Public Disclosure Authorized

Public Disclosure Authorized

WPS4803

Policy Research Working Paper 4803

Sources of Welfare Disparities across and within Regions of Brazil

Evidence from the 2002–03 Household Budget Survey

Emmanuel Skoufias
Roy Katayama

The World Bank
Poverty Reduction and Economic Management Network
Poverty Reduction Group
December 2008

Policy Research Working Paper 4803

Abstract

Brazil's inequalities in welfare and poverty across and within regions can be accounted for by differences in household attributes and returns to those attributes. This paper uses Oaxaca-Blinder decompositions at the mean as well as at different quantiles of welfare distributions on regionally representative household survey data (2002–03 Household Budget Survey). The analysis 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.

This paper—a product of the Poverty Reduction Group, Poverty Reduction and Economic Management Network—is part of a larger effort in the department to analyze poverty and monitor and evaluate the effectiveness of poverty reduction programs. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The author may be contacted at eskoufias@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Produced by the Research Support Team

**Sources of Welfare Disparities across and within Regions of Brazil:
Evidence from the 2002-03 Household Budget Survey**

Emmanuel Skoufias
Roy Katayama

The World Bank (MC4-415)
1818 H Street NW
Washington DC, 20433

JEL classification: O15, O18, R10, R23, R58, I31
*Keywords:*Brazil, Leading and Lagging Regions, Welfare, Poverty, Oaxaca-Blinder decompositions.

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.
This paper is part of a regional World Bank study in the LAC region financed by the LAC Chief Economist office and the PREM Poverty anchor unit. 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 São 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[1]. 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) account 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.

[1] 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 (*Pesquisa 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 decomposition of mean welfare differentials and the quantile regression decomposition along with a brief summary of the 2002-03 Pesquisa 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, we outline the methodology for our investigation of the factors behind Brazil's spatial disparities in the standard of living. We being with a brief discussion of the measure used for the standard of living of households and their members and then summarize the Oaxaca-Blinder methodology used to decompose differences in mean welfare within regions (i.e. between metro and urban areas, and between urban and rural areas) and between regions (i.e. across metro, across urban, and across rural areas). The Oaxaca-Blinder decomposition allows us to estimate the relative contributions of differences in household characteristics and in returns in accounting for differences in living standards. Next, we describe the quantile regression decomposition methods used to determine the relative importance of covariate and returns effects at different quantiles of the distributions. Lastly, we describe the data used in our analysis.

Comparing Living Standards Within and Across Regions

The measure of standard of living that we use in our analysis is the welfare ratio[2]. The welfare ratio is the nominal consumption expenditure per capita deflated by the appropriate region-specific poverty line. The region-specific poverty lines are assumed to incorporate all the cost of living differences faced by the poor in different regions and areas.[3] As a result, this allows us to make comparisons both within and across regions.

There are both conceptual and pragmatic reasons why consumption expenditures available from household surveys are preferable for the purpose of poverty and inequality analysis to an indicator such as household income. It is argued, for example, that consumption expenditures reflect not only what a household is able to command based on its current income, but also whether a 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) to smooth consumption. In this way, a consumption measure is thought to provide a better picture of a household's longer run standard of living than a measure of current income. Furthermore, consumption expenditures for the poor are often better captured than household incomes. 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.

[2] The welfare ratio and its theoretical properties is discussed by Blackorby and Donaldson (1987). More practical applications of the welfare ratio in the measurement of poverty can be found in Ravallion (1998) and Deaton and Zaidi (2002).

[3] Since the poverty lines are constructed based on the cost of basic needs (CBN) approach, which in essence is a Laspeyres price index with fixed weights, the welfare ratio is also analogous to "real expenditures".

Oaxaca-Blinder Type Decomposition

We classify the variety of determinants of the welfare ratio into two broad groups: a set of "covariates" that summarize the portable or non-geographic attributes of the household, denoted by the vector X , and a set of structural parameters, denoted by the vector β that summarize the marginal effects or "returns" to these household attributes. The variables in vector X 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.

Specifically, given any two sub-populations, A and B, defined by the region (and area) of residence, we assume that the logarithm of the welfare ratio of each sub-population, denoted by $\ln C$ can be summarized by the linear regression

$$\ln C_A = \beta_A X_A + \varepsilon_A, \text{ and} \qquad (1)$$

$$\ln C_B = \beta_B X_B + \varepsilon_B, \qquad (2)$$

where ε is a random disturbance term with the usual properties, for summarizing the influence of all other factors on the standard of living.

In this specification, the "returns" to characteristics β summarize the influence of a variety of factors on the standard of living for the sub-population living in a particular region (and area). Basic infrastructure, and ease of access to markets and other basic services are some of the most important of these factors. In addition, returns to characteristics are also affected by the role of institutions, social customs and other cultural factors that are typically too difficult to quantify.

Based on the specifications above (1 and 2), and given that estimated regression lines always cross through the mean values of the sample, the mean difference in the standard of living between groups A and B can then be expressed as

$$\overline{\ln C_A} - \overline{\ln C_B} = \beta_A \overline{X_A} - \beta_B \overline{X_B} \qquad (3)$$

where the bar over the relevant variables denotes the sample mean values of the respective variables. We have assumed that $E(\varepsilon_j) = 0$ for $j = \{A, B\}$.

After adding and subtracting the term $\beta_B \overline{X_A}$ to the above differences in (3), we can express the difference above as

$$\overline{\ln C_A} - \overline{\ln C_B} = \beta_A \overline{X_A} - \beta_B \overline{X_B} + \beta_B \overline{X_A} - \beta_B \overline{X_A} \Rightarrow$$

$$\overline{\ln C_A} - \overline{\ln C_B} = (\overline{X_A} - \overline{X_B})\beta_B + (\beta_A - \beta_B)\overline{X_A} \text{, or} \tag{4}$$

$$\Delta(\overline{\ln C}) = (\Delta\overline{X})\beta_B + (\Delta\beta)\overline{X_A} \tag{4a}$$

Alternatively, if one were to add and subtract the term, $\beta_A \overline{X_B}$, the difference in (3) could be expressed as

$$\overline{\ln C_A} - \overline{\ln C_B} = (\overline{X_A} - \overline{X_B})\beta_A + (\beta_A - \beta_B)\overline{X_B} \text{, or} \tag{5}$$

$$\Delta(\overline{\ln C}) = (\Delta\overline{X})\beta_A + (\Delta\beta)\overline{X_B} \tag{5a}$$

Both expressions (4) and (5) imply that the differential in the mean log welfare ratios between regions A and B, can be decomposed into two components: a component that consists of the differences in average characteristics summarized by the term $\Delta\overline{X}$ and another component that is due to the differences in the coefficients or returns to characteristics in different regions (and areas) of a country summarized by the term, $\Delta\beta$. This is the decomposition method first proposed by Oaxaca (1973) and Blinder (1973).

The decompositions given by expressions (4) and (5) are equally valid. The only difference between them lies in how the differences in the characteristics $\Delta\overline{X}$ and the differences in coefficients $\Delta\beta$ are weighted. In expression (4) the differences in the characteristics $\Delta\overline{X}$ are weighted by the returns to characteristics in region B, whereas the differences in the returns $\Delta\beta$ are weighted by the average characteristics of households in region A. In contrast, in expression (5) the differences in the characteristics $\Delta\overline{X}$ are weighted by the returns to characteristics in region A, whereas the differences in the returns $\Delta\beta$ are weighted by the average characteristics of households in region B.

Since the original decompositions by Oaxaca and Blinder, there have been numerous papers extending the method by proposing alternative weights for the differences in the characteristics $\Delta\overline{X}$ and the differences in returns $\Delta\beta$, (e.g., Reimers, 1983; Cotton, 1988; and Neumark, 1988).[4] The general expression allowing for these alternative weights is

$$\overline{\ln C_A} - \overline{\ln C_B} = \left(\overline{X_A} - \overline{X_B}\right)\left(D\beta_A + (I-D)\beta_B\right) + \left(\beta_A - \beta_B\right)\left((I-D)\overline{X_A} + D\overline{X_B}\right), \qquad (7)$$

where I is the identity matrix and D is a matrix of weights. The traditional Oaxaca-Blinder decompositions can be considered to be special cases, in which D=0 yields (4) and D=1 yields (5). In addition to using D=0 and D=1, we have followed Reimers (1983) and used as weights the average of the coefficients and the average of the characteristics, that is, the diagonal of the D matrix = 0.5.

At this point, it is important to point out that the use and interpretation of the decomposition method discussed above involves a number of caveats. First of all, these decompositions are simple descriptive tools that provide a useful way of summarizing the role of endowments and returns in explaining existing welfare differentials. For this reason, we refrain from attributing causality to either endowments or returns for the welfare differences between or within regions.

Secondly, the variables that we use in the vector *X* above are composed only of portable non-geographic household characteristics. Our specification intentionally excludes infrastructure and access to basic services. The influence of infrastructure, as well as other omitted variables, is captured by default by the estimated coefficients of the portable characteristics of the household. As the formula for omitted variable bias suggests, the estimated coefficients of the household characteristics can be considered as including the direct effect of the omitted variables (such as infrastructure, local institutions and other household variables possibly correlated with the location of the household) on welfare and their correlation with the included household characteristics.

