Economia Aplicada, v. 17, n. 4, 2013, pp. 647-666

EFFECTS OF GROWTH AND REDUCTION OF INCOME INEQUALITY ON POVERTY IN NORTHEASTERN BRAZIL, 2003-2008

Vladimir Faria dos Santos [*]
Wilson da Cruz Vieira [†]

Abstract

This article aimed to investigate the main determinants of individual income inequality and analyze the characteristics of the interrelation among poverty, inequality and economic growth for the states of the Northeast region of Brazil, from 2003 to 2008. For that, we used decompositions based on the Shapley value. The results show that growth was the main factor leading to the drop in the number of poor and indigents (extremely poor). In addition, the most important factors to explain the level of individual income inequality in the Northeastern states are education level and professional experience.

Keywords: Income inequality, poverty, decomposition, Northeast, Brazil.

Resumo

Este artigo teve como objetivo investigar os principais determinantes da desigualdade de renda pessoal e analisar as características da interrelação entre pobreza, desigualdade e crescimento econômico nos estados da região Nordeste do Brasil, de 2003 a 2008. Para tal, foram utilizadas decomposições baseadas no valor de Shapley. Os resultados mostram que o crescimento foi o principal fator que levou à queda do número de pobres e indigentes (extremamente pobres). Além disso, os fatores mais importantes para explicar o nível de desigualdade de renda pessoal nos estados da região Nordeste são nível educação e experiência profissional.

Palavras-chave: Desigualdade de renda, pobreza, decomposição, Nordeste, Brasil.

JEL classification: I32, O15.

[*] Universidade Federal Fluminense (UFF). E-mail: vladi_fs@yahoo.com.br
[†] Universidade Federal de Viçosa (UFV). E-mail: wvieira@ufv.br

Recebido em 21 de setembro de 2012 . Aceito em 11 de setembro de 2013.

1 Introduction

Although the amount of wealth produced by Brazil would be enough to eradicate poverty, the country has a significant number of people who do not have sufficient resources to meet their minimum needs. Wealth concentration, therefore, can be considered as one of the main difficulties in eradicating poverty and misery.

Poverty in Brazil has a regional character, and its incidence is higher in the Northeast (Rocha 2000). In 2008, the proportion of poor in this region was 42%, the highest among the five regions of Brazil.

A high rate of poor people has persisted for decades in the region, despite economic growth. For example, from 1960 to 1969, the GDP expanded at an annual rate of 4.4%. Between 1970 and 1979, the growth rate was 9.4% per year. Between 1980 and 1989, the growth rate fell to 4.3% per year, while between 1990 and 2000, the growth rate decreased to approximately 2.6% (Guimarães Neto (2004)). Although the northeastern economy has been growing since the 1960s, social indicators, particularly poverty, remain at regrettable levels. Growth, therefore, did not provide significant reduction in the number of people below the poverty line and this may be related to the high level of inequality observed in the region.

However, in recent years, there has been a downward trend in the level of poverty in the region. The proportion of poor people has decreased, especially since 2003. Comparison between the years 2003 and 2008 shows a decline of approximately 31%.

This significant decrease can be related to a less unequal income distribution in the Northeast, as the number of people below the poverty line may vary due to changes in economic growth and/or income concentration. Regarding income distribution, it is observed that in the Northeast, there was an improvement between the years 2003 and 2008. This decline, measured by the Gini coefficient, was 4.04%.

Although a less unequal distribution may contribute to poverty decline, growth in earnings is also an important factor in this regard. Between 2003 and 2008, the growth in per capita household income in Northeastern Brazil was approximately 43%, which is the highest among all regions of the country. This significant increase in income suggests that the decline in poverty may be more related to growth in earnings than to a decrease in its concentration.

The main objective of this article was to investigate, for all states in the Northeastern Brazil, the contribution of two phenomena — economic growth and individual income inequality — to the decrease in absolute poverty from 2003 to 2008, and analyze the main determinants of these phenomena.

2 Methodology

Two decomposition methods were used to achieve the proposed objectives, both based on the Shapley value. The first one divides a given change in poverty into two components, growth and distribution (inequality). The second method is applied to an econometric model (regression), specifically in an earning equation.

It is important to highlight that the Shapley value is extremely important for the conducted decompositions. Thus, a brief discussion about this branch of game theory is presented in Appendix A.

2.1 Poverty, growth and inequality: Shapley decomposition[1]

The reduction in poverty is determined by the rate of growth of the average population income and/or its distribution (drop in inequality). There are several methods that seek to quantify the contribution of these two components, including the decomposition based on the Shapley value.

To demonstrate the decomposition, firstly, it is necessary to define a poverty line, which in this case will be denoted by z. Thus, the level of poverty at time t can be expressed as a function, $P(\mu_t, L_t)$, of the mean income (μ_t) and the Lorenz curve (L_t). The growth factor in the variation of poverty between the period t and $t+n$ is denoted by $G = \frac{\mu_{t+n}-\mu_t}{\mu_t} = \frac{\mu_{t+n}}{\mu_t} - 1$, while the distribution factor (or inequality factor), which is related to the Lorenz curve, is given by $D = L_{t+n} - L_t$.

In this article, we adopted the P_α poverty measures developed by Foster et al. (1984). These measures are widely used in the literature on poverty. Next, the variation in the class P_α is expressed by:

$$\Delta P_\alpha = P_\alpha(\mu_t(1+G), L_t + D) - P_\alpha(\mu_t, L_t) = \nu(G, D) \tag{1}$$

The Shapley value relating to the components growth (G) and distribution (D) may be obtained as follows, respectively:

$$\phi_G^S(\nu) = \frac{1}{2}[\nu(G, D) - \nu(D) + \nu(G)] \tag{2}$$

$$\phi_D^S(\nu) = \frac{1}{2}[\nu(G, D) - \nu(G) + \nu(D)] \tag{3}$$

In the absence of growth, G is equal to zero, and the variation in poverty only due to income distribution becomes:

$$\nu(D) = P(\mu_t, L_{t+n}) - P(\mu_t, L_t) \tag{4}$$

On the other hand, assuming that $D = 0$, one has:

$$\nu(G) = P(\mu_{t+n}, L_t) - P(\mu_t, L_t) \tag{5}$$

Using the equations from (1) to (5), we obtain the full expression of the contributions of growth and distribution effects on changes in the P_α classes:

$$\phi_{\alpha G}^S(\nu) = \frac{1}{2}[P_\alpha(\mu_{t+n}, L_{t+n}) - P_\alpha(\mu_t, L_{t+n}) + P_\alpha(\mu_{t+n}, L_t) - P_\alpha(\mu_t, L_t)] \tag{6}$$

and

$$\phi_{\alpha D}^S(\nu) = \frac{1}{2}[P_\alpha(\mu_{t+n}, L_{t+n}) - P_\alpha(\mu_{t+n}, L_t) + P_\alpha(\mu_t, L_{t+n}) - P_\alpha(\mu_t, L_t)] \tag{7}$$

The Shapley-based decomposition has some advantages over the others, such as lack of a residue (the method is accurate) and symmetry between the initial and final periods.[2]

[1] This section is based on (Baye 2006).

