the same size is used to obtain $\tilde{m}^{s-}$. The simulated central distance estimator of the derivative $\partial\tilde{\mu}^{s} / \partial b \,|_{\hat{x}}$ is $(\tilde{m}^{s+} - \tilde{m}^{s-})/(2d \,|\, \hat{b}_k \,|)$. Having thus derived an estimate of the gradient vector, we can calculate $\hat{V}_M = \nabla^T V(\hat{x})\nabla$.

IV. Implementation

The first-stage estimation is carried out using the PPV survey. As described in section II this survey is stratified into ten regions and is intended to be representative at that level. Within each region there are several levels of clustering. At the final level, 8 households are randomly selected from a census enumeration area. Such groups we call a 'cluster' and denote with a subscript c. Expansion factors, l_{ch}, allow the calculation of regional totals.

Our first concern is to develop an accurate empirical model of household consumption. Consider the following model:

$$\ln y_{ch} = E[\ln y_{ch} \,|\, x_{ch}^T] + u_{ch} = x_{ch}^T b + h_c + e_{ch} \tag{9}$$

where h and e are independent of each other and uncorrelated with observables, x_{ch}. This specification allows for an intra-cluster correlation in the disturbances. One expects location to be related to household income and consumption, and it is certainly plausible that some of the effect of location might remain unexplained even with a rich set of regressors. For any given disturbance variance, s_{ch}^2, the greater the fraction due to the common component h_c the less one enjoys the benefits of aggregating over more households within a UF. Welfare estimates become less precise. Further, the greater the part of the disturbance which is common, the lower will be inequality. Thus, failing to take account of spatial correlation in the disturbances would result in underestimated standard errors on welfare estimates, and upward biased estimates of inequality.

Since unexplained location effects reduce the precision of poverty estimates, the first goal is to explain the variation in consumption due to location as far as possible with the choice and construction of x_{ch} variables. We try to tackle this in three ways. First, we estimate different models for each of the ten regions in the PPV. Second, we include in our specification household level indicators of connection to various networked infrastructure services, such as connection to electricity, piped water, telephone. To the extent that all or most households within a given neighborhood or community are likely to enjoy similar levels of access to such infrastructure, these variables might capture unobserved latent location effects. Third, we have merged both the PPV and the PNAD datasets with an independently compiled municipio-level database (BIM) of variables (such as employment rates, school attendance rates, etc.) and also consider these variables as candidate variables for inclusion in our household expenditure models.[10]

We apply a selection criterion when deciding on our final specification requiring a significance level of 5% of all household-level regressors. To select location variables (cluster means and BIM variables), we estimate a regression of the total residuals, $\hat{u}$, on cluster fixed effects. We then regress the cluster fixed-effect parameter estimates on our location variables and select those five that best explain the variation in the cluster fixed-effects estimates. These five location variables are then added to our household level variables in the first-stage regression model.

[10] A municipio represents a higher level of aggregation than the census EA, and as such BIM variables are intended to capture loactional effects at this higher level, rather than the more local cluster-level means.

We apply a Hausman test described in Deaton (1997) to determine whether each regression should be estimated with household weights. In seven out of ten regions we find that weighting has no significant effect on the coefficients, and these first-stage regressions are thus estimated without weights. $\overline{R}^2$'s on our models are generally high, ranging between 0.45 and 0.77.[11]

We next model the variance of the idiosyncratic part of the disturbance, $s^2_{e,ch}$. Note that the total first-stage residual can be decomposed into uncorrelated components as follows:

$$\hat{u} = \hat{u}_{c.} + (\hat{u}_{ch} - \hat{u}_{c.}) = \hat{h}_c + e_{ch} \qquad (10)$$

where a subscript '.' indicates an average over that index. To model heteroskedasticity in the household-specific part of the residual, we choose the twenty variables, $\mathbf{z}_{ch}$, that best explain variation in e^2_{ch} out of all potential explanatory variables, their squares, and interactions.[12] We estimate a logistic model of the variance of e_{ch} conditional on $\mathbf{z}_{ch}$, bounding the prediction between zero and a maximum, *A*, set equal to $(1.05)*\max\{e^2_{ch}\}$:

$$\ln\left[\frac{e^2_{ch}}{A - e^2_{ch}}\right] = z^T_{ch}\hat{a} + r_{ch}. \qquad (11)$$

Letting $\exp\{z^T_{ch}\hat{a}\} = B$ and using the delta method, the model implies a household specific variance estimator for e_{ch} of

$$\hat{s}^2_{e,ch} = \left[\frac{AB}{1+B}\right] + \frac{1}{2}Var(r)\left[\frac{AB(1-B)}{(1+B)^3}\right]. \qquad (12)$$

[11] For reasons of space we do not reproduce here the parameter estimates and full set of diagnostics for all ten regression models. These can be furnished upon request.
[12] We limit the number of explanatory variables to twenty to be cautious about overfitting.

Finally, we check whether η and ε are distributed normally, based on the cluster residuals $\hat{\eta}_c$ and standardized household residuals $e^*_{ch} = \frac{e_{ch}}{\hat{s}_{e,ch}} - [\frac{1}{H}\sum_{ch}\frac{e_{ch}}{\hat{s}_{e,ch}}]$, respectively where H is the number of households in the survey. The second term in e^*_{ch} is not needed when first stage regressions are not weighted. In many cases normality is rejected, although the standard normal does occasionally appear to be the better approximation even if formally rejected. Elsewhere we use t distributions with varying degrees of freedom (usually 5), as the better approximation.

Before proceeding to simulation, the estimated variance-covariance matrix, $\hat{\Sigma}$, is used to obtain GLS estimates of the first-stage parameters, $\hat{b}_{GLS}$, and their variance, Var($\hat{b}_{GLS}$).

V. Poverty and Inequality at the Regional Level

We begin our examination of empirical results at the level of the representative region in the PPV survey. Table 1a reports estimates of the incidence of poverty for the ten representative regions of the PPV. We report poverty estimates based on the three possible combinations of welfare concept and data-source that are available. In the first column we present estimates of the incidence of poverty in the combined 1996-97 PNAD survey based on the PNAD income measure of welfare. Column two provides our calculation of the standard error on this income-based poverty measure. This standard error comprises the *sampling error* described earlier, and our calculations of this have taken into account the complex sample design of the PNAD survey. The second set of poverty estimates and standard errors (columns 3 and 4) are based on the PPV survey and the per capita consumption aggregate that can be constructed from that survey. The standard errors are once again sampling errors incorporating the complex sample design of the PPV. Finally, columns 5 and 6, provide estimates of poverty from the PNAD using the consumption indicator of welfare that we have imputed into the PNAD. The standard errors on these poverty estimates comprise both a *sampling error* as well as the *model error* described in sections III and IV. The *idiosyncratic error* is vanishingly small at the levels of disaggregation that we are concerned with in this paper, and is therefore not reported.[13] *Computational error* has been pushed close to zero by employing at least 100 simulations in all our calculations.

We use the same poverty line to measure the incidence of poverty across all three cases. We employ the poverty line of R$65.07 in 1996 São Paulo reais which was derived in Ferreira, Lanjouw and Neri (2000) as an extreme poverty line, sufficient to permit consumption of a minimum bundle of food items only. Both the income and consumption measures of welfare have been adjusted to capture spatial price variation (see Ferreira et al, 2000).

Considering first a comparison of poverty based on PNAD income versus PPV consumption we are immediately struck by the much higher levels of measured poverty in the PNAD. This point has already been discussed at length by Ferreira, et al (2000) and it is perhaps useful to add only that the differences in measured poverty between PNAD income and PPV consumption are generally statistically significant. It is important to note that even though the PPV is designed to be representative at the level of

[13] Calculation of this component is very computationally intensive as it requires using expansion factors to explode the PNAD sample up to a meta-census level, and then carrying out simulations to estimate the idiosyncratic error on the point estimate of poverty. Elbers et al (2002) document that that idiosyncratic error becomes negligible when welfare estimates are for populations of 10,000 households or more. In no case do we estimate welfare measures for populations below this size. (Note that the criterion is *population* not sample size).

these ten regions, the standard errors on the poverty estimates at this level are generally higher than 10% of the point estimate – indicating that confidence bounds around these point estimates are quite wide.[14] Even so, measured poverty in the PNAD is so much higher than in the PPV that one can generally rule out that they are statistically indistinct.