Thirdly, the decomposition results may be biased because of the presence of selection bias. To the extent there is free internal migration within and between different regions, then the current place of residence may not be exogenous. The role of selection bias in the decomposition results was explored, as in Ravallion and Wodon (1999), and the decomposition results did not change significantly after correcting for selection bias (see Appendix A).

Lastly, the decomposition formula in equation (7) holds only at the mean of the two regions being compared, and the decompositions are performed at only one point in time. The extent to which the results of these decompositions change substantially over time is a question that is worth looking into. The potential variation at different points of distributions is examined through quantile decompositions.

---

[4] For a comprehensive summary, see O'Donnell, et al. (2008) or Ben Jann (2008). In our study, the decompositions employed are done using the Stata command "oaxaca" written by Ben Jann using the "weight ()" option.

Quantile Regression Decomposition

Since the role of household characteristics and returns can vary across distributions, we look beyond averages and explore differences across the entire spectrum of welfare distributions. We explore the heterogeneity in characteristics and returns by applying the quantile regression decomposition methodology used in recent studies (Machado and Mata, 2005; Nguyen et al., 2007; Shilpi, 2008) that extend the Oaxaca-Blinder decomposition to any quantile of the distribution of living standards. We focus on the leading Southeast and lagging Northeast regions of Brazil, and estimate the relative importance of the returns and covariate effects in accounting for welfare differences across distributions.

While the basic idea is the same as traditional Oaxaca-Blinder type decompositions, the quantile regression decomposition technique requires the construction of a counterfactual distribution to separate covariate and returns effect. Following Nguyen et al. (2007), we construct a counterfactual distribution by running quantile regressions at each percentile (θ) for sub-population A to estimate coefficient vectors ($\hat{\beta}_\theta^A$). Each of these coefficient vectors is then used to generate fitted values ($y_i^{*BA_\theta}$) of the natural logarithm of welfare ratios using the covariates of sub-population B (X^B) for each household i. From each set of fitted values, a sample of 100 values is randomly selected with replacement, and these are combined to yield a counterfactual welfare distribution of households that possess Group B's characteristics but receive Group A's returns; the distribution can be denoted by $F(y^{*BA} \mid X_i \in X^B, \beta_i \in \hat{\beta}_\theta^A)$.

The counterfactual distribution is then used to decompose the difference between two welfare distributions, for example, between metro areas of leading and lagging regions, or between metro and urban distributions of the same region. For any given quantile (q) of the distributions, we can estimate the covariate effect that accounts for differences in household characteristics and the returns effect that accounts for differences in returns and the constant terms. The decomposition can be expressed as:

$$\text{(a)} \quad y_A(q) - y_B(q) = \underbrace{\left[y_A(q, X_A, \beta_A) - y^{*BA}(q, X_B, \beta_A)\right]}_{\textit{Covariate effect}} + \underbrace{\left[y^{*BA}(q, X_B, \beta_A) - y_B(q, X_B, \beta_B)\right]}_{\textit{Returns effect}}$$

where the first set of brackets represents the covariate effect and the second represents the returns effect.

Since an alternative counterfactual distribution, denoted by $F(y^{*AB} \mid X_i \in X^A, \beta_i \in \hat{\beta}_\theta^B)$, could be constructed such that households possess Group A's characteristics and receive Group B's returns, an alternative specification of the decomposition would be:

$$\text{(b)} \quad y_A(q) - y_B(q) = \underbrace{\left[y_A(q, X_A, \beta_A) - y^{*AB}(q, X_A, \beta_B)\right]}_{\textit{Returns effect}} + \underbrace{\left[y^{*AB}(q, X_A, \beta_B) - y_B(q, X_B, \beta_B)\right]}_{\textit{Covariate effect}}$$

where the first set of brackets represents the returns effect and the second represents the covariate effect. We use both specifications to see whether results are sensitive to the choice of the counterfactual distribution used.

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).[5] The regional breakdown of households in our sample is shown in Table 2.1 below.

Table 2.1 Sample Households by Region

	Metro	**Urban**	**Rural**	**Total**
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
Total	**13,848**	**24,072**	**10,648**	**48,568**

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).[6]

[5] 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.
[6] 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)	Mean Welfare Ratio
1	North	Metro Belem	299.0	105	2.8
2		Urban	238.2	102	2.3
3		Rural	135.0	93	1.5
4	Northeast	Metro Fortaleza	309.4	99	3.1
5		Metro Recife	331.3	104	3.2
6		Metro Salvador	386.8	108	3.6
7		Urban	207.6	100	2.1
8		Rural	111.9	92	1.2
9	Southeast	Metro Rio De Janeiro	547.7	107	5.1
10		Metro São Paulo	525.3	115	4.6
11		Metro Belo Horizonte	429.1	103	4.2
12		Urban	381.3	109	3.5
13		Rural	207.0	97	2.1
14	South	Metro Curitiba	522.8	105	5.0
15		Metro Porto Alegre	485.0	111	4.4
16		Urban	368.3	99	3.7
17		Rural	236.9	90	2.6
18	Center West	Brasilia	596.2	109	5.5
19		Goiania municipality	425.9	103	4.1
20		Urban	268.5	105	2.6
21		Rural	217.7	100	2.2
	Total		335.9	103	

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

While mean welfare ratios and poverty rates vary considerably both within and across regions, there are clear patterns. Within regions, mean welfare ratios are the lowest in rural areas, as one might expect, and highest in metropolitan areas (see Table 2.2). Likewise, the poverty rates are the highest in the rural areas and lowest in metropolitan areas (see Figure 2.1). Across regions, the Northeast and North exhibit the lowest welfare ratios while the Southeast, South, and Center West regions have the higher welfare ratios (Table 2.2.). The regional pattern of poverty is similar for metropolitan, urban, and rural areas, that is, 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 behind these trends.

Figure 2.1: Poverty Within Regions

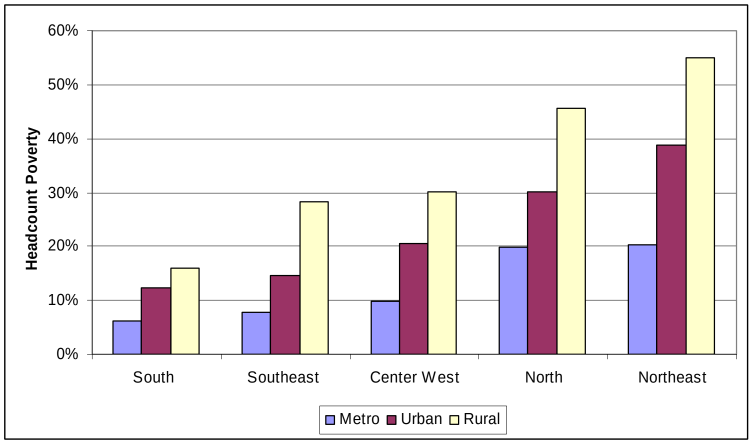
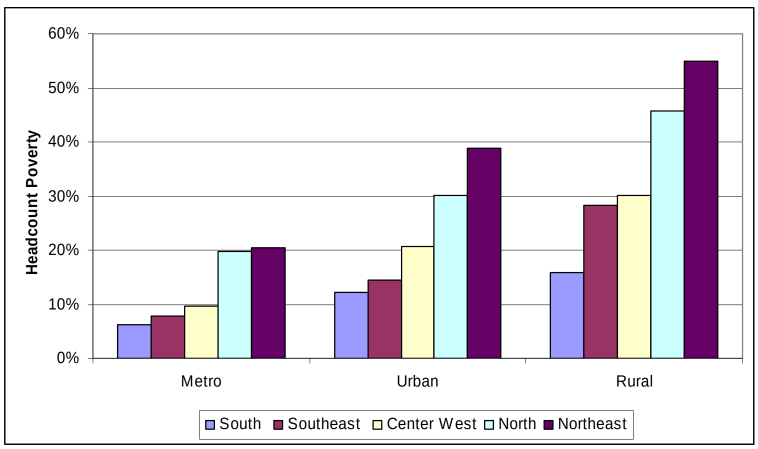


Figure 2.2: Poverty Across Regions



3. RESULTS

In this section, we begin with a discussion of regression results for metropolitan, urban, and rural areas. Then, we present the results of the Oaxaca-Blinder decompositions of differences in means consumption expenditures within and across regions. Lastly, we discuss the quantile regression decompositions, within and between the Northeast and Southeast regions.

Regression Results

The regression results for metropolitan, urban, and rural areas are presented in Table 3.1 and summarized below. 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, four binary variables identifying the North, Northeast, South and Center-West regions were included in the regression. It should be noted the Oaxaca-Blinder decompositions discussed subsequently are not based on the coefficients presented in Table 3.1 since the groups used in the decompositions are divided not only by regions but by (metro, urban, and rural) areas as well. Also, the equality of coefficients between metropolitan and urban areas, and between urban and rural areas, can be rejected for all categories of variables (see Appendix A).

Region of residence: Comparing coefficients for the regional dummy variables, 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.