[2] For more details on decomposition methods, see (Baye 2006).

2.2 Regression-based decomposition

The decomposition method based on regression allows the inclusion of any explanatory variables, including economic, social and demographic variables (Gunatilaka & Chotikapanich 2009). According to Wan & Zhou (2005), the procedure for its application depends on an earning equation (income generating function), which can be written as:

$$ln(Y_i) = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ji} + \varepsilon_i \tag{8}$$

where Y is the income (or earning), X_{ji} are the explanatory variables and ε is the stochastic error term. The semi-log specification is justified because Y has an approximately log-normal distribution (Wan & Zhou 2005).

The decomposition based on the Shapley value can be explained using the following earning equation:

$$lnY = f(X_1,\ldots,X_k) \tag{9}$$

It must be highlighted that the explanatory variables $(X_1,\ldots,X_k)$ differ among individuals. Thus, the variable X_k is replaced by its sample mean. This replacement eliminates the difference in X_k among the observations (individuals). The expected income generated after this replacement $(\hat{Y_k})$ still varies among individuals, but not due to the variable X_k, as before. In other words, the inequality in the expected variable $(\hat{Y_k})$, which can be measured by any index and generally denoted by $I(\hat{Y_k})$, is caused by the differences in the other variables, and no longer by X_k. Following this rationale, the contribution of X_k for the total inequality (denoted by C_k) can be measured as follows: $C_k = I(\hat{Y}) - I(\hat{Y_k})$, for $k = 1,2,...,K$. The term $I(Y)$ is the measure of inequality achieved when applying any index (for example, the Gini, Theil-T or Theil-L indexes) on the expected variable $\hat{Y}$. In the second stage, besides X_k, the variable X_j $(i \neq j)$ is also replaced by its sample mean. Therefore, as in the first stage, it is necessary to obtain the expected income, $\hat{Y_{kj}}$. The contribution of the second stage or, according to Shorrocks (1999), "second round", is measured by $C_k = I(\hat{Y_j}) - I(\hat{Y_{kj}})$, for k and $j = 1,2,\ldots K$. Similarly, the contribution of the third stage (or "third round") is obtained as follows: $C_k = I(\hat{Y_{ij}}) - I(\hat{Y_{ijk}})$, for i, j and $k = 1,2,\ldots K$ $(i \neq j \neq k)$. The process is completed when all explanatory variables are replaced by their sample mean.

Since equation (9) is of the semi-log type, it is necessary to transform it, after being estimated, into a linear model so that the Shapley decomposition is applied in Y rather than in log (Y).

There are other decomposition methods based on one regression.[3] However, Wan (2004) claims that these methods have limitations, namely: (i) the earning equations are estimated using specific functional forms; (ii) only one measure of inequality is used, the coefficient of variation; and (iii) a large residue can be generated.

[3] See, for exemple, Morduch & Sicular (2002), Fields & Yoo (2000) and Fields (2003).

Heckman procedure

The estimation of the earning equations, using the method of Ordinary Least Squares (OLS), may generate biased coefficients (Heckman 1979, Kassouf 1994). The explanation for this is in the likely sample selectivity, since only individuals who are employed, i.e., who have some income from work, are commonly used in the estimation process. Individuals who do not receive labor income would not be considered in the calculations. To account for this possibility, Heckman (1979) developed a method that allows consistent estimates. Since there is no consensual specification, according to Resende & Wyllie (2006), the variables were defined according to the literature (Kassouf 1994, 1997, Hoffmann & Kassouf 2005, Cirino 2008). So, the selection equation (Probit model) was defined as follows:

$$Z_i = \beta_0 + \beta_1 Sexo_i + \beta_2 Cor_i + \beta_3 Id_i + \beta_4 Id_i^2 + \beta_5 Educ_i + \beta_6 Crian_i + \beta_7 Chefe + \varepsilon_i \quad (10)$$

where

Z_i = binary variable that reflects the activity status of the ith individual, i.e., 1 if it is part of the labor force and 0, otherwise;

$Sexo$ = binary variable for gender, which assumes value 1 for men and 0, otherwise;

Cor = binary variable for race that takes value 1 for white and 0 otherwise;

Id = Age, in tens of years;

Id^2 = Squared age, in tens of years;

$Educ$ = Education in years;

$Crian$ = Number of children in the family aged 0-4 years, and

$Chefe$ = binary variable for head, which takes value 1 for the individual who is the head of the family and 0 otherwise.

By using equation (10), the inverse of the Mills' ratio was calculated, which is inserted in the following earning equation:

$$\begin{aligned} Y_i = &\beta_0 + \beta_1 Educ_i + \beta_2 Id_i + \beta_3 Id_i^2 + \beta_4 Sexo + \beta_5 Cor_i + \beta_6 Ocupl_i + \\ &\beta_7 Ocup2_i + \beta_8 Industry_i + \beta_9 Service_i + \beta_{10} Construction_i + \beta_{11} lambda_i + \varepsilon_i \end{aligned} \quad (11)$$

where Y_i is the natural logarithm of the monthly income from all jobs, $Ocupj$ (j = 1 and 2) are dummy variables that reflect workers' type of occupation, and the base-category is formed by agricultural laborers and workers producing goods and services, repair, and maintenance; $Ocup1$, leaders in general, practitioners of sciences and arts as well as technicians; $Ocup2$, service workers, vendors and trade service providers, members of the armed forces and aides and occupations poorly defined; the dummy variables Industry, Service and Construction indicate workers' activity sector, and the base-category is formed by the agricultural sector; the other variables have been specified.

The variables included in the model (11) were also defined according to the literature (Salardi 2002, Alejos 2003, Cirino 2008, Hoffmann 2000).

2.3 Data

The data used in this study were obtained from the Pesquisa Nacional por Amostra de Domicílios (PNAD) (National Survey by Household Sampling) for 2003 and 2008. The analysis started in 2003 because that was the year, in recent times, when poverty began to decline more sharply.

The poverty line, the monthly income from all jobs and per capita household income were adjusted to reflect 2008 values, using the INPC from IBGE.

PNAD is not a simple random sample but rather a complex one. Therefore, it was necessary to consider the effects of the sample design to prevent estimates of standard errors from being calculated incorrectly.

The Stata 10.1 software system was used to perform most estimates. The measures of the P_α class poverty and the decomposition of poverty changes were calculated using the Distributive Analyse Stata Package (DASP), version 2.1, developed by Araar & Duclos (2009).

