

**The Sources of Welfare Disparities Across and Within Regions of Brazil:  
Evidence from the 2002-03 POF**

Roy Katayama  
Emmanuel Skoufias

The World Bank  
1818 H Street NW  
Washington DC, 20433

**Abstract:**

Brazil's inequalities in welfare and poverty between and within regions can be accounted by differences in household attributes and returns to those attributes. Using Blinder-Oaxaca decompositions at the mean as well as at different quantiles of welfare distributions on regionally representative household survey data (2002-03 *POF*), this paper finds that household attributes account for most of the welfare differences between urban and rural areas within regions. However, comparing the lagging Northeast region with the leading Southeast region, differences in returns to attributes account for a large part of the welfare disparities, in particular in metropolitan areas, supporting the presence of agglomeration effects in booming areas.

*JEL classification:*

*Keywords:* Brazil, Leading and Lagging Regions, Welfare, Poverty,

*Corresponding author:* Emmanuel Skoufias, The World Bank, 1818 H Street NW, Washington DC 20433-USA. tel: (202)458-7539. fax: (202) 522-3134. e-mail: [eskoufias@worldbank.org](mailto:eskoufias@worldbank.org).

This paper is part of a regional World Bank study "Poor Areas and/or Poor People in LAC" financed by the LAC Chief Economist office and the PREM Poverty anchor unit. The authors share equal responsibility for the contents of this paper. The findings, interpretations, and conclusions in this paper are entirely those of the authors and they do not necessarily reflect the view of the World Bank.

## 1. INTRODUCTION

Although Brazil is now one of the world's top ten largest economies, high levels of poverty and inequality continue to pose major challenges for the country. Like in most countries, Brazil's development is spatially uneven. People in metropolitan areas tend to be better off than people in urban (non-metropolitan) areas, and people in urban areas tend to be better off than people in rural areas. Between the five major regions, the North and Northeast regions lag behind the South and Southeast where economic centers like Sao Paulo are booming. In 2002, the Northeast was home to 28% of the country's population but accounted for 50% of the country's poor, making its poverty rate of 38% the highest in the country. On the other hand, the Southeast was home to 43% of the population but accounted for 25% of the poor and had a poverty rate of 13%. Although recent decreases in inequality have been encouraging, the Gini index has nonetheless remained high – above 55 – over the last 25 years.

These persistently high levels of inequality in Brazil have raised many economic and social concerns. First, there are concerns that high levels of inequality may compromise economic efficiency and growth. For instance, credit and insurance market failures may prevent poorer households from investing in and contributing to the economy at an optimal level, thereby undermining efficiency and growth. Also, inequality in political influence may lead to an inefficient allocation of resources for the public services necessary for greater output. Furthermore, lower social cohesion and greater crime may increase the cost of doing business (World Bank, 2003). Second, there are concerns that high initial levels of inequality may undermine the poverty reduction potential of growth. Based on empirical studies, the growth elasticity of poverty tends to be low in countries with high levels of initial inequality. (World Bank 2003; Ravallion 2004). Third, there are concerns that addressing the inequality of opportunities is fundamental to the pursuit of social justice, and doing so constitutes a development objective in itself (Roemer 1998, World Bank 2003 & 2005).

Several complementary factors are important in determining the observed spatial distribution of welfare across the major regions and areas of Brazil. One factor is the concentration and availability of skilled and unskilled labor in a particular location. Since migration is not restricted, the sorting of certain attributes may occur over time to produce variation in the concentration of household attributes in a region and area. Other factors, such as the quality of infrastructure, the distance to and size of markets, and the ability of local government to finance public investments and to create the right incentives for private sector development, can influence the scope of opportunities available and the rate of return to attributes, such as occupational specialties or education.

In exploring the factors that may be driving this spatial inequality of welfare within Brazil, a fundamental question is whether the observed differences are due primarily to the spatial concentration of

individuals with characteristics that tend to leave them in poverty or the geographical differences in the returns to these characteristics<sup>1</sup>. In other words, would individuals who live in different regions of the country, but are otherwise identical, have comparable standards of living, indicating that the returns to identical characteristics are similar? Or are their returns and standard of living quite different across regions?

Several studies have investigated this question using income as a measure of household welfare typically obtained from the annual PNAD surveys and employing different methods. Duarte et al. (2004) utilized a semi-parametric model (following DiNardo, Fortin & Lemieux, 1996), to investigate the educational disparities between the Northeast and Southeast regions as a partial determinant of income differences. They concluded that “more than 50% of the income difference is explained by the difference in schooling.” However, factors other than education were not considered in their study. Bourguignon et al. (2002) find that “most of Brazil’s excess income inequality is due to underlying inequalities in the distribution of two key endowments: access to education and to sources of nonlabor income, mainly pensions.” This conclusion is based on a comparison of the relative roles of three components – the distribution of population characteristics, the returns to these characteristics, and the occupational structure of the population – in accounting for the income distributions between Brazil and the United States.

Guimarães et al. (2006) accounts for differences in labor income between the metropolitan areas of the Southeast and Northeast regions using the quantile regression decomposition method in Machado and Mata (2004). The paper finds that the difference in the returns to education accounts for a larger share of the income gap than the differences in the distribution of education, suggesting that policy interventions focused on education alone are not likely to be sufficient in decreasing regional inequality. It is possible that the higher returns to education that Guimarães et al. (2006) finds in the Southeast may be due to agglomeration effects in booming metropolitan areas. As described in the New Economic Geography literature, agglomeration economies are characterized by increasing economies of scale. With well developed infrastructure, a high degree of market specialization, greater competition, information exchange, and more efficient matching in the labor market, the environment is conducive to lowering costs and producing higher returns (Venables 2005, Krugman 1998). Thus, one could expect metropolitan areas of leading regions to have both high returns – from increasing economies of scale – and a higher concentration of individuals with valuable human capital assets (both observable education and unobservable ability and motivation) as talented workers are attracted to the higher rates of return and wider range of employment opportunities.

---

<sup>1</sup> This question was posed by Ravallion and Wodon (1999) for Bangladesh.

This study deepens our understanding of poverty and the spatial welfare disparities in Brazil through an investigation of the role of demographics, human capital, occupation, and structural geographical differences in returns. This is done by employing a recent survey in Brazil, the 2002-2003 Household Budget Survey (*Pequisa de Orcamentos Familiares* or *POF*) that allows us to use consumption, rather than income, as a measure of household welfare. For a variety of reasons, consumption is considered by economists to provide a more accurate measure of household welfare. There are both conceptual and pragmatic reasons why consumption expenditures from household surveys might be preferred, for the purpose of poverty and inequality analysis, over an indicator such as household income. First, consumption tends to fluctuate less in the short term than income, which can be affected by the seasonality of employment. Consumption expenditures reflect not only what a household is able to command based on its current income, but also whether that household can access credit markets or household savings at times when current incomes are low or even negative (due perhaps to seasonal variation or a harvest failure). In this way, consumption is thought to provide a better picture of a household's longer run standard of living than a measure of current income. Second, income measures may not accurately capture in-kind, seasonal, or informal income. While poor households are probably purchasing and consuming only a relatively narrow range of goods and services, their total income may derive from multiple different activities with strong seasonal variation and with associated costs that are not always easily assigned. Third, income surveys are susceptible to under-reporting as respondents may perceive incentives to do so.

While it is clear that regional disparities in income and welfare have existed for some time in Brazil, there is not a clear consensus on whether it is mainly due to returns to characteristics or the distribution of characteristics. Our paper explores these issues in more depth. Aside from the fact that we use consumption instead of income as a measure of welfare, our approach is different from previous studies in that we (i) disaggregate urban areas into metropolitan and non-metropolitan areas within regions to obtain a more refined picture, and (ii) analyze both differences in mean welfare and differences between distributions, using Oaxaca-Blinder and quantile regression decomposition techniques.

This paper is organized as follows. Section 2 presents the methodology for the mean decomposition and quantile regression decomposition along with a brief summary of the 2002-03 *Pequisa de Orcamentos Familiares* household budget survey and the variables used in the analysis. Section 3 presents the results of these decompositions, and Section 4 concludes and discusses some of the implications for policy.

## 2. METHODOLOGY

In this section of the paper we outline the methodology we use for our investigation of the factors behind Brazil's spatial disparities. We first apply a slight variation of the Oaxaca-Blinder methodology used in Ravallion and Wodon (1999) to explore differences in mean welfare between metropolitan, non-metropolitan urban, and rural areas, within and between regions. By comparing observed mean welfare and poverty rates with simulated conditional means and poverty rates, we account for the relative importance of "returns" and "covariate" effects.<sup>2</sup> Next, we focus on the leading (Southeast) and lagging (Northeast) regions of Brazil, and repeat the analysis by estimating quantile regression decompositions as in Machado and Mata (2005) and Nguyen et al. (2007) to determine the relative importance of these geographic and concentration effects at different quantiles of the distribution.

### *Metro-Urban-Rural Mean Decomposition*

The empirical framework consists of estimating three separate regression equations – one for metropolitan (denoted by M), one for urban areas (denoted by U), and one for rural areas (denoted by R) of the form

$$\ln C_i = a_M + \beta_M X_i + \gamma_M D_i + \varepsilon_{Mi} \quad (i \in M) \quad (1)$$

$$\ln C_i = a_U + \beta_U X_i + \gamma_U D_i + \varepsilon_{Ui} \quad (i \in U) \quad (2)$$

$$\ln C_i = a_R + \beta_R X_i + \gamma_R D_i + \varepsilon_{Ri} \quad (i \in R) \quad (3)$$

where  $\ln C_i$  is the natural logarithm of the welfare ratio of household  $i$  (per capita expenditure of household  $i$  divided by the area-specific poverty line),  $X_i$  is a vector of household-specific portable (i.e. individuals or households can migrate) non-geographic characteristics, such as the age and composition of the members of the household, the education level of the head and his/her spouse, the main occupation of the household head, etc.,  $D_i$  is a vector of binary variables identifying the different regions of the country and  $\varepsilon_{M,U,R}$  is a random disturbance term with the usual properties, summarizing the influence of all other

---

<sup>2</sup> The use of more elaborate methods developed by Jalan and Ravallion (2002) that control for latent heterogeneity in measured consumption growth was considered, but without panel data of consumption expenditures covering two time periods (i.e. three points in time), applying such methods was not feasible.

factors on the standard of living. In this specification,  $\alpha_{M,U,R}$ ,  $\beta_{M,U,R}$  and  $\gamma_{M,U,R}$  are parameter vectors.<sup>3</sup> Specifically, the parameter vector  $\beta_{M,U,R}$  may be considered as summarizing the “returns” to non-geographic household characteristics  $X_i$ , in terms of how these characteristics translate in household living standard, summarized here by the welfare ratio, in metropolitan (M), urban (U) and rural areas (R).

Based on the above specification differences in living standards between different geographic areas may be attributed to two main factors:

1. differences in the portable non-geographic characteristics between metropolitan, urban, and rural areas, holding all else constant, i.e. differences between  $X_{Mi}$  and  $X_{Ui}$ , and between  $X_{Ui}$  and  $X_{Ri}$ ; and
2. differences in the returns to characteristics, holding all else constant, i.e. differences between  $\beta_M$  and  $\beta_U$ , and between  $\beta_U$  and  $\beta_R$ .

The estimates of equation (1-3) can be used to conduct a number of useful comparisons. The first question concerns the overall differential in living standards between metropolitan, urban and rural areas within Brazil. This entails a comparison of the metro-urban and urban-rural differentials in mean welfare ratios,

$$\begin{aligned} E(\ln C_i | i \in M, X_i = X_M) - E(\ln C_i | i \in U, X_i = X_U) = \\ = (\alpha_M - \alpha_U) + (\beta_M X_M - \beta_U X_U) + \sum_k (s_{Mk} \gamma_{Mk} - s_{Uk} \gamma_{Uk}) \end{aligned} \quad (4)$$

$$\begin{aligned} E(\ln C_i | i \in U, X_i = X_U) - E(\ln C_i | i \in R, X_i = X_R) = \\ = (\alpha_U - \alpha_R) + (\beta_U X_U - \beta_R X_R) + \sum_k (s_{Uk} \gamma_{Uk} - s_{Rk} \gamma_{Rk}) \end{aligned} \quad (5)$$

where  $X_{M,U,R}$  are the sample means for metro, urban, and rural areas respectively, and  $s_{Mk,Uk,Rk}$  are the proportions of region  $k$ 's population in each sector.

