

VLADIMIR FARIA DOS SANTOS

**EFEITOS DO CRESCIMENTO E REDUÇÃO DA DESIGUALDADE DE
REND A NA POBREZA DA REGIÃO NORDESTE DO BRASIL - 2003-2008**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Economia Aplicada, para obtenção do título de *Doctor Scientiae*.

**VIÇOSA
MINAS GERAIS – BRASIL
2011**

VLADIMIR FARIA DOS SANTOS

**EFEITOS DO CRESCIMENTO E REDUÇÃO DA DESIGUALDADE DE
RENDA NA POBREZA DA REGIÃO NORDESTE DO BRASIL – 2003/2008**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Economia Aplicada, para obtenção do título de *Doctor Scientiae*.

APROVADA: 21 de junho de 2011.

Marília Fernandes Maciel Gomes

Silvia Harumi Toyoshima

Marlon Gomes Ney

João Eustáquio de Lima
(Co-orientador)

Wilson da Cruz Vieira
(Orientador)

*Dedico aos meus pais e a minha
esposa Dayane.*

AGRADECIMENTOS

Acima de tudo, eu devo agradecer a Deus por mais uma grande conquista em minha vida.

Agradeço intensamente aos meus pais Adão Lucindo dos Santos e Marta Verônica de Faria Santos. Sem o apoio deles eu não teria alcançado todos os meus objetivos.

Outra pessoa de suma importância no desenvolvimento desse trabalho foi a minha esposa Dayane, que esteve sempre ao meu lado, acreditando e me apoiando em todos os momentos difíceis.

Ao Departamento de Economia Rural da Universidade Federal de Viçosa que me forneceu todas as condições para finalizar o curso de Doutorado em Economia Aplicada.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo fornecimento da bolsa de doutorado. Essa bolsa me possibilitou dedicação exclusiva aos estudos.

Ao professor Wilson da Cruz Vieira, por tudo que me ensinou, pela paciência, pelo respeito e pela brilhante orientação. Gostaria de expressar minha gratidão por tudo isso e por ele sempre acreditar em mim.

Aos membros da banca de defesa: professor João Eustáquio de Lima, professora Marília Fernandes Maciel Gomes, professor Marlon Gomes Ney e professora Silvia Harumi Toyoshima. Todos eles contribuíram para que o trabalho tornasse melhor.

Aos demais professores do DER que contribuíram intensamente para o meu crescimento pessoal e intelectual.

À professora Ramani Gunatilaka, ao professor Steven M. Helfand e ao professor Guanghua Wan. Os dois primeiros por valiosas sugestões e o terceiro por ter disponibilizado o programa java, que foi imprescindível a essa pesquisa.

Agradeço a todos os funcionários do DER, em especial à Carminha, Cida, Tedinha, Brilhante e Helena. Todos eles sempre estavam dispostos a ajudar.

Aos amigos Airton Lopes Amorin e Matheus Wemerson Gomes Pereira, pela amizade e pelo tempo em que passamos estudando.

Aos demais amigos da pós-graduação: Alan Figueiredo de Arêdes, Jean dos Santos Nascimento, João Ricardo F. de Lima, Lucas Oliveira de Souza, Marcelo Dias Paes Ferreira, Mirelle Cristina de Abreu Quintela, Reisoli Bender Filhos, Ricardo Bruno Nascimento dos Santos, Talles Girardi de Mendonça e Vanessa da Fonseca Pereira.

BIOGRAFIA

VLADIMIR FARIA DOS SANTOS, filho de Adão Lucindo dos Santos e Marta Verônica de Faria Santos, nasceu em Viçosa, MG, em 11 de Julho de 1980.

Em Maio de 2006, graduou-se em Ciências Econômicas pela Universidade Federal de Viçosa, em Viçosa, MG.

Em Maio de 2006, iniciou o Programa de Pós-Graduação em Economia Aplicada, em nível de Mestrado, no Departamento de Economia Rural na Universidade Federal de Viçosa, defendendo a dissertação em 21 de fevereiro de 2008.

Em Março de 2008, foi selecionado para cursar o Programa de Pós-Graduação em Economia Aplicada, em nível de Doutorado, no Departamento de Economia Rural. Em Agosto de 2010 tomou posse no cargo de professor assistente do curso de economia da Universidade Federal Fluminense (pólo de Campos dos Goytacazes – RJ). Em junho de 2011 defendeu sua tese de doutorado e venceu mais uma etapa de sua vida.

