NEREUS
Núcleo de Economia Regional e Urbana
da Universidade de São Paulo
The University of São Paulo
Regional and Urban Economics Lab

INCOME INEQUALITY REDUCTION IN BRAZIL: A PSEUDO-PANEL APPROACH IN THE SEARCH OF ITS DETERMINANTS

Ana Maria B. Barufi

TD Nereus 11-2012
São Paulo
2012

Income Inequality Reduction in Brazil: A Pseudo-panel Approach in the Search of its Determinants

Ana Maria Bonomi Barufi[1]

Abstract: Ever since the late 1990`s personal income inequality has shown a steady decrease in Brazil. Most of the investigations of this phenomenon are restricted to the analysis of aggregated information along the years. This paper tries to identify the main factors for it, using microdata of the National Household Sample Survey (PNAD) for the period 1995-2009. The years have been chosen in order to allow an evaluation of different economic and political contexts and a pseudo-panel based on age-state cohorts is built to compare the effect of transfers (social programs and retirement), wage increases and regional specific factors in the decrease of income inequality. The pseudo-panel is essential to control for unobservable characteristics of the head of the household and is an option for the lack of true panel data information in Brazil. In spite of the overall improvement of households' lives in the country, regional inequalities persist and remain to be assessed in the following decades. However, it is noticeable the essential role of social transfers to reduce poverty (at least in the short-run) and induce the inequality reduction.

1. Introduction

Inequality has been widely studied in Brazil due to its impressive reduction in the last decade. Even though the country still presents huge income disparities, both in personal and regional dimensions, advances have been made along the recent period. This paper aims to investigate the relative importance of each income component in the explanation of that movement in the last 15 years, identifying regional heterogeneities in the relationships studied.

The relation among income inequality and economic growth in a macroeconomic perspective has been explored in the literature, as described by Bénabou (1996). Kuznets' Curve theory states that inequality and income per capita have an inverted U relationship (the initially proposed idea can be found in KUZNETS, 1955). However, there has been some criticism over this framework due to the lack of empirical support. Such discussion led to the work of Ravaillon (2003), who explored the possibility of inequality convergence among countries, finding empirical evidence of this movement.

[1] The author is very thankful to the comments by Professor Carlos Azzoni and Professor Naercio Menezes Filho, which were crucial to the set up of the study.

Inequality reduction when simultaneous to income growth is a clear sign of development of a country. In the Latin America context, Brazil has shown one of the major and consistent decreases in the Gini coefficient[2] in the last 15 years, as can be seen in Table 1.

It is important to notice that even though this movement has occurred, Brazil still presents one of the highest inequality levels among the countries analyzed. Another relevant aspect to be considered when studying this issue is the timing of such process (Graph 1). During the 1980's and the early 1990's, Brazil had a mix of high inflation and low economic growth, leading to a very unstable inequality level[3]. After that, with the stabilization plan (*Plano Real*) and a proper democracy, income inequality got less erratic. During the second half of the 1990's, important economic policies where conducted, mainly related to the deregulation of labor market and education expansion. Moreover, cash transfer programs were created in order to ease extreme poverty in the country and stimulate poor families to take their children to school[4].

[2] The Gini coefficient ranges from 0 to 1, 0 meaning perfect equality (everybody has the same income level) and 1 meaning complete inequality (only one person has all income of the economy).
[3] Another aspect to be considered is that prior to 1992 the indicators provided by PNAD where less trustable than later on.
[4] Some of the transfer programs created in that period are Bolsa Escola, Auxílio Gás, Auxílio Alimentação, BCP (Benefício de Prestação Continuada) – LOAS, PETI (Programa de Erradicação do Trabalho Infantil). In the following decade. Fome Zero and Bolsa Família Program were created and the later substituted a major part of these programs.

Table 1. Inequality in Latin America: Gini Coefficient in 1996, 2003 and 2009

	1996	2003	2009
Argentina	0.495	0.547	0.461
Brazil	0.606	0.588	0.547
Bolivia	0.583 [2]	0.599 [4]	0.563 [6]
Chile	0.549	0.546	0.521
Colombia	0.569	0.579	0.567
Costa Rica	0.465	0.497	0.507
Ecuador	0.512 [1]	0.551	0.494
Mexico	0.485	0.497 [4]	0.483 [7]
Paraguay	0.582 [1]	0.569	0.510
Peru	0.562	0.552	0.491
Uruguay	0.427	0.462	0.463
Venezuela, RB	0.472 [1]	0.481	0.448 [5]

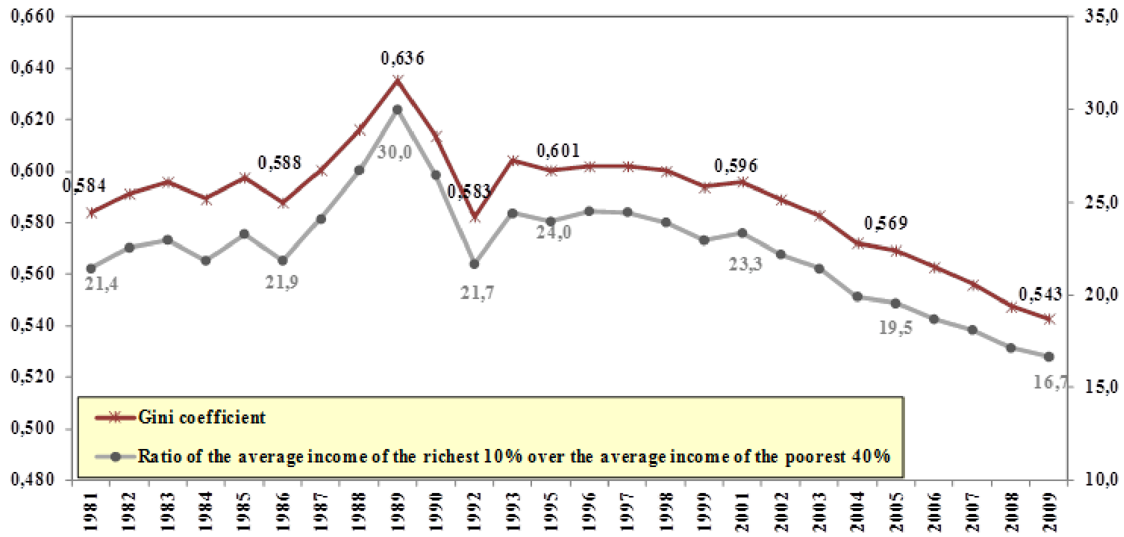
1 - 1995, 2 - 1997, 3 - 1998, 4 - 2002, 5 - 2005, 6 - 2007, 7 - 2008

Source: World Bank.

