UNIVERSITA' CATTOLICA DEL SACRO CUORE

MILANO

Dottorato di Ricerca in

Modelli Quantitativi per la Politica Economica

Ciclo XIX

SECS – P/01 SECS – P/02

ESSAYS ON POVERTY, INEQUALITY AND WELL-BEING IN BRAZIL

Coordinatore: Ch.mo Prof. Maurizio BAUSSOLA

Tesi di Dottorato di Paola Salardi

Matricola: 3280100

Anno Accademico 2007-2008

Table of Contents

Introduction

This thesis is a collection of three essays on poverty, inequality and well-being for Brazil.

Brazil is a continent-sized nation with profound contrasts and remarkable diversity and is well-known for its very high level of inequality. In 2002 Brazil was the eighth most unequal country in the world, based on a Gini value equal to 59.1 (UNDP, 2002). Inequality in Brazil has been high and stable during a period that covers more than twenty years. On average one out of every three persons was considered poor according to the international US$1 a day poverty line by the end of the twentieth century (Wodon, 2000). It is for these reasons we argue that for the analysis and measurement of income distribution and poverty trends Brazil presents an interesting case, particularly in order to understand more deeply the determinants of its current situation by applying decomposition techniques.

The first chapter aims at understanding the key determinants of the Brazilian inequality. In order to reach this purpose, the chapter firstly sketches a poverty and inequality analysis for Brazil and then investigates the main determinants of inequality by applying several decomposition techniques. The study has been conducted by employing the annual Brazilian household survey for 2002.

The decomposition techniques applied in this study split into two approaches: inequality decomposition by indexes and regression-based inequality decomposition. Using the first methodology, a decomposable class of inequality measures is analysed by considering households characteristics such as geographic location, gender, age and ethnicity (Cowell and Jenkins, 1995).
For regression-based decomposition analysis, due to the large number of such methodologies, we limit the analysis to only a few. In particular, the first technique we apply was developed by Field and aims to capture the main determinants of income variability. This decomposition estimates the factor shares that mainly contribute to determine income inequality (Fields, 2002). After providing a general overview on the main causes of income inequality we

focus on racial heterogeneity and spatial differences by adopting the first and the second moment decomposition techniques. The Oaxaca decomposition method (also called the first moment decomposition) splits income differential between given groups into two effects: endowment effect, which accounts for differences in characteristics, and treatment effect, which computes differences in structure (Oaxaca, 1973). The second moment decomposition method enriches the study by analysing variance differentials, following the Dolton and Makepeace formula (Dolton and Makepeace, 1985; Callan and Reilly, 1993).

According to our results, in 2002 one third of the Brazilian population was considered poor and the Gini index was equal to 58.1 with an income distribution sharply skewed on the right. While Brazil is experiencing an improvement of its macroeconomic situation, the country is failing in the fight against inequality. We confirm the findings of several well-know studies developed by Ferreira and Paes de Barros (1999), Ferreira and Litchfield (2001), Bourguignon et al (2002), Elbers et al (2004), Rocha (2004). Brazilian inequality is primarily rooted in the differences across regions, educational levels and races.

Going deeper into inequality analysis by race and region, the application of regression-based decomposition techniques offers some clarity. Both first and second moment decompositions reveal that income inequality among race is mainly due to differences in characteristics: income discrimination between black and white people seems to be caused not by a "direct" discrimination against black. Indeed it is mainly caused by differences in assets, which might reflect more structural and hence long-standing features. The decomposition by region is conducted by comparing the poorest region, the North-East with each of the other regions. Comparing the North-East region to the North, the differential seems to be due to different and less favourable endowments for the North-East. By contrast, differentials between the North-East and the other three wealthier regions reveal that differences in structure, and not in the assets, are the key determinants. In other words, the North-East region has lower returns than the other three regions, holding characteristics constant. As understanding of the determinants of inequality deepens, it becomes a matter for the policy-makers to define possible interventions.

If by applying income differentials decomposition techniques we aim to deepen the understanding of the determinants of inequality, the employment of a methodology to decompose poverty measures can help in deducing what lie beyond poverty level differentials.

Indeed the second chapter investigates Brazilian poverty by exploiting geographical differences and questions whether the standard approach in measuring poverty is informative enough taking into consideration that the population is clearly heterogeneous. To do so, we apply the reformulation of the FGT class of poverty measures proposed by Chiappero and Civardi (2006). This poverty decomposition technique aims at computing poverty within groups, using group-specific poverty lines, and poverty between groups by adopting a community-wide poverty line.

This alternative conceptual and analytical approach to poverty measurement has potentially remarkable implications especially where the differentiation among poverty lines is very significant. Since geographical location is one of the most relevant determinants of Brazilian heterogeneity, the study exploits this criterion to establish geographically homogenous groups and assign to each of them their related poverty lines provided by Rocha (Rocha, 2003).

By employing the same data of the first chapter, we run two empirical exercises: for the entire country and for each Brazilian region. The North and the Central-West reveal a dominance of the within component. The North-East shows the highest level of poverty, even higher than the North and the Central-West, but the high within-group component is counterbalanced by a higher between-group component, attributable to the high level of inequality of the North-East. The South and the South-East have between-group components that dominate over within group ones. These empirical findings suggest that the analysis of poverty between- and within-groups is more exhaustive than the standard methodology when differentiated poverty lines are exploited.

This is particularly important with regard to policy implications. When a rise in inequality is detected, policy makers should be more focused on redistributive policies and particularly on policies related to social mobility that could improve income distribution in the long run. By contrast, increase in poverty may demand more immediate intervention to combat destitution and to increase access to basic needs and income. Behind our analysis of the dominance of the between- or the within-components of poverty lies a deep understanding of the complex relationship between poverty levels, income distribution and the robustness of poverty lines.

The first two chapters of this work apply techniques able to measure and decompose both poverty and inequality within the context of the standard monetary approach. In fact, well-being is conceptualized only in terms of income without taking into consideration possible other dimensions. The purpose of the last chapter is to enlarge the perspective of our analysis by adopting the capability approach developed by Sen (1985). The capability approach is an intrinsically complex framework, not only because it pays attention to a plurality of well-being dimensions in a similar fashion as other approaches, but also takes into account a multiplicity of personal, social and institutional contexts crucially important in the process of well-being. Individual well-being is not described as a static and materialistic condition defined by the possession of material resources but can be viewed as a process in which resources available are instruments to obtain well-being.

Under a capability perspective the well-being of a person can indeed be defined by a set of a person's functionings. From our point of view, the concept of functionings is a more comprehensive way of identifying personal well-being. Functionings is defined by what a person manages to do or to be with a given package of assets. It thus embodies the state of a person not as a mere possessor of goods or utility. Focusing on functionings allows us to observe what a person succeeds in doing or being with the resources that she or he is able to command.

The third chapter aims to model and estimate the health functioning production function as a relation that conveys to what extent people are able to convert private and public resources into the achievement of the specific functioning "being healthy". Hence, in our model the achievement of the health functioning is determined by private resources, given by an indicator of wealth, as well as public resources, identified by an index of public services, and controlled for a set of internal and external conversion factors.

The first conceptualization of the conversion process as a tool for assessing individual well-being is given by Sen (1985). This conversion process is affected by a set of internal and external conversion factors identified by given individual, social and environmental characteristics. The construction of the model is based on the conceptual analysis for modelling individual well-being provided by Chiappero-Martinetti et al (2007).

The estimation of the health functioning production function has been made by employing Brazilian data, in particular the households survey for 2003. The choice of this year is due to the fact that the 2003 version of the same dataset exploited in previous chapters contains a special section on health that is functional for our investigation. The econometric methodologies applied depend on the nature of the variables that identifies the health functioning. We estimate the health functioning production function by applying both probit and ordered probit regression models due to the categorical nature of the dependent variables that identify functionings achievement. The computations have been made for the entire Brazilian sample and by gender and race, recognizing the relevance of our empirical findings in terms of policy implications.

According to our findings, when the health functioning is identified by the self-reported morbidity index, public resources are more relevant in the health functioning achievement process. On the other hand, when a health status indicator identifies the health functioning, private resources become predominant.

Looking at our empirical results disaggregating by gender and race, Brazilian black people might be considered one of the most vulnerable groups. The

Brazilian policy maker should protect this part of the population that records the lower ability to convert their private resources and a good efficiency in using public resources. Another interesting result is the fact that women record a greater impact of public resources while for men private resources are more relevant. The Brazilian policy maker should protect these weaker sub-groups of the population. Possible directions of policy intervention might be to promote black-targeted public provision of medical assistance and prevention. Moreover, the public health services should be aware of the fact that the highest portion of its policyholders is female. We conclude that our empirical findings might be relevant for policy making, for example in the health public sector, once a more comprehensive approach of assessing individual well-being is accepted.

We conclude by listing some fundamental remarks that need to be solved in order to further the applied methodologies.

The analysis of inequality decomposition needs to employ more refined econometric techniques that are able to deal with some of the limits of the first and second moment decomposition techniques, such as selection bias and error measurement among others. Moreover, we would like to extend the income differential analysis by decomposing it into its sources and then applying decomposition techniques to each of the income sources in order to understand in which income source creates the greatest "discrimination effect" and, hence, ultimately causes most income inequality.

In the context of the operationalization of the capability approach, we would like to estimate conversion rates for more than one functioning as well as employ more data and more appropriate econometric techniques to deal with problems such as endogeneity and omitted variables. Finally, we believe that not only the assessment of more than one functioning is necessary, but also the investigation of the possible interrelations existing among functionings is a key priority for a more comprehensive view of individual well-being.

Chapter 1

How much of Brazilian Inequality can be explained?

Abstract: Brazil is well-known for its very high level of inequality. Understanding the key determinants of this inequality is the principal aim of this study. In order to reach this purpose, the present work firstly sketches a poverty and inequality analysis for Brazil and then investigates the main determinants of inequality by applying several decomposition techniques by using the annual Brazilian household survey for 2002. Numerous techniques are developed, split into two approaches: inequality decomposition by indexes and regression-based inequality decomposition. Using the first methodology, a decomposable class of inequality measures is analysed by considering households characteristics such as geographic location, gender, age and ethnicity. For regression-based decomposition analysis, the present work employs the Field decomposition and the Oaxaca decomposition. We confirm the findings of other studies by verifying that Brazilian inequality is primarily rooted in the differences across regions, education levels and races. After investigating more deeply the differentials by race and region, inequality seems not to be caused by a "direct" discrimination against most marginalized groups, but spring from a group of structural problems stemming from both Brazilian culture and habits and also related to the structure of the Brazilian economy and society.

1.1 Introduction

Brazil is a continent-sized country and it occupies half of the entire area of South America. According to the UNDP report (2002), Brazil's population of 176.3 million makes it the sixth most populous country in the world.
Brazil is not only a giant but also a country of striking diversities: probably Brazil is home to remarkable geographical and climatic variety, to a hugely diverse population of indigenous tribes, white people of European descents, black people who arrive during the era of slavery, and Asians and Europeans, who arrived in successive waves of immigration.
All this diversity has the potential to form the basis of a great and powerful nation. However, sharp diversities are also a fertile soil for social and economic inequalities. Indeed, Brazil is well-know for its very high levels of inequality.

Using 2002 as reference year, Brazil was the eighth most unequal country in the world, based on UNDP-Gini index calculations which found a Brazilian Gini value of 59.1 (UNDP, 2002). The six most unequal countries are all very small African countries with US$ GDPs less than a thousandth of Brazil's GDP.[1] The only large country more unequal than Brazil is South Africa, where inequality is also the product of the apartheid era which only came to an end a decade ago.

Exacerbating the situation is the fact that Brazil records the smallest share of income owned by the 10% poorest population. Together with Lesotho, Sierra Leone and Namibia, the poorest decile of the population distribution owns only 0.5% of the GDP. While this population group might be considered negligible for very small countries, for Brazil the poorest decile accounts for a consistent, and large, part of the population that is totally interdicted from Brazilian wealth. More broadly, in 2002 the poorest half of the Brazilian

[1] These six countries are Namibia (12.3), Lesotho (4.3), Botswana (14), Sierra Leone (2.7), Central African Republic (4.5) and finally Swaziland (4.5). In brackets, GDP of each country is reported in billions of US$. South Africa is the seventh most unequal country with a US$ GDP equal to 456.8 billions. Brazil's GDP is 1,355 US$ billions. All these values come from the UNDP report for 2002.

population owned only 13.42% of the total GDP, while the richest 10% held half of the Brazilian GDP.

Brazilian inequality is thus something that cannot be ignored. The main aim of this work is to investigate inequality and poverty of this country and to determine the possible causes of its considerable inequality.

The fundamental steps of any analysis and study of inequality are, first, the definition of concepts of inequality and wealth, and, second, the choice of methods to implement those concepts. In this sense the study of inequality embraces different aspects that are worth highlighting in this introduction.

First, inequality is generally used to refer to income. However, income inequality is not the only and more comprehensive way to look at inequality. In fact, there are other aspects such as financial and land assets, or health and education, which should be taken into account. It may be argued that investigating income inequality is nonetheless quite effective because it is strictly correlated with other inequalities in areas such as land and education (World Bank, 2003). This may not always hold true and an independent investigation might help to better detect the cause-effect relationship that leads these variables. In particular, several studies have outlined a significant connection between income inequality and inequality in land assets, as well as in educational attainment, for Brazil (Ferreira and Paes de Barros, 1999; Ferreira and Litchfield, 2000).

Second, the concept of welfare is frequently associated with economic growth, but this might be too shallow of an approach. An inclusive concept of welfare should consider not only income growth, but also the issue of income distribution.[2] Looking at the GDP growth of a country is fundamental to better understanding its development process, but it is never sufficient to sketch a reliable picture of the welfare situation in that country. As already pointed out, Brazil is a middle-income country, but under other aspects considered essential for a complete concept of welfare, such as educational attainment, it falls behind this standard (UNDP, 2002).

[2] There is a large body of economic literature that refers to growth with redistribution issues. Related to Brazil, one of the most important studies is Datt and Ravallion (1992).

Third, the complex linkages among inequality, poverty and growth can help us to deeply understand the composite and multidimensional Brazilian reality. According to significant economics literature, defining and conceptualizing all the linkages in the well-know inequality-poverty-growth triangle (Bourguignon 2004; Lopez, 2004) is doubly important. It is not only valuable by itself in term of ethics, but also because poverty and inequality affect economic performances just as economic performance might worsen poverty and inequality. While the complex cause-effect connections among these variables are difficult to detect, the general wisdom agrees that a high level of structural and persistent inequality jeopardizes potential economic growth (Deininger and Squire, 1998).

For this reason, Brazil is often called the "sleeping giant". The country has all of the characteristics needed to become a powerful country in the international panorama, with large potential in the industrial and manufacturing sectors and a wide range of disposable natural resources (Graham, 2004). As such, Brazilian struggles to achieve consistent economic development cannot be totally explained without taking into consideration the issue of inequality.

According to Litchfield's studies (2001), while macroeconomic instability that has characterized Brazil in the last thirty years has certainly undermined economic growth, Brazil has also suffered from the economic and social illness called inequality. This inequality grew during the decades of economic stagnation and contributed to a vicious loop of economic collapses and social deterioration.

As such, studying the main determinants of inequality should contribute to better understanding the economic and social situation in Brazil and ultimately might provide useful insights for further policy making. This is the principal purpose of this study.

The data come from the annual Brazilian household survey, called the *Pesquisa Nacional por Amostra do Domicilios* (PNAD). The Author's elaborations are only based on the survey for 2002, while comparisons with previous years are possible by using Litchfield's earlier computations (2001).

In order to facilitate comparison, this study tries to apply the same methodological choices for constructing variables as Litchfield's work.[3]

Section 1.2 presents poverty and inequality analysis of Brazil for 2002. These results are then compared with Litchfield's findings for previous years to sketch possible evolutions. Section 1.3 employs inequality decomposition techniques to identify the potential determinants of inequality. Numerous techniques are developed, split into two approaches: inequality decomposition by indexes and regression-based inequality decomposition.

Due to the large number of such methodologies, we limit the analysis to only a few of them. In the first part of this section, inequality decomposition by population sub-groups is conducted. Using this methodology, a decomposable class of inequality measures is analysed by considering households characteristics such as geographic location, gender, age and ethnicity.

The second part of this section presents three regression-based decomposition techniques. First, Field's decomposition which identifies key determinants of Brazilian income inequality for 2002 by regressing an income generating function (Field, 2002). Then, by applying Shorrocks' formula, it is possible to compute inequality shares.

Second, the Oaxaca decomposition technique divides the estimated income differential into two different effects: the effect of differences in characteristics and the effect of differences in structure (Oaxaca, 1973). This methodology is useful for understanding the potential role of discrimination behind any income differentials between races, genders or regions. Moreover, this technique is deepened at the end of the section by considering not only the mean income differentials, the so-called first moment decomposition, but also the variance differentials, the so-called second moment decomposition, following the Dolton and Makepeace formula (Dolton and Makepeace, 1985; Callan and Reilly, 1993).

Conclusions focusing on the policy implications are provided in section 1.4.

[3] A detailed description of the dataset that has been used for all of the empirical exercises reported in this study is provided in the appendix 1.B.

1.2 Poverty, inequality and wealth across 1981, 1990 and 2002

This section presents a comprehensive analysis of the level and composition of Brazilian poverty and inequality over the period 1981-2002. The study uses the 2002 PNAD data to compute a wide battery of poverty and inequality indexes in order to sketch a complete poverty and inequality profile for 2002. The empirical results are subsequently compared with the Litchfield's calculations for 1981 and 1990 to allow a more detailed and reliable analysis of Brazilian welfare conditions during the last two decades.

1.2.1 Poverty analysis

The poverty analysis is performed by applying the FGT class of measures (Foster, Greer and Thorbecke, 1984). As the basis for the computation of the summary statistics shown below, several methodological assumptions have been made. These assumptions play a crucial role in the outcome of this study. Hence, it is useful to highlight the most important of them.

First, real per capita income is adopted as welfare measure. The choice of income instead of consumption is largely pragmatic.[4] Moreover, since the variable comes from a survey and not from national accounts, poverty might be overestimated.[5] Similarly, the per capita adjustment might cause an

---

[4] The majority of studies that refers to Brazil adopt income instead of consumption: for example the analysis on Latin American countries developed by Wodon (2000) and specifically for Brazil the last study of Rocha (2004). To the best of our knowledge, the only study that employs a consumption variable is the analysis provided by Elbers et al (2004) where data from the PNAD are compared and then merged with data from the PPV (the *Pesquisa sobre Padrões de Vida*). This survey is similar to the LSMS and collects data on consumption in addition to information on incomes.

[5] In his work, Lluch (1982) highlights how the under-reporting of capital incomes in Brazil is likely to lead to underestimates of both the mean and the dispersion of the income distribution. Altimir (1977) provides a complete review on the household survey for LAC and proposes a methodology to overcome the problem of underreporting for surveys versus national accounts data. There is plenty of works that highlights the problem of equating between surveys and national accounts such as Meja and Vos (1997), Szekely et al (2000), Wodon et al (2000).

upward bias in the estimations.[6] As such, the interpretation of the empirical results should be conscious of these shortcomings.

A second notable choice is that the real per capita income is weighted by a deflator with 1995 as base year. The choice of the 1995 as base year has been done because the real value of income should be harmonized with the real values for the poverty lines in order to be comparable.[7]

Table 1.1: Brazilian per capita poverty lines, in 1995 prices

PNAD Regions		Value
Region I	Metropolis of Rio de Janeiro	100.73
	Urban	62.45
	Rural	45.33
Region II	Metropolis of São Paulo	107.33
	Urban	67.62
	Rural	42.93
Region III	Metropolis of Curitiba	86.27
	Metropolis of Porto Alegre	59.89
	Urban	54.81
	Rural	36.54
Region IV	Metropolis of Belo Horizonte	82.78
	Urban	55.46
	Rural	32.28
Region V	Metropolis of Fortaleza	62.94
	Metropolis of Recife	83.79
	Metropolis of Salvador	96.19
	Urban	56.68
	Rural	34.01
Region VI	Brasilia	102.98
Region VII	Metropolis of Belem	58.36
	Urban	51.94
	Rural[1]	38.22
Region VIII	Goinia	97.86
	Urban	74.37
	Rural[1]	38.22

Source: Rocha, 1993, re-adapted by Litchfield, 2001.

[6] Per-capita adjustment is generally adopted in the literature on poverty measurement for Brazil (Rocha, 2004). Ferreira and Paes de Barros is one of the few studies that employs two different adjustments in order to take into account economies of scale and heterogeneity of needs within households.

[7] Indeed, in this study we have adopted poverty lines at 1995 prices and incomes from different years have been coherently adjusted to their real values for 1995.

Finally, the absolute poverty lines adopted by this study follow Rocha (1993) as shown in table 1.1. Rocha constructed a range of region specific absolute poverty lines by using a variant of the cost of basic needs approach, recognizing that the cost of the required basket of food varies by region and between urban and rural areas.

Table 1.2 shows poverty estimates for 1981, 1990 and 2002 using Rocha's set of poverty lines. Looking to the results, poverty seems to have decreased during the last twenty years. This table shows that the entire FGT group of indicators generally displays downward trends that become even sharper as sensitivity to the bottom of the income distribution increases.

Table 1.2: Summary statistics of FGT(α) class of measures across 1981, 1990 and 2002

	1981[(a)]	1990[(a)]	2002[(b)]
Headcount ratio	0.445	0.450	0.336
s.e.	0.002	0.0024	0.0019
C.I.	(0.441,0.449)	(0.445,0.455)	(0.334, 0.338)
Poverty Gap	0.187	0.199	0.136
s.e.	0.001	0.0012	0.001
C.I.	(0.185,0.189)	(0.196,0.202)	(0.135, 0.137)
Squared Poverty Gap	0.104	0.114	0.074
s.e.	0.0007	0.0009	0.0007
C.I.	(0.103,0.105)	(0.112,0.116)	(0.073, 0.075)

(a) Source: Litchfield's calculations from PNAD 1981-1995;
(b) Source: Author's calculations from PNAD 2002;

The Headcount Ratio (HC) decreased by 24.5% between 1981 and 2002, while the Poverty Gap (PG) and the Squared Poverty Gap (SPG) shrank by 27.3% and 28.6% respectively.[8] These figures confirm the downward trend identified by Litchfield for the period 1981 to 1995. Referring to her work (Litchfield, 2001), during the period 1981-1995 the HC index decreased by 15.3%, whereas the PG and the SPG diminished by 16.6% and 17.3%.

[8] All the estimated changes in the poverty indicators are statistically significant at 95%. The exception to this is the change between 1981 and 1990 in the HC ratio whose increase is not statistically significant at 95% confidence. However this does not affect the results concerning the trend between 1981 and 2002.

However the main cause of this decrease is the reduction in poverty found by Litchfield (2001) that occurred between 1993 and 1995 mainly due to the consequences of the Plano Real in 1994[9].

This decrease in poverty between 1993 and 1995 coupled with the estimates shown in table 1.2 indicate a clear downward trend from the mid '90s to 2002. The massive increase in poverty recorded along the lost decade of the '80s has been offset by the improvement of the last decade, also called "the decade of the reforms", shown by this work which updates Litchfield's study (2001). Furthermore, the figures in these poverty indexes are evidence of a strong link between poverty and macroeconomic performances. The fact that poverty increased with recession and shrank when the market witnessed an economic boom supports the view of the anti-cyclical behaviour of this phenomenon.

Although the decrease in poverty during the last decade might be imputed to an effective economic improvement, the analysis of these summary statistics should be conducted while keeping in mind the controversial effects of the macroeconomic adjustment, and in particular of devaluation.

The above analysis made by summary statistics is further confirmed by the stochastic dominance analysis. By plotting the Poverty Incidence Curves, it is possible to check graphically which year shows a higher level of poverty: each point of these poverty incidence curves gives the proportion of the population consuming less than the amount given as the horizontal axis of the graph.

In the appendix 1.A, the figures A1.1, A1.2 and A1.3 confirm the previous results obtained by computing poverty indexes: the level of poverty in the 2002 is lower than in either 1981 or 1990 while the comparison between 1990 and 1981 is ambiguous; in fact these two poverty incidence curves are almost

[9] The Plano Real was a new stabilization programme that was supposed to overcome some weaknesses of previous plans. As pointed out by Baer (2001), one of the major problems of previous stabilization programme was to stop inflation only temporarily. The new plan meant to work on fiscal stabilization as well as to lead to a new currency only gradually through a new indexing system. The results were initially positive. By the end of the 1980s, the mean income of the poorest 40% had fallen to below 1981 levels. Only when inflation began to fall again after the 1994 *Plano Real* did real incomes recover to levels similar to the beginning of the 1980s. Rocha (1996, 2000) provides a detailed analysis of the impact of *Plano Real* on the poor populations explaining the changes in income distribution and labour market.

coinciding. Brazilian poverty gradually increased during ‘80s and then, during ‘90s, it decreased noticeably until 2002, leaving a final level, slightly lower than twenty years earlier.

1.2.2 Inequality analysis

The analysis of inequality involves the study of the levels and the shares of income for different economic groups across years. In order to drawn a comprehensive inequality analysis, summary statistics for the most important inequality indicators are presented along with the stochastic dominance analysis. Table 1.3 shows the inequality indicators adopted.

Table 1.3: Summary Statistic of the inequality indexes across 1981, 1990 and 2002

	1981[(a)]	1990[(a)]	2002[(b)]
Mean income[(c)]	136.2	149.8	198.70
Median income	71.4	72.2	103.27
Inequality			
Gini	0.574	0.606	0.581
s.e.	0.0014	0.0022	0.0019
C.I.	(0.571, 0.577)	(0.601, 0.610)	(0.5791, 0.5829)
GE(0)	0.613	0.705	0.631
s.e.	0.0034	0.0058	0.0046
C.I.	(0.605, 0.619)	(0.691, 0.717)	(0.6264, 0.6356)
GE(1)	0.647	0.745	0.688
s.e.	0.0048	0.0119	0.0117
C.I.	(0.637, 0.655)	(0.722, 0.771)	(0.6763,0.6997)
GE(2)	1.336	2.019	2.058
s.e.	0.0287	0.2523	0.5353
C.I.	(1.282, 1.390)	(1.591, 2.618)	(1.5227, 2.593)

(a) Source: Litchfield’s calculations from PNAD 1981-1995;
(b) Source: Author’s calculations from PNAD 2002.
(c) Mean and median income values are shown in Brazilian Reais at 1995 real values.

The first notable trend is the massive increase in the mean income over the last twenty years. This trend should be read with caution. Although it implies sharp Brazilian economic growth, it might lead to wrong and distorted conclusion on the wealth situation in Brazil. Demographic trends, as well as changes in income shares, should be investigated, as analysis of mean income is not sufficiently reliable.

The data show that the mean income increased by 45.88% during the last twenty years, primarily the last decade, as only 9.9% of the increase occurred between 1981 and 1990. The median value of income dramatically rose as well. This tells about the degree to which the income distribution is skewed. By comparing the median to the mean, the distribution of income appears to be skewed to the right across all of the years considered. However the ratio of mean to median tells us that it is becoming less skewed.

The inequality indicators computed are the Gini index as well as the three most well-know indexes from the General Entropy class of measures, the Mean Log Deviation, the Theil index and the Coefficient of Variation, respectively GE(0), GE(1) and GE(2).[10] The overall trend shows that inequality has increased from 1981 until 2002. However, a more detailed observation of the data reveals that after a constant and striking increase in inequality during the '80s, the last decade has traced a regular decrease, although it was not enough to return inequality to the level in 1981. During the period 1981-2002, the Gini index shows an overall increase of 1.22%, while the GE class of measures shows respective increases of 2.9%, 6.3% and 54%.[11]

The comparison with Litchfield's calculation over the period 1981-1995 confirms the previous results: inequality diminished during the last decade, but still not enough to offset damage done in the '80s. Particularly, all of these inequality measures show a slight, but statistically significant decrease between 1990 and 2002, with exception for the GE(2) measure that keeps on increasing but this increase is not statistically significant at 95% confidence giving confidence to the conclusion of a downward trend in the '90s.

Given the weak decrease in inequality during the '90s, the calculations presented in the next two tables allow us to better understand how the increase in overall welfare has been shared among the different decile groups.

[10] To test for statistical significance of the estimated changes in the inequality indicators, the standard errors for each indicator have been computed by using the bootstrapping procedure with replacement over 100 replications.

[11] All of these estimated inequality increases are statistically significant at 95% confidence.

Table 1.4 reports mean incomes per decile groups, i.e. the absolute variation in income for each decile group, while Table 1.5 displays income shares by decile groups to show the relative variation.

Table 1.4: Mean Incomes per decile groups across 1981, 1990 and 2002

Decile	1981[(a)]	1990[(a)]	2002[(b)]
1	13.3	11.6	18.58
2	25.1	22.8	36.03
3	35.7	33.9	53.26
4	47.9	46.5	71.81
5	62.2	62.9	91.92
6	80.6	83.0	117.22
7	106.4	111.9	151.72
8	146.4	158.1	208.55
9	225.8	250.6	321.78
10	613.9	719.1	923.72
Overall	136.2	149.8	198.70

(a) Source: Litchfield's calculations from PNAD 1981-1995;
(b) Source: Author's calculations from PNAD 2002.

The interpretation of these two tables is straightforward. During the period 1981-2002, the bottom of the Brazilian income distribution gained in term of absolute terms, but lost in relative terms.

Table 1.5: Income shares by decile groups across 1981, 1990 and 2002

Decile	1981[(a)]	1990[(a)]	2002[(b)]
1	0.97	0.77	0.93
2	1.85	1.52	1.80
3	2.63	2.26	2.68
4	3.53	3.10	3.61
5	4.59	4.19	4.62
6	5.94	5.53	5.88
7	7.84	7.46	7.63
8	10.78	10.54	10.49
9	16.64	16.70	16.19
10	45.23	47.93	46.48

(a) Source: Litchfield's calculations from PNAD 1981-1995;
(b) Source: Author's calculations from PNAD 2002.

Specifically, the first decile of the distribution experienced a 39.7% increase in mean income, but lost 4.1% of the income share. The changes for the top of the distribution are more unambiguous. Their mean income increased by 50.5% while their income share increased 2.7%.

To summarize the main conclusions form these two tables, the absolute variation of mean income for each decile groups is unequivocally positive, while the relative variation of income computed by income shares gives evidence that the Brazilian population did not benefit equally from economic growth.

In her study covering the years 1981-1995 (Litchfield, 2001), Litchfield referred to Datt and Ravallion's analysis of the Brazilian growth and redistribution (Datt and Ravallion 1992, quoted in Litchfield 2001) and highlighted the ongoing debate about the effectiveness of economic growth for fighting poverty when it is not followed by income redistribution. As long as Brazilian economic growth excludes the poorest part of the population, the overall level of poverty and inequality may not improve. A more equal distribution of the benefits coming from economic growth is needed.

Litchfield (Litchfield, 2001) drew an insightful table to illustrate the winners and losers through this period by classifying the winning and losing deciles during each period. Table 1.6 replicates the same idea by adding the information about 2002.

Table 1.6: Brazilian economic performances: winners and losers

	1981-1990		1990-2002		1981-2002	
	Winners	Losers	Winners	Losers	Winners	Losers
Absolute terms	5-10	1-4	1-10	None	1-10	None
Relative terms	9-10	1-8	1-7	8-10	3-5 and 10	1,2 and 6-9
Both	9-10	1-4	None	None	1 and 10	None

Source: Author's calculation from PNAD 2002.

To complete this inequality profile, stochastic dominance analysis is useful. This provides added clarity when the indicators provide contradictory results due to differing sensitivity to different parts of the income distribution. The stochastic dominance analysis can be carried out using the Lorenz Curve

and the Generalized Lorenz Curve. Deaton (1997) has argued it is essential to investigate them all in order to obtain a clear picture not only of inequality, but also of social welfare. Intuitively, welfare is considering a broader concept than inequality, since it embraces both income levels and income shares.
While the Lorenz Curve provides information on income shares, the Generalized Lorenz Curve sums up both income shares effect with income levels effect for more comprehensive information.

When we compare the Lorenz Curves for the dominance analysis, the most noticeable finding is huge inequality in all years.
In the appendix 1.A, the figures A1.4, A1.5, A1.6 confirm the trends in inequality illustrated by the Gini index values in Table 2.3. In 2002 the poorest 50% of the population received only 13.42% of total income.
The Lorenz curve for 1981 dominates 1990 indicating the increase in inequality, while the Lorenz curve for 2002 dominates 1990 showing the opposite. When comparing the Lorenz curves 2002 and 1981, there is no clear dominance indicating no substantial change in inequality during the last twenty years.

Finally, the Generalized Lorenz Curves summarise the effect of both income levels and income shares on inequality. As already stated, the comparison among Generalized Lorenz Curves is a second-order stochastic dominance analysis.
The figures A1.7, A1.8, A1.9 combine the previous stochastic dominance analyses. Clearly, 2002 dominates both previous years. The main reason for the dominance of 2002 over 1981 is the increase in income levels as we have already seen inequality changed little over this time period. However if we have not conducted the previous stochastic analysis we cannot draw this conclusion.
In contrast, the dominance of 2002 over 1990 is mainly due to the decrease in inequality with the changes in income levels having a smaller effect than the previous comparison as a result of the shorter time period.
When comparing GL curves for 1990 and 1981, there is no clear dominance. This result is maybe due to the rise in income levels being offset by the rise in

income inequality. As we discussed, the accurate interpretation of GL curves requires knowledge of the evolution of the income levels and inequality over the time.

1.3 The determinants of income inequality in Brazil for 2002

The previous section highlighted the crucial role of Brazilian inequality in affecting welfare, suggesting the importance of understanding the determinants of that inequality. Such understanding is a key tool for policy making, as it helps to uncover structural challenges and so to identify which direction interventions should take.

The analysis of the determinants of inequality exploits well-know inequality decomposition techniques. These techniques fall into two broad categories: inequality decomposition by indexes and the regression-based inequality decomposition.

1.3.1 Inequality decomposition by population sub-groups

The methodologies of inequality decomposition by indexes decompose inequality into two parts: an explained between-groups inequality and a residual within-groups inequality. To be able to distinguish these two components, the detection of each group is made by considering specific characteristics. Inequality may be due to the heterogeneity of households or the heterogeneity of income sources.

In the first case, the inequality is decomposed based on differences among households due to factors including geographic location, gender, age and race. This technique, developed by Cowell and Jenkins (1995) is called Inequality Decomposition by Population Sub-groups.

This methodology is based on the assumption that inequality can be divided into an explained component between selected groups and an unexplained component representing within-group inequality. In a static decomposition, each inequality measure that has the property of decomposability, such as the General Entropy class of measures, can be decomposed as follows:

$$I_{tot} = I_b + I_w \tag{1}$$

where the between-group inequality can be written as:

$$I_b = \frac{1}{\alpha^2 - \alpha}\left[\sum_{j=1}^{k} n_j \left(\frac{\mu_j}{\mu}\right)^{\alpha} - 1\right] \tag{2}$$

where the term α is the weight of the GE measure, μ is the overall mean income, μ_j is the mean income for each partition j and n_j is the share of population of each partition j.

The residual within-group inequality is given by the following formula:

$$I_w = \sum_{j=1}^{k} w_j GE(\alpha)_j \tag{3}$$

where $w_j = y_j^{\alpha} n_j^{1-\alpha}$.

The term w_j is a weight given to each subgroup that depends on y_j, the income share, and n_j , the population share for each partition j.

An intuitive and summary measure, R_b, is given by the ratio of the amount of explained between-group inequality, I_b, divided by the total inequality, I_{tot}, as follows:

$$R_b = \frac{I_b}{I_{tot}} \tag{4}$$

The main determinants of inequality in Brazil for 2002 are illustrated by applying this methodology: after elaborating the static decomposition by sub-groups, the estimated results for 2002 are compared with the previous results for 1981 and 1990 calculated by Litchfield (Litchfield, 2001).

To be able to compare the outcomes of inequality decompositions, it is important to apply the same criteria across years in defining population sub-groups.[12]

[12] The sub-groups used are:

- urban and rural, on the basis of the PNAD classification of urban and rural areas;
- region, by aggregating the PNAD municipalities in five regions: North, North-East, South-East, South and Central-West.
- gender of the household head, male or female;

The tables in the appendix 1.A provide the results of the inequality decomposition. Each table reports the values across years of the decomposition as well as the mean incomes and the population shares for every sub-group.