SUMÁRIO

LISTA DE TABELAS	viii
LISTAS DE FIGURAS.....	x
RESUMO	xi
ABSTRACT	xiii
1. INTRODUÇÃO	1
1.1. Considerações iniciais	1
1.2. O problema e sua importância.....	4
1.3. Hipótese.....	11
1.4. Objetivos	11
1.4.1. Geral.....	11
1.4.2. Específicos	11
2. REFERENCIAL TEÓRICO	13
2.1. Teoria do capital humano	16
2.2. Teoria da discriminação no mercado de trabalho.....	20
2.2.1. Preferência por discriminação	21
2.2.1.1. Discriminação pelos empregadores.....	21
2.2.1.2. Discriminação pelos consumidores	22
2.2.1.3. Discriminação pelos trabalhadores.....	23
2.2.2. Discriminação estatística.....	24
2.3. Teorias dos diferenciais compensatórios e salário-eficiência	25
2.3.1. Teoria dos diferenciais compensatórios.....	25
2.3.1.1. Oferta de trabalho	27
2.3.1.2. Demanda de trabalho	27
2.3.2. Teoria salário-eficiência	28

2.4.	Teoria do mercado de trabalho dual	29
2.5.	Pobreza, crescimento e desigualdade	30
3.	METODOLOGIA	35
3.1.	Valor de Shapley	35
3.2.2.	Decomposição de Kakwani.....	37
3.2.3.	Decomposição de Shapley	39
3.3.	Decomposição baseada em um modelo de regressão.....	41
3.3.1.	Decomposição de Shapley: exemplo simplificado	43
3.4.	Procedimento de Heckman.....	46
3.5.	Variáveis explicativas endógenas.....	48
3.6.	Medidas de desigualdade de renda.....	53
3.6.1.	Medidas de pobreza	57
3.7.	Fonte de dados e tratamento dos dados	57
4.	RESULTADOS E DISCUSSÃO	59
4.1.	Medidas de pobreza e decomposição	59
4.2.	Decomposição baseada numa regressão.....	76
4.2.1.	Equação de rendimentos estimada pelo procedimento de Heckman	77
4.2.2.	Viés de seletividade e endogeneidade da variável educação	90
5.	CONCLUSÕES	96
	REFERÊNCIAS BIBLIOGRÁFICAS.....	100
	ANEXOS	109
	ANEXO A	110
	ANEXO B	114
	ANEXO C	115

LISTA DE TABELAS

Tabela 1 – Proporção da renda total apropriada pelos décimos da PEA (Pessoas Economicamente Ativas). Brasil – período 1960/1990	3
Tabela 2 – Índice de Gini da Região Nordeste, por Estados, período 2003-2008	8
Tabela 3 – Decomposição do valor de Shapley	40
Tabela 4 – Valor de Shapley aplicada à análise de regressão	44
Tabela 5 – Medidas de pobreza da classe FGT (α), 2003 e 2008.....	60
Tabela 6– Medidas de extrema pobreza da classe FGT (α), 2003 e 2008	62
Tabela 7 – Contribuição relativa dos estados do Nordeste para a pobreza (P_0) regional, 2003-2008	63
Tabela 8 – Rendimento domiciliar per capita médio (por mês), em valores de setembro de 2008, dos estados da Região Nordeste, 2003 e 2008.....	65
Tabela 9 – Taxa de crescimento da renda média por décimo da distribuição (%), 2003-2008	66
Tabela 10 – Parcela dos rendimentos apropriados pelos décimos da distribuição, 2003 e 2008.....	69
Tabela 11 – Decomposição da pobreza e extrema pobreza (P_0) em dois componentes: crescimento (ϕ_G^S) e redistribuição (ϕ_D^S), 2003-2008.....	70
Tabela 12 – Composição da renda domiciliar <i>per capita</i> (rdpc) nos estados do Nordeste do Brasil, 2003 e 2008	74
Tabela 13 – Equações de participação (modelo <i>Probit</i>) no mercado de trabalho nordestino, 2008.....	78
Tabela 14 – Equações de rendimentos para os estados da Região Nordeste, procedimento de Heckman, 2008.....	81
Tabela 15 – Decomposição (Shapley) baseada na equação de rendimentos estimada pelo procedimento de Heckman (%), 2008.....	86