The second question relates to the **inter**-regional differential in living standards within metro, urban, and rural sub-populations of Brazil. This can be answered simply by comparing the estimates of

---

<sup>3</sup> In the above specification, the returns to attributes are assumed to be the same across regions for metropolitan or urban or rural areas. As a means of examining the sensitivity of results, we also estimate equations (1), (2) and (3) separately by region.

the coefficients of the region binary variables in the metro, urban and rural areas (i.e. comparing  $\gamma_{Mk}$ ,  $\gamma_{Uk}$  and  $\gamma_{Rk}$  for the  $k$ th region)

The third question has to do with the **intra**-regional differential in living standards, that is the difference between metro, urban, and rural areas within a region. The answer to this question involves a comparison of the expected gain in the welfare ratio from living in, say, a metropolitan area of a given region over living in an urban area of the same region, holding household characteristics constant at some level (e.g. at the national mean).

$$\begin{aligned} E(\ln C_i | i \in M, D_i = D^k, X_i = X^*) - E(\ln C_i | i \in U, D_i = D^k, X_i = X^*) = \\ = (\alpha_M - \alpha_U) + (\beta_M - \beta_U)X^* + (\gamma_{Mk} - \gamma_{Uk}) \end{aligned} \quad (6)$$

$$\begin{aligned} E(\ln C_i | i \in U, D_i = D^k, X_i = X^*) - E(\ln C_i | i \in R, D_i = D^k, X_i = X^*) = \\ = (\alpha_U - \alpha_R) + (\beta_U - \beta_R)X^* + (\gamma_{Uk} - \gamma_{Rk}) \end{aligned} \quad (7)$$

where  $D^k$  identifies the specific region examined and  $X^*$  is the sample mean value of household observed characteristics over the full sample (metro, urban, and rural areas pooled).

### ***Estimating the importance of the geographic (returns) effect in explaining observed patterns of welfare and poverty***

The relative importance of the geographic (returns) effect in explaining observed patterns of welfare and poverty can be estimated by simulating two alternative profiles

- *The geographic/returns profile* : the welfare and poverty profile for each metro, urban, and rural area in each region that is based on the simulated welfare ratios that are obtained holding observed household characteristics constant at  $X^*$ , the sample mean value of household observed characteristics over the full sample (metro, urban and rural areas pooled), while allowing the returns to characteristics to differ between areas (using  $\alpha_M$ ,  $\beta_M$  and  $\gamma_{Mk}$  for metro areas of region  $k$ ,  $\alpha_U$ ,  $\beta_U$  and  $\gamma_{Uk}$  for urban areas, and  $\alpha_R$ ,  $\beta_R$  and  $\gamma_{Rk}$  for the rural areas). Thus in this profile any differences in poverty between the urban and rural areas and across regions are due to differences in the intercept terms and/or slopes. Assuming a log normal distribution of welfare ratios, the geographic poverty profile can be estimated separately for the metro, urban, and rural areas or each region  $k$  as follows

$$\text{Prob}[\ln C_i < 0 | i \in M, D_i = D^k, X_i = X^*] = \Phi\left[-(\alpha_M + \beta_M X^* + \gamma_{Mk})/\sigma_M\right]$$

$$\begin{aligned} \Pr ob[\ln C_i < 0 | i \in U, D_i = D^k, X_i = X^*] &= \Phi[-(\alpha_U + \beta_U X^* + \gamma_{Uk})/\sigma_U] \quad (8) \\ \Pr ob[\ln C_i < 0 | i \in R, D_i = D^k, X_i = X^*] &= \Phi[-(\alpha_R + \beta_R X^* + \gamma_{Rk})/\sigma_R] \end{aligned}$$

where  $\Phi$  is the cumulative density of the standard normal, the term inside the parentheses is the simulated log welfare ratio, and  $\sigma_M$ ,  $\sigma_U$ , and  $\sigma_R$  are the standard deviations of the errors in the metro, urban, and rural regressions, respectively.

- *The concentration/covariate profile*: the welfare and poverty profile for each rural and urban area in each region that is based on the simulated welfare ratios that are obtained holding the returns to household characteristics constant while allowing the mean characteristics of households in metropolitan, urban and rural areas of each region to vary. Again, assuming a log normal distribution, the concentration poverty profile for the urban and rural areas or each region  $k$  can be constructed as<sup>4</sup>

$$\begin{aligned} \Pr ob[\ln C_i < 0 | i \in M, D_i = D^k, X_i = X_M^k] &= \Phi[-(\alpha_N + \beta_N X_M^k)/\sigma_M] \\ \Pr ob[\ln C_i < 0 | i \in U, D_i = D^k, X_i = X_U^k] &= \Phi[-(\alpha_N + \beta_N X_U^k)/\sigma_U] \quad (9) \\ \Pr ob[\ln C_i < 0 | i \in R, D_i = D^k, X_i = X_R^k] &= \Phi[-(\alpha_N + \beta_N X_R^k)/\sigma_R] \end{aligned}$$

where the term inside the parentheses represents the simulated log welfare ratio,

$$\alpha_N = s_U \left( \alpha_U + \sum_k s_{Uk} \gamma_{Uk} \right) + s_R \left( \alpha_R + \sum_k s_{Rk} \gamma_{Rk} \right)$$

and

$$\beta_N = s_U \beta_U + s_R \beta_R.$$

### ***Alternative method allowing for variation of coefficients across regions***

Instead of including binary variables for each region as in equations (1-3), it is possible to estimate a regression equation separately for metropolitan, non-metropolitan urban, and rural areas of each region  $k$  since the POF data are representative at the regional level for each of these areas.

$$\ln C_i = a_{Mk} + \beta_{Mk} X_i + \varepsilon_{Mki}. \quad (1A)$$

$$\ln C_i = a_{Uk} + \beta_{Uk} X_i + \varepsilon_{Uki}. \quad (2A)$$

---

<sup>4</sup> Given that  $\ln C_i$  is the logarithm of the welfare ratio of household  $i$  (PCE of household  $i$  divided by the area-specific poverty line) it follows that  $\ln C_i < 0$  identifies a poor household (i.e. a household with PCE less than poverty line).

$$\ln C_i = a_{Rk} + \beta_{Rk} X_i + \varepsilon_{Rki}. \quad (3A)$$

Using the constants and coefficients from (1A-3A) in equations (6) to (9) permits variation of the returns to characteristics across regions in the estimations. The alternative specifications are below:

- Intra-regional differences

$$\begin{aligned} E(\ln C_i | i \in M_k, X_i = X^*) - E(\ln C_i | i \in U_k, X_i = X^*) &= \\ &= (\alpha_{Mk} - \alpha_{Uk}) + (\beta_{Mk} - \beta_{Uk}) X^* \end{aligned} \quad (6A)$$

$$\begin{aligned} E(\ln C_i | i \in U_k, X_i = X^*) - E(\ln C_i | i \in R_k, X_i = X^*) &= \\ &= (\alpha_{Uk} - \alpha_{Rk}) + (\beta_{Uk} - \beta_{Rk}) X^* \end{aligned} \quad (6A)$$

- Simulated geographic poverty profile

$$\begin{aligned} \Pr ob[\ln C_i < 0 | i \in M_k, X_i = X^*] &= \Phi[-(\alpha_{Mk} + \beta_{Mk} X^*)/\sigma_{Mk}] \\ \Pr ob[\ln C_i < 0 | i \in U_k, X_i = X^*] &= \Phi[-(\alpha_{Uk} + \beta_{Uk} X^*)/\sigma_{Uk}] \\ \Pr ob[\ln C_i < 0 | i \in R_k, X_i = X^*] &= \Phi[-(\alpha_{Rk} + \beta_{Rk} X^*)/\sigma_{Rk}] \end{aligned} \quad (8A)$$

where  $\sigma_{mk}, \sigma_{uk}, \sigma_{rk}$  are the standard deviation of errors for each region and area (1A-3A).

Simulated concentration poverty profile

$$\begin{aligned} \Pr ob[\ln C_i < 0 | i \in U_k, X_i = X_M^k] &= \Phi[-(\alpha_N + \beta_N X_M^k)/\sigma_{Mk}] \\ \Pr ob[\ln C_i < 0 | i \in U_k, X_i = X_U^k] &= \Phi[-(\alpha_N + \beta_N X_U^k)/\sigma_{Uk}] \\ \Pr ob[\ln C_i < 0 | i \in R_k, X_i = X_R^k] &= \Phi[-(\alpha_N + \beta_N X_R^k)/\sigma_{Rk}] \end{aligned} \quad (9A)$$

where

$$\alpha_N = s_M \left( \sum_k s_{Mk} \alpha_{Mk} \right) + s_U \left( \sum_k s_{Uk} \alpha_{Uk} \right) + s_R \left( \sum_k s_{Rk} \alpha_{Rk} \right)$$

and

$$\beta_N = s_M \left( \sum_k s_{Mk} \beta_{Mk} \right) + s_U \left( \sum_k s_{Uk} \beta_{Uk} \right) + s_R \left( \sum_k s_{Rk} \beta_{Rk} \right).$$

### ***Quantile decomposition***

With the method employed above, we impose homogenous returns to characteristics for metropolitan, non-metropolitan urban, or rural sub-populations (and regions using the alternative method). Recent studies have extended the Oaxaca-Blinder decomposition to any quantile of the

distribution of living standards using quantile regressions. We explore the heterogeneity in the returns to characteristics within a sub-population (i.e. metropolitan and non-metropolitan urban and rural areas) of a geographic region by applying the quantile regression decomposition methodology of Machado and Mata (2005) and Nguyen et al. (2007) to explore the distribution of returns and covariate effects for a region.

First, we explore the interquantile difference in coefficients to determine whether returns vary between the 5<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup>, and 95<sup>th</sup> quantiles (numerical estimates available upon request). The standard errors for each of the coefficients are estimated using bootstrap methods that account for the complex survey design of the POF survey.

Next, we decompose the difference between two distributions of log welfare ratios (e.g between metro and urban distributions, or between leading and lagging regions) into two components: the first component accounts for differences in the distribution of returns to characteristics and the second accounts for differences in the distribution of covariates. By using the quantile regression coefficients from one region (area) and covariates from another region (area), we construct a counterfactual welfare distribution. For instance, a counterfactual distribution of rural log welfare ratios where households possess *urban characteristics* but receive *rural returns* can be denoted by  $F(y_R^* | Z_i \in Z^U, \beta_i \in \hat{\beta}_\theta^R)$ . An alternative counterfactual distribution, denoted by  $F(y_U^* | Z_i \in Z^R, \beta_i \in \hat{\beta}_\theta^U)$ , would use *rural characteristics* and *urban returns*. The steps taken to construct the counterfactual are described in detail in Nguyen et al (2007).

The counterfactual distribution is then used to decompose the difference between two sample distributions of log welfare ratios into returns and covariate effects. For any given percentile of a distribution  $\theta$ , we can calculate the difference between the observed urban and rural distributions of welfare and the counterfactuals, that is,

$$y_U(\theta) - y_R(\theta) = [y_U(\theta) - y_R^*(\theta)] + [y_R^*(\theta) - y_R(\theta)]$$

where the first set of brackets represents the returns effect and the second represents the covariate effect. Or equivalently, using an alternative counterfactual distribution, we have

$$y_U(\theta) - y_R(\theta) = [y_U(\theta) - y_U^*(\theta)] + [y_U^*(\theta) - y_R(\theta)]$$

where the first set of brackets represents the covariate effect and the second represents the returns effect. We use both specifications of the decomposition (i.e. using different counterfactual distributions) to see whether our findings are robust.

As noted in Machado and Mata, any unobservable characteristics that are correlated with any of our covariates may produce biases in the coefficient estimates. They point out, “The method only

provides accounting decompositions, conditional on a given model. Thus, changes in unobserved ability or in the labour market institutions, for example, would also be reflected in coefficient changes.”

### *Data*

The 2002-2003 Pesquisa de Orcamentos Familiares (POF) survey in Brazil is a household budget survey designed to measure consumption, expenditures, and income. Unlike other Brazilian surveys, POF is representative at both the national and regional levels for metropolitan, urban (i.e. non-metropolitan urban), and rural areas. The 48,568 households (181,747 individuals) in the POF represented 48,534,638 households (175,331,798 individuals).<sup>5</sup> The regional breakdown of households in our sample is shown in Table 2.1.