Demographics (household size and structure): Welfare tends to decrease as family size increases (i.e. higher dependency ratio) 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 welfare ratio = β_1(# in age group) + β_2(# 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 a generous pension system.

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

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

*p<0.10, ** p<0.05, ***p<.01

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.[7]

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).

Comparing the mean log welfare ratios across the different areas (table 3.2), we find that the mean welfare exceeded the poverty line by 177% in metro areas, 97% for urban areas, and 22% for rural areas. The largest negative factor is the household size and composition, and the largest positive factors include education and other demographics. Also, the constant term accounts for a substantial part of the expected mean welfare for metro and rural areas. However, these factors do not distinguish between the difference in returns and in characteristics.

Table 3.2: Contributing Factors to Average Levels of Living and Metro-Urban and Urban-Rural Disparities

	Metro	**Urban**	**Rural**
Mean log welfare ratio	1.02	0.68	0.20
Constant	0.77	0.01	0.47
Geographic dummies	-0.06	-0.07	-0.09
HH Composition	-0.77	-0.72	-0.81
HH Demographics	0.52	0.85	0.45
Education	0.47	0.57	0.31
Occupation	0.10	0.04	-0.13

[7] Elementary education is compulsory and eight years, divided into two four-year cycles. Secondary education is another three years.

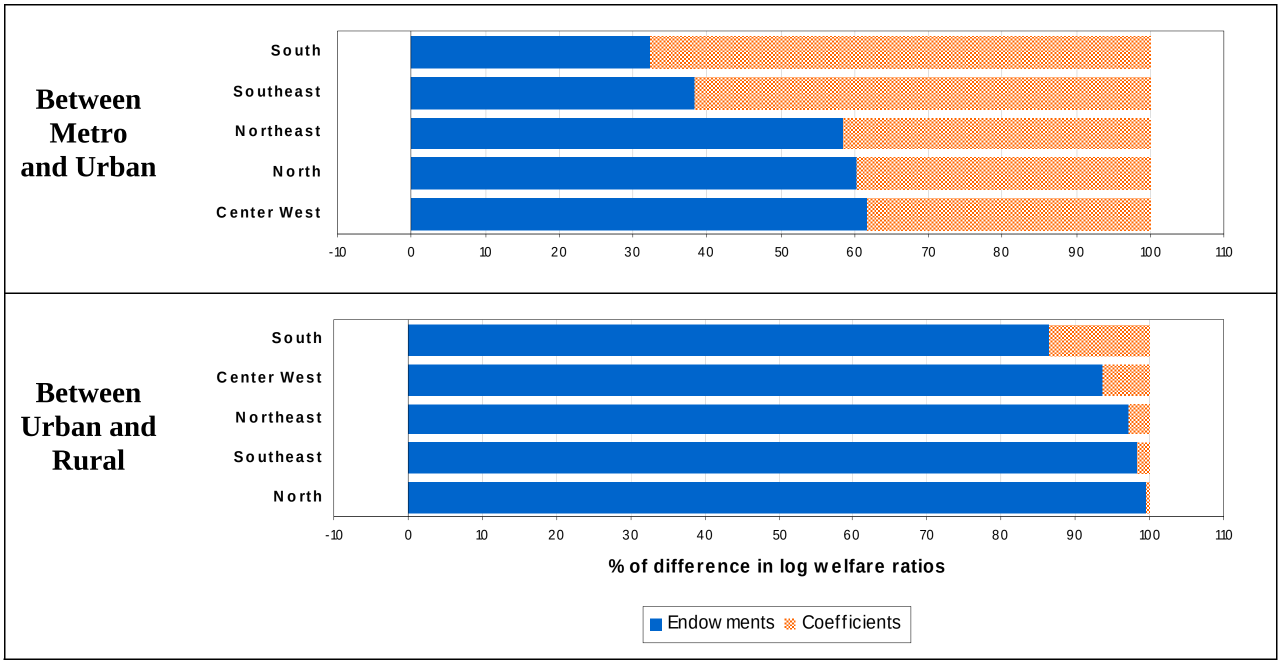
Oaxaca-Blinder Decomposition - Within Regions

In explaining metropolitan-urban welfare differentials, both characteristics and returns play a major role. Oaxaca-Blinder decompositions[8], shown in the top panel of figure 3.1, indicate that in the South and Southeast regions, returns account for about 60 to 70 percent of the log welfare difference, whereas in the North, Northeast, and Center West regions, characteristics account for about 60% of the log welfare difference.

In explaining urban-rural welfare differentials within regions, differences in household characteristics seem to be the dominant explanation. Oaxaca-Blinder decomposition results, shown in the bottom panel of figure 3.1, indicate that characteristics account for nearly all of the urban-rural differences. The results from poverty simulations, following Ravallion and Wodon (1997), are consistent with these findings (see Appendix B).

It should be noted that the results shown in figures 3.1 and 3.2 are for decompositions using a weighting matrix of D=0.5, as in Reimers (1983). Results for other weighting specifications (i.e. D=0 and D=1) are included in Appendix C. While the choice of the weighting specification can shift these results slightly, the relative importance of either the characteristics or returns components in accounting for welfare differences is fairly robust to the choice of the weighting matrix *D*.

Figure 3.1: Oaxaca-Blinder decompositions of welfare differential <u>within</u> regions

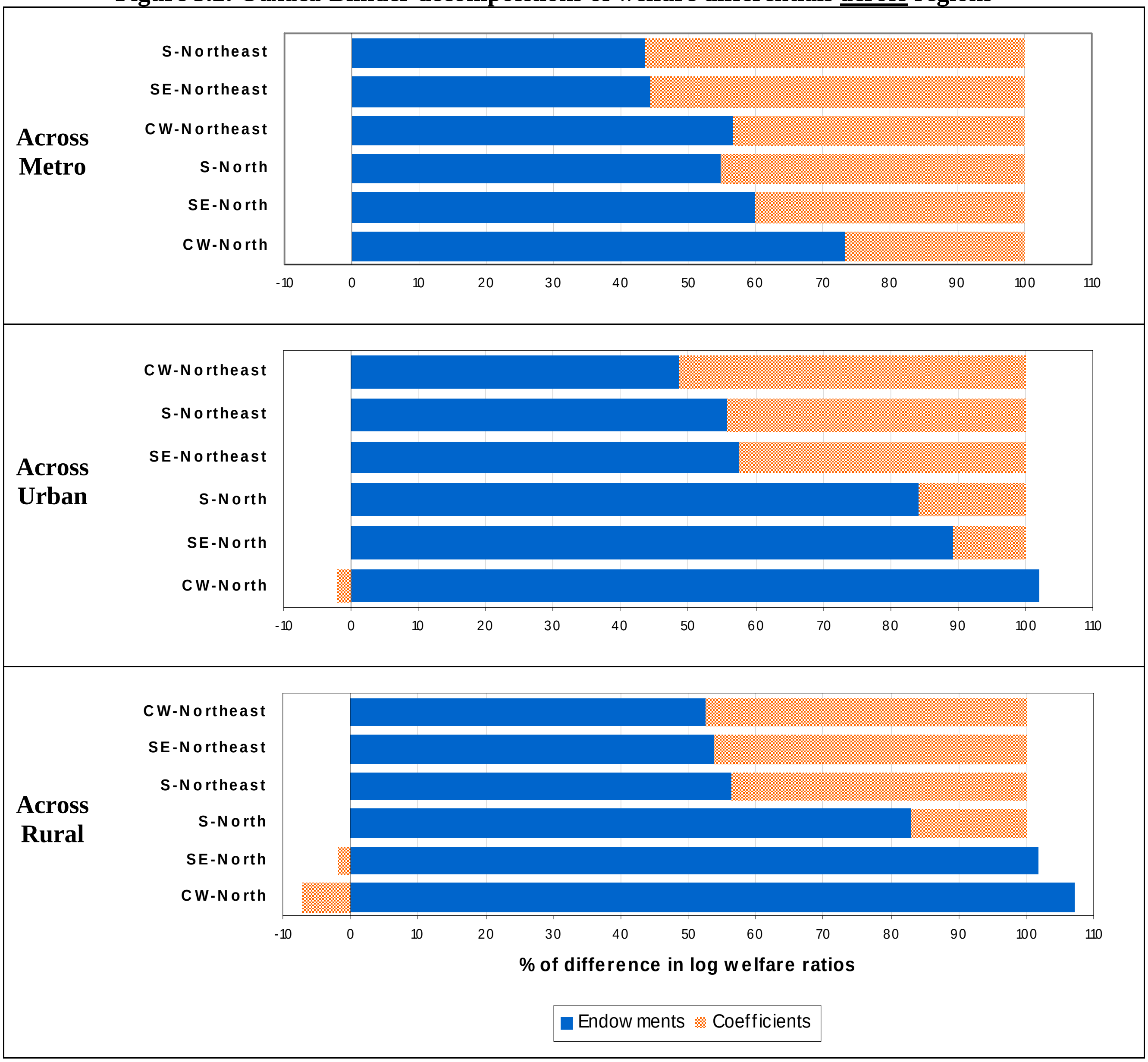


[8] The decomposition was done as in Reimers (1983) using the average of the two groups' coefficients for the so-called "nondiscriminatory" coefficient vector.

