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

and $\mathit{Treatment} = (\hat{\beta}_A - \hat{\beta}_B)'\overline{X}_B$ 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:

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

[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.

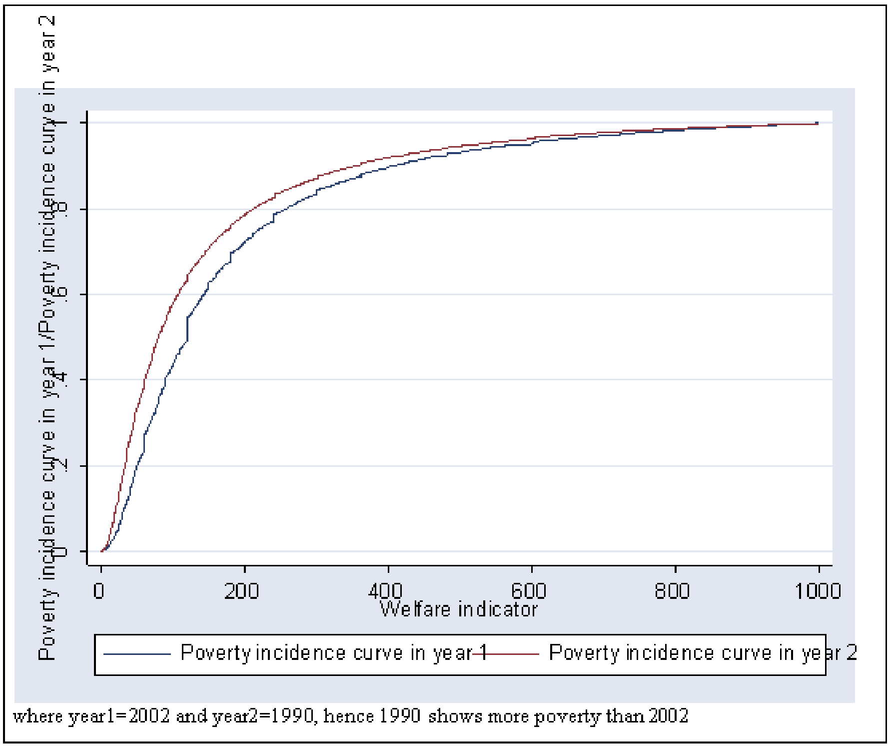
Wodon, Q., (2000), “Poverty and Policy in Latin America and the Caribbean”, *World Bank Technical Paper*, No. 467, Washington, D.C.: World Bank.