During the following decade these processes have been further stimulated: social transfers increased consistently, average education level of the population continued to grow and minimum wage has increased relatively to the average income of the country. More than that, the country has seen abnormally high economic growth (when compared to the previous decades).The combination of all these elements resulted in inclusive income growth, reducing inequality.

Graph 1 also shows that inequality measured in different ways has the same pattern along the years. Therefore it is possible to consider different inequality measures to understand the process that occurred in the past 15 years.

Graph 1. Gini Coefficitent and the Ratio of the Average Income of the Richest 10% over the Average Income of the Poorest 40% (based on per capita income), 1981-2009*



* In 1991, 1994 and 2000 the survey was not conducted by IBGE.

Source: IPEAdata, PNAD/IBGE.

This paper presents a new way to study inequality, at least as far as is of the knowledge of the author, applying regression methods instead of decomposition analysis to understand the relative importance of each income component in the evolution of inequality inside state-age cohorts. An analysis of income inequality convergence is also presented in order to understand the timing of this process in Brazil.

It is important to notice that the objective here is to explain the movement of an inequality indicator measured inside each cohort. The underlying hypothesis of this approach is that if inequality decreases in most state-age groups, then it is expected that the general indicator will show the same trend. As will be discussed in the following sections, the cohort analysis allows the comparison of nearly homogeneous groups along the years, building a pseudo-panel. Furthermore, when the dimension analyzed goes this deep in the decomposition of the population it is possible to reach conclusions that aggregated data do not conceive.

The remaining of the paper is organized as follows: section 2 reviews the inequality literature in Brazil; section 3 discusses methodological issues; data is presented in section 4; section 5 brings the results; and section 6 develops the conclusions and future steps in this work.

2. Literature Review

The study of inequality in Brazil has been based in two main approaches: one of the decomposition of inequality and the other of income convergence, where a negative signal of the initial level of income represents inequality reduction. The later will be further discussed in the methodology session, while the former is briefly presented in the following paragraphs.

Considering the first methodological approach, Souza e Osorio (2011) identify the labor income as the main factor of the reduction of disparities among metropolitan and non-metropolitan regions from 1981 to 2009. Silveira-Neto and Azzoni (2011) explored regional income convergence and an inequality decomposition, finding that minimum wage changes, income transfer programs and the increase in labor productivity were relevant for the regional inequality reduction seen from 1995 to 2005.

Soares (2006) shows that 2004 presented the lowest level of income inequality for different measures (Gini coefficient, Theil T, ratio 10/40 and ratio 20/20) in the period of 1976-2004. Moreover, applying inequality decomposition the author shows that from 1995 to 2004 transfer programs such as Bolsa Família were responsible for ¼ of the whole inequality reduction. Hoffman (2000) discusses the most appropriated income to be considered depending on the objective of the analysis. If the researcher is interested in the labor market dynamics, the income of the economically active population should be used. On the other hand, if the main concern is with the welfare of families, then the recommendation is to analyze per capita familiar income.

Initially this paper aimed to identify the importance of cash transfer programs to the reduction of income inequality. However, many difficulties were found to fulfill this task, in such a way that most studies of this subject apply only two years of the PNAD database (2004 and 2006)[5], in which there was a specific group of questions to

[5] See for instance Soares et al. (2007a), Soares et al. (2007b), Pedrozo (2010),

investigate the issue of transfers. Hoffman (2007) recognizes this limitation but still tries to assess the impact of transfers on income inequality considering the more aggregated income group available annually (transfers, dividends, interest and other incomes). According to Paes de Barros and Carvalho (2003) the problem to identify this income component was even worst before 2003, as social transfer programs did not have a very accurate focus.

Another strand of research in the area of inequality reduction can be found in the work of Ravaillon (2003), who proposes the identification of absolute convergence in this indicator. Therefore, the first part of this study will explore this issue, applying the idea, which was originally considered for countries, for state-age-cohorts. Then, the second part will be dedicated to understand the main factors for this income inequality reduction.

3. Methodology

Considering the relevant issues to be discussed in order to develop the analysis proposed here, firstly the cohorts' composition and its potential advantages and disadvantages are presented. After that, the econometric model is proposed, both for the study of inequality convergence and for the explanation of the inequality level.

3.1. Cohorts

According to Ryder (1965), birth cohorts constitute aggregates that minimize the attrition present in the society. Each cohort has contact both with innovative and conservative forces, what results in a unique way of relating with the remaining of the society. Therefore, birth cohorts usually differ in their cultural references, meaning that they will value differently education, the propensity of women to participate in the labor force, as well as they may have different attitudes toward risk, discount factors, and preferences over the lifetime path of consumption. As Ryder highlights, each cohort is different from the others because it embodies a temporally specific version of the heritage.

The approach of birth cohorts allows the construction of a pseudo-panel, which is especially useful when annual surveys constitute the main source of information about the socioeconomic evolution of a country. In Brazil, the only comprehensive survey that provides annually a wide range socioeconomic information is the National Household Sample Survey (PNAD), by IBGE.As the survey has no longitudinal structure, a pseudo-panel approach is the instrument applied to compare homogeneous groups along the years.

This method has been used earlier by Deaton and Paxson (2001), who consider age-cohorts to investigate the impact of income inequality over mortality probability, identifying that more than inequality, socioeconomic characteristics should be analyzed when explaining mortality propensity. Another application can be found in studies of family consumption patterns over household head age cohorts in the United States (BROWNING et al., 1985) and in the relationship of consumption patterns and business cycles in the United Kingdom (ATTANASIO and BROWNING, 1995).

According to Browning et al. (1985), as one-year cohorts usually yield samples that are too small to give accurate estimates of sample means, it is possible to apply five-year age bands. Other treatments to the database are necessary in order to achieve cohorts with a minimum size sufficient to provide a robust analysis. Therefore, some age groups are excluded from the analysis whenever at some point of time their cohort size would be too small. This treatment is applied to families whose household head's age is above a certain upper limit at the first year considered in the panel (as they would be really old in the last year of the panel and probably this cohort would be too small due to mortality) or below a lower bound age at the last year of the panel (because at the first year of the panel only a few households would be headed by a under-20 person, for instance).