While levels of poverty across the two surveys and welfare concepts are clearly different, qualitative conclusions between the PNAD-income poverty profile and the PPV-consumption profile, are much more similar. Both approaches find clear evidence that rural poverty in the northeast is highest of all ten regions, followed by poverty in the urban northeast and then the rural southeast. Poverty in the metropolitan areas of the southeast is clearly lowest.

These conclusions are echoed when we return to the PNAD data but base our poverty estimates on consumption imputed according to the method described in sections III and IV. The ranking of poverty across regions is identical to that obtained with the PNAD-income approach, but point estimates of poverty are now quite close to those obtained in the PPV. The method we have employed seems to work in that it provides us with estimates of poverty in the large PNAD dataset that are in close accordance with the PPV survey. An indirect implication is that the two data sources are reasonably good samples of, and are describing, the same underlying population. Without that strict comparability of the surveys we could not have expected to obtain such close agreement between the PPV and the PNAD consumption-based estimates.

Although the standard errors for the imputed-consumption based PNAD profile incorporate several error components that do not affect the PPV estimates, the precision of the PNAD consumption estimates is generally greater than that of the PPV estimates,

[14] For this reason one would be very reluctant to disaggregate the PPV down below this level.

even at this high level of aggregation. This is because the PNAD consumption estimates are calculated over the much larger PNAD dataset and sampling errors are thus commensurately smaller. Although the model errors on the PNAD consumption estimates are not negligible (as reflected in the higher total standard errors for these estimates than for the PNAD income estimates), it appears that they are not so large as to invalidate the exercise.

Table 1b repeats the exercise for the Poverty Gap, rather than the headcount measure. Once again results of the imputation exercise are quite encouraging, although the regions of Metropolitan Recife, Belo Horizonte and Rio de Janeiro, do not match as well as for the headcount in Table 1a.

In Table 2 we turn to a similar examination of the three alternatives, but consider measured inequality. The inequality measure we employ for this purpose is the General Entropy class measure:

$$GE_c = \frac{1}{c(1-c)}\left\{1-\frac{1}{N}\sum_{h\in H_v} m_h \left(\frac{y_h}{\bar{y}}\right)^c\right\}. \quad (13)$$

This class of inequality measures has the attractive feature of being sub-group decomposable, and the choice of *c* allows the analyst to weight changes in inequality differently depending on which segments of the income distribution are affected (see Bourguignon, 1979, Cowell, 1980, and Shorrocks, 1980). We employ here a value for *c* of 0.5 in Table 2a, which corresponds to a fairly high weighting to changes in inequality amongst the lower tail of the distribution, and a value of 1.0 in Table 2b.

In implementing the methodology described above to the measurement of inequality, and important issue arises with respect to the adoption of *trimming* protocols.

We first trim the standardized residuals e^*_{ch} and $\hat{h}_c$ outside of (-3,3) bounds. The rationale for trimming the disturbance draws is that we do not think that the true distribution has support on (-8, +8) and this seems to be a conservative place to put truncation. More importantly for the purpose of inequality measurement, we trim the total estimated $\ln y_{ch}$ in each simulation by throwing out those draws that result in an estimate that exceeds the top 99.5% of the PPV consumption distribution or that fall below the 0.5% of the PPV consumption distribution. We then also compare our PNAD estimate of inequality against the PPV-based estimate for the same trimmed distribution. This approach attempts to deal with problems of mismeasured data as well as of a misspecified model. The impact of this will obviously be more pronounced for those inequality measures that are sensitive to the tails of the distribution. With our methodology the estimates of the tails of the distribution will be largely driven by functional form assumptions. In a small survey such as the PPV one is unlikely to include the tails in the sample.

In Table 2a we can see that, as with poverty measurement, measured inequality is much higher with the PNAD income concept than based on consumption. Consumption based estimates in both the PPV and PNAD are much lower than those in the first column. We are unable to state at this point whether these differences are statistically significant because we have not yet been able to calculate sampling errors on inequality measures which properly take into account the complex design of both the PNAD and PPV surveys. This means also that we are only able to report model errors on the PNAD consumption-based inequality estimates. Future work will address this concern.

There is considerable disagreement between all three data-source/welfare-concept combinations in terms of the relative ranking of inequality across regions. We do note that the range of values of the inequality estimates is more compressed than the values of poverty estimates in Table 1, and that it is therefore quite possible that rankings would not be statistically significant if we had information on sampling errors. We will find below that inequality differences at lower levels of disaggregation are more pronounced.

The PNAD consumption-based estimates are generally of the same order of magnitude as those estimated in the PPV, and are similarly lower than the PNAD income-based estimates. The main outlier in this regard is inequality in the Metropolitan Fortaleza which is estimated at 0.330 with the PNAD consumption criterion and 0.402 with the PPV. Table 2b provides comparable results for the General Entropy measure with c=1.

The broad conclusion on which all three alternatives agree is that inequality in the rural northeast tends to be particularly low compared to urban areas in the northeast. A similar uniform finding is that metropolitan areas in the southeast tend to be more unequal than other urban areas in the southeast, which in turn are more unequal than rural areas in the southeast.

VII. Poverty and Inequality at Lower Levels of Disaggregation

The discussion in the preceding section indicated that our methodology appears to allow us to impute consumption into the PNAD survey and obtain estimates of poverty and inequality that are not out of line with what we would expect (based on analysis at the representative region in the PPV). The next step is to produce PNAD consumption-based estimates at levels of disaggregation that are below what we would be able to produce with the PPV. This step represents an important objective with the whole exercise: to employ a concept of welfare we are more comfortable with in a dataset which offers much more scope for disaggregation than the PPV.

In Table 3 we produce estimates of poverty at the level of the Union Federação, breaking these states up, in turn, into metropolitan, other urban, and rural areas. Once again, in order to compare broad qualitative conclusions, we produce both PNAD income estimates as well as estimates in the PNAD based on our imputed consumption measure. We report not only point estimates of the incidence of poverty, but also caclulate relative contribution to overall poverty. A first point to note is that there is considerable heterogeneity of estimated poverty rates across states, in total as well as within urban and rural areas. The two approaches both clearly identify the Maranhão and Piauí as the two poorest states in total. The high poverty in these two states is attributable to high rates in both rural and urban areas. Our two approaches to measuring poverty also agree that Céara, Alagoas and Bahia are among the next poorest group of four states; although the precise ranking within this group is not the same for the two approaches.

Considering rural areas only, the two approaches reach some common conclusions (on the high poverty in Ceara and Piauí) but they also indicate some clear differences (rural Paraiba is less poor than average in the rural northeast according to the PNAD consumption criterion, but is the third poorest state according to the PNAD income criterion). Considering non-metropolitan urban areas only, Maranhao stands out as most poor according to both criteria.

In terms of poverty contribution, broad conclusions across the PNAD income and imputed consumption criteria are remarkably stable. Although the two welfare criteria indicate very different *levels* of poverty, the composition of the poor across state and sector is quite stable. One exception to this conclusion can be observed with respect to the relative importance of the poor in metropolitan urban areas relative to other urban areas. In Ceara for example, 24% of the state's poor are estimated to reside in Metropolitan Fortaleza according to the PNAD income criterion, but only 17% according to the consumption criterion. Similarly in Pernambuco the income criterion suggests that 29% of the state's poor reside in metropolitan Recife, while the consumption criterion suggests it is only 25%. And again in Bahia, a similar observation holds. The difference is largely made up by the relatively higher poverty contribution observed in non-metropolitan urban areas according to the consumption criterion.

Metropolitan areas are clearly less poor than both other urban and rural areas in the northeast, but this is less markedly the case in the southeast, according to both the income and consumption criteria. In the southeast other urban areas are particularly low in the UF of Sao Paulo. This finding is also common to both criteria.

Table 4 reports UF-level point estimates of inequality from the PNAD based on the income and imputed consumption criteria. Within both the northeast and southeast, both the income and consumption-based approaches find that inequality is generally lower in rural areas than in metropolitan or other urban areas (except Maranhao). The highest levels of inequality are generally observed in metropolitan areas of both the northeast and southeast. Once again, although levels of measured inequality are markedly different, broad qualitative conclusions across the two approaches tend to be broadly similar, with only a few subtle differences.