The decomposition based on regression (Shapley) involves many calculations, depending on the number of variables inserted into the earning equation, which makes it extremely cumbersome if done manually. Thus, the decomposition was implemented by a Java program developed by the World Institute for Development of Economics Research of the United Nations University (UNU-WIDER).

3 Results

Aiming to favor the understanding of the analysis, the results were separated into two sections. The first presents, for each state in the Northeast region of Brazil, the decomposition of poverty change into two components: economic growth (income) and inequality (distribution). The second section presents and analyzes the determinants of income of workers and the main factors contributing to income inequality in the Northeastern states.

3.1 Decomposition of variation in poverty

Estimates of absolute poverty depend on the selection of a line that separates the poor from the non-poor. There is no official poverty line (PL) in Brazil. So, we decided to use, as Hoffmann (2000) and Helfand et al. (2009), the lines of poverty and extreme poverty based on the minimum wage. Thus, half 2003 minimum wage (MW) was used to define the poverty line and a quarter of it, for extreme poverty. These lines were adjusted to reflect 2008 values. The minimum wage in 2003 was R$240,00, which generated a line of poverty and indigence of R$120,00 and R$60,00, respectively. In 2008, these values corresponded to, respectively, R$154,30 and R$77,15. These lines were used to generate the results of Shapley methodology, which can be seen in Table 1.

In Table 1, considering the Northeast as a whole, it is possible to observe that when income distribution remains constant — that is, at 2003 level — the growth of per capita average household income would have been responsible for a reduction of approximately 15 percentage points in poor people ratio. For extreme poverty, the component "growth" would have reduced the level of poverty, given the constant inequality, in 10 pp. The factor distribution reinforced, although to a lesser extent, the growth effect. In a scenario where economic growth had remained unchanged, the decreased inequality would

Table 1: Decomposition of poverty and extreme poverty (P_0) into two components: growth (ϕ_G^S) and distribution (ϕ_D^S), 2003-2008

States	Poverty			Extreme poverty		
	$\phi_G^S(\nu)$	$\phi_D^S(\nu)$	ΔP_0	$\phi_G^S(\nu)$	$\phi_D^S(\nu)$	ΔP_0
Maranhão (MA)	−0.1387	−0.0679	−0.2066	−0.0921	−0.0512	−0.1434
Piauí (PI)	−0.1773	−0.0175	−0.1948	−0.1105	−0.0433	−0.1538
Ceará (CE)	−0.1635	−0.0323	−0.1958	−0.1053	−0.018	−0.1233
Rio G. do Norte (RN)	−0.1838	−0.0247	−0.2086	−0.0958	−0.0043	−0.1001
Paraíba (PB)	−0.1925	0.0184	−0.1741	−0.1155	0.0201	−0.0953
Pernambuco (PE)	−0.1456	−0.0318	−0.1774	−0.0982	−0.0175	−0.1157
Alagoas (AL)	−0.1289	−0.0469	−0.1758	−0.111	−0.0029	−0.1138
Sergipe (SE)	−0.1003	−0.057	−0.1573	−0.0598	−0.0483	−0.108
Bahia (BA)	−0.1601	−0.0427	−0.2028	−0.1036	−0.0168	−0.1205
Northeast (NE)	−0.1557	−0.0366	−0.1924	−0.103	−0.0179	−0.1209

have reduced poverty by 3.6 pp, and extreme poverty, by 1.7 pp. Therefore, it was observed that the increase in the per capita household income explained approximately 81% of poverty reduction and 85% of extreme poverty reduction in the Northeast. Marinho & Soares (2003), using methodology and period different from that used in this study (1985-1999), decomposed the variation of poverty in the components of growth and concentration (distribution). According to their results, the decline in poverty in the Northeast region was explained entirely by the growth factor (−55.73%). Inequality, unlike what was observed in this study, also played an important role (44.27%), though in order to increase the proportion of poor.

The state of Rio Grande do Norte showed the best result for the drop in the number of poor people (20.86 pp). Most of this reduction was due to the growth in earnings, namely, 88% of the change in the ratio of poor people was due to the factor income. Distribution, however, contributed less significantly, having participated with only 12% of the total variation. This result is strongly associated with a small reduction in income inequality observed in the state. Between 2003 and 2008, the Gini coefficient dropped only 1.9%. In the case of extreme poverty, the same situation was observed, i.e., the growth component explained the largest portion of the variation in poverty (96%).

Maranhão was the second state in the drop in poverty ranking (20.66 pp). The Shapley decomposition showed that income growth was primarily responsible for this reduction. In other words, if inequality had remained constant, poverty would have fallen — as a result of rising earnings — 13.87 pp, which is 67% of the drop in the P_0 index. The contribution of the redistributive factor was lower, i.e., 33% reduction in the proportion of poor people resulted from the decreased inequality degree. With regard to extreme poverty, income distribution had a slightly better performance, and was responsible for 36% of the poverty change, while the factor growth contributed with 64%.

Orair (2006) quantified the contribution of economic growth (income) and income inequality to poverty change, using a methodology different from that applied in this article. This author used two poverty lines and several subperiods for all states of Brazil, from 1992 to 2004. Considering Maranhão and the period between 1993 and 1995, its poverty change being similar to that observed in 2003/2008 (−0.20128), Orair (2006) found estimates similar to those obtained in this article. Income growth explained 64% of the change in the proportion of poor people, while inequality was responsible for 36%.

Sergipe presented the second largest drop in the index of inequality, according to the Gini coefficient (6.5%). This decline led the component distribution to have greater participation in the explanation of the decline of 15.73 percentage points in the proportion of poor. In other words, reduced distributive inequality contributed with 36.2% of the variation of poverty. The importance of lessening the degree of inequality is greater for extreme poverty. Approximately 45% of the variation in extreme poverty was explained by improved income distribution. The contribution of the growth factor is relatively small compared to other states. One explanation for this would be the lower income growth (27%), considering, of course, the rates seen in other states of the Northeast region, with average of 43.3%.

Another point that should be mentioned is related to Paraíba. According to Table 1, poverty could have fallen even more if there had been some distribution of income, i.e., in a situation without change in inequality, the number of poor would fall 19.25 pp. On the other hand, if earnings remained constant, poverty would have increased — due to increased inequality — 1.8 pp. For indigence (extreme poverty), the result is the same, and only the growth factor is responsible for the decreased number of people below the extreme poverty line.

The results showed that the growth factor was mainly responsible for the fall in the number of poors and indigents. It is important to note that the rdpc is composed of various sources of income, namely labor income, pensions, rents and donations, interest income, cash transfer (Bolsa Família, for example) and other incomes sources. Among these categories, several studies have highlighted the importance of the Bolsa Família Program in combating poverty. As an example, Soares et al. (2006) emphasized that, without direct cash transfer programs, poverty and income inequality in Brazil would hardly fall to tolerable levels within a relatively short period of time.