**Table 2.1 Sample Households by Region**

	<b>Metro</b>	<b>Urban</b>	<b>Rural</b>	<b>Total</b>
North	2,472	2,452	1,957	6,881
Northeast	5,524	8,921	4,218	18,663
Southeast	2,578	4,254	1,835	8,667
South	1,423	3,666	1,023	6,112
Center West	1,851	4,779	1,615	8,245
<b>Total</b>	<b>13,848</b>	<b>24,072</b>	<b>10,648</b>	<b>48,568</b>

The measure of standard of living that we use is the log of the “welfare ratio”, defined as the nominal per capita consumption deflated by the region-specific poverty line that summarizes the cost of meeting minimum livelihood needs. The 2007 World Bank study on measuring poverty in Brazil estimates separate poverty lines for twenty one different metropolitan, urban, and rural areas of each of Brazil’s five major regions using the Cost of Basic Needs method (see Table 2.2).<sup>6</sup>

---

<sup>5</sup> When the sample size is a small fraction of the population, the finite population correction is close enough to unity and can be ignored (Deaton, pg 43). Thus, for the Brazil POF where 1 out of every 1,000 households were sampled, we chose to ignore the finite population correction for simplicity with the understanding that the calculated standard errors may be slightly larger than if the finite population correction were accounted for.

<sup>6</sup> Specifically the CBN poverty lines used are based on the lower estimate of the adjustment to the food poverty line for basic nonfoods. For a detailed discussion of the construction of the region-specific poverty lines in Brazil see World Bank, 2007. Note that a welfare ratio equal to 1, or equivalently a log welfare ratio equal to 0, represents a household per capita consumption equal to the poverty line.

**Table 2.2: Regional poverty lines and mean expenditures**

Region		Mean Per Capita Expenditures (R\$/month)	Lower Poverty Line (R\$/month)
1	Metro Belem	299.0	105
2	North Urban	238.2	102
3	Rural	135.0	93
4	Metro Fortaleza	309.4	99
5	Metro Recife	331.3	104
6	Northeast Metro Salvador	386.8	108
7	Urban	207.6	100
8	Rural	111.9	92
9	Metro Rio De Janeiro	547.7	107
10	Metro Sao Paulo	525.3	115
11	Southeast Metro Belo Horizonte	429.1	103
12	Urban	381.3	109
13	Rural	207.0	97
14	Metro Curitiba	522.8	105
15	South Metro Porto Alegre	485.0	111
16	Urban	368.3	99
17	Rural	236.9	90
18	Brasilia	596.2	109
19	Center Goiania municipality	425.9	103
20	West Urban	268.5	105
21	Rural	217.7	100
Total		335.9	103

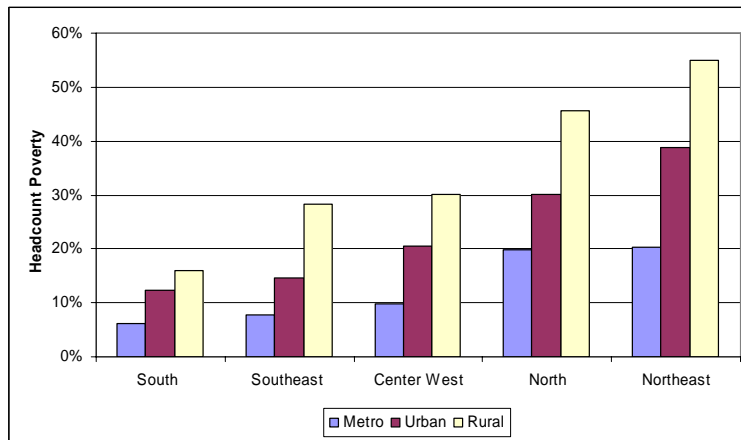
Source: Table 12 in World Bank (2007) and authors' estimates using the 2002-03 POF.

The dependent variables include the number and age of household members (excluding housekeepers and renters), education levels, marital status, ethnicity, gender of the household head, and occupations. Education is categorized into 5 groups: no education, incomplete elementary I (1-3 years), incomplete elementary II (4-7 years), incomplete secondary (8-10 years), and at least secondary completed (11 years or more). The various occupations are included in the following categories: (1) professional or military, (2) technician, (3) administrative services, (4) service workers and vendors

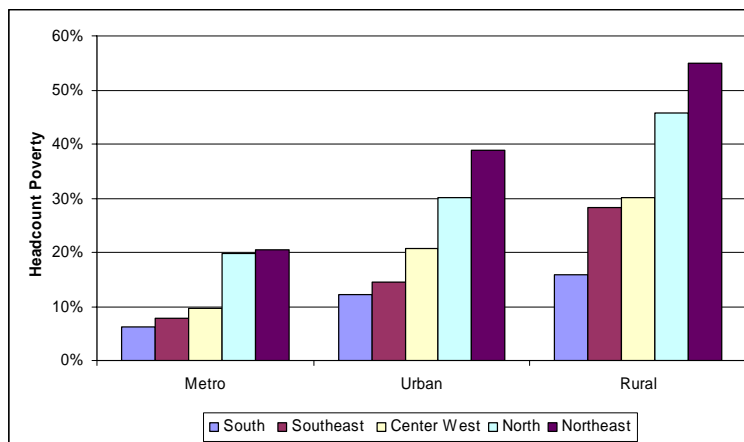
(reference case), (5) agriculture, (6) manufacturing and industrial services, and (7) other occupation or missing. We would have liked to include land ownership variables, but they were not available in the POF. A variable for religion (Roman Catholic) was included initially but was insignificant and subsequently dropped.

Poverty rates vary considerably both within and across regions, but there are clear trends. Within regions, poverty rates are the highest in rural areas and lowest in metropolitan areas (see Figure 2.1). Across regions, the regional pattern of poverty is similar for metropolitan, urban, and rural areas: the Northeast consistently has the highest poverty rates in each of the areas while the South has the lowest (see Figure 2.2). In the next section, we will investigate the factors driving these trends.

**Figure 2.1: Poverty Within Regions**



**Figure 2.2: Poverty Across Regions**



### 3. RESULTS

We now present the results of the decomposition of metro-urban-rural means and quantile regression decomposition. When interpreting the regression results, it is important to keep in mind that the reference household in our analysis is comprised of a single person of mixed race ancestry (“parda”) who has no kids, has no schooling, is a vendor / service worker, and lives in the Southeast region. With the Southeast region being the reference region included in the constant term, the vector  $D$  includes 4 binary variables identifying the North, Northeast, South and Center-West regions of Brazil. The regression results (equations 1-3) are in Table 3.1 and summarized below. Equality of coefficients between metropolitan and urban areas as well as between urban and rural areas can be rejected for all categories of variables (see Appendix for F-values).

*Demographics – household size and structure:* Welfare tends to decrease with larger families (i.e. more dependents) as indicated by the negative coefficients on variables representing the number of household member. In general, additional household members will decrease per capita welfare ratios, but given the functional form ( $\log \text{welfare ratio} = \beta_1(\# \text{ in age group}) + \beta_2(\# \text{ in age group squared})$ ), the incremental change for each additional household member is not as great.

*Other demographics:* Controlling for other factors, households with a head who is White or Asian tend to have a higher welfare than households with heads of other ethnicities. The ethnicity of the head is categorized into three groups: Parda (reference case), White or Asian, or Other (Black, Indigenous, or Missing). Also, controlling for other factors, having a spouse tends to decrease per capita welfare ratios by about 9-13%. Since the coefficients for the age of head and age of head squared are both positive, welfare tends to increase with the age of household. The regression estimates indicate a slight convex shape to the curve but relationship is nearly linear. This is may be due to the generous pension system in Brazil.

*Education:* Education plays an important role in household welfare. *Ceteris paribus*, higher educational attainment is positively correlated to higher welfare. Attainment of at least 11 years of education by the household head tends to increase the welfare ratio by 63% in metro areas, 81% in urban areas, and 60% in rural areas. Having a spouse with education tends to increase household welfare (e.g. by 50% on average with 11 or more years of education), although to a lesser extent than the head. When household members (e.g. children) have more education than either the head or spouse, household welfare also tends to be higher, in particular when the difference is large. For instance, with a difference of 10

years or more, per capita welfare ratio increases by over 50% in metro and urban areas and 38% in rural areas.<sup>7</sup>

*Occupation:* The primary occupation of the household head is classified in one of the following categories: (1) professional or military, (2) technician, (3) administrative services, (4) service workers and vendors (reference case), (5) agriculture, (6) manufacturing and industrial services, and (7) other occupation or missing. There is a large disparity in returns for professional and technical occupations between metro/urban and rural sectors (i.e. 0.548 and 0.477 versus 0.188). And a household head whose primary occupation in agriculture tends to have a lower welfare ratio by about 9% in urban areas and 16% in rural areas (relative to service workers).

### **Decomposition of means**

Expected mean welfare exceeded the poverty line by 177% for metro households, 97% for urban households, and 22% for rural households. *Ceteris paribus*, the welfare gap between metro areas of the Southeast and metro areas of other regions is on average the same as the difference between urban areas (-0.07) and similar to rural differences (-0.10). In other words, controlling for other factors, the mean welfare in metro, urban, and rural areas of the Southeast region is slightly higher than those in other regions. Differences in education contribute considerably to the urban-rural gap (about half of the difference). Household size and structure and the primary occupation of the head in rural areas tends to make them worse off on average than metro or urban households. However, this decomposition cannot distinguish between the difference in returns or in characteristics.

---

<sup>7</sup> Elementary education in Brazil is compulsory and eight years, divided into two four-year cycles. Secondary education is another three years. Thus, we group educational attainment of household members into the following four groups: incomplete elementary I (1-3 years), incomplete elementary II (4-7 years), incomplete secondary (8-10 years), and at least secondary completed (11 years or more).

**Table 3.1 Regressions for Log Welfare Ratios for Metro, Urban and Rural areas of Brazil**

	<b>Metro</b>		<b>Urban</b>		<b>Rural</b>		<b>Pooled</b>	
<b>Constant</b>	0.773	***	0.012		0.469	***	0.291	***
<b>Geographic Regions</b>								
North	-0.137	***	-0.040		0.017		-0.070	***
Northeast	-0.242	***	-0.242	***	-0.248	***	-0.264	***
South	0.029		0.018		0.134	**	0.007	
Center West	-0.081	***	-0.053	**	-0.007		-0.095	***
<b>Demographics</b>								
# age 0-2	-0.378	***	-0.358	***	-0.281	***	-0.339	***
# age 0-2 squared	0.050	***	0.060	***	0.034	**	0.046	***
# age 3-11	-0.339	***	-0.292	***	-0.229	***	-0.297	***
# age 3-11 squared	0.035	***	0.030	***	0.013	***	0.028	***
# age 12-17	-0.245	***	-0.208	***	-0.196	***	-0.220	***
# age 12-17 squared	0.031	***	0.021	***	0.022	***	0.024	***
# age 18-59	-0.171	***	-0.183	***	-0.212	***	-0.176	***
# age 18-59 squared	0.012	***	0.018	***	0.020	***	0.016	***
# age 60+	-0.199	***	-0.155	***	0.004		-0.144	***
# age 60+ squared	0.016		0.035	**	-0.019		0.021	
Female head	-0.060	**	-0.054	**	-0.113	***	-0.058	***
Spouse	-0.163	***	-0.166	***	-0.144	***	-0.192	***
White or Asian	0.171	***	0.158	***	0.108	***	0.149	***
Black, Indigenous, or Missing	-0.020		0.007		-0.029		-0.003	
Age of head	0.018	***	0.032	***	0.019	***	0.026	***
Age of head squared / 100	-0.010	**	-0.023	***	-0.015	***	-0.019	***
<b>Education of head</b>								
1-3 years	-0.051		0.130	***	0.112	***	0.100	***
4-7 years	0.139	***	0.299	***	0.241	***	0.276	***
8-10 years	0.276	***	0.490	***	0.444	***	0.453	***
11+ years	0.629	***	0.806	***	0.605	***	0.800	***
<b>Education of spouse</b>								
1-3 years	0.010		0.080	***	0.096	***	0.079	***
4-7 years	0.095	**	0.181	***	0.198	***	0.174	***
8-10 years	0.189	***	0.302	***	0.351	***	0.279	***
11+ years	0.438	***	0.518	***	0.483	***	0.507	***
<b>Education differential</b>								
1-3 years	0.048		0.005		0.031		0.022	
4-6 years	0.125	***	0.115	***	0.098	***	0.105	***
7-9 years	0.140	***	0.168	***	0.212	***	0.168	***
10+ years	0.531	***	0.545	***	0.385	***	0.550	***
<b>Occupation</b>								
Professional	0.542	***	0.443	***	0.177	***	0.479	***
Technician	0.234	***	0.213	***	0.028		0.204	***
Administrative	0.088	**	0.083	**	-0.006		0.084	***
Agriculture	-0.088		-0.099	***	-0.170	***	-0.122	***
Manufacturing / Industry	0.001		0.027		-0.033		0.006	
Missing / not defined	0.060	**	-0.047	*	-0.222	***	-0.026	
N	13,848		24,072		10,648		48,568	
R-squared	0.583		0.534		0.438		0.568	