Oaxaca-Blinder Decomposition – Across Regions

Although both household characteristics and returns play important roles in accounting for welfare differences between regions, characteristics tend to be a slightly greater factor in most cases, in particular when comparing metropolitan areas. When we compare the lagging Northeast and North regions with the leading southern regions, the Oaxaca-Blinder decomposition results indicate that 45 to 75 percent of the difference in log welfare ratios is explained by differences in household characteristics (figure 3.2). However, it should be noted that in comparisons involving the Northeast region, both returns and characteristics are about equally important, with returns playing a slightly more dominant role in Northeast-Southeast and Northeast-South comparisons of metro areas.

Figure 3.2: Oaxaca-Blinder decompositions of welfare differentials across regions

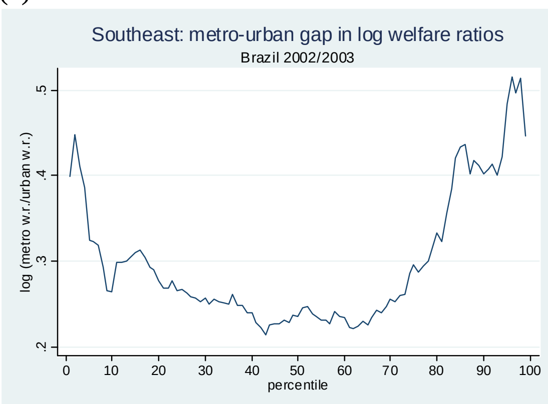


Differences in Welfare Distributions

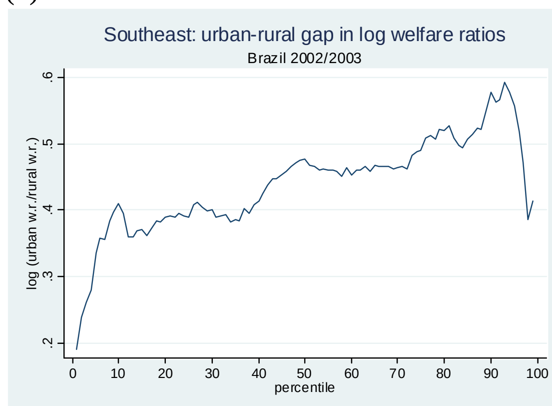
We begin with a set of figures (figure 3.3) summarizing the distribution if welfare differences within the leading Southeast region and within the lagging Northeast region of Brazil and specifically between metropolitan and urban areas and between urban and rural areas.

Figure 3.3: Welfare differences within the Southeast and Northeast regions

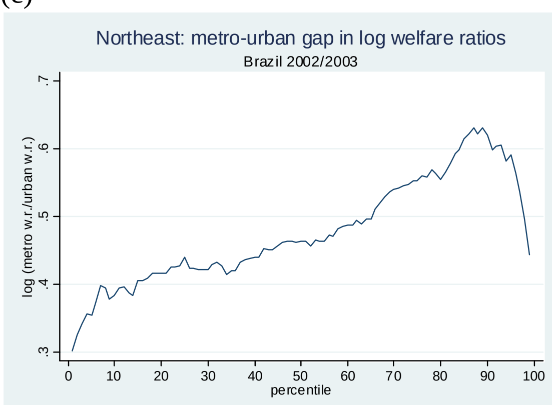
(a)



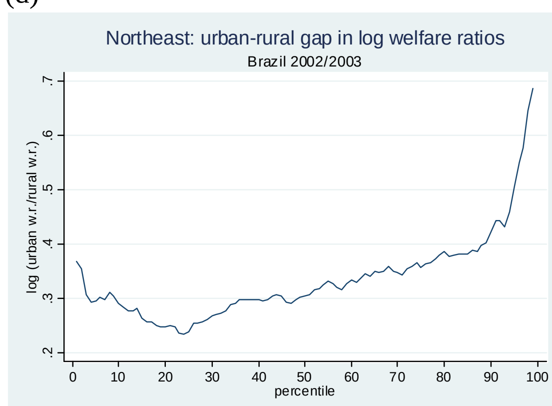
(b)



(c)



(d)



As it can be easily inferred, the difference in welfare ratios between metropolitan and urban areas within the SE region has a U shape, meaning that welfare differences are higher at the bottom and at the top of the distribution of welfare in the leading region. In contrast, the differences in welfare between urban and rural areas in the SE region increase almost monotonically with the level of welfare. Within the NE region, the lagging region of Brazil, the difference in welfare between metropolitan areas and urban areas is also higher at higher level of welfare suggesting that the wealthier households in the metropolitan areas of the NE are much better off than the wealthier households in the urban areas of the NE. Lastly, the differences in welfare between households in the urban and rural areas of the Northeast suggest that the differences are larger primarily among relatively wealthier households (above the 50^{th}

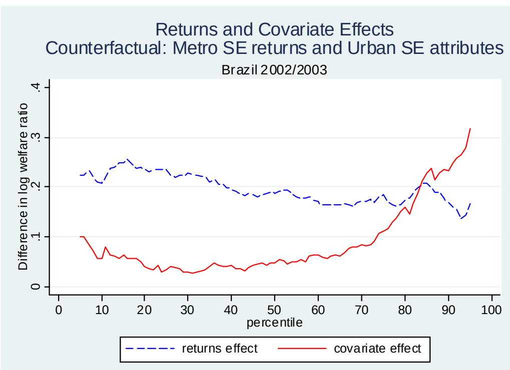
percentile of the distribution in per capita consumption). See Appendix D for welfare differences for other regions. The graphs in figure 3.3 suggest that decompositions of the welfare differences at the mean may yield a misleading picture about the relative role of characteristics and their returns in explaining these differences.

Quantile Decompositions Within Regions (Southeast and Northeast)

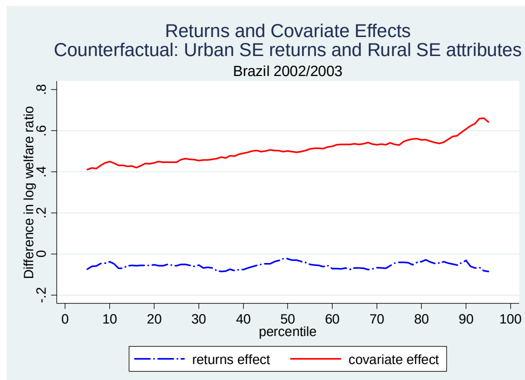
In figure 3.4, we present the decompositions of the metro-urban and urban-rural differences within the Southeast region, and within the Northeast region (see Appendix D for decompositions using alternative counterfactual distributions). As revealed by the simple Oaxaca decompositions, we find that both returns and covariate effects play a role, in explaining the large differential in living standards between metro and urban areas.

Figure 3.4: Quantile decomposition of welfare differences within Southeast and Northeast regions

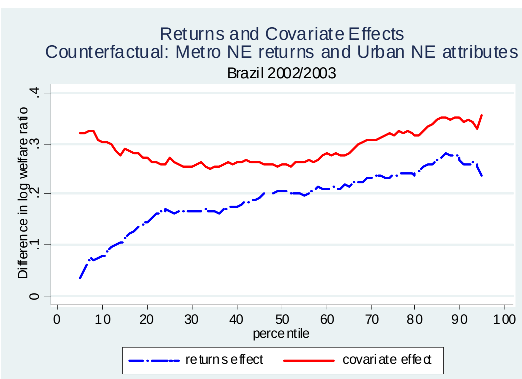
(a) Between metro and urban areas of SE



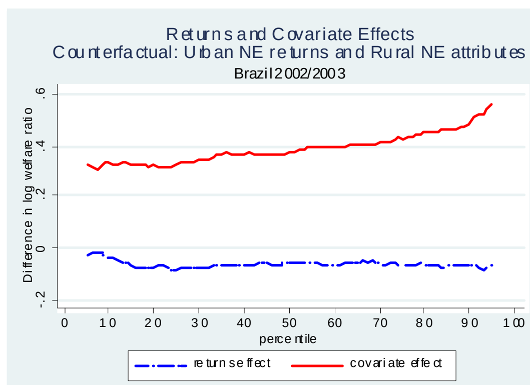
(b) Between urban and rural areas of SE



(c) Between metro and urban areas of NE



(d) Between urban and rural areas of NE



In the Southeast region (figure 3.4a) , the returns effect is dominant throughout most of the distribution, which is consistent with the result in the previous section, and further supports the idea of agglomeration effects in the leading metropolitan area.[9]

In the Northeast region, the metro-urban comparison (figure 3.4c) 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. The returns effect increases for households at the middle to the top of the welfare ratio distribution, indicating that those in the metro area are able to obtain higher returns, in particular for those better off. However, the returns effect is smaller for the poorer relative to those better off.

Decomposing welfare differences between urban and rural areas (figures 3.4b and 3.4d), 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.

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 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.

In conclusion, the results of within region decompositions indicate (i) a large returns effect in accounting for higher welfare in the metro Southeast relative to the 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 (iii) a dominant covariate effect and little to no returns effect in accounting for higher welfare in urban over rural areas. These results are consistent with the findings from the analysis of the means presented earlier in this section.