By looking at the table A1.1, it seems that geographic location is a key factor explaining Brazilian inequality between sub-groups of the population. The decomposition between urban and rural areas shows that mean income is much greater in urban than in rural areas. The urban population has increased over time and accounts for 84% of the population in 2002.
The values of the GE class of measures also tell an interesting story. For 1981 and 1990, GE(0) and GE(1) yield higher values in urban areas, while GE(2) is higher in rural areas. In 2002, by contrast, all three indicators yield higher values in urban areas. This suggests a reversed trend from previous years. Knowing that GE(0) and GE(1) are more sensitive at the bottom of the distribution, whilst GE(2) is more sensitive at the top, we can conclude that in 1981 and 1990 inequality was greatest among poor people in urban areas, however in rural areas the presence of a small number of very rich households was the primary source of inequality. By 2002, though, this structure no longer seems to hold, suggesting an increase in inequality between the bottom and the top of the distribution mostly in urban areas.
The decomposition of inequality among regions is equally telling. Mean income varies a lot between regions with, for example, income was twice as high in the South-East than the North-East.
The wealthier regions of Brazil are the South-East and the South and in 2002 43% of the overall population lived in the wealthiest region, the South-East.
Looking at the values of the GE indicators for 2002, the most interesting value is the GE(2) for the North-East, which is the highest of all of the

---

- race of the household head, white, black or Asian, where black also includes mixed and indigenous ethnicities;
- age of the household head, by aggregating into six groups, younger than 25, between 25 and 34, between 35 and 44, between 45 and 54, between 55 and 64 and finally over 65;
- educational attainment of the household head, by aggregating into five groups, illiterate, elementary, intermediate, high school and college.

The only criteria that is not applied for 2002 which was applied in Litchfield's previous work (Litchfield, 2001) is the decomposition by family type, because the classification of family type differs between 2002 (IBGE, 2002a) and the previous years.

regions and is higher than the overall value as well as the value of GE(1). This result reinforces the previous conclusions about rural inequality. The North-East region is the poorest region of Brazil and together with the North, is the most rural region. The high level of inequality explained by GE(2) highlights the existence of very wealthy households among a very poor rural population.

Generally speaking, the decomposition outcomes by geographic location are able to explain income inequality mainly through between-groups inequality rather than within-group inequality.

When considering the decompositions by characteristics of the household head we find equally interesting results. Table A1.2 reports decompositions by gender and by race of the household head, while table A1.3 shows decompositions by age and education level.

The household heads are mainly male, 78% in 2002. However, the comparison across years reveals an increase in the households headed by women. This could be interpreted either as an arbitrary willingness of women to set up their own family or as a voluntary recognition among household members of a female head, although in the majority of the cases it could be an increase of widows, divorced or single women due to the increased instability of familiar relationships and to the biological differences in survival across gender.

Looking at the mean income values, the mean income for male headed households is higher than for female headed households. That said, the values of the GE measures do not tell of dramatic discrepancies between gender: gender does not seem to be critical to decompose inequality. This may reflect that female heads are not a homogenous category.

By contrast, the decompositions by race give more significant results. Mean incomes vary enormously among races: the mean income for white population is twice that for black people. Meanwhile, the mean income of Asians is four times the average black income, though Asians are only 0.5% of the overall population. The GE measures for Asians are very small, suggesting that Asians are a relatively wealthy and homogenous group.

Finally table A1.3 describes inequality decompositions by age and education level of the household head. Generally, the outcome of the decomposition based on age is not significant. Perhaps the most interesting observation is related to the values of GE indexes for household heads over 65: the high value of inequality reveals the presence of a small group of very wealthy retired people.

Inequality decomposition by education displays wide differentials in mean incomes among sub-groups. People with a university degree are only 0.7% of the overall population and earn on average roughly ten times the Brazilian mean income. The big variances in the GE measures convey that between group inequalities are able to explain the main part of overall inequality.

Essentially, looking at the household characteristics, race and education seems to be able to explain overall inequality mainly through between-group inequality, while age and gender explain a tiny amount of between-group inequality.

After examining the summary statistics shown in tables A1.1, A1.2 and A1.3, the table 1.7 here below provides the decomposition results, i.e. the proportion of inequality explained by each factor and for the three GE measures of inequality. In order to compute these values the formula (4) described in the previous section has been used.

It is also important to highlight that the ability to explain inequality by each factor depends on the measure employed. As already pointed out, the three GE measures are sensitive to different parts of income distribution.

Looking at the table 1.7, the most significant determinant of inequality is the education level of the household head. Then geographic location, in term of both urban and region, as well as race have major explanatory power.

Finally, as deduced from the previous summary statistics, age and gender have a negligible importance in explaining the overall inequality.

To sum up, the key determinants of between-group inequality in Brazil are geographic location and the race and education level of the household heads, while age and gender of household heads do not appear to be significant factors.[13]

Table 1.7: The percentage of Total income Inequality explained by Household Differences

	1981(a)			1990(a)			2002(b)		
	GE(2)	GE(1)	G(0)	GE(2)	GE(1)	G(0)	GE(2)	GE(1)	G(0)
Urban	5%	13%	17%	3%	11%	15%	2%	6%	8%
Region	4%	10%	12%	3%	8%	10%	2%	7%	10%
Age	0%	1%	1%	0%	0%	0%	1%	0%	3%
Education	30%	42%	37%	21%	40%	37%	18%	32%	24%
Gender	0%	0%	0%	0%	0%	0%	0%	0%	0%
Race(c)	n.a.	n.a.	n.a.	4%	11%	13%	4%	10%	13%

(a) Source: Litchfield's calculations from PNAD 1981-1995;
(b) Source: Author's calculations from PNAD 2002.
(c) Racial characteristics are not available in 1981.

After examining the results provided by this inequality decomposition by population sub-groups, it is important to highlight that this conventional approach has two fundamental shortcomings (Wan, 2004). First of all, this methodology supplies a high percentage of between components, particularly in decomposing by characteristics such as urban-rural or male-female. Second, this decomposition generates spurious results. In order to be able to compute the impact of a variable on inequality, the decomposition methodology must control for other factors. This limit should be overcome using regression based decompositions, since these methodologies need an identity where the whole income is given by a sum of several income determinants.
These conclusions are the basis for the decomposition analysis presented in the next section.

[13] These results are very much in line with ones produced by Ferreira and Litchfield (2001) by looking at the same type of data over the period 1981-1995. They conclude by claiming that behind Brazilian inequality lies the unequal distribution of education, spatial differences and heterogeneities across ethnicities.

1.3.2 Regression-based inequality decomposition

This section examines three different regression-based decomposition analyses that share the same aim of investigating the main determinants of Brazilian inequality.

The starting point in each regression-based decomposition analysis is the income generating function: to set this function, the factors that contribute to determining income need to be isolated in order to find the explanatory variables for the income regressions. Following this, all of the information given by the econometric estimation of these functions is plugged into specific formulas used in each particular decomposition analysis. There are several decomposition techniques and each of them stresses different elements in establishing the main determinants of inequality.

This study focuses on three techniques.[14] First, the Field's decomposition technique computes the inequality shares, i.e. the contribution of each regressor in determining income inequality. Second, the Oaxaca decomposition explains income differentials by decomposing them into two different effects, the differences in characteristics and the differences in structure. Finally, the Dolton and Makepeace's decomposition exploits the Oaxaca's approach, but focuses on the second moment decomposition instead of the first.

1.3.2.1 Field's decomposition

The regression-based decomposition method developed by Field (Field, 2002) allows for identifying the main factors that determine income differentials.

With this technique it is possible to compute not only the income shares covered by each factor, but also the changes of these income shares.

[14] Due to the large amount of regression-based decomposition techniques, we decide to apply only a small selection of them. In relation to Brazil, there are two important studies looking at income differentials through econometric techniques of decomposition. Ferreira and Paes de Barros (1999) apply a decomposition technique that account for labour incomes, occupational choices and educational decisions. Bourguignon et al (2002) employ the Oaxaca-Blinder decomposition but compare Brazil with U.S. and Mexico.

Essentially, Field's decomposition computes the levels of, and the changes in, income inequality.

This methodology starts with an income generating equation:

$$\ln(Y_i) = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ji} + u_i \tag{5}$$

where Y_i is the income for each observation *i* with *i=1…n*, and X_{ji} are the factors that generate income. The income generating equation is a semi-log function, since the *ln* denotes the natural logarithmic operators applied to per capita real income. This equation may be re-expressed in matrix notation:

$$\ln(Y_i) = \beta' X \tag{6}$$

where $\beta = [\beta_0, \beta_1, \beta_2, ...\beta_k, 1]$ is the vector of the estimated coefficients and the vector of the regressors is given by $X = [1, X_1, X_2, X_3, ...X_k, u]$.

With this income generating equation, the contribution of each factor to determining income may be isolated, so it is possible to quantify the main determinants of the level of income inequality (Krstić and Reilly, 2004).

Once the inequality index is defined on the vector of natural logarithm of income, the levels of income inequality are computed by applying the Shorrocks formula (Shorrocks, 1982). This work uses the Shorrocks' formula as rearranged by Krstić and Reilly (2004) in their more recent paper:

$$S_j[\ln(Y)] = \frac{\operatorname{cov}[\beta_j X_j, \ln(Y)]}{\sigma^2[\ln(Y)]} = \frac{\beta_j * \sigma(X_j) * cor(X_j, \ln(Y))}{\sigma[\ln(Y)]} \tag{7}$$

where $S_j[\ln(Y)]$ defines the share of the j_{th} factor in the inequality of the income measure, β_j are the estimated coefficients, $\sigma(X_j)$ and $\sigma[\ln(Y)]$ are the standard deviations respectively for the regressors and for the dependent variable (i.e. the estimated inequality of the income measure) and, finally, the term $cor(X_j, \ln(Y))$ is the vector of the correlation indexes between regressors and the estimated dependent variable.

Roughly speaking, the estimation of the above shares provides the level decomposition in the Field's framework, and therefore it gives the estimated determinants of income inequality. As already explained, in order to be able to compute these shares, the outcomes of the basic OLS regression for log per capita monthly income are needed.

Looking at the R-squared value at the bottom of the table A1.4 provided in the appendix 1.A, we see that this OLS regression is able to explain more than half of the total variation in income and that the joint statistical significance is good. Regarding the statistical significance of each covariate, only the construction sector does not yield significant results at the 95% confidence level. Nonetheless the interpretation of the estimated coefficients is straightforward and confirms all of the common features of the Brazilian economy.

For example, being male increases income by only 3.9%, again indicating that the gender gap is not a dramatic problem, as is the case in most of Latin America. By contrast, working in the formal sector raises the income by a more significant 24.4%. the same income gain exists for those living in an urban area.

Perhaps, the most interesting results concerns geographic location and ethnicity.

Relative to the North region, living in the North-East decreases income by 10.6%. More strikingly, living in the other three regions increases the income respectively by 32.7% for the South-East, 33.9% for the South and 33.7% for the Central-West.

Regarding race, if we use Asians, a very small relatively rich sub-group, as the comparison group, being white decreases income by only 5.1%, while being black decreases income by 20.4%.

A further striking result relates occupation type: adopting blue-collar workers as the reference category, being a professional increases income by 71.29%, but white collar workers have incomes only 3.8% higher.

Our final variable of interest is the continuous variable for years of school attended. The regression reveals that one more year of education raises income by 10.6%: this is a central result that highlights the crucial role played by education in determining income gaps.

Tables A1.5 and A1.6 report the income ratios for selected groups. The first table shows raw values, based on the ratio of per capita incomes in selected category relative to the base category, while the second table provides *ceteris paribus* relative values that are based on the antilog of the estimated coefficients from the OLS regression described above.

Comparing the raw values with the outcomes of the OLS regression, the sign and the magnitude of the value are similar, with the latter values slightly smaller for each selected group: it seems that the estimation "smoothes" the effects of each factor on the variation of income.

From the OLS regression shown in Table A1.4, it is possible to compute the factor shares in income inequality by applying the Shorrocks' formula already defined in equation (7).

Below, table 1.8 reports a selection of factor inequality shares, so the sum of the listed values is not equal to unity.

Table 1.8: Selected Factor inequality Shares, 2002

Category	2002[(a)]
Region North-East	0.024
Region South-East	0.038
Region South	0.025
Region Central-West	0.0091
Male	0.0004
Whites	0.013
Blacks	0.058
Formal	0.027
Agriculture	0.026
Industry	0.0007
Professionals/ Technicians	0.104
Intermediates	0.0006
Urban	0.031

(a) Source: Author's calculations from PNAD 2002. The sample uses only household head aged between 15 and 80.

However, the reported values are enough to highlight the most important determinants of income inequality for Brazil in 2002: education, ethnicity and geographic location. The inequality share for professionals is further able to explain 10.4% of inequality. This is in turn related to the gain from a graduate or postgraduate education.

Regarding ethnicity, the inequality share for black people explains 5.8% of overall income inequality. As to geographic location, when added together the inequality shares for regions explain 9.6% of overall inequality.

1.3.2.2 The first and second moment decomposition

The Field's decomposition technique allowed us to quantify the effects on inequality of each single factor. The mean and variance decomposition techniques developed by Oaxaca (1973) and Dolton and Makepeace (1985) have slightly different purposes.

In conducting this decomposition, first of all a factor need to be identified as a key determinant of income inequality: in this case, the decomposition uses race and geographic location as the main determinants of Brazilian inequality. Next, the income differential between two sub-categories of each given factor is estimated using OLS regressions: for example, we may calculate the difference between two regions, such as North-East and South-East. Finally, the two decomposition techniques, either for the mean, or for the variance of income, try to disaggregate the estimated differential into two effects: the endowment effect, which identifies differences in characteristics, and the treatment effect, which accounts for differences in structure.

The income equation separately estimated for each sub-group has a semi-log functional form:

$$\ln(Y_i) = X_i'\beta + u_i \tag{8}$$

where Y_i is the household income for each household i, where $i=1...n$, X_i is a vector of the household characteristics, β is a vector of coefficients and u_i is the disturbance term.

By taking two given groups, called group A and group B, the percentage change in the difference between the mean of Y_i for group A and the mean of Y_i for group B is given by the following formula:

$$\Delta\% = \frac{\overline{Y}_A - \overline{Y}_B}{\overline{Y}_B} \tag{9}$$

where $$\ln(\Delta\% + 1) = \overline{\ln(Y_i)}_A - \overline{\ln(Y_i)}_B \tag{10}$$

From the estimation of the income regression for each group, A and B, respectively:

$$\overline{ln(Y)}_A = \overline{X}_A{}'\hat{\beta}_A \tag{11}$$

$$\overline{\ln(Y)}_B = \overline{X}_B{}'\hat{\beta}_B \tag{12}$$

with $\overline{X}_A$ and $\overline{X}_B$ as the vectors of the mean values of the characteristics for group A and group B and $\hat{\beta}_A$ and $\hat{\beta}_B$ as the related vectors of the estimated coefficients.

Then, by subtracting (11) from (12):

$$\overline{\ln(Y)}_A - \overline{ln(Y)}_B = \overline{X}_A{}'\hat{\beta}_A - \overline{X}_B{}'\hat{\beta}_B \tag{13}$$

And by adding and subtracting the additional term $\overline{X}_B\hat{\beta}_A$:

$$\overline{\ln(Y)}_A - \overline{\ln(Y)}_B = \overline{X}_A{}'\hat{\beta}_A + \overline{X}_B\hat{\beta}_A - \overline{X}_B\hat{\beta}_A - \overline{X}_B{}'\hat{\beta}_B \tag{14}$$

The following decomposition of the first moment is obtained:

$$\ln(\Delta\% + 1) = \overline{\ln(Y)}_A - \overline{\ln(Y)}_B = \hat{\beta}_A{}'(\overline{X}_A - \overline{X}_B) + (\hat{\beta}_A - \hat{\beta}_B)'\overline{X}_B \tag{15}$$

where $Endowment = \hat{\beta}_A{}'(\overline{X}_A - \overline{X}_B)$ is the differences in characteristics (16)

and $Treatment = (\hat{\beta}_A - \hat{\beta}_B)'\overline{X}_B$ is the differences in structure (17)

the two terms on the right side respectively identify the differences in characteristics and the differences in structure.[15]

By looking at the (16) and (17) terms, the sampling variances[16] are respectively:

$$Var(Endowment) = (\overline{X}_A - \overline{X}_B)'Var(\hat{\beta}_A)(\overline{X}_A - \overline{X}_B) \tag{18}$$

[15] Both effects need to be tested for statistical significance. Since the standard errors are required to assess the statistical significance, it is important to determine the sampling variances.

[16] As Reilly points out (Paci and Reilly, 2004, p.17), these sampling variances are constructed by recognizing that the income differential is expressed in log points. If it was in percentage points, a delta method to estimate variances would be applied, as Oaxaca and Ransom did (Oaxaca and Ransom, 1998 quoted in Paci and Reilly, 2004).

$$Var(Treatment) = \overline{X}_B' \left[Var(\hat{\beta}_A) - Var(\hat{\beta}_B) \right] \overline{X}_B \qquad (19)$$

Before explaining the second moment decomposition, some critical comments need to be highlighted.

First, the mean decomposition explained earlier uses separate models for each given group. This technique is more complex than the analysis of income differentials by estimating a single pooled equation. In this latter case, the differential is identified simply by a parallel shifting of the regression line, hence the only part of the model that can change is the intercept term (Paci and Reilly, 2004, p.4).

Second, the Oaxaca decomposition is based on the "index number" approach so it has all of the shortcomings related to such an approach. Specifically, equation (15) provides differences in characteristics between group A and group B evaluated at the returns to group A and differences in the estimated relationship between group A and group B evaluated at the mean of the characteristics of group B. Evidently, the formula could be recomputed by looking at differences in characteristics at the returns of group B and differences in structure at the mean characteristics for group A.

This would yield different values due to the fact that techniques involving the index number approach are subject to the usual index number problem[17] (Paci and Reilly, 2004, p.6).

Third, the Oaxaca decomposition is a static methodology, as it analyzes the endowment and treatment effects at a given time. Juhn, Murphy and Pierce (1991) introduced a dynamic dimension of the decomposition of the first moment, as Reilly stated (Paci and Reilly, 2004).

Finally, this first moment decomposition can also be carried out with the *quantile regression methodology*, developed by Koenker and Hallock (2001). This technique estimates income differentials at a given quantile of the conditional income distribution instead of taking average values. This estimation method is called Least Absolute Deviation (LAD) and aims to

[17] A decomposition methodology that attempts to overcome this limitation is the Cotton-Neumark Decomposition technique (Cotton and Neumark, 1988 quoted in Paci and Reilly, 2004, p.6).

minimize the absolute sums of the errors rather than the sum of squared errors, as in the OLS method.

Variance differentials are explored by applying the second moment decomposition developed by Dolton and Makepeace (Dolton and Makepeace, 1985, quoted in Callan and Reilly, 1993).

With the analysis of the second moment, it is possible to examine the differences between the variances of the income distributions for two given sub-groups.[18] Therefore, if the first moment decomposition studies between-group inequality, the second moment decomposition looks into within-group inequality.

As Reilly points out (Callan and Reilly, 1993), the variance decomposition might give a considerable residual, hence it might happen that a portion of the variance differential cannot be explained by this decomposition technique. This is due to the non-linearity associated with the variance decomposition. By contrast, the mean decomposition is able to explain all of the values of the income differentials between the two effects.

Using the Dolton and Makepeace approach, the variance differential decomposition is:

$$\hat{s}_A - \hat{s}_B \cong \left[\hat{\sigma}_A^2 - \hat{\sigma}_B^2 + (\hat{\beta}_A - \hat{\beta}_B)'\Omega(X_B)(\hat{\beta}_A - \hat{\beta}_B)\right] + \hat{\beta}_A'\left[\Omega(X_A) - \Omega(X_B)\right]\hat{\beta}_A \qquad (20)$$

where $\hat{s}_A$ and $\hat{s}_B$ are the estimated variances for group A and group B, $\hat{\sigma}_A^2$ and $\hat{\sigma}_B^2$ are the estimated variances of the errors and, finally, $\Omega(X_A)$ and $\Omega(X_B)$ are the variance-covariance matrix of characteristics respectively for groups A and B.

On the right side of equation (20) the first term in square brackets accounts for differences in structure, while the second term indicates differences in characteristics.

Decomposition by race Tables A1.7, A1.8 and A1.9 in the appendix 1.A present the results of the decomposition by race. The ethnicity variable has been chosen as one of the main factors that may determine income inequality

[18] As in the case of mean decomposition, the variance is decomposed in two effects: the differences in characteristics and the differences in structure.

in Brazil on the basis of the results provided above. It is worthwhile to remember how this variable has been aggregated.

While the categories of white people and of Asians embrace only one ethnicity, the black category includes not only black people, but also mixed and indigenous populations. Because the category “Asian” used earlier is negligible size, it has seemed convenient to drop this category and analyse the decomposition simply between whites and blacks.

Table A1.7 reports the two OLS regressions, for black and for white population respectively. The differentials between the two categories are listed in the last column of the table. The coefficients on the regressors for region and for education provide important information about the differences between whites and blacks.

In the North-East region, the entire population earns less with respect to the North region, but the black population earns 3% more than the white population. In the Central-West region, the black population income is 0.1% greater than white population income, holding the North region as baseline category. On the contrary, compared with the white population, black population income is by 1% lower in the South-East and by 2% lower in the South with respect to the North.

The black population earns less than the white population in all regions, but, compared with the base category North, the discrepancies seems to be sharper in the two wealthier regions, the South-East and the South. This indicates that the effect of discrimination by race is even more pronounced when geographic disparities are taken into account.

Looking at the coefficients for years of education, one year of schooling increases the income of black people by 1.8% less than for whites: returns to schooling are higher for whites than for blacks.

In the regression results for black people, only two variables are not statistically significant at 95% confidence, the “construction” economic sector and the “intermediate” occupation type, while the regression results for white people report only gender as non statistically significant.

Table A1.8 presents the Oaxaca decomposition results. As already described, mean income differentials are disaggregated into an endowment effect and a treatment effect. In the case of the decomposition between blacks and whites, the mean income differential is equal to -0.65: being black means having an income that is on average 48% less than that for whites. Of this effect, -0.458 represents an endowment effect while -0.192 is the treatment effect.

This suggests that differences in characteristics are more relevant than differences in structure in the determination of income differentials between whites and blacks. In other words, black people earn less than white people primarily due to their characteristics, such as education or family structure, rather than due to direct discrimination indicated by smaller returns for black people holding other characteristics constant.

This is an important finding, but we must be careful in understanding how the concept of direct discrimination has been defined: more complex form of discrimination may lay behind differences in characteristics across races and this discrimination is more difficult to detect as well as to eradicate.

Finally, table A1.9 summarizes the main findings. The endowment and treatment effects on income differentials are tested and are found to be statistically significant. Then, at the bottom of the table information about the variance decomposition is provided: we find that the variance for whites is greater than the variance for blacks. This could suggest greater inequality within the white population than within the black population as this analysis explores within group inequality. Once again, the main part of this latter gap is explained by differences in characteristics. If the main part of the income differential is due to differences in characteristics, there is a high probability that the variance gap will be primarily explained by differences in characteristics as well.

Decompositions by region The decomposition by region has been conducted by comparing the poorest region in Brazil, the North-East, with each other region in order to quantify the regional income gaps.

For each comparison, three tables are provided following the same structure as the decomposition analysis by races.

Looking first at the OLS regression results and, in particular, by looking at the coefficient differentials,[19] some common features can be depicted. The greater differentials in the coefficients are given by the regressors related to race or education.

Compared to Asians, the white population has less income throughout all of the regions. The same pattern holds for the black population, but with even greater income differences. Regarding the education variable, the North-East region has less return to education than any other region.

However it is important to outline that the coefficients for race have been found to be not statistically significant in most of the regressions by regions. Consequentially, these coefficients and their impacts should be analysed taking into account this limitation.

The coefficient gaps on the regressors for economic sectors vary significantly across regions: this large variety may be due to the fact that economic activities themselves vary a lot between regions.

Similarly to race coefficients, looking at all of these regression results by region, some coefficients for sectors have been found to be not statistically significant at 95% confidence. Again, care should be applied in interpreting their effects.

The estimated income differentials are the following:[20] the North East region has a mean income that is smaller by 22%, 52%, 54% and 44% relative to the North, the South-East, the South and the Central-West respectively. Even the first moment decomposition by region confirms that the North-East region is the poorest region in Brazil.

While these results are not new, the decomposition between endowment and treatment effects may be more insightful. We find that while income differentials are due mainly to differences in characteristics when the North-

---

[19] The coefficient differentials are given by tables A1.10, A1.13, A1.16 and A1.19 in the appendix 1.A.

[20] These percentages are computed by taking the antilog of the values shown in tables A1.12, A1.15, A1.18 and A1.21.

East is compared to the North, in all of the other cases, the income differentials are primarily explained by differences in structure.

The most obvious explanation is that the North-East and the North have many common features, and as such income differentials are likely to stem from differences in the characteristic of people. Indeed, the key components of the endowment effect are education, sector and urban: being more educated, working in some sectors such as public administration and social services, or living in urban areas increase income in the North with respect to the North-East.

The comparison between North-East and South-East, South and Central-West highlights the crucial role played by the treatment effect, which is representative of structural differences between regions that generate different returns for the same characteristics. The factors that play a key role in the determination of the treatment effect are years of education as well as economic sectors, occupational type and household type.

Each of these factors generates smaller returns in the North-East relative to the richer regions. Finally, when we consider the variance decomposition by region, the variance for the North-East region is always greater than for all of the other regions. This suggests that within region inequality in the North-East is much bigger than the rest of the country. This finding seems to confirm previous observations. In fact, the GE class of measures in the decomposition by regions shows a sharp increase in inequality at the top of the income distribution for the North-East region. Hence, these results related to the variance decomposition are in line with the findings generated by different methodology.

Here again, the variance differentials seems to be generated by the same effect as their respective mean differentials, so the same explanations can be applied. The variance gap between North-East and North is principally due to differences in characteristics, while the variance gap between the North-East and all of the other regions is mainly the result of differences in structure.

Regarding the reliability of this methodology in explaining variance differentials, some final critical comments need to be emphasized. While the mean income differentials are totally explained by the sum of the two estimated effects, endowment and treatment, the variance decomposition can be explained only partially. In the decomposition by races, the variance gap is equal to 0.198 and the sum of the two decomposition effect is only 0.1189. Hence the variance decomposition was unable to explain 40% of the variance differential.

In the decomposition by regions, the variance gap between North-East and North, as well as between North-East and South-East, is equal to 0.147. However, the decomposition is able to explain respectively 0.0937 and 0.054 of the variance differential. Hence "the degrees of explanation" are respectively 64% and 36%.

Similarly, the variance gap between North-East and South is equal to 0.218 and the two effects together explain only 45%. Finally the estimation for the decomposition between North-East and Central-West is even worse: the gap is equal to 0.033, but the sum of the estimated endowment and treatment effects is equal to -0.002.

The previous findings provide evidence of the shortcomings of the decomposition methodology when analysing non-linear variables, as the estimated differentials fall drastically short of 100%.

1.4 Conclusions

This work has tried to throw new light on the determinants of inequality in Brazil. Here we summarize the main findings.

After quantifying Brazilian inequality and recognizing that the recovering from the accumulated inequality during the last two decades is still too weak, inequality decomposition techniques have been applied. Some of the most well-known inequality decomposition methodologies aim to categorize possible determinants of inequality. Hence, ultimately these techniques and their findings can play a crucial role in the identification of policies.

Although poverty and inequality have declined over the past decade, after a sharp rise in the '80s, poverty remains a terrible concern in Brazil. Indeed, in 2002 one third of its population was considered poor.

The inequality situation is still a deeper concern. In 2002, the Gini index was equal to 58.1 and the income distribution was sharply skewed on the right. The unequal distribution of Brazilian income is even clearer when looking at the Lorenz curve for 2002: half of the Brazilian population owns only 13.42% of total GDP, while the richest 10% of the population holds 45.5% of total Brazilian GDP.

A further finding is that inequality followed a similar pattern to poverty, particularising and falling in tandem in response to unstable economic growth, a depressed employment situation and volatile inflation during last two decades. However, it seems that poverty is more sensitive to economic performances than inequality. Poverty grew faster in the 1980s and recovered faster in the '90s, while inequality remained relatively more stable, albeit high, across the last twenty years. To some degree, it is not surprising that an absolute indicator, poverty, is more dependent on changes in prices or to a devaluation process, than a more structural, relative variable like inequality.

Nonetheless, the results highlight that while Brazil is experiencing an improving macroeconomic situation, with a more stable inflation and higher economic growth, the country is failing in the fight against inequality (Bourguignon and Ferreira, 2000). Hence we have sought to examine the deeper causes of this inequality.

By applying inequality decompositions by population sub-groups and a regression-based inequality decomposition technique developed by Field, this work confirmed the findings of several well-know works on the determinants of Brazilian inequality (Ferreira and Litchfield, 2001). Brazilian inequality is rooted primarily in the differences across regions, educational levels and races. More precisely, the inequality factors shares computed with the Field's decomposition have shown that the main portion of Brazilian inequality is explained by these three factors.

Moving to the first and second moment decomposition methods, we find that the initial results are confirmed, while additional detail is revealed. As was already explained, these two techniques allow us to split mean income differentials and variance differentials into two effects: the endowment effect, which detects differences in characteristics, and the treatment effects, which accounts for differences in structure.

In other words, these decompositions quantify the portion of income differentials which are the results of the differences in endowments of income generating factors and the portion which are the results of differing returns to the same factors.

This allows us to investigate the effect of discrimination of various kinds on different groups. Referring to several studies (Oaxaca, 1973 and Callan and Reilly, 1993), estimation of the discrimination effect is not always straightforward.

According to Oaxaca (1973), differences in returns to the same characteristics are clearly a strong sign of discrimination; hence the treatment effect could be interpreted as the "pure" discrimination effect. However, differences in characteristics between two given population groups often involve more subtle forms of discrimination, which are even more difficult to eradicate.

Looking at the empirical findings revealed by the application of these two decomposition techniques offers some clarity.

On the question of race, in computing income differentials between blacks and whites, the Brazilian black population was found to be poorer than Brazilian whites, earning 48% less on average. The income gap between these two ethnicities is even sharper in the wealthier regions such as the South and the South-East. Moreover, returns to education are weaker for black people with respect to whites.

Finally, the difference between their respective variances show a higher income variance among white people, implying that the black population is poorer and more homogenous than whites, as the latter category embraces both very wealthy and extremely indigent people.

Both the first and second moment decompositions revealed that income inequality among races is mainly due to differences in characteristics. On the line of the previous interpretation, income discrimination between black and white Brazilian populations seems to be caused not by a "direct" discrimination against blacks.

There may though be more subtle discrimination if it is the case that this part of the population is interdicted at the first stages from the possibility to reach a wealthier status: being black may increase the probability of living in less wealthy areas or in more troubled family situations, and may imply access to less remunerated jobs and, lower quality education.

The computations of regional income differentials provide even more interesting results. The decomposition analysis was conducted by comparing the poorest region, the North-East with each of the other regions. Using both the first moment and the second moment decomposition methods, the determinants of regional income inequality varies significantly among comparisons depending on the regions being compared.

In particular, when comparing the North-East with the second poorest Brazilian region, the North, the mean income differentials as well as the variances differentials, are due to differences in characteristics. The North-East region is poorer than the North, this seems to be due to different and less favourable characteristics for the North-East.

By contrast, differentials between the North-East and the other three wealthier regions reveal that differences in structure are the key determinants. The North-East region has lower returns than the other three regions by holding characteristics constant.

For example, a white man in the North-East probably earns less than a comparable white man with the same level of education living in the South-East or a household in the North-East consisting of a single parent and two children has a higher probability of being indigent than a household with the same structure in the South.

These findings may be crucial in policy making targeting regional differences, particularly in recognizing that different kinds of discrimination generate income differentials in different regions.

Although this study focused on decomposition techniques by race and region, further research could involve other inequality factors to improve upon this exploration of the determinants of Brazilian inequality, such as the level of educational attainment.

Further research should be focused not only in improvement involving other inequality factors, but also in overcoming the highlighted shortcoming in second moment decomposition analysis, since it has been shown that it generates unreliable results with consistently large residuals due to the non-linearity of the variables.

As understanding of the determinants of inequality deepens, it becomes a matter for the politics to define possible policy interventions. That says, a few comments are possible. In implementing policies, it should be borne in mind that several studies (Litchfield, 2001 and Ferreira and Paes de Barros, 1999), including this one, have confirmed the strong correlation among poverty and inequality and macroeconomic variables such as GDP growth, employment and inflation.

Structural adjustments in the long run, as well as stabilization programmes in the short run, do affect poverty and inequality phenomena and this is exactly what happened in Brazil in the last two decades. Nevertheless, this study intends to conclude by highlighting that political institutions and social infrastructure are an important part of this story.

There is no doubt that macroeconomic performance affects and is affected by, poverty and inequality. It is equally clear that political institutions and social conditions are strictly interconnected with macroeconomic variables and with poverty and inequality issues.

Due to the high level of inequality, Brazil suffers failures in good governance in political institutions: an economically and socially unequal society cannot guarantee an effective government partnership.

Economics and econometrics are useful and essential to identify and quantify the welfare and inequality profile as well as the inequality determinants of a complex country such as Brazil. However, the complexity of a phenomenon like inequality has deep roots not only in economic reasons, but primarily in historical and sociological explanations.

Therefore, the key conclusions emanating from this study are the strong belief that further economic investigations and more complex econometric techniques are surely required to move ahead the analysis of Brazilian inequality determinants. Nevertheless the indisputable *raison d'être* of inequality lays in more anthropological explanations, which can only be deduced from further explorations.

References 1

Altimir, O., (1977), “Income distribution Statistics in Latin America and their Reliability”, *Review of Income and Wealth*, Vol. 33, No. 2, pp. 111-155.

Baer, W., (2001), “The Brazilian Economy, Growth and Development”, 5th Edition, Wesport: Praeger Publishers.

Bourguignon, F., (2004), “The Poverty-Growth-Inequality Triangle”, Paper prepared for a Conference on Poverty, Inequality and Growth, Agence Française de Développement /EU Development Network, Paris, November 13, 2003, The World Bank.

Bourguignon, F., and F.H.G. Ferreira, (2000), “Understanding inequality in Brazil: a conceptual overview”, Departamento de Economia PUC-Rio, Texto para Discussão, No.434.

Bourguignon, F., F. Ferreira, and Phillipe Leite, (2002), “Beyond Oaxaca-Blinder: Accounting for Differences in Household Income Distribution Across Countries”, *DELTA Working Paper*, No. 04, Paris: DELTA.

Callan, T., and B. Reilly, (1993), “Unions and the Wage Distribution in Ireland”, *The Economic and Social Review*, Vol. 24, No. 4, pp. 297-312.

Cowell, F.A., and S.P. Jenkins, (1995), “How much inequality can we explain? A methodology and an application to the USA”, *Economic Journal*, No. 105, pp. 421-430.

Datt, G., and M. Ravallion, (1992), “Growth and Redistribution Components of Changes in Poverty Measures”, *Journal of Development Economics*, Vol. 38, No.2, pp. 275-95.

Deaton, A., (1997), “The Analysis of Household Survey: A Microeconometric Approach to Development Policy”, Baltimore and London: John Hopkins Press and World Bank.

Deininger, K., and L. Squire, (1998), “New Ways of Looking at Old Issues: Asset Inequality and Growth”, *Journal of Development Economics*, Vol. 57, No.2, pp. 259-297.

Dolton, P.J., G.H. Makepeace, (1985), “The Statistical Measurement of Discrimination”, *Economics Letters*, Vol. 18, No. 4, pp. 391-395.

ECLAC, (2004), “Statistical Yearbook for Latin America and the Caribbean”, Santiago de Chile: ECLAC.

Elbers, C., J.O. Lanjouw, P. Lanjouw, and P. Leite, (2004), “Poverty and Inequality in Brazil: New Estimates from Combined PPV-PNAD Data”, mimeo, Washington, D.C.: World Bank.

Ferreira, F.H.G, and J.A. Litchfield, (2001), “Education or Inflation? The Micro and Macroeconomics of the Brazilian Income Distribution during 1981-1995”, in N. Hicks and A. Valdés (eds) Special Edition of Cuadernos de Economia [Latin American Journal of Economics], Vol. 114, pp. 209-238.

Ferreira, F.H.G., and J.A. Litchfield (2000), “Desigualdade, Pobreza e Bem-Estar Social no Brasil: 1981-1995”, in Henriques, Ricardo (ed) Desigualdade e Pobreza no Brasil, Rio de Janeiro: Instituto de Pesquisa Economica Aplicada.

Ferreira, F.H.G., and R. Paes de Barros (1999), “The Slippery slope: Explaining the increase in Extreme Poverty in urban Brazil, 1976-1996”, *The Brazilian Review of Econometrics*, Vol. 19, No. 2.

Field, G., (2002), “Accounting for Income Inequality and its change: a New Method, with Application to the Distribution of Earnings in the United States”, mimeo, Department of Labor Economics, Cornell University, Ithaca.

Foster, J., J. Greer, and E. Thorbecke, (1984), “A Class of Decomposable Poverty Measures”, *Econometrica*, No. 52, pp. 761-65.

Graham, B., (2004), “Development, Inequality and Poverty in Brazil”, *Student Economic Review*, Vol.18.

IBGE, (2002a), "Perquisa Nacional por Amostra de Domicilio: Sintese de Indicatores- 2002", Rio de Janeiro: IBGE.

IBGE, (2002b), "Perquisa Nacional por Amostra de Domicilio: Notas Tecnicas- 2002", Rio de Janeiro: IBGE.