Another issue that must be highlighted refers to labor income. This source of income is extremely important to explain the growth of earnings of the population in the Northeast region and, consequently, to poverty. For Santos (2011), labor income was the main component of household income in the Northeast, representing 72% of the total. Moreover, in the Northeast region, labor income explained 70% of the increase in household income, whose growth had been of 43.33%.

The importance of labor income in the composition of the household income is also observed when considering the country as a whole. For example, Soares et al. (2006) found, for the year 2004, the proportion of 72.6%, which shows the relevance of the work to the total income. Given this, the main determinants of income from work are presented below.

3.2 Decomposition based on regression

Selectivity bias is one of the problems that may occur in the process of estimating earning equations.[4] The Heckman procedure was used to solve this

[4]Another problem that may arise during the estimation of earning equations is the endogeneity of the education variable. In order to solve this possible problem, the equations of income were estimated using the two-stage least squares method, with the following variables as instruments: provision of elementary schools and the law 5692 of 1971. By Durbin-Wu-Hausman test, it was not possible to reject the hypothesis that the variable education is correlated with the error term, indicating that the model estimated by the Heckman procedure is the most appropriate.

problem. The estimates of earning equations are shown in Table 2.[5]

Table 2 shows that the inverse of the Mills' ratio (lambda) is statistically significant (at 1%) in all states of the Northeast, which indicates that their inclusion was necessary to avoid selectivity bias.

In most cases, the coefficients of the equations are statistically significant and have the expected signals, with few exceptions.

These results demonstrate that the variable years of education (Educ), whose estimates are highly significant, has the ability to positively impact the income received from work. This observation is valid for all states in the Northeast, which confirms the benefits of education to the income of individuals. Estimates of the variable Educ can also be interpreted as the education return rate. Thus, for each additional year of study, in the Northeast as a whole, there is an increase of 8.58% in income.[6] Ceará and Bahia are the states that present the best return rates, with values above 8.8%. With regard to Bahia, the coefficient found is very similar to that verified in Lacerda (2008), whose value was 0.0873.

Maranhão is the state where the education return rate presents the lowest level, showing that increased income, provided by each additional year of education, is lower than that observed in other states of the Northeast region.

With regard to age, a proxy for experience, it is observed that the coefficient associated with the variable age is positive, but that related to age squared is negative. This shows that the relationship between experience and income have inverted U behavior. In other words, as the individuals gain more experience in their work environment, earnings tend to increase. However, upon reaching a certain point, income begins to decrease. Considering the Northeast as a whole, when income is maximized, the age would be around 55 years.[7] In Brazil, it would be, according to Kassouf (1994), 50 years for men and approximately 46 years for women. According to the estimates of Hoffmann (2000), for the year 1997, the age that would maximize the Brazilians' income would be around 50 years.

The variable gender, similarly to years of education, has a positive sign, which was expected. The results show that, for all states of the Northeast, men earn higher incomes than women. Considering the whole region, men receive on average 62.2% more than women. On the other hand, Alagoas has a much lower (though high) differential. The difference is around 33%. Ueda & Hoffmann (2002) also found high estimates for the gender coefficient for Brazil. The values obtained by these authors for 2002 ranged from 0.4415 to 0.4506, depending on the model. This suggests that, in Brazil, the income gap between men and women is also high.

Regarding the variable race, the signal is displayed in accordance with what was expected and, in most states, the coefficients are statistically significant at 1% level. Then, as presented in Table 2, the income gap between white people and others (black, brown, yellow and Indian) is not as great when compared to the variable gender. The results suggest that income discrepancy in

[5]The participation equations, used in the Heckman procedure, can be seen in Table B.1 in the Appendix B.

[6]To avoid any inaccuracy, Wooldridge (2006) suggested using the formula $100[exp(x)-1]$ to calculate the percentage increase in income generated by each additional year of education, where x is the estimated coefficient of the variable education.

[7]This result was obtained by simply deriving the earning equation related to age and equating it to zero. Since age is in decades, the value found was multiplied by ten.

Table 2: Equations of income for the states of the Northeast region, Brazil, 2008

Variables	MA	PI	CE	RN	PB	PE	AL	SE	BA	NE
Cons.	3.8432* (0.2978)	3.0763* (0.3697)	3.1347* (0.1739)	4.0239* (0.6491)	3.6189* (0.5644)	3.3389* (0.2139)	3.8387* (0.3064)	4.0931* (0.5068)	3.1884* (0.1471)	3.3923* (0.1071)
Educ.	0.0652* (0.0067)	0.0713* (0.0107)	0.0899* (0.0036)	0.0793* (0.0095)	0.0808* (0.0103)	0.0811* (0.0042)	0.0769* (0.0064)	0.0852* (0.0091)	0.0889* (0.0037)	0.0823* (0.0021)
Id	0.4336* (0.1037)	0.4722** (0.1420)	0.5996* (0.0776)	0.3007 (0.2287)	0.4948* (0.1861)	0.6835* (0.0784)	0.4999* (0.1081)	0.327*** (0.1872)	0.7184* (0.0536)	0.5918* (0.0397)
Id^2	−0.0348* (0.0115)	−0.0402* (0.0162)	−0.0571* (0.0091)	−0.0162 (0.0264)	−0.0405*** (0.0224)	−0.0669* (0.0097)	−0.0383* (0.0119)	−0.0175 (0.0237)	−0.0669* (0.0064)	−0.0538* (0.0047)
Sexo	0.5118* (0.0599)	0.6261* (0.0705)	0.5755* (0.0493)	0.4181* (0.0799)	0.3661* (0.0825)	0.4226* (0.0357)	0.2863* (0.0608)	0.4345* (0.1018)	0.4932* (0.0283)	0.4839* (0.0211)
Cor	0.1599* (0.0352)	0.0852 (0.0852)	0.1381* (0.0222)	0.1175* (0.0368)	0.0241 (0.0389)	0.1485* (0.0249)	0.1046*** (0.0546)	0.0926** (0.0389)	0.1259* (0.0273)	0.1223* (0.0123)
Occup1	0.5427* (0.0729)	0.8395* (0.0939)	0.6224* (0.0635)	0.5434* (0.1013)	0.5233* (0.0826)	0.497* (0.0536)	0.6046* (0.0785)	0.5652* (0.0887)	0.4754* (0.0485)	0.5529* (0.0259)
Occup2	0.0773* (0.0538)	0.1845*** (0.0472)	0.0917 (0.0554)	0.0223 (0.0621)	0.0457 (0.0555)	−0.0324 (0.0324)	0.0325 (0.067)	0.0782 (0.084)	−0.0375 (0.0272)	0.0315*** (0.0938)
Industry	0.3204* (0.1053)	0.6052* (0.1407)	0.2211* (0.0817)	0.3849* (0.0664)	0.2894*** (0.1546)	0.2777* (0.0621)	0.3262* (0.1109)	−0.0699 (0.1663)	0.2555* (0.0595)	0.2514* (0.0397)
Service	0.3665* (0.0907)	0.7038* (0.1096)	0.3783* (0.0522)	0.4587* (0.0764)	0.4482* (0.1151)	0.2569* (0.0541)	0.23 (0.1189)	0.2122** (0.0974)	0.2254* (0.0482)	0.3373* (0.0275)
Construction	0.3427* (0.0729)	0.6865* (0.1492)	0.4937* (0.0586)	0.4379* (0.0856)	0.5214* (0.1198)	0.2321* (0.0555)	0.1556* (0.1142)	0.2822* (0.0667)	0.2406* (0.0532)	0.3602* (0.0288)
Lambda	−0.3464* (0.0478)	−0.37874* (0.1001)	−0.3347* (0.0525)	−0.5102* (0.1513)	−0.4133* (0.1302)	−0.1606* (0.0607)	−0.2981 (0.0663)	−0.4136* (0.1343)	−0.1966* (0.0463)	−0.2937* (0.0291)