[9] It is important to keep in mind that we have also estimated bootstrapped standard errors around the estimated returns effect and covariate effects. The standard error bands do not overlap for the most part of the distribution of welfare, which implies that there are significant differences in the estimated returns and characteristics effects. More details are available upon request from the authors.

Quantile Decompositions Across Regions (Southeast and Northeast)

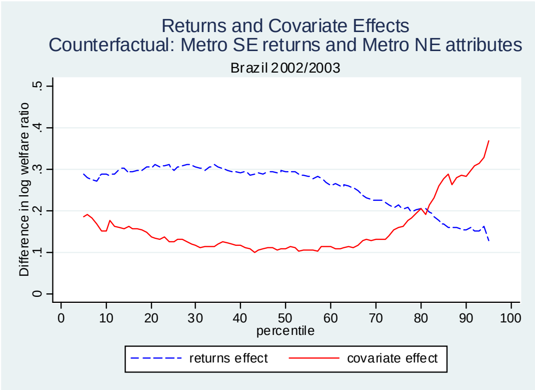
The next set of decompositions explores differences between similar areas of the leading Southeast and the lagging Northeast regions. The contribution of the returns and covariate effects in accounting for differences in the distributions of welfare are shown in Figure 3.5.

In comparing the metro areas of the Southeast and Northeast regions (Figure 3.5a), the returns effect is greater than the covariate effect in accounting for the differences between welfare distributions. Considering that metro Southeast (i.e. São 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.

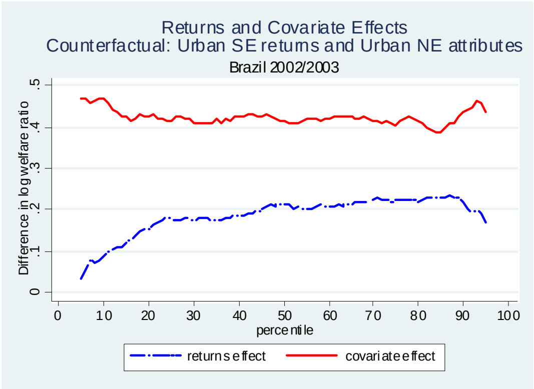
Between urban areas (Figure 3.5b) and between rural areas (Figure 3.5c) of these two regions, both the covariate and returns effects contribute to the difference. The upward slope of the returns effects indicates that the returns effect is smaller for the poor relative to those better off.

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

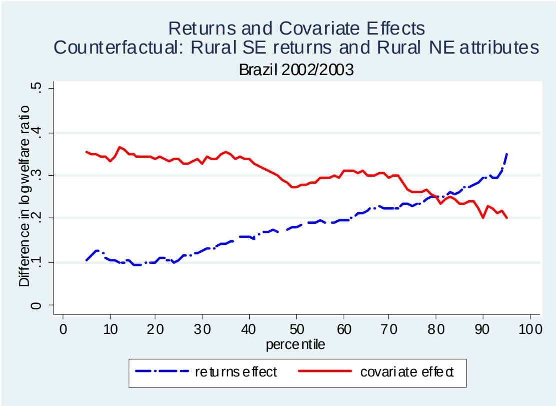
(a) Between SE and NE metro areas



(b) Between SE and NE urban areas



(c) Between SE and NE rural areas



In conclusion, the results indicate (i) a large returns effect in accounting for higher welfare in the metro Southeast relative to the metro Northeast, and (ii) a combination of covariate and returns effects in accounting for higher welfare in the urban and rural Southeast relative to urban and rural Northeast respectively. Again, these results are broadly consistent with the findings from the analysis of the means presented earlier.

4. CONCLUSIONS AND POLICY IMPLICATIONS

Welfare disparities between metropolitan, urban, and rural areas within any of the five regions of Brazil as well as disparities across regions continue to remain large, in spite of many federal and state government programs devoted to alleviating such disparities. In the development literature and in the policy arena, one encounters two opposing 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” to 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., São Paolo and Rio de Janeiro) where a high density of economic activity and better infrastructure exist. 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. On the other hand, 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 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 should concentrate on 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.

Cotton, J. (1988) "On the Decomposition of Wage Differentials." *The Review of Economics and Statistics*, 70: pp. 236-243.

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" Ensaios 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.

Jann, B. (2008) "A Stata implementation of the Blinder-Oaxaca decomposition" ETH Zurich Sociology Working Paper No. 5.

Kanbur, Ravi. 2006. "The policy significance of inequality decompositions" *Journal of Economic Inequality*, 4(3): pp 367-374.

Koenker and Bassett. 1978. "Regresion 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.

Neumark, D. (1988) "Employers' Discriminatory Behavior and the Estimation of Wage Discrimination" *The Journal of Human Resources* 23: pp. 279-295.

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.

Reimers, C. W. (1993) "Labor Market Discrimination Against Hispanic and Black Men." *The Review of Economics and Statistics*, 65(4), pp. 570-579.

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).

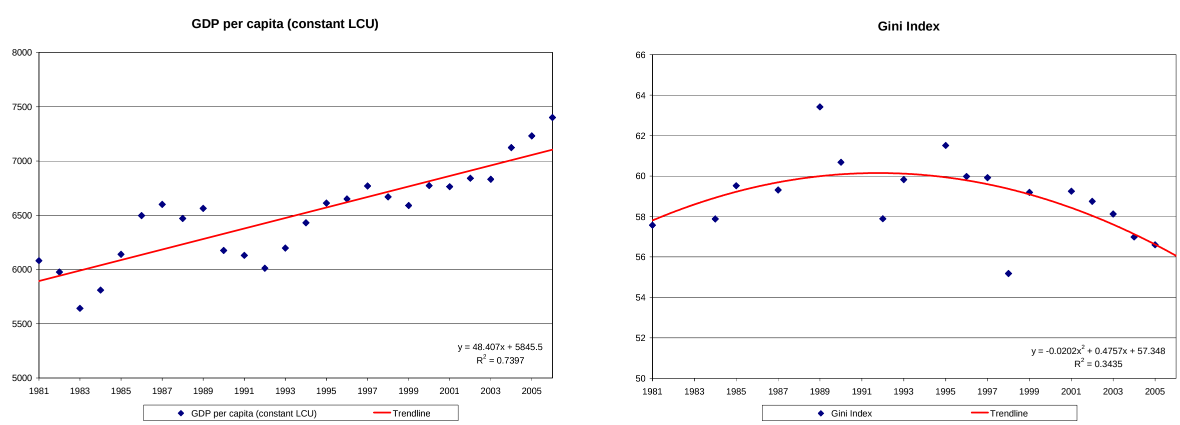
APPENDIX A

Appendix A contains: a map of Brazil's major regions, GDP per capita and inequality trends, various descriptive statistics on population and poverty distributions, and a brief discussion of selection bias.

Figure A1: Major Regions and States of Brazil

Source: http://gosouthamerica.about.com

Figure A2: GDP per capita and Gini Index: 1981-2005



Source: WDI 2008

Table A1: Descriptive Statistics of Regional Distribution of Population and Poverty

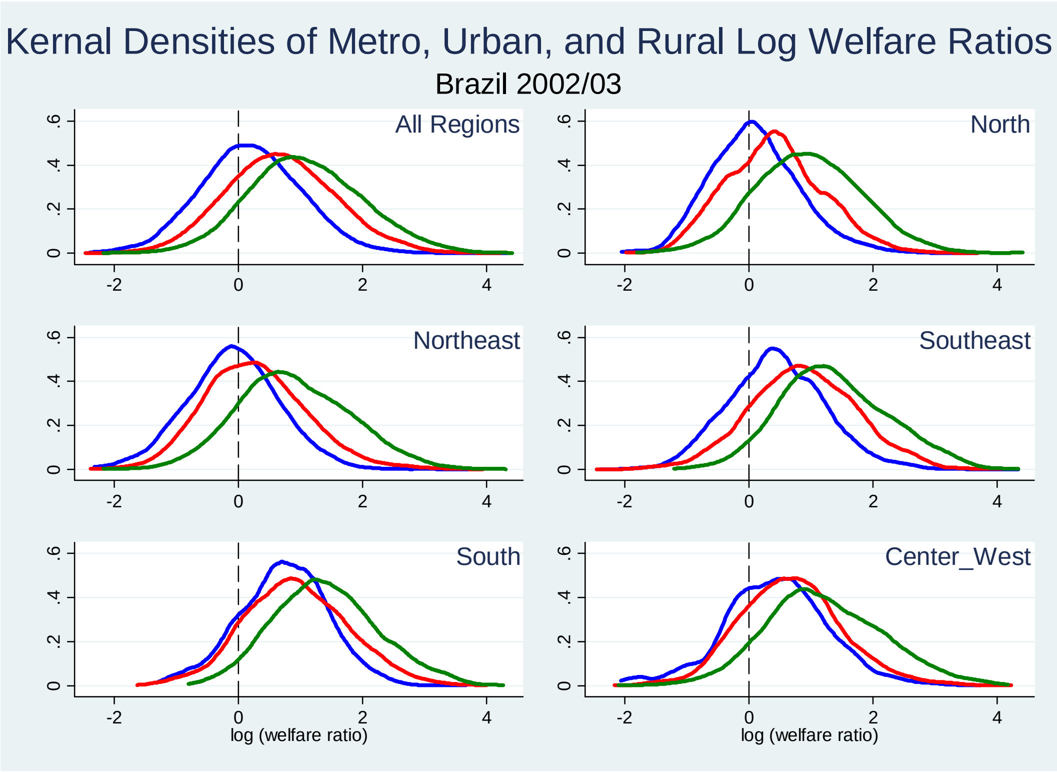
	Region		Total Population	Population Share	Number of Poor	Distribution of the Poor	Headcount Poverty	Poverty Gap	Poverty Severity
1	North	Metro Belem	1,845,708	1.05%	307,931	0.8%	16.7%	4.2%	1.7%
2		Urban	8,229,439	4.69%	2,252,145	6.0%	27.4%	8.9%	4.0%
3		Rural	3,533,713	2.02%	1,615,266	4.3%	45.7%	16.3%	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	**14.59%**	9,420,720	**25.0%**	36.8%	13.3%	6.4%
8		Rural	13,940,461	**7.95%**	7,664,897	**20.3%**	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 São Paulo	17,696,179	**10.09%**	1,278,930	**3.4%**	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	**19.97%**	5,045,726	**13.4%**	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 A2 : 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 A3: Distributions of welfare ratios by region