Juhn, C., K. Murphy, and B. Pierce, (1993), "Wage Inequality and the Rise in Returns to Skill", *The Journal of Political Economy*, Vol. 101. No. 3, pp. 410-442.

Koenker, R., K. Hallock, (2001), "Quantile Regression", *Journal of Economic Perspectives*, Vol. 15, No. 4.

Krstić, G., and B. Reilly, (2004), "An Anatomy of Labour Market Earnings Inequality in Serbia – 1995 to 2000", Brighton: University of Sussex.

Litchfield, J. A., (2001), "Welfare and Income Distribution in Brazil, 1981-1995", Ph.D. Dissertation, London School of Economics, London: LSE.

Lluch, C., (1982), "Sobre Medicões de Renda a Partir dos Censos e da Contas Nacionais no Brasil", *Pesquisa e Planejamento Econômico*, Vol. 12, No. 1, pp. 133-48.

Lopez, H., (2004), "Pro-poor growth: a review of what we know (and of what we don't)", mimeo, The World Bank (PRMPR).

Mejia, J.A., and R. Vos, (1997), "Poverty in Latin America and the Caribbean. An Inventory: 1990-95", INDES Working Paper Series I-4 prepared for the Program "Improvement if Surveys and the Measurement of Living Conditions in Latin America and the Caribbean" co-sponsored by IDB, WB and CEPAL, Washington D.C.: IADB.

Mejia, J.A., and Rob Vos, (1997), "Poverty in Latin America and the Caribbean. An Inventory: 1990-95", INDES Working Paper Series I-4 prepared for the Program "Improvement if Surveys and the Measurement of Living Conditions in Latin America and the Caribbean" co-sponsored by IDB, WB and CEPAL, Washington D.C.: IADB.

Oaxaca, R., (1973), "Male-Female Wage Differentials in Urban Labor Markets", *International Economic Review*, Vol. 14, No. 3, pp. 693-709.

Oaxaca, R., M. Ransom, (1994), "On Discrimination and the Decomposition of Wage Differentials", *Journal of Econometrics*, Vol. 61, pp. 5-21.

Paci, P., and B. Reilly, (2004), "Does Economic Liberalization Reduce Gender Inequality in the Labor Market: the Experience of the Transition Economies of Europe and Central Asia", Report prepared for the Poverty Reduction & Economic Management Group. Eastern Europe & Central Asia Region, World Bank, Washington D.C.: World Bank.

Rocha, S., (1993), "Poverty Lines for Brazil: New Estimates from Recent Empirical Evidence", Instituto de Pesquisa Economica Aplicada, Unpublished Paper, Rio de Janeiro: IPEA.

Rocha, S., (1996), "Renda e Pobreza: Os impactos do Plano Real", Texto Para Discussão, No. 439, Rio de Janeiro: IPEA.

Rocha, S., (2000), "Opções metodologicas para a estimaç ão de linhas de indigencia e de pobreza no Brasil", Texto Para Discussão, No. 720, Rio de Janeiro: IPEA.

Rocha, S., (2004), "Pobreza no Brasil: O que mudou nos ultimos 30 anos?", Estudos e Pesquisas, No. 83, Rio de Janeiro: INAE.

Shorrocks, A. F., (1982) "Inequality Decomposition by Factor Components", *Econometrica*, Vol. 50, No. 1, pp. 193-212.

Szekely, M., N. Lustig, M. Cumpa, and J.A. Meja, "Do we Know How Much Poverty There Is?", *IADB Working Paper*, No. 437, Washington, D.C.: IADB.

UNDP, (2002), "Human Development Report 2002: Deepening Democracy in a Fragmented World", New York: UNDP.

Wan, G., (2004), "Accounting for income inequality in rural China: a regression-based approach", Journal of Comparative Economics, No. 32, pp. 348-363.

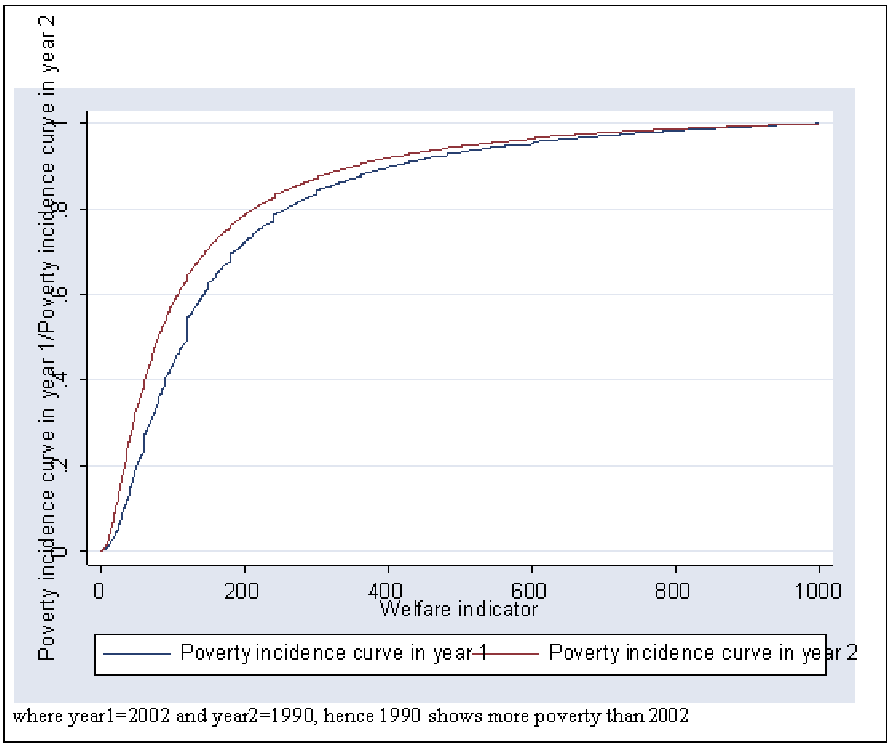
Wodon, Q., (2000), "Poverty and Policy in Latin America and the Caribbean", *World Bank Technical Paper*, No. 467, Washington, D.C.: World Bank.

World Bank, (2003), "Brazil Inequality and Economic Development", Vol.1: Policy Report, A Joint Report by IPEA and Brazil Country Management Unit Poverty Reduction and Economic Management Sector Unit Latin American and the Caribbean Region, Report No. 24487-BR, Washington D.C.: World Bank.

Appendix 1.A

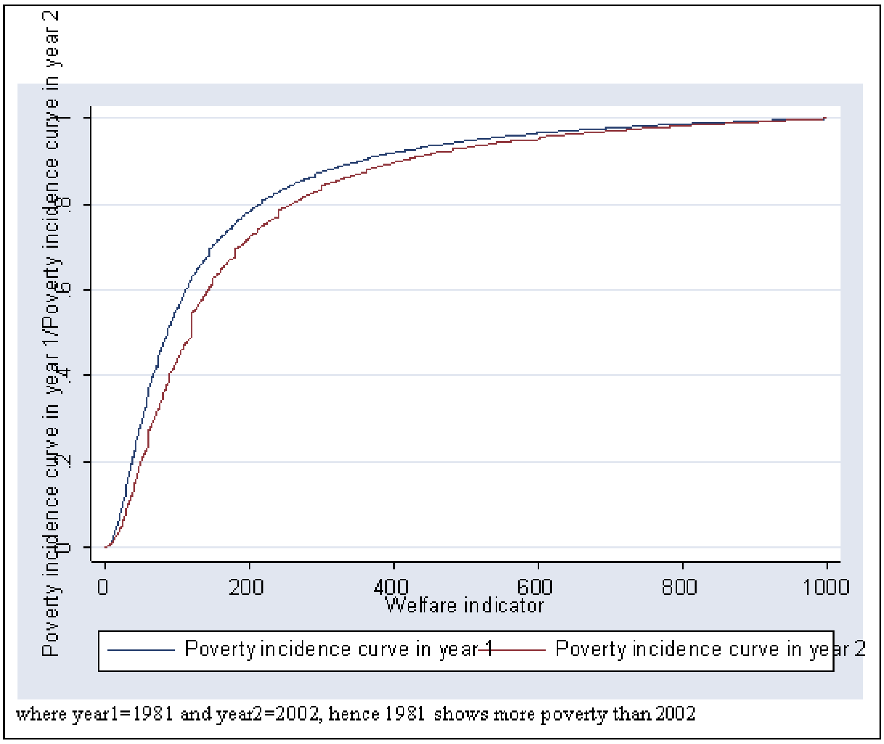
Graphs and Tables

Figure A1.1: Poverty Incidence between 2002 and 1990



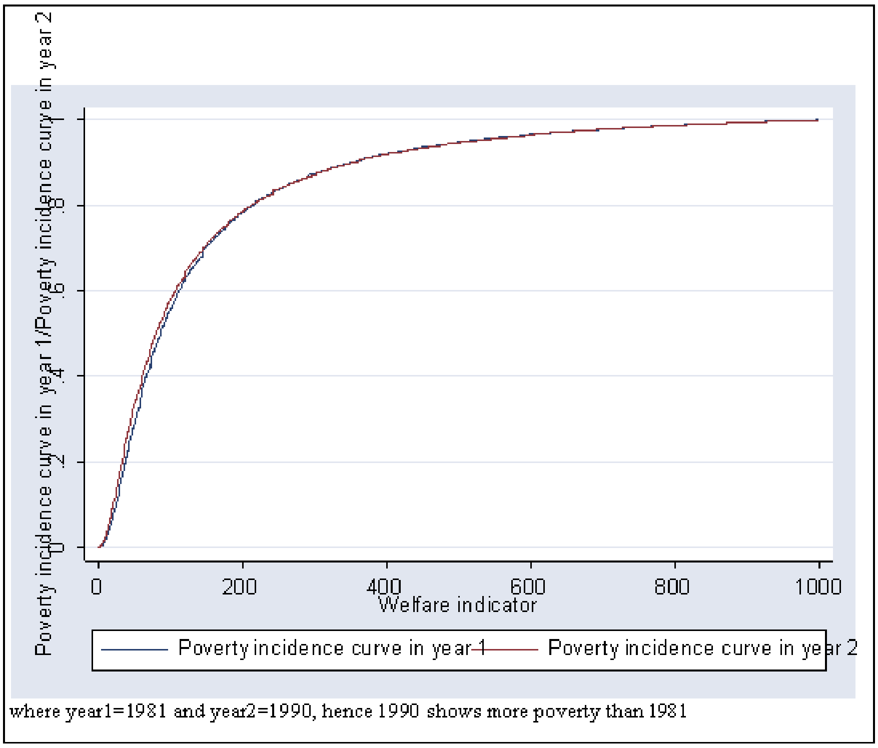
Source: Author's calculation from PNAD 1990 and 2002.

Figure A1.2: Poverty Incidence between 2002 and 1981



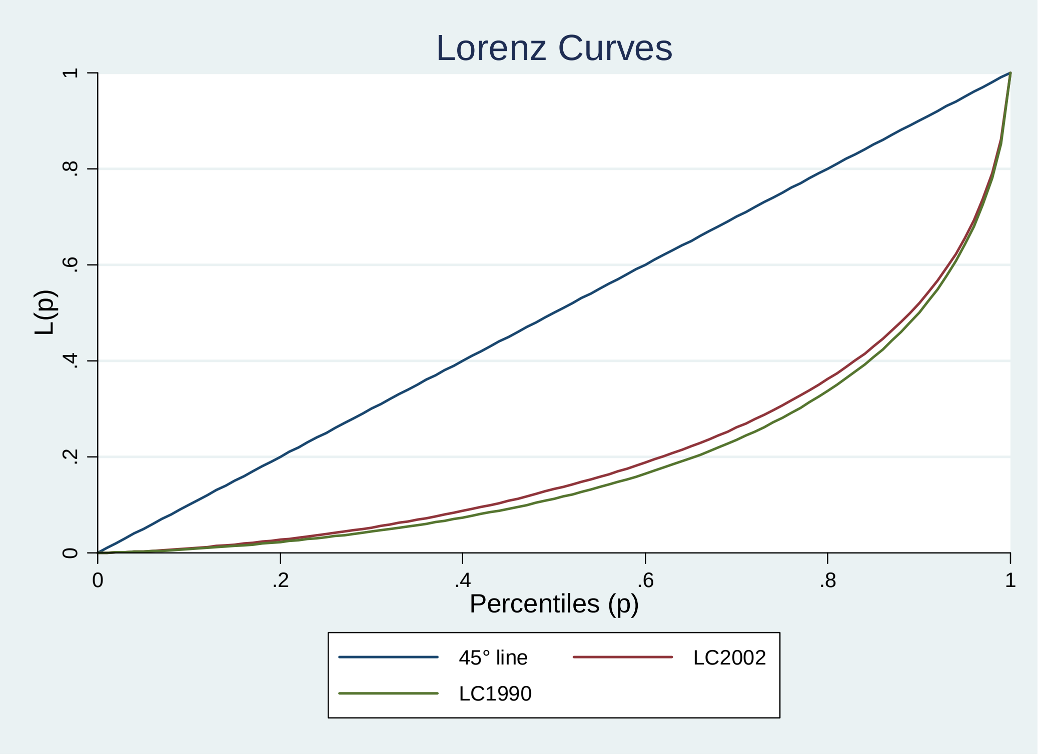
Source: Author's calculation from PNAD 1981 and 2002.

Figure A1.3: Poverty Incidence between 1990 and 1981



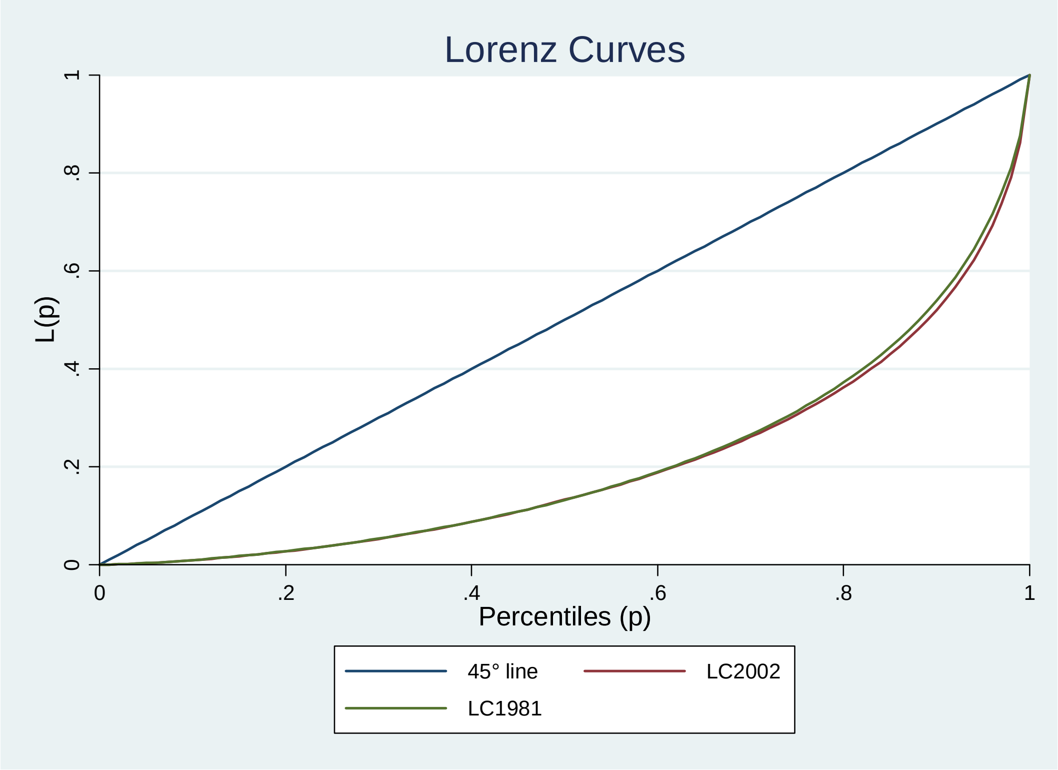
Source: Author's calculation from PNAD 1990 and 2002.

Figure A1.4: Lorenz dominance between 2002 and 1990



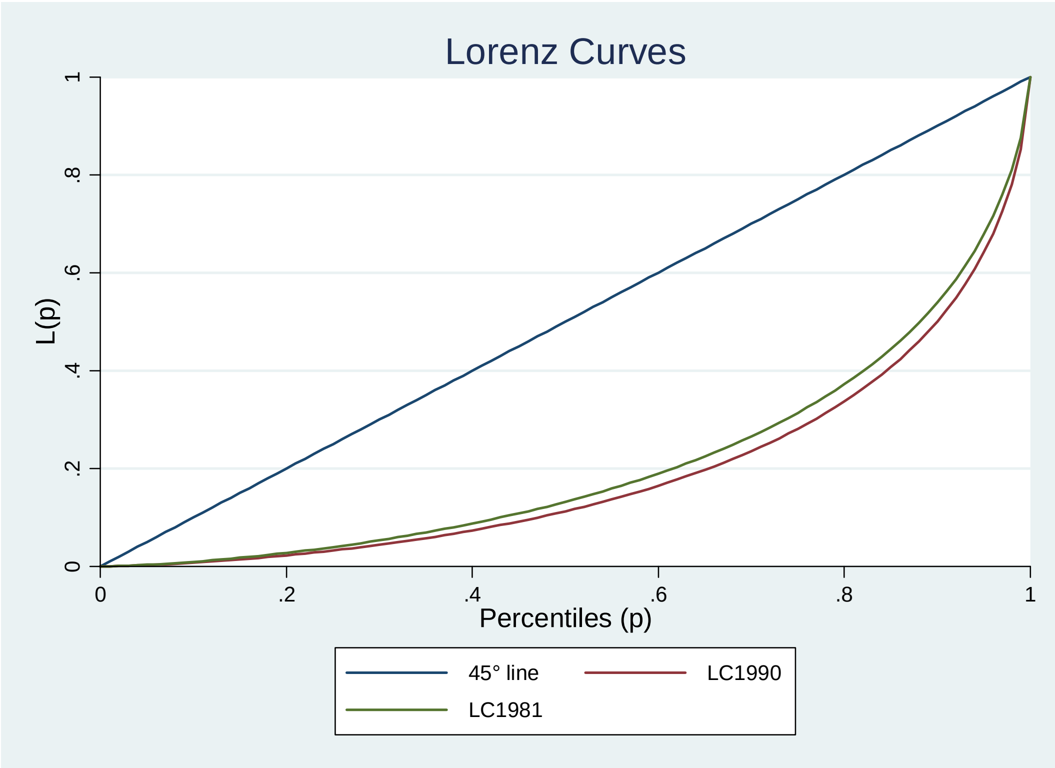
Source: Author's calculation from PNAD 1990 and 2002.

Figure A1.5: Lorenz dominance between 2002 and 1981



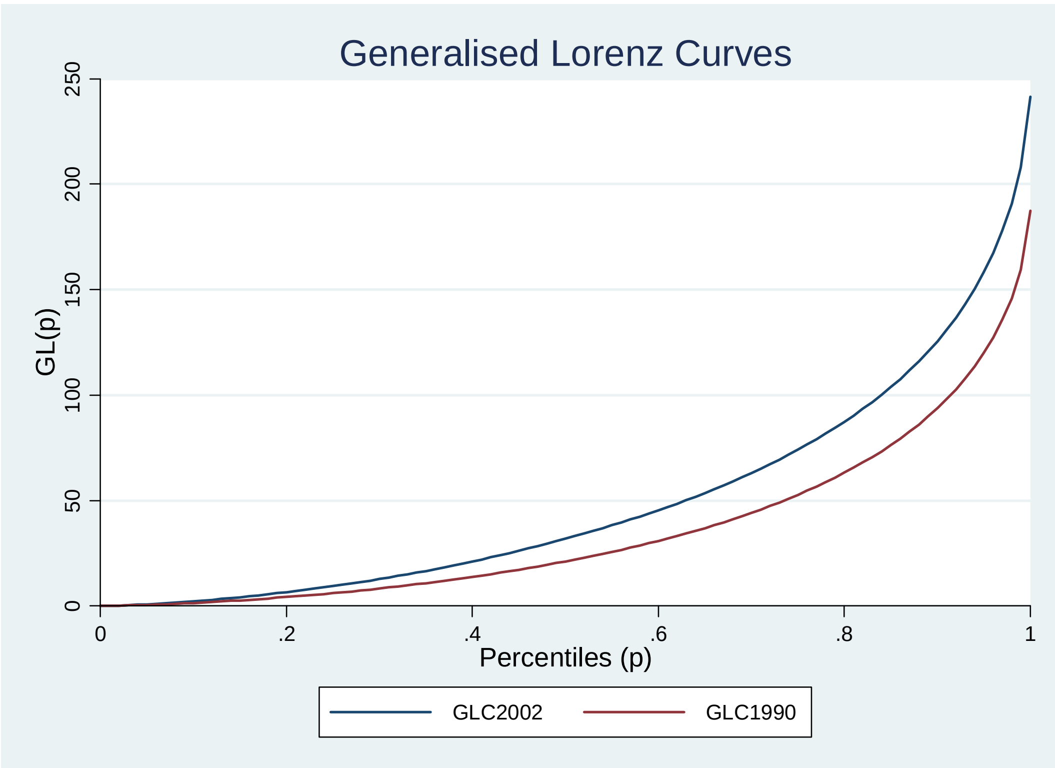
Source: Author's calculation from PNAD 1981 and 2002.

Figure A1.6: Lorenz dominance between 1990 and 1981



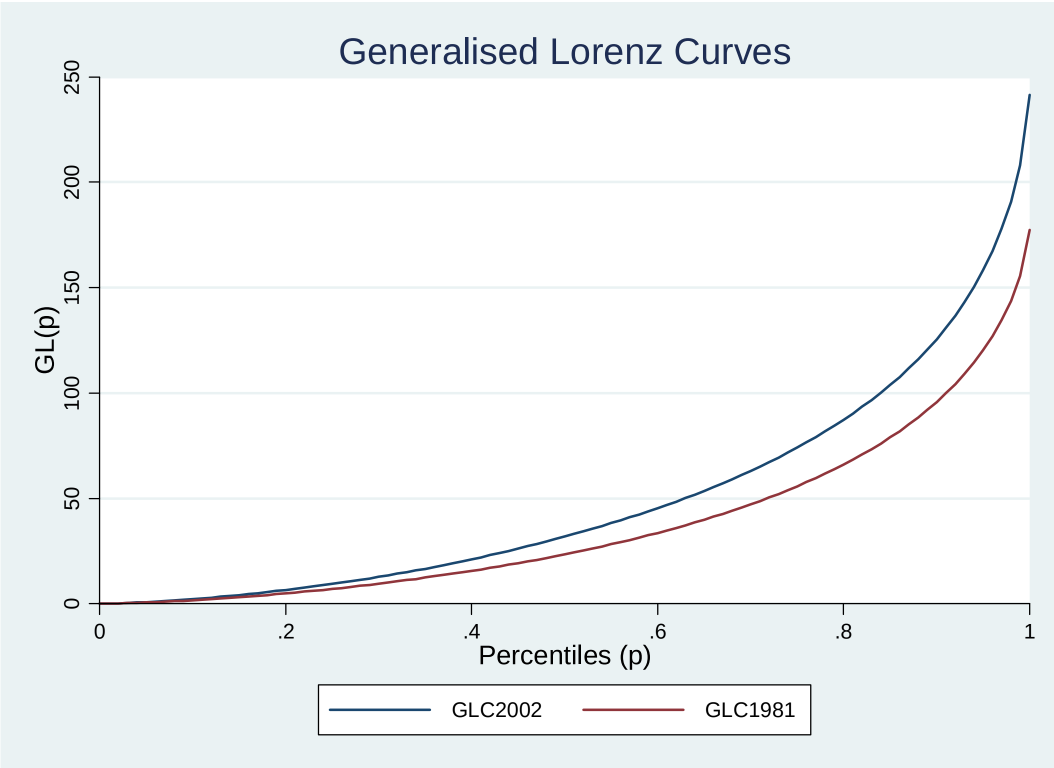
Source: Author's calculation from PNAD 1981 and 1990.

Figure A1.7: Second order stochastic dominance between 2002 and 1990



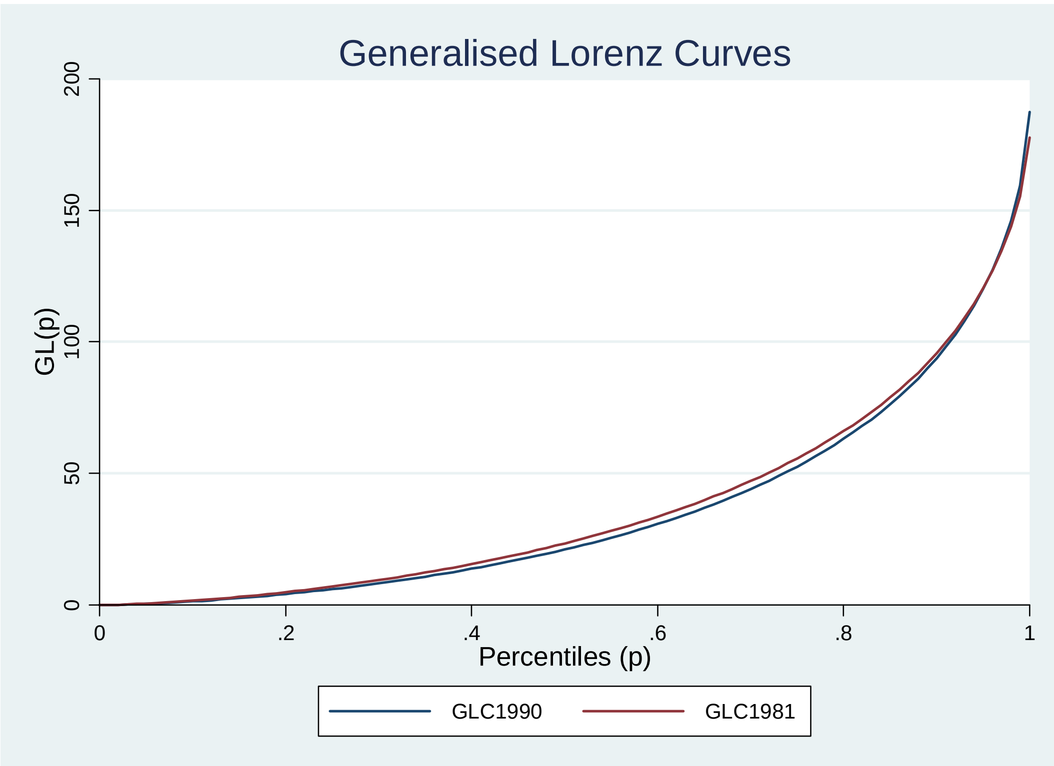
Source: Author's calculation from PNAD 1990 and 2002.

Figure A1.8: Second order stochastic dominance between 2002 and 1981



Source: Author's calculation from PNAD 1981 and 2002.

Figure A1.9: Second order stochastic dominance between 1990 and 1981



Source: Author's calculation from PNAD 1981 and 1990.

Table A1.1: Summary Statistics of Households Income per Capita, by urban and by region, 2002

	1981[a]					1990[a]					2002[b]				
Urban/Rural	**Mean**	**Pop %**	**GE2**	**GE1**	**GE0**	**Mean**	**Pop %**	**GE2**	**GE1**	**GE0**	**Mean**	**Pop %**	**GE2**	**GE1**	**GE0**
Urban	168	71	1.09	0.57	0.54	183	74	1.71	0.67	0.62	221.03	84	1.91	0.65	0.58
Rural	56	29	1.64	0.53	0.44	57	26	1.83	0.59	0.53	80.67	16	1.31	0.52	0.47
All	136	100	1.34	0.65	0.61	150	100	2.02	0.74	0.70	198.7	100	2.06	0.69	0.63
Region	**Mean**	**Pop %**	**GE2**	**GE1**	**GE0**	**Mean**	**Pop %**	**GE2**	**GE1**	**GE0**	**Mean**	**Pop %**	**GE2**	**GE1**	**GE0**
Southeast	182	44	1.06	0.56	0.53	192	45	1.74	0.64	0.59	250.53	43	2.04	0.62	0.55
South	139	16	1.09	0.55	0.51	156	16	1.38	0.64	0.61	228.06	15	1.08	0.53	0.49
Northeast	70	30	1.84	0.68	0.57	76	29	2.55	0.84	0.70	109.57	29	2.15	0.76	0.63
C-West	128	7	1.47	0.65	0.58	173	7	1.83	0.74	0.68	227.45	7	1.66	0.69	0.61
North	121	3	1.09	0.51	0.44	160	3	2.48	0.72	0.62	143.08	6	1.59	0.65	0.55
All	136	100	1.34	0.65	0.61	150	100	2.02	0.74	0.70	198.7	100	2.06	0.69	0.63

(a) Source: Litchfield's calculations from PNAD 1981-1995;
(b) Source: Author's calculations from PNAD 2002.

Table A1.2: Summary Statistics of Households Income per Capita, by gender and by race, 2002

	1981[a]					1990[a]					2002[b]				
Gender	**Mean**	**Pop %**	**GE2**	**GE1**	**GE0**	**Mean**	**Pop %**	**GE2**	**GE1**	**GE0**	**Mean**	**Pop %**	**GE2**	**GE1**	**GE0**
Male	137	89	1.35	0.65	0.62	152	86	2.07	0.75	0.71	200.5	78	2.23	0.69	0.64
Female	126	11	1.24	0.59	0.55	136	14	1.59	0.71	0.65	192.2	22	1.34	0.64	0.59
All	136	100	1.34	0.65	0.61	150	100	2.02	0.74	0.70	198.7	100	2.06	0.69	0.63
Race	**Mean**	**Pop %**	**GE2**	**GE1**	**GE0**	**Mean**	**Pop %**	**GE2**	**GE1**	**GE0**	**Mean**	**Pop %**	**GE2**	**GE1**	**GE0**
White	n.a.	n.a.	n.a.	n.a.	n.a.	201	54	1.73	0.68	0.66	267.97	52	1.86	0.64	0.60
Black	n.a.	n.a.	n.a.	n.a.	n.a.	85	45	1.46	0.60	0.56	120.37	47.5	1.14	0.53	0.49
Asian	n.a.	n.a.	n.a.	n.a.	n.a.	385	1	0.71	0.44	0.47	473.28	0.5	0.70	0.47	0.54
All	n.a.	n.a.	n.a.	n.a.	n.a.	150	100	2.02	0.74	0.70	198.7	100	2.06	0.69	0.63

(a) Source: Litchfield's calculations from PNAD 1981-1995;
(b) Source: Author's calculations from PNAD 2002.

Table A1.3: Summary Statistics of Households Income per Capita, by age and by education, 2002

	1981[a]					1990[a]					2002[b]				
Age	**Mean**	**Pop %**	**GE2**	**GE1**	**GE0**	**Mean**	**Pop %**	**GE2**	**GE1**	**GE0**	**Mean**	**Pop %**	**GE2**	**GE1**	**GE0**
$<$ 25 yrs	118	4	0.81	0.45	0.43	115	4	1.36	0.61	0.56	123.96	4	1.13	0.52	0.48
25-34 yrs	141	22	1.17	0.63	0.62	143	22	1.54	0.69	0.68	155.21	19.5	1.40	0.65	0.60
35-44 yrs	121	28	1.38	0.67	0.64	149	29	1.67	0.74	0.73	183.59	28	1.49	0.67	0.63
45-54 yrs	139	24	1.32	0.63	0.60	154	22	1.67	0.72	0.70	224.33	22.5	1.30	0.64	0.62
55-64 yrs	154	13	1.38	0.65	0.61	166	14	1.71	0.74	0.70	240.77	14	1.71	0.71	0.65
65 yrs +	144	8	1.65	0.70	0.61	150	10	5.41	0.94	0.73	231.20	12	4.77	0.70	0.54
All	136	100	1.34	0.65	0.61	150	100	2.02	0.74	0.70	198.7	100	2.06	0.69	0.63
Education	**Mean**	**Pop %**	**GE2**	**GE1**	**GE0**	**Mean**	**Pop %**	**GE2**	**GE1**	**GE0**	**Mean**	**Pop %**	**GE2**	**GE1**	**GE0**
Illiterate	57	30	0.71	0.39	0.38	52	25	1.33	0.45	0.42	116.97	16	1.69	0.65	0.56
Elementary	104	46	0.71	0.41	0.40	104	40	1.08	0.50	0.47	138.78	27.7	0.67	0.40	0.41
Intermediate	176	14	0.80	0.43	0.40	153	18	2.26	0.52	0.45	172.19	49	1.03	0.52	0.50
High School	311	7	0.53	0.35	0.36	272	10	0.79	0.44	0.43	744.3	7	1.09	0.39	0.37
College +	592	5	0.39	0.28	0.29	608	7	0.62	0.36	0.35	1.37e+03	0.3	0.32	0.26	0.27
All	136	100	1.34	0.65	0.61	150	100	2.02	0.74	0.70	198.7	100	2.06	0.69	0.63

(a) Source: Litchfield's calculations from PNAD 1981-1995;
(b) Source: Author's calculations from PNAD 2002.

Table A1.4: OLS Regression estimates for Log per Capita Monthly Income, 2002[(a)]

Regressors	Coefficients[(b)]	S.e.[(c)]
Regions:		
North	*base*	
North-East	-0.1115282**	(0.00898)
South-East	0.2836071**	(0.008927)
South	0.2924355**	(0.00993)
Central-West	0.29101**	(0.010588)
Gender	0.0381541**	(0.0092)
Age	0.0163881**	(0.00102)
Squared Age	0.0000453**	(1.07E-05)
Race:		
White	-0.0523253**	(0.038053)
Black	-0.2278619**	(0.038124)
Asian	*Base*	
Edu years	0.0501255**	(0.001942)
Squared Edu years	0.0042389**	(0.000135)
Formal	0.2183031**	(0.006098)
Economic sectors:		
Agriculture	-0.1768621**	(0.011497)
Industry	0.0295589**	(0.009477)
Construction	-0.0204712*	(0.010758)
Trade	*Base*	
Tourism	-0.0705659**	(0.015016)
Transports	0.1511405**	(0.012177)
Public Administration	0.1339801**	(0.011961)
Health, Education, Social Services	-0.1134392**	(0.009979)
Others	-0.0755563**	(0.00934)
Occupation type:		
Professionals and Technicians	0.5382941**	(0.010863)
Intermediates	0.0377208**	(0.007078)
Blue Collars	*Base*	
HH type:		
couple without kids	-0.0853562**	(0.00956)
couple with kids	-0.5452731**	(0.008213)
Single parent with kids	-0.4898715**	(0.009917)
others	*Base*	
Urban	0.2186183**	(0.008661)
Constant	3.463263**	(0.046931)
F-stat for Joint test of Significance	4253.7	
R^2	0.5415	
S.e. of Estimate	0.73751	
Number of obs.	99,945	

(a) Source: Author's calculations from PNAD 2002. The sample uses only household head aged between 15 and 80.
(b) Only age and edu year are continuous variables and their squared values are considered as well. The term *base* denotes category omitted in estimation.
(c) The estimation procedure is OLS and White (1980) estimated standard errors are reported in parentheses.

Table A1.5: Raw income ratios for Selected Groups

Category[b]	Base	2002[a]
Region North-East	Region North	0.776
Region South-East	Region North	1.619
Region South	Region North	1.693
Region Central-West	Region North	1.477
Male	Female	0.958
Whites	Asians	0.639
Blacks	Asians	0.333
Formal	Informal	1.402
Agriculture	Trade	0.441
Industry	Trade	1.003
Professionals/Technicians	Blue Collars	4.491
Intermediates	Blue Collars	1.291
Urban	Rural	2.254

(a) Source: Author's calculations from PNAD 2002. The sample uses only household head aged between 15 and 80.
(b) The income ratios are based on the ratio of per capita monthly incomes in the selected category relative the base category.

Table A1.6: Ceteris Paribus Relative income ratios for Selected Groups

Category[b]	Base	2002[a]
Region North-East	Region North	0.894
Region South-East	Region North	1.328
Region South	Region North	1.339
Region Central-West	Region North	1.337
Male	Female	1.038
Whites	Asians	0.949
Blacks	Asians	0.796
Formal	Informal	1.243
Agriculture	Trade	0.838
Industry	Trade	1.029
Professionals/Technicians	Blue Collars	1.713
Intermediates	Blue Collars	1.038
Urban	Rural	1.244

(a) Source: Author's calculations from PNAD 2002. The sample uses only household head aged between 15 and 80.
(b) The relative income effects are based on the anti-log of the estimated coefficients for the relevant regressors from the regression showed in the Table A1.4. Hence these effects are adjusted for other characteristics included in the income generating equation.