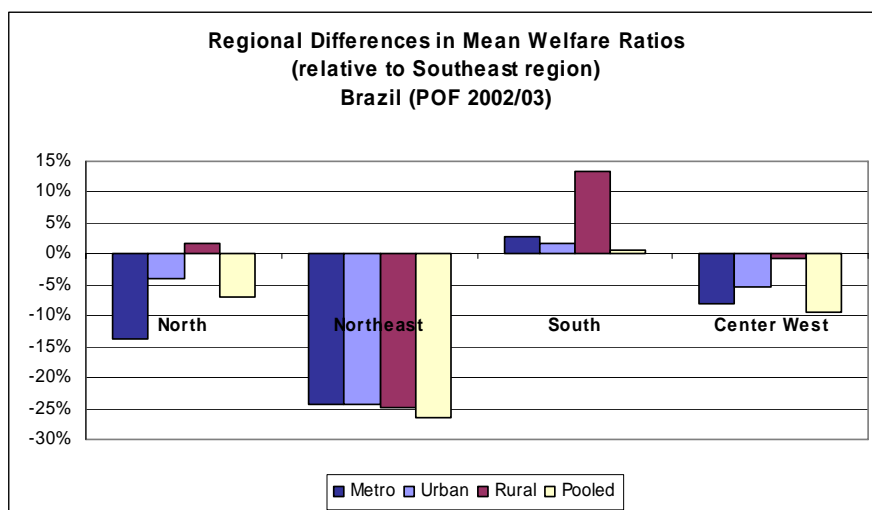
\*p&lt;0.10, \*\* p&lt;0.05, \*\*\*p&lt;.01

The constant term controlling for other factors accounts for more than the entire metropolitan-urban difference in log welfare ratios. For our reference household described above, the welfare ratio is much higher in metropolitan areas relative to both urban and rural areas, and accounts for much of the inter-regional difference.

	<b>Metro</b>	<b>Urban</b>	<b>Rural</b>	<b>M-U difference</b>	<b>U-R difference</b>
<b>Mean log welfare ratio</b>	1.02	0.68	0.20	0.34	0.49
Constant	0.77	0.01	0.47	0.76	-0.46
Geographic dummies	-0.06	-0.07	-0.09	0.00	0.03
HH Composition	-0.77	-0.72	-0.81	-0.06	0.09
HH Demographics	0.52	0.85	0.45	-0.34	0.40
Education	0.47	0.57	0.31	-0.09	0.26
Occupation	0.10	0.04	-0.13	0.06	0.17

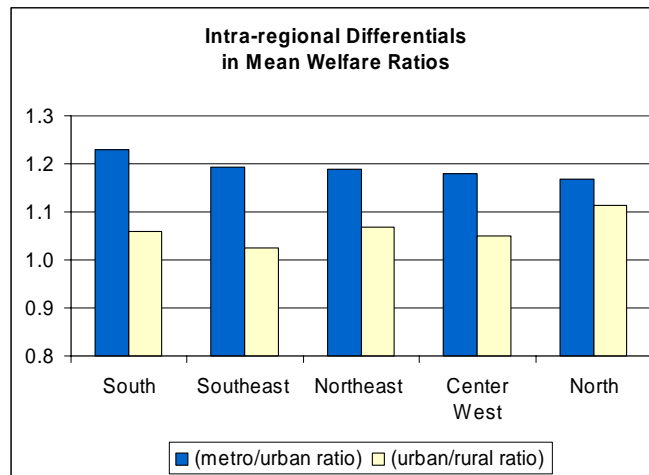
Comparing coefficients for the regional dummy variables in Figure 3.1, the metro, urban, and rural areas of the Northeast region appear to have structural regional disadvantages relative to the Southeast region, the reference region. The rural area of the South has the largest positive regional advantage. The metro and urban areas of the North and Center West have slight disadvantages. All other areas exhibit little difference relative to corresponding areas of the Southeast.

**Figure 3.1: Regional Differences in Mean Welfare Ratios**



In the intra-regional comparison of welfare, we find that the conditional mean welfare gap between metropolitan and urban areas is greater than the urban-rural gap. To control for differences in household characteristics, mean welfare ratios were conditioned on national means of household characteristics. Using this method, expected conditional mean welfare ratios of metropolitan households are 15% to 21% greater than the means of urban households in the same region. As for intra-regional differences between urban and rural areas, the Southeast region exhibits very little difference between the two areas, and the gap in other regions ranges from 3% to 11% (see Figure 3.2).

**Figure 3.2: Intra-regional Differentials**



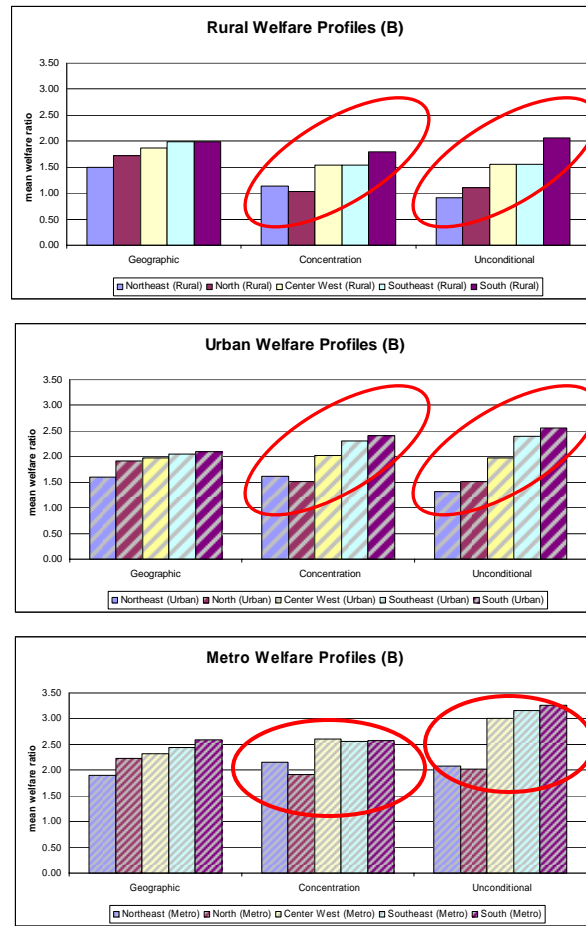
We now turn to the results of the simulation profiles (Figure 3.3) and start with a comparison across regions. First of all, we can see a repeating regional pattern in the geographic and concentration profiles for metro, urban, and rural areas. With the geographic profile, each region's simulated mean welfare ratio is conditioned on national mean characteristics, so the observed pattern indicates that the Northeast and North have the lowest returns and the Southeast and South have the highest. With the concentration profile, we see a similar pattern but for characteristics. Secondly, differences across regions in household attributes / characteristics are more important in explaining the observed regional welfare differences in rural and urban areas. We can see that the concentration profiles in these two areas are quite similar to the unconditional profiles, implying that differences in returns do not play as a big of role in explaining the observed welfare. Along the same lines, conditioning the simulated welfare in the geographic profiles on national mean characteristics tends to make the regional differences more equitable. The correlation coefficients between the geographic profile and actual log welfare ratio (.81-.93) and between the concentration profile and actual log welfare ratio (.93-.96) are high.

We can also compare profiles within regions. Comparing the rural and urban geographic profiles, we find that these profiles are similar (both with a range from ~1.5 to ~2.0) indicating similar returns between the rural and urban areas for each of the regions. When comparing the metro and urban profiles, we find that the welfare ratios for the metro geographic profile are greater by about 0.3 to 0.5 than those in the corresponding urban profile (i.e. larger returns for metro regions). Looking at the concentration profiles, we find that the rural areas have less of the welfare enhancing attributes and that the metropolitan areas have more of them. The mean welfare ratios of the urban concentration profile exceed those in the rural profile by at least 0.5 indicating more favorable attributes in urban populations than in rural ones. We also find substantial differences in attributes for the metro areas of Northeast, North, and Center West relative to urban areas, but smaller differences for Southeast and South regions.

When we compare the metro areas of the lagging Northeast and leading Southeast regions, we find that the difference in welfare ratios in the geographic profile is much larger than the difference in the concentration profile, implying that the difference in returns contributes to the difference in welfare more than differences in attributes. This result is consistent with the quantile results that will be discussed later in this section. It is worth noting that the actual welfare means for the Southeast and South regions are much greater (0.6 to 0.7 higher) than the simulated concentration profile due to the higher returns. One explanation for this large difference would be that agglomeration effects associated with large metropolitan areas is leveraging higher returns.

Since the difference in returns does not vary as much across regions as do the differences in characteristics, the geographic profiles exhibit a much more equitable distribution of welfare ratios relative to the unconditional profiles. The ratio of the variance of the simulated welfare ratios (poverty rates) to the variance of the actual values can be used to indicate the relative importance of the two profiles in determining living standards (Ravallion and Wodon, 1999). Using this summary statistic, concentration profile accounts for 58 percent of the variance of the actual welfare ratios, whereas the geographic profile only accounts for 15 percent.

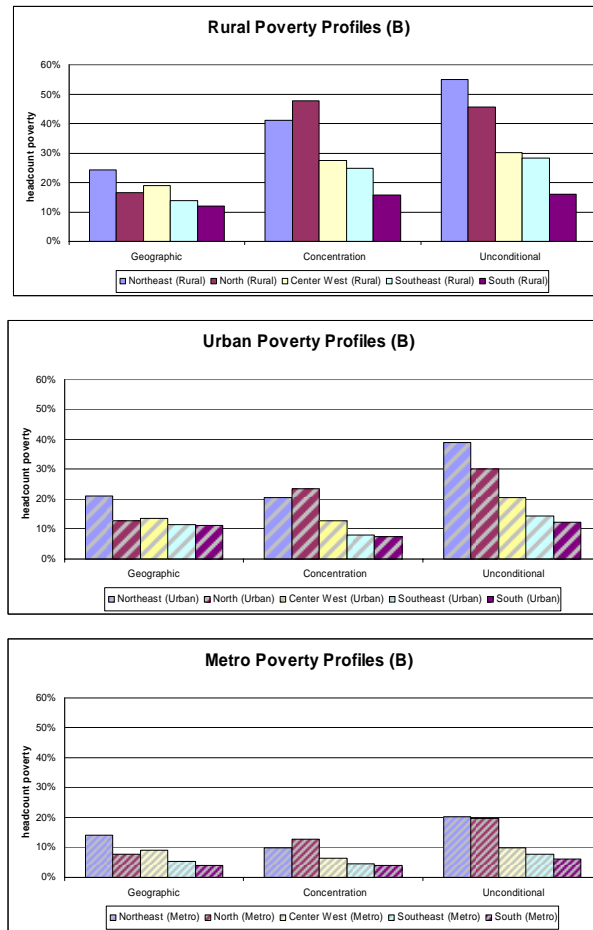
**Figure 3.3: Simulated Welfare Profiles<sup>8</sup>**



Assuming a log normal distribution of welfare ratios, poverty rates are simulated based on the mean welfare ratios for each region and area (see figure 3.3). By the ratio of the variance of the simulated poverty rates to the variance of the actual values, the geographic profile accounts for only 15% of unconditional variance whereas the concentration profile accounts for 85%. Correlation between the simulated profiles and unconditional estimates is high, with correlation coefficients ranging from 0.77 to 0.95.

<sup>8</sup> The welfare profiles in the figure are based on coefficients obtained from separate regressions by sector (M, U, and R) and by region. See figure 5 in the Appendix for a comparison of the welfare profiles obtained using coefficients obtained from separate regressions by sector only (including regional dummies as covariates).