Azzoni et al. (2000) describe the advantages of using cohort-level data, highlighting that this method allows the researcher to control for changes in the composition of the population in each state, shows the effects of life cycle evolution in the behavior of families and permits the control of unobserved effects with random effects and/or fixed effects estimation strategies because of its panel data structure. Moreover, with this method it is possible to combine time-series and cross-section data, resulting in a simultaneous analysis of behavior over both life and business cycles (ATTANASIO and BROWNING, 1995).

There are also relevant disadvantages that cohort-level data provide (AZZONI et al., 2000): measurement errors at the household level are likely to be carried out to the cohort level unless the cell sizes are big; if there is a strong migration flow inside a country, the composition of age-state-cohorts may change over time. Nevertheless, in Brazil it is more common to find intrastate migration instead of interstate. Important interregional flows were observed during the 1970's, with the extraordinary growth of the Southeast region attracting people from all over the country (but especially from the poorer Northeast). During the period analyzed here this is not such a relevant issue, as recent migration usually accounts for less than 9% of the resident population in each state/year.

In the Brazilian context this method has been applied to investigate regional economic convergence (AZZONI et al., 2000, AZZONI et al., 2003) controlling for the age structure of the population. Within the same framework, MENEZES et al. (2011) find that regional inequality is diminishing for the older cohorts and is increasing, or non-diminishing, for the younger cohorts. The authors call the attention for the fact that returns to experience differ among regions in Brazil and age structures are different from one state to the other, what justifies the usage of state-age cohort level data. It is important to notice that most of these works identify inequality reduction in the regional level as a consequence of a regression of income convergence.

As Brazil has a continental dimension, each region has a particular cultural background, apart from general infrastructure, education and health access. Therefore, following the works mentioned above, cohorts here will be based not only on household head age groups but also on states, resulting in 20 states[6] x 8 age groups = 160 cohorts for each year.

Differently from the income convergence approach, this paper assesses the issue of income inequality decrease exploring the possibility of reduction of the index of each

[6] As prior to 2004 PNAD covered only urban areas of most states of the North region, all states from this region were excluded from the analysis.

state-age-cohort. The underlying hypothesis of this reasoning is that as far as income inequality decreases inside most cohorts, the highest the probability that the accumulated index (for the whole country) will show the same movement. The advantage of this approach is that possibly personal characteristics will have a higher role in explaining the process.

However, constructing a dispersion variable poses a higher difficulty than mean variables. Therefore, even if the Gini coefficient is commonly applied in most income inequality studies, here another measure is considered, the ratio of the average income of the richest 10% (p90) over the average income of poorest 40% (p40) in each state-age-cohort (II_{it}).

$$II_{it} = \frac{\bar{w}_{p90_{it}}}{\bar{w}_{p40_{it}}} \tag{1}$$

where *i* represents each cohort based on one age group and one state.

Another relevant aspect of the construction of this indicator is that the income considered here is the total income of the household head[7] (and the household head is also the reference for the construction of the cohorts). This is a strategy to avoid large variations in the income analyzed, what is needed in a situation where the dependent variable, income inequality, is already difficult to build in a robust and consistent way.

3.2. Econometric Issues

In this session the basic econometric model will be described considering the problems to be studied here, followed by a comparison among different estimations methods, tests to be conducted in order to define the best model to be analyzed and potential econometric issues to be dealt with.

[77] Here household head is the person who is the reference in the family.

3.2.1. Inequality convergence

The first problem to be studied here is based on Ravaillon (2003), who presents the idea of testing the existence of inequality convergence among countries (inspired by Bénabou, 1996). His results show that even when treating the Gini index for misspecification, absolute convergence is found.

Considering the panel data structure of the data, the model to be estimated has the following structure[8]:

$$II_{it} - II_{i,t-1} = \alpha + \beta II_{i,t-1} + u_{it} \qquad (2)$$

where II_{it} is the inequality measure, α and β are the parameters to be estimated, u_{it} is the zero-mean error term, *i* represents the state-age cohorts $(i = 1,2,\ldots,N)$ and *t* stands for time intervals available$(t = 1,2,\ldots,T)$. If β is found to be negative (positive) there is inequality convergence (divergence).

3.2.2. Inequality level

The counterpart for the inequality convergence study is the analysis of which factors affect the level of inequality along time for the cohorts constructed here. In this sense, the equation to be estimated is the following:

$$II_{it} = \beta X_{it} + \delta_t + c_i + u_{it} \qquad (3)$$

It is noticeable that now the dependent variable is the indicator of income inequality II_{it} and that a contemporaneous set of explanatory variables is considered (X_{it}). The objective is to identify the vector of parameters β, acquiring information about the relative effect of each explanatory variable considered. This vector will contain the average participation of each income component (see Data subsection) in the average income of the cohort, state and year dummies and other relevant controls.

[8] This methodological session is also highly based on Wooldridge (2002).

3.2.3. Estimation methods[9]

In the panel data literature the choice of which econometric procedure to apply may depend on the sample size, the number of time periods and other characteristics of the database. Here three different approaches will be discussed and compared, based on the unobserved effects model (UEM): Pooled Ordinary Least Squares (POLS), Random Effects (RE) and Fixed Effects (FE). Moreover, potential econometric issues will be explored in the following subsection. Finally, in order to establish which method is more appropriate, a Hausman test will be applied, comparing RE and FE.

The first aspect to notice is that in this paper time (T = 14) and cross-section dimensions (N = 160) are such that N is sufficiently larger than T, meaning that an asymptotic analysis can provide suitable approximations. In this sense, there is no need to employ time series techniques to treat data properly.

3.2.3.1. The Unobserved effects model

One of the main advantages provided by panel data model is that there are alternatives to obtain consistent estimators in the presence of omitted variables. Considering cross-section units (cohorts) i = 1,…,N and time periods t = 1,…,T, y_{it} represents the dependent variable for each cohort in each time period, X_{it} represents the set of K explanatory variables, c_i is the unobserved heterogeneity that is constant for each cohort over time and δ_t is the set of time dummies (constant among cohorts). The basic model can be written as follows:

$$y_{it} = X_{it}\beta + \delta_t + c_i + u_{it} \quad (4)$$

However, as c_i is unobserved, the estimation of the vector β is consistent only if some special conditions are satisfied, as will be seen in the following sections.