Decomposition by race

Table A1.7: OLS Regression with Decomposition by Black and White

Variable	Black	White	$\Delta\beta$
North-East	-0.10199**	-0.1301**	0.028107
	(0.010549)	(0.017193)	
South-East	0.274777**	0.285262**	-0.01049
	(0.011102)	(0.015987)	
South	0.272275**	0.293354**	-0.02108
	(0.015768)	(0.01631)	
Central-West	0.292676**	0.291123**	0.001553
	(0.013048)	(0.018686)	
Male	0.052284**	0.020214	0.03207
	(0.013119)	(0.01298)	
Age	0.012291**	0.020715**	-0.00842
	(0.001422)	(0.001467)	
Age^2	8.03E-05**	7.70E-06**	7.26E-05
	(0.000015)	(1.53E-05)	
Edu yrs	0.042265**	0.060309**	-0.01804
	(0.002822)	(0.002845)	
Edu yrs^2	0.004544**	0.003741**	0.000803
	(0.000216)	(0.000186)	
Formal	0.21095**	0.230689**	-0.01974
	(0.008291)	(0.009051)	
Agriculture	-0.22067**	-0.13772**	-0.08296
	(0.015816)	(0.016979)	
Industry	0.02973**	0.025194**	0.004536
	(0.014217)	(0.012759)	
Construction	0.001317	-0.06054**	0.061855
	(0.015086)	(0.015503)	
Tourism	-0.04087*	-0.09278**	0.051911
	(0.021944)	(0.020599)	
Transports	0.147376**	0.152417**	-0.00504
	(0.01805)	(0.01655)	
Public Administration	0.182354**	0.09437**	0.087984
	(0.01741)	(0.016537)	
Health, Education, Social Services	-0.07189**	-0.14811**	0.076218
	(0.014417)	(0.01393)	
Other sectors	-0.11288**	-0.04186**	-0.07102
	(0.013604)	(0.012928)	
Professionals/Technicians	0.503286**	0.556233**	-0.05295
	(0.018039)	(0.01388)	
Intermediates	0.017394**	0.048204**	-0.03081
	(0.010249)	(0.009858)	
Couples without children	-0.09098**	-0.07382**	-0.01716
	(0.013816)	(0.013336)	
Couples with children	-0.57753**	-0.51195**	-0.06558
	(0.01137)	(0.011871)	
Single parent with children	-0.52454**	-0.45682**	-0.06773
	(0.014244)	(0.013878)	
Urban	0.213527**	0.226496**	-0.01297
	(0.011502)	(0.013123)	
Constant	3.404509**	3.228782**	0.175727
	(0.037633)	(0.040303)	
F-stat for Joint Test of Significance	1559.41	2284.48	
R^2	0.4512	0.5325	
S.e. of estimate	0.73175	0.74104	
Number of obs.	48,534	50,984	

Table A1.8: Effects of the decomposition estimated from the income regressions

Variable	(1)	(2)
Income differential (Y_b-Y_w)	**-0.651724**	**100.00%**
Regions	-0.116050544	25.34917122
Male	-0.000193698	0.042309796
Age	-0.022054289	4.817366134
Age^2	-0.013565159	2.963067077
Edu yrs	-0.087028226	19.0097635
Edu yrs^2	-0.129067322	28.19245425
Formal	-0.010020479	2.188794865
Economic sector	-0.010816612	2.362695844
Occupation type	-0.039534015	8.635500434
Household type	-0.020704593	4.522548946
Urban	-0.008773103	1.916327935
Constant	0	0
Endowment effect (X_b-X_w)'b_b	**-0.45780804**	**100**
Regions	-0.004728133	2.437194146
Male	0.02368223	-12.20739553
Age	-0.390140887	201.1045493
Age^2	0.17121091	-88.253485
Edu yrs	-0.126413164	65.16174858
Edu yrs^2	0.056815894	-29.28668889
Formal	-0.012186023	6.281486587
Economic sector	-0.012592481	6.491001844
Occupation type	-0.015854713	8.17257301
Household type	-0.048001261	24.7430412
Urban	-0.011518407	5.937352631
Constant	0.175727	-90.5813779
Treatment effect X_w'(b_b-b_w)	**-0.193999036**	**100**

Table A1.9: Decomposing the mean and the variance of the Log of Income

	Mean	t-stat
Total Income Differential	-0.651724	
Differences in Characteristics	-0.45780804	-117.38
Differences in Structure	-0.193999036	-31.77
	Variance	F-stat
Total Differential	0.198	1.2
Differences in Characteristics	0.161	
Differences in Structure	0.0189	

Decomposition by region

North-East versus North

Table A1.10: OLS Regression with Decomposition by North-East vs North

Variable	North-East	North	Δβ
Male	0.0560844**	0.0529515**	0.003133
	(0.017563)	(0.026661)	
Age	0.0054198**	0.0108164**	-0.0054
	(0.001422)	(0.003324)	
Age^2	1.68E-04**	0.0000822**	8.58E-05
	(1.94E-05)	(3.59E-05)	
White	-0.0142798	-0.1940016	0.179722
	(0.078829)	(0.131021)	
Black	-0.140475*	-0.3711457**	0.230671
	(0.078591)	(0.130588)	
Edu yrs	0.0354847**	0.0338889**	0.001596
	(0.003455)	(0.006522)	
Edu yrs^2	0.0060365**	0.0049834**	0.001053
	(0.000262)	(0.000478)	
Formal	0.1970317**	0.2491202**	-0.05209
	(0.010916)	(0.019391)	
Agriculture	-0.2315047**	-0.0954758**	-0.13603
	(0.020598)	(0.041237)	
Industry	0.0093817	-0.0116425	0.021024
	(0.020371)	(0.031839)	
Construction	0.0382776*	-0.0606348*	0.098912
	(0.021842)	(0.034469)	
Tourism	-0.0642508*	-0.138895**	0.074644
	(0.028559)	(0.046145)	
Transports	0.1764088**	0.0888174**	0.087591
	(0.024754)	(0.041467)	
Public Administration	0.1287046**	0.20831**	-0.07961
	(0.022484)	(0.034116)	
Health, Education, Social Services	-0.1053078**	-0.1212735**	0.015966
	(0.019126)	(0.029539)	
Other sectors	-0.1049381**	-0.112829**	0.007891
	(0.018564)	(0.031529)	
Professionals/Technicians	0.5477983**	0.5985229**	-0.05072
	(0.022995)	(0.038708)	
Intermediates	0.0395058**	0.0093148	0.030191
	(0.014521)	(0.02312)	
Couples without children	-0.0722223**	-0.0163747**	-0.05585
	(0.019034)	(0.032567)	
Couples with children	-0.6004047**	-0.5068357**	-0.09357
	(0.015919)	(0.026282)	
Single parent with children	-0.5143901**	-0.4932187**	-0.02117
	(0.018543)	(0.031932)	
Urban	0.1998368**	0.3423687**	-0.14253
	(0.013234)	(0.042492)	
Constant	3.549894**	3.664657**	-0.11476
	(0.092511)	(0.158053)	
F-stat for Joint Test of Significance	1304.71	322.59	
R^2	0.5162	0.4328	
S.e. of estimate	0.75461	0.76441	
Number of obs.	29,939	9,922	

Table A1.11: Effects of the decomposition estimated from the income regressions

Variable	(1)	(2)
Income differential (Y_{ne}-Y_n)	**-0.252553**	**100.00%**
Regions	0.001441307	-0.988919183
Male	0.012623148	-8.66107612
Age	0.038654448	-26.52184085
Age^2	0.001798412	-1.233938119
Edu yrs	-0.044618959	30.61424956
Edu yrs^2	-0.071303259	48.92305485
Formal	-0.005395023	3.701668549
Economic sector	-0.034642063	23.76883694
Occupation type	-0.012588644	8.637401918
Household type	0.003780104	-2.593629342
Urban	-0.035495192	24.35419179
Constant	0	0
Endowment effect (X_{ne}-X_n)'b_{ne}	**-0.14574572**	**100**
Regions	0.002202061	-2.060296494
Male	-0.233265607	218.2484509
Age	0.177376485	-165.957355
Age^2	0.215089203	-201.2422061
Edu yrs	0.009738208	-9.111282684
Edu yrs^2	0.059642698	-55.80302453
Formal	-0.030643076	28.67033832
Economic sector	0.00278039	-2.601394115
Occupation type	0.00624576	-5.843670682
Household type	-0.063866013	59.75445173
Urban	-0.13741787	128.5711924
Constant	-0.114763	107.3747963
Treatment effect X_n'(b_{ne}-b_n)	**-0.106880762**	**100**

Table A1.12: Decomposing the mean and the variance of the Log of Income

	Mean	t-stat
Total Income Differential	-0.252553	
Differences in Characteristics	-0.14574572	-67.45
Differences in Structure	-0.106880762	-11.73
	Variance	**t-stat**
Total Differential	0.147	1.14
Differences in Characteristics	0.09762	
Differences in Structure	-0.0039	

North-East versus South-East

Table A1.13: OLS Regression with Decomposition by North-East vs South-East

Variable	North-East	South-East	Δβ
Male	0.0560844**	0.0436739**	0.012411
	(0.017563)	(0.01651)	
Age	0.0054198**	0.0235852**	-0.01817
	(0.001874)	(0.001776)	
Age^2	1.68E-04**	-0.0000301**	0.000198
	(1.94E-05)	(1.85E-05)	
White	-0.0142798	-0.1607519**	0.146472
	(0.078829)	(0.057235)	
Black	-0.140475*	-0.3641336**	0.223659
	(0.078591)	(0.057521)	
Edu yrs	0.0354847**	0.0540926**	-0.01861
	(0.003455)	(0.003529)	
Edu yrs^2	0.0060365**	0.0036179**	0.002419
	(0.000262)	(0.000235)	
Formal	0.1970317**	0.2239625**	-0.02693
	(0.010916)	(0.011098)	
Agriculture	-0.2315047**	-0.1820422**	-0.04946
	(0.020598)	(0.021734)	
Industry	0.0093817	0.0577225**	-0.04834
	(0.020371)	(0.015682)	
Construction	0.0382776*	0.0024328	0.035845
	(0.021842)	(0.018226)	
Tourism	-0.0642508*	0.0017556	-0.06601
	(0.028559)	(0.02537)	
Transports	0.1764088**	0.1420719**	0.034337
	(0.024754)	(0.020035)	
Public Administration	0.1287046**	0.0794256**	0.049279
	(0.022484)	(0.021464)	
Health, Education, Social Services	-0.1053078**	-0.0676302**	-0.03768
	(0.019126)	(0.017421)	
Other sectors	-0.1049381**	-0.0179879	-0.08695
	(0.018564)	(0.01577)	
Professionals/Technicians	0.5477983**	0.5280803**	0.019718
	(0.022995)	(0.017801)	
Intermediates	0.0395058**	0.0301251**	0.009381
	(0.014521)	(0.011816)	
Couples without children	-0.0722223**	-0.0790166**	0.006794
	(0.019034)	(0.016304)	
Couples with children	-0.6004047**	-0.5136287**	-0.08678
	(0.015919)	(0.014129)	
Single parent with children	-0.5143901**	-0.4573801**	-0.05701
	(0.018543)	(0.017219)	
Urban	0.1998368**	0.2528541**	-0.05302
	(0.013234)	(0.017601)	
Constant	3.549894**	3.635103**	-0.08521
	(0.092511)	(0.074556)	
F-stat for Joint Test of Significance	1304.71	1375.07	
R^2	0.5162	0.505	
S.e. of estimate	0.75461	0.71363	
Number of obs.	29,939	31,707	

Table A1.14: Effects of the decomposition estimated from the income regressions

Variable	(1)	(2)
Income differential (Y_{ne}-Y_{se})	**-0.734478**	**100.00%**
Regions	-0.00078159	0.23263823
Male	-0.00498719	1.48442043
Age	-0.01190381	3.54312754
Age^2	-0.03989762	11.8753907
Edu yrs	-0.06092851	18.1351608
Edu yrs^2	-0.11042575	32.8678464
Formal	-0.00683131	2.03331468
Economic sector	-0.02793157	8.31373516
Occupation type	-0.02151963	6.40524285
Household type	-0.02409266	7.17109767
Urban	-0.0266693	7.9380255
Constant	0	0
Endowment effect (X_{ne}-X_{se})'b_{ne}	**-0.33596894**	**100**
Regions	0.009215012	-2.31243951
Male	-0.84421535	211.84963
Age	0.469153693	-117.730667
Age^2	0.174618356	-43.8191915
Edu yrs	-0.12210543	30.6414594
Edu yrs^2	0.152653204	-38.3071984
Formal	-0.0160394	4.02496944
Economic sector	-0.03804519	9.54716164
Occupation type	0.004890436	-1.22721896
Household type	-0.05464026	13.7115703
Urban	-0.04877351	12.2393532
Constant	-0.085209	21.3825716
Treatment effect X_{se}'(b_{ne}-b_{se})	**-0.39849744**	**100**

Table A1.15: Decomposing the mean and the variance of the Log of Income

	Mean	t-stat
Total Income Differential	-0.734478	
Differences in Characteristics	-0.33596894	-90.78
Differences in Structure	-0.39849744	-56.13
	Variance	**t-stat**
Total Differential	0.147	1.14
Differences in Characteristics	-0.0189	
Differences in Structure	0.0729	

North-East versus South

Table A1.16: OLS Regression with Decomposition by North-East vs South

Variable	North-East	South	Δβ
Male	0.0560844**	-0.0251699	0.0812543
	(0.017563)	(0.020783)	
Age	0.0054198**	0.0249127**	-0.0194929
	(0.001874)	(0.002454)	
Age^2	1.68E-04**	-0.0000421	0.0002101
	(1.94E-05)	(2.59E-05)	
White	-0.0142798	-0.0282841	0.0140043
	(0.078829)	(0.092673)	
Black	-0.140475*	-0.2397829**	0.0993079
	(0.078591)	(0.09356)	
Edu yrs	0.0354847**	0.0583436**	-0.0228589
	(0.003455)	(0.004962)	
Edu yrs^2	0.0060365**	0.0032825**	0.002754
	(0.000262)	(0.000326)	
Formal	0.1970317**	0.220588**	-0.0235563
	(0.010916)	(0.015173)	
Agriculture	-0.2315047**	-0.1995409**	-0.0319638
	(0.020598)	(0.026449)	
Industry	0.0093817	0.036086	-0.0267043
	(0.020371)	(0.020095)	
Construction	0.0382776*	-0.10494**	0.1432176
	(0.021842)	(0.024345)	
Tourism	-0.0642508*	-0.1223749**	0.0581241
	(0.028559)	(0.036915)	
Transports	0.1764088**	0.1534512**	0.0229576
	(0.024754)	(0.027333)	
Public Administration	0.1287046**	-0.0149309	0.1436355
	(0.022484)	(0.028945)	
Health, Education, Social Services	-0.1053078**	-0.1678895**	0.0625817
	(0.019126)	(0.023888)	
Other sectors	-0.1049381**	-0.1098678**	0.0049297
	(0.018564)	(0.021464)	
Professionals/Technicians	0.5477983**	0.5060351**	0.0417632
	(0.022995)	(0.023947)	
Intermediates	0.0395058**	0.0710603**	-0.0315545
	(0.014521)	(0.016138)	
Couples without children	-0.0722223**	-0.0707244**	-0.0014979
	(0.019034)	(0.02144)	
Couples with children	-0.6004047**	-0.4861478**	-0.1142569
	(0.015919)	(0.019256)	
Single parent with children	-0.5143901**	-0.4845624**	-0.0298277
	(0.018543)	(0.023278)	
Urban	0.1998368**	0.1779843**	0.0218525
	(0.013234)	(0.019777)	
Constant	3.549894**	3.626171**	-0.076277
	(0.092511)	(0.112496)	
F-stat for Joint Test of Significance	1304.71	632.06	
R^2	0.5162	0.4797	
S.e. of estimate	0.75461	0.70638	
Number of obs.	29,939	17,141	

Table A1.17: Effects of the decomposition estimated from the income regressions

Variable	(1)	(2)
Income differential (Y_{ne}-Y_s)	**-0.779588956**	**100.00%**
Regions	-0.001028655	0.286480018
Male	-0.001366332	0.380522741
Age	-0.001090992	0.30384079
Age^2	-0.066772761	18.59618434
Edu yrs	-0.067810552	18.88520877
Edu yrs^2	-0.118650486	33.04410796
Formal	-0.014983295	4.17284113
Economic sector	-0.020406756	5.68327246
Occupation type	-0.022227727	6.190412138
Household type	-0.030743803	8.562135534
Urban	-0.013985638	3.894994115
Constant	0	0
Endowment effect (X_{ne}-X_s)'b_{ne}	**-0.359066997**	**100**
Regions	0.060690673	-14.43222445
Male	-0.892886514	212.3281542
Age	0.484050441	-115.1070544
Age^2	0.027817741	-6.615050846
Edu yrs	-0.154433928	36.72434338
Edu yrs^2	0.177574753	-42.22722475
Formal	-0.015004239	3.568003794
Economic sector	0.019365593	-4.605132332
Occupation type	-0.003602289	0.856623377
Household type	-0.066533516	15.82165079
Urban	0.018716328	-4.450737261
Constant	-0.076277	18.13864848
Treatment effect X_s'(b_{ne}-b_s)	**-0.420521959**	**100**

Table A1.18: Effects of the decomposition estimated from the income regressions

	Mean	t-stat
Total Income Differential	-0.779588956	
Differences in Characteristics	-0.359066997	-43.67
Differences in Structure	-0.420521959	-46.72
	Variance	**t-stat**
Total Differential	0.2184	1.23
Differences in Characteristics	0.0079787	
Differences in Structure	0.09021641	

North-East versus Central-West

Table A1.19: OLS Regression with Decomposition by North-East vs Central-West

Variable	North-East	Central-West	Δβ
Male	0.0560844**	0.0376659	0.018419
	(0.017563)	(0.028625)	
Age	0.0054198**	0.0209011**	-0.01548
	(0.001874)	(0.003112)	
Age^2	1.68E-04**	-0.00003	0.000198
	(1.94E-05)	(3.39E-05)	
White	-0.0142798	0.1769437*	-0.19122
	(0.078829)	(0.101618)	
Black	-0.140475*	0.0032787	-0.14375
	(0.078591)	(0.101745)	
Edu yrs	0.0354847**	0.037193**	-0.00171
	(0.003455)	(0.005999)	
Edu yrs^2	0.0060365**	0.0052475**	0.000789
	(0.000262)	(0.00041)	
Formal	0.1970317**	0.2428364**	-0.0458
	(0.010916)	(0.017659)	
Agriculture	-0.2315047**	-0.0119832	-0.21952
	(0.020598)	(0.033745)	
Industry	0.0093817	-0.0336098	0.042992
	(0.020371)	(0.029142)	
Construction	0.0382776*	-0.0723544**	0.110632
	(0.021842)	(0.030654)	
Tourism	-0.0642508*	-0.1460396**	0.081789
	(0.028559)	(0.043852)	
Transports	0.1764088**	0.1663534**	0.010055
	(0.024754)	(0.036557)	
Public Administration	0.1287046**	0.250929**	-0.12222
	(0.022484)	(0.033243)	
Health, Education, Social Services	-0.1053078**	-0.1769953**	0.071688
	(0.019126)	(0.029492)	
Other sectors	-0.1049381**	-0.0769874**	-0.02795
	(0.018564)	(0.027129)	
Professionals/Technicians	0.5477983**	0.5392471**	0.008551
	(0.022995)	(0.029865)	
Intermediates	0.0395058**	0.042598**	-0.00309
	(0.014521)	(0.020874)	
Couples without children	-0.0722223**	-0.1807565**	0.108534
	(0.019034)	(0.027311)	
Couples with children	-0.6004047**	-0.5975079**	-0.0029
	(0.015919)	(0.023495)	
Single parent with children	-0.5143901**	-0.5100085**	-0.00438
	(0.018543)	(0.031108)	
Urban	0.1998368**	0.2116313**	-0.01179
	(0.013234)	(0.027446)	
Constant	3.549894**	3.527203**	0.022691
	(0.092511)	(0.129868)	
F-stat for Joint Test of Significance	1304.71	499.86	
R^2	0.5162	0.5021	
S.e. of estimate	0.75461	0.75512	
Number of obs.	29,939	11,236	

Table A1.20: Effects of the decomposition estimated from the income regressions

Variable	(1)	(2)
Income differential (Y_{ne}-Y_{cw})	**-0.642181333**	**100.00%**
Regions	-0.00143585	0.589700527
Male	0.010051724	-4.128220358
Age	0.03212748	-13.19468395
Age^2	-0.018547976	7.617612124
Edu yrs	-0.05492847	22.55899954
Edu yrs^2	-0.104868312	43.06918023
Formal	-0.011620004	4.772309412
Economic sector	-0.022367972	9.186475769
Occupation type	-0.029199798	11.99229133
Household type	-0.025004273	10.269199
Urban	-0.017694609	7.267136379
Constant	0	0
Endowment effect (X_{ne}-X_{cw})'b_{ne}	**-0.24348806**	**100**
Regions	0.01389092	-3.675007139
Male	-0.676517329	178.9806641
Age	0.417022848	-110.3283288
Age^2	-0.16426587	43.45847964
Edu yrs	-0.010921047	2.889292323
Edu yrs^2	0.049072415	-12.98268808
Formal	-0.028393532	7.51184484
Economic sector	-0.010921047	2.889292323
Occupation type	0.049072415	-12.98268808
Household type	-0.028393532	7.51184484
Urban	-0.010320712	2.730466572
Constant	0.022691	-6.003172538
Treatment effect X_{cw}'(b_{ne}-b_{cw})	**-0.377983472**	**100**

Table A1.21: Effects of the decomposition estimated from the income regressions

	Mean	**t-stat**
Total Income Differential	-0.642181333	
Differences in Characteristics	-0.24348806	-61
Differences in Structure	-0.377983472	-14.44
	Variance	**t-stat**
Total Differential	0.033	1.028
Differences in Characteristics	-0.0403	
Differences in Structure	0.01549	

Appendix 1.B

The PNAD 2002

The data for this study is drawn from an annual national households survey, *Pesquisa Nacional por Amostra do Domicilios* (PNAD), for the years 1981, 1990 and 2002. The data was collected by the *Instituto Brasileiro de Geografia e Estatistica* (IBGE). The PNAD is based on a nationally representative random sample of households and uses a three level multi-stage sampling procedure.

The household survey consists of two sections, *Archivo de Domicílios* and *Archivo de Pessoas*. The first section contains information at the household level, such as characteristics of the dwellings and the geographical locations of the households. The second section provides data at the individual level, focusing more on the characteristics of household members.

The survey covers every state in the Brazilian Federation and the sample size varies in each year, ranging from 289,783 to 514,569 individual observations during the past twenty years of surveys (Litchfield, 2001, p.42).

The table below displays the numbers of observations for the three years considered in this study.

PNAD Sample sizes

	1981[a]	1990[a]	2002[b]
Individuals	482,568	309,146	409,152
Households	103,955	73,165	102,500
Weighted individuals (in millions)	117.83	144.01	166.27

(a) Source: Litchfield's calculations from PNAD 1981-1995;
(b) Source: Author's calculations from PNAD 2002.

Sampling methodology

The sampling methodology is based on a three level procedure, which includes municipalities, census sectors and, finally, households (IBGE, 2002b, p.21)

At the first level, municipalities are identified for each state of the Federation. Some municipalities are automatically included in the sample. These include capital municipalities, metropolitan municipalities and municipalities with high population density or particular social and economical characteristics.

While all of the urban municipalities are taken into account, certain rural municipalities are not included in the survey sample. Specifically excluded are the Northern states of Rondônia, Acre, Amazonas, Roraima, Parà and Amapà, all of which are located in the Amazon area of Brazil and comprise just 2.1% of the Brazilian population (IBGE, 2002a): this is because population densities are very low here making the survey costs very high.

Within each municipality, census sectors are selected by considering the population proportion in the Brazilian demographic census. For this reason, population weights are used in order to ensure the representativeness of the sample.

Finally, within each census sector, households are randomly selected from the resident population[21].

A final important feature of the sampling methodology is the reference period[22], i.e. when the household survey is conducted. The reference period is normally one week in the last quarter of the year: from 8th to 14th November for 1981, from 23rd to 29th September for 1990 and from 22nd to 28th September for 2002 (IBGE, 2002a, p.1).

The questionnaire

The questionnaire consists of two sections, one at household level and one at the individual level. The household level questionnaire investigates characteristics of the household dwellings such as the property type, the estimated value of the dwelling, the number of rooms, the physical assets, the existence of water and electricity connections and the geographic location.

The individual level questionnaire involves information about each member of the household, such as age, gender, race, level of educational attainment and current employment or activity. All of the members aged 10 or above answer specific questions about employment, such as the nature or their income, i.e. salaries and business revenues, private remittances, pensions, private insurance, savings or investments.

During the years considered in this study, only small changes were made to the questionnaire. The most important and significant changes were the introduction of questions related to race and ethnicity since 1988 and the expansion of the questions related to the durable physical assets of the dwelling, and the addition of questions

---

[21] The only part of the population that is deliberately excluded is the armed forces, prisoners, residents of religious institutions, residents of diplomatic institutions and interns in schools, orphanages, hospitals and asylums (Litchfield, 2001).

[22] The reference period is very important, particularly for the conversion into real term of nominal values.

related to secondary and additional activities, in addition to the primary activity (Litchfield, 2001, p.43).

Finally, the PNAD for 2002 underwent a substantial change in its classification of the economic and occupational sectors in order to harmonize with the international classifications (IBGE, 2002b, p.2). The changes were from the Brazilian classification of occupations, the *CBO-Domiciliar*, and to the national classification of economic activities, the *CNAE-Domiciliar*.

Income definition

This study chooses, mostly for pragmatic reasons, the nominal household income. In practice, the choice between income or consumption is driven by data availability: measuring income seems to be more difficult, particularly related to self-owned activities and non-employment incomes. These factors imply that household surveys underreport income, as already explained. On the contrary, at least for Latin American countries and for Brazil as well, the reason for using household income is simply that income data are more available at great frequency (Mejia and Vos, 1997, p.20).

In the PNAD survey for 2002[23] (IBGE, 2002) the nominal household income is labelled *v4614* and considers:

- Income from employment or self-employment, i.e. first, second, third and fourth jobs with payment in cash or in-kind;
- Social insurance receipts for old-age, disability or survivors pensions, sickness and maternity benefits, work injury and unemployment benefits and family allowances paid in cash through the National Institute of Social Security;
- Other incomes, such as rental incomes, dividends or interest payments on savings and investments.

Necessary adjustments have been conducted in order to obtain the variable required from the empirical analysis.

First of all, household income has been transformed into per capita income by dividing by household size. In order to facilitate comparisons with estimates of poverty and inequality in earlier years which use per capita income definitions, e.g. Litchfield (2001), an equivalence scale was not used.

Secondly, the conversion to real values has been conducted by diving the nominal value by the deflator value. The deflator adopted in this work for 2002 is equal to 166.1 and has been constructed by ECLAC (ECLAC, 2004) with 1995 as base year. 1995 was

[23] For 1981 and 1990, the variable identifying the nominal household income is labelled v410.

chosen as the reference year as Litchfield's estimates are all reported in 1995 prices as the set of poverty lines to be used in the analysis (Rocha, 1993).

Finally, contamination due to the presence of zero incomes, missing incomes and misreported incomes needs to be dealt with.[24] The misreporting of income is not only a problem of rural households or of the informal sector. As Litchfield reports (Litchfield, 2001, p.55), a massive amount of misreporting is imputable to the richest percentiles of the Brazilian population.

There are three different possible solutions to avoid this data contamination. The first two solutions involve simply substituting these "dirty" observations with either the mean of income or with the predicted value of income. The third solution could be to drop all the zero, missing, and misreported value observations.

Zero and Missing Incomes

	1981[(a)]	1990[(a)]	2002[(b)]
Zero Incomes	1.07%	0.99%	1.33%
Missing Incomes	0.80%	1.25%	1.95%

(a) Source: Litchfield's calculations from PNAD 1981-1995;
(b) Source: Author's calculations from PNAD 2002.

The above table shows the share of zero and missing values in the three samples. These shares seem to be negligible. However, the dataset for 2002 identifies as missing income all the observations coded 999,999,999,999 that may include both missing and misreported values. Clearly these values need to be dropped from the sample or substituted with more reliable values.

The table below provides a sensitivity analysis that compares summary measures of income distribution resulting from applying the three suggested solutions in dealing with dirty data.

In the first column, the whole distribution drops already missing values, but retains zero income values. The second column shows the measures of income distribution when zero income values are dropped from the sample, while the third column provide the results if all of zero income values are replaced by 25% of the mean.

[24] Generally speaking, misreported incomes were coded as 999,999,999,999. This obviously affects estimations of any poverty, inequality and welfare measures. Moreover, some of the zero incomes, missing income and misreported incomes are due to the problem of top-coding of upper income as well as bottom-coding.

Sensitivity analysis for Zero Income of 2002[a]

	Whole distribution[b] 103,915 obs.	Zero Incomes dropped 102,500 obs.	Zero Income Imputed[c] 103,915 obs.
Mean[d]	196.5	198.7	197.07
Median	101.2	103.27	101.2
Gini	0.581	0.581	0.581
s.e.	(0.0019)	(0.0019)	(0.0019)
GE(0)	0.613	0.613	0.630
s.e.	(0.0046)	(0.0046)	(0.0046)
GE(1)	0.688	0.688	0.690
s.e.	(0.0125)	(0.0125)	(0.0125)
GE(2)	2.058	2.058	2.072
s.e.	(0.5353)	(0.5353)	(0.5353)

(a) Source: Author's calculations from PNAD 2002;
(b) Whole distribution drops 2069 obs. as missing values;
(c) 25% of mean income of whole distribution, equal to 49.12, has been imputed to zero income obs;
(d) All incomes are shown in 1995 reais. Standard errors are estimated by boot-strapping procedure.

Since from this table a small difference between outcomes with and without this bunch of zero and missing values has been found, it has been decided to drop these observations for this study.

Chapter 2

Brazilian Poverty between and within Groups: Decomposition by Geographical, Group-specific Poverty Lines

Abstract: This study investigates Brazilian poverty by exploiting geographical differences and questions whether the standard approach in measuring poverty is informative enough when the population is heterogeneous. To do so, we apply the reformulation of the FGT class of poverty measures proposed by Chiappero and Civardi (2006). This decomposition aims to compute poverty within groups, using group-specific poverty lines, and poverty between groups by adopting a community-wide poverty line. We run two empirical exercises, for the entire country and for each Brazilian region. The North and the Central-West reveal a dominance of the within component. The North-East shows the highest level of poverty, even higher than the North and the Central-West, but the high within group component is counterbalanced by a higher between group component, attributable to the high level of inequality of the North-East. The South and the South-East have between group components that dominate over within group ones. Our findings suggest that the analysis of poverty between and within groups is more exhaustive than the standard methodology when differentiated poverty lines are exploited.

2.1 Introduction

The purpose of this study is twofold: first, to investigate Brazilian poverty by exploiting geographical differences. Second, it questions whether the standard approach to measuring poverty is informative enough considering that the population is clearly not homogenous.

Brazil is a country with huge regional disparities. In 2002, 56% of the real Brazilian GDP was generated by the most economically developed region of Brazil, the South-East, including metropolitan areas such as Rio de Janeiro and São Paulo. By contrast, the two most depressed regions of the country, the North and the North-East, together produced only 0.6% of national GDP.[1] Regional differences are sharp not only in terms of GDP values or income distribution data, but also in terms of social and demographic variables, such as ethnicity and family structures. Hence, the study of these geographically-specific discrepancies becomes crucial for the understanding of causes of poverty and targeting more focused policies.

The adoption of differentiated poverty lines provides a more complex picture of the poverty situation, and it has been applied in the literature on poverty measurement.[2] However, so far empirical studies adopting differentiated poverty lines have provided poverty evaluations simply as a result of a simple aggregation of poverty outcomes for each homogenous group, defined by the set of group-specific poverty lines adopted.

The implementation of this approach recognizes the importance of applying group-specific poverty thresholds. What is lacking in this kind of application is the detection of poverty resulting from comparison between groups, using a community-wide poverty threshold.[3]

[1] These values are taken from the IBGE publication, *Conta Regionais do Brasil*, 2002, IBGE(2005).

[2] Regarding Brazil, Ferreira, Leite and Litchfield (2006) and Ferreira and Litchfield (2001) analyze poverty adopting differentiated poverty lines (Litchfield 2001). Bottiroli-Civardi and Chiappero-Martinetti (1999) study the Italian poverty situation by applying a set of differentiated poverty lines.

[3] The importance of investigating on differentials not only within groups but also between groups has been widely explored by Stewart (2001) in her paper on horizontal inequality.

Chiappero and Civardi (2006) propose a reformulation of the three most famous poverty indexes, better known as the Foster, Greer and Thorbecke (FGT) class of measures,[4] that aims at decompose poverty within and between homogenous groups by implementing differentiated poverty lines. After comparing each individual position within its homogenous group using the group-specific poverty line, people belonging to different groups are compared to each other by adopting a community-wide poverty line in order to capture poverty between groups.

This alternative conceptual and analytical approach to poverty measurement has potentially remarkable implications especially where the differentiation among poverty lines is very significant.

To the best of our knowledge, this is the first work that applies Chiappero and Civardi's 2006 poverty indexes reformulation to Brazilian data. We aim to discover whether the computation of poverty between and within groups provides valuable information on Brazilian heterogeneity. The attraction of this reformulated measures is that it allows us to look at poverty situation for each group singularly, captured by the within-group component, but also to get a rough measure of the importance of poverty across groups, as the between-group component tell us how poor people are relative to other groups. The significance of poverty between groups is sometimes overlooked also when differentiated poverty lines are adopted. This has significant negative implications for our understanding and for making policy. As such this paper seeks to investigate the value of a more inclusive approach.

To run our empirical exercises we use the 2002 Brazilian households survey, *Pesquisa Nacional por Amostra do Domicilios* (PNAD). The dataset contains information on incomes and other socio-economic data available for Brazil and is collected annually by the *Instituto Brasileiro de Geografia e Estatística* (IBGE).

[4] In their work, Foster, Greer and Thorbecke (1984) aggregated in an unique formula the most common well-known poverty indexes, such as the Headcount Ratio, the Poverty Gap and the Squared Poverty Gap by weighing for a parameter α. Later on in this section, this procedure of aggregation is better described.

Since geographical location is one of the most relevant determinants of Brazilian heterogeneity, we exploit this criterion in our empirical analysis to establish homogenous groups and their related poverty lines. The construction of differentiated poverty lines based on this criterion divides the population into geographically-specific homogeneous groups. To do so, we apply Rocha's 2003 estimation of absolute poverty lines.

In this respect, two important remarks need to be highlighted. By adopting geographically-specific poverty lines we recognize the geographical feature, such as living in a specific region and in an urban or rural area, as the only source of heterogeneity of the Brazilian population. We understand that this approach might be too narrow and we recognize that the geographically-specific groups are far from being homogenous in terms of other criteria, such as household type, educational level or ethnicity. However, this work aims to investigate poverty within and between groups by focusing on geographical disparities as a typical feature of the Brazilian society. Moreover, Rocha's geographically-specific poverty lines are absolute poverty lines. Hence this study only looks at absolute poverty within and between groups and does not consider any notions of relative poverty, but it analyzes how the persistence of inequality might have an impact on the levels of absolute poverty, in particular on the between-groups component.

Starting from geographically-specific absolute poverty lines, we investigate Brazilian poverty using standard methodology. Then, by applying Rocha's differentiated poverty lines and the reformulation of FGT class of poverty indexes, we focus on the extent to which the between- or within-group component of poverty is able to explain the pattern of regional disparities in Brazil. Hence, we run two different empirical exercises, first at the national level and then at the regional level.

Our findings suggest that when differentiated poverty lines are exploited the analysis of poverty between and within groups is more exhaustive than the standard methodology. In the empirical exercise at the regional level, we find that in the North and the Central-West the within-group component is dominant because of the high level of absolute poverty within all homogenous groups. On the other hand, the South-East and the South show the dominance of the between-group component. Finally, the North-East follows a pattern similar to the North and the Central-West, though with a lower contribution of the within-group component: this might be due to the high level of inequality which causes the between-group component diminish the within-group effect.

These results throw new light on the intricate relation existing between poverty and inequality. By looking at absolute poverty levels within and between groups it becomes clear how inequality affect the level of poverty between groups.

The structure of this paper is as follows. Section 2.2 depicts Brazilian poverty analysis. Section 2.3 explains the conceptual and analytical framework that we have adopted. Section 2.4 proposes the empirical results by computing poverty between and within groups. Finally, Section 2.5 concludes.

2.2 The profile of Brazilian poverty

Brazil is a country characterized by dramatic differences among geographical regions and these gaps have persisted across more than fifty years of Brazilian history (Baer, 2001).