Frequency weighted Epanechnikov, halfwidth of kernel = 0.15

----- Rural ----- Urban (non-metro) ----- Metropolitan

NOTE: 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 A.3 represents the population living in poverty.

Table A3: Tests of Equality of Coefficients between Metro, Urban, and Rural Regressions

		Metro=Urban		**Urban=Rural**	
Year 2002-2003	**Restrictions**	**F-value**	**F-test (1 percent level)**	**F-value**	**F-test (1 percent level)**
Nongeographic Variables					
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
Geographic dummies	4	9.75	Rejected	3.91	Rejected

Table A4: Selection Bias Corrected and Uncorrected Coefficients

	Urban Northeast		Rural Northeast	
Dependent: log welfare ratio	**movestay**	**ols**	**movestay**	**ols**
# age 0-2	-0.322***	-0.313***	-0.285***	-0.296***
# age 0-2 squared	0.050**	0.046**	0.032	0.037**
# age 3-11	-0.292***	-0.289***	-0.203***	-0.204***
# age 3-11 squared	0.033***	0.033***	0.009*	0.009*
# age 12-17	-0.181***	-0.180***	-0.206***	-0.204***
# age 12-17 squared	0.018**	0.016*	0.023**	0.024**
# age 18-59	-0.168***	-0.169***	-0.206***	-0.205***
# age 18-59 squared	0.017***	0.017***	0.018***	0.018***
# age 60+	-0.122**	-0.097*	0.083	0.059
# age 60+ squared	0.023	0.019	-0.014	-0.010
Female head	-0.042	-0.028	-0.064	-0.074
Spouse	-0.146***	-0.158***	-0.169***	-0.159***
White or Asian	0.083***	0.075***	0.039	0.050
Black, Indigenous, or Missing	0.019	0.012	-0.076	-0.066
Age of head	0.023***	0.025***	0.023**	0.021***
Age of head squared / 100	-0.016***	-0.018***	-0.018*	-0.016***
Education of head				
1-3 years	0.133***	0.147***	0.112***	0.105***
4-7 years	0.224***	0.269***	0.260*	0.224***
8-10 years	0.373***	0.468***	0.497	0.380***
11+ years	0.655***	0.750***	0.725	0.588***
Education of spouse				
1-3 years	0.125***	0.126***	0.134***	0.134***
4-7 years	0.231***	0.254***	0.247***	0.228***
8-10 years	0.334***	0.392***	0.513**	0.450***
11+ years	0.592***	0.655***	0.715**	0.630***
Education differential				
1-3 years	0.036	0.032	0.046	0.048
4-6 years	0.095***	0.116***	0.138*	0.120***
7-9 years	0.191***	0.260***	0.327	0.261***
10+ years	0.351***	0.436***	0.383	0.298***
Occupation				
Professional	0.468***	0.461***	0.113	0.114
Technician	0.103*	0.107**	-0.036	-0.051
Administrative	0.115**	0.124**	0.027	0.010
Agriculture	0.046	-0.152***	-0.347	-0.188***
Manufacturing / Industry	-0.027	-0.023	0.029	0.027
Missing / not defined	-0.088***	-0.096***	-0.263***	-0.253***
_cons	0.092	-0.116	0.326	0.128
rho 1 (urban)	-0.506**			
rho 2 (rural)	0.382			

note: *** p<0.01, ** p<0.05, * p<0.1; Urban excludes metro areas

Since households may be making the decision to migrate or not based on expected welfare gains or losses given household characteristics, the location of residence should not be assumed to be exogenous. Using a full information maximum-likelihood estimation of an endogenous switching regression model, we explore the direction of the bias for urban and rural areas of the Northeast region. However, as in most cases, the correct specification of the switching equation always poses a challenge. Without good instrumental variables at hand, the model was simply identified by non-linearities. A comparison of the selection bias-corrected and uncorrected (OLS) coefficients are presented in Table A4 (below). In the urban areas, we find that the uncorrected education coefficients are higher than corresponding selection bias corrected coefficients. In rural areas, we find the converse. Since the urban coefficients are lower and rural coefficients higher, the difference in the coefficients, that is, urban minus rural, would be smaller, and since the returns effect is essentially the difference in coefficients, the returns effect would be smaller as well. Assuming the direction of the bias is correct, these results do not undermine the finding that the urban-rural differences in the Northeast are accounted for by the covariate effect and not the returns effect.

APPENDIX B

Simulated Poverty Profiles

We have also followed Ravallion and Wodon (1988) and simulated the poverty profiles that would prevail if the mean values for individual portable characteristics for different regions and areas were fixed at national means and thereby did not vary between and within regions (geographic or returns poverty profile) and poverty profiles that would prevail if the returns to characteristics did not differ between and within regions (concentration or endowment poverty profiles). In many cases this alternative approach allows one to determine visually whether it is differences in the returns to household characteristics or differences in the characteristics themselves that can better explain the differences in the standard of living between and within regions. First, in each region k a regression equation is estimated separately for metro, urban, and rural areas, as in

$$\begin{aligned} \ln C_i &= a_{Mk} + \beta_{Mk} X_i + \varepsilon_{Mki} . \\ \ln C_i &= a_{Uk} + \beta_{Uk} X_i + \varepsilon_{Uki} . \\ \ln C_i &= a_{Rk} + \beta_{Rk} X_i + \varepsilon_{Rki} . \end{aligned} \qquad (1)$$

The simulated geographic poverty profile is constructed using the constants and coefficients from equations (1) using the expressions

$$\begin{aligned} \Pr ob\left[\ln C_i < 0 \mid i \in M_k, X_i = X^N\right] &= \Phi\left[-\left(\alpha_{Mk} + \beta_{Mk} X^N\right)/\sigma_{Mk}\right] \\ \Pr ob\left[\ln C_i < 0 \mid i \in U_k, X_i = X^N\right] &= \Phi\left[-\left(\alpha_{Uk} + \beta_{Uk} X^N\right)/\sigma_{Uk}\right] \\ \Pr ob\left[\ln C_i < 0 \mid i \in R_k, X_i = X^N\right] &= \Phi\left[-\left(\alpha_{Rk} + \beta_{Rk} X^N\right)/\sigma_{Rk}\right] \end{aligned} \qquad (2)$$

where $\sigma_{mk}, \sigma_{uk}, \sigma_{rk}$ are the standard deviation of errors for each region/area and X^N is the national mean value of the individual endowments (urban and rural areas pooled).