[9] This section is largely based on Wooldridge (2002) and Baltagi (2008).

3.2.3.2. Pooled OLS (POLS)

In the estimation by POLS, c_i becomes a component of the residual. Therefore, the composed residual will be:

$$v_{it} = c_i + u_{it} \tag{5}$$

In order to identify β, some hypotheses need to be satisfied:

POLS.1: i) $E(X_t{}'u_t) = 0$ (6)

ii) $E(X_{it}{}'c_i) = 0$

POLS.2: $rank[\sum_t E(X_t{}'X_t)] = K$ (7)

Equation (6) represents the condition of weak exogeneity and (7) states for the rank condition, which allows the inversion of the matrix X'X. The estimation by POLS of this model has weak exogeneity conditions. However, this estimator is inefficient because of the presence of the unobserved effect in the residuals, which generates serial autocorrelation. Robust variance matrices and robust test statistics are needed to deal with this issue.

3.2.3.3. Random effects (RE)

The RE model considers that the behavior of the unobserved effect is unknown and can be represented as a random variable, being treated as part of the error term (as in (5)). Therefore, the error term is composed by two random elements, the idiosyncratic term that varies among all individuals and time periods and the unobserved effect, which is specific for each individual and constant over time. The main difference from RE and POLS is that the former explores the serial correlation found in the error term when the unobserved effect composes it. This advantage does not come without a cost: identification assumptions are more restrictive in the RE case.

The basic idea is to estimate Generalized Least Squares (GLS) model weighting the observations by the correlation among shocks in each time period, taking into account the specific error structure. As already mentioned, RE estimation imposes more restrictive assumptions than POLS, because it includes the unobserved heterogeneity in the residual in a deliberated way, meaning that its hypotheses are sufficient to obtain consistent and efficient estimators even in this situation. The residuals are composed as showed in (5) and the identification of β in this case is based on the following assumptions:

RE.1: i) $E(u_{it}|X_{it}, c_i) = 0$ (8)

ii) $E(c_i|X_i) = E(c_i) = 0$

RE.2: $rank\left(E(X_i{}'\Omega^{-1}X_i)\right) = K$

where $\Omega = E(v_i v_i{}')$, the unconditional variance matrix of v_i. This matrix has a special form, meaning that its diagonal elements are $\sigma_c^2 + \sigma_u^2$ and off-diagonal are all equal to σ_c^2. There are additional assumptions that are necessary to ensure efficiency:

RE.3: i) $E(u_i u_i{}'|X_i, c_i) = \sigma_u^2 I_T$ (9)

ii) $E({c_i}^2|X_i) = \sigma_c^2$

If RE.3 fails, the estimator is still consistent, but it is necessary to use a robust variance matrix estimator. It is important to notice that GLS has a feasible counterpart, which in this case is simply the estimation of β considering the weights matrix $\Omega = \frac{\sum \hat{v}_i \hat{v}_i}{N}$, with $\hat{v}_i$ being the residuals of a first stage estimation by OLS of the basic model. The estimation is conducted using a maximum likelihood framework because of its GLS nature.

3.2.3.3. Fixed effects (FE)

The FE model assumes that the intercept varies among each cross-section but is constant over time. Then, the basic idea is that a set of specific intercepts can capture all the unobserved heterogeneity, while the slope related to the partial correlation of each explanatory variable with the dependent variable is the same for all cross-sections.

Because of that, this estimator is also known as dummy variable estimator. Comparing the method described below with including one dummy for each cross-section, both estimations produce the same coefficients, the main difference being the fact that the later is more inefficient as it results in a loss of N degrees of freedom.

Considering its underlying hypotheses, the FE estimator can be applied when hypothesis RE.1.ii is not satisfied. In this sense, if the unobserved effect is correlated with the exogenous variables, the RE estimator will be inconsistent. A possible solution for this failure is to include c_i as an explanatory variable. As this term is not observed, transformations are necessary in order to eliminate it. The FE estimator is also known as the within estimator, due to its usage of time variation within each cross-section to estimate the vector of coefficients β.

There is a main advantage of the FE estimator when compared to the RE estimator: X_i and c_i can be arbitrarily correlated (or $E(c_i|X_i)$ can be any function of X_i). On the other hand, the price to be paid is that X_{it} cannot have time-constant factors, because it is not possible to differentiate their effects from the unobserved heterogeneity's. The basic assumptions of the FE estimator are:

FE.1: $E(u_{it}|X_{it}, c_i) = 0$ (10)

FE.2: $rank\ \left(E\left(\ddot{X}_i{}'\ddot{X}_i\right)\right) = K$

As mentioned above, a transformation is necessary to eliminate c_i. The one explored here is not the unique alternative, but is widely applied. The within transformation is based on averaging equation (4) over time and subtracting it from the original equation:

$$\bar{y}_i = \bar{X}_i\beta + \bar{\delta} + c_i + \bar{u}_i \quad (11)$$

$$\ddot{y}_{it} = \ddot{X}_{it}\beta + \ddot{\delta}_t + \ddot{u}_{it} \quad (12)$$

where for instance $\ddot{y}_{it} = y_{it} - \bar{y}_i$. It is important to notice that the unobserved effect disappears with this transformation, as well as any other time-constant variable present in X_{it}. The following steps are based on a pooled OLS estimation of (12). Additionally, efficiency of the FE estimator requires that:

FE.3: i) $E(u_i u_i' | X_i, c_i) = E(u_i u_i')$ (13)

ii) $E(u_i u_i') = \sigma_u^2 I_T$

These assumptions mean that the error term has a constant variance across time and is serially uncorrelated. The later characteristic is different from the RE case because the original error term in FE does not present the unobserved effect. However, it is possible that the error term u_i shows heteroskedasticity, leading to serial correlation that fades over time.

3.2.3.4. Additional issues

In the case of the first model to be estimated, related to income inequality convergence, there is a potential endogeneity issue in $II_{i,t-1}$, as it can be determined by $(II_{i,t-1} - II_{i,t-2})$, what would violate the strict exogeneity assumption. The solution applied here consists of considering a dynamic model following the Arellano-Bond framework. Therefore, both the lag of the dependent variable as well as $II_{i,t-2}$ will be used as instruments to generate a model with sequential moment restrictions. Estimation will be conducted by GMM.