Note: Educ.= years of education, Occup1 = professionals/technicians; Occup2 = workers who have graduated from high school; Occup. (base) = Workers (blue collars). * significant at 1%; ** significant at 5%; *** significant at 10%. The standard-deviations (linearized) are between brackets. The values which have not been marked are not significant.

Maranhão, in terms of race, is the highest in the region. In that state, white people receive, on average, 17.34% more than individuals in the base category. In the Northeast as a whole, this difference was 13%.

Discrimination in the labor market occurs when individuals or groups with equal average qualifications (skills) do not receive the same mean income (Lundberg & Startz 1983). Therefore, income inequalities observed between genders and between races, after assessing the other variables, suggest that there is some kind of discrimination in the labor market in the Northeast region of Brazil.

The variables $Ocup1$, and $Ocup2$ and $Ocup3$ (base) were inserted into the model to analyze the effects of allowing the analysis of the type of workers' occupation on income. According to the results, the sign of the variable Ocup1 — which refers to managers in general, practitioners of sciences and arts and technicians — is in line with what was expected. In other words, workers in this group have incomes well above those that form the base group, the manual laborers (blue collars). For the Northeast as a whole, the group Ocup1 receives on average 73.8% more than workers of the base category.

The income of mid-level workers ($Ocup2$) is higher than that observed in $Ocup3$, except for Pernambuco and Bahia. Piauí has the highest coefficient related to variable $Ocup2$, suggesting that the income of workers that occupy intermediate positions is, on average, 20.3% higher than that found in the base category.

According to the theories of compensating wage differentials and efficiency wages, income inequality may be partially explained by the characteristics peculiar to each economic activity. By virtue of being controlled, in earning equations, the effects of the variables education, age, gender, race and professional activity, the estimates of the variables $Ocup1$ and $Ocup2$ may reflect what is established by the mentioned theories, i.e., the advantages or disadvantages existing in jobs (theory of compensating differentials) as well as the internal policies adopted by some firms that set wages higher than in the labor market (efficiency wages theory) may be important to determine the salary of workers.

The regressors Industry, Service, Construction and Agriculture (base) were included in the model to analyze the impacts of working in their respective sectors on income. All these variables have expected signs and were statistically significant, except for the variable industry in the state of Bahia. According to estimates, the occupations in industry, service and construction generate income well above that recorded in the base category (agriculture). Given this and based on the dual labor market theory, which assumes the existence of two market segments, primary and secondary, we can say that agriculture in the Northeast is the industry that best fits the secondary segment, which, according to the dual labor market theory, is characterized by low wages, lack of stability and career advancement and higher manpower turnover.

It is important not only to identify the main determinants of people's income, but also to know which factors explain income inequality. Thus, the contributions of each variable to income inequality were calculated by earning equations (Table 2). To calculate the levels of inequality, the three main measures were used, namely, the Gini index and the two indices of the generalized entropy class: Theil-L and Theil-T. Similar to the work of Gunatilaka & Chotikapanich (2009), this study also used some aggregate variables. Therefore, inequality was decomposed into six components: Education ($Educ$), Age

(age and age squared), Gender, Race (Color) and Occupation (*Ocup*1 and *Ocup*2). The results are presented in Table 3.

Table 3: Decomposition (Shapley) based on the earning equation estimated by the Heckman procedure (%), 2008

Variables	Gini Coefficient									
	MA	PI	CE	RN	PB	PE	AL	SE	BA	NE
Educ	17.58	22.36	20.53	22.66	21.53	19.68	23.01	25.08	20.05	20.49
Age	18.86	18.78	21.78	14.66	19.26	22.33	19.33	15.87	24.99	22.24
Gender	6.5	1.67	8.01	4.15	2.95	5.19	1.72	6.13	6.76	6.12
Color	2.07	1.43	1.74	1.7	0.31	2.21	1.47	1.12	1.23	1.53
Occupation	10.4	12.94	11.58	11.39	10.17	10.6	13.32	13.02	8.72	10.55
Sector	9.7	5.03	7.14	9.76	9.00	4.3	5.18	5.28	4.26	6.55
Residue	34.88	37.79	29.21	35.69	36.77	35.69	35.98	33.5	33.99	32.51
Total	100	100	100	100	100	100	100	100	100	100
	Theil Coefficient - L									
Educ	12.31	12.59	16.07	15.36	14.99	14.44	15.74	17.74	14.45	15.28
Age	6.25	4.95	7.5	4.67	6.00	7.34	6.18	5.5	10.11	7.97
Gender	3.35	– 0.15	4.83	1.09	0.49	2.45	– 0.18	2.58	3.88	3.11
Color	1.04	0.58	0.94	0.75	0.15	1.2	0.73	0.44	0.68	0.8
Occupation	7.73	7.85	9.17	7.98	7.36	8.29	9.81	9.13	7.00	8.37
Sector	5.89	2.09	3.69	5.2	4.78	2.1	2.62	2.95	2.00	3.6
Residue	63.43	72.09	57.79	64.94	66.24	64.18	65.11	61.65	61.88	60.86
Total	100	100	100	100	100	100	100	100	100	100
	Theil Coefficient - T									
Educ	10.77	11.66	13.33	14.06	12.54	10.74	13.51	15.85	11.08	12.49
Age	6.76	5.52	8.54	5.29	6.3	6.5	6.39	5.76	9.95	8.66
Gender	2.96	– 0.02	4.45	1.08	0.64	2.14	– 0.02	2.25	3.51	2.85
Color	1.02	0.57	0.94	0.73	0.13	1.06	0.66	0.46	0.69	0.78
Occupation	7.47	8.05	8.82	8.27	6.85	7.28	9.33	9.00	6.59	7.96
Sector	4.54	1.69	2.44	4.12	3.35	1.22	1.96	2.34	1.09	2.33
Residue	66.49	72.53	1.48	66.47	70.18	71.07	68.17	64.34	67.09	64.93
Total	100	100	100	100	100	100	100	100	100	100

According to the results, when the Gini coefficient is used, the models explain between 62.2 (Piauí) and 70.8% (Ceará) of total inequality, which shows the good explanatory power of the decomposition method. According to Wan & Zhou (2005), models that explain only 30 to 40% of inequality, leaving the rest to the residue, cannot provide reliable estimates.