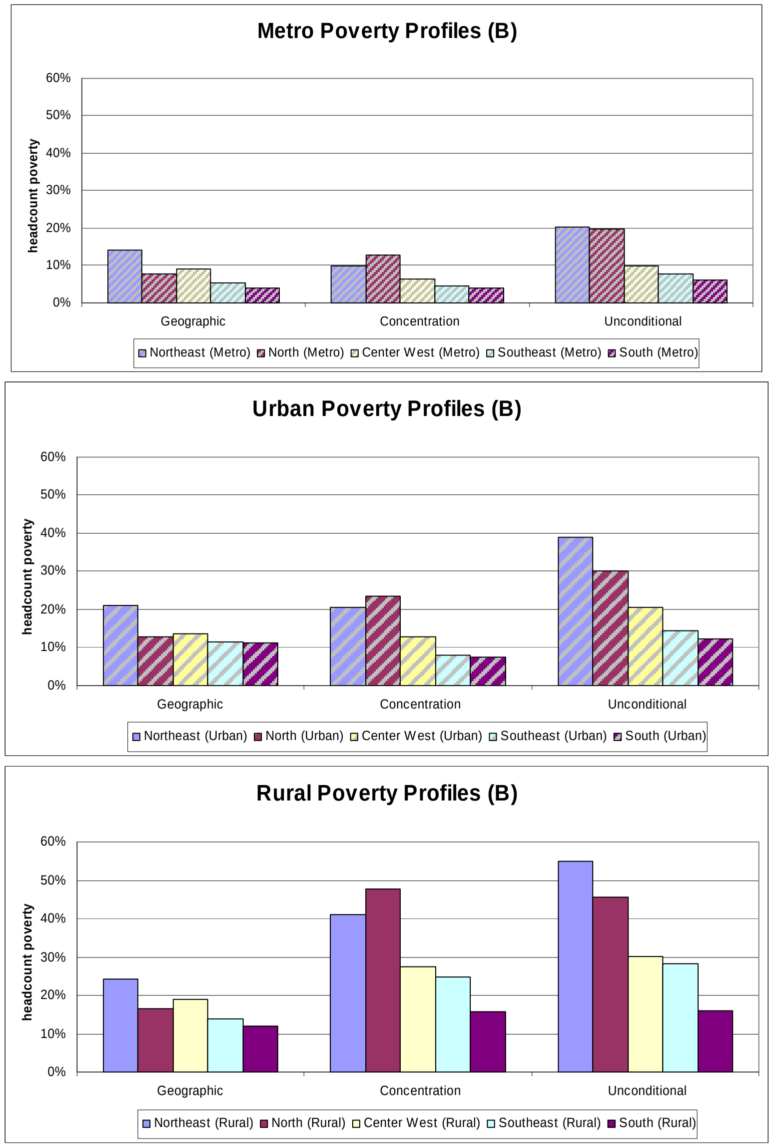
The simulated concentration poverty profiles are constructed as:

$$\begin{aligned} \Pr ob\left[\ln C_i < 0 \mid i \in M_k, X_i = X_M^k\right] &= \Phi\left[-\left(\alpha_N + \beta_N X_M^k\right)/\sigma_{Mk}\right] \\ \Pr ob\left[\ln C_i < 0 \mid i \in U_k, X_i = X_U^k\right] &= \Phi\left[-\left(\alpha_N + \beta_N X_U^k\right)/\sigma_{Uk}\right] \\ \Pr ob\left[\ln C_i < 0 \mid i \in R_k, X_i = X_R^k\right] &= \Phi\left[-\left(\alpha_N + \beta_N X_R^k\right)/\sigma_{Rk}\right] \end{aligned} \qquad (3)$$

where $\alpha_N = s_M (\sum_k s_{Mk} \alpha_{Mk}) + s_U (\sum_k s_{Uk} \alpha_{Uk}) + s_R (\sum_k s_{Rk} \alpha_{Rk})$
and $\beta_N = s_M (\sum_k s_{Mk} \beta_{Mk}) + s_U (\sum_k s_{Uk} \beta_{Uk}) + s_R (\sum_k s_{Rk} \beta_{Rk})$.

Figure B1 below presents the simulated geographic and concentration poverty profiles together with the actual poverty profile in the metropolitan, urban and rural areas of Brazil. In rural areas, the similarity of the concentration poverty profile across regions to the actual or unconditional poverty profile suggests that it is household endowments that are primarily responsible for the level and the variation of poverty across regions.

Figure B1: Simulated Poverty Profiles



APPENDIX C

Table C1: Oaxaca-Blinder Decompositions *Within* Regions:

(a) Difference in log welfare ratios explained by differences in endowments and coefficients

	Metro-Urban				Urban-Rural			
	threefold	**twofold**			**threefold**	**twofold**		
North		**D=0**	**D=0.5**	**D=1**		**D=0**	**D=0.5**	**D=1**
endowments	0.181	0.175	0.178	0.181	0.311	0.356	0.333	0.311
coefficients	0.121	0.121	0.118	0.115	-0.021	-0.021	0.002	0.024
interaction	-0.006	-	-	-	0.045	-	-	-
Northeast								
endowments	0.273	0.284	0.278	0.273	0.391	0.464	0.427	0.391
coefficients	0.191	0.191	0.197	0.203	-0.024	-0.024	0.012	0.049
interaction	0.011	-	-	-	0.073	-	-	-
Southeast								
endowments	0.105	0.125	0.115	0.105	0.185	0.248	0.216	0.185
coefficients	0.176	0.176	0.186	0.195	-0.028	-0.028	0.004	0.035
interaction	0.019	-	-	-	0.063	-	-	-
South								
endowments	0.098	0.093	0.096	0.098	0.227	0.273	0.250	0.227
coefficients	0.203	0.203	0.201	0.198	0.016	0.016	0.040	0.063
interaction	-0.005	-	-	-	0.046	-	-	-
Center West								
endowments	0.280	0.299	0.290	0.280	0.168	0.284	0.226	0.168
coefficients	0.170	0.170	0.179	0.189	-0.043	-0.043	0.015	0.073
interaction	0.019	-	-	-	0.116	-	-	-

(b) Percentage of difference in log welfare ratio explained by endowments and coefficients

	Metro-Urban				Urban-Rural			
	threefold	**twofold**			**threefold**	**twofold**		
North		**D=0**	**D=0.5**	**D=1**		**D=0**	**D=0.5**	**D=1**
endowments	61%	59%	60%	61%	93%	106%	99%	93%
coefficients	41%	41%	40%	39%	-6%	-6%	1%	7%
interaction	-2%	-	-	-	13%	-	-	-
Northeast								
endowments	57%	60%	59%	57%	89%	106%	97%	89%
coefficients	40%	40%	41%	43%	-6%	-6%	3%	11%
interaction	2%	-	-	-	17%	-	-	-
Southeast								
endowments	35%	41%	38%	35%	84%	113%	98%	84%
coefficients	59%	59%	62%	65%	-13%	-13%	2%	16%
interaction	6%	-	-	-	29%	-	-	-
South								
endowments	33%	31%	32%	33%	78%	94%	86%	78%
coefficients	69%	69%	68%	67%	6%	6%	14%	22%
interaction	-2%	-	-	-	16%	-	-	-
Center West								
endowments	60%	64%	62%	60%	70%	118%	94%	70%
coefficients	36%	36%	38%	40%	-18%	-18%	6%	30%
interaction	4%	-	-	-	48%	-	-	-

Table C2: Oaxaca-Blinder Decompositions Between Northeast and Southeast:

(a) Difference in log welfare ratios explained by differences in endowments and coefficients

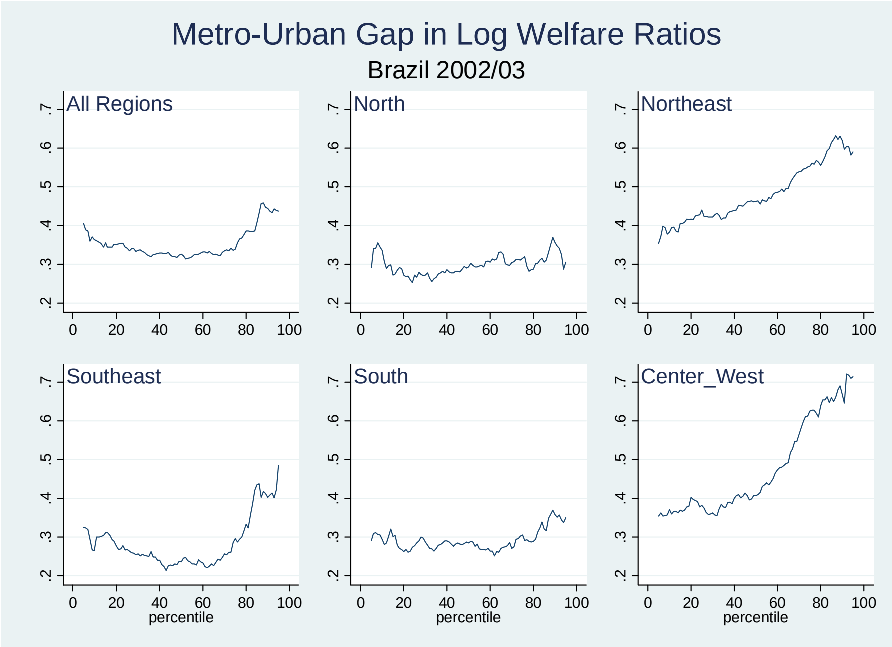
	Southeast - Northeast			
	threefold	twofold		
Metro		**D=0**	**D=0.5**	**D=1**
endowments	0.185	0.188	0.187	0.185
coefficients	0.233	0.233	0.234	0.235
interaction	0.003	-	-	-
Urban				
endowments	0.307	0.379	0.343	0.307
coefficients	0.216	0.216	0.252	0.288
interaction	0.072	-	-	-
Rural				
endowments	0.239	0.288	0.263	0.239
coefficients	0.202	0.202	0.227	0.251
interaction	0.049	-	-	-

(b) Percentage of difference in log welfare ratio explained by endowments and coefficients

	Southeast - Northeast			
	threefold	twofold		
Metro		**D=0**	**D=0.5**	**D=1**
endowments	44%	45%	44%	44%
coefficients	55%	55%	56%	56%
interaction	1%	-	-	-
Urban				
endowments	52%	64%	58%	52%
coefficients	36%	36%	42%	48%
interaction	12%	-	-	-
Rural				
endowments	49%	59%	54%	49%
coefficients	41%	41%	46%	51%
interaction	10%	-	-	-