3.2.3.5. Comparison among estimation methods

In order to compare the two main models discussed above, RE and FE, a Hausman test is applied. The basic hypothesis being tested is whether FE and RE estimates are statistically different or not. This is so because FE is consistent when X_i and c_i are correlated but RE is inconsistent and in the opposite case both are consistent but RE is more efficient. If the estimates are not statistically different, the null hypothesis will not be rejected and RE is recommended. Otherwise FE is more appropriate. The original form of the Hausman test can be written as follows:

$$H = (\hat{\delta}_{FE} - \hat{\delta}_{RE})' \left(\widehat{Avar}(\hat{\delta}_{FE}) - \widehat{Avar}(\hat{\delta}_{RE}) \right)^{-1} (\hat{\delta}_{FE} - \hat{\delta}_{RE}) \quad (14)$$

which is distributed as X^2_M, M being the number of coefficients being tested. As already mentioned, if H_0 is not rejected, RE must be chosen.

4. Data Description

The database used here was obtained from the microdata of the National Household Sample Survey (PNAD), by IBGE. These data are publicly available or can be obtained directly with IBGE.

In the study of income inequality, the main variables are total income of the household head and the share of each component of this income. There are nine different incomes investigated in PNAD, that will be classified here in the following groups:

- ✓ FORMAL LABOR (income obtained by the household head who has a formal job as the main one)
- ✓ INFORMAL LABOR AND OTHER (income obtained by the household head who has an informal job or other precarious jobs as the main one)
- ✓ EMPLOYER (income of the household head who owns an enterprise)
- ✓ OWN BUSINESS (income of the household head who has an own business)
- ✓ RETIREMENT AND PENSION _ MINIMUM WAGE (income obtained through pension and retirement that amounts to one minimum wage – social pension program)
- ✓ RETIREMENT AND PENSION_OTHER (income obtained through pension and retirement that is different from a minimum wage)
- ✓ TRANSFERS+INTEREST+DIVIDENDS (savings interest, dividend, social transfer programs and other incomes)
- ✓ OTHER (rent, donations, work allowances)

Considering all these income sources, the ones that are of particular interest here are "retirement and pension – minimum wage" and "transfers + interest + dividends". They are the main ones related to social transfers in Brazil. However, as the later is composed by different elements and the last two are more related to high income household heads, it will be less likely to identify any specific effect of this group to income inequality reduction.

Moreover, the inequality measure discussed previously (ratio 10/40 – see section 3.1) will be calculated over the total income of the household head (all incomes will refer only to the household head), inside each cohort. Therefore, here income inequality refers to intra-cohort's disparities.

As mentioned previously, an inequality measure can be very sensitive to outliers. In the context of cohort analysis, as each cohort can have a not very large size, it is important to control for these outliers to achieve reliable measures of income dispersion. Table 2 presents relevant descriptive statistics for the cohorts created in this study.

It is important to notice that minimum, average and maximum measures refer to the dispersion of the number of observations among states and years. Moreover, the exclusion of outliers does not seem to raise a significant problem considering that the percentage of observations excluded does not go higher than 15.4% and that even after excluding them, the minimum number of observations in all the sample is 50. Comparing this result with other studies that apply this methodology, this does not seem to raise a problem (see for instance MENEZES et al., 2011).

Table 2. Descriptive Statistics of the State-Age-Cohorts
(the number of observations refers to the samples after excluding outliers).

	Average age		Number of observations in the cohorts among states			% of observations lost excluding outliers	
	1995	2009	Minimum	Average	Maximum	Maximum	Average
cohort 1971 - 1975	22	36	50	424	1,475	15.3%	4.8%
cohort 1966 - 1970	27	41	79	503	1,570	9.4%	5.1%
cohort 1961 - 1965	32	46	105	532	1,624	11.8%	5.5%
cohort 1956 - 1960	37	51	95	493	1,512	12.8%	5.9%
cohort 1951 - 1955	42	56	83	431	1,350	13.4%	5.9%
cohort 1946 - 1950	47	61	68	360	1,103	13.8%	6.7%
cohort 1941 - 1945	52	66	53	287	871	15.4%	6.7%
cohort 1936 - 1940	57	71	52	252	741	14.8%	7.0%

Source: Author's preparation, PNAD microdata.

Table 3 shows that the total sample size for each year does not change much, ranging from 58,600 to 71,334 (without weighting). It is important to remember that this sample size refers to the number of household heads that presented income higher than 0 and without any declaration problem. Because of the fact that PNAD prior to 2004 did not covered rural areas in six of seven states of the North region, all states localized in this region were excluded from this study.

Table 3. Total Sample Size in Each Year of the Study.

Year	Sample size	Year	Sample size
1995	58,600	2003	68,122
1996	57,773	2004	69,646
1997	61,753	2005	70,692
1998	62,155	2006	71,334
1999	63,714	2007	67,622
2001	66,767	2008	66,579
2002	67,471	2009	66,709