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

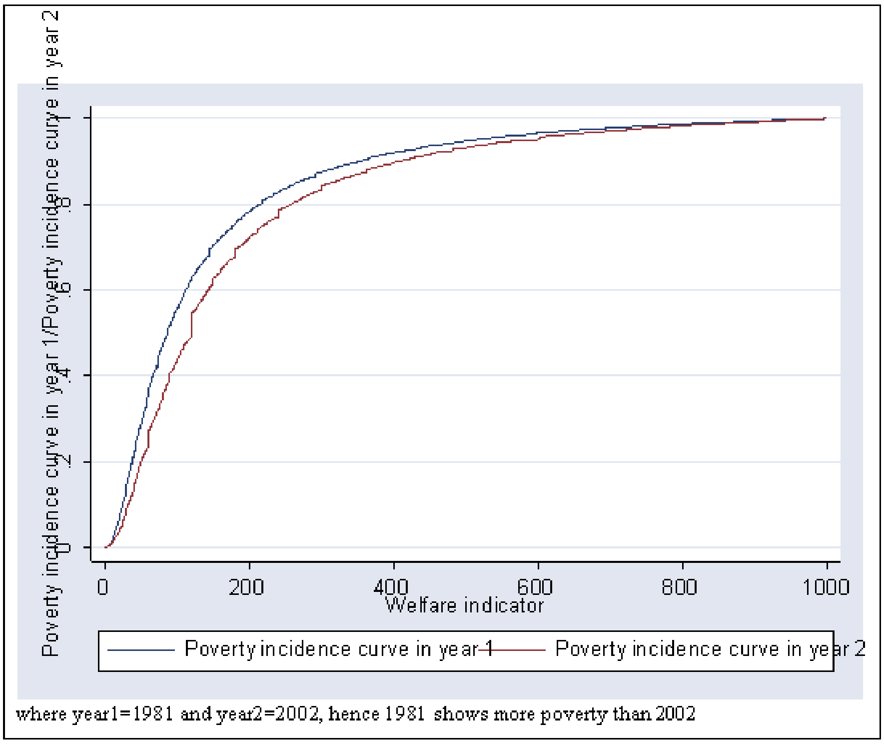
Graphs and Tables

Figure A1.1: Poverty Incidence between 2002 and 1990



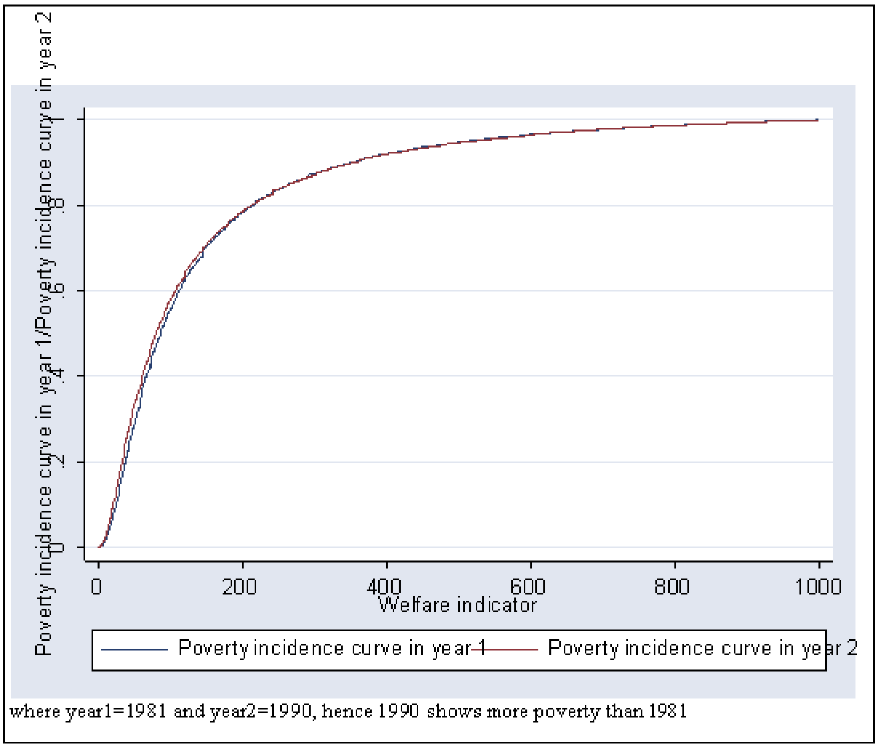
Source: Author's calculation from PNAD 1990 and 2002.