We briefly report, in Table 5, a further attempt at disaggregation. In this table we break up urban and rural areas further. In urban areas we draw a distinction between actually urbanized settlements, and those which have been delineated as urban but which may are still rather sparsely populated (a proxy for peri-urban areas). In rural areas we draw a distinction between those areas which are rather more remote and dispersed (designated as *Rural*) and those designated as rural but which are in fact somewhat built up, with certain minimal facilities and infrastructure (designated *other Rural*).[15]

Rural poverty is unambiguously highest in dispersed, remote areas. This is also where the bulk of the rural population resides. In those states which have a sizable rural population residing in built-up areas, poverty rates are generally markedly lower in those areas than in the dispersed regions. In states where the built-up rural sector is small, poverty rates are not particularly low. There is considerable variation across states in the distribution of poverty across these locations, with the unambiguous results that metropolitan areas are always least poor, followed by urbanized urban areas. The definition of peri-urban areas we have used does not appear to work terribly well, as it generally represents only a very small fraction of the urban population.

In Table 6 we break the population into 7 mutually exclusive occupation groups, and estimate poverty and inequality for these groups, at the all Brazil level as well as separately for the northeast and southeast. As with the geographic disaggregations, the poverty profile does not change radically as a result of changing welfare definition, even though *levels* of poverty and inequality are quite different. However, some subtle differences can be discerned. The consumption criterion tends to find that the unemployed are a relatively less important sub-group of the poor , than does the income criterion. But the consumption criterion finds that the self-employed in rural areas comprise a relatively larger share of the poor than does income, markedly so in the southeast.

VIII. Inequality Decompositions

We next decompose inequality across different population subgroups, based on the PNAD data and our two different welfare concepts. As mentioned earlier, the General Entropy class of inequality can be readily decomposed into a within-group and between group component. With our parameter c=0.5 our decomposition takes the following form:

$$W_{0.5} = 4\left\{1 - \sum_{j=1}^{J} f_j \left(\frac{y_j}{\overline{y}}\right)^{0.5}\right\} + \sum_{j=1}^{J} w_j f_j \left(\frac{\overline{y}_j}{\overline{y}}\right)^{0.5}, \qquad (14)$$

where N individuals are placed in one of J groups subscripted by j, and the proportion of the population in the jth group, denoted f_j, has weighted mean per-capita expenditure (or income) $\overline{y}_j$ and inequality w_j. The first term in this expression is the inequality between groups and the second is within groups. One can think of the share of the between group inequality to total inequality as the amount of inequality that is due simply to differences

[15] This exercise is essentially intended as a cross-check on results reported in Ferreira and Lanjouw (2001) which implemented a very basic version of the methodology employed here.

in average expenditures between the groups. That portion of inequality that would remain if all differences across individuals within each group were to be eliminated.[16]

Table 7 reports the results from our decomposition exercise. We first take the country as a whole and ask how much of overall inequality is attributable to the between-group component in a series of settings. We observe that if one breaks Brazil down into an urban and rural sector, that only around 13 percent of overall inequality can be attributed to the difference in average consumption or income between these two sectors. Most of inequality would remain if this difference in averages would be removed. The conclusion holds irrespective of the welfare concept that is being used. If the country were broken down into Northeast and Southeast only, then the between component rises slightly to 18% (once again remarkably similar across the two welfare concepts). When the country is divided into four – urban northeast, rural northeast, urban southeast and rural southeast – the between-group component continues to rise slowly. Again, the two approaches give essentially the same result. Turning to the question of whether inequality is largely attributable to differences between metropolitan areas and the rest of the country, we find that only 8-10 percent of inequality is due to the difference in average welfare across these two sectors. Adding a further subgroup, other urban, raises the between group component to 15-17%. At the national level, it is evident that much of overall inequality remains within the groups that have been considered here. An important point to note, given the purpose of this paper, is that the decomposition results at the national level are qualitatively the same whether we use the PNAD income measure or our imputed consumption measure.

[16] One minus this proportion can then be attributed to the share of inequality that is due to heterogeneity within the groups.

When we look at rural areas only, we see that the two approaches to appear to give rather different results. The PNAD income approach suggests that rural inequality would fall by around 13% if the difference in average income between the northeast and southeast were removed. The consumption based approach in the PNAD suggests that the reduction in inequality would be about half that. If differences in average income across all states were removed, the PNAD income approach suggests that inequality would fall by approaximately 18%. The consumption approach suggests that the fall in inequality would be about half as much: 8.9%. The two approaches depart here in a quite significant way, with the consumption based approach suggesting that a much smaller source of overall rural inequality is due to differences across states in average rural incomes. In urban areas, the decomposition of inequality is, again, remarkably similar across the two welfare indicators.

Decomposing inequality across occupation groups finds, again, a fairly stable picture across the welfare criteria. In general, PNAD consumption indicates that the within-occupation inequality accounts for a larger share of total inequality than does PNAD income, but the differences are slight. The same patterns hold at the level of the northeast and southeast respectively, but are somewhat more pronounced in the northeast.

VIII. Conclusions

This paper had two objectives. The first objective has been to demonstrate a methodology to impute a measure of consumption, as defined in the PPV household survey, into the much large PNAD household survey. The purpose of this exercise has been to estimate measures of welfare, such as poverty and inequality, defined in terms of consumption, at levels of disaggregation that are permitted by PNAD dataset. Although the results are still to be finalized we have shown that the methodology works quite well. We are able to validate the exercise at the representative region level in the PPV, and find that at that level, point estimates are very similar across the PPV and the PNAD. We have also shown that standard errors on the consumption-based point estimates in the PNAD are quite reasonable – certainly compared to the standard of typical household surveys.

Our second objective has been to shed some light on the question of whether the analysis of poverty and inequality based on the PNAD income indicator yields different conclusions than an analysis based on consumption. We referred to the concern in the literature on PNAD-based distributional analysis that the income measure in the PNAD might suffer from serious biases.

We have found that poverty and inequality, estimated on the basis of consumption in the PNAD, tend to be much lower than estimates based on the income concept. This is not necessarily an indictment of income based analysis, however, as the two concepts of welfare are different and should not be expected to yield the same quantitative estimates. We demonstrated however, that differences in estimates of poverty and inequality between the PNAD and the PPV are not attributable to non-comparability of these two surveys. Our PNAD consumption-based estimates are very close to those which obtain with the PPV.

We pursued the comparability of income and consumption-based results further by examining whether there are important qualitative differences in the geographic profile of welfare across the two approaches. We found that, in fact, the two reach broadly similar findings. In only a few cases do we note differences across the two approaches that may need to pursued further. First, according to the consumption criterion, there is a clear basis for viewing metropolitan areas in the northeast as less poor than other areas. This distinction is less clear-cut according to the income criterion. Second, within rural areas in the northeast, rural Paraiba is least poor state according to the PNAD consumption criterion, but is found to be the third poorest state according to the PNAD income criterion. Third, the PNAD consumption criterion finds that metropolitan areas in the northeast are markedly more equal than other urban areas in this region. The PNAD income criterion finds the reverse. Fourth, the consumption-based approach reflects much more strongly than the income-based one the contribution of differences in average incomes across states to overall rural inequality.

Looking for differences in qualitative conclusions regarding the spatial distribution of poverty and inequality, may not be the best way to examine whether the income-based PNAD measures introduce important biases into distributional analysis in Brazil. As described in Section II, the PNAD income measure is thought to be inadequately capturing income levels of certain population subgroups, notably those who are engaged in informal sector self-employment activities. A more effective direction to take might thus be to compare consumption-based estimates of poverty and inequality amongst population subgroups defined in terms of occupations and education levels, rather than along geographic lines.[17] Initial exploration along these lines in Section VI appears to suggest that here too, the profile of poverty is fairly robust. But these findings

[17] Note the spatial dimension was a natural one to pursue in this paper given our interest to also validate results against the representative region level estimates in the PPV.

are still tentative, and it seems an important next step to pursue this question in greater detail.

Still within a geographic focus, there would seem to be two promising directions for further work. First, it is important to examine whether the conclusions of Ferreira et al (2000) regarding the distribution of urban poverty across city-size is robust to the application of a consumption-based indicator of welfare. It is possible to link urban households in the PNAD to the size of conurbation in which they reside. Ferreira et al observe a much higher incidence of poverty in smaller towns relative to large cities and metropolitan areas. Second, the results presented here suggest that there may be considerable variation in poverty rates within rural areas. So far we have split rural areas up quite crudely into dispersed and built-up areas. An important additional direction to take would be to divide rural areas into agro-ecological and climatic zones, as well as areas demarcated by differences in access to facilities and infrastructure.

The analysis in this paper has concentrated on the northeast and southeast of Brazil. As a result, some 25% of the population have not been included in the analysis. In principle it would be possible to extend the analysis carried out here, to regions such as the south and the center west of the country. But to do so would require making some important, unverifiable, assumptions. Because there are no PPV data applicable to these regions one would have to select a set of parameter estimates from the PPV data and impose the assumption that they are applicable for these regions which lie outside the PPV sampling domain. One might, for example, assume that the appropriate model to apply to the rural center west region is a first stage model based on the combined rural northeast and southeast sample of the PPV. Similarly one might impose the rural southeast parameter estimates on the rural south PNAD data. This exercise is possible but still pending.