**Figure 3.4: Simulated Poverty Profiles<sup>9</sup>**



The similarity of the concentration and unconditional profiles suggest that regional differences in concentration of poverty related household attributes explain much of the observed regional inequality. However, the previous decomposition cannot capture the variation of characteristics and returns of a distribution. It is possible that the relative importance of the geographic (returns) and concentration (covariate) effects is not constant across a distribution. In the next section, we will dig a bit deeper into these variations.

### Quantile Regression Decomposition

To investigate the relative importance of the returns and covariate effects at various points in a distribution, we utilize quantile regression decompositions as in Machado and Mata (2005) and Nguyen et al. (2007). We first compare differences between Brazil's leading regions (the Southeast) and the lagging

<sup>9</sup> The poverty profiles in the figure are based on coefficients obtained from separate regressions by sector (M, U, and R) and by region. See figure 6 in the Appendix for a comparison of the poverty profiles obtained using coefficients obtained from separate regressions by sector only (including regional dummies as covariates)

region (the Northeast), and then compare differences within each of these regions. The results indicate (i) a large returns effect in accounting for higher welfare in the metro Southeast relative to the metro Northeast or urban Southeast, (ii) a combination of covariate and returns effect in accounting for higher welfare in the metro Northeast relative to the urban Northeast, and in the urban and rural Southeast relative to urban and rural Northeast respectively, and (iii) a dominant covariate effect and little to no returns effect in accounting for higher welfare in urban over rural areas.

***Between NE and SE regions***

The first set of decompositions explores differences between metro, urban, and rural areas of the Southeast and Northeast. The contribution of the returns and covariate effects in accounting for differences in the distribution of welfare are shown in Figure 3.5. As a check on the robustness of the results, for each of these quantile decompositions, two different counterfactual distributions were used. For example, the counterfactual distribution,  $y^*$  below, can be constructed using NE attributes and SE returns as in (a) or SE attributes and NE returns as in (b):

$$(a) \quad y_{SE} - y_{NE} = \underbrace{\left[ y_{SE}(X_{SE} \cdot \beta_{SE}) - y^*(X_{NE} \cdot \beta_{SE}) \right]}_{\text{Covariate effect}} + \underbrace{\left[ y^*(X_{NE} \cdot \beta_{SE}) - y_{NE}(X_{NE} \cdot \beta_{NE}) \right]}_{\text{Returns effect}}$$

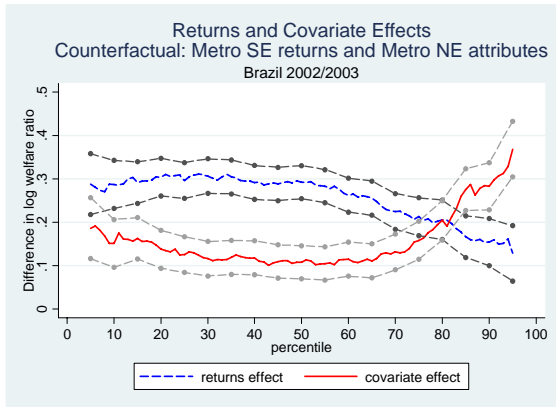
$$(b) \quad y_{SE} - y_{NE} = \underbrace{\left[ y_{SE}(X_{SE} \cdot \beta_{SE}) - y^*(X_{SE} \cdot \beta_{NE}) \right]}_{\text{Returns effect}} + \underbrace{\left[ y^*(X_{SE} \cdot \beta_{NE}) - y_{NE}(X_{NE} \cdot \beta_{NE}) \right]}_{\text{Covariate effect}}$$

In comparing the metro areas of the Southeast and Northeast regions (graph 1), the returns effect is greater than the covariate effect in accounting for the differences between welfare distributions. Considering that metro Southeast (i.e. Sao Paolo and Rio De Janeiro) has a high density of economic activity, better infrastructure, and serves as a hub for trade, a large returns effect suggests the presence of agglomeration effects.

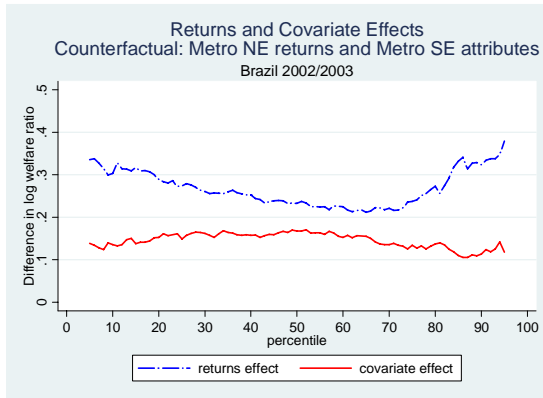
Between urban areas (graph 2) and between rural areas (graph 3) of these two regions, both the covariate and returns affect contribute to the difference. The upward slope of the returns effects indicates that the returns effect is smaller for the poor relative to those better off.

**Figure 3.5: Quantile decomposition of welfare differences Between Southeast and Northeast regions**

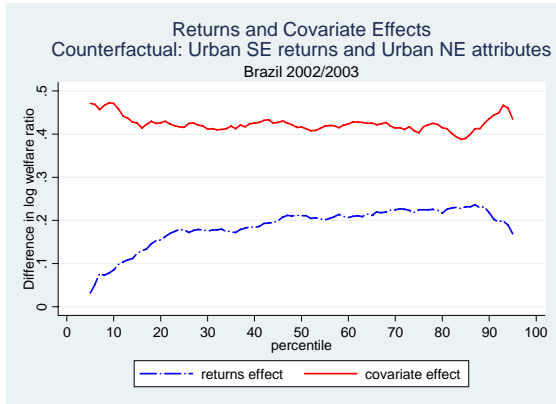
(1a) Between SE and NE metro areas



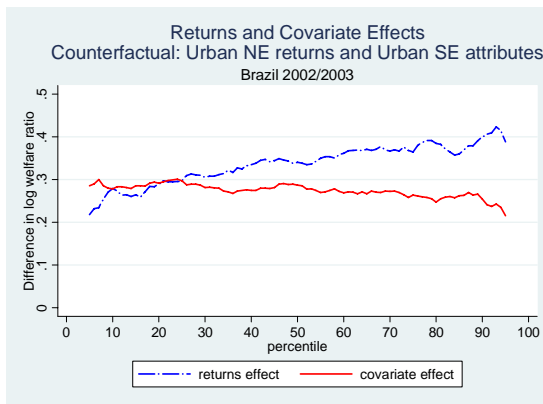
(1b) using alternative counterfactual



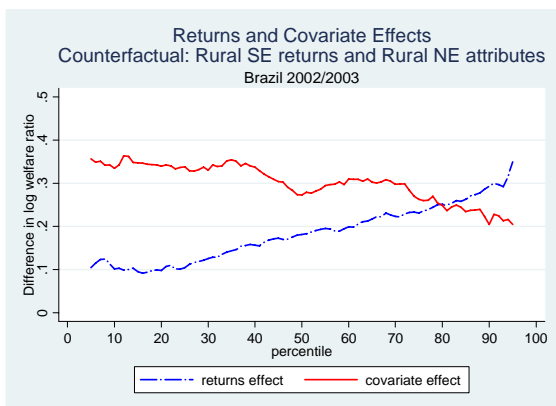
(2a) Between SE and NE urban areas



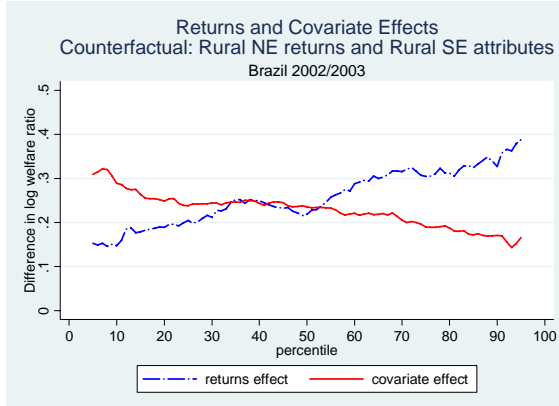
(2b) using alternative counterfactual



(3a) Between SE and NE rural areas



(3b) using alternative counterfactual



### Within NE and SE regions

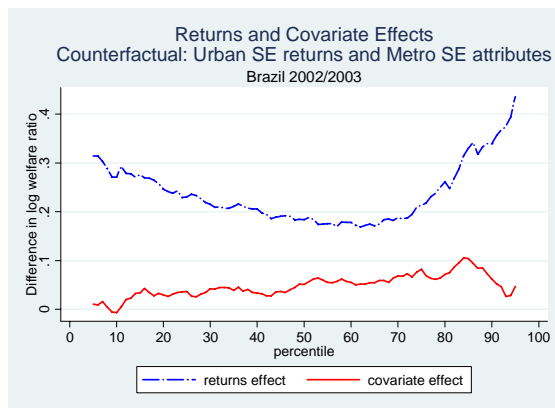
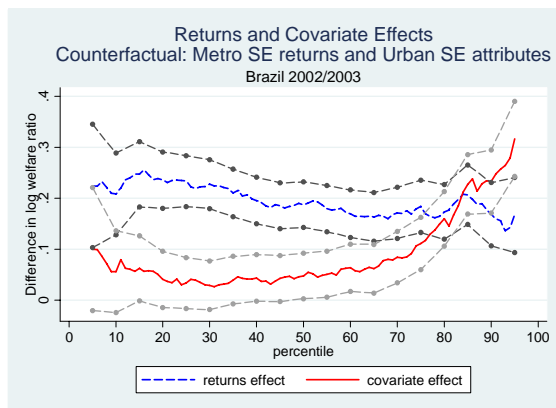
The second set of decompositions explores metro-urban and urban-rural differences within regions (Figure 3.6). Between metro and urban areas, we find that both returns and covariate effects play a role. In the Southeast region (graph 4), the returns effect is dominant throughout most of the distribution, which is consistent with the result above (NE metro vs. SE metro) and further supports the idea of agglomeration effects in the leading metropolitan area.

In the Northeast region, the metro-urban comparison (graph 5) shows that both the returns effect and covariate effects play a role in the observed difference, with the covariate effect at least as large as the returns effect. As in the NE-SE comparison of urban areas or rural (graphs 2 and 3), the returns effect increases as welfare ratio increases, indicating that the returns effect is smaller for the poor relative to those better off. Those in the metro area are able to obtain higher returns, in particular for those better off, but not quite like that in the metro Southeast. These results are similar to that of the NE-SE comparison of urban or rural areas (graphs 2 and 3),

**Figure 3.6: Quantile decomposition of welfare differences Within Southeast and Northeast regions**

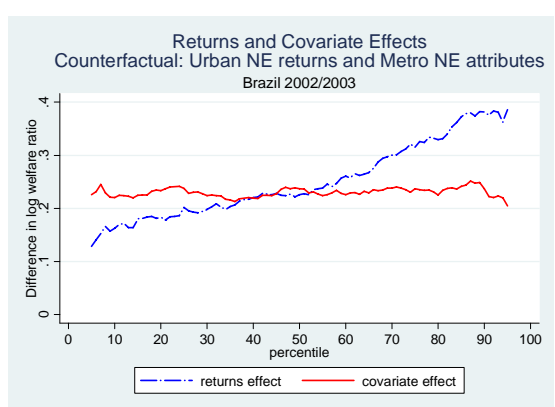
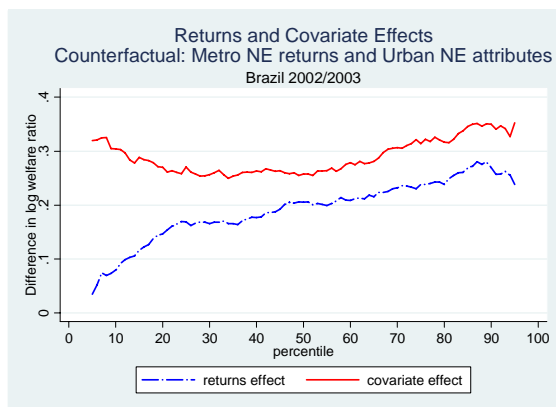
(4a) Between metro and urban areas of SE