APPENDIX D

Figure D1: Metro-Urban and Urban-Rural differences in log welfare ratios by region



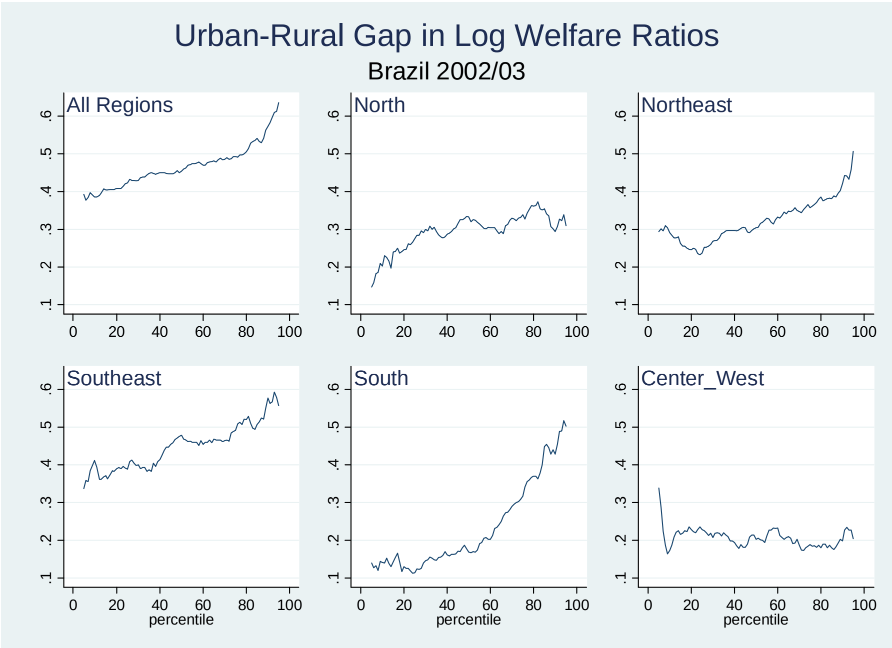
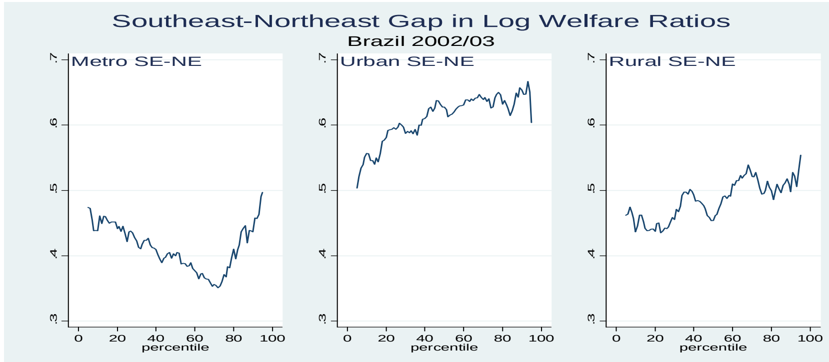


Figure D2: Differences in log welfare ratios between Southeast and Northeast regions

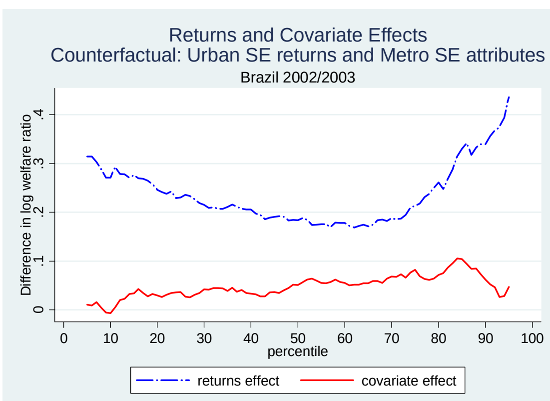


As mentioned in the methodology section, alternative counterfactual distributions can be constructed for the quantile decompositions. The results using these alternative counterfactual distributions are presented in Figure D3 below.

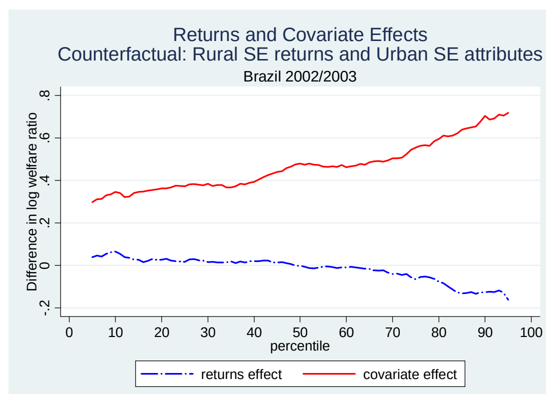
Figure D3: Quantile decompositions using *alternative* counterfactual distributions

Within Regions

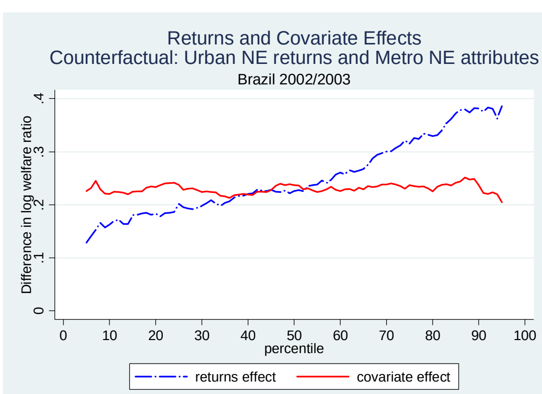
(a) Between metro and urban areas of SE



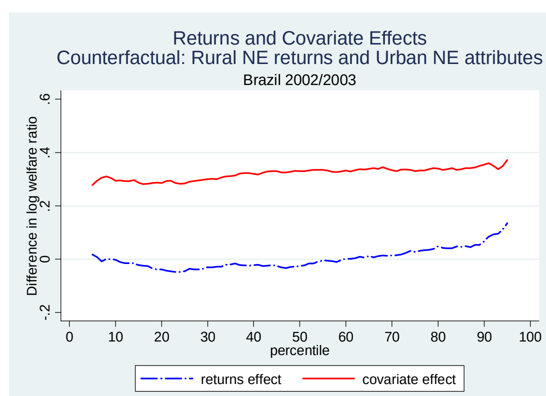
(b) Between urban and rural areas of SE



(c) Between metro and urban areas of NE

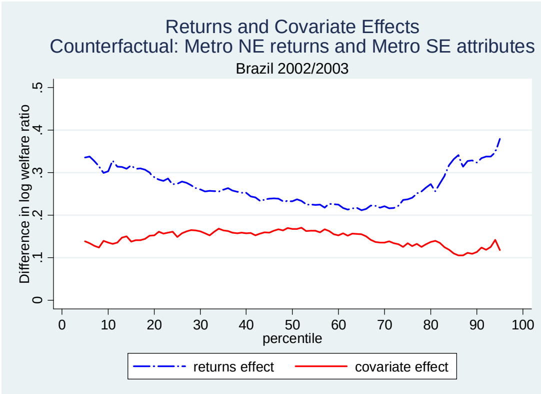


(d) Between urban and rural areas of NE

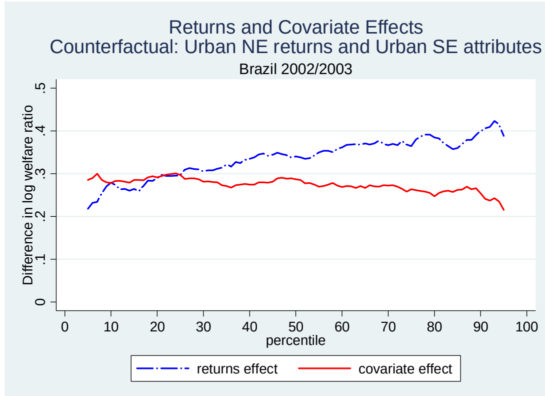


Between Regions

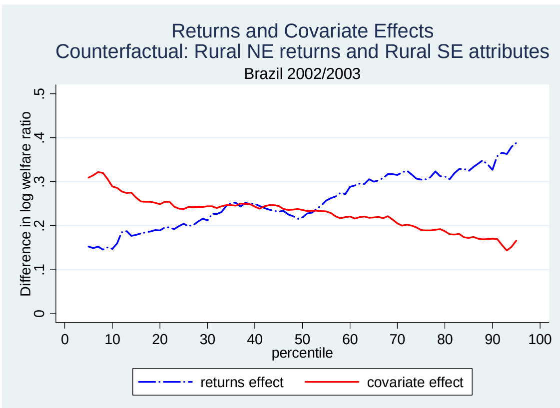
(e) Between SE and NE metro areas



(f) Between SE and NE urban areas



(g) Between SE and NE rural areas



Emmanuel Skoufias
C:\Documents and Settings\wb17224\My Documents\My Papers\LAC-RegionalStudies\LACstudy_Gladys\A_CountryCaseStudies\Brazil_LACEA_RK_ES.Nov16.doc
12/04/2008 9:27:00 AM