Figure A1.2: Poverty Incidence between 2002 and 1981



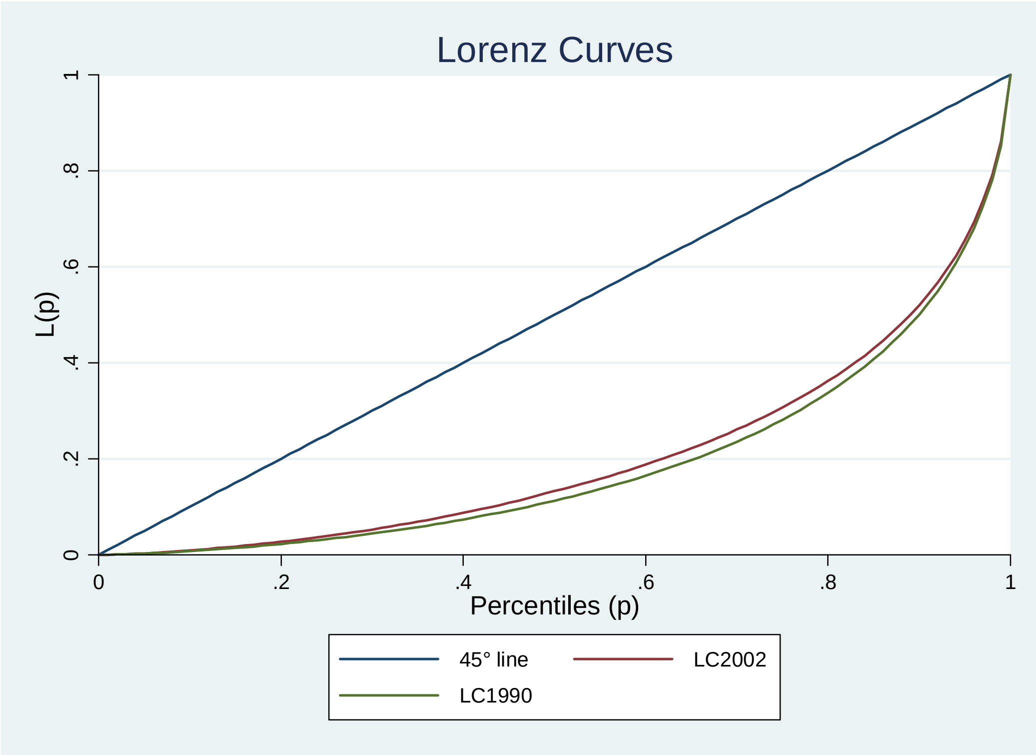
Source: Author's calculation from PNAD 1981 and 2002.

Figure A1.3: Poverty Incidence between 1990 and 1981



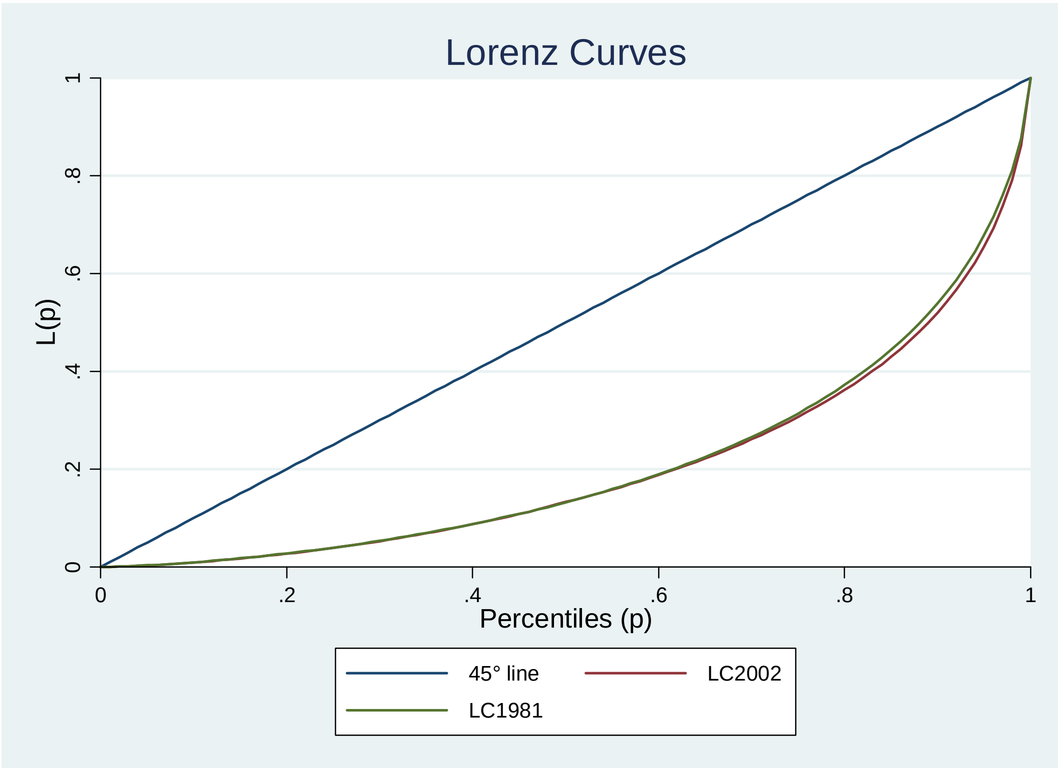
Source: Author's calculation from PNAD 1990 and 2002.