In the medium run there is a potential to apply the methodology reported here to the 2001 population census for Brazil. There are initiatives underway to implement a large new consumption survey in Brazil, covering both rural and urban areas. While the mooted sample size of 50,000 is very large in absolute terms, it is clear that these data will not permit disaggregations of poverty and inequality significantly below the UF level. If this new survey were to serve a basis for estimating first stage consumption models with which to impute consumption into the population census, it would then be possible to measure inequality and poverty, based on a consumption measure of welfare, at the town or village (and possibly neighborhood) level across the entire country. Such initiatives are being actively implemented and/or explored in a number countries in the last few years. Some, such as Mexico, Indonesia and China have total populations that, like Brazil, are very large.

References

Bianchini, Z. M., and Albieri, S. (1998) "A Review of Major Household Sample Survey Designs Used in Brazil", proceedings of the Joint IASS/IAOS Conference, Statistics for Economic and Social Development, September.

Bourguignon, F. (1979) 'Decomposable Income Inequality Measures' *Econometrica*, 47:901:920.

Camargo, J. M., and Ferreira, F.H.G. (1999) "A Poverty Reduction Strategy of the Government of Brazil: A Rapid Appraisal", mimeo, Dept. of Economics, Catholic University of Rio de Janeiro.

Cowell, F. (1980) 'On the Structure of Additive Inequality Measures' *Review of Economic Studies* 47:521-531.

Deaton, A. (1997) The Analysis of Household Surveys: A Microeconometric Approach to Development Policy (World Bank: Johns Hopkins University Press).

Elbers, C., Lanjouw, J.O., and Lanjouw, P. (2002) "Welfare in Villages and Towns: Micro-Level Estimation of Poverty and Inequality", Policy Research Working Paper No. 2911, Development Economics Research Group, the World Bank.

Elbers, C., Lanjouw, J.O. and Lanjouw, P. (2003) 'Micro-level Estimation of Poverty and Inequality' *Econometrica*, 71:355-364.

Ferrerira, F.H.G., Lanjouw, P., and Neri, M. (2000) "A New Poverty Profile For Brazil Using PPV, PNAD and Census Data", Departamento de Economia PUC-Rio, TD #418, March.

Ferreira, F.H.G. and Lanjouw, P. (2001) "Rural Nonfarm Activities and Poverty in the Brazilian Northeast", *World Development* Vol 29, No. 3, pg 509-528.

Ferreira, F.H.G. and Litchfield, J. (1996) "Growing Apart: Inequality and Poverty Trends in Brazil in the 1980s", LSE – STICERD –DARP Discussion Paper No. 23, London (August).

Ferreira, F.H.G. and Paes de Barros, R. (1999) "The Slippery Slope: Explaining Increases in Extreme Poverty in Urban Brazil, 1976-1996", Brazilian Review of Econometrics, 19(2), 1999.

Howes, Stephen and J. O. Lanjouw (1998) "Does Sample Design Matter for Poverty Comparisons?" *Review of Income and Wealth.* Series 44, no. 1. pp. 99-109.

Ravallion, M. (1994) Poverty Comparison (Chur: Harwood Press).

Shorrocks, A. (1980) ‘The Class of Additive Decomposable Inequality Measures’ *Econometrica* 48: 613-625.

Soares de Freitas, M., Nogueira Duarte, R., Carneiro Pessoa, D., Albieri, S. and do Naschimento Silva, P. (1997) ‘Comparando Distribuições Etárias em Pesquisas por Amostragem: PNAD 1995 e PPV 96/97”, mimeo Fundação Instituto Brasileiro de Geografia e Estatística, DPE, DEMET.

World Bank (2001a) “Attacking Brazil’s Poverty: A Poverty Report with a Focus on Urban Poverty Reduction Policies”, World Bank, Washington D.C..

World Bank (2001b) “Brazil: Rural Poverty Report”, World Bank, Washington D.C.

Figure 1
Inequality in Brazil Compared to the World: PNAD Per Capita Income Concept of Welfare

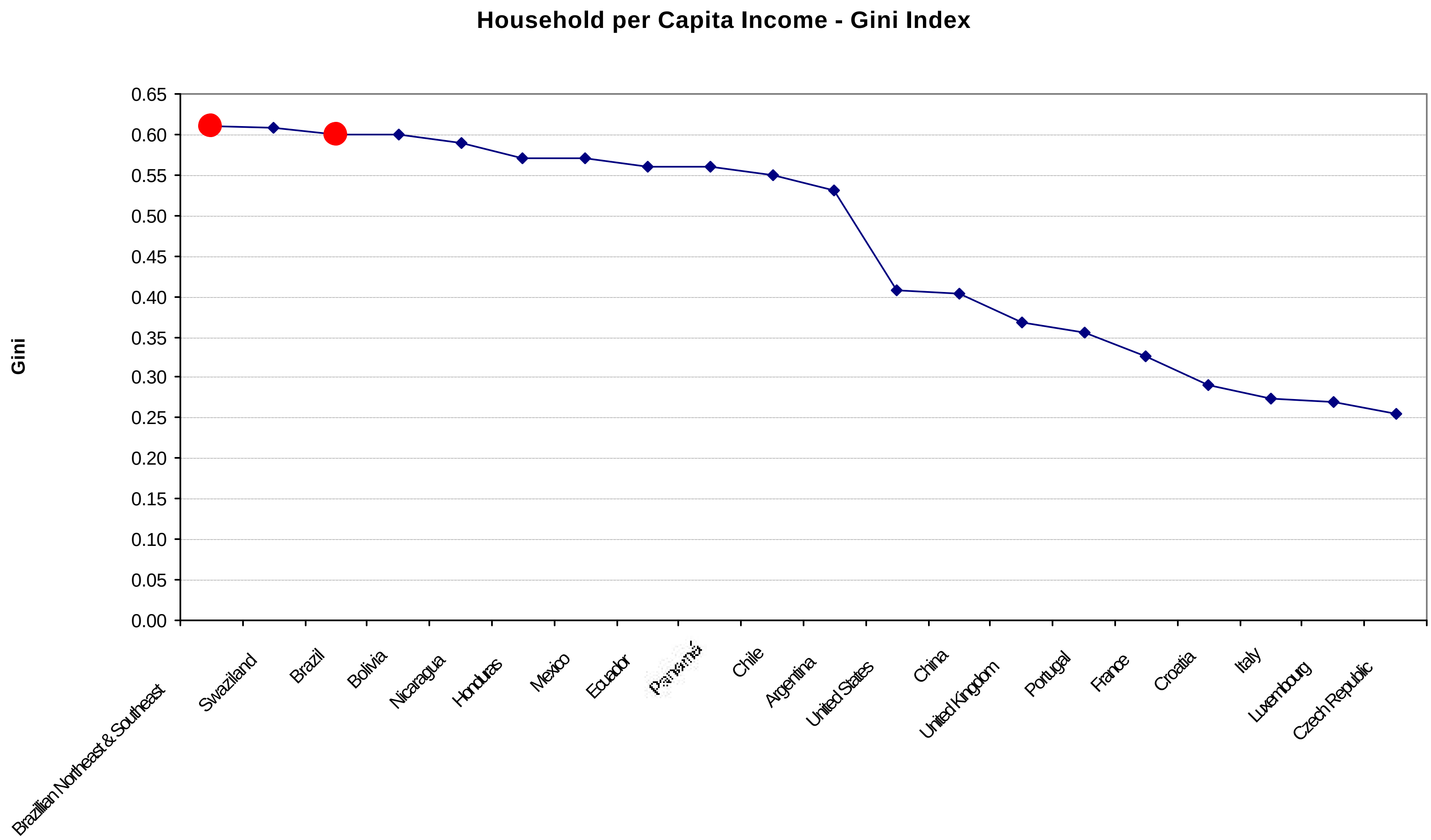

Figure 2
Inequality in Brazil Compared to the World: PNAD Per Capita Income Concept of Welfare