The dataset employed is constructed on the basis of the annual Brazilian household survey, *Pesquisa Nacional por Amostra do Domicilios* (PNAD) for 2002.[5]

From this survey we take nominal household monthly income as the measure of welfare, as it includes income from employment or self-employment, social insurance transfers for old-age, disability or survivor's pensions, sickness and

[5] The PNAD is based on a nationally representative random sample of households and adopts a three-stage sampling procedure, by selecting municipalities, census sectors and, finally, households. While some municipalities are automatically included, some rural municipalities in the Northern states of Rondônia, Acre, Amazonas, Roraima, Parà, Amapà, are excluded because of their very low population density and their location in remote areas of the Amazonas. Moreover, it is estimated that these excluded municipalities count just for the 2.1% of the entire Brazilian population. In order to guarantee the representativeness of the sample, population weights are estimated. Hence, the PNAD for 2002 counts 409,152 individuals aggregated into 102,500 households, but the weighted individuals are 166,270,000.

maternity benefits, work injury and unemployment benefits and family allowances. Finally, monthly income also considers other sources of income such as rental incomes, dividends or interest payments on savings and investments.

Since income data refer to households rather than to individuals, technical adjustments should be applied in order to evaluate intra-household welfare. The adjustment of household income by adopting equivalence scales[6] improves the reliability of the data because it takes into account the potential heterogeneity of individuals within households and the effect of economies of scale.

However, the majority of studies on Brazilian poverty have tended to avoid adjustment via equivalence scales and to prefer per capita values, although the simple per capita adjustment tends to overestimate poverty, as stressed by Glewwe and Van der Gaag (1990). For comparative reasons, in this study we adopt per capita income following the mainstream in the Brazilian literature (Rocha, 1997).

Before going deeper into Brazilian poverty issues, it is worth looking at general economic indicators for Brazil and its regions.[7] Table 2.1 provides some summary statistics for the entire nation and for each geographical region showing mean and median income values as well as the most common inequality indicator, the Gini coefficient.

The huge differences across Brazilian regions are strikingly portrayed in Figure 2.1. Looking at the level of income, the poorest region is the North-

[6] When expenditure data are used, equivalence scales are mostly estimated by the adoption of two different techniques: the Rothbarth method, based on expenditure data on goods consumed by children versus adults, and the Engel method, based on the relation of food expenditure versus total expenditure. For further discussion, see Deaton (1997, section 4.3). When income data are exploited, the most common and simplest technique is to compute per capita income. Besides that, the most common equivalence scales applied to income data requires weighting the household size, n, to a parameter θ that is defined among [0,1] (Buhnmann et al., 1988)

[7] In the PNAD survey, the choice of geographic locations is among 27 different municipalities. To analyze Brazilian situation by region, these municipalities have been aggregated in the five geographical regions: the North (Rondônia, Acre, Amazonas, Roraima, Parà, Amapà and Tocantis), the North East (Maranhão, Piauì, Cearà, Rio Grande do Norte, Paraìba, Pernambuco, Alagoas, Sergipe, Bahia), the South East (Minas Gerais,pìrito Santo, Rio de Janeiro, São Paulo), the South (Paranà, Santa Catarina, Rio Grande do Sul) and the Central West (Mato Grosso do Sul, Mato Grosso, Golàs, Distrito Federal).

East followed by the North, the South and the Central West.[8] The South-East is the richest geographical region of the country with a median per-capita income twice that of the North-East region.

This pattern of regional disparities is well-known in Brazilian history. During the last century, the South-East region has always dominated of the regional distribution of national income, while the North and the Central-West were typically the most deprived regions.[9] This allows us to recognize the important jump in terms of contribution to Brazilian GDP made by the North and the Central-West regions and, at the same time, to detect a worrying depression for the North-East.

The distribution of income among regions tracks a trend similar to the one obtained from the levels of income. In fact, the most unequal region is the North-East with a Gini coefficient even higher than the value for the whole country. The Central-West ranks second, followed by the North, then the South-East and, finally, the South.[10] In order to deepen the investigation of Brazilian distribution of income, Table 2.2 shows mean incomes per decile by region.

One additional important issue should be stressed before moving to poverty indexes analysis. As reported in many publications,[11] the data coming from national household surveys are often very different to data elaborated by the National Accounts system.

---

[8] The ranking between the South and the Central-West varies with the definition of income we look. Using per capita income the South is richer than the Central West, but if we use other two equivalent income values, we find the reversal.

[9] A detailed description of changing in regional differences during the past century is well reported in Baer' book (Baer, 2001, chapter 14).

[10] In particular, if we use per capita income, the ranking is clear: from the most unequal we have the North-East, the Central-West, the North, the South-East and, finally, the South. When we use both equivalent incomes, the ranking is, always starting from the most unequal: the North-East, then the Central-West and the North come together and, finally, together again, the South-East and the South.

[11] For further discussions on discrepancies between National Account data and Household Survey data, see Deaton (1997, section 1.2). Litchfield discussed this issue specifically for Brazil stressing the problem in comparing incomes coming from these two types of dataset (Litchfield, 2001, page 51).

Table 2.3 reports total GDP and monthly per capita GDP in 2002 Reais[12] provided by National Account data. National accounts reveal sharp differences in regional contributions to GDP, which is consistent with the findings coming from survey data. But, in terms of value, the Brazilian per capita income reported by the National Accounts, is roughly twice the per capita income of that computed using survey data.

Finally, in the last row of Table 2.3, the growth of total value added is provided accumulated by period 1994-2002. The reported values confirm what we have already seen, i.e. the North and the Central-West are the two regions showing greater economic improvements.

The investigation of Brazilian welfare through levels and distribution of income among regions should provide a more informative analysis when coupled with a detailed poverty profile study. Moving toward poverty analysis, the identification of poor people can be conducted only when poverty lines are set. In this study we adopt a set of absolute poverty lines constructed by Rocha (2003) on the basis of geographical differences, in order to highlight regional differences.

Studies of Brazilian poverty have used several definitions of poverty lines, mostly based on the concept of absolute poverty. Although the 1$ a day poverty line set by the World Bank has sometimes been used for international poverty comparison, the most common method for defining Brazilian poverty lines has been the adoption of the minimum wage or its multiples.[13]

With more available consumption data, poverty lines can be assessed by using information on the structure of household consumption. The only two expenditure surveys that are available in Brazil are the *Pesquisa de Orçamentos Familiares* (POF) for 1987/88 and the *Estudo Nacional de Despensa Familiar* (ENDEF) for 1974/1975.

[12] In the reference week of the 2002 PNAD survey, the exchange rate US dollar against Brazilian Reais was 3.12.

[13] Referring to Rocha (2003), among the most famous studies that constructed poverty lines on the basis of the minimum wage, we should remember Pfeffermam and Webb (1983), Hoffmann(1984), Fox and Morley(1991) and Tolosa (1991).

Looking at the literature that has tried to estimate Brazilian poverty lines based on consumption data,[14] the choice of measuring poverty taking by geographically differentiated poverty lines is well-established and it provides more reliable results.

Rocha estimates geographically-specific poverty lines on the basis of the cost of basic needs approach.[15] This approach estimates the minimum cost of food required to achieve the recommended calorie intake.[16] Obviously, food baskets vary across geographical locations, such as municipalities, metropolitan areas, urban and rural areas, since preferences and prices change substantially. Rocha (2003) estimates the minimum cost of food baskets for nine metropolitan areas by using the POF survey and then she estimates values for urban and rural areas by the implementation of conversion factors provided by Fava (1984) and based on the ENDEF survey. For the non-food expenditure component, Rocha estimates adjusted values for each metropolitan area, avoiding the standard method that exploits the inverse of the Engel coefficient (Rocha, 1997). Thus, the final value of each geographically-specific poverty line is the sum of the food and non-food components. In her recent book (2003), Rocha reports 24 specific poverty lines at 1990-99 current prices.

In order to measure poverty by region, we need to match Rocha's poverty line areas with the five geographical regions, as reported in Table 2.4. The values of these poverty lines are in 2002 prices: the conversion has been made using the CEPAL deflator equal to 166.1 with 1995 as base year (ECLAC, 2004).

By applying Rocha's poverty lines, we are able to compute the poverty indexes for Brazil and each of its regions, together with their standard errors shown in Table 2.5. Looking at regional differences, the pattern that we find in income distribution analysis is reproduced.

[14] Referring to Rocha (2003), the first poverty lines estimations based on consumption data are Thomas (1982) and Fava (1984). Rocha (1988) estimates poverty lines using consumption data derived from ENDEF. Then, following studies adopted consumption data coming from the POF, such as Rocha (1993) and Rocha (2003).

[15] On the Basic Need approach, see Streeten (1981).

[16] The minimum caloric requirement is estimated by FAO (1985), as Rocha indicated in her book (Rocha, 2003, page 54).

The North-East region is not only the most unequal region but also the poorest. The North and the Central-West follow, both with values substantially above the Brazilian average. Finally, the South-East and the South are the regions that contain the fewest poor people. Figures 2.2, 2.3 and 2.4 give an even clearer picture of regional differences by poverty index.

After computing Brazilian poverty and income distribution via simple descriptive statistics, the investigation on the main characteristics of poor people by geographical region has been found necessary. The poverty profile for Brazilian households is provided in table 2.6. It follows the methodology previously used by Fishlow (1972) and simply takes the Headcount ratio and analyzing the characteristics of household heads below the poverty line for each region.

We explore several individual characteristics of the household head, such as gender, age, race and level of schooling, as well as characteristics of the household head related to her employment situation, i.e. whether she is active, whether she works in the formal sector, and if so, in which economic sector and in which position. More general characteristics related to the whole family are also considered. The first one is the geographical location within regions, including urban or rural status. We also consider other family characteristics, i.e. the family size, the number of workers and children per family.

The personal characteristics of the household head do not vary much by region. On average, household heads among the poor are men aged between 35 and 45 years with an intermediate level of schooling.

The main difference among regions when looking at personal characteristics of the household head is *race*. Not surprisingly, the majority of the Brazilian poor are black, while the non-poor are white: hence skin colour can be considered as a crucial determinant of poverty in Brazil.[17] Focusing on regional patterns, in the North and North-East the majority of the population is black, so both poor and non-poor people are predominantly black. The

[17] Giving the fact that racial discrimination is a fundamental problem in Brazil, a number of papers have investigated Brazilian income inequality and poverty by race, such as Lovell (1999), Telles (2006) and Wood (1991).

reverse is true in the South, where the population is primarily white. The South-East and Central-West follow a similar pattern to that of the country as a whole: the majority of the black population is poor, while the majority of non-poor population is white.

Level of education is another crucial characteristic of the Brazilian poverty profile. Almost all the household heads among the poor have mid level education. But very few people have attended high school and in the profile we produce, no poor household heads have attended college. These findings are in line with other empirical studies on social conditions in Brazil showing that low returns to secondary school education and a lack of access to graduate and postgraduate education for the majority of the population are the most important determinants of Brazilian inequality and poverty.[18]

As a likely consequence, the majority of the poor household heads work in blue collar professions without any significant variations across regions.

Moving to other characteristics related to the labour market, we notice that the majority of the poor household heads are economically active. Obviously, having a job cannot be deemed as a cause of poverty; the mechanism behind our empirical findings can be hypothesized to be that it depends primarily on the position of occupation and on the economic sector. While occupational position is almost constant across regions, the *economic sectors* where poor household heads are employed varies across regions. We can individuate two main groups: in the North and in the North-East, poor household heads predominantly work in agriculture and trade; while in the South, the South-East and the Central-West, poor people are employed not only in the agricultural and commercial sectors, but also in construction and industry, particularly in the South.

The characteristic *formal* identifies if the household head works in the formal sector. The percentage of people working in the informal sector is always more than one third and is higher for poor people. Particularly, it is noticeable that

[18] A large literature on Brazilian welfare focuses on education as the major determinant of income inequality and poverty, for example Ferreira and Paes de Barros (1999) and Ferreira and Litchfield (2001).

in the North we find that the majority of poor people are employed in the informal sector.

The variable *urban* shows how the Brazilian poor are concentrated in urban areas.

Looking at characteristics related to family structure among poor people, the *family size* variable considerably varies across regions: in the North and in the North-East the majority of poor families have over 6 members, while, in the rest of Brazil, poor families consist on average of four or five individuals.

Although the majority of Brazilian households have two or three workers, families with one worker are more likely to be poor than families with two or three workers. As a consequence, poor families are likely to show higher dependency ratios computed as family size over number of worker because poor individuals belong to larger households with fewer workers. Also the number of children per family varies considerably between poor and non poor families. On average poor families tend to have two or three children while the majority of non-poor Brazilian families do not have children or have only one.

2.3 A reformulation of the FGT class of poverty measures

The standard approach to measuring poverty consists of computing the well-know FGT class of measures by using a unique poverty line, i.e. the critical threshold below which one can be considered poor.[19]

The definition of a poverty line implies crucial methodological choices that significantly affect the overall figures of poverty analysis as well as the sketched poverty profile. This threshold can be set by adopting a one-dimensional indicator of welfare, such as income or consumption. However, there is a growing consensus within the economics community in favor of the

[19] See the World Development Report 2000/2001: Attacking Poverty (World Bank, 2000).

adoption of a wider concept of welfare that might include more subjective criteria, from education, health and housing to vulnerability and dignity.[20]

In this study we have chosen to measure poverty using a one-dimensional indicator of welfare, but this still involves several important choices. First of all, we take into account the often debated choice between income and consumption. As stressed by Deaton (1997) and by Ray (1998), consumption is generally preferred to income for two fundamental reasons: consumption accounts for self-owned production and non-employed income and is a long-term measure of welfare not affected by fluctuations in income.[21] For studies of Latin American countries income is generally used due to the greater availability of data, whereas in other developing countries consumption data is more often available. The underreporting of overall welfare implied by the adoption of income as an indicator instead of consumption characterizes Latin American household surveys, including the Brazilian survey, and should be taken into account when interpreting data and outcomes (Wodon et al, 2000).

A second and even more contentious issue related to the definition of the poverty line is the choice between absolute versus relative poverty lines. The absolutist concept of poverty embraced by Sen (1983a) starts from the fundamental assumption that there is a certain level of needs below which it is not possible to survive, while the relative concept is anchored to the income levels, or consumption levels, of other individuals in a given country.

The choice between a unique poverty line and a set of differentiated poverty lines is the third critical issue. The limitations in adopting a unique poverty line are well-explored by poverty literature and Chiappero and

[20] Plenty of economists have explored different notions of well-being in contrast with the money-metric approach. Surely, the most important references are Sen's works (1976, 1983b, 1985, 1992). The literature spans from Lipton and Ravallion (1995) and Baulch (1996) to the new multidimensional poverty approach, such as Bibi (2003), Atkinson (2003) and Bourguignon and Chakravarty (2003).

[21] Although consumption is generally preferred because its consistency with the life-cycle theories of consumption, it might not hold when a lack of access to insurance and credit markets is detected, as is likely in developing countries and more broadly speaking in the most vulnerable and deprived part of the population (Lipton and Ravallion, 1995).

Civardi (2006) suggest the implementation of differentiated poverty lines for homogenous population groups.

The most evident weakness in considering the whole population as an homogenous group, and using an unique threshold for poverty measurement, is that it fails to acknowledge one of the most important characteristics of the real world. The heterogeneity of individuals and households among the entire population cannot be ignored: differences in personal characteristics and in the social environment affect the level and composition of needs and, as a consequence, the level of deprivation.

The hypothesis of the “representative agent” in the context of poverty analysis does not take into account the existence of many dissimilar personal and household characteristics as well as different socio-economic contexts. In studying levels of poverty and welfare we should keep in mind that individuals usually compare their condition to other analogous situations, thus the idea of relative deprivation cannot be ignored and methodological tools should take this into account in order to sketch more reliable poverty profiles.

In their work, Chiappero and Civardi (2006) propose a conceptual framework that considers the potential heterogeneity of individual and households and advances a new analytical approach by reformulating the FGT class of measures for absolute, relative and hybrid[22] poverty lines.

Their methodology can be summarized in four steps. A set of homogenous groups can be identified following a specific criterion. Then a specific (absolute or relative) poverty line has to be defined for each homogenous group. The third step involves the choice of a common community-wide threshold. Finally, the level of poverty is measured via this reformulation of the FGT class of poverty indexes that is able to capture the within- and between-group components.

This method for computing poverty generates a poverty analysis that conveys not only how much poverty there is within each homogeneous group, but also how much poverty exists between different groups.

[22] For further information on the notion of hybrid poverty lines, see Citro and Michael (1995).

The within-group component identifies poverty existing in each homogenous group once its own group-specific poverty line is applied. The outcomes from the within component computation are equal to poverty outcomes resulting from the standard FGT class of measures using differentiated poverty lines.
The between-group component tells us to what extent individuals from each homogenous group are deprived relatively to a community-wide poverty line. This community-wide poverty line is basically a poverty line taken as a reference for comparison between groups. This reference point can be a conventional threshold computed as a given percentage of the mean or median income or estimated from consumption behavior, or it can be a poverty line chosen from the set of differentiated poverty lines assigned to the homogenous groups (Chiappero and Civardi, 2006).

There are many criticisms that might arise once this new approach is analyzed. The problem of "subjectivity" in defining the criteria employed to identify homogenous group is an unsolved topic. The problem in choosing relative versus absolute poverty lines is still present. When relative poverty lines are adopted, poverty outcomes are affected by the degree of inequality existing in the society. Similarly, if all the individuals are above an absolute level of needs, the poverty issue vanishes for even higher level of inequality.

Below we briefly outline the analytical framework of this reformulation, restricted to the case of purely absolute poverty lines. The reason of this restriction is the fact that the empirical exercises proposed in Section 2.4 adopt only differentiated absolute poverty lines.
We start from the standard FGT class of measures that incorporates the three most common poverty indexes, such as the Headcount Ratio (H), the Poverty Gap (PG) and the Squared Poverty Gap (SPG).
For each $\alpha \geq 0$, this class of measures is usually formulated by

$$P_\alpha\left(y_j;z\right)=\frac{1}{n}\sum_{j=1}^{q}\left(\frac{z-y_j}{z}\right)^{\alpha}, \qquad \text{for } y_j<z, \tag{1}$$

where y_j is a vector of the income of each individual or household j with $j=1....q$ poor individuals among a population of n individuals. The poverty line is identified by z, while the term α is the weight given to income gaps below the poverty line.

When $\alpha=0$ the above formula becomes the Headcount Ratio, P_0, The *Headcount Ratio* gives the incidence of poverty as follows

$$P_0 = H = \frac{q}{n}. \tag{2}$$

If $\alpha=1$ the formula becomes the Poverty Gap, P_1, which describes the intensity of poverty as follows

$$P_1 = PG = \frac{1}{n}\sum_{j=1}^{q}\frac{z-y_j}{z}. \tag{3}$$

Finally, if $\alpha=2$ the measure becomes the *Squared Poverty Gap* or P_2, which gives the severity of the poverty, i.e. the inequality among poor people as follows

$$P_2 = SPG = \frac{1}{n}\sum_{j=1}^{q}\left(\frac{z-y_j}{z}\right)^2. \tag{4}$$

The greater the α term, the greater the weight given to the lower part of the income distribution, hence in the Squared Poverty Gap, incomes far from the poverty line carry more weight.

We assume that the population size, n, can be divided into k groups, mutually exclusive, following a specific criterion that is able to define homogenous groups, i.e. gender, ethnicity or regional location.

For each k group a specific absolute poverty line, z_i, with $i=1\ldots k$, is identified; in this case, an absolute poverty line, z_k, defines a minimum level of basic needs that should be reached for the specific k-group of the population in order to be considered non-poor. Differences in this minimum level of basic needs among groups might depend on differences in their availability and differences in their prices.

This reformulated poverty measures aims to identify a within-group component, i.e. the number of people living below the group-specific poverty line, and the between-group component, which captures the level of poverty within each group when measured against a community wide poverty line.

Let y_j be a vector of household incomes and z_i be the set of differentiated poverty lines, both ranked in a non-decreasing order, the overall poverty $P_{WB\alpha}$ is the sum between the within component $P_{W\alpha}$ and the between component $P_{B\alpha}$ as follows

$$P_{WB\alpha}(y_j;z_i) = P_{W\alpha}(y_j;z_i) + P_{B\alpha}(y_j;z_i). \tag{5}$$

The within component is given by

$$P_{W\alpha}(y_j;z_i) = \sum_{i=1}^{k} P_{\alpha i}(z_i)\frac{n_i}{n}. \tag{6}$$

The within component is then equal to the overall poverty if there is no difference among poverty lines, i.e. $z_1 = z_2 = \ldots = z_k$.

The between component is formulated by

$$P_{B\alpha}(y_j;z_i) = \sum_{i=1}^{k-1} \left[P_{\alpha i}(z^*) - P_{\alpha i}(z_i)\right]\frac{n_i}{n}. \tag{7}$$

where z^* represents the reference point, i.e. the threshold used as a community-wide poverty line. As Chiappero and Civardi (2006) highlight, the between component is positive when $z_i<z^*$ and it is negative when $z_i>z^*$. The reference point z^* can be a conventional value, such as a poverty line taken from the given set of k poverty lines.

In our empirical analysis, we find reasonable for the purpose of this study to compare each group to the group with the highest poverty line in order to compute the between-group component, hence $z^*=z^k$. This means that each group is compared with the kth poverty line after having arranged this set in a non-decreasing order and that the between-group component is always positive. The choice to use the group with the highest poverty line as the community-wide threshold is motivated by the extent to the possibility of income redistribution at the national or regional level.

Although differentiated poverty lines do not necessarily correspond to different standards of living, we can look at them as a frame of reference in detecting those groups that are more privileged than others. Hence comparing each group to the "luckiest" one can give the extent to how far away they are from the reference group, i.e. the group with the highest poverty line.

From the policy-maker's perspective, this approach reflects the need for an estimate of the effort needed to reach a convergence among different groups toward a common desirable relatively higher threshold. For this reason, we find appropriate to set the community-wide threshold at the level of highest poverty line.

Now, we can write the reformulation of the three poverty indexes and individuate the within- and between-group components in each case.

The *Headcount ratio* can be written as follows:

$$H_{WB}\left(y_j;z_i\right)=\sum_{i=1}^{k}H_i(z_i)\frac{n_i}{n}+\sum_{i=1}^{k-1}\left[H_i(z^k)-H_i(z_i)\right]\frac{n_i}{n} \tag{8}$$

where the first term identifies the within component, H_W, as a weighted average of the headcount ratios, and the second term represent the between component, H_B, where each headcount ratio is compared with the headcount ratio of the k^{th} group taken as reference group.

Similarly, the *Poverty Gap* is defined by the following formula:

$$PG_{WB}\left(y_j;z_i\right)=\sum_{i=1}^{k}PG_i(z_i)\frac{n_i}{n}+\sum_{i=1}^{k-1}\left[PG_i(z^k)-PG_i(z_i)\right]\frac{n_i}{n} \tag{9}$$

and the *Squared Poverty Gap* is defined as:

$$SPG_{WB}\left(y_j;z_i\right)=\sum_{i=1}^{k}SPG_i(z_i)\frac{n_i}{n}+\sum_{i=1}^{k-1}\left[SPG_i(z^k)-SPG_i(z_i)\right]\frac{n_i}{n} \tag{10}$$

where, for both indexes, it is possible to identify the within-group component, which is the first term, and the between-group component, which is the second term at the right hand side of both equations.

By computing the values of the additive terms as percentages of the overall indexes, it is possible to check which component is dominant.

When the within-group component is dominant, it means that poverty exists primarily within homogenous groups. Conversely, if the between-group component dominates, poverty between groups is greater than within groups due to significant heterogeneity between groups with respect to the community-wide threshold.

2.4 Empirical exercises on decomposability of the FGT class of measures

The empirical exercises we present in this section are based on the conceptual and analytical reformulation of the FGT class of poverty indexes carried on by Chiappero and Civardi (2006). The data come from the Brazilian households survey for 2002 and have been summarized in section 2.2.

Starting from Rocha's 2003 definition of group-specific absolute poverty lines by geographical location, the computation of poverty between and within these groups should provide additional information on poverty in Brazil.

As already mentioned, this poverty decomposition allows us not only to compute absolute poverty levels within each homogeneous group, but also to capture the between-group component that is otherwise ignored.

The within-group component is the sum of the poverty levels calculated for each homogeneous group by adopting its group-specific absolute poverty line. The between-group component emerges by applying the same community-wide threshold to each homogenous group.

Table 2.7 shows the results of this poverty decomposition after adopting homogenous geographically specific poverty lines, while using the Brazilian group with the highest poverty line used as the community-wide reference group. As a consequence of this empirical design, the between-group component is always positive and provides the aggregate value of additional poverty experienced by each group when compared with the reference group. In particular, this group for Brazil, following Rocha's estimations, is the metropolitan area of São Paulo and its poverty line is adopted as the community-wide threshold for this exercise.

As discussed in the previous section, the choice of setting the community-wide threshold at the level of the highest poverty lines is driven by a specific *ratio*: the policy maker should be interested in working for the convergence of each group toward a common desirable level of welfare. For this reason it is worthwhile to compute how far each group is from this community-wide threshold that is captured by the between-group component, following the methodology we have adopted.

The table reports the total values of the reformulated FGT class of measures together with their within- and between-group components. The absolute value of both components shown in the table is followed by the share of that component as a percentage of the total value.

The table also records the contribution to both components provided by each region. It is important to highlight that each region is not a homogenous group, since we adopt 25 geographically specific groups. Each region has more than one homogenous group. Analyzing the contribution of each Brazilian region to either the within- or the between-group component might help us to better understand Brazilian regional disparities in analyzing poverty.

The overall values for the reformulated FGT class of measures are greater than the standard FGT values shown in table 2.5 because of the positive between-group components. The within-group component is dominant for the Headcount ratio, but looking at the Poverty Gap and the Squared Poverty Gap, the between-group component becomes increasingly significant. The measurement of the depth and severity of poverty is more sensitive to the between-group component than is the poverty incidence.

Again, the contribution of each Brazilian region to determining both components can help us to get a more complete picture of the situation. Because the North-East is the region with the highest poverty and inequality levels, it is also the region that makes the largest contribution to both the within- and between-group components.

The second region largest contribution comes from the South-East: this is a quite surprising result. Our previous investigations convey that the South-East is the richest region in terms of mean income, GDP values and traditional poverty measures. Clearly using the reformulated poverty measures adds some important information.

Such differing results are likely due to the fact that both components are weighted to the population share of each region, and the fact that the between-group component is very sensitive to the heterogeneity of the poverty line values. The South-East is the most populated region, and as such its poverty levels are weighted more heavily when the poverty measure takes population shares into account. Moreover, the between-group component of this region is noticeably inflated by the great variability of its set of poverty lines.

A final comment is that the contribution of each region varies across poverty measures. In particular, the contribution of the North-East becomes increasingly significant as we move from the Headcount Ratio to the Poverty Gap and Squared Poverty Gap, and it diverges increasingly from the South-East and other regions. It seems that when we consider poverty depth and severity the North-East is the region that performs worst.

It is important to highlight a primary reason why between-group components are so dominant in this poverty decomposition exercise. We are using an estimated population from a sample that covers the entirety of Brazil.

Hence we are comparing a large number of geographically homogenous groups with respect to a unique reference for the entire country. Having analyzed the huge differences in poverty and income distribution across the country, the between-group component is predictably dominant when we use a large number of different poverty lines.

In order to run a more realistic and refined exercise, it could be useful to apply this poverty decomposition by region; this means applying the same procedures to each of the five geographical regions separately taken, always using the group with the highest poverty line in each region as the community-wide reference group.

Poverty analysis that considers the notion of relative deprivation is very significant and often overlooked. As such it seems sensible to assume that a person not only compares her own situation to that of a group of people with similar personal and socio-economical characteristics, but that she also compares herself with people with different characteristics that she has seen, or with whom she experiences some kind of relationship.

As geographic location is one of the main sources of heterogeneity in Brazil, we find it more reasonable to assume that an individual living in, say, Amazons, compares herself with people living there. If she wants to compare herself with different people, she is more likely to compare herself to the wealthiest people living in Belem, the capital of that region, rather than with the wealthiest in São Paulo.

Table 2.8 provides findings from the poverty decomposition by region following the same structure as table 2.7. The within-group component dominates for all of the indexes in the North, North-East and Central-West. The pattern changes for the remaining Brazilian regions, where the within-group component gets noticeably smaller, while the between-group component dominates when looking at the depth of poverty for the South and at the severity of poverty for both remaining regions.

So, what we find is that in the North the within-group component dominates due to the high level of poverty in all of the homogenous groups. The North-East has a very consistent within-group component, but the sharp differences among groups generate large values for the between-group components, and noticeably shrink the within-group component, although the latter is still dominant. The South-East shows a small within-group component because of the low level of poverty in this region compared to the two previous ones. Hence the variation given by the between-group component does not have to be very large to dominate the within-group component. The South shows an even more dramatic situation. Since this region has the lowest level of poverty, it is within-group component is very low. Finally, the Central-West presents a situation similar to the North because of the high level of poverty within each homogenous group.

These findings cannot be immediately intuitive, but we can suggest some observations that might be useful in interpreting this pattern. The dominance of the between-group component is not dependent on the size of the sample for each region, nor on the number of groupings within each region, because the reformulation of the poverty indexes is still weighted by population. That said, the population size of each group belonging to each region is important in determining the weight of both components.

The mapping of the differentiated poverty lines, i.e. the delineation of each homogenous group, also plays a crucial role in determining the dominance of the between or of the within-group component. In particular, the definition of the reference group, and its size in terms of population, is fundamental in establishing the value of the between-group component.

The sensitiveness of poverty lines for each homogeneous group to shifts towards the wealthiest poverty threshold as well as the poverty levels of the homogenous groups with a significant weight in term of population size are crucial factors that affect the extent to which between or within components dominate. The between-group component tends to be large when the community-wide poverty line is significantly higher than the group-specific poverty lines, and when the population of the lower income groups is very large. This circumstance generates the sharpest changes in the poverty measures.

Finally the relationship between inequality and the dominance of the between-group component does not seem to be so straightforward. Inequality among different homogenous groups within the regions determines the dominance of one or the other component.

In the exercise at national level, at the beginning of this section, we infer the existence of a relationship between inequality and the between-group component because inequality deepens potential discrepancies in welfare among heterogeneous groups. This second empirical exercise which decomposes poverty by region provide no evidence for a strong relationship between inequality and the dominance of the between-group component. The North-East, the most unequal region in Brazil, shows a pattern similar to the two other regions with the highest inequality, the Central-West and the North. Were there a strong relationship between inequality and the between-group component, these three regions are expected to have the highest values for the between-group component. However, the within-group component dominates in these three regions.

By contrast, the most egalitarian regions of Brazil, the South and the South-East, show the highest dominance of the between-group component. In these two regions, the between-group component easily dominates due to the low level of poverty within homogenous groups. When the within-group component is huge, the between-group component needs to be large in order to be able to dominate. When the within-group component is small, the between-group component does not need to be very large to dominate.

To sum up, the within-group component is dominant in the North and the Central-West due to the high level of poverty within each group. By contrast, in the South-East and the South, where poverty levels are lower, the between-group component dominates. The North-East follows a pattern similar to the North and the Central-West but with a lower contribution of poverty within groups. This may be surprising given that the North-East is the region recording the highest level of poverty, and thus would be expected to have the highest contribution of the within-group component across regions. Nonetheless it is also the region with the highest level of inequality and this inequality allows the between-group component to shrink the within-group term. Thus the within-group component is still dominant in the North-East due to the high levels of poverty, but not to the same extent as in the North and Central-West, as the North-East also has a very high level of between-group poverty.

2.5 Conclusions

The aim of this paper is to apply and interpret the empirical findings arising from the application of Chiappero and Civardi's 2006 poverty measures reformulation to Brazilian household survey data.

The reformulation aims to decompose poverty into between- and within-group components by applying group-specific poverty lines. The empirical exercises have been conducted using Brazilian data and applied geographically specific absolute poverty lines provided by Rocha (2003) to identify homogenous groups. This choice is mainly due to the fact that Brazil is a country characterized by sharp regional discrepancies. Thus geographic location plays a significant role in dividing the country into homogeneous groups.

We run two empirical poverty decomposition exercises. First we consider the whole country and we refer to a unique reference group, the metropolitan area of Brazil, São Paulo. We find that the between-group component dominates due to the huge differences in income between all of the Brazilian homogenous groups and the metropolitan area of São Paolo.

Then, being aware of the deep differences among Brazilian regions, we run the poverty decomposition by region, assigning a reference group to each region.

The North and the Central-West analysis reveals a dominance of the within-group component, due to the high level of poverty in these two regions. The North-East shows the highest level of poverty, even higher than the North and the Central-West, but the high within-group component is counterbalanced by a higher between-group component, attributable to the high level of inequality of the North-East. The other two regions both reveal a dominant between-group component. More precisely, the South and the South-East have the lowest levels of poverty, and the between-group component therefore easily dominates the within-group component.

Looking at these findings, we believe that this poverty decomposition approach, using both between-and within-group measures, is more informative than the standard approach when differentiated poverty lines are adopted.

This alternative way of measuring poverty highlights the importance of keeping poverty and inequality analysis separate. Indeed, both analyses are important and they cannot substitute for one other, as argued by Sen (1983a). This is particularly important with regard to policy implications. When a rise in inequality is detected, policy makers should be more focused on fiscal policies and particularly on policies related to social mobility that could improve income distribution in the long run. By contrast, increases in poverty may demand more immediate interventions to combat destitution and to increase access to basic needs and income.

In summary, we should be aware that behind our analysis of the dominance of the between- or the within-group components of poverty lies a deep understanding of the complex relationship between poverty levels, income distribution and the robustness of poverty lines. This last remark renews the importance of having a critical eye in interpreting the many different indexes of poverty.

References 2

Atkinson, A.B., (2003), "Multidimensional Deprivation: Contrasting Social Welfare and Counting Approaches", *Journal of Economic Inequality*, Vol. 1, No. 1, pp. 51-65.

Baer, W., (2001), "The Brazilian Economy, Growth and Development", 5th Edition, Wesport: Praeger Publishers.

Baulch, R., (1996), "Neglected Trade-Offs in Poverty Measurement", *IDS Bulletin*, Vol. 27, No. 1, pp. 36-42.

Bibi, S., (2003), "Measuring Poverty in a Multidimensional Perspective: A Review of Literature", mimeo, CIRPEE, Université Laval, Quebec, Canada.

Bottiroli-Civardi, M., and E. Chiappero-Martinetti, (1999), "Strutture Familiari e Povertà: un Rapporto tra diversi Metodi di Analisi", in Banca d'Italia, (ed.) Ricerche Qualitative per la Politica Economica 1997, Roma: Banca d'Italia.

Bourguignon, F., and S.R. Chakravarty, (2003), "The Measurement of Multidimensional Poverty", *Journal of Economic Inequality*, Vol. 1, pp. 25-49.

Buhmann, B., L. Rainwater, G. Schmaus and T. Smeeding, (1988), "Equivalence Scales, Well-being, Inequality and Poverty: Sensitivity Estimates across Ten Countries using the Luxembourg Income Study database", *Review of Income and Wealth*, Vol. 34, pp. 115-142.

Chiappero, E. And M. Civardi, (2006), "Poverty between and within groups: a reformulation of the FGT class of index", *IRISS Working Paper*, No. 6, CEPS/INSTEAD, Differdange, Luxemburg.

Citro, C.F., and R.T. Michael, (1995), "Measuring Poverty: a new approach", National Academy Press: Washington D.C.

Deaton, A., (1997), "The Analysis of Household Survey: A Microeconometric Approach to Development Policy", Baltimore and London: John Hopkins Press and World Bank.

ECLAC, (2004), “Statistical Yearbook for Latin America and the Caribbean”, Economic Commission for Latin America and Caribbean, Santiago de Chile: ECLAC.

FAO (1985), “Necessidades de energia e proteinas”, *Série Informes Técnicos*, No.724, Geneva: FAO/WHO.

Fava, V. L., (1984), “Urbanização, Custo de Vida e Pobreza no Brasil”, *Série Ensaios económicos*, Vol. 37, São Paulo: IPE/USP.

Ferreira, F. H. G., P. G. Leite and J. A. Litchfield, (2006), “The Rise and Fall of Brazilian Inequality: 1981-2004”, *World Bank Policy Research Working Paper*, No. 3867, Washington, D.C.: World Bank.

Ferreira, F.H.G, and J.A. Litchfield, (2001), “Education or Inflation? The Micro and Macroeconomics of the Brazilian Income Distribution during 1981-1995”, in N. Hicks and A. Valdés, (eds.), Especial Edición de Cuadernos de Economía, Vol. 114, pp. 209-238.

Ferreira, F.H.G. and R. Paes de Barros, (1999), “The Slippery Slope: Explaining the Increase in Extreme poverty in Urban Brazil, 1976-1996”, *Brazilian Review of Econometrics*, Vol. 19, No. 2.

Fishlow, A., (1972), “Brazilian Size Distribution of Income”, *The American Economic Review*, Vol. 62, No. 1/2, pp. 391-402.

Foster, J., J. Greer, and E. Thorbecke, (1984), “A Class of Decomposable Poverty Measures”, *Econometrica*, Vol. 52, pp. 761-65.

Fox, M.L. and S.A. Morley, (1991), “Who Paid the Bill? Adjustment and Poverty in Brazil, 1980-95”, *World Bank Policy, Research and External Affairs Working Paper*, No.648, Washington, D.C.: World Bank.