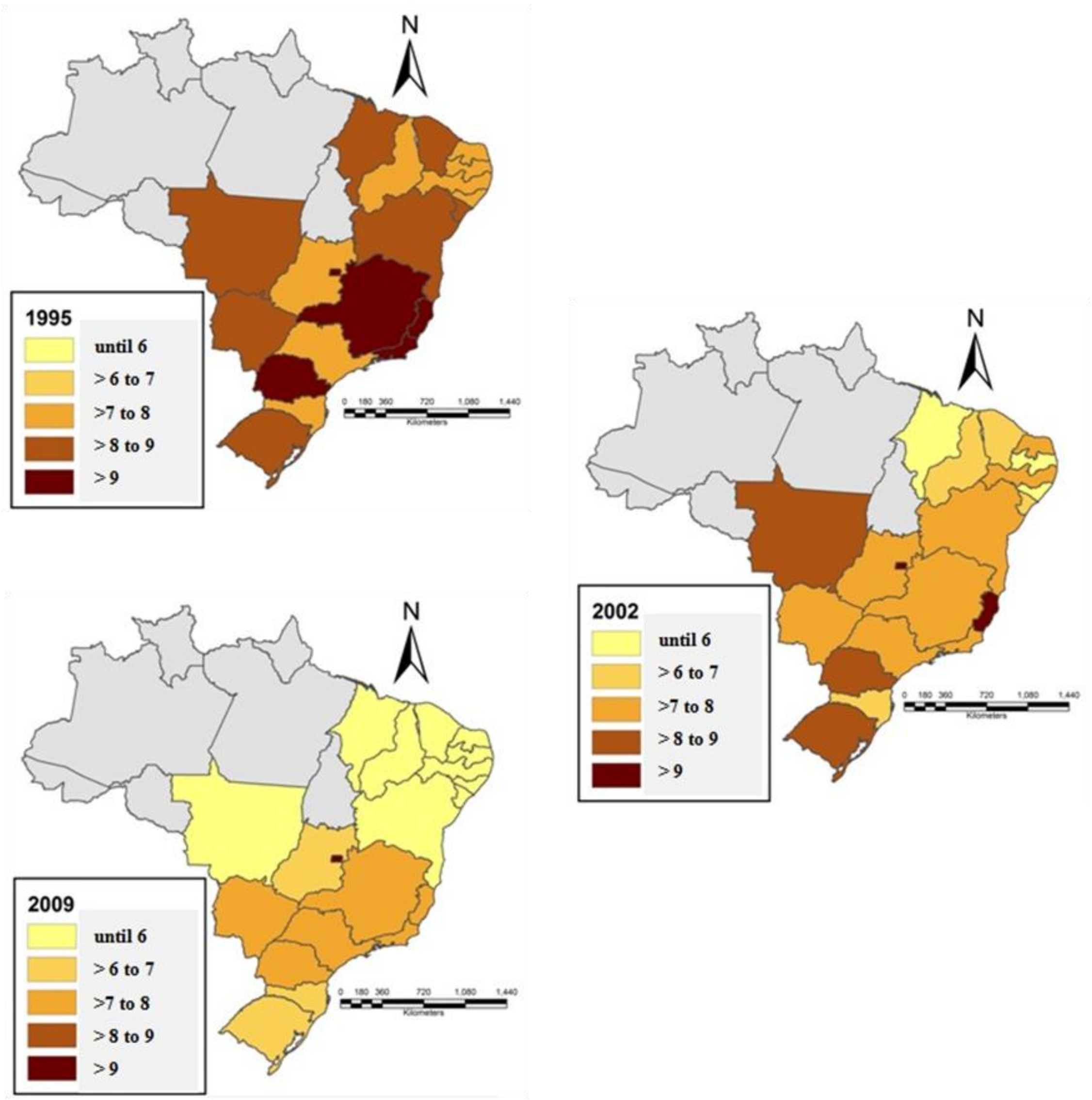
Source: Author's preparation, PNAD microdata.

After describing the database composition process, it is important to understand the patterns that can be observed both in regions as well as in cohorts. Figure 1 presents the evolution of income inequality in the states studied here. A few considerations must be made:

- ✓ This is not a regular income inequality analysis. Therefore, the information represented in the maps refers to the ratio of average income of the upper decile over the average income of the 40% poorest share of the population. Then, this is a measure of intrastate inequality, not interstate;
- ✓ As a consequence of that, the maps above do not show the commonly known pattern of interregional disparities usually described in studies over this subject. In such cases, the information plotted refers to the income level, showing a North-South paradox (North-Northeast poor, South-Southeast rich);
- ✓ With these considerations in mind, it is possible to notice that income inequality has decreased in almost every state. One of the only exceptions is the Distrito Federal, where public administration jobs combined with low-qualification services generates this result;

- ✓ In spite of presenting the lowest income levels in the country, Northeast states do not detach from the others: they are more equal in poverty. The opposite rationality applies to South-Southeast states.

Figure 1. Income Inequality Inside Each State
(ratio of the average income of the richest 10% over the average income of the poorest 40% - household head income) - 1995, 2002 and 2009

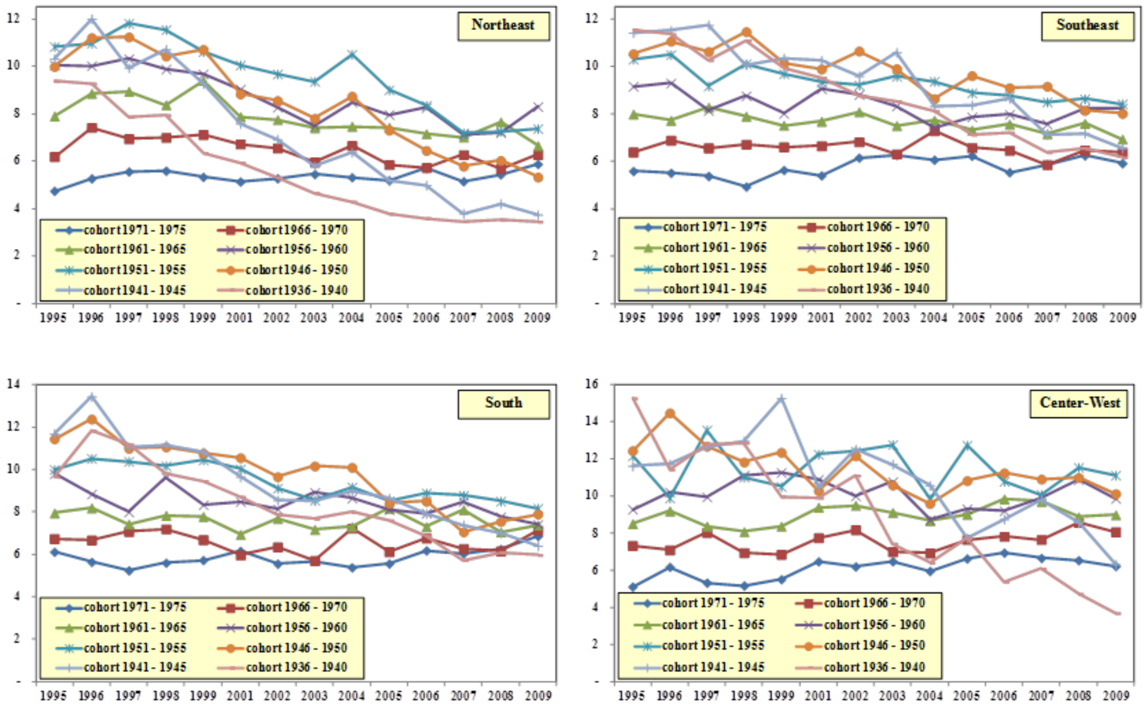


Source: Author's preparation, PNAD microdata.