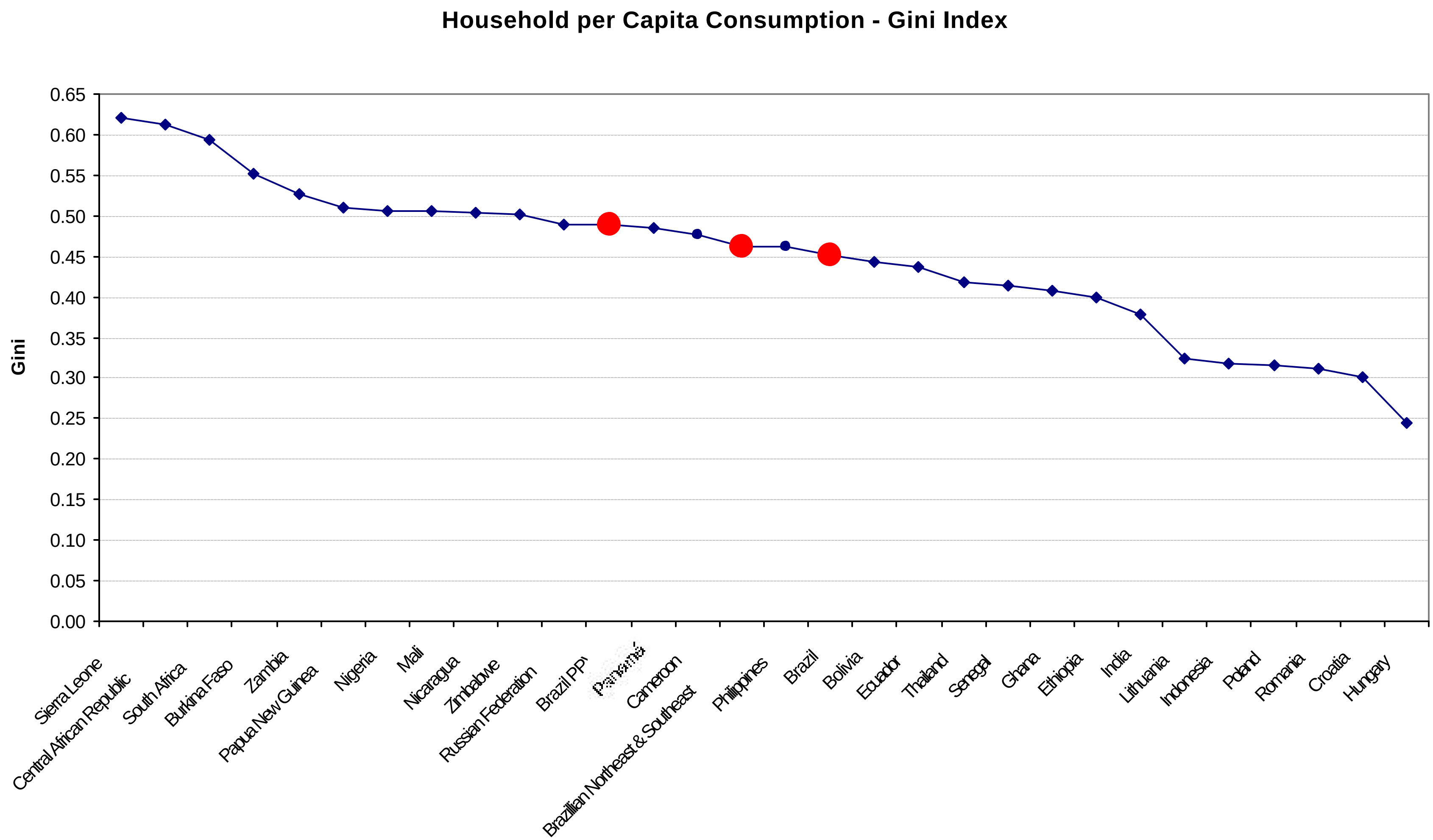

Table 1a: Poverty Measures by Region for Different Data Sets

Headcount

Region	PNAD Income				PPV Consumption		PNAD Imputed Consumption			
	FGT 0	S.E	Composition of the poor %	Composition of the poor by region%	FGT 0	S.E	FGT 0	Total S.E	Composition of the poor %	Composition of the poor by region%
RM Fortaleza	0.363	0.014	0.03	0.04	0.184	0.022	0.166	0.014	0.02	0.03
RM Recife	0.352	0.012	0.03	0.04	0.221	0.024	0.173	0.016	0.02	0.03
RM Salvador	0.372	0.014	0.03	0.04	0.191	0.023	0.187	0.020	0.02	0.03
Urban NE	0.498	0.017	0.30	0.42	0.376	0.027	0.337	0.031	0.34	0.44
Rural NE	0.710	0.019	0.33	0.46	0.497	0.025	0.455	0.029	0.36	0.47
Northeast	**0.549**		**0.71**	**1.00**	**0.387**		**0.350**		**0.77**	**1.00**
RM B.Horizonte	0.165	0.011	0.02	0.06	0.079	0.022	0.072	0.018	0.01	0.06
RM Rio de Janeiro	0.116	0.005	0.04	0.12	0.030	0.011	0.036	0.012	0.02	0.08
RM Sao Paulo	0.072	0.005	0.03	0.12	0.037	0.012	0.033	0.008	0.03	0.12
Urban SE	0.132	0.008	0.12	0.40	0.047	0.014	0.050	0.014	0.08	0.33
Rural SE	0.404	0.022	0.08	0.29	0.261	0.025	0.260	0.023	0.09	0.40
Southeast	**0.145**		**0.29**	**1.00**	**0.066**		**0.067**		**0.23**	**1.00**

Source: PNAD 1996 and 1997, and PPV 1996

1: PNAD per capita income, PPV per capita consumption and PNAD imputed per capita consumption have been adjusted for spatial price variation (see Ferreira, Lanjouw and Neri, 2000)

2: PNAD per capita income, PPV per capita consumption and PNAD imputed per capita consumption have been trimmed by the mean of the 0.5 and 99.5 percentile.

3: Poverty Line of R$ 65.07 in 1996 Sao Paulo reais (see Ferreira, Lanjouw, and Neri, 2000).

4: Sampling errors incorporate adjustments for complex survey design (see text and also Howes and Lanjouw)

Table 1b: Poverty Measures by Region for Different Data Sets

FGT 1

Region	PNAD Income		PPV Consumption		PNAD Imputed Consumption	
	FGT 1	S.E	FGT 1	S.E	FGT 1	Total S.E
RM Fortaleza	0.137	0.007	0.054	0.007	0.048	0.006
RM Recife	0.137	0.006	0.071	0.011	0.045	0.006
RM Salvador	0.146	0.007	0.055	0.008	0.058	0.007
Urban NE	0.217	0.010	0.124	0.014	0.108	0.004
Rural NE	0.370	0.013	0.176	0.013	0.154	0.014
RM B.Horizonte	0.053	0.004	0.014	0.004	0.020	0.005
RM Rio de Janeiro	0.035	0.002	0.003	0.001	0.008	0.004
RM Sao Paulo	0.022	0.002	0.007	0.003	0.007	0.002
Urban SE	0.041	0.003	0.010	0.004	0.011	0.003
Rural SE	0.162	0.013	0.074	0.009	0.083	0.010

Source: PNAD 1996 and 1997, and PPV 1996

1: PNAD per capita income, PPV per capita consumption and PNAD imputed per capita consumption have been adjusted for spatial price variation (see Ferreira, Lanjouw and Neri, 2000)

2: PNAD per capita income, PPV per capita consumption and PNAD imputed per capita consumption have been trimmed by the mean of the 0.5 and 99.5 percentile.

3: Poverty Line of R$ 65.07 in 1996 Sao Paulo reais (see Ferreira, Lanjouw, and Neri, 2000).

4: Sampling errors incorporate adjustments for complex survey design (see text and also Howes and Lanjouw)

Table 2a: Inequality Measures by Region for Different Data Sets
General Entropy Class c=0.5

Region	PNAD Income	PPV Consumption	PNAD Imputed Consumption	
	GE 0.5	GE 0.5	GE 0.5	Total S.E
RM Fortaleza	0.528	0.402	0.33	0.02
RM Recife	0.538	0.353	0.33	0.02
RM Salvador	0.595	0.338	0.32	0.02
Urban NE	0.481	0.357	0.29	0.01
Rural NE	0.391	0.254	0.21	0.01
RM B.Horizonte	0.473	0.323	0.27	0.02
RM Rio de Janeiro	0.491	0.338	0.30	0.02
RM Sao Paulo	0.410	0.325	0.27	0.02
Urban SE	0.398	0.273	0.24	0.01
Rural SE	0.412	0.217	0.22	0.02

Source: PNAD 1996 and 1997, and PPV 1996

1: PNAD per capita income, PPV per capita consumption and PNAD imputed per capita consumption have been adjusted for spatial price variation (see Ferreira, Lanjouw and Neri, 2000)

2: PNAD per capita income, PPV per capita consumption and PNAD imputed per capita consumption have been trimmed by the mean of the 0.5 and 99.5 percentile (see text).