(4b) using alternative counterfactual

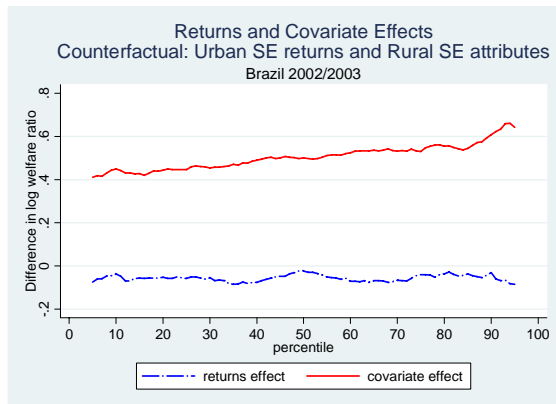


(5a) Between metro and urban areas of NE

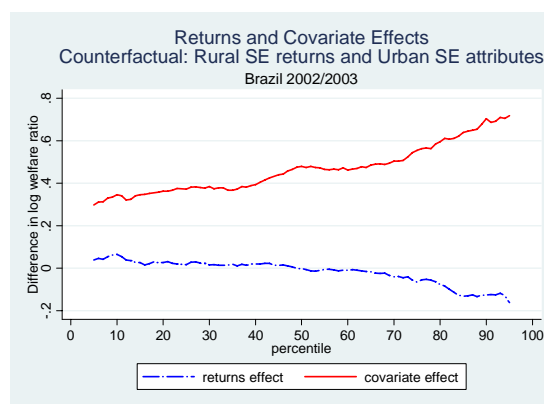
(5b) using alternative counterfactual



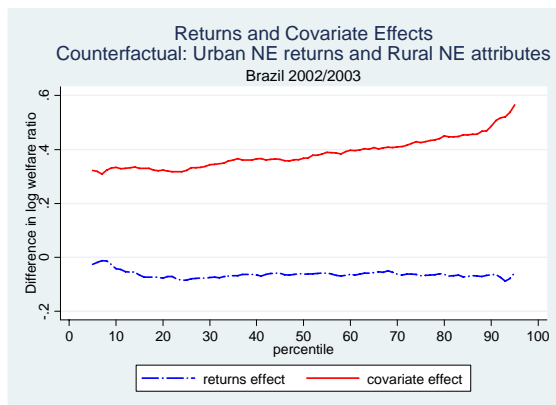
(6a) Between urban and rural areas of SE



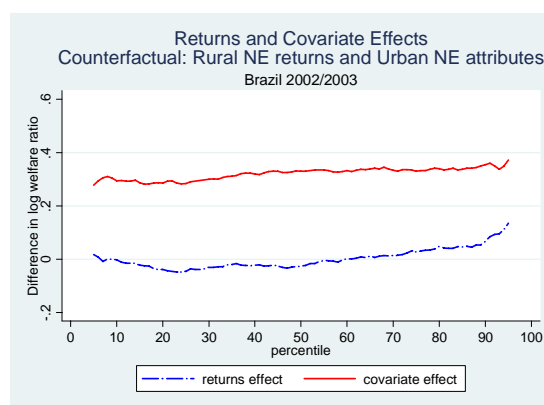
(6b) using alternative counterfactual



(7a) Between urban and rural areas of NE



(7b) using alternative counterfactual



Between urban and rural areas (graphs 6 and 7), we find that most of the welfare disparity is accounted for by the covariate effect, with little difference in returns, throughout the distribution. Thus, households possess more favorable attributes in urban areas vis-à-vis rural areas while the differences in returns are relatively small. These results are consistent with the results from the mean decomposition and with the results using the alternative counterfactual (6b and 7b).

One possible explanation for little to no returns effect between these areas is that migration may be equalizing returns across areas. It may be that migration between urban and rural areas within the Southeast and within the Northeast region involves lower costs/risk (e.g. more likely to have social networks to assist with the transition, less distance from home, temporary employment) than between regions. Relatively cost-free migration is expected to facilitate migration flows and thereby equalize the returns of portable characteristics between the origin and destination areas, as long as there are no agglomeration effects in the destination region. Also, through migration, sorting based perhaps on education or some unobserved ability may occur such that we arrive at the observed concentration of poor people / covariate effects.

#### 4. CONCLUSIONS AND POLICY IMPLICATIONS

Welfare disparities between urban and rural areas within any of the five regions of Brazil as well as between regions are large and continue to remain large in spite of many government programs, at the federal and at the state level, devoted to alleviating such disparities. In the development literature as well as in the policy arena, one encounters two extreme views offering alternative explanations for these spatial disparities. One view is that poor areas arise from the persistent concentration in these areas of individuals with personal attributes that inhibit growth in their living standards. According to this view, otherwise identical individuals will have the same growth prospect independent of where they live. Thus, geography would not play a causal role in explaining the level of and growth in living standards. At the other extreme is the view that geography itself is the cause of the high level of poverty and weak growth of living standards over time. In areas better endowed with local public goods, such as better infrastructure and other basic services (electricity, water and sanitation), there may be geographic externalities that facilitate higher returns and in turn the exit of poor households from poverty. According to this view, given two identical individuals, the one living in an area with lower endowments of these public goods or some other geographically correlated attribute important to productivity may be condemned to stagnation and poverty over time.

In this paper, we classify the variety of factors associated with spatial differences in the standard of living into two broad groups: a set of “covariates” that summarize the portable or non-geographic attributes of the household, such as age, level of education, type of occupation etc., and a set of parameters that summarize the marginal effects or “returns” of these characteristics (either at the mean or at different points of the welfare distribution). Based on this framework, we then address the question of whether the spatial disparities in welfare and poverty are better explained by the sorting of people with low portable characteristics in some areas (e.g. less educated people being concentrated in the rural areas of any given region) or by persistent spatial differences in the returns to portable characteristics such as human capital. This in turn provides more guidance for the design and prioritization of policies aimed to reduce poverty in different areas.

The decomposition of means indicates that the welfare disparities across regions are associated more with the concentration of people with similar observable household attributes than the differences in returns to those attributes. Moreover, in the Northeast region, both household characteristics and returns are much lower than in other regions of Brazil.

Our findings reveal that, except for the Southeast region, the “covariate” effect is the primary explanation for the differences in the standard of living between metropolitan and urban areas or between urban and rural areas within the remaining four regions of Brazil. While returns effects also accounted for

a non-trivial part of the difference in living standards, covariate effects were generally the dominant explanation.

In the Southeast region, the “returns” effect turns out to be the dominant explanation for differences in the standard of living between metropolitan and urban areas. This result is supportive of the existence of agglomeration effects in the metropolitan areas of the Southeast (i.e., Sao Paulo and Rio de Janeiro) due to the high density of economic activity and better infrastructure. However, even in the Southeast region, the covariate effect is the dominant explanation for living standards differences between urban and rural areas.

Another result from our within region analysis is that within the Southeast and the within the Northeast regions we find that differences in returns between urban and rural areas play a very small role in explaining living standards differentials. Thus, most of the welfare disparity between urban and rural areas in both the Southeast and the Northeast is explained almost exclusively by the covariate effect, that is, households in urban areas possess more favorable attributes than households in rural areas.

Our comparisons between metropolitan areas in the Southeast and the Northeast regions provide further validation of the presence of substantial agglomeration effects in the metropolitan areas of the Southeast. These agglomeration effects lead to substantially higher returns to portable assets, such as education, in the metropolitan areas of the Southeast in comparison to the returns in metropolitan areas of the Northeast. Further comparisons between urban areas (and rural areas) of these two regions, suggests that both the covariate and returns affect contribute to the difference, with the covariate effect dominating especially for the poorer households.

The results also shed some light on the role of labor migration on welfare disparities across and within regions of Brazil. On the one hand, the absence of any significant role in the returns effect as an explanation for welfare disparities between urban and rural areas in the Southeast or the Northeast regions suggests that migration of labor between urban and rural areas within regions is able to equalize returns to individual attributes within regions. Therefore, welfare differences between urban and rural regions seem to be primarily due to the sorting or concentration of people with higher attributes in the urban areas of these regions.

On the other hand, the dominant role of the returns effect in explaining living standards between metropolitan areas in the Southeast region and metropolitan areas in the Northeast, suggests that the persistent and large migration of workers from the lagging Northeast region to the metropolitan areas of the Southeast region is not able to diminish the differences in the returns across regions. In fact, migration from the Northeast to the Southeast may also be a leading cause of these welfare inequalities. To the extent that the migration of workers from the Northeast to the Southeast enhances overall productivity and economic growth due to the positive externalities associated with clustering human capital in the

metropolitan areas of the Southeast region, then it should be encouraged and facilitated in spite of the magnitude and apparent persistence of inequality in the living standards in the Northeast region.

Overall the findings of our study validate the recent change in strategy towards poverty alleviation in Brazil encapsulated by the *Bolsa Familia* program. Since the 1950s, government policies in Brazil have been focusing at diminishing regional inequality in Brazil through direct government investments in infrastructure, public credits subsidizing private initiative, and related territorial development program in the Northeast region. The key characteristic of the recent *Bolsa Familia (BF)* program is that it considers the lack of sufficient human capital rather than geography as the primary cause of extreme poverty, and it employs monetary and in-kind benefits as instruments for encouraging poor families to invest in the education, health and nutrition of their children. In the *BF* program, as in other conditional cash transfer (CCT) programs, geographic targeting is only a means to finding the areas where these poor households are likely to be located. Overall, the new emphasis of the Brazilian government towards investment in human capital, as exemplified by the *Bolsa Familia* program, represents a major step towards the right direction not only in the fight against poverty but also towards reducing spatial disparities in welfare in the long-run.

Although quite tentative, the inferences regarding the role of migration in Brazil, also suggest a set policies complementary to the *BF* program. As long as some that people migrate out of the Northeast because of push rather than pull factors, such as limited access to or low quality of basic social services such as health and education, then programs focusing on the Northeast region the should be concentrating at increasing access to and quality of these basic services. More empirical evidence on the determinants of migration between the lagging and leading regions of Brazil, along the lines of Lall, Timmins and Yu (2008) can be particularly helpful in guiding the design and nature of government interventions that can enhance both equity and efficiency (i.e. increase equality of opportunity in the lagging regions as well as aggregate productivity).

## References

- Bourguignon, Francois, Francisco Ferreira, and Phillippe Leite. 2002. "Beyond Oaxaca-Blinder: Accounting for Differences in Household Income Distributions across Countries." World Bank Policy Research Working Paper 2828.
- Deaton, Angus. 1997. *The Analysis of Household Surveys*.
- Duarte, Angelo Jose Mont'Alverne, Pedro Cavalcanti Ferreira, and Marcio Antonio Salvato. 2003. "Regional or educational disparities? A counterfactual exercise" *Ensaio Economicos da EPGE* No 532.
- Efron, Bradley, and Robert J. Tibshirani. 1993. *Introduction to the Bootstrap*.
- Guimarães, Juliana Ferraz, Tiago Cavalcanti, and Raul Silveira Neto. (2006). "Accounting for Labor Income Differences in Brazil: the role of Human Capital" [Anais do XXXIV Encontro Nacional de Economia \[Proceedings of the 34th Brazilian Economics Meeting\]](#) 35, ANPEC - Associação Nacional dos Centros de Pósgraduação em Economia [Brazilian Association of Graduate Programs in Economics].
- Hendricks, Wallace, and Roger Koenker. 1992. "Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity," *Journal of the American Statistical Association*, **87**(417): pp 58-68.
- Jalan, Jyotsna and Martin Ravallion (2002) "Geographic Poverty Traps? A Micro Model of Consumption Growth in Rural China," *Journal of Applied Econometrics*, **17**: pp. 329-346.
- Kanbur, Ravi. 2006. "The policy significance of inequality decompositions" *Journal of Economic Inequality*, 4(3): pp 367-374.
- Koenker and Bassett. 1978. "Regression Quantiles," *Econometrica*, **46**(1): pp 33-50.
- Koenker and Hallock. 2001. "Quantile Regression," *Journal of Economic Perspectives*, **15**: pp 143-156.
- Krugman, P. 1998. "What's New about the New Economic Geography." *Oxford Review of Economic Policy* 14(2): 7-17.
- Lall, Somik, and Z. Shalizi. 2003. "Location and Growth in the Brazilian Northeast." *Journal of Regional Science*, 43(4): 663-681.
- Lall, S., C. Timmins, S. Yu (2008) "Moving to Opportunity? Successful Integration of Bright Lights," Presentation at World bank, June 2008.
- Machado, Jose A. F., and Jose Mata, 2005. "Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression", *Journal of Applied Econometrics*, **20**: 445 – 465.
- Nguyen, Binh T., James W. Albrecht, Susan B. Vroman, and M. Daniel Westbrook, 2007. "A Quantile Regression Decomposition of Urban-Rural Inequality in Vietnam", *Journal of Development Economics*, **83**: 466 – 490.
- Oaxaca, Ronald, L. (1973) "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, vol. 14, pp. 693-709.
- Oaxaca Ronald, and Michael Ransom 1994) "On Discrimination and the Decomposition of Wage Differentials," *Journal of Econometrics*, vol. 61. pp. 2-21.
- Ravallion, Martin. 2004. "Pro-Poor Growth: A Primer." World Bank Policy Research Working Paper 3242.