Glewwe, P., and J. Van der Gaag, (1990), “Identifying the Poor in Developing Countries: Do Different Definitions Matter?”, *World Development*, Vol. 18, No. 6, pp. 803-14.

Hoffmann, R., (1984), “Pobreza no Brasil”, *Série Estudos e Pesquisas*, No.43, Piracicaba: Esalq.

IBGE, (2005), "Conta Regionais do Brasil, 2002", Rio de Janeiro: IBGE.

Lipton, M., and M. Ravallion, (1995), "Poverty and Policy", in Behrman J., and Srinivasen T. N., (eds.), Handbook of Development Economics, Vol. 3B, Amsterdam: Elsevier.

Litchfield, J. A., (2001), "Welfare and Income Distribution in Brazil, 1981-1995", PhD Dissertation, London: London School of Economics.

Lovell, P., (1999), "Development and Persistence of Racial Inequality in Brazil: 1950-1991", *Journal of Developing Areas*, Vol. 33, No. 3, pp. 395-417.

Pfefferman, G. and R. Webb, (1983), "Pobreza e Distribuição de Renda no Brasil", *Revista Brasileira de Economia*, Vol. 37, No. 2.

Ray, D., (1998), "Development Economics", Princeton University Press.

Rocha, S., (1997), "Do Consumo observado à Linha de Pobreza", *Pesquisa e Planejamento Economico*, Vol. 27, No. 2, pp. 313-52, Rio de Janeiro: IPEA.

Rocha, S., (2003), "Pobreza no Brasil, Afinal, de que se trata?", Rio de Janeiro: Editora FGV.

Sen, A., (1976), "An Ordinal Approach to Measurement", *Econometrica*, Vol. 44, pp. 219-232.

Sen, A., (1983a), "Poor, Relatively Speaking", *Oxford Economic Papers, New Series*, Vol. 35, No. 2, pp. 153-169.

Sen, A., (1983b), "Poverty and Famines: An Essay on Entitlement and Deprivation", Oxford: Oxford University Press.

Sen, A., (1985), "Commodities and Capabilities", Amsterdam: North Holland.

Sen, A., (1992), "Inequality Reexamined", Oxford: Oxford University Press.

Streeten, P., (1981), "First things first: Meeting basic needs in developing countries", New York : Oxford University Press,.

Telles, E., (2006), "Race in another America: The Significance of Skin Colour in Brazil", Princeton Press.

Thomas, (1982), "Differences in income, nutrition and poverty within Brazil", *World Bank Staff Working Paper*, No. 505, Washington, D.C.: World Bank.

Tolosa, H.C., (1991), "Pobreza no Brasil: Uma Avaliação dos Anos 80", in J.P. dos Reis Veloso, (ed.), A Questão Social no Brasil, São Paulo: Nobel.

Wodon Q., (2000), "Poverty and Policy in Latin America and the Caribbean", *World Bank Technical Paper*, No.467, mimeo, Washington D.C.: World Bank.

Wood, C., (1991), "Categorias Censitarias e Classificacoes Subjectivas de Raca no Brasil", in P. A. Lovell (ed.), Desigualidade Racial no Brazil Contemporaneo, Belo Horizonte: Univesidade Federal de Minas Gerais, Centro de Demografia e Planejamento Regional, pp. 93-113.

World Bank, (2001), "World Development Report 2000/2001: Attacking Poverty", Washington D.C.: World Bank.

Appendix 2

Table 2.1: Summary statistics for Brazilian regions using per capita income, 2002

	Brazil	North	North-East	South-East	South	Central-West
Sample size	102,500	10,126	30,886	32,504	17,572	11,412
Weighted individuals	166,270,000	9,837,205	47,676,831	71,678,789	25,285,970	11,790,515
Mean	329.85	237.51	181.89	415.89	378.59	377.57
Median	171.43	126.67	92.50	226.67	225.00	187.50
Gini Index	0.58	0.56	0.59	0.55	0.52	0.58

Source: Author's calculations from the PNAD 2002.

Table 2.2: Mean incomes per Decile by Region, 2002

	Brazil	North	North-East	South-East	South	Central-West
1	30.83	30.54	18.81	48.50	48.04	42.82
2	59.80	53.08	36.81	89.36	89.38	76.67
3	88.42	71.59	50.76	124.13	125.13	103.68
4	119.20	92.63	65.06	161.69	163.55	135.63
5	152.59	115.21	81.88	204.23	204.03	169.40
6	194.59	142.80	102.63	255.11	253.74	207.57
7	251.85	184.12	133.08	328.15	319.08	268.48
8	346.19	243.07	180.06	443.98	428.39	381.09
9	534.15	368.71	264.45	672.95	625.54	607.84
10	1533.37	1078.60	894.35	1834.12	1556.51	1798.50

Source: Author's calculations from the PNAD 2002.

Table 2.3: General indicators from National Accounts, 2002

	Brazil	North	North-East	South-East	South	Central-West
Total GDP (in millions of $R)	1,346,028	67,790	181,933	758,374	237,729	100,202
Per capita GDP, monthly (in $R)	635,91	411,58	307,83	840,50	763,08	680,50
Value Added[(a)] (percent)	0,24	0,51	0,22	0,20	0,23	0,36

(a) The evolution of the total value added is accumulated by period 1994-2002;
Source: IBGE, (2005), *Conta Regionais do Brasil*, 2002, Rio de Janeiro: IBGE ed.

Table 2.4: Brazilian per capita poverty lines, in 2002 prices

Geographical Regions matched with Rocha's Regions		Value (in $R)
Region 1: North		
Region VII	Metropolis of Belem	119.99
	Urban	104.59
	Rural[(a)]	77.64
Region 2: North-East		
Region V	Metropolis of Fortaleza	119.82
	Metropolis of Recife	163.97
	Metropolis of Salvador	153.43
	Urban	102.83
	Rural	62.02
Region 3: South-East		
Region I	Metropolis of Rio de Janeiro	164.79
	Urban	102.53
	Rural	74.84
Region II	Metropolis of São Paulo	198.57
	Urban	126.88
	Rural	79.83
Region IV	Metropolis of Belo Horizonte	136.38
	Urban	91.69
	Rural	54.28
Region 4: South		
Region III	Metropolis of Curitiba	134.03
	Metropolis of Porto Alegre	103.45
	Urban	89.16
	Rural	60.11
Region 5: Central-West		
Region VI	Brasilia	189.06
Region VIII	Goiania	177.53
	Urban	135.17
	Rural[(a)]	77.64

Source: Rocha, 2003, re-adapted by the Author.
(a) We impute to the rural poverty line for Region VII, the same value of the rural poverty line for Region VIII, following Ferreira and Litchfield (2001).

Table 2.5: Summary statistics of FGT class of measures by region, 2002

	Brazil	North	North-East	South-East	South	Central-West
Headcount	0.3359	0.4225	0.5156	0.2582	0.1455	0.4173
s.e	0.0019	0.0060	0.0035	0.0030	0.0035	0.0053
Poverty Gap	0.1357	0.1681	0.2247	0.0968	0.0480	0.1729
s.e.	0.0010	0.0032	0.0021	0.0014	0.0015	0.0029
Squared Poverty Gap	0.0742	0.0897	0.1292	0.0500	0.0236	0.0236
s.e.	0.0007	0.0022	0.0015	0.0010	0.0009	0.0009

Source: Author's calculations from the PNAD 2002.

Table 2.6: The profile of Poverty in Brazil for 2002, values in percentages of poor and non-poor population

	North		North-East		South-East		South		Central-West		Brazil	
				non		non		non		non		non
	poor	non poor	poor	poor	poor	poor	poor	poor	poor	poor	poor	poor
Gender of Head of HH												
Male	71	74.6	78.8	77.5	75	79	77.7	81	78.2	79	76.8	78.9
Female	29		21.2	22.5	25	21	22.3	19	21.8	21	23.2	21.1
Age of Head of HH												
age<25	6		5.4	3.2	4.7	2.9	6.1	3.6	6.5	4.1	5.3	3.3
25≤age≤34	27	20.4	23.5	15.4	26.1	15.9	25.8	18	27	19.9	25.1	16.7
35≤age≤44	28.4	27.1	29.1	22	32.9	26.6	34.7	28.7	31.4	27.9	30.9	26.2
45≤age≤54	20.9	22.6	21.5	21.8	19.1	25.2	18.9	23.8	18.5	23.8	20.2	24
55≤age≤64	11	14.1	12.5	18.3	10.2	15.6	10	14	10.2	14.3	11.3	15.6
x≥65	6.7	11.3	8	19.3	7.0	13.8	4.5	11.9	6.4	10	7.2	14.2
Race of Head of HH												
White	21.3	31.3	23.9	34.3	46.6	67	68.7	83.7	32.9	50.5	35	60.6
Black	78.5	68.3	75.9	65.5	53.3	32.2	31.2	15.9	66.9	48.7	64.9	38.8
Asian	0.2	0.4	0.2	0.2	0.1	0.8	0.1	0.4	0.2	0.8	0.1	0.6
Education of Head of HH												
illiterate	21.9	16.4	29.8	22.8	14.6	9.3	14.2	9.2	17.8	12.1	22.1	12.6
elementary	22	19.4	28.1	28	27.1	28.7	28.8	30.6	24	20.9	27	28
intermediate	55.8	57	41.9	41.4	57.4	49.4	56.8	50.2	57.7	53.8	50.5	48.5
high school	0.3	6.8	0.2	7.4	0.9	11.9	0.2	9.4	0.5	12.4	0.4	10.3
college plus	0	0.4	0	0.4	0	0.7	0	0.6	0	0.8	0	0.6
Head of HH Economically Active												
active	81.5	83.1	83	78.3	81.9	78.5	85.3	84.4	84.7	84.9	82.9	80.3
no active	18.5	16.9	17	21.7	18.1	21.5	14.7	15.6	15.3	15.1	17.1	19.7
Head of HH in Formal Sector												
formal	49.1	65.8	52.1	61.7	51.2	64.7	52.7	68.9	53.1	69.4	51.7	65.3

informal	50.9	34.2	47.9	38.3	48.8	35.3	47.3	31.1	46.9	30.6	48.3	34.7
Sectoral Distribution												
agriculture	15.3	8.8	35.7	24.5	11.3	8.1	27.8	17.8	19	14.2	24.1	13.8
industry	11.5	11.9	6.1	7.5	11.1	15.1	9.9	15.9	9.3	9.3	8.7	13.1
construction	11	8.2	8.7	5.7	13.1	8.3	13.5	8.4	14	7.1	11.1	7.7
trade	12.6	16.5	10.5	12.7	11.7	13.2	8.5	13.5	11.3	17.1	11	13.6
tourism	3.3	2.6	2.3	2.3	3.2	2.8	1.9	2.3	3.2	2.7	2.8	2.6
transports	4.2	6	2.8	5.1	4.4	6.8	2.6	5.8	4.3	6.3	3.6	6.2
public adm	3.7	10.4	2.3	6.3	2.3	4.8	1.5	4.8	2.7	9.5	2.4	5.7
health, educ, etc.	12	12.5	7.8	9.1	11	10.3	9.6	8.7	11.4	10.5	9.5	9.9
others	26.4	23.1	23.8	26.8	31.9	30.6	24.7	22.8	24.8	23.3	26.8	27.4
Occupation of Head of HH												
professional/technicians	1.8	12	1.6	9.9	1.3	13.7	0.8	11.9	2.1	17.4	1.5	12.7
intermediate	32.3	34.1	22.6	24.5	30.9	27.9	21.2	24	28.9	29.3	26.5	26.8
blue collars	65.9	53.9	75.8	65.6	67.8	58.4	78	64.1	69	53.3	72	60.5
Region of Family												
North	-	-	-	-	-	-	-	-	-	-	7.4	5.1
North-East	-	-	-	-	-	-	-	-	-	-	44	20.9
South-East	-	-	-	-	-	-	-	-	-	-	33.1	48.2
South	-	-	-	-	-	-	-	-	-	-	6.6	19.6
Central-West	-	-	-	-	-	-	-	-	-	-	8.9	6.2
Location of Family												
urban	96	97.2	70.2	71	90.5	92.2	75.8	82.4	85.4	88.7	80.6	85.9
rural	4	2.8	29.8	29	9.5	7.8	24.2	17.6	14.6	11.3	19.4	14.1
Family Size												
1	0.2	2.6	0.4	3.8	0.4	4	0.3	3.5	0.4	4.8	0.4	3.8
2-3	11.8	27.6	14	32.4	18.3	37.6	15.5	39	20.4	35.4	16	36.1
4-5	37.2	43.3	40.4	42	46.6	45.9	47	46.4	50.7	46.6	43.5	45.1
over 6	50.8	26.5	45.2	21.8	34.7	12.5	37.2	11.1	28.5	13.2	40.1	15

Numbers of Workers per Family												
0	4.6	3.2	4.9	6.2	5.3	6.3	4.7	5.1	4.1	4.3	4.9	5.7
1	39.2	23.7	31.8	22.9	37.7	25.2	35.1	23.4	37	24.5	35	24.2
2-3	42.7	56	47.7	54.9	47.7	56.4	51.4	60.1	49.9	58.9	47.8	57
4-5	11.2	14.3	12.2	13.1	8	10.8	7.9	10.6	7.6	11.7	10	11.5
over 6	2.3	2.8	3.4	2.9	1.3	1.3	0.9	0.8	1.4	0.6	2.3	1.6
Number of Children per Family, 0-14												
0	8.7	32.6	12.4	43.3	15	48.1	9.4	42.3	17.2	45.9	13.2	45
1	17.4	31.1	22.3	30.4	23.1	29.3	21.1	31.7	25.1	28.8	22.4	30.1
2-3	47.6	32.3	45.1	23.9	48.3	21.4	48.9	24.5	46.4	24	46.7	23.3
over 4	26.3	4	20.2	2.4	13.6	1.2	20.6	1.5	11.3	1.3	17.7	1.6

Source: Author's calculations from the PNAD 2002.

Table 2.7: Poverty decomposition between and within groups with a unique reference group for the entire country[(a)], 2002

Brazil	Hwb= 0.5447				PGwb= 0.2807				SPGwb= 0.1774			
	Hw	%	Hb	%	PGw	%	PGb	%	SPGw	%	SPGb	%
	0.3358	61.66	0.2088	38.34	0.1357	48.33	0.1450	51.67	0.0742	41.85	0.1031	58.15
	Contribution of each region:											
North	0.0250	7.44	0.0145	6.95	0.0099	7.33	0.0111	7.62	0.0053	7.15	0.0080	7.71
North-East	0.1478	44.02	0.0738	35.33	0.0644	47.49	0.0689	47.48	0.0370	49.92	0.0550	53.31
South-East	0.1113	33.14	0.0710	34.02	0.0417	30.77	0.0393	27.09	0.0215	29.04	0.0245	23.76
South	0.0221	6.59	0.0425	20.33	0.0073	5.38	0.0211	14.57	0.0036	4.83	0.0126	12.21
Central-West	0.0296	8.81	0.0070	3.36	0.0123	9.04	0.0047	3.23	0.0067	9.07	0.0031	3.00

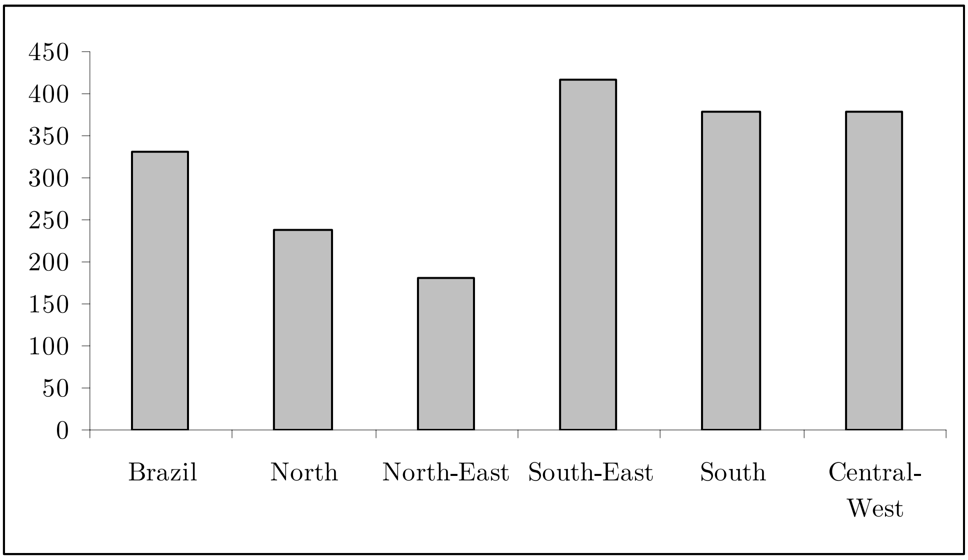
Source: Author's calculations from the PNAD 2002.
(a) The unique reference group for the entire country is the metropolitan area of São Paulo.

Table 2.8: Poverty decomposition between and within groups with a reference group for each Brazilian region[(a)], 2002

North	Hwb= 0.4670				PGwb= 0.2013				SPGwb= 0.1113			
	Hw	%	Hb	%	PGw	%	PGb	%	SPGw	%	SPGb	%
	0.4225	90.46	0.0445	9.54	0.1681	83.49	0.0332	16.51	0.0897	80.55	0.0216	19.45
North-East	Hwb= 0.7078				PGwb= 0.3825				SPGwb= 0.2490			
	Hw	%	Hb	%	PGw	%	PGb	%	SPGw	%	SPGb	%
	0.5156	72.84	0.1922	27.16	0.2247	58.74	0.1578	41.26	0.1292	51.88	0.1198	48.12
South-East	Hwb= 0.4230				PGwb= 0.1880				SPGwb= 0.1068			
	Hw	%	Hb	%	PGw	%	PGb	%	SPGw	%	SPGb	%
	0.2582	61.04	0.1648	38.96	0.0968	51.51	0.0912	48.49	0.0500	46.79	0.0569	53.21
South	Hwb= 0.2797				PGwb= 0.1052				SPGwb= 0.0555			
	Hw	%	Hb	%	PGw	%	PGb	%	SPGw	%	SPGb	%
	0.1455	52.01	0.1342	47.99	0.0480	45.60	0.0572	54.40	0.0236	42.46	0.0319	57.54
Central-West	Hwb= 0.5034				PGwb= 0.2256				SPGwb= 0.1291			
	Hw	%	Hb	%	PGw	%	PGb	%	SPGw	%	SPGb	%
	0.4173	82.89	0.0861	17.11	0.1729	76.66	0.0526	23.34	0.0950	73.57	0.0341	26.43

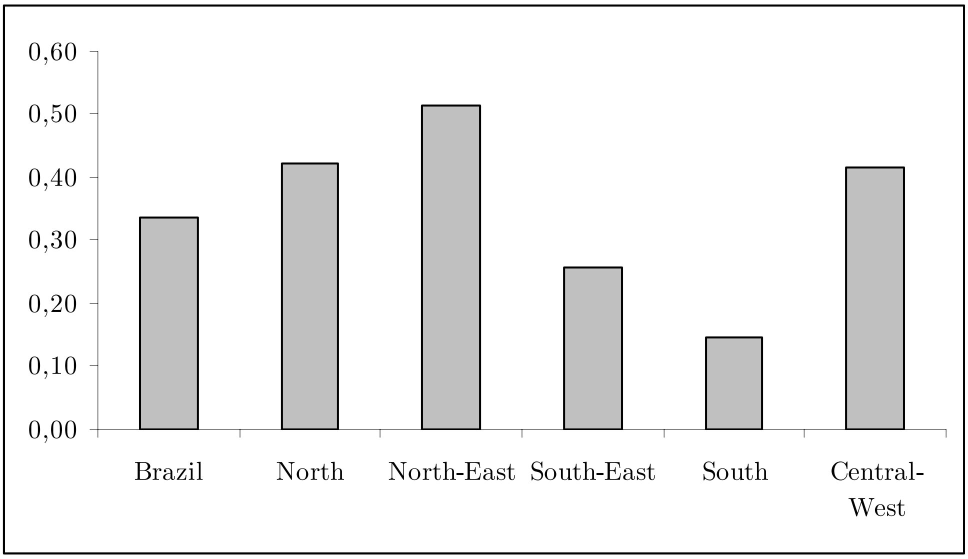
Source: Author's calculations from the PNAD 2002.
(a) The reference groups for each Brazilian region are the metropolitan area of Belem for the North, the metropolitan area of Recife for the North-East, the metropolitan area of São Paulo for the South-East, the metropolitan area of Curitiba for the South and Brasilia for the Central-West.

Figure 2.1: Regional differences in mean values, 2002



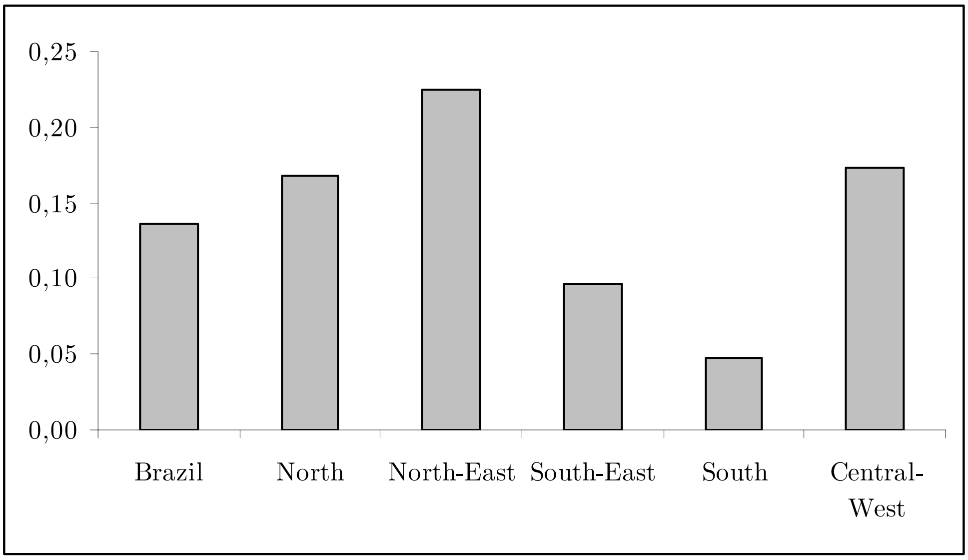
Source: Author's calculations from the PNAD 2002.

Figure 2.2: Regional differences in the Headcount ratio, 2002



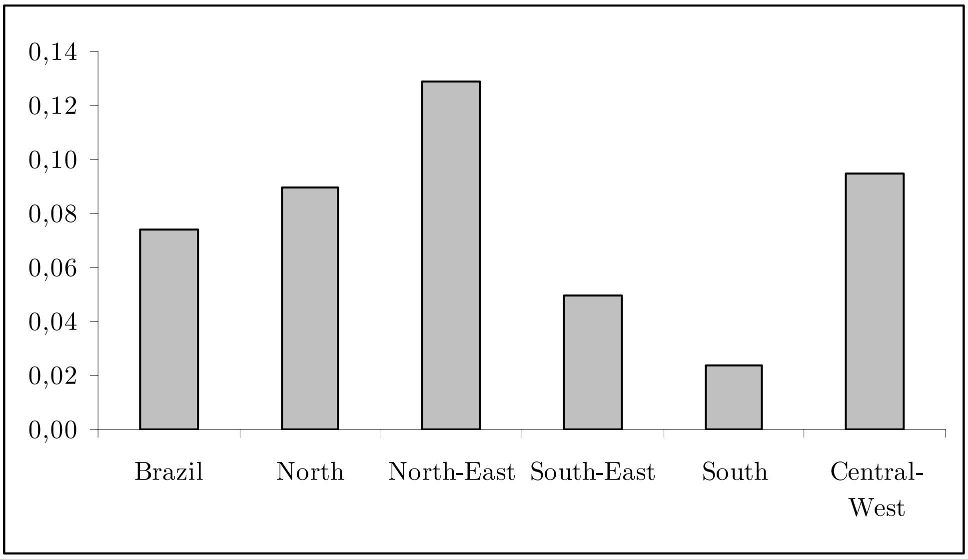
Source: Author's calculations from the PNAD 2002.

Figure 2.3: Regional differences in the Poverty Gap, 2002



Source: Author's calculations from the PNAD 2002.

Figure 2.4: Regional differences in the Squared Poverty Gap, 2002



Source: Author's calculations from the PNAD 2002.

Chapter 3

The Estimation of the Health Functioning Production Function for Brazil

Abstract: This paper aims to model and estimate the health functioning production function as a relation that conveys to what extent people are able to convert private and public resources into the achievement of the specific functioning "being healthy". This conversion process is affected by a set of internal and external conversion factors identified by exogenous individual, social and environmental characteristics. The estimation of this function has been made by employing Brazilian data. By applying probit and ordered probit regression models, we find that Brazilian young people are the most vulnerable group that convert resources less efficiently into the achieved functioning. Moreover, women are the most relevant policyholder for the Brazilian public health services. We conclude that our empirical findings might be relevant for policy making once a more comprehensive approach of assessing individual well-being is accepted.

3.1 Introduction

The aim of this paper is to construct and assess the health functioning production function. The health functioning production function defines the relationship between the achievement of the functioning "being healthy" and a set of private and public resources needed to achieve this specific functioning controlling for a set of exogenous characteristics.

The definition and estimation of the functioning production functions can be considered a valuable technique for assessing individual well-being in the context of the capability approach developed by Amartya Sen.

In fact, by adopting the Senian framework, the well-being of a person can be conceptualized by a set of achieved functionings, where these functionings are beings or doings that a person manages to achieve, such as being well nourished, being well sheltered, being educated, or living in a safe and healthy environment.

If the achievement of functionings is determined by a set of available resources subject to individual, social and environmental characteristics, we can conceive this relationship as a functionings-resources conversion process. Indeed, the individual, social and environmental characteristics might be viewed as internal and external conversion factors that affect the conversion process. The estimation of the functionings-resources conversion process conveys to what extent a person is able to convert her or his set of resources in order to achieve functionings.

The first conceptualization of the conversion process as a tool for assessing individual well-being was given by Sen (1985). After that, little has been done to deepen the analysis of the functionings-resources conversion process from both the theoretical and empirical perspectives.

Kuklys's (2005) book is the main contribution that aims to investigate if the capability approach is more comprehensive than standard welfare economics in assessing individual well-being. Kuklys' contribution is a pioneer work in the econometric estimation of a functioning production function.

Starting from these previous studies, we define and construct a functioning production function for the specific functioning "being healthy". In our perspective, the achievement of the health functioning is determined by private resources, given by an indicator of wealth, as well as public resources, identified by an index of public services, and controlled for a set of internal and external conversion factors. In order to construct the model, we exploit the conceptual analysis for modelling individual well-being provided by Chiappero-Martinetti et al (2007).

To the best of our knowledge, this paper is the first study that aims to model and compute the impact of both private and public resources on health functioning achievement in the context of the capability approach.

Understanding the functionings-resources conversion process might provide valuable results for policy purposes. The policy maker might indeed be interested to know how individuals are able to convert their resources into achieved functionings.

If the estimation of this function conveys to what extent people convert their resources, by aggregating the population into specific sub-groups we are able to estimate the ability of each population sub-group to achieve functionings.

Hence the estimation of the conversion process by population sub-groups is equally relevant for the policy maker because it helps in understanding which population-sub-groups can be considered more efficient in converting resources or more vulnerable in that well-being process. It is undoubtedly a more comprehensive way to look at individual well-being and it might be supportive in defining policy interventions.

The estimations of the health functioning production function are made by employing the Brazilian household survey, *Pesquisa Nacional por Amostra do Domicilios* (PNAD), and a specific regional dataset on Brazilian public health services, *Datasus*, for 2003. The econometric methodologies applied depend on the nature of the variables that identify the health functioning. We estimate the health functioning production function by applying both probit and ordered probit regression models. The computations have been made for the entire Brazilian sample and by population sub-groups, recognizing the relevance of our empirical findings in terms of policy implications.

When the health functioning is identified by the self-reported morbidity index, public resources are more relevant in the health functioning achievement process. White people are the least efficient in using public resources. On the other hand, when a health status indicator identifies the health functioning, private resources become predominant. White men are generally the most efficient in employing their private resources in order to achieve better health conditions.

Looking at our empirical results, Brazilian black people might be considered a vulnerable group. The Brazilian policy maker should protect this part of the population that demonstrate a lower ability to convert their private resources and a higher efficiency in using public resources. Another interesting result is the fact that women record a greater impact of public resources while for men private resources are more relevant. The Brazilian policy maker should protect these weaker sub-groups of the population. A possible policy intervention might be to promote black-targeted public provision of medical assistance and prevention. Moreover, the public health services should be aware of the fact that the highest portion of its policyholders is female.

The paper is structured as follows. Section 3.2 presents a review of the previous literature. Section 3.3 describes our economic conceptualization of the functioning production function. Data and variables are explained in section 3.4, while section 3.5 explicates the econometric methodologies employed. Section 3.6 proposes empirical results. Final remarks and conclusions are provided in section 3.7.

3.2 Previous contributions

Sen's (1985) book "Commodities and Capabilities" is considered the first theoretical contribution of personal well-being assessment in the context of the capability approach.

From a capability perspective the well-being of a person can be defined by a set of a person's functionings. The concept of functionings is a more comprehensive way of identifying personal well-being with respect to a traditional money-metric approach. Functionings is defined by what a person manages to do or to be. It thus embodies the state of a person not as a mere possessor of goods or utility.[1] Focusing on functionings means to pay attention to what a person succeeds in doing or being with the resources that she or he is able to command.

Sen conceptualizes this process analytically through the "utilization function" $f_i(\cdot)$ with the set of functioning b_i of a person *i* given by

$$b_i = f_i(c(x_i)) \tag{1}$$

where x_i is the commodities vector of person *i* and $c(\cdot)$ is the function that converts the commodities vector into the characteristics vector.

The utilization function is indeed a function that conveys how a set of commodities, particularly the characteristics of these commodities, are employed by the person *i* in order to achieve functionings.

Sen defines a more general construction of the previous formula that considers not only a particular set of functionings, but various combinations of them. The capabilities set $Q_i(X_i)$ represents the space of all possible functionings that a person values to do or to be and the general formula of $Q_i(X_i)$ of a person *i* is given by

$$Q_i(X_i) = [b_i | b_i = f_i(c(x_i)), \text{ for some } f_i \subset F_i \text{ and for some } x_i \subset X_i] \tag{2}$$

where x_i is the commodities vector selected from a given group of commodities X_i and f_i is the utilization function chosen from a given set of possible utilization functions F_i.

This more general reformulation tells how each person is able to achieve a combination of functionings b_i that she or he values from a capabilities set

[1] In his book of 1985, Sen considers three different approaches: utility, opulence and functionings. Formal economists have adopted a unique measure of person's state and interests called utility, reflected by satisfaction, happiness or desire-fulfilment. The opulent approach focuses on good possession as a more commodities-fetishist view. The well-being evaluation based on functionings aims looks at commodity-commands of a person.

$Q_i(X_i)$ given a bounded set of commodities x_i and a particular utilization function f_i that is affected by her or his personal attributes.

The conversion process of commodities into functionings is subject to the availability of commodities and to the type of the utilization function which largely depends on what Sen defines as personal and social factors (Sen, 1985). Examples of personal and social factors are respectively age, activity levels, health conditions, and the role within the family or the social conventions and rules.

After Sen's (1985) fundamental contribution little has been done in order to define and to estimate the conversion process between commodities and functionings. Some studies embracing the capability approach highlight the intricacy of translating the complex Senian conceptual framework into empirical applications.[2]

Robeyns (2003, 2005) redefines the importance of this conversion process from goods in order to achieve functionings. Goods represent means to achieve functionings while capabilities, i.e. different combinations of functioning that a person values, represent the freedom to achieve functionings. She stresses the crucial role played by conversion factors in this goods-functionings conversion process. Conversion factors are personal, social and environmental characteristics that inevitably affect person's ability to achieve functionings. The utilization function introduced by Sen (1985) has been redefined "conversion function" by Robeyns and Kyklys (2004) in order to value the conversion process of commodities into functionings or, more generally, into capabilities. In the same study, they refresh the role of conversion factors in affecting conversion processes.

[2] For more on the complexity of the operationalization of the capability approach, see Chiappero-Martinetti (2000), Robeyns (2000) and Comin (2001). Chiappero-Martinetti (2000) highlights the fact that this approach is more challenging because of the greater need of information with respect to standard approaches in assessing well-being. This could be the reason for the relatively low number of empirical applications in the context of the capability approach. Robeyns (2000) stresses several key difficulties related to theoretical and empirical applications and again underlines the lack of empirical works embracing this approach. Comin (2001) defines the concept of operationalizing Sen's capability approach and suggests possible alternatives that can be considered as operationalization strategies. Also Comin claims the absence of studies by citing the papers of Chiappero-Martinetti (2000) and Robeyns (2000).

In 2005, Kuklys wrote an insightful book whose aim was to contribute to the well-being assessment by connecting welfare economics to the capability approach literature and to understand whether the capability perspective is more informative and comprehensive than the standard approach. In this book, the novelty of the conversion function with respect to the utilization function is the inclusion of conversion factors in the analytical formulation as follow

$$b_i = f_i(c(x_i)) | z_i, z_s, z_e \tag{3}$$

where z_i, z_s and z_e are the set of individual, social and environmental conversion factors.

Subsequently, Kuklys (2005) provides a regression approach to model and measure the achievement of functionings. The statistical formulation of the conversion function is given by the so-called "functioning production function" where the achievement of functioning is subject to resources employed and a set of conversion factors. The functioning production function is given by

$$b_i = f(y_h, z_i, z_s, z_e) + \varepsilon_i \tag{4}$$

where the achieved functionings vector b_i of person i is a function of the household income y_h and the conversion factors z_i, z_s and z_e. It is important to point out that the household income is taken as a proxy for the available resources that are otherwise difficult to quantify and the conversion factors are personal, social and environmental characteristics that simply enter in the regression function as exogenous variables.

Kuklys' estimation of the functionings production function is the pioneering study in applying regression methodology to estimate the achievement of functionings by proposing a structural equation model as an alternative. Thus it provides an important contribution in the quantification and estimation of conversion process between commodities and functionings. On the other hand, although she refers to x_i as a vector of market and non-market goods and services, namely both private and public resources, she employs household income as unique proxy of resources that can be exploited in the conversion process. We can imagine that in the personal well-being assessment, the functioning achievement is subject to a wider set of resources such as goods and services that are available on the free market as well as available publicly. Household income is a reasonable proxy for all private resources. However the accessibility of public resources is independent to household income level and hence income cannot be a reliable proxy for all resources indispensable for functionings achievement.

Chiappero-Martinetti et al (2007) offer a more complex conceptual framework that explains well-being assessment generating from private and public resources. In their work, they explain how the conversion process toward functioning achievement depends to an initial asset of resources that are partially available on the market and partially are public. In line with previous studies explaining conversion processes, once again the conversion factors are considered crucial in these processes. Chiappero-Martinetti et al (2007) essentially distinguished into internal factors that are more related to personal characteristic of each person and external factors that are instead depending on the social and institutional context where each individual operates.

3.3 The economic framework

The main contribution of this study is to model and to estimate a functioning production function for the functioning "being healthy". The health functioning production function is a relation where the achievement of a good health status is explained by a set of private and public resources controlling for conversion factors, say personal, social and environmental characteristics. The estimation of this function indeed conveys the impact of these private and public resources in determining the achievement of a specific functioning given a set of exogenous characteristics.

If the estimation of this function can provide the extent to which each individual can convert resources into functionings, then disaggregating the population into specific groups can tell how much the ability of converting resources into the functioning "being healthy" varies across several groups.

In a policy maker's perspective it might be useful to know which population groups are more or less efficient in converting their available resources into functionings achievement. An example can clarify the issue. Imagine to consider "being healthy" as the selected achieved functioning and to aggregate the female population by geographical location as well as by age. We might find out that in achieving a good health status two women of the same age living in the same place differ in their ability of converting their set of resources, because the woman with a higher level of education is more efficient in converting her set of private and public goods than the other. This example is too reductive because it avoids considering other important observed determinants, but it gives a bit of flavour of the influence of this estimation.

As already said, the functioning production function refers to the utilization function introduced by Sen (1985) that reflects the way by which each individual uses commodities in order to generate functionings. However in defining the health functioning production function, some of Sen's assumptions have been dropped.

First the function that transforms commodities into characteristics is not considered. We simply suppose to take directly goods' characteristics instead of the goods themselves in order to avoid defining this function as well. The reason for dropping the fundamental assumption that individuals use goods only for the characteristics that goods embodies is a simple practical reason, although we agree that considering the function transforming goods into characteristics of goods is crucial if one wants to embrace the functioning approach rather than a hedonistic or utilitarian approach[3].

Second the problem of the choice of the functioning among a set of possible functionings is not taken into account. We define the functioning production function for the specific functioning "being healthy". The opportunity to choose functionings into a capability set is fundamental in the capability approach framework, but this study aims to measure the achievement of the health functioning instead to analyse the capability set.

[3] *Ibidem* 1.