Table 2b: Inequality Measures by Region for Different Data Sets
General Entropy Class c=1

Region	PNAD Income	PPV Consumption	PNAD Imputed Consumption	
	GE 1	GE 1	GE 1	Total S.E
RM Fortaleza	0.589	0.421	0.343	0.018
RM Recife	0.599	0.381	0.357	0.021
RM Salvador	0.654	0.355	0.336	0.025
Urban NE	0.532	0.384	0.312	0.016
Rural NE	0.419	0.270	0.225	0.016
RM B.Horizonte	0.522	0.343	0.279	0.020
RM Rio de Janeiro	0.537	0.360	0.319	0.023
RM Sao Paulo	0.436	0.341	0.282	0.017
Urban SE	0.422	0.282	0.250	0.016
Rural SE	0.452	0.226	0.224	0.018

Source: PNAD 1996 and 1997, and PPV 1996

1: PNAD per capita income, PPV per capita consumption and PNAD imputed per capita consumption have been adjusted for spatial price variation (see Ferreira, Lanjouw and Neri, 2000)

2: PNAD per capita income, PPV per capita consumption and PNAD imputed per capita consumption have been trimmed by the mean of the 0.5 and 99.5 percentile (see text).

Table 3: Poverty Estimates By UF: Headcount

State	PNAD Income				PNAD Imputed Consumption				Total Population
	FGT 0	S.E.	Composition of the poor%	Composition of the poor by state%	FGT 0	Total S.E	Composition of the poor%	Composition of the poor by state%	
Maranhão									
Urban	0.597	0.040	0.04	0.41	0.454	0.034	0.05	0.43	2,209,726
Rural	0.688	0.090	0.06	0.59	0.492	0.032	0.07	0.57	2,744,847
Total	0.648	0.062	0.10	1.00	0.475	0.023	0.12	1.00	4,954,573
Piauí									
Urban	0.523	0.076	0.02	0.47	0.360	0.039	0.03	0.47	1,445,881
Rural	0.789	0.020	0.03	0.54	0.542	0.041	0.03	0.53	1,094,075
Total	0.633	0.070	0.05	1.00	0.438	0.028	0.06	1.00	2,539,956
Ceará									
Metropolitan area	0.363	0.014	0.03	0.24	0.166	0.014	0.02	0.17	2,385,430
Other Urban	0.560	0.031	0.03	0.28	0.416	0.032	0.04	0.32	1,775,096
Rural	0.780	0.018	0.05	0.47	0.544	0.036	0.06	0.51	2,153,505
Total	0.561	0.020	0.11	1.00	0.365	0.016	0.12	1.00	6,314,031
Rio Grande do Norte									
Urban	0.374	0.065	0.02	0.52	0.238	0.053	0.02	0.57	1,561,487
Rural	0.637	0.040	0.02	0.48	0.325	0.040	0.01	0.43	855,456
Total	0.465	0.072	0.03	1.00	0.269	0.037	0.03	1.00	2,416,943
Paraíba									
Urban	0.397	0.050	0.02	0.50	0.231	0.042	0.02	0.52	1,998,321
Rural	0.744	0.033	0.02	0.51	0.391	0.040	0.02	0.48	1,087,765
Total	0.519	0.063	0.05	1.00	0.287	0.030	0.05	1.00	3,086,086
Pernambuco									
Metropolitan area	0.352	0.012	0.03	0.29	0.173	0.016	0.03	0.25	2,906,921
Other Urban	0.519	0.028	0.04	0.39	0.345	0.027	0.05	0.45	2,638,267
Rural	0.710	0.023	0.03	0.30	0.421	0.033	0.03	0.31	1,492,429
Total	0.496	0.018	0.11	1.00	0.290	0.014	0.11	1.00	7,037,616
Alagoas									
Urban	0.464	0.066	0.02	0.54	0.289	0.026	0.02	0.55	1,562,163
Rural	0.694	0.035	0.02	0.46	0.401	0.042	0.02	0.45	901,092
Total	0.549	0.070	0.04	1.00	0.330	0.022	0.04	1.00	2,463,255
Sergipe									
Urban	0.422	0.042	0.01	0.61	0.258	0.029	0.01	0.63	1,089,891
Rural	0.688	0.025	0.01	0.40	0.383	0.056	0.01	0.37	435,365
Total	0.495	0.054	0.02	1.00	0.294	0.026	0.02	1.00	1,525,256
Bahia									
Metropolitan area	0.372	0.014	0.03	0.15	0.187	0.020	0.03	0.12	2,562,443
Other Urban	0.521	0.026	0.08	0.39	0.352	0.027	0.09	0.41	4,895,099
Rural	0.678	0.017	0.09	0.47	0.439	0.028	0.10	0.47	4,463,861
Total	0.546	0.016	0.20	1.00	0.349	0.016	0.22	1.00	11,921,403
Northeast									
Urban	**0.459**	**0.011**	**0.38**	0.53	**0.290**	**0.027**	**0.41**	0.53	27,030,724
Rural	**0.707**	**0.019**	**0.33**	0.46	**0.455**	**0.029**	**0.36**	0.47	15,228,393
Total	**0.549**	**0.013**	**0.71**	**1.00**	**0.350**	**0.027**	**0.77**	**1.00**	42,259,117

State	PNAD Income				PNAD Imputed Consumption				Total Population
	FGT 0	S.E.	Composition of the poor%	Composition of the poor by state%	FGT 0	Total S.E	Composition of the poor%	Composition of the poor by state%	
Minas Gerais									
Metropolitan area	0.165	0.011	0.02	0.14	0.072	0.018	0.01	0.12	3,515,749
Other Urban	0.214	0.017	0.06	0.45	0.074	0.019	0.03	0.30	8,755,279
Rural	0.519	0.028	0.05	0.42	0.369	0.032	0.07	0.58	3,379,350
Total	0.268	0.017	0.13	1.00	0.138	0.013	0.11	1.00	15,650,378
Espírito Santo									
Urban	0.203	0.029	0.01	0.58	0.066	0.033	0.01	0.49	2,048,317
Rural	0.441	0.025	0.01	0.42	0.204	0.032	0.01	0.51	687,811
Total	0.263	0.032	0.02	1.00	0.101	0.026	0.01	1.00	2,736,128
Rio de Janeiro									
Metropolitan area	0.116	0.005	0.04	0.66	0.036	0.012	0.02	0.60	9,910,950
Other Urban	0.148	0.020	0.01	0.21	0.046	0.014	0.01	0.19	2,502,826
Rural	0.395	0.037	0.01	0.13	0.207	0.037	0.01	0.20	584,301
Total	0.134	0.008	0.05	1.00	0.046	0.009	0.03	1.00	12,998,076
São Paulo									
Metropolitan area	0.072	0.005	0.04	0.43	0.033	0.008	0.03	0.41	15,955,104
Other Urban	0.071	0.006	0.03	0.40	0.035	0.011	0.03	0.40	14,852,472
- Rural	0.210	0.026	0.01	0.17	0.115	0.035	0.01	0.19	2,068,606
Total	0.080	0.005	0.08	1.00	0.039	0.007	0.07	1.00	32,876,182
Southeast									
Urban	**0.113**	**0.004**	**0.20**	0.69	**0.044**	**0.012**	**0.13**	0.59	58,228,506
- **Rural**	**0.392**	**0.021**	**0.08**	0.31	**0.260**	**0.023**	**0.09**	0.41	6,720,068
Total	**0.145**	**0.006**	**0.29**	**1.00**	**0.067**	**0.013**	**0.23**	**1.00**	64,948,574

Source: PNAD 1996 and 1997, and PPV 1996

1: PNAD per capita income, PPV per capita consumption and PNAD imputed per capita consumption have been adjusted for spatial price variation (see Ferreira, Lanjouw and Neri, 2000)

2: PNAD per capita income, PPV per capita consumption and PNAD imputed per capita consumption have been trimmed by the mean of the 0.5 and 99.5 percentile.

3: Poverty Line of R$ 65.07 in 1996 Sao Paulo reais (see Ferreira, Lanjouw, and Neri, 2000).