- Ravallion, Martin, and Quentin Wodon (1999) "Poor Areas, or Only Poor People?" *Journal of Regional Science*, vol. 39, no. 4, pp. 689-711.
- Venables, A. 2005. "Spatial disparities in developing countries: cities, regions, and international trade." *Journal of Economic Geography* 5(1): 3-21.
- World Bank. 2003. "Brazil: Inequality and Economic Development" Washington, D.C.: The World Bank (Report# 24487-BR).
- World Bank. 2005. *World Development Report 2006: Equity and Development*. New York: The World Bank and Oxford University Press.
- World Bank (2007). "Brazil: Measuring Poverty Using Household Consumption" Washington, D.C.: The World Bank (Report # 36358-BR).

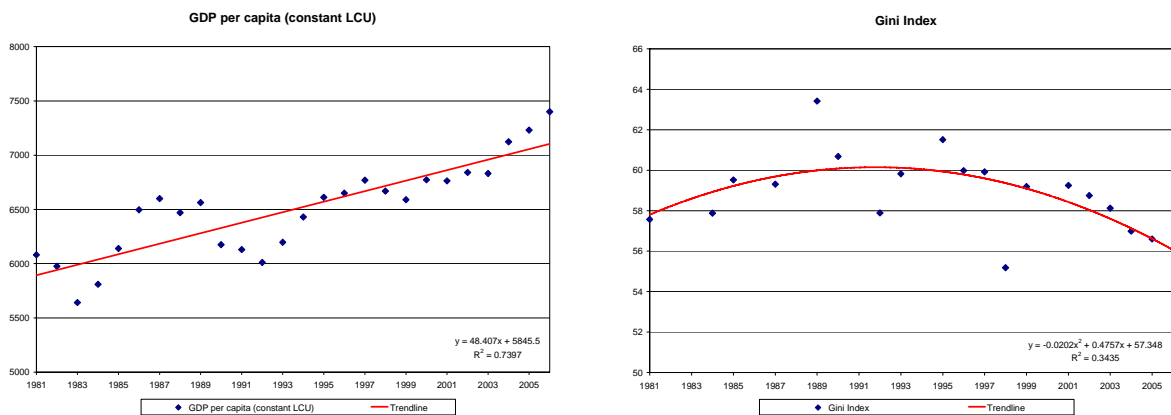
## Appendix -1

Figure 1: Major Regions and States of Brazil



Source: <http://gosouthamerica.about.com>

Figure 2 : GDP per capita and Gini Index: 1981-2005



Source: WDI 2008

**Table 1: Descriptive Statistics of Regional Distribution of Population and Poverty**

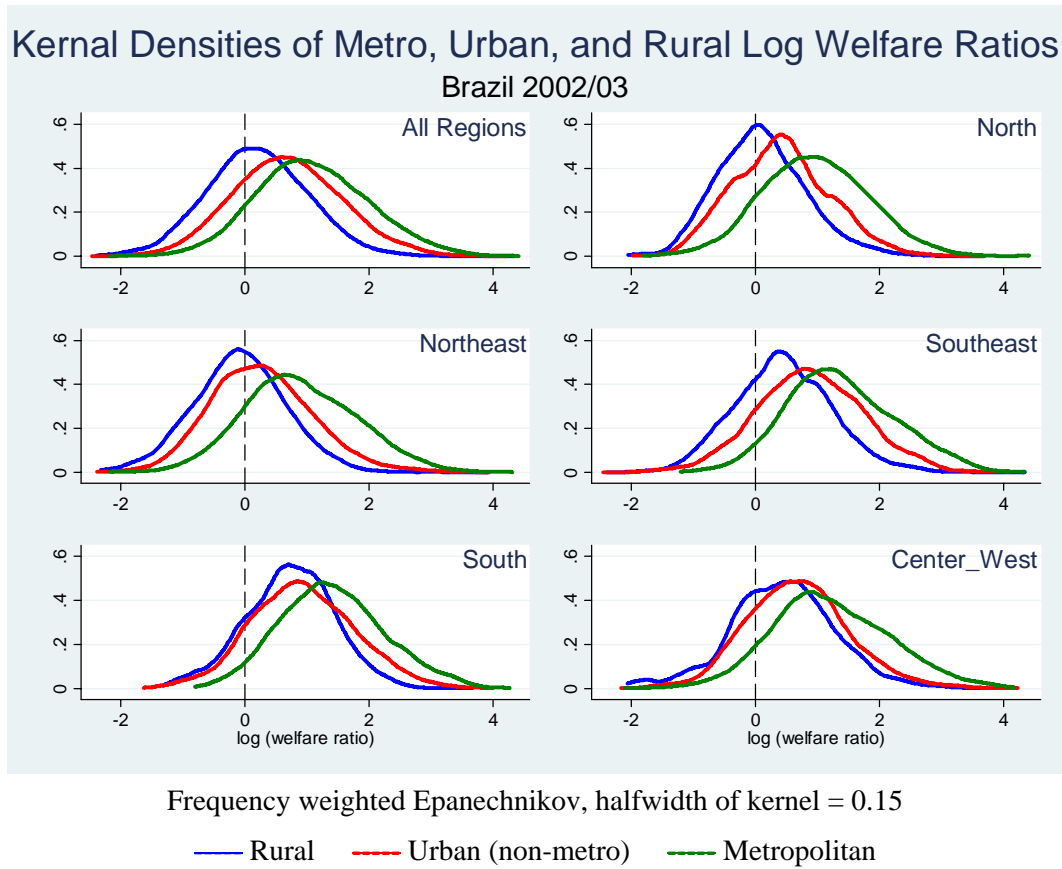
Region			Total Population	Population Share	Number of Poor	Distribution of the Poor	Headcount Poverty	Poverty Gap	Poverty Severity
1	North	Metro Belem	1,845,708	1.05%	307,931	0.8%	16.7%	4.2%	1.7%
2		Urban	8,229,439	4.69%	2,252,145	6.0%	27.4%	8.9%	4.0%
3		Rural	3,533,713	2.02%	1,615,266	4.3%	45.7%	16.3%	7.7%
4	Northeast	Metro Fortaleza	2,985,823	1.70%	620,452	1.6%	20.8%	6.4%	2.7%
5		Metro Recife	3,331,278	1.90%	588,597	1.6%	17.7%	5.4%	2.5%
6		Metro Salvador	3,088,893	1.76%	444,994	1.2%	14.4%	4.4%	2.1%
7		Urban	25,579,176	<b>14.59%</b>	9,420,720	<b>25.0%</b>	36.8%	13.3%	6.4%
8		Rural	13,940,461	<b>7.95%</b>	7,664,897	<b>20.3%</b>	55.0%	22.3%	12.0%
9	Southeast	Metro Rio De Janeiro	11,052,249	6.30%	970,581	2.6%	8.8%	2.3%	0.9%
10		Metro Sao Paulo	17,696,179	<b>10.09%</b>	1,278,930	<b>3.4%</b>	7.2%	1.7%	0.7%
11		Metro Belo Horizonte	4,437,346	2.53%	316,681	0.8%	7.1%	1.4%	0.5%
12		Urban	35,016,773	<b>19.97%</b>	5,045,726	<b>13.4%</b>	14.4%	4.7%	2.2%
13		Rural	6,586,851	3.76%	1,859,310	4.9%	28.2%	9.4%	4.4%
14	South	Metro Curitiba	2,641,166	1.51%	125,209	0.3%	4.7%	1.2%	0.4%
15		Metro Porto Alegre	3,663,574	2.09%	263,083	0.7%	7.2%	1.6%	0.5%
16		Urban	15,083,301	8.60%	1,825,518	4.8%	12.1%	3.6%	1.6%
17		Rural	4,438,516	2.53%	704,361	1.9%	15.9%	4.7%	2.0%
18	Center West	Brasilia	2,151,035	1.23%	173,966	0.5%	8.1%	1.5%	0.4%
19		Goiania municipality	1,121,683	0.64%	57,457	0.2%	5.1%	1.4%	0.6%
20		Urban	7,526,053	4.29%	1,722,950	4.6%	22.9%	7.2%	3.3%
21		Rural	1,382,581	0.79%	427,710	1.1%	30.9%	10.9%	5.7%
Total			175,331,798	100.00%	37,686,485	100.0%	21.5%		

Source: Authors' estimates using the 2002-03 POF; World Bank (2007).

**Table 2 : Poverty by Geographic Regions**

	Poverty Headcount Rate	Distribution of the Poor	Distribution of Population
	2002	2002	2002
Urban	17.5	67.4	82.9
Rural	41.0	32.6	17.1
North	30.7	11.1	7.8
Northeast	38.3	49.7	27.9
Southeast	12.7	25.1	42.7
South	11.3	7.7	14.7
Center West	19.6	6.3	6.9
Total	21.5	100.0	100.0

**Figure 3: Distributions of welfare ratios by region**



Since a welfare ratio of one, or equivalently a log welfare ratio equal to zero, means that an individual is living at the poverty line, the area to the left of the dashed vertical line in Figure 2.X represents the population living in poverty.

**Table 3. Tests of Equality of Coefficients between Metro, Urban, and Rural Regressions**

Year 2002-2003	Restrictions	Metro=Urban		Urban=Rural	
		F-value	F-test (1 percent level)	F-value	F-test (1 percent level)
<b>Nongeographic Variables</b>					
HH Structure (hhmem)	10	14.46	Rejected	24.03	Rejected
HH Demographics (hhdem)	5	15.39	Rejected	24.83	Rejected
Education	12	25.84	Rejected	20.89	Rejected
Occupation	6	5.83	Rejected	12.13	Rejected
<b>Geographic dummies</b>	4	9.75	Rejected	3.91	Rejected

**Table 4a. Simulated Welfare Ratios by Region and by Metro/Urban/Rural Areas**

Geographical region	Geographic Profile			Concentration Profile			Mean Welfare Ratios (log)		
	Metro	Urban	Rural	Metro	Urban	Rural	Metro	Urban	Rural
Southeast (region 3)	0.91	0.73	0.65	0.95	0.85	0.45	1.15	0.88	0.45
North (region 1)	0.77	0.69	0.67	0.66	0.43	0.05	0.70	0.41	0.10
Northeast (region 2)	0.66	0.49	0.41	0.78	0.50	0.15	0.73	0.27	-0.09
South (region 4)	0.93	0.75	0.79	0.95	0.89	0.61	1.18	0.94	0.72
Center West (region 5)	0.82	0.68	0.65	0.97	0.72	0.45	1.10	0.68	0.44
<b>Correlation with mean log welfare ratios</b>	<b>0.89</b>	<b>0.84</b>	<b>0.85</b>	<b>0.95</b>	<b>0.96</b>	<b>0.93</b>			
Var ()	0.02			0.08			0.1441		
Var (sim) / var (actual)	0.13			0.57					

**Table 4b. Simulated Welfare Ratios by Region and by Metro/Urban/Rural Areas  
(using regressions by region and sector)**

Geographical region	Geographic Profile			Concentration Profile			Mean Welfare Ratios (log)		
	Metro	Urban	Rural	Metro	Urban	Rural	Metro	Urban	Rural
North (region 1)	0.80	0.65	0.54	0.65	0.41	0.03	0.70	0.41	0.10
Northeast (region 2)	0.64	0.47	0.40	0.77	0.48	0.13	0.73	0.27	-0.09
Southeast (region 3)	0.89	0.71	0.69	0.94	0.83	0.43	1.15	0.88	0.45
South (region 4)	0.95	0.74	0.68	0.94	0.88	0.59	1.18	0.94	0.72
Center West (region 5)	0.84	0.68	0.63	0.96	0.70	0.43	1.10	0.68	0.44
<b>Correlation with mean log welfare ratios</b>	<b>0.81</b>	<b>0.89</b>	<b>0.93</b>	<b>0.95</b>	<b>0.96</b>	<b>0.93</b>			
Var ()	0.02			0.08			0.14		
Var (sim) / var (actual)	<b>0.15</b>			<b>0.58</b>					