The results obtained by the Gini index reveal that, in five states (Piauí, Rio Grande do Norte, Paraíba, Alagoas and Sergipe), education was primarily responsible for income inequality, explaining at least 21% of it. In the Northeast region as a whole, the Gini coefficient in 2008 was approximately 0.55. Out of this total, education contributed with 20.49%.

Maranhão was the state with the lowest value for the contribution of education to inequality. This result may be related to the rate of return to education in that state. Table 2 shows that Maranhão has the lowest coefficient in the Northeast for the variable years of education, which generated the lowest return rate of one additional unit of study. According to Ney (2006), one of the reasons for the explanatory power of education on income inequality is the rate of return to schooling.

Age is another important variable to explain inequality in the Northeast, a proxy for experience. Table 3 shows the contribution of age using the Gini index, which ranged from 14.7 (Rio Grande do Norte) to 25% (Bahia). In the Northeast region as a whole, the variable experience was mainly responsible for the high Gini index, contributing with approximately 22%.

Overall, the results related to education and age are consistent with the human capital theory, which establishes the link between income inequality and qualification of individuals. The positive relationship between education and inequality, observed in Table 3, is justified by the fact that most educated people (with more years of education) tend to earn higher incomes. Nevertheless, depending on where one lives and on financial conditions, not everyone has the same access to education.

The variables that characterize the jobs, joined in occupation (Table 3), were included in the analysis to capture the likely effects established by the compensating wage differentials and efficiency wages theories. According to these theories, the earning differentials can be explained both by the advantages and disadvantages present in each job (compensating wage differentials) and by the wages above the market price (efficiency wages theory) that firms pay to their employees. For these theories, the differences between jobs were extremely important in explaining income inequality in the Northeast region of Brazil. The results demonstrate that these variables accounted, on average, for 10.55% of the Gini index, i.e., the group is ranked as the third largest contributor to inequality in the region. In Alagoas, the group Occupation explained 13.32% of the Gini coefficient.

The variables related to workers' sector of activity, which were included to take into account the theory of dual labor market, had the ability to explain 6.55% of the Gini index, which is a relatively significant value. To some extent, these results agree with the theory of labor market segmentation, and agriculture is the sector that best fits the profile of the secondary segment. The difference between the sectors can explain part of income inequality observed in the Northeast.

The variables gender and race show that discrimination is a major factor in income inequality. In the Northeast, these two variables accounted for 7.65% of the distributive inequality, measured by the Gini coefficient. However, gender discrimination is more pronounced, and is responsible for 6.12% of total inequality. While in Piauí the income of men was approximately 87% higher than women's, the contribution of gender to the Gini coefficient was not high (1.67%), compared to other states. On the other hand, Alagoas had the lowest income gap between the genders (Table 2) and this difference was not so relevant to explain income inequality.

The decomposition was also conducted using the Theil-L and Theil-T indices (Table 3). As observed, the share of inequality that was not explained by the variables included in the earning equation — i.e., the residue — was much greater than that observed when the Gini coefficient was used. This difference in the residues was also found in Wan & Zhou (2005). These authors used the Gini and Theil-L coefficients in the decomposition process. In all their estimates, the residue generated when the Theil-L coefficient was used as the inequality measure was much greater than that observed with the use of the Gini index.

In spite of the large amount of residues observed when the Theil-L and Theil-T coefficients are used, the conclusions about which variables contributed

most to explain the inequality have not changed, i.e., the variables related to human capital remain the leading cause of income inequality. Differences in values, observed when different measures are used, can be explained, according to Wan & Zhou (2005), by the sensitivity of each index in the different segments of the Lorenz curve and by the use of different functions of welfare. For example, for Hoffmann (1998), the Theil-T coefficient is more sensitive to changes in income located at the right end of the distribution, namely, the relatively richer. On the other hand, the Theil-L index is more sensitive to changes at the left end of the distribution (the poorest). Finally, the Gini coefficient is more sensitive to changes in incomes located at the interval with the greatest frequency density, i.e., around mode or median of the distribution.

4 Conclusions

The results above lead to some conclusions. The first is that the per capita household income in the Northeast region of Brazil was primarily responsible for the decline in the number of poor, considering the years 2003 and 2008. These results are strongly related to the significant growth in income in the period observed. Although income growth has been fundamental to reduce the number of people below the poverty line, it cannot be said that it is more important than the reduction in inequality in fighting poverty, especially when it comes to the Northeast, where the gap between rich and poor is very significant. The difference for income inequality among the Northeast and the other Brazilian regions is extremely high, suggesting that inequality has much to fall. Therefore, given the relationship between income distribution and poverty reduction, the number of poor and indigent may be reduced even more if income distribution is intensified.

In a scenario where inequality may fall more heavily, it is very important to know the main variables that can influence it. In this context, it can also be concluded that education in the Northeast was primarily responsible for the high inequality in the region. In other words, the lack of skills and low levels of education are the main factors that led the Northeastern states to have the worst income distribution in the country. Considering the prospect that the Brazilian economy will continue to grow, investment in the expansion and improvement of education will be crucial to significantly reduce the gap between those at the bottom and those at the top of the distribution. On the contrary, we may have the same condition as the one observed in the 1960s, when economic growth was accompanied by increasing inequality. At that time, there was a great demand for skilled labor, but there was not enough supply; therefore, those who were educated (higher education level) began to earn large incomes, thus worsening income distribution.

Besides the human capital variables, those related to gender and race also contributed positively to income inequality, corroborating what the theory of discrimination in the labor market states. Thus, it can be concluded that there may be some form of discrimination in the Northeast (by employers), especially against women. The difference in income between men and women is substantial — much more pronounced than the differences in income between races. In Ceará, this problem is more pronounced. Discrimination was one of the factors that explained most income inequality in the state. In this sense, it would be very important to adopt education policies or programs (curric-

ula) aimed at discussing with children the issue of prejudice, to prevent the appearance of any type of discrimination in future generations. Moreover, public policies — or laws providing for stiffer penalties — against discrimination would also be extremely important at reducing income concentration in the Northeastern states.

Bibliography

Alejos, L. (2003), Contribution of the determinants of income inequality in guatemala, Instituto de investigaciones económicas y sociales. Universidad Rafael Kandívar: Guatemala.

Aliprantis, C. D. & Chakrabarati, S. K. (2000), Games and decision making, Oxford university press.

Araar, A. & Duclos, J. Y. (2009), Dasp: Distributive analysis stata package, Pep & cirpÉe e world bank, laval university.