Figure A1.4: Lorenz dominance between 2002 and 1990



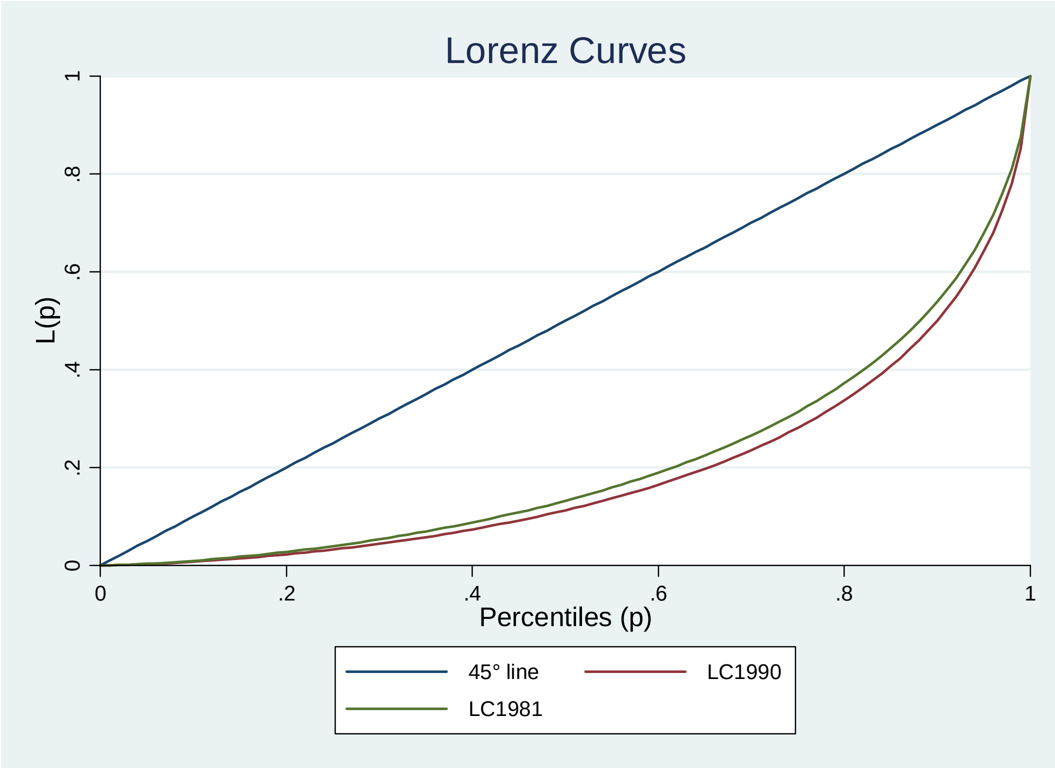
Source: Author's calculation from PNAD 1990 and 2002.