The cohort-level data allows a more disaggregated analysis, as can be seen in Graph 2:

- ✓ Income inequality seems to have decreased more in older cohorts. Younger cohorts (1966-1970 and 1971-1975) presents a more stable income inequality level;
- ✓ The Northeast shows a stronger reduction in income inequality in the period analyzed, reaching a similar level to the South and the Southeast;
- ✓ The South and the Southeast showed a convergence in the income inequality measure among the cohorts.

Graph 2. Income Inequality (ratio p10/p40) among Cohorts and Regions, 1995-2009



Source: Author's preparation, PNAD microdata.

This brief analysis already allows preliminary and interesting conclusions. The next session will go further in the real measurability of inequality convergence and of the factors that may have caused this process.[10]

[10] As the description of independent variables to be used in the next session would be too extensive for this work, they were not exposed here, but are available under request to the author (ana.barufi@gmail.com).

5. Results

The first model discussed here is related to the literature of inequality convergence presented above. Table 4 brings the main results.

Table 4. Income Inequality Convergence

(dependent variable: $II_{it} - II_{i,t-1}$).

	Pooled OLS				Random Effects				Fixed Effects		Arellano-Bond
$I_{i,t-1}$ (lag of inequality)	-0.193***	-0.201***	-0.336***	-0.361***	-0.193***	-0.201***	-0.336***	-0.361***	-0.595***	-0.687***	-0.994*
year dummies		x		x		x		x		x	
state dummies			x	x			x	x			
constant	1.401***	2.122***	2.032***	2.931***	1.401***	2.122***	2.032***	2.931***	4.605***	6.232***	
N	2,080	2,080	2,080	2,080	2,080	2,080	2,080	2,080	2,080	2,080	1,440
Adjusted R2	0.09	0.11	0.17	0.21					0.24	0.31	
AIC	8,448	8,405	8,265	8,192					7,868	7,673	
BIC	8,460	8,484	8,383	8,378					7,879	7,752	
Log-Likelihood	-4,222	-4,189	-4,111	-4,063					-3,932	-3,822	
F	39.6	11.6	11.0	13.8					821.1	86.0	

note: .01 - ***; .05 - **; .1 - *;

Source: Author's preparation.

In fact, the convergence of income inequality is found in the Brazilian cohorts' context. This result was expected based on the graphs analyzed previously. It is important to notice the robustness of the signal and significance of the coefficient to different model specifications and estimations strategies. The potential endogeneity issue is treated with GMM estimation (Arellano-Bond) and still there is a significant sign of inequality convergence.

Now moving to the regression that aims to explain income inequality levels among the cohorts and years, Tables 5 and 6 show the results.

Table 5. Regression for Income Inequality Level (dependent variable: II_{it}) – omitted: formal jobs

	Pooled OLS		FE		RE	
	coef	se	coef	se	coef	se
Pension & retirement minim. wage	-13.128***	0.462	-14.534***	0.678	-12.141***	0.506
Transfers, interests & dividends	5.553*	3.166	-6.057*	3.228	3.118	3.154
Pension & retirement - other	8.387***	0.343	0.664	0.746	7.596***	0.410
Informal jobs & others	-17.696***	0.966	-10.944***	1.145	-15.691***	1.031
Own business	-1.375*	0.728	-6.204***	0.841	-2.030***	0.747
Employer	16.233***	1.081	10.573***	1.145	15.351***	1.083
Other income sources	1.500	2.339	-8.118***	2.424	-0.592	2.348
Constant	10.545***	0.484	11.218***	0.415	10.550***	0.521
State dummies	x				x	
Year dummies	x		x		x	
Number of observations	2,240		2,240		2,240	

note: .01 - ***; .05 - **; .1 - *;

Source: Author's preparation.

In terms of the regression exposed in Table 5, all signs are expected, except from the one referring to transfers, interests and dividends, which only in the pooled model is significant but has a positive sign (probably the part of income that is related to high income class is dominant in this effect).However when controlling for unobserved effects of each cohort (FE and RE) this coefficient becomes not significant. Another important result is that pension and retirement related to the minimum wage have a negative and significant coefficient, meaning that this component is associated with a smaller level of inequality. On the other hand, other pension and retirement and employer are associated with higher inequality levels. The Hausman test applied for FE and RE in Table 5[11] indicates that the fixed effects estimation seems more appropriated (consistent, while random effects is inconsistent) – see Table 7 for these results.

It is important to notice that in the FE formulation state dummies are omitted in order to avoid multicollinearity with unobserved effects. Controlling for year dummies allows the consideration of events that occurred in specific years as for instance economic recessions and external shocks.

[11] To do this comparison, the RE model was estimated without state dummies (that are omitted in the FE model).

Table 6. Regression for Income Inequality Level (dependent variable: II_{it}) – omitted: formal jobs.

	Pooled OLS	FE	RE
Pension & retirement minim. wage	Pooled OLS	-14.5345***	-13.0879***
Transfers, interests & dividends	8.0863*	-60.5690	-12.9320
Pension & retirement - other	8.3044***	0.6642	4.8784***
Informal jobs & others	-22.3418***	-10.9440***	-14.3289***
Own business	-5.9413***	-6.2043***	-4.9258***
Employer	11.9847***	10.5731***	12.2949***
Other income sources	-0.3393	-8.1182***	-46,049.0000
Region dummies			
Northeast	0.2246		-0.1184
Southeast	-2.3012***		-1.9215***
South	-2.5590***		-2.1243***
Year dummies			
1996	0.6027**	0.6316***	0.6080***
1997	0.3685	0.5181***	0.4213**
1998	-0.1265	0.2245	0.0348
1999	-0.1276	0.3156*	0.0746
2001	-0.7253***	-0.1890	-0.4936**
2002	-0.6869**	-0.0742	-0.4275**
2003	-0.8893***	-0.1470	-0.5736***
2004	-1.1373***	-0.1988	-0.6905***
2005	-1.5039***	-0.5011**	-1.0406***
2006	-1.5884***	-0.4674*	-1.0578***
2007	-1.5146***	-0.4302*	-1.0358***
2008	-1.6975***	-0.4470*	-1.1260***
2009	-1.8896***	-0.6075**	-1.3157***
Constant	12.4521***	11.2182***	11.4748***
N	2,240	2,240	2,240
R2	0.5934	0.4737	
Adjusted R2	0.5892	0.4279	

Source: Author's preparation.

Table 7. Hausman Tests Results*

	Chi2 statistic	Prob.	Model indicated
Regression Table 5	60.53	0.00	FE
Regression Table 6	170.78	0.00	FE

* RE models have been adapted to be compared to FE models
(excluding from them all variables that are omitted in the FE models).