Baye, F. M. (2006), 'Growth, redistribution and poverty changes in cameroon: a shapley decomposition analysis', *Journal of African Economics* **15**(4), 543–570.

Cirino, J. F. (2008), Participação feminina e rendimento no Mercado de trabalho: análise de decomposição para o Brasil e as regiões metropolitanas de Belo Horizonte e Salvador. 188f. Tese (Doutorado em Economia Aplicada), PhD thesis, Departamento de Economia Rural, Universidade Federal de Viçosa, Minas Gerais.

Fields, G. (2003), 'Accounting for income inequality and its change: a new method, with application to the distribution of earnings in the united states', *Research in Labour Economics* pp. 1–38.

Fields, G. & Yoo, G. (2000), 'Falling labour income inequality in korea's economic growth: patterns and underlying causes', *Review of Income and Wealth* pp. 139–159.

Foster, J., Greer, J. & Thorbecke, E. (1984), 'A class of decomposable poverty measures', *Econometrica* **52**(3), 761–766.

Guimarães Neto, L. (2004), O nordeste, o planejamento regional e as armadilhas da macroeconomia., *in* 'Desigualdades Regionais', Superintendência de Estudos Econômicos e Sociais da Bahia, pp. 153–175.

Gunatilaka, R. & Chotikapanich, D. (2009), 'Accounting for sri lanka's expenditure inequality 1980-2002: regression-based decomposition approaches', *Review of Income and Wealth* **55**(4), 882–905.

Heckman, J. J. (1979), 'Sample selection bias as a specification erro', *Econometrica* **47**(1), 153–161.

Helfand, S. M., Rocha, R. & Vinhais, H. E. F. (2009), 'Pobreza e desigualdade de renda no brasil rural: uma análise da queda recente', *Pesquisa e Planejamento Econômico* **39**(1), 67–88.

Hoffmann, R. (1998), Distribuição de renda, medidas de desigualdade e pobreza, Edusp, São Paulo.

Hoffmann, R. (2000), *Mensuração da desigualdade e da pobreza no Brasil*, number 3, IPEA, chapter Desigualdade e pobreza no Brasil.

Hoffmann, R. & Kassouf, A. L. (2005), 'Deriving conditional and unconditional marginal effects in log earnings equations estimated by heckman's procedure', *Applied Economics* **37**(11), 1303–1311.

Kassouf, A. L. (1994), 'The wage rate estimation using the heckman procedure', *Revista de Econometria* **14**(1), 89–107.

Kassouf, A. L. (1997), 'Retorno à escolaridade e ao treinamento nos setores urbanos e rural do brasil.', *Revista de Economia e Sociologia Rural* **35**(2), 59–76.

Lacerda, F. C. C. (2008), Desigualdade de rendimentos na bahia: estimação de equações de rendimentos com base nos microdados da pnad 2005, *in* 'IV Encontro de Economia Bahiana', Salvador, BA.

Lundberg, S. J. & Startz, R. (1983), 'Private discrimination and social intervention in competitive labor market', *The American Economic Review* **73**(3), 340–347.

Marinho, E. & Soares, F. (2003), Impacto do crescimento econômico e da concentração de renda sobre a redução da pobreza nos estados brasileiros, *in* 'XXXI Encontro Nacional de Economia', Vol. 31, ANPEC, Porto Seguro.

Montet, C. & Sierra, D. (2003), Game theory and economics, Palgrave macmillan. 487.

Morduch, J. & Sicular, T. (2002), 'Rethinking inequality decomposition, with evidence from rural china', *Economic Journal* pp. 93–106.

Ney, M. G. (2006), Educação e desigualdade de renda no meio rural brasileiro, PhD thesis, Instituto de Economia da Unicamp, Campinas, São Paulo. 117f. Tese (Doutorado em Economia Aplicada).

Orair, R. O. (2006), Como crescimento e desigualdade afetam a pobreza?, Master's thesis, Instituto de economia da UNICAMP, São Paulo. 125f. Dissertação (Mestrado em Economia).

Osborne, M. J. & Rubinstein, A. (1994), A course in game theory, Cambridge, ma.

Resende, M. & Wyllie, R. (2006), 'Retornos para educação no brasil: evidências empíricas adicionais', *Economia Aplicada* **10**(3), 349–365.

Rocha, S. (2000), Pobreza e desigualdade no brasil: o esgotamento dos efeitos distributivos do plano real, Texto para discussão, Instituto de Pesquisa Econômica Aplicada. N^o 721.

Salardi, P. (2002), How much of brazilian inequality can explain? an attempt of income differentials using pnad 2002. 2005, Acesso em: 06 dez. 2009.
URL: *http://www-3.unipv.it/webdept/prin/workpv05.pdf*

Santos, V. F. (2011), Efeitos do crescimento e redução da desigualdade de renda na pobreza da região nordeste do Brasil - 2003-2008, PhD thesis, Universidade Federal de Viçosa, Viçosa - MG. 140f. Tese (Doutorado em Economia Aplicada.

Shapley, L. S. (1953), *A value for n-person game*, Vol. 28, Ann. Math. Stud., chapter Contributions to the theory of games II, pp. 307–317.

Shorrocks, A. F. (1999), Decomposition procedures for distributional analysis: a unified framework based on the shapley value, University of essex. Mimeogr.

Soares, F. V., Soares, S., Medeiros, M. & G., O. R. (2006), Cash transfer programmes in brazil: impacts on inequality and poverty. Working Paper n.21, IPC-IG.

Ueda, E. M. & Hoffmann, R. (2002), 'Estimando o retorno da educação no brasil.', *Revista Economia Aplicada* **6**(2), 209–238.

Wan, G. (2004), 'Accounting for income inequality in rural china: a regression-based approach', *Journal of Comparative Economics.* **32**(2), 348–363.

Wan, G. & Zhou, Z. (2005), 'Income inequality in rural china: regression-based decomposition using household data', *Review of Development Economics* **9**(1), 107–120.

Wooldridge, J. M. (2006), *Introdução à econometria: uma abordagem moderna*, Thompson Learning, São Paulo.

Appendix A Shapley value

Shapley (1953) introduced one of the most used solution concepts in cooperative game theory. This concept (value), according to Montet & Sierra (2003), has the ability to summarize, in a coalition game, the complex possibilities faced by economic agents (players) in a single number, called the Shapley value. Therefore, in an allocation problem, the Shapley value allows or suggests the division of profits or common costs arising from the grand coalition. Let $N = \{1, 2, \ldots, m\}$ be a finite set of players. Then, a nonempty subset of N is called a coalition.[8]

For each coalition $S \subseteq N$, $v(S)$ is a value, either positive or negative, which is available for division among members of the coalition S.[9] This is carried out based on the average marginal contribution of each player to the coalition.