Bearing in mind these restrictions on Sen's assumptions, we adopt Kuklys' formulation of conversion function for the health functioning rewritten as follows

$$H_i = f_i\left(x_i \middle| z_i, z_s, z_e\right) \qquad \forall f_i \subset F_i \text{ and } \forall x_i \subset X_i \tag{5}$$

where H_i is the vector of health functioning for person *i*, x_i is a generic vector of all resources that might be exploited to achieve a good health status given the conversion factors z_i, z_s, z_e.

The statistical representation of the previous conversion function is the health functioning production function given by

$$H_{ij} = f_i\left(W_{ij}, G_j \middle| z_{ij}, z_j\right) + \varepsilon_{ij} \tag{6}$$

where H_{ij} is the achievement of the health functioning for person *i* living in the geographical area *j*. This health functioning achievement is given by employing the wealth indicator W_{ij} of a person *i* living in the geographical area *j*, as a proxy for goods and services available on the market, and an index for public goods and services G_j located in the geographical area *j*. The estimation of the achievement of the functioning via private and public resources is controlled for internal conversion factors z_{ij} related to person *i* living in the geographical area j and external conversion factors z_j related to the geographical area *j*.

Formally, we model the health functioning production function following the simplification introduced on Sen's assumptions, including conversion factors as Kuklys' approach and, particularly, adding a specific variable for public resources in line with the more comprehensive conceptual framework provided by Chiappero-Martinetti et al (2007).

In health economics literature many studies define and model the individual and social determinants of health and estimate the impact of these personal, households and community characteristics on individual health.[4]

[4] In general, the literature on health economics refers to "Social Determinants of Health" SDH to identify all social and economic factors that might have an impact on health and health inequalities (Marmot and Wilkinson, 2006). In their report, Wilkinson and Marmot (2003) discuss the social gradient of health and analyze psychological and social determinants of

Other empirical studies assess the impact of public policies, public interventions and health-care utilization on health status.[5] However, to the best of our knowledge, this is the first study aiming to estimate the impact of both private and public resources on health conditions in the context of the capability approach.

A very interesting work by Martin (2006) models individual and collective resources and their impact on women's health in Morocco. This study differs to our model substantially in two assumptions. First it models the impact of public goods and services only through private resources that are represented by an assent index and the educational level attainment. Second, the capability perspective is employed only to identify education as an instrumental capacity in the conversion process of private and public resources into health.

Finally, some clarifications need to be added on the concept of conversion factors. In the already quoted Sen's (1985) book, by introducing the utilization function concept he writes that "the conversion of commodity-characteristics into personal achievements of functionings depends on a variety of factors". He sets the general outline without revealing how factors should be analytically conceptualized. Other studies we already cited generally refer to conversion factors as some personal, social and environmental characteristics which affect the conversion process between resources and functionings.

Robeyns (2005) says that "the relation between a good and the functionings to achieve certain beings and doings is influenced by three groups of conversion factors […] personal, […] social and […] environmental conversion factors". In the same line Kuklys (2005) writes that "the achievement of these functionings depends on resources at the disposal of the

---

longevity and physical health. Wagstaff (2002) reflect upon the relationship between poverty and health and analyze the possible determinants of health disparities. Healtzman et al (1994) claim the need for a broad conceptual framework in the investigation of heterogeneities in population health status and they sketch possible sources of heterogeneity. Finally, Frenk et al (1994) provide a comprehensive analysis of the determinants of health.

[5] For example, Rivera (2001) employs an ordered probit model to assess the impact of public health spending on health status using Spanish data. Earlier, Thomas et al (1996) study how health services and facilities are able to improve child health in Côte d'Ivoire.

individual, such as her income or education, as well as conversion factors, such as age, marital status and region of living".

Thus conversion factors are identified by exogenous characteristics and Kuklys (2005) has econometrically estimated the achievement of functionings considering these variables in the regression models and quantifying the impact of them on the estimated functioning.

This study wants to highlight that the focus in conversion processes has clearly to be on the role played by conversion factors not for themselves, but rather in affecting the impact of resources on functionings achievement given conversion factors. In other words, the focus has to be on the rates of conversion rather than on the conversion factors.[6]

Estimating the health functioning production function means assessing to what extent people are able to convert their resources into a good health status subject to their internal and external characteristics. Consequentially, looking at the health functioning production function will quantify the rates of conversion of private and public resources specified in the regression equation controlling for other exogenous variables identifying conversion factors.

3.3.1 Modelling Issues

The representation of conversion process between resources and functioning into an econometric estimation of the functioning production function can be viewed as comparable to estimating a reduced-form demand equation.

Ruggeri Laderchi (1999) highlights the essential advantage in adopting a reduced-form demand function by stressing that "such relation reduces responses of the household to depend only on the exogenous or predetermined variable and parameter from the point of view of the household".

Referring to Schultz (1984), it is possible either to estimate a reduced-form equation between health and its determinants that are assumed as

[6] On these aspects, see also the working paper by Chiappero-Martinetti and Salardi (2007) that aims at developing the same conceptual and methodological framework to the study of three different functionings, say "being healthy", "being educated" and "living in a safe and healthy environmental", applied to the Italian reality.

exogenous or to estimate simultaneously a demand equation for health inputs and a production function that is a relation between health outcomes and inputs. The estimation of parameters in the health production function is demanding of data since information on inputs, outputs and related instruments, namely prices are needed. For this reason, reduced-form demand functions have been applied often in the health economics literature.[7] As already said, these functions are derived from models where the household utility function is maximized subject to both the total budget constraint (including time constraint) and the health production function.

The most important and pioneering contribution on the demand and production of health has been provided by Grossman (1972). Following the traditional model of household behaviour of Becker (1965), this model proposes to maximize household utility constrained to resources consumption and time and resources allocation as well as to the best utilization of household endowment, namely economic and biological endowment. From this utility-maximization and its constraints a reduced-form demand function is derived and depends on exogenous variables, proxies of prices, income and preferences.

Since the functioning production function proposed in this study is derived from a model that estimates the health functioning with respect to individual and household characteristics as well as to monetary resources, we are able to assimilate our functioning production function to a reduced-form health demand equation.

3.4 Data and Variables Description

Our main data source is the annual Brazilian households survey, *Pesquisa Nacional por Amostra do Domicilios* (PNAD), for 2003 collected by the *Instituto Brasileiro de Geografia e Estatistica* (IBGE). The PNAD is based on

[7] Examples for studies in the economic literature that present this typology of models are Lavy et al (1996); Thomas et al (1991), Schultz (1984) and Rosenzweig and Schultz (1982, 1983).

a nationally representative random sample of households. The household survey consists of two sections, *Archivo de Domicílios* and *Archivo de Pessoas*. The first section contains information at the household level, such as characteristics of the dwellings and the geographical locations of the households. The second section provides data at the individual level, focusing more on the characteristics of household members. Particularly for 2003 the PNAD devotes an entire section to health conditions at individual level. This special section includes information on health status, presence of chronic diseases, availability of health services, health-care utilization, degrees of satisfaction in health-care provision, health insurance coverage and financial assistance in health-care utilization.

The sample used includes 128,028 Brazilians and is based on individuals aged from 10 to 65 years that have self-reported their health conditions.[8]

This Brazilian household survey has been integrated with regional data on public health services coming from the so-called *Datasus* dataset provided by the Brazilian Minister of Health.[9] The *Datasus* is a specific dataset provided by the Brazilian Government that offers geographically aggregated information related to the Brazilian public health services, the health conditions of Brazilian population and financial aspects of public health-care system.

3.4.1 The dependent variable

The functioning "being healthy" is measured by exploiting two different indicators on heath conditions.

First, we construct an index of self-reported morbidity (SRMI). This index accounts for twelve chronic diseases: vertebral column dysfunctions, arthritis and rheumatisms, cancer, diabetes, chronic bronchitis and bronchial asthma, hypertension, hearth dysfunctions, chronic kidney diseases, depression,

---

[8] It means that people whose health conditions have been reported by other respondents have been dropped from the sample. The underlining reason is to assure the reliability of the reported health conditions.

[9] Source: Ministério da Saúde - CGRH-SUS/SIRH (2006) available on the website http://w3.datasus.gov.br/datasus/datasus.php.

tuberculosis, tendinitis and cirrhosis. Moreover the extent of these chronic diseases is matched with the information on invaliding consequences that lead to inactivity. The SRMI is a dummy variable that takes value 1 if individuals suffer of one of these chronic diseases and this sufferance involves invalidity.

Second, we create an indicator of subjective health status (SHSI). We consider the question "Value your health status from your personal point of view" where the possible answers are "very good", "good", "fair", "bad" and "very bad". By aggregating these answers we construct the categorical variable SHSI that takes value 1 if the health status is considered bad or very bad, 2 if the health status is considered fair and finally 3 if the health status is considered good or very good.

Table 3.1: Frequencies for SRMI and SHSI

	SRMI		
SHSI	0	1	Total
1	3,351	1,955	5,306
2	25,854	3,721	29,575
3	92,226	1,685	93,911
Total	121,431	7,361	128,792

Table 3.1 reports the frequencies for both SRMI and SHSI. The incidence of chronic diseases has more observations where the health status is subjectively judged as bad or fair. On the other hand who is not affected by chronic disease is more likely to values her or his health conditions as fair or good.

Referring to the SRMI, 5.7% of the sample is affected by chronic and invalidating illnesses and among them 73.5% are women where women account for 65.8% of the entire sample. Individuals affected by chronic and invalidating disease are for 45.6% aged between 30 and 50 and for 37.5% aged over 50. Comparing SRMI across different level of educational attainment there is a negative relationship between chronic disease incidence and education: if 10% of individuals with primary education are affected by chronic and invalidating illnesses, only 4.5% of individuals with graduate education report the same. 56.7% of ill individuals live in the North-East and

South-East of Brazil which are the most populated regions and also the most numerous ones in our sample. Looking at the occupational levels, 68% of ill people are blue-collars while only 5.2% are professionals.

The SHSI reveals that 4.1% of the sample judges their health status as "bad" or "very bad" while 23% as "fair". The majority, say 72.9% of the sample, considers their health status "good" or "very good".

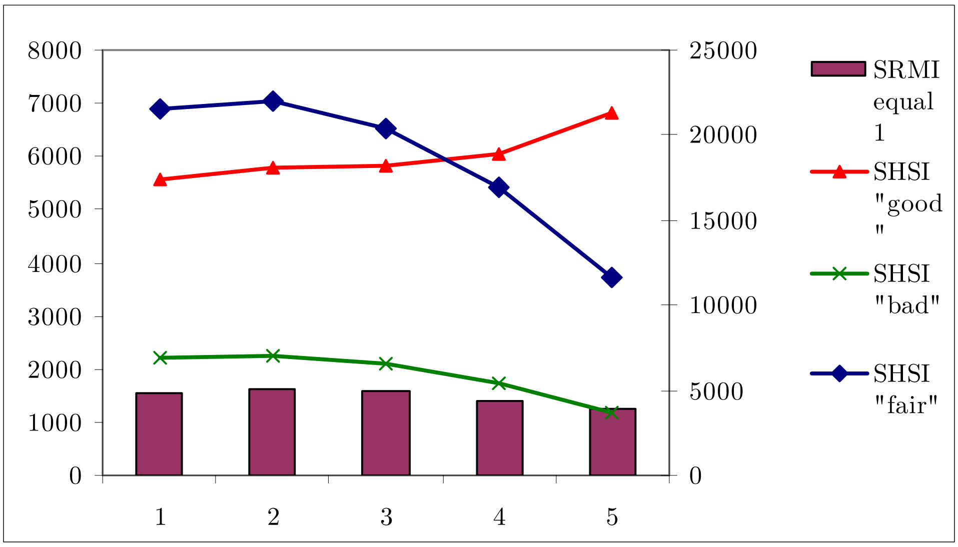
As for the SRMI, women are the majority of the sample across all three categories of the SHSI, but their relative share changes. In particular, if women that judge their health status as bad and fair count respectively for 69.6% and 71.5% of the category, their share reduces to 63.3% in the good health status category. As regarding to age groups, the incidence in the bad health status category increases moving to elderly people. Looking at the levels of attained education, individuals with good health status are likely to be more educated than individuals judging their health status as bad or very bad. Again, the distribution across three occupational levels is interesting and follows the pattern shown in the SRMI. Professionals and intermediates are the minority across all SHSI categories, but their relative shares increase moving to better health status. The relationship between occupational stratification and health indicators shows a pattern very similar to the one drawn by the educational attainment. [10]

Analyzing the distribution of both SRMI and SHSI across income quintiles, an interesting pattern emerges that can be easily understood by looking at the following graph.

In Figure 3.1, people affected by chronic and invalidating diseases for which the SRMI take value 1 are represented by the columns across quintiles. Although moving from the first to the second quintile, the self-reported morbidity index slightly raises, after the second quintile the index decrease as we move to higher income quintiles.

[10] The computations are available on request from the Author.

Figure 3.1: The incidence of SRMI and SHSI



The three categories of SHSI are instead plotted via lines. It is clear how the two lowest categories, namely “bad” and “fair”, decrease while the “good health status” category noticeably increases as soon as we move to the top of the income distribution.

The pattern of both indexes, SRMI and SHSI, with respect to income quintiles testifies a clear relationship between income and health, usually well-known as the income-health gradient.

Both SRMI and SHSI are employed in our regression analysis as dependent variables and, due to the nature of these variables, we need to apply a qualitative dependent variable regression models. In particular as we already have seen, SRMI is a dichotomous variable while SHSI has a categorical character. With regard to the category variable, the temptation is either to ignore the problem by adopting a linear regression model, but this can bring in heteroscedasticity, or to dichotomize it by setting a critical threshold upon which health conditions are judged good. The latter technique reduces the difficulties of the model since a binary model for discrete choice is more intuitive than an ordinal probability model. However the loss of information might be relevant especially in the context of the capability perspective where a dichotomist approach excludes the existence of some kind of complexity and fuzziness central for this field of research (Chiappero-Martinetti, 2004). Moreover the choice of the cut-off point above which health status might be considered good is subject to the critique of excessive arbitrariness.

Moreover, it should be reminded that the standard OLS method cannot be used unless a cardinalisation of the dependent variable is applied. Basically, the cardinalisation of a categorical variable implies the assumption that this variable is a latent variable with a standard lognormal distribution and then a score is assigned to each category.[11]

A final remark is needed on the intrinsic nature of these two variables identifying the health functioning. SRMI provides information on individual morbidity that has been self-reported by each respondent; indeed neither doctors nor professional personnel have checked for these diseases. SHSI conveys opinions given by each individual to its own health conditions. Both variables are thus subjective indicators of health and might be criticized. First, the employment of a mortality index instead of a morbidity index is preferred following the epidemiological literature (Sen, 1998), because it gives more reliable information related to the level of health and illness of a population at the aggregate level. Nevertheless the core of this study is to assess the individual health functioning achievement that exemplifies the concept of being able to be healthy. Hence indicators on morbidity or health status might be more suitable at individual level with respect to a mortality index.

Second, it might be claimed that even if adopting a morbidity index it should not be self-reported. In fact the self-reporting nature of an indicator increases the degree of subjectivity. The internal assessment of health status generally differs from the external view provided by medical experts (Sen, 2002). A person's evaluation of her or his own health situation is clearly affected by her or his social experience. People's awareness on health and

[11] An example of cardinalisation of a self-assessed health variable is provided in Wagstaff et al (2001). This study follows a previous study by Wagstaff and Van Doorslaer (1994) that proposes a methodology to construct a latent variable to overcome some limits of health studies using multiple-category morbidity indicators.

illness diverges across different social context and is highly associated to their medical understanding and the presence of health-care services systems. Sen's case study on Kerala[12] is clarifying. In India the State of Kerala reports higher rates of self-assessed morbidity in comparison to other states such as Bihar and Uttar Pradesh where there is a low life expectancy. Better health conditions in Kerala make people's perception of illness much higher and, as a consequence, the comparison of morbidity levels across these Indian states is mistaken. However the unreliability of the health assessment subsists much more in the illusion of low rates of morbidity in Bihar and Uttar Pradesh rather than in erroneous estimations in Kerala.

Following Sen's 1993 debate on the adoption of subjective or objective indicators,[13] we argue that our health indicators should be viewed as not subjective, but positionally objective indicators. If subjectivity is generally perceived as a rejection of objectivity, positional objectivity asserts objectivity of perceptions and understandings from a certain position (Sen, 1993).

The subjectivism should be kept separated from the notion of positional objectivity. Sen (1993) embraces Nagel's notion of objectivity, however he claims that "this conception of objectivity is in some tension with the inescapable positionality of observation". And he remarks that the role of positionality plays a crucial role in identifying illusions or misunderstanding in socio-economic investigations, as Kerala case study shows.

We accept this view claiming that subjective assessment can be explicated by specification of the positional constraints affecting her or his understanding. By applying this approach the demands of objectivity of values can be reinterpreted.

3.4.2 Wealth and public goods

The health functioning production function investigates the relationship between health functioning and the resources employed in the conversion

---

[12] Several Sen's publications (1993, 1998 and 2002) explain the case study on Kerala morbidity and the different problem related to the self-perception of health conditions.
[13] See Sen (1993).

process. As already mentioned, in our model we consider two main types of resources, private and public.

Following Kuklys' methodological choice, we take income as a proxy for all private resources, i.e. all commodities and services that are available on the private market. The relationship between health and income is well-known in the economics and health economics literature and it generally called "gradient" because it better exemplifies the gradual relationship existing between health status and income levels. Moving to the top of the income distribution health status is usually improving.

Strauss and Thomas (1995) provide a review on the interrelationships between health/nutrition and income/productivity by mentioning empirical works that estimate the effects of income on nutrient intakes and, conversely, how nutrition affects income and labour productivity.

The first original empirical study reporting the existence of a socio-economic gradient is the Whitehall study conducted among British male civil servants (Marmot and Shilpey, 1996).

The existence of a reverse causal relationship between health and income is well explained in an empirical work by Case (2000). In particular, Case stresses different channels through which money provides health: medical care, water and sanitation, nutrition and psychosocial stress. Deaton (2002) reports an exhaustive analysis of the gradient health-income. He claims that the gradient is affected by health-related behaviour and that it changes considering different pathologies and different access to medical care. Moreover Deaton argues that not only income, but also socioeconomic status (SES) is intimately correlated to health. Following our conceptual framework, other variables that determine socioeconomic status except income are considered personal characteristics and enter in the model as conversion factors.

In spite of the wide literature on the positive relationship between income and health, the reversal causality is subject to controversy. Nevertheless the reverse causal relationship between health and income might cause endogeneity problem in our regression model. The application of a two-stage procedure helps in overcoming this problem. In her study of 1996, Ettner estimates the impact of income on health status both with ordinary and instrumental variables (IV) estimates. Conversely it seems difficult to find the right instruments where the residuals are not correlated to the health variable. For this reason we decide to construct a long-run indicator of wealth to substitute the income variable because a long-run wealth index is less exposed to reversal causality with health conditions.[14]

The wealth indicator has been constructing using the principal components analysis.[15] In order to construct the wealth indicator we exploit variables regarding to housing characteristics, facilities access and durables ownership. Table 3.2 reports the scoring factors from the principal components analysis that are used to compute the wealth indicator.

Table 3.2: Scoring factors and summary statistics for variables entering in the computation of the first principal component for computing the wealth indicator

	Scoring factors	Mean	SD
Having good walls	0,0478	0,8696	0,3367
Number of Rooms	0,0784	5,8020	2,1995
Number of Bedrooms	0,0272	2,0900	0,6035
Garage	0,0825	0,4588	0,4983
House property	0,0825	0,6957	0,4601
Piped water	0,1217	0,8708	0,3355
Well water	-0,0837	0,0510	0,2204
Flush toilet	0,1084	0,9340	0,2483
Garbage collection	0,0968	0,7480	0,4342
Electricity as energy source	0,0940	0,9655	0,1825
Gas as energy source	0,1197	0,8876	0,3159
Wood as energy source	-0,0994	0,8336	0,2764
Coal as energy source	-0,0469	0,0171	0,1295

[14] Martin (2006) also adopt a wealth indicator to solve the endogeneity problem in estimating their health production function. They also underline that the introduction of a morbidity variable as covariate allow them to partially control for endogeneity.

[15] To apply the principal component analysis in constructing the wealth indicator we refer to the relevant literature on this topic, such as Filmer and Pritchett (2001), Montgomery (2003), Montgomery et al. (2000) and Sahn and Stifel (2000, 2003).

Kitchen (one cooker)	-0,0685	0,0139	0,1171
Kitchen (more cookers)	0,0785	0,9743	0,1582
Own Telephone	0,1092	0,4745	0,4994
Own Water Filter	0,0398	0,5053	0,5000
Own Radio	0,0525	0,8705	0,3357
Own Colour Television	0,1169	0,8520	0,3551
Own Black/white Television	-0,0568	0,0457	0,2089
Own Fridge	0,1196	0,8573	0,3498
Own Freezer	0,0568	0,1707	0,3762
Own Washing Machine	0,0901	0,3004	0,4584
Own Computer	0,0796	0,1378	0,3447
Own Internet Access	0,0734	0,1010	0,3014

Each scoring factor gives its contribution in determining the wealth indicator. The check for the robustness of the wealth index constructed by using principal components procedure can be done by comparing this index with another one constructed using a different procedure for deriving weights. We obtain a 0.9931 Spearman rank correlation between our wealth indicator and a similar one developed by applying factor analysis. This result conveys that the constructed wealth indicator is robust. Finally in order to ensure that the wealth indicator can substitute the income variable in our regression analysis we compute the Spearman rank correlation between wealth and income: the values of 0. 6372 is good in comparison with the results in the relevant literature (Sahn & Stifel, 2003).

The use of a wealth indicator as a proxy for private goods instead of income might be viewed as a more comprehensive and appropriate variable because it is a long-run indicator and embodies more information that is able to determine health conditions. Furthermore we think the reversal causality is weaker between health and wealth than health and income since detrimental health conditions are more likely to affect income levels in the short-run rather than long-run wealth.

Public resources are the second type of resources we consider in the conversion process to health functioning achievement. As specified above, data referring to public goods and services at local level are drawn from *Datasus* dataset. We decide to consider the number of doctors, nurses and hospital beds available at local level plus the per capita public expenditure in health-care imputed by geographical area. We aggregate these four variables by constructing an indicator of availability of public resources via principal component analysis as shown in table 3.3. The constructed variable representing public resources has a geographical variability and has been merged with the individual dataset by adopting a geographical criterion.

Table 3.3: Scoring factors and summary statistics for variables entering in the computation of the first principal component for computing the public resources index

	Scoring factors	Mean	SD
Number of Doctors	0,3686	1,2697	0,6847
Number of Nurses	0,3711	0,5419	0,2094
Number of Beds	0,1483	0,8612	0,3142
Per capita public expenditure in health care	0,3902	268,0110	61,2206

The main purpose of modelling the conversion process from resources to functioning is to estimate the impact that the wealth indicator and the public resources index have on health conditions controlling for conversion factors. Referring to equation (6) the conversion process is conceptualized as a production function, where these two variables W_{ij} and G_j enter into the conversion processes as production factors subject to z_{ij} and z_j. We might be interested not only in how these factors singularly contribute to the conversion process, but also in the effect of the interaction of these resources.
The individual impact of the wealth indicator as well as the public resources index in achieving health functioning can be shown through a simple mathematical expression as follow

$$\frac{\partial H_{ij}}{\partial W_{ij}} = f\left(\partial z_{ij}, \partial z_j, \varepsilon_{ij}\right) \tag{7}$$

$$\text{and } \frac{\partial H_{ij}}{\partial G_j} = f\left(\partial z_{ij}, \partial z_j, \varepsilon_{ij}\right). \tag{8}$$

In equations (7) and (8), the impact of the wealth and public goods is given by the first derivatives with respect to these variables where the function is a function of conversion factors as well.

In order to investigate the interactions occurring among private and public resources, second-order derivatives provide the joint impact of these resources. Hence the sign of these second-order derivates conveys in which relationship these resources jointly determine the health functioning achievement. If the first derivative is positive and

$$\frac{\partial H_{ij}^2}{\partial W_{ij} \partial G_j} > 0 \quad \text{then the private and public resources are complements;} \tag{9}$$

$$\frac{\partial H_{ij}^2}{\partial W_{ij} \partial G_j} < 0 \quad \text{then the private and public resources are substitute.} \tag{10}$$

When the first derivative is negative, the reverse is true. Hence private and public resources are complements if the second derivative is negative and they are substitute if the second derivative is positive.

3.4.3 Individual characteristics

The estimation of the health functioning production function aims to quantify the impact of private and public resources in the functioning achievement subject to so-called conversion factors.

Conversion factors are individual, social and environmental characteristics that unavoidably enter into the conversion process. Indeed the specification of the set of these characteristic is clearly crucial.

In our model, we identify two sets of conversion factors, z_{ij} and z_j. As we have already explained, the first set of internal conversion factors consists of characteristics for individual *i* living in the *j-th* geographical area, while the set of external conversion factors is a group of community characteristics of the *j-th* geographical area.

In order to identify individual and community characteristics, we refer to previous studies in health economics aiming to classify the determinants of health outcomes.

Frenk et al (1994) provide a clear diagram where health status is affected by proximate, structural and basic determinants. Basic determinants have a systemic character and refer to population genome, environment and social organization. Structural determinants have a more societal attribute and look at the level of wealth, social stratification and occupational structure as well as the redistribution mechanisms. Proximate determinants are institutional or household factors that directly affect health status such as working and living conditions, the health care system as well as individual life-style.

Hertzman et al (1994) stress the importance of a comprehensive framework to analyze health outcomes and reject the analysis of the health of a population only explained by individual characteristics. They highlight that the heterogeneity in health conditions depends on life cycle stages, individual characteristics and other sources of heterogeneity. The individual characteristics involve socioeconomic status, ethnicity, migration status, geography and gender. Other sources of heterogeneity might be the individual life-style, physical and social environmental and differences in access to health care services.

Referring to Wagstaff (2002), the main determinants of health outcomes are grouped into three groups: households and community factors, health system and government policies. In particular the households and community factors are household actions and risk factors, such as utilization of health services, sanitary, sexual practices, dietary and lifestyle, household assets, namely human, physical and financial, and community factors like social capital, environment, infrastructure, cultural norms and community institutions.

Bearing in mind all possible determinants of health status, at this stage of our empirical analysis we consider only individual characteristics, particularly personal characteristics such as gender, race or education, labour market characteristics and geographical characteristics.

Table 3.4: Summary statistics for individual characteristics

Variable	Mean or Percentage in the Category	Std. Dev
Personal characteristics		
Male	0.3409	0.474
White	0.4687	0.499
Age group:[(a)]		
Mature people	38.8835	5.6640
Elderly people	56.6904	4.5828
Educational attainment:[(b)]		
Primary school	0.1538	0.3607
Secondary school	0.549	0.4975
College	0.0665	0.2492
Post-graduate	0.0035	0.0597
Labour market characteristics		
Farmer	0.1198	0.3248
Occupational level:[(c)]		
Intermediate	0.2882	0.4529
Blue collar	0.6454	0.4783
Formal sector	0.1968	0.3976
Geographical characteristics		
Region:[(d)]		
North-East	0.3278	0.4694
South-East	0.2872	0.4524
South	0.1598	0.3665
Central-West	0.1177	0.3222
Urban	0.8468	0.3601
Brasilia	0.0282	0.1657
São Paulo	0.1087	0.3113
Roraima	0.0044	0.0667
Acre	0.0071	0.0841

(a) For the category variable Age group, the base category are your people;
(b) For the category variable Educational attainment, the base category is illiterate;
(c) For the category variable Occupational level, the base category is professional/technician;
(d) For the category variable Region, the base category is the North.

Table 3.4 reports summary statistics for individual characteristics that were selected and employed in our regression analysis. Personal characteristics embrace *male*, *white*, *age group* and *educational attainment*. The majority of our sample is female and black. The black category is the majority because it also covers brown people and mulattos. The sample considers only people aged from 10 to 65 because children generally do not report their health status by themselves and this is also true for very elderly people. Moreover at the tails of age distribution is more likely to find outliers. The population has been divided in three age groups: your people aged from 10 to 29, mature people aged from 30 to 49 and elderly people aged from 50 to 65. We decide to adopt the maximum level of educational attainment instead of the years of education because it is usually considered a more informative variable on the real level of education achieved.

In the set of personal characteristics we also include a selection of variables related to the labour market. In particular *farmer* identifies people who work as farmers; and *occupational level* groups individuals into three categories: professional/technician, intermediate and blue collar. The variable *formal* classifies people as working in the formal sector when they own a work card.[16]

Finally with the geographical characteristics we control for geographical differences in health status and health provision. It is important to remind that Brazil is a country with huge geographical disparities. We control for *region*, where Brazil is divided into five regions: North, North-East, South-East, South and Central-West. The dummy variable *urban* identifies people who lives in an urban area where wealth tends to be higher and health care provision better. Finally, some metropolises showing particular trends[17] with respect to the wealth indicator and the public resources index are added into the geographical controls.

3.5 Econometric methodologies

The econometric estimation methodology depends on the distribution of the indicator adopted.[18] The self-assessed morbidity index is estimated as a

[16] The possession of the work card guarantees legal rights through labour legislation. Hence the definition of formal and informal sector used to construct this dummy variable refers to the state regulation of work as indicated by social security payment.
[17] Brasilia and São Paulo have very high levels for the wealth indicator, while Roraima and Acre are very poor cities placed in the North region. They show particular bad performances in term of health provision.
[18] See Maddala (2001).

probit model, while the subjective health status indicator as an ordered probit model.

In the probit model, the binary dependent variable y_i is replaced by a latent continuous dependent variable y_i^* such that if $y_i^* \geq 0$ then $y_i = 1$ and if $y_i^* \leq 0$ then $y_i = 0$. In other words, in the first case the event occurs, while in the latter not. We assume the following regression model in matrix form

$$y_i^* = x_i^{'}\beta + u_i \qquad \text{with } i=1,\ldots,n \tag{11}$$

where $u_i \approx N(0, \sigma^2)$ and $y_i^* \approx N(x_i^{'}\beta, \sigma^2)$.

Then, the probability that the event occurs is

$$prob[y_i = 1] = prob[y_i^* \geq 0] = prob\left[\frac{u_i}{\sigma} \leq \frac{x_i^{'}\beta}{\sigma}\right] \tag{12}$$

Equation (12) shows the probability that the cumulated probabilities from -∞ to the point delineated by $\frac{x_i^{'}\beta}{\sigma}$. We can rewrite equation (12) as follow

$$prob(y_i = 1) = \Phi(x_i^{'}\beta) \tag{13}$$

where $\Phi(\cdot)$ is the cumulative distribution function for a standard normal random variable.

In order to interpret the regressor's impact on the probability of an event occurring, we need to compute marginal effects if the regressor is a continuous variable or impact effects if the regressor is a binary variable.

Instead of using the matrix expression of the index, we use the following simple expression

$$x_i^{*}\beta = \alpha + \beta X_i + \delta D_i \tag{14}$$

where the index contains a constant term, a continuous regressor X_i and a dummy variable D_i. We can express the model as follow

$$prob[y_i = 1] = P_i = \Phi(\alpha + \beta X_i + \delta D_i). \tag{15}$$

The marginal effect is then given by

$$\frac{\partial P}{\partial X_i} = \phi(\alpha + \beta X_i + \delta D_i) \times \beta \tag{16}$$

The impact effect is given by

$$\Delta = \Phi(\alpha + \beta X_i + \delta) - \Phi(\alpha + \beta X_i). \tag{17}$$

The ordered probit model is an extension to the binary probit model that provides a way of modelling ordered discrete data. We express again the model following equation (14). In this model, the latent continuous dependent variable y_i^* replaces the ordinal variable in the following way:

$$y_i = \begin{cases} 1 & if \quad y_i^* \leq \theta_1 \\ 2 & if \quad \theta_1 < y_i^* \leq \theta_2 \\ . & \\ . & \\ M & if \quad \theta_{M-1} \leq y_i^* \end{cases} \tag{18}$$

where M represents the number of alternatives where j=1,…,m and θ_j are the cut-off points between alternatives.

Then, the probability of observing y_i is given by

$$\begin{aligned} & prob[y_i = j] = prob[\theta_{j-1} < \alpha + \beta X_i + \delta D_i + u_{ii} \leq \theta_j] \\ & = \Phi[\theta_j - (\alpha + \beta X_i + \delta D_i)] - \Phi[\theta_{j-1} - (\alpha + \beta X_i + \delta D_i)] \end{aligned} \tag{19}$$

where $\Phi(\cdot)$ has a normal distribution.

If the ordered dependent variable has three categories, marginal effects are computed as follow

$$\frac{\partial prob\,[y_i = 1]}{\partial X_i} = -\phi(\alpha + \beta X_i + \delta D_i) \times \beta \tag{20}$$

$$\frac{\partial prob[y_i = 2]}{\partial X_i} = \phi(-(\alpha + \beta X_i + \delta D_i)) \times \beta - \phi(\theta_1 - (\alpha + \beta X_i + \delta D_i)) \times \beta \tag{21}$$

$$\frac{\partial prob[y_i = 3]}{\partial X_i} = \phi(\theta_1 - (\alpha + \beta X_i + \delta D_i)) \times \beta \tag{22}$$

Finally, the impact effects are given by

$$\Delta = prob[y_i = j | D = 1] - prob[y_i = j | D = 0]. \tag{23}$$

3.6 Empirical results

The estimations of the health functioning production function are provided in Tables 3.5 and 3.6. The econometric model implemented depends on the nature of the dependent variable. When the self-reported morbidity index (SRMI) is the variable used as health functioning, due to the binary nature of this variable, the health functioning achievement is estimated as a probit model. When the health functioning is specified by the indicator of subjective health status (SHSI), the ordinal categorical nature of the health status imposes the utilization of an ordered probit model.

Table 3.5 in the appendix shows the marginal effects from the probit estimates employing four different models. In particular, the first model estimates only the impact of the private resources, identified by the wealth indicator, on the health functioning achievement controlling for a set of individual characteristics The second model provides only the impact of the public resources, namely the public resources index, again controlling for a set of individual characteristics. The third model considers simultaneously private and public resources and estimates the conversion of these two resources into health functioning achievement. Finally, the fourth model is the most comprehensive because it considers not only both resources, but also the interaction between them.

The same model specifications, except for the last one, are applied for the ordered probit model and results are presented in Table 3.6.[19]

Some clarifications might be useful in interpreting the results of our estimations. First, we adopt two different variables for identifying the dependent variable. SRMI takes value 1 when the respondent is affected by an invalidating and chronic illness and 0 if not. Hence the estimates of the resources-functioning conversion process should be negative: the more we employ private and public resources in the converting process, it should be less likely to get an invalidating and chronic disease. On the other hand, SHSI evaluates health status from 1 to 3 and the best health status is associated

[19] Marginal effects from the ordered probit models are provided by the Author on request.

with the higher categorical value, i.e. when SHSI takes value 3. In this case the estimated conversion rates from the health functioning achievement process should be positive.

Second, the impact of private and public resources in the achievement of functioning is given by the estimated coefficients of the wealth indicator and the public resources index. The interpretation of these coefficients is not straightforward as in the case of the individual characteristics used as controls that are binary or continuous variables.

In fact the wealth indicator and the public resources index are variables constructed using the principal component analysis. The procedure involves scoring the factors and retaining the first score as the latent common factor. Both the wealth indicator and the public resources index are expressed in terms of standard deviation and consequently in order to assess the impact of these variables we should consider the effect of one standard deviation increase of these two indexes.

In order to be able to interpret these impacts in a more intuitive way, we employ the five quantiles of the wealth index distribution instead of the wealth index itself. Thus we can directly understand the effect of being in a specific part of the wealth index distribution on the dependent variable. For the public resources index we construct a dummy variable that takes value 1 when these index shows a value that lies in the highest fifth quantile of its distribution, say when people benefit of the highest level of public resources in the health sector.

Looking at Table 3.5 we can see that the impact of private and public resources in reducing the probability of contracting an invalidating and chronic illness is statistically significant across four different models.

When we take private or public resources separately, public resources seem to have a greater impact: having a good access to public resources decreases the probability of getting ill by 2.4 percentage points, *ceteris paribus*. The impact of the private resources is greater as soon as we move to the highest quantiles of the wealth index distribution. Indeed, being in the second quantile of the wealth distribution decreases the probability of getting ill by 0.7 percentage points with respect to the first quantile, while in the highest quantile the probability diminishes on average by 2.1 percentage points, *ceteris paribus*. Once both resources are simultaneously considered in the regression, their effects do not vary significantly. The fourth model is the most complete because it adds the interaction terms between wealth quantiles and the dummy variable for the good level of public resources. These interaction terms are crucial in order to understand whether private and public resources are substitutes or complements in the health functioning achievement process. Giving the negative sign of the first derivative for both resources, the negative interaction terms allow us to infer that private and public resources are complements in reducing the probability of contracting an invalidating and chronic disease. In particular, public resources have a greater impact than private resources in reducing the probability of contracting invalidating and chronic diseases at the bottom of the wealth index distribution, i.e. for the least wealthy part of the population.