4: Sampling errors incorporate adjustments for complex survey design (see text and also Howes and Lanjouw)

Table 4: Inequality Estimates By UF in the Northeast: General Entropy 0.5

		PNAD Income	PNAD Imputed Consumption	
State		**GE 0.5**	**GE 0.5**	**Total S.E**
Maranhão				
	Urban	0.424	0.263	0.021
	Rural	0.566	0.330	0.040
	Total	0.511	0.301	0.024
Piauí				
	Urban	0.472	0.309	0.020
	Rural	0.358	0.178	0.025
	Total	0.504	0.288	0.016
Ceará				
	Metropolitan area	0.528	0.326	0.016
	Other Urban	0.439	0.252	0.021
	Rural	0.328	0.175	0.014
	Total	0.551	0.351	0.010
Rio Grande do Norte				
	Urban	0.536	0.304	0.033
	Rural	0.352	0.182	0.016
	Total	0.526	0.281	0.022
Paraíba				
	Urban	0.508	0.302	0.025
	Rural	0.328	0.173	0.016
	Total	0.545	0.290	0.017
Pernambuco				
	Metropolitan area	0.538	0.331	0.018
	Other Urban	0.415	0.217	0.014
	Rural	0.318	0.182	0.014
	Total	0.518	0.315	0.010
Alagoas				
	Urban	0.534	0.345	0.034
	Rural	0.305	0.206	0.019
	Total	0.526	0.322	0.023
Sergipe				
	Urban	0.528	0.313	0.022
	Rural	0.293	0.186	0.016
	Total	0.525	0.300	0.017
Bahia				
	Metropolitan area	0.595	0.317	0.022
	Other Urban	0.429	0.267	0.015
	Rural	0.321	0.188	0.014
	Total	0.511	0.289	0.009
Northeast				
	Urban	**0.521**	**0.322**	**0.011**
	Rural	**0.392**	**0.215**	**0.014**
	Total	**0.531**	**0.311**	**0.009**

State		PNAD Income	PNAD Imputed Consumption	
		GE 0.5	**GE 0.5**	**Total S.E**
Minas Gerais				
	Metropolitan area	0.473	0.272	0.019
	Other Urban	0.405	0.226	0.015
	Rural	0.398	0.215	0.016
	Total	0.463	0.281	0.010
Espírito Santo				
	Urban	0.463	0.306	0.034
	Rural	0.436	0.187	0.020
	Total	0.498	0.320	0.026
Rio de Janeiro				
	Metropolitan area	0.491	0.300	0.021
	Other Urban	0.373	0.229	0.019
	Rural	0.350	0.200	0.021
	Total	0.483	0.296	0.016
São Paulo				
	Metropolitan area	0.410	0.272	0.016
	Other Urban	0.358	0.232	0.015
-	Rural	0.335	0.183	0.018
	Total	0.392	0.262	0.010
Southeast				
	Urban	**0.428**	**0.269**	**0.009**
-	**Rural**	**0.407**	**0.220**	**0.017**
	Total	**0.450**	**0.286**	**0.009**

Source: PNAD 1996 and 1997, and PPV 1996

1: PNAD per capita income, PPV per capita consumption and PNAD imputed per capita consumption have been adjusted for spatial price variation (see Ferreira, Lanjouw and Neri, 2000)

2: PNAD per capita income, PPV per capita consumption and PNAD imputed per capita consumption have been trimmed by the mean of the 0.5 and 99.5 percentile.

3: Poverty Line of R$ 65.07 in 1996 Sao Paulo reais (see Ferreira, Lanjouw, and Neri, 2000).

4: Sampling errors incorporate adjustments for complex survey design (see text and also Howes and Lanjouw)

Table 5: Incidence of Poverty in Brazil: By State and Location Type

State		PNAD Income			PNAD Imputed Consumption			
		FGT 0	S.E.	Composition of the poor %	FGT 0	Total S.E	Composition of the poor %	Total Population
Northeast								
Maranhão								
	Urban area 'urbanizadas'	0.597	0.040	0.41	0.454	0.034	0.43	2,209,727
	Other urban							
	Rural	0.802	0.035	0.41	0.612	0.046	0.42	1,620,278
	Other Rural (extensao urbana+povoado+nucleo)	0.526	0.146	0.18	0.319	0.026	0.15	1,124,569
	Total	0.648	0.062		0.475	0.022		4,954,574
Piauí								
	Urban area 'urbanizadas'	0.523	0.076	0.47	0.360	0.039	0.47	1,445,882
	Other urban							
	Rural	0.811	0.018	0.45	0.585	0.036	0.46	883,755
	Other Rural (extensao urbana+povoado+nucleo)	0.700	0.058	0.09	0.363	0.076	0.07	210,320
	Total	0.633	0.070		0.438	0.026		2,539,956
Ceará								
	Metropolitan area	0.363	0.014	0.24	0.166	0.014	0.17	2,385,430
	Urban area 'urbanizadas'	0.558	0.031	0.28	0.415	0.032	0.32	1,753,893
	Other urban	0.705	0.000	0.00	0.478	0.187	0.00	21,204
	Rural	0.792	0.018	0.45	0.562	0.037	0.49	2,013,454
	Other Rural (extensao urbana+povoado+nucleo)	0.605	0.059	0.02	0.286	0.044	0.02	140,051
	Total	0.561	0.020		0.365	0.016		6,314,031
Rio Grande do Norte								
	Urban area 'urbanizadas'	0.374	0.065	0.52	0.238	0.053	0.57	1,561,487
	Other urban							
	Rural	0.681	0.046	0.34	0.369	0.034	0.31	554,552
	Other Rural (extensao urbana+povoado+nucleo)	0.555	0.071	0.15	0.244	0.060	0.11	300,904
	Total	0.465	0.072		0.269	0.036		2,416,943
Paraíba								
	Urban area 'urbanizadas'	0.397	0.050	0.50	0.231	0.042	0.52	1,998,321
	Other urban							
	Rural	0.765	0.030	0.46	0.411	0.035	0.45	969,603
	Other Rural (extensao urbana+povoado+nucleo)	0.576	0.092	0.04	0.225	0.102	0.03	118,162
	Total	0.519	0.063		0.287	0.029		3,086,086
Pernambuco								
	Metropolitan area	0.352	0.012	0.29	0.173	0.016	0.25	2,906,921
	Urban area 'urbanizadas'	0.526	0.025	0.38	0.342	0.029	0.42	2,520,364
	Other urban	0.367	0.000	0.01	0.394	0.224	0.02	117,903
	Rural	0.717	0.022	0.29	0.432	0.036	0.30	1,423,761
	Other Rural (extensao urbana+povoado+nucleo)	0.580	0.082	0.01	0.192	0.110	0.01	68,668
	Total	0.496	0.018		0.290	0.015		7,037,616

State	PNAD Income			PNAD Imputed Consumption			Total Population
	FGT 0	S.E.	Composition of the poor %	FGT 0	Total S.E	Composition of the poor %	
Alagoas							
Urban area 'urbanizadas'	0.460	0.066	0.52	0.286	0.026	0.54	1,534,823
Other urban	0.670	0.000	0.01	0.424	0.151	0.01	27,341
Rural	0.758	0.025	0.36	0.470	0.044	0.37	640,834
Other Rural (extensao urbana+povoado+nucleo)	0.538	0.058	0.10	0.232	0.043	0.07	260,258
Total	0.549	0.070		0.330	0.020		2,463,255
Sergipe							
Urban area 'urbanizadas'	0.416	0.041	0.58	0.254	0.033	0.60	1,051,686
Other urban	0.572	0.000	0.03	0.374	0.229	0.03	38,206
Rural	0.682	0.028	0.29	0.382	0.065	0.28	323,889
Other Rural (extensao urbana+povoado+nucleo)	0.709	0.049	0.10	0.385	0.043	0.10	111,476
Total	0.495	0.054		0.294	0.027		1,525,256
Bahia							
Metropolitan area	0.372	0.014	0.15	0.187	0.020	0.12	2,562,443
Urban area 'urbanizadas'	0.521	0.026	0.39	0.352	0.027	0.41	4,895,100
Other urban			0.00			0.00	0
Rural	0.686	0.017	0.40	0.464	0.027	0.42	3,794,764
Other Rural (extensao urbana+povoado+nucleo)	0.635	0.041	0.07	0.300	0.046	0.05	669,097
Total	0.546	0.016		0.349	0.015		11,921,404
Southeast							
Minas Gerais							
Metropolitan area	0.165	0.011	0.14	0.072	0.018	0.12	3,515,749
Urban area 'urbanizadas'	0.213	0.017	0.44	0.074	0.019	0.30	8,698,918
Other urban	0.386	0.125	0.01	0.145	0.099	0.00	56,362
Rural	0.522	0.029	0.38	0.381	0.032	0.54	3,028,433
Other Rural (extensao urbana+povoado+nucleo)	0.493	0.061	0.04	0.267	0.049	0.04	350,917
Total	0.268	0.017		0.138	0.013		15,650,378
Espírito Santo							
Urban area 'urbanizadas'	0.203	0.029	0.58	0.066	0.033	0.49	2,048,317
Other urban			0.00			0.00	0
Rural	0.443	0.027	0.39	0.211	0.035	0.49	636,281
Other Rural (extensao urbana+povoado+nucleo)	0.408	0.000	0.03	0.122	0.109	0.02	51,531
Total	0.263	0.032		0.101	0.026		2,736,128
Rio de Janeiro							
Metropolitan area	0.116	0.005	0.66	0.036	0.012	0.60	9,910,950
Urban area 'urbanizadas'	0.147	0.021	0.20	0.046	0.014	0.19	2,424,810
Other urban	0.178	0.043	0.01	0.050	0.092	0.01	78,016
Rural	0.394	0.038	0.13	0.208	0.038	0.20	567,439
Other Rural (extensao urbana+povoado+nucleo)	0.406	0.000	0.00	0.176	0.057	0.00	16,862
Total	0.134	0.008		0.046	0.009		12,998,076