Figure A1.5: Lorenz dominance between 2002 and 1981



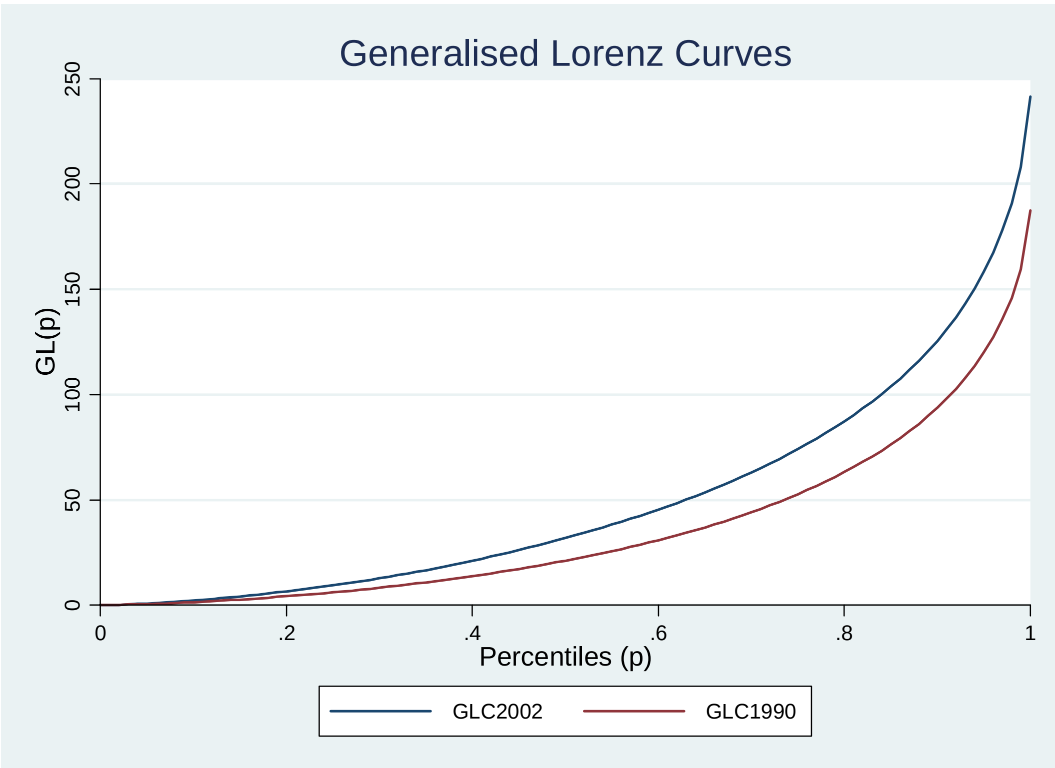
Source: Author's calculation from PNAD 1981 and 2002.

Figure A1.6: Lorenz dominance between 1990 and 1981



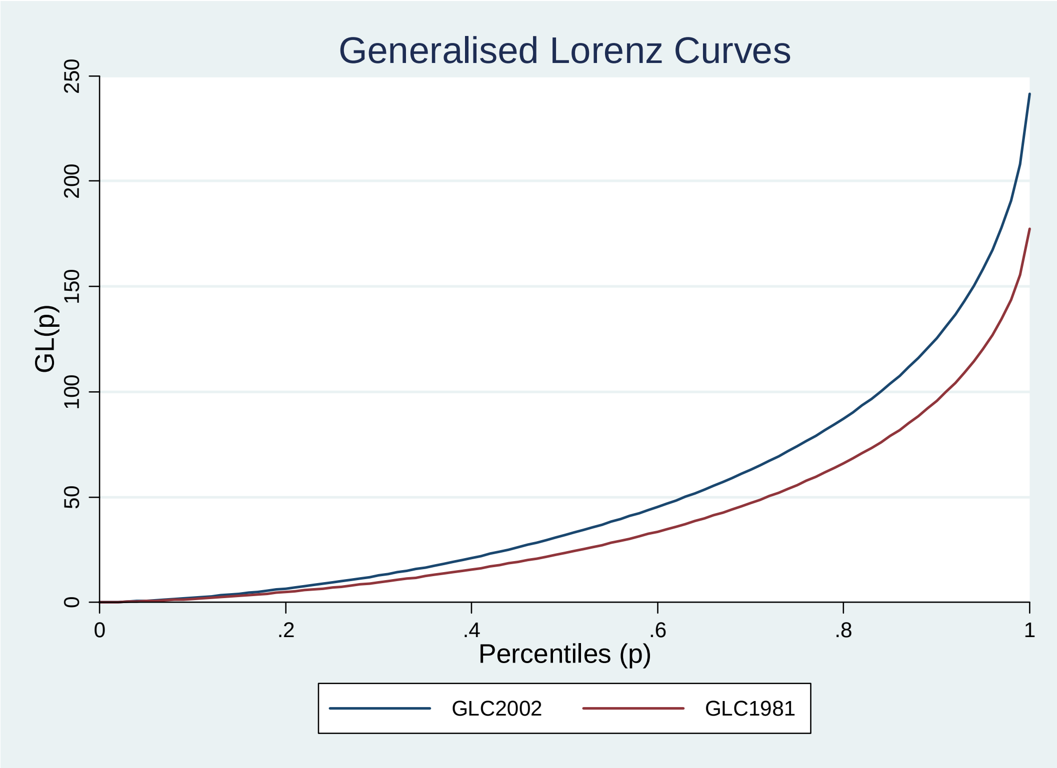
Source: Author's calculation from PNAD 1981 and 1990.