Personal, labour market and geographical characteristics enter into our model as individual conversion factors affecting the resource-functioning conversion process. Nonetheless it is interesting to analyze how these variables influence the process. Since the impact of these variables is similar across model specifications, we comment on results from the last and most complete model specification from Table 3.5. Considering personal characteristics we find that being male decreases the probability of getting ill by 1.9 percentage points, while being white increases the probability by 0.3, *ceteris paribus*. As a consequence, we can infer that white people are more likely to report invalidating and chronic diseases although it is less clear whether whites are actually more likely to contracting illness. Age is robustly statistically significant and increases the probability of getting ill. Being elderly increases the probability by 9 percentage points, while for mature people the probability increases by 4.2 percentage point taking young people as reference group. The maximum attained educational level is an interesting variable. Only secondary school and college are statistically significant and decrease the probability of getting ill by respectively 0.8 and 0.9 percentage points, *ceteris paribus*. It means that primary school is not sufficient to acquire those standards of living and life-styles able to prevent invalidating and chronic diseases. On the other hand, postgraduate education is no more functional than college education, since it is not statistically significant.

Among labour market characteristics, being a blue collar worker is statistically significant and increases the probability of contracting illness by 0.6 percentage points with respect to working as a professional or technician, *ceteris paribus*.

The geographical characteristics involve dummy variables for regions, for living in an urban area and four specific dummies for the "Unidade de Federação" of Brasilia, São Paulo, Roraima and Acre. Living in an urban area raises the probability of being chronically ill by 1.6 percentage points, *ceteris paribus*. Regional dummies are not statistically significant or particularly informative, while the dummies for some "Unidade de Federação" are interesting. Living in some of these geographical areas increases the probability of getting invalidating and chronic diseases, in particular by 3.5 percentage points in Brasilia as well as in Roraima and 7.6 percentage points in Acre, *ceteris paribus*. On the other hand, in São Paolo the probability shrinks by 0.6 percentage points, *ceteris paribus*.

Table 3.6 illustrates the estimated impact of private and public resources on SHSI controlling for the same set of individual characteristics and with the same model specifications as that of the probit estimations, but employing an ordered probit model due to the ordered categorical nature of the health status variable. As said earlier, the model with interaction terms is not applied due to their statistical insignificance.

The analysis of ordered probit estimations is less intuitive than a binary model and to quantify the impact of each covariate we should refer to the marginal effects.

Nonetheless by looking at the estimated coefficients in Table 3.6 we find some interesting patterns. The impact of private and public resources is strongly statistically significant across different model specifications. On the contrary of estimated coefficients for SRMI, the private resources have a greater impact on health status than public resources. In fact the impact of having good public service is greater only than the impact of the second and the third quantile of wealth taking the first quantile as reference group. It means that at the top of wealth distribution private resources are more effective than public resources in increasing the probability of having a good health status.

The effects of individual characteristics are analyzed by considering only the third model specification provided in Table 3.6.

Generally speaking the estimated coefficients for individual characteristics from ordered probit models are all in line with the probit estimates. Male individuals are more likely to judge their health status as good or very good. White people are more likely to judge their health status as bad or very bad. Age is negatively associated with good health status.

All categories related to the maximum attained educational level with illiterate people as the reference category are statically significant. Having a college degree has the greatest impact on the probability of having a good health status followed by the postgraduate degree and the secondary school degree. Having attended primary school affects negatively the achievement of a good health status with respect to being illiterate as a reference group. Although this result could appear atypical, apparent better health conditions of illiterate people compared to people who attend primary education might reflect a lack of awareness by illiterate people in reporting their health status.

Moving to labour market conditions, being a farmer increases the probability of reporting good health conditions. The previous remark referring to those who have attended only primary school can help in interpreting the estimated coefficient of this dummy variable. In fact, this might mean that farmers are less likely to report bad health status rather than being effectively healthier than people working in other economic sectors.

Looking at the occupation levels, intermediates and blue collars are less likely to report good health status with respect to professionals. The last labour market characteristic, namely *formal*, tells us that working in the formal sector, i.e. owing a working card, increases the probability of having good health status probably due to better guaranteed working conditions.

Finally, with regard to geographical characteristics, the South region seems to be the region where individuals are more likely to report better health status. Living in an urban area decreases the probability of having good health and in particular living in the districts of Roraima and Acre has the worst impact. Again this result is in line with the ones obtained from the previous probit analysis. The fact that people living in metropolitan areas are likely to report worse health conditions might be due to a more conscious perception of their health and, more in general, to a greater awareness of the health-care system as we have already explained for *primary school* and *farmer* variables.

3.6.1 Aggregating by race

Tables 3.5 and 3.6 provide probit and ordered probit estimates considering the entire Brazilian sample. We have already highlighted that the main purpose of this econometric analysis is to assess the health functioning achievement in order to understand to what extent individuals are able to convert private and public resources into health functioning achievement.

We might also be interested in understanding how this ability to convert resources into functioning might vary across population sub-groups. Policy makers might be interested in understanding which population sub-groups are more "efficient" in converting their available resources and which ones are more vulnerable and which factors might affect the conversion process more.

To do that, our Brazilian sample has been aggregated into four different population sub-groups by gender and race: white women, white men, black women and black men. Before proceeding with the estimation, we check whether the model allows for an intercept shift for gender and race but not other gender-race effects. In other words, we test whether the separation by gender-race is supported by our data and we conclude that there are some gender-race differentials in the effect of covariates on the two dependent variables.

Tables 3.7 and 3.8 again in the appendix present results of the probit and ordered probit estimations across the four population sub-groups by employing the last model specification of both regression models.[20]

For probit estimates, Table 3.7 provides the marginal effects.

White women show a statistically significant impact of private resources only at the top end of the wealth distribution, while among black women all wealth quantiles have a statistically significant impact on decreasing the probability of getting ill.

White men show the highest impact of wealth in reducing the probability of getting ill across wealth quantiles. For both white women and men, being in the highest wealth quantile decreases the probability of getting ill by 1.9 percentage points, *ceteris paribus*.

At the bottom of wealth distribution black people have a higher impact of wealth on the probability of getting ill, in particular black women show a higher impact than black men across all wealth quantiles.

The impact of having good public resources is statistically significant only for black women and decreases the probability of getting ill by 2.9 percentage points, ceteris paribus.

If for black women having good public resources show an intercept shift and does not permit for other effects given by interaction terms, for black men interaction terms are statistically significant and it means that they benefit by the interaction of public and private resources.

Among white people public resources affect only women through their interaction with the highest quantile of wealth, while men do not show any impact of public resources.

Ordered probit estimates provided in Table 3.8 show the crucial role played by the private resources in improving the health status.[21] Across all population

[20] For the probit regression Table 3.7 already reports the marginal effects. The model specification adopted for the probit model across population sub-groups is the one that takes into account both private and public resources and their interaction terms, while for the ordered probit model it is the one that considers both private and public resources but not their interaction terms.

[21] To do that we compare the marginal effects of ordered probit estimations with the marginal effects of probit estimates provided with table 3.7. In this work we show only ordered probit

sub-groups wealth quantiles have a greater impact in comparison with having good public resources except for the lowest quantile.

Amongst women, black women show a greater impact of private resources than white women in lowest wealth quantiles while white women perform better at the top of the wealth distribution. In general, moving to the highest quantile of wealth distribution across all regression estimations white people show greater effect of private resources than black people.

Similarly to the results of the previous table, white men benefit from the highest impact of private resources in reaching a good health status, but not from the access to good public resources. In fact black people show a higher impact of having good public resources in improving their health status than white people.

Again, we infer that if white people benefit from a greater impact of private resources, black people show a greater effect in having good public resources. Across race black people perform better in lower wealth quantiles while white people in higher ones. Finally, women seem to benefit less from private resources, although it is not true at the top of wealth distribution especially among white people.

3.7 Final remarks and conclusions

Our probit and ordered probit regression estimations provide interesting patterns about the ability of the Brazilian population in converting private and public resources into the achievement of the health functioning.

When the self-reported morbidity index (SRMI) is employed, public resources seem to have a greater impact than private resources in reducing the probability of contracting an invalidating and chronic disease. Once interactions between private and public resources are added, the effect of private resources in the process is strengthened by the role played by public

estimates and we decide to omit tables with the related marginal effects due to the unnecessary amount of information.

resources. The interaction terms also tell us that these resources are complementary in achieving health functioning.

When examining the role played by conversion factors, among personal characteristics, we notice that men are less likely in getting ill and that whites generally display higher probability of getting ill. Age obviously increases the probability as well. Achieving a college degree seems to be a fundamental determinant in lowering the probability of contracting an invalidating and chronic disease. In fact, working as blue collar increases the probability of getting ill. Among geographical characteristics, living in an urban area has the greatest effect on the probability of poor health. The fact that the urban population is subject to more illnesses than rural people might not be entirely true. As we have already highlighted, the urban population might be more aware of their own health conditions as a consequence of living in an environment where health-care provision is dispensed more.

The utilization of the subjective indicator of health status (SHSI) gives different results. Private resources have a more relevant role in achieving health functioning with respect to public resources. When the health functioning is measured with health status rather than a morbidity index, the strong positive relationship between wealth and health status is even more clearly noticeable. Looking at the personal, labour market and geographical characteristics we control for, a pattern similar to the one for SRMI emerges. In particular, we want to focus on two noticeable differences: the negative impact of having a primary school education compared to being illiterate and the positive effect of being a farmer. Both cases might be misinterpreted. It is difficult to believe that illiterate people are effectively healthier than Brazilians who have attended primary school or that farmers are in better health than the urban population. It is easier to accept that the illiterate population and those who live in rural areas are less informed about health and health services and, consequentially, have a different perception about their health conditions. The perception of illness varies with what people experience and with their knowledge about health and medical provision. As for the urban variable with SRMI, the evaluation of their own health status depends on the Brazilian population's understanding of health.

Having analyzed the resources-functioning process for the entire sample, we estimate the effect of private and public resources by aggregating the Brazilian population into four sub-groups.

We employ again both dependent variables, namely SMRI and SHSI. With the self-reported morbidity index and considering white people, men are more efficient in converting private resources than women. In particular, for white women private resources have an impact in lowering the probability to get ill only in the highest quantile. White men are also more efficient than black men. In general, at the top of wealth distribution white people are more efficient than black people.

With regard to having good public resources, when statistically significant they have greater impact than private resources in health functioning achievement. Black people are more efficient in converting public resources in lowering the probability to get ill. Among white people, only white women are able to convert public resources, but exclusively at the top of the wealth distribution.

As highlighted by the results for the entire sample, the use of the subjective health status indicator highlights the significant impact of private resources on the health functioning. Across all population sub-groups private resources have a greater effect than public resources.

White men are again the most efficient group in converting private resources in health functioning achievement, but at the top of wealth distribution white women show a greater impact of private resources in achieving a good health status. Across both races, black people are more efficient in lower wealth quantiles while white people are more efficient as we move to the top of the wealth distribution. Public resources show greater impact again for black people than for white people.

To summarize, by identifying the health functioning with the self-reported morbidity index, public resources are more crucial in the health functioning achievement process. White people are the least efficient in using public resources. On the other hand, when the health status indicator is used to identify the health functioning, the role played by private resources becomes predominant. White men are generally the most efficient in employing their private resources in order to achieve better health conditions.

These econometric estimations of the health functioning production function aim to assess the extent to which Brazilians are able to convert a set of private and public resources into the health functioning, controlling for individual characteristics. Moreover, we think that the definition of population sub-groups and the estimation of conversion processes for each sub-group might be of considerable interest for policy making because it helps in identifying population categories that are more or less efficient in exploiting private and public resources.

Looking at our empirical results, black people might be considered a vulnerable group. The Brazilian policy maker should protect this part of the population that records the lower ability into converting their private resources and good efficiency in using public resources. Possible directions of intervention might be to promote black-targeted public provision of medical assistance and prevention considering that private resources of black people are on average more limited. Another interesting result that might affect policy makers is the fact that across race, women record a greater impact of public resources while for men private resources are more relevant. A possible explanation might be the weaker power of the women in managing private resources of the household that pushes women in exploiting more efficiently public services. Indeed the public health services should be aware of the fact that the highest portion of its policyholders is female and thus the creation of more female-centric policies may help to most efficiently improve health functioning.

Modelling and quantifying the resources-functioning conversion process is the main purpose of this paper. With our empirical analysis we want to focus on the conversion process not to define and estimate the variable identifying the health functioning, but to assess the conversion process for itself giving the health functioning, the private and public resources and the conversion factors, i.e. personal, labour market and geographical characteristics.

Little has been done in order to operationalize the capability approach and this study might be considered a contribution to assessing individual well-being in the Senian context of capabilities and functionings.

We want to conclude by listing some fundamental remarks that need to be solved in order to forward the operationalization of the capability approach.

First, the definition of the variable that can best identify the functioning is important, but problematic. In our paper we analyze the functioning "being healthy" and we adopt two different variables to identify this functioning: a morbidity index and an indicator of health status. Furthermore the investigation should go deeper and handle the definition and measurement of other functionings, such as "being educated" or "living in a safe and healthy environment". Nevertheless the lack of statistical data constrains empirical applications of well-being assessment that wish to employ the concepts of capabilities and functionings in their analysis.

Second, our functioning production function conceives the functioning achievement as a production of the health functioning where private and public resources are the main resources that identify production factors. However the definition of which type of resources can be considered in the model is open to discussion. We take the wealth indicator as proxy of income where income is a proxy of all goods freely acquirable from the market. Martin (2006) considers not only a long-term indicator of wealth, but also education as resources that can be employed in the well-being production.

Thirdly, the resources-functioning conversion process is controlled by a set of individual characteristics that we have called internal conversion factors. We consider several characteristics, namely personal, labour market and geographical characteristics, but the extension of the set of conversion factors we control for is a needed step toward a more precise estimation of the conversion process. Although we try to classify exogenous characteristics that might affect the functionings achievement, there are several factors not easy to quantify or to add into regression equations, such as genetic background.

Moreover there are differences in norms and expectations that affect the functioning "being healthy" and related to self-reported and subjective indicator of health status that are ignored.[22] Generally speaking, the problem of the omitted variables tends to overestimate the model.

Finally, the estimation of the health production function has been made by employing a probit and an ordered probit regression model. The potential endogeneity problem related to the reversal causality existing between health and income has been partially overcome by substituting income with a long-term indicator of wealth, the wealth indicator. However, using a two-stage instrumental variables estimation might be more consistent. Ettner (1996) estimates the effect of income on self-assessed health status by applying both ordinary and IV estimates. She highlights that this method is reliable as long as the instruments for income are valid. She uses unemployment rate, work experience, parental education and spouse characteristics as potential instruments for household income. We question, however, whether or not these are valid instruments and if IV estimation procedure is able to control for endogeneity problem better than using a long-term indicator of wealth instead of income and, hence, if it is judged more appropriate.

A study of the identification of the variables, the definition of the model and the improvement of the econometric strategies as well as to explore different functioning and their interactions in order to assess individual well-being in the context of the capability approach would be the major contribution to the existing literature.

[22] Hildebrand and Van Kerm (2005) remark that the problems related to the omitted variable and to the differences in norms and expectations are partially controlled by the adoption of panel data since it control for the effects of unobservable fixed effects in the income-health relationship.

References 3

Becker, G. S., (1965), "A Theory of the Allocation of Time", *The Economic Journal*, Vol. 75, No. 299, pp. 493-517.

Case, A., (2002), "Health, Income, and Economic Development", *Annual World Bank Conference on Development Economics, 2001/2002,* pp. 221-241.

Chiappero-Martinetti, E., (2000), "A Multidimensional assessment of well-being based on Sen's Functioning Approach", *Rivista Internazionale di Scienze Sociali*, No. 2.

Chiappero-Martinetti, E., (2007), "Complexity and vagueness in the capability approach: strengths or weaknesses?", in S. Alkire, F. Comin, M. Qizilbalsh, (Eds.), The capability approach in human development: concepts, applications and measurement, Cambridge: CUP.

Chiappero-Martinetti, E., M. Grasso, and S. Pareglio, (2007), "La qualità della vita nelle società avanzate: linee di ricerca per un modello basato sull'approccio delle capacità di Amartya Sen", *Economia e Politica Industriale*, Vol. 34, No. 1.

Chiappero-Martinetti, E., P. Salardi, (2007), "Well-being process and conversion factors: an estimation of the micro-side of the well-being process", *HDCP-IRC Working paper*, Paper presented at the 2007 International Conference of the Capability Approach "Ideas Changing History", September 17-20, New York City.

Comin, F., (2001), "Operationalizing Sen's Capability Approach", paper prepared for the Conference Justice and Poverty: examining Sen's Capability Approach, Cambridge 5-7 June 2001.

Deaton, A., (2002), "Policy Implications Of The Gradient Of Health And Wealth", *Health Affairs*, Vol. 21, No.2.

Ettner, S. L., (1996), "New Evidence on the relationship between income and health", *Journal of Health Economics*, Vol. 15, pp. 67-85.

Filmer, D., and L. H. Pritchett, (2001), “Estimating Wealth Effect without Expenditure Data or Tears: An Application to Educational Enrolments in State of India, *Demography*, Vol. 38, No. 1, pp. 115-132.

Frenk, J., J. L. Bobadilla, C. Stern, T. Frejka, and R. Lozano, (1994), “Elements for a theory of the health transition”, in: L. C. Chen, A. Kleinman, N. C. Ware (Eds.), Health and social change in international perspective, Harvard Series on population and international health, Harvard: Harvard University Press.

Grossman, M. (1972), “The demand for health: A theoretical and empirical investigation”, *National Bureau of Economic Research, Occasional paper 11.*

Healtzman, C., J. Frank, and R. G. Evans, (1994), “Heterogeneities in Health Status and the Determinants of Population Health”, in R. G. Evans (Ed.), Why are some People Healthy and Others Not, Berlin-New York: Walter de Gruyter.

Hildebrand, V., and P. Van Kerm, (2005), “Income Inequality and Self-rated Health Status: Evidence from the European Community Household Panel”, *IRISS Working Paper*, No. 2005-01.

Kuklys, W. (2005), *“Amartya Sen’s Capability Approach: Theoretical Insights and Empirical Applications”*, Berlin: Springer Verlag.

Lavy, V., J. Strauss, D. Thomas, and P. de Vreyer, (1996), “Quality of care, survival and health outcomes in Ghana”, *Journal of Health Economics*, Vol. 15, No. 3, pp. 333-357.

Marmot, M., and M. Shilpey, (1996), “Do Socio-Economic Differences Persist after Retirement? 25 Year Follow-up of Civil Servants from the First Whitehall Study”, *British Medical Journal*, Vol. 313, pp. 1177-1180.

Marmot, M., and R. Wilkinson, eds (2006), “*Social Determinants of Health”*, Oxford: Oxford University Press.

Martin, M.-C., (2006), “Individual and collective resources and women’s health in Morocco”, paper prepared for the 7th HD-CA Conference in Groeningen.

Montgomery, M.R. (2003), “Measuring living standards: household consumption and wealth indices”, Quantitative Techniques for Health Equity Analysis, Technical Note No. 4.

Montgomery, M.R., M. Gragnolati, K. A. Burke, and E. Paredes, (2000), “Measuring living standards with proxy variables”, *Demography*, Vol. 37, No. 2, pp. 155-174.

Rivera, B., (2001), “The effects of public health spending on self-assessed health status: an ordered probit model”, *Applied Economics*, Vol. 33, pp. 1313-1319.

Robeyns, I., (2000), "An unworkable idea or a promising alternative? Sen’s capability approach re-examined", Discussion paper 00.30, Center for Economic Studies, University of Leuven.

Robeyns, I., (2003), “The Capability Approach: An Interdisciplinary Introduction”, paper written for the Training Course preceding the 3rd International Conference on the Capability Approach on 6 September, Pavia, Italy.

Robeyns, I., (2005), "The Capability Approach: A Theoretical Survey", *Journal of Human Development*, Vol. 6, No. 1, pp. 93-114.

Robeyns, I., and W. Kuklys, (2004), “Sen's capability approach to welfare economics”, *Cambridge Working Papers in Economics* 0415.

Rosenzweig, M. R., and T. P. Schultz, (1982), “The Behaviour of Mothers as Inputs to Child Health: The Determinants of Birth Weight, Gestation, and Rate of Fetal Growth”, in V. Fuchs (Ed.), Economic Aspects of Health, Chicago: The University of Chicago Press.

Rosenzweig, M. R., and T. P. Schultz, (1983), “Estimating a Household Production Function: Heterogeneity, the Demand for Health Inputs, and Their

Effects on Birth Weight", *The Journal of Political Economy*, Vol. 91, No. 5, pp. 723-746.

Ruggeri Laderchi, C., (1999), "The Many Dimensions of Deprivation in Peru: Theoretical Debates and Empirical Evidence", QEH Working Paper Series 29.

Sahn, D., and D. Stifel, (2000), "Poverty comparisons over time and across countries in Africa", *World Development*, Vol. 28, No. 12, pp. 2123-2155.

Sahn, D., and D. Stifel, (2003), "Exploring alternative measures of welfare in the absence of expenditure data", *Review of Income and Wealth*, Vol. 49, No. 4.

Schultz, T. P., (1984), "Studying the Impact of Household Economic and Community Variables on Child Mortality", *Population and Development Review*, Vol. 10, Child Survival: Strategies for Research, pp. 215-235.

Sen, A., (1985), "*Commodities and Capabilities"*, New Delhi: Oxford University Press.

Sen, A., (1993), "Positional Objectivity", *Philosophy and Public Affairs*, Vol. 22, No. 2, pp. 126-145.

Sen, A., (1998), "Mortality as an Indicator of Economic Success and Failure", *The Economic Journal*, Vol. 108, No. 446, pp. 1-25.

Sen, A., (2002), "Health: perception versus observation. Self reported morbidity has severe limitations and can be extremely misleading", *British Medical Journal,* Vol. 324, pp. 860–1.

Strauss, J., and D. Thomas, (1995), "Human resources: Empirical modelling of household and family decisions", in: J. Behrman, T. N. Srinivasan (Eds.), Handbook of Development Economics, Vol. 3A, Amsterdam: Elsevier.

Thomas, D., J. Strauss, and M. H. Henriques, (1991), "How Does Mother's Education affect Child Height?", *The Journal of Human Resources*, Vol. 26, No. 2, pp. 183-211.

Thomas, D., V. Lavy, and J. Strauss, (1996), “Public policy and anthropometric outcomes in the Côte d’Ivoire”, *Journal of Public Economics,* Vol. 61, pp. 155-192.

Wagstaff, A., (2002), “Poverty and health sector inequalities”, *Bulletin of the World Health Organization*, Vol. 80 No. 2.

Wagstaff, A., and E. Van Doorslaer, (1994), “Measuring inequalities in health in the presence of multiple-category morbidity indicators”, *Health Economics*, Vol. 3, pp. 281-291.

Wagstaff, A., P. Paci, and H. Joshi, (2001), “Causes of inequality in health: Who you are? Where you live? Or who your parents were?”, *World Bank Working Paper 2713*, Washington, D.C.: World Bank.

Wilkinson, R., and M. Marmot, eds (2003), “*Social Determinants of Health: the Solid Facts”*, (2nd edition), Copenhagen: WHO Europe.

Appendix 3

Table 3.5: Marginal effects of Probit estimates using SRMI

	(1)	(2)	(3)	(4)
Private and public resources				
Wealth2	-0.007		-0.006	-0.006
	(0.002)***		(0.003)**	(0.003)**
Wealth3	-0.010		-0.009	-0.008
	(0.003)***		(0.003)***	(0.003)***
Wealth4	-0.015		-0.015	-0.015
	(0.002)***		(0.002)***	(0.002)***
Wealth5	-0.021		-0.020	-0.020
	(0.004)***		(0.004)***	(0.004)***
Public		-0.024	-0.023	-0.014
		(0.002)***	(0.002)***	(0.005)***
Wealth2*Public				-0.013
				(0.005)**
Wealth3*Public				-0.013
				(0.005)***
Wealth4*Public				-0.009
				(0.006)
Wealth5*Public				-0.011
				(0.007)
Personal characteristics				
Male	-0.019	-0.019	-0.019	-0.019
	(0.002)***	(0.001)***	(0.002)***	(0.002)***
White	0.003	0.005	0.003	0.003
	(0.001)***	(0.001)***	(0.001)**	(0.001)**
Mature people	0.041	0.041	0.042	0.042
	(0.003)***	(0.002)***	(0.002)***	(0.002)***
Elderly people	0.089	0.086	0.090	0.090
	(0.005)***	(0.004)***	(0.004)***	(0.004)***
Primary school	0.008	0.009	0.008	0.008
	(0.002)***	(0.002)***	(0.002)***	(0.002)***
Secondary school	-0.008	-0.008	-0.008	-0.008
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
College	-0.009	-0.014	-0.009	-0.009
	(0.003)***	(0.002)***	(0.003)***	(0.003)***
Post-graduate	-0.009	-0.014	-0.009	-0.009
	(0.009)	(0.009)	(0.009)	(0.009)
Labour market characteristics				
Farmer	-0.002	0.000	-0.003	-0.002
	(0.003)	(0.003)	(0.003)	(0.003)
Intermediate level	0.000	0.004	0.001	0.001
	(0.003)	(0.003)	(0.003)	(0.003)
Blue collar level	0.006	0.010	0.006	0.006
	(0.003)**	(0.003)***	(0.003)**	(0.003)**
Formal	0.002	0.001	0.002	0.002
	(0.003)	(0.003)	(0.003)	(0.003)
Geographical characteristics				

North-East	-0.015	-0.015	-0.015	-0.015
	(0.008)*	(0.008)*	(0.008)*	(0.008)*
South-East	-0.012	-0.008	-0.004	-0.004
	(0.010)	(0.008)	(0.008)	(0.008)
South	-0.002	-0.006	-0.002	-0.002
	(0.008)	(0.008)	(0.008)	(0.008)
Central-West	-0.002	-0.004	-0.002	-0.002
	(0.008)	(0.008)	(0.008)	(0.008)
Urban	0.016	0.011	0.016	0.016
	(0.003)***	(0.002)***	(0.002)***	(0.002)***
Brasilia	0.001	0.032	0.035	0.035
	(0.003)	(0.004)***	(0.005)***	(0.005)***
São Paolo	0.003	-0.007	-0.006	-0.006
	(0.008)	(0.002)***	(0.002)**	(0.002)**
Roraima	0.002	0.037	0.037	0.035
	(0.008)	(0.013)***	(0.012)***	(0.013)***
Acre	0.034	0.088	0.083	0.076
	(0.012)***	(0.018)***	(0.017)***	(0.017)***
Observations	128,028	128,028	128,028	128,028
Pseudo-R^2	0.0436	0.0431	0.0449	0.0450

Robust standard errors adjusted for clustering on *Unidade de Federacão* in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3.6: Ordered probit estimates using SHSI

	(1)	(2)	(3)
Private and public resources			
Wealth2	0.071		0.067
	(0.021)***		(0.019)***
Wealth3	0.165		0.161
	(0.026)***		(0.026)***
Wealth4	0.275		0.271
	(0.024)***		(0.024)***
Wealth5	0.513		0.508
	(0.024)***		(0.025)***
Public		0.244	0.226
		(0.031)***	(0.032)***
Personal characteristics			
Male	0.201	0.195	0.201
	(0.010)***	(0.010)***	(0.010)***
White	-0.081	-0.122	-0.077
	(0.016)***	(0.018)***	(0.016)***
Mature people	-0.542	-0.517	-0.545
	(0.020)***	(0.019)***	(0.020)***
Elderly people	-0.996	-0.941	-1.001
	(0.027)***	(0.026)***	(0.027)***
Primary school	-0.267	-0.285	-0.266
	(0.018)***	(0.019)***	(0.018)***
Secondary school	0.028	0.037	0.028
	(0.014)**	(0.015)**	(0.014)**
College	0.350	0.491	0.347
	(0.036)***	(0.035)***	(0.037)***
Post-graduate	0.345	0.507	0.341

	(0.083)***	(0.086)***	(0.084)***
Labour market characteristics			
Farmer	0.045	0.002	0.053
	(0.027)	(0.029)	(0.026)**
Intermediate level	-0.144	-0.229	-0.146
	(0.017)***	(0.016)***	(0.017)***
Blue collar level	-0.217	-0.307	-0.219
	(0.021)***	(0.021)***	(0.021)***
Formal	0.149	0.166	0.150
	(0.014)***	(0.015)***	(0.014)***
Geographical characteristics			
North-East	0.046	0.057	0.045
	(0.102)	(0.105)	(0.102)
South-East	0.199	0.200	0.127
	(0.113)*	(0.102)*	(0.098)
South	0.160	0.243	0.163
	(0.104)	(0.108)**	(0.104)
Central-West	0.027	0.083	0.028
	(0.095)	(0.099)	(0.096)
Urban	-0.083	0.021	-0.085
	(0.024)***	(0.021)	(0.024)***
Brasilia	-0.048	-0.211	-0.271
	(0.022)**	(0.033)***	(0.036)***
São Paolo	0.001	0.112	0.077
	(0.060)	(0.031)***	(0.033)**
Roraima	-0.234	-0.454	-0.459
	(0.094)**	(0.103)***	(0.100)***
Acre	-0.220	-0.478	-0.446
	(0.094)**	(0.103)***	(0.100)***
/Cut1	-2.365	-2.525	-2.367
	0.105	0.108	0.105
/Cut2	-1.098	-1.27	-1.099
	0.109	0.111	0.109
Observations	128,028	128,028	128,028
Pseudo-R^2	0.0977	0.0908	0.0984

Robust standard errors adjusted for clustering on *Unidade de Federacão* in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3.7: Marginal effect of Probit estimates with SRMI by race

	(1)White women	(2)White men	(3)Black women	(4)Black men
Private and public resources				
Wealth2	-0.001	-0.006	-0.005	-0.009
	(0.007)	(0.005)	(0.003)*	(0.004)***
Wealth3	-0.002	-0.013	-0.010	-0.007
	(0.005)	(0.006)**	(0.004)**	(0.003)**
Wealth4	-0.009	-0.018	-0.017	-0.014
	(0.006)	(0.004)***	(0.003)***	(0.004)***
Wealth5	-0.019	-0.019	-0.017	-0.017
	(0.006)***	(0.006)***	(0.005)***	(0.005)***
Public	-0.005	-0.006	-0.029	-0.004
	(0.016)	(0.020)	(0.010)***	(0.010)
Wealth2*Public	-0.023	-0.015	-0.005	-0.013
	(0.015)	(0.013)	(0.012)	(0.010)
Wealth3*Public	-0.014	-0.011	-0.008	-0.024
	(0.014)	(0.018)	(0.012)	(0.007)***
Wealth4*Public	-0.018	-0.000	0.007	-0.026
	(0.013)	(0.024)	(0.018)	(0.005)***
Wealth5*Public	-0.022	-0.013	-0.002	-0.012
	(0.013)*	(0.018)	(0.018)	(0.009)
Personal characteristics				
Mature people	0.035	0.031	0.050	0.045
	(0.003)***	(0.004)***	(0.004)***	(0.005)***
Elderly people	0.082	0.061	0.105	0.098
	(0.005)***	(0.008)***	(0.007)***	(0.008)***
Primary school	0.011	0.008	0.009	0.003
	(0.004)***	(0.005)*	(0.004)**	(0.003)
Secondary school	-0.005	-0.006	-0.009	-0.008
	(0.003)	(0.004)	(0.003)***	(0.002)***
College	-0.009	-0.005	-0.013	0.006
	(0.005)*	(0.005)	(0.005)**	(0.007)
Post-graduate	-0.016	-0.012	0.018	0.025
	(0.013)	(0.009)	(0.034)	(0.035)
Personal characteristics				
Farmer	0.001	-0.013	0.002	-0.008
	(0.006)	(0.004)***	(0.005)	(0.004)*
Intermediate level	0.002	0.003	-0.014	-0.000
	(0.004)	(0.004)	(0.007)*	(0.005)
Blue collar level	0.010	0.016	-0.019	0.014
	(0.004)**	(0.004)***	(0.010)*	(0.004)***
Formal	0.005	0.006	-0.004	-0.001
	(0.005)	(0.003)**	(0.005)	(0.004)
Geographical characteristics				
North-East	-0.013	-0.016	-0.018	-0.010
	(0.011)	(0.007)**	(0.010)*	(0.006)*
South-East	-0.005	-0.010	-0.004	-0.002
	(0.011)	(0.008)	(0.009)	(0.006)
South	0.001	-0.011	-0.004	-0.003
	(0.012)	(0.007)	(0.009)	(0.006)
Central-West	-0.002	-0.016	0.003	0.001

	(0.012)	(0.006)**	(0.010)	(0.006)
Urban	0.021	0.006	0.020	0.008
	(0.004)***	(0.004)	(0.004)***	(0.005)
Brasilia	0.039	0.014	0.052	0.024
	(0.006)***	(0.007)*	(0.009)***	(0.004)***
São Paolo	-0.009	-0.001	-0.002	-0.004
	(0.001)***	(0.005)	(0.003)	(0.002)
Roraima	0.066	-0.023	0.063	0.012
	(0.021)***	(0.005)***	(0.016)***	(0.008)
Acre	0.075	0.036	0.110	0.060
	(0.022)***	(0.022)	(0.023)***	(0.014)***
Observations	39,857	20,144	44,519	23,508
Pseudo-R^2	0.0357	0.0398	0.0455	0.0627

Robust standard errors adjusted for clustering on *Unidade de Federacão* in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3.8: Ordered probit estimates using SHSI by race

	(1)White women	(2)White men	(3)Black women	(4)Black men
Private and public resources				
Wealth2	0.045	0.103	0.067	0.105
	(0.033)	(0.032)***	(0.021)***	(0.032)***
Wealth3	0.134	0.237	0.156	0.205
	(0.043)***	(0.047)***	(0.030)***	(0.036)***
Wealth4	0.277	0.319	0.255	0.289
	(0.044)***	(0.033)***	(0.029)***	(0.035)***
Wealth5	0.546	0.536	0.444	0.451
	(0.041)***	(0.039)***	(0.033)***	(0.044)***
Public	0.172	0.110	0.279	0.369
	(0.024)***	(0.045)**	(0.035)***	(0.026)***
Personal characteristics				
Mature people	-0.536	-0.570	-0.534	-0.564
	(0.023)***	(0.036)***	(0.028)***	(0.030)***
Elderly people	-0.993	-0.957	-0.998	-1.060
	(0.035)***	(0.036)***	(0.038)***	(0.032)***
Primary school	-0.279	-0.394	-0.255	-0.171
	(0.025)***	(0.037)***	(0.027)***	(0.029)***
Secondary school	0.048	-0.050	0.029	0.046
	(0.022)**	(0.037)	(0.014)**	(0.027)*
College	0.420	0.234	0.297	0.233
	(0.030)***	(0.066)***	(0.066)***	(0.072)***
Post-graduate	0.409	0.229	0.270	0.167
	(0.120)***	(0.163)	(0.283)	(0.279)
Personal characteristics				
Farmer	-0.046	0.139	0.024	0.147
	(0.026)*	(0.043)***	(0.038)	(0.048)***
Intermediate level	-0.141	-0.090	-0.133	-0.200
	(0.031)***	(0.037)**	(0.044)***	(0.037)***
Blue collar level	-0.226	-0.258	-0.183	-0.263
	(0.035)***	(0.033)***	(0.040)***	(0.043)***
Formal	0.122	0.187	0.126	0.181
	(0.028)***	(0.026)***	(0.021)***	(0.022)***

Geographical characteristics				
North-East	0.040	0.053	0.049	0.019
	(0.123)	(0.097)	(0.103)	(0.098)
South-East	0.168	0.156	0.133	0.052
	(0.110)	(0.103)	(0.098)	(0.094)
South	0.206	0.169	0.136	0.141
	(0.113)*	(0.099)*	(0.116)	(0.119)
Central-West	0.069	0.042	-0.000	0.027
	(0.108)	(0.102)	(0.095)	(0.092)
Urban	-0.076	-0.050	-0.118	-0.063
	(0.028)***	(0.062)	(0.033)***	(0.040)
Brasilia	-0.257	0.035	-0.353	-0.358
	(0.027)***	(0.058)	(0.042)***	(0.025)***
São Paolo	0.095	0.068	0.023	0.019
	(0.025)***	(0.044)	(0.036)	(0.026)
Roraima	-0.529	-0.322	-0.530	-0.526
	(0.113)***	(0.104)***	(0.100)***	(0.096)***
Acre	-0.390	-0.163	-0.522	-0.587
	(0.113)***	(0.106)	(0.099)***	(0.096)***
/Cut1	-2.279	-2.386	-2.262	-2.356
	0.121	0.120	0.105	0.102
/Cut2	-0.984	-1.216	-0.944	-1.173
	0.121	0.113	0.112	0.106
Observations	39,857	20,144	44,519	23,508
Pseudo-R^2	0.1075	0.1113	0.0792	0.0884

Robust standard errors adjusted for clustering on *Unidade de Federacão* in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%