State		PNAD Income FGT 0	PNAD Income S.E.	PNAD Income Composition of the poor %	PNAD Imputed Consumption FGT 0	PNAD Imputed Consumption Total S.E	PNAD Imputed Consumption Composition of the poor %	Total Population
São Paulo								
	Metropolitan area	0.072	0.005	0.43	0.033	0.008	0.41	15,955,104
	Urban area 'urbanizadas'	0.070	0.006	0.39	0.034	0.011	0.39	14,589,900
	Other urban	0.135	0.033	0.01	0.068	0.034	0.01	262,573
-	Rural	0.214	0.028	0.15	0.119	0.036	0.17	1,872,010
-	Other Rural (extensao urbana+povoado+nucleo)	0.172	0.045	0.01	0.077	0.035	0.01	196,596
	Total	0.080	0.005		0.039	0.007		32,876,182

Source: PNAD 1996 and 1997, and PPV 1996

1: PNAD per capita income, PPV per capita consumption and PNAD imputed per capita consumption have been adjusted for spatial price variation (see Ferreira, Lanjouw and Neri, 2000)

2: PNAD per capita income, PPV per capita consumption and PNAD imputed per capita consumption have been trimmed by the mean of the 0.5 and 99.5 percentile.

3: Poverty Line of R$ 65.07 in 1996 Sao Paulo reais (see Ferreira, Lanjouw, and Neri, 2000).

4: Sampling errors incorporate adjustments for complex survey design (see text and also Howes and Lanjouw)

Table 6: Incidence of Poverty in Brazil: By Occupation Group

	PNAD Income							PNAD Imputed Consumption							
	FGT 0	**Composition of the poor %**	**Composition of the poor by Occup. Choice %**	FGT 1	GE 0.5	GE 1	Mean	FGT 0	**Composition of the poor %**	**Composition of the poor by Occup. Choice %**	FGT 1	GE 0.5	GE 1	Mean	**Total Population**
Total															
Out of labor force	0.211		0.53	0.077	0.496	0.528	261.28	0.118		0.51	0.033	0.320	0.334	238.86	82,431,739
Unemployeed	0.507		0.05	0.252	0.486	0.506	109.39	0.141		0.02	0.040	0.295	0.308	203.81	3,192,585
Formal sector	0.482		0.04	0.182	0.256	0.264	91.93	0.257		0.04	0.078	0.198	0.202	121.67	2,640,427
Informal sector	0.708		0.12	0.355	0.308	0.320	59.60	0.440		0.13	0.154	0.229	0.239	92.92	5,518,118
Goverment	0.098		0.00	0.022	0.335	0.356	299.40	0.060		0.00	0.018	0.200	0.197	227.30	57,477
Self employeed Urban	0.584		0.06	0.284	0.500	0.531	97.79	0.397		0.06	0.141	0.316	0.338	114.97	3,115,177
Self employeed Rural	0.697		0.20	0.381	0.431	0.454	64.22	0.476		0.24	0.164	0.197	0.203	84.05	9,564,370
			1.00												
Northeast															
Out of labor force	0.428	0.36	0.50	0.172	0.497	0.556	139.41	0.265	0.38	0.49	0.080	0.309	0.332	146.32	27,175,783
Unemployeed	0.764	0.02	0.03	0.440	0.543	0.610	62.66	0.316	0.02	0.02	0.097	0.270	0.291	123.02	1,010,301
Formal sector	0.752	0.02	0.03	0.345	0.222	0.240	54.97	0.445	0.02	0.03	0.150	0.173	0.176	85.39	874,175
Informal sector	0.847	0.08	0.12	0.452	0.230	0.236	41.24	0.547	0.09	0.12	0.197	0.174	0.178	73.16	3,230,302
Goverment	0.150	0.00	0.00	0.033	0.258	0.267	229.61	0.082	0.00	0.00	0.025	0.240	0.235	228.37	37,295
Self employeer Urban	0.723	0.05	0.07	0.362	0.357	0.374	61.43	0.514	0.06	0.08	0.186	0.201	0.208	79.32	2,282,795
Self employeer Rural	0.764	0.18	0.25	0.428	0.359	0.365	50.23	0.514	0.21	0.27	0.177	0.176	0.180	77.36	7,648,469
		0.71	1.00						0.77	1.00					
Southeast															
Out of labor force	0.105	0.18	0.61	0.031	0.422	0.453	320.52	0.046	0.13	0.59	0.010	0.271	0.284	284.37	55,255,956
Unemployeed	0.389	0.03	0.09	0.167	0.411	0.430	130.76	0.060	0.01	0.03	0.014	0.250	0.262	241.22	2,182,284
Formal sector	0.349	0.02	0.07	0.102	0.207	0.214	110.08	0.164	0.02	0.07	0.043	0.175	0.179	139.63	1,766,252
Informal sector	0.512	0.04	0.12	0.217	0.286	0.297	85.61	0.289	0.03	0.15	0.094	0.227	0.232	120.82	2,287,816
Goverment	0.000	0.00	0.00	0.000	0.318	0.330	430.90	0.019	0.00	0.00	0.003	0.127	0.126	225.32	20,183
Self employeer Urban	0.205	0.01	0.02	0.070	0.343	0.350	196.63	0.077	0.00	0.01	0.018	0.227	0.235	212.75	832,382
Self employeer Rural	0.432	0.03	0.09	0.194	0.423	0.450	119.68	0.326	0.03	0.15	0.110	0.219	0.223	110.77	1,915,901
		0.29	1.00						0.23	1.00					

Table 7: Decomposing Inequality: PNAD Income Versus PNAD Consumption
General Entropy Class (0.5)

	PNAD Income	*PNAD Consumption*
1. National Level		
Total Inequality	0.563	0.352
% Attributable to BETWEEN Group Component		
All regions	25.3	26.6
Rural vs Urban	12.5	12.9
Northeast vs Southeast	17.5	18.0
Rural NE vs Urban NE vs Rural SE vs Urban SE	23.7	24.5
Metropolitan Regions vs Rest	7.6	9.5
Metropolitan Regions vs Other Urban vs Rural	15.3	16.8
2. Rural Areas		
Total Inequality	0.454	0.229
% Attributable to BETWEEN Group Component		
Northeast vs Southeast	13.4	5.9
By State	18.4	8.9
3. Urban Areas		
Total Inequality	0.509	0.323
% Attributable to BETWEEN Group Component		
Northeast vs Southeast	12.5	13.8
By State	15.1	15.6
4. Occupational status of the head		
Total Inequality	0.563	0.352
% Attributable to BETWEEN Group Component	16.1	13.9
Northeast Total Inequality	0.531	0.311
% Attributable to BETWEEN Group Component	17.5	12.7
Southeast Total Inequality	0.450	0.286
% Attributable to BETWEEN Group Component	9.2	7.5
5. Occupational status of the head **Formal x Informal**		
Total Inequality	0.309	0.226
% Attributable to BETWEEN Group Component	7.1	3.7
Northeast Total Inequality	0.235	0.176
% Attributable to BETWEEN Group Component	3.2	1.2
Southeast Total Inequality	0.257	0.206
% Attributable to BETWEEN Group Component	3.1	1.3