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



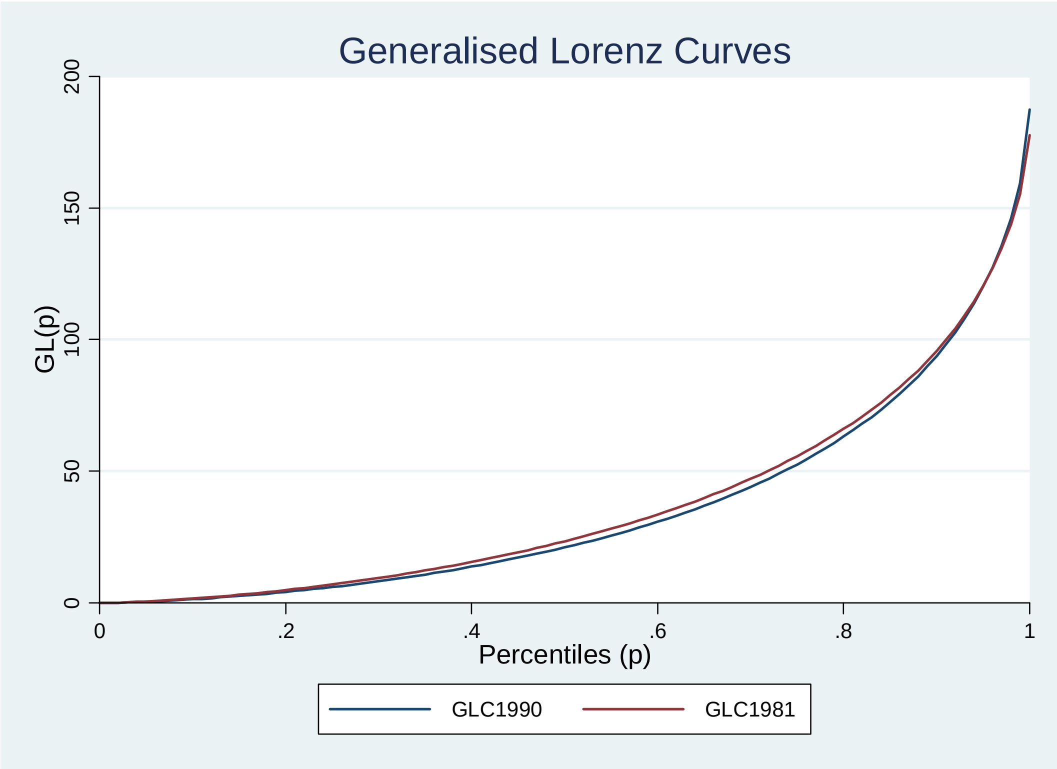
Source: Author's calculation from PNAD 1990 and 2002.

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	Δβ
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.