According to Osborne & Rubinstein (1994), the marginal contribution from player i to any coalition S, with $i \notin S$ in a given game, can be measured as follows:

[8]The method developed by Shapley can be applied to various allocation problems, such as water resource management, allocation of taxes, landing fees at airports, etc. (Aliprantis & Chakrabarati 2000).

[9]In other words, v is a characteristic function that associates to each subset S a real number $v(S)$.

$$\Delta_i(S) = v(S \cup \{i\}) - v(S) \tag{12}$$

Thus, the Shapley value of the player i, denoted by $\phi_i^S(v)$ is given by:

$$\phi_i^S(v) = \frac{1}{|N|!} \sum_{R \in \Re\mathfrak{c}} \Delta_i(S_i(R)), \text{ for each } i \in N \tag{13}$$

where $\Re\mathfrak{c}$ denote the set of all orders of N and $S_i(R)$ denote the set of the players preceding i in the order R. By conversion, $0! = 1$ and $v(\varnothing) = 0$.

Intuitively, the Shapley method can be understood by considering, for simplicity, three players: 1, 2 and 3. Therefore, we have $N = \{1,2,3\}$. Thus, the number of permutations that can be made from the set N is 6, that is, $N! = 6$. These permutations or orderings can be observed in the first column of Table A.1.

Table A.1: Shapley value in a game with three players

Permutations	Marginal contribution of players		
	Player 1	Player 2	Player 3
(1,2,3)	$v(1) - v(\varnothing)$	$v(1,2) - v(1)$	$v(1,2,3) - v(1,2)$
(1,3,2)	$v(1) - v(\varnothing)$	$v(1,3,2) - v(1,3)$	$v(1,3) - v(1)$
(2,1,3)	$v(2,1) - v(2)$	$v(2) - v(\varnothing)$	$v(2,1,3) - v(2,1)$
(2,3,1)	$v(2,3,1) - v(2,3)$	$v(2) - v(\varnothing)$	$v(2,3) - v(2)$
(3,1,2)	$v(3,1) - v(3)$	$v(3,1,2) - v(3,1)$	$v(3) - v(\varnothing)$
(3,2,1)	$v(3,2,1) - v(3,2)$	$v(3,2) - v(3)$	$v(3) - v(\varnothing)$

To calculate the marginal contribution of each player in any coalition, it is necessary to define the set of players preceding player i. Taking column 2 of Table A.1 (player 1) as an example, it is observed that, in the first ordering, the set of players preceding player 1 is $\varnothing$ (empty set). Therefore, the empty set is one of the possible coalitions. The other sets, referring to other orderings, are $\{2\}$, $\{2,3\}$, $\{3\}$ and $\{3,2\}$. The logic is the same for the other players.

After defining the set of players that precede player i, it is possible to measure the marginal contributions of player i when he joins the coalition S. Thus, according to the equation 12, the marginal contribution of player 1, taking into account the first ordering, is equal to $v(1) - v(\varnothing)$, where $v(1)$ and $v(\varnothing)$ are, respectively, the values associated with the coalition 1 ($\varnothing \cup \{1\}$) and the coalition $\varnothing$.

From equation 14, and knowing that $N! = 6$, it is possible to measure the Shapley value for each player. Considering the player 1, one has

$$\phi_1 = \frac{1}{6} \sum_{R \in \Re\mathfrak{c}} \Delta_i(S_i(R)), \tag{14}$$

where $\sum(.)$ is the sum of column 2 of Table A.1.

Appendix B

Table B.1: Participation equations (Probit model) for labor market in the Northeast region, Brazil, 2008

Variables	MA	PI	CE	RN	PB	PE	AL	SE	BA	NE
Cons.	−4.4269*	−4.1358*	−3.5940*	−3.9817*	−4.1205*	−4.1933*	−4.1580*	−3.9905*	−3.9024*	−3.9719*
	(0.1839)	(0.1925)	(0.0332)	(0.1730)	(0.0802)	(0.1016)	(0.1984)	(0.2276)	(0.0874)	(0.0467)
Educ.	0.0734*	0.0462*	0.0637*	0.0659*	0.0819*	0.0769*	0.0612*	0.0561*	0.0693*	0.0689*
	(0.0066)	(0.0067)	(0.0034)	(0.0060)	(0.0062)	(0.0040)	(0.0049)	(0.0076)	(0.0032)	(0.0016)
Sexo	0.978*	0.9284*	0.7417*	0.8943*	0.9254*	0.9118*	0.8397*	0.7362*	0.8801*	0.8737*
	(0.0595)	(0.0563)	(0.0332)	(0.0559)	(0.0615)	(0.0339)	(0.0602)	(0.0697)	(0.0246)	(0.0151)
Cor	−0.060	−0.0654	− 0.0528***	−0.0438	0.0309	−0.0215	0.1308*	0.0568	− 0.0765**	−0.0476*
	(0.0533)	(0.6165)	(0.0305)	(0.0444)	(0.0375)	(0.0312)	(0.0409)	(0.0493)	(0.0308)	(0.0133)
Id	2.0123*	2.0556*	1.6515*	1.8421*	1.7657*	1.8639*	1.8286*	1.9355*	1.8117*	1.8083*
	(0.1105)	(0.0876)	(0.0506)	(0.1063)	(0.0483)	(0.0489)	(0.1167)	(0.1151)	(0.0497)	(0.0254)
Id^2	−0.2421*	−0.2396*	−0.1945*	−0.0220*	−0.2158*	−0.2275*	−0.2252*	−0.2400*	−0.2173*	−0.2170*
	(0.0141)	(0.0098)	(0.0674)	(0.0141)	(0.0076)	(0.0063)	(0.0162)	(0.0175)	(0.0067)	(0.0034)
Children	0.1158**	0.0468	−0.0089	−0.0764	0.0688	0.0518***	−0.0131	−0.0543	0.0039	0.0245***
	(0.0438)	(0.0645)	(0.0314)	(0.0639)	(0.0436)	(0.0278)	(0.0469)	(0.0627)	(0.0281)	(0.0131)
head of house	0.5633*	0.5161*	0.6020*	0.4287*	0.6252*	0.4818*	0.7162*	0.5877*	0.4867*	0.5391*
	(0.0730)	(0.0811)	(0.0278)	(0.0936)	(0.0541)	(0.0336)	(0.0664)	(0.0608)	(0.0343)	(0.0171)
Number of observation	4,814	3,881	18,095	4,809	5,534	18,157	3,884	4,521	24,635	88,330
Considered population	4,228,031	2,130,021	6,028,229	2,267,435	2,759,061	6,064,189	2,177,670	1,487,961	9,885,248	37,027,845
F statistic (prob.)	31.00*	100.58*	216.19*	43.67*	25.44*	144.10*	38.57*	53.73*	159.32*	448.70*

Note: Educ.= years of education. * significant at 1%; ** significant at 5%. *** significant at 10%. The (linear) standard-deviations are in parenthesis. Values that are not specified are not significant.