Source: Author's preparation.

In Table 6 similar results are presented, controlling this time for region instead of state. For the Northeast, after controlling for all the other variables, the coefficient is positive but statistically non-significant. In opposition, the South and the Southeast show negative and significant coefficients, meaning that in the former the initial inequality level was smaller than in the later (as discussed earlier) and the reduction of this indicator has been higher in these more developed states.

The main conclusions that can be drawn from these regressions is that there is an indication that transfers (only when considered as pension and retirement connected with one minimum wage) have a significant role to lead to income inequality reduction (negative sign, significant in all regressions). In the opposite direction, private retirement and some labor income components, even when controlling for year idiosyncrasies and cohorts specific effects, are significant and positive, meaning that probably the main gains in labor productivity may have benefited the upper share of the population.

Finally, year dummies capture an aggregated movement of increase in income inequality within cohorts in 1996-1997 and a strong decrease after 2001. Therefore, even after controlling for unobserved effects and household head income composition, there is an underlying movement that affects all cohorts in the same way, increasing inequality at the beginning of the analyzed period and decreasing it in the end. Again the FE model seems to be more suitable for the specification presented in Table 6 (see Table 7).

These results represent only the initial findings of this study. There are significant improvements to be made, which will be explored in the next section.

6. Conclusion

The decrease of income inequality in Brazil in the recent period has been subject of a large number of studies. Most of them explore income inequality decomposition and others discuss income convergence, where the negative sign for initial income stands for income inequality reduction.

In this study a new approach is proposed, based on state-age cohorts aggregation, generating a pseudo-panel. Instead of analyzing average income, a dispersion measure is built and two different methods are applied in order to explain the intra-cohort income inequality. In the first case, it is found an indication of income inequality convergence, robust to different specifications and estimation methods. Therefore, the movement of aggregated income inequality reduction has embedded a decrease of the differences inside each state-age cohort.

The second approach leads to the potential conclusions that, at least with the framework here applied, one of the main income components to reduce inequality was the pension and retirement – minimum wage. The separation of transfers and other incomes is not done here, what may explain why the expected effect of direct social cash transfer programs over inequality is not found here.

Some pitfalls can be pointed: building dispersion measures inside cohorts may be questionable, as they are supposed to be more homogeneous - however, it has been shown here that they definitely present significant income heterogeneity; it is important to understand not only intra-cohorts but also inter-cohorts inequality evolution, what will be done in a next study; other models should be developed in order to account for other factors than the ones considered here. These are the main aspects to be developed in further studies.

References

ATTANASIO, O.; BROWNING, M. "Consumption over the Life Cycle and the Business Cycle", American Economic Review, 85: 1, 118-136, 1995.

AZZONI, C.; MENEZES-FILHO, N.; MENEZES, T. Opening the convergence black box: measurement problems and demographic aspects. United Nations University, Discussion Paper No. 2003/56, 2003.

AZZONI, C. et al. Geography and income convergence among Brazilian States. Inter-American Development Bank, Research Working Paper R-395, 2000.

BALTAGI, B.H. Econometric Analysis of Panel Data. John Wiley & Sons Ltd, 4th edition, Chichester, UK, 2008.

BÉNABOU, R.. Inequality and growth. In: Bernanke, B., Rotemberg, J. (Eds.), National Bureau of Economic Research, Macroeconomics Annual. MIT Press, Cambridge, pp. 11–74, 1996.

BROWNING, M.; DEATON, A.; IRISH, M. A Profitable Approach to Labour Supply and Commodity Demands Over the Life Cycle, Econometrica 53(3):503-44, 1985.

DEATON, A.; PAXSON, C. Mortality, education, income, and inequality among American cohorts. In: WISE, D.A. (Ed.) Themes in the Economic of Aging. University of Chicago Press, pp. 129-170, 2001.

HOFFMAN, R. Mensuração da desigualdade e da pobreza no Brasil. In: HENRIQUES, R. M. Desigualdade e Pobreza no Brasil. Rio de Janeiro, IPEA, 2000.

HOFFMAN, R. Transferências de renda e redução da desigualdade no Brasil e em cinco regiões, entre 1997 e 2005. In: PAES DE BARROS, R. et al (Orgs.). Desigualdade de Renda no Brasil: uma análise da queda recente. IPEA, Brasília, pp.17-40, 2007.

KUZNETS, S. Economic Growth and Income Inequality. The American Economic Review, Vol. 45, No. 1, pp. 1-28, 1955.

MENEZES, T; SILVEIRA-NETO, R.M.; AZZONI, C. Demography and evolution of regional inequality. The Annals of Regional Science, 1, pp.1-13, 2011.

PAES DE BARROS, R.; CARVALHO, M. Desafios para a política brasileira. Texto para Discussão 985, IPEA, 2003.

PEDROZO, E. Efeitos da eligibilidade e condicionalidade do Programa Bolsa Família sobre a alocação de tempo dos membros do domicílio. Tese de Doutorado, FGV, 2010.

RYDER, N.B. The cohort as a concept in the study of social change. American Sociological Review, Vol. 30, No. 6, pp. 843-861, 1965.

SILVEIRA-NETO,R.M.; AZZONI, C. Non-spatial government policies and regional income inequality in Brazil. Regional Studies, Vol. 45, Issue 4, pp. 453-461, 2011.

SOARES, F.V. et al. Programas de transferência de renda no Brasil: impactos sobre a desigualdade. In: BARROS, R.P. et al (Orgs.). Desigualdade de Renda no Brasil: uma análise da queda recente. IPEA, Brasília, pp.87-129, 2007b.

SOARES, S. Análise de bem-estar e decomposição por fatores da queda na desigualdade entre 1995 e 2004. Econômica, Rio de Janeiro, v.8, n.1, pp.83-115, 2006.

SOARES, S. et al. Programas de transferência condicionada de renda no Brasil, Chile e México: impactos sobre a desigualdade.Texto para Discussão 1293, IPEA, 2007a.

SOUZA, P.H.G.F.; OSORIO, R.G. A redução das disparidades regionais e a queda da desigualdade nacional de renda (1981-2009). Texto para Discussão 1648, IPEA, 2011.

WOOLDRIDGE, J.M. Econometric Analysis of Cross-section and Panel Data.The MIT Press, Cambridge, MA, 2002