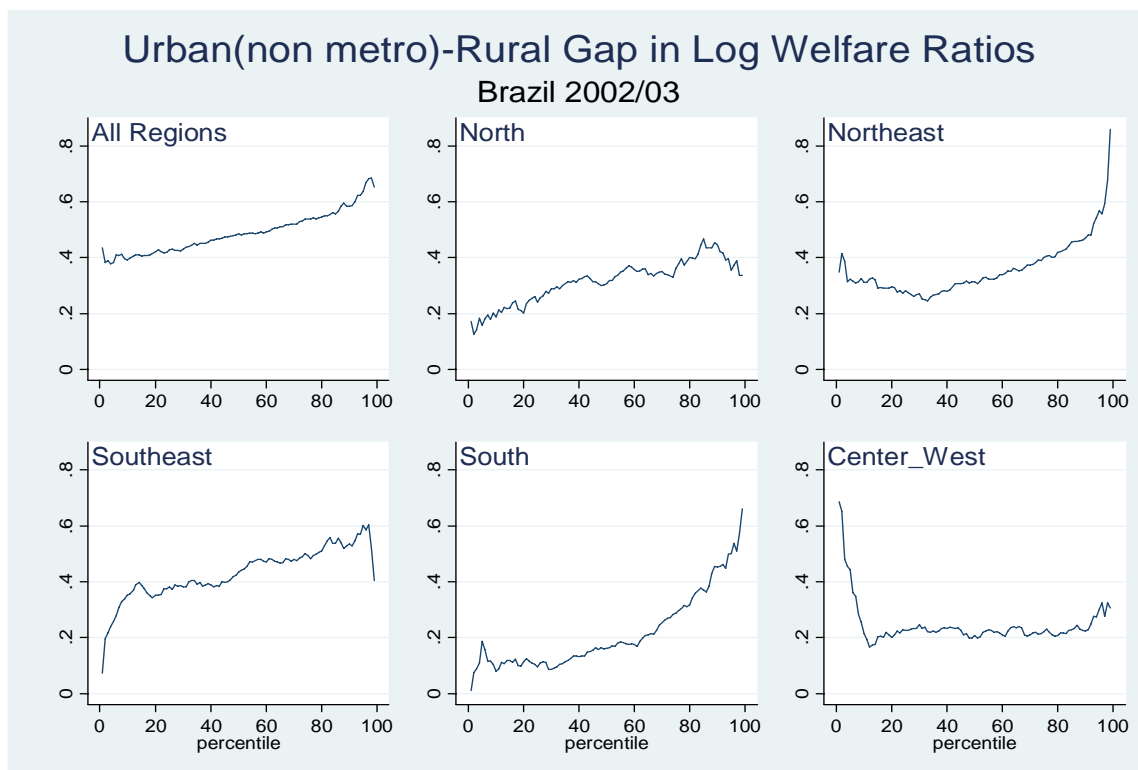
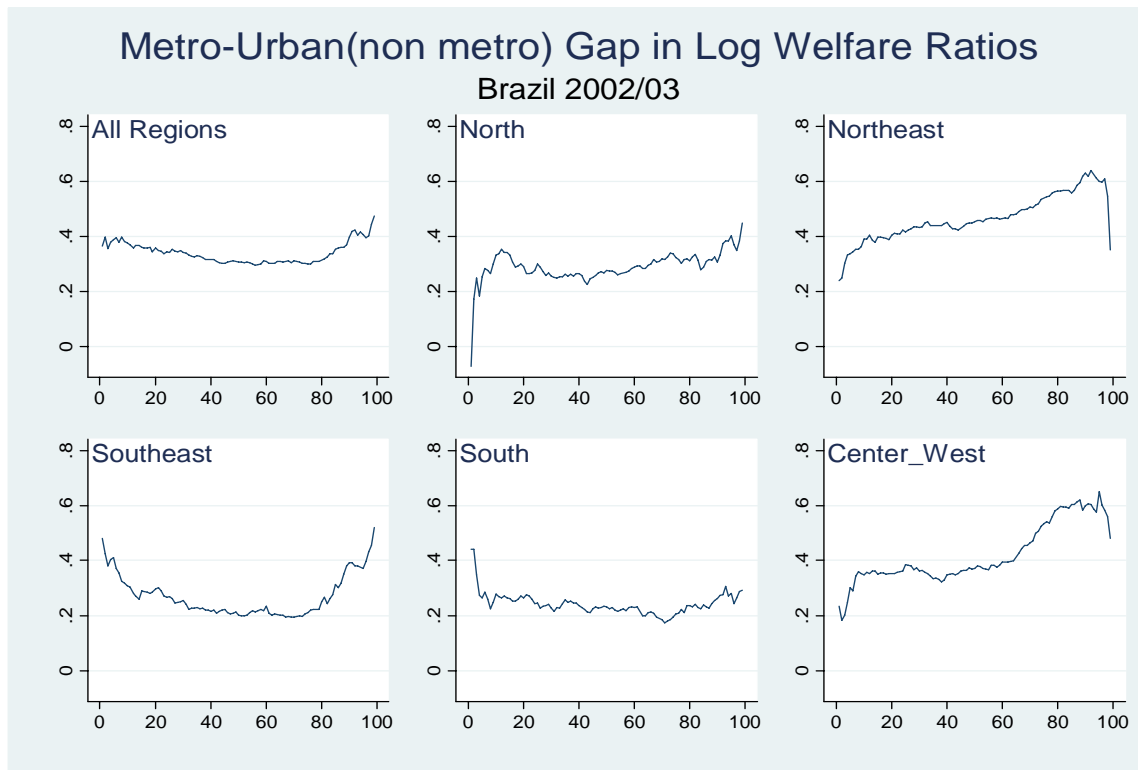
**Table 5a. Simulated Poverty Measures by Region and by Metro, Urban, and Rural Areas**

Year	Geographic Profile			Concentration Profile			Headcount Poverty		
	POF 02-03 2002-2003			POF 02-03 2002-2003			POF 02-03 2002-2003		
	Metro	Urban	Rural	Metro	Urban	Rural	Metro	Urban	Rural
<b>Geographical region</b>							P0	P0	P0
Southeast (region 3)	0.06	0.11	0.14	0.05	0.08	0.23	0.08	0.14	0.28
North (region 1)	0.09	0.13	0.13	0.12	0.24	0.46	0.20	0.30	0.46
Northeast (region 2)	0.12	0.21	0.25	0.09	0.20	0.40	0.20	0.39	0.55
South (region 4)	0.05	0.11	0.10	0.05	0.07	0.16	0.06	0.12	0.16
Center West (region 5)	0.07	0.13	0.14	0.04	0.12	0.23	0.10	0.21	0.30
<b>Correlation with headcount poverty (P0) estimates</b>									
	0.91	0.88	0.83	0.89	0.90	0.92			
Var ()	0.0028			0.0163			0.0205		
Var (sim) / var (actual)	0.14			0.80					

**Table 5b. Simulated Poverty Measures by Region and by Metro, Urban, and Rural Areas**

<b>(using regressions by region and sector)</b>									
Year	Geographic Profile			Concentration Profile			Headcount Poverty		
	POF 02-03 2002-2003			POF 02-03 2002-2003			POF 02-03 2002-2003		
	Metro	Urban	Rural	Metro	Urban	Rural	Metro	Urban	Rural
<b>Geographical region</b>							P0	P0	P0
North (region 1)	0.08	0.13	0.17	0.13	0.24	0.48	0.20	0.30	0.46
Northeast (region 2)	0.14	0.21	0.24	0.10	0.21	0.41	0.20	0.39	0.55
Southeast (region 3)	0.05	0.12	0.14	0.05	0.08	0.25	0.08	0.14	0.28
South (region 4)	0.04	0.11	0.12	0.04	0.08	0.16	0.06	0.12	0.16
Center West (region 5)	0.09	0.14	0.19	0.06	0.13	0.27	0.10	0.21	0.30
<b>Correlation with headcount poverty (P0) estimates</b>									
	<b>0.77</b>	<b>0.87</b>	<b>0.84</b>	<b>0.95</b>	<b>0.91</b>	<b>0.92</b>			
Var ()	0.0031			0.0173			0.0205		
Var (sim) / var (actual)	<b>0.15</b>			<b>0.85</b>					

**Figure 4: Metro-Urban and Urban-Rural differences in log welfare ratios by region**



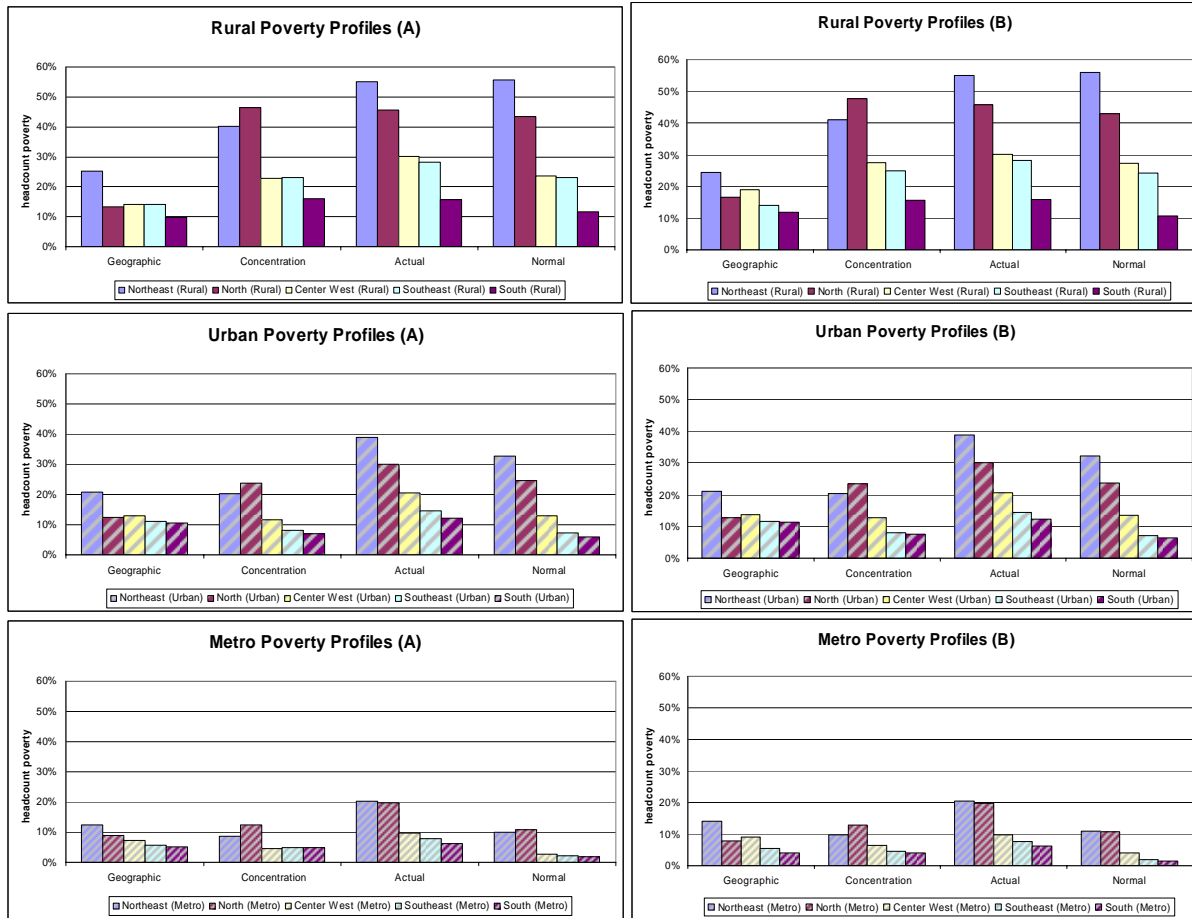
**Figure 5: Simulated Welfare Profiles Using Different Models**



(A) uses coefficients from regressions by sector (as in Ravallion and Wodon), and (B) uses coefficients from regressions by sector and region.

**Figure 6 (alternative): Simulated Poverty Profiles Using Different Models**

(with additional simulated poverty profile “Normal” – where poverty is estimated using actual means and assuming a normal distribution)



(A) uses coefficients from regressions by sector (as in Ravallion and Wodon), and (B) uses coefficients from regressions by sector and region.