Tabela 16 – Equações de rendimentos para os estados da Região Nordeste, metodologia de Wooldridge, 2008.....	91
Tabela 17 – Decomposição (Shapley) baseada na equação de rendimentos estimada pelo método de Wooldridge (%), 2008.....	94
Tabela A. 1 – Estratificação geográfica da PNAD	111
Tabela B. 1 – Decomposição da pobreza e extrema pobreza (P1) em dois componentes: crescimento (ϕ_G^S) e redistribuição (ϕ_D^S), 2003 e 2008....	114
Tabela B. 2 – Decomposição da pobreza e extrema pobreza (P2) em dois componentes: crescimento (ϕ_G^S) e redistribuição (ϕ_D^S), 2003 e 2008.....	114

LISTAS DE FIGURAS

Figura 1 – Curva de Lorenz para o Brasil, 2003 e 2008.	4
Figura 2 – Proporção de pessoas com renda domiciliar per capita inferior à linha de pobreza	5
Figura 3 – Evolução da proporção de pessoas na Região Nordeste do Brasil com renda domiciliar per capita inferior à linha de pobreza, período 2001-2008	6
Figura 4 – Curva de Lorenz para a Região Nordeste, 2003 e 2008	7
Figura 5 – Educação como investimento	18
Figura 6 – Política ótima no combate à pobreza	32
Figura 7 – Curva de Lorenz para uma distribuição contínua	55
Figura A. 1 – Ilustração do plano amostral da PNAD.	113

RESUMO

SANTOS, Vladimir Faria dos, D.Sc., Universidade Federal de viçosa, junho de 2011. **Efeitos do crescimento e redução da desigualdade de renda na pobreza da Região Nordeste do Brasil - 2003-2008** Orientador: Wilson da Cruz Vieira. Co-Orientadores: Steven M. Helfand e João Eustáquio de Lima.

Na Região Nordeste, uma grande parte da população vive na pobreza absoluta. O número elevado de seres humanos nessa situação está estreitamente relacionado ao alto grau de desigualdade da região, visto que no País existe renda mais do que suficiente para erradicar a pobreza. Embora a proporção de pobres e de indigentes seja grande, o que vem ocorrendo nos últimos anos, sobretudo a partir de 2003, é a queda significativa da pobreza e da indigência. Essa redução expressiva pode estar relacionada tanto ao crescimento dos rendimentos das pessoas como por uma distribuição menos desigual; visto que, o número de pessoas pobres pode variar conforme haja alterações no crescimento econômico e/ou na concentração de renda. Levando em consideração que há dois fatores (crescimento e redistribuição de renda) que podem reduzir a quantidade de indivíduos vivendo na pobreza e na miséria, este trabalho teve como objetivo investigar, para todos os estados da Região Nordeste, a contribuição do crescimento econômico e da desigualdade pessoal de rendimentos para a queda na pobreza absoluta e analisar os seus principais determinantes. O referencial teórico é baseado em dois grupos de teorias: aquelas que focam as características dos trabalhadores; e aquelas que enfatizam a estrutura do mercado de trabalho. Podem-se classificar, dentro do primeiro grupo, as teorias do capital humano e da discriminação no mercado de trabalho. No segundo grupo, estão as teorias dos diferenciais compensatórios e salário-eficiência e a teoria do mercado de trabalho dual. Como modelos analíticos foram utilizados dois métodos de decomposição, ambos baseados no valor de Shapley, um conceito de solução para jogos cooperativos. O primeiro teve como função mensurar a importância de cada

um dos componentes (crescimento e redistribuição) para a variação na pobreza no período de 2003 a 2008. O segundo método é baseado em uma regressão (equação de rendimentos) e teve como propósito quantificar a contribuição de cada variável inserida no modelo econométrico para a desigualdade de rendimentos. Como medida de desigualdade utilizou-se as três mais importantes: Coeficiente de Gini, Coeficiente de Theil-T e Theil-L. Os resultados da primeira decomposição mostraram que o crescimento da renda foi o componente que mais explicou a queda na pobreza no período de 2003 a 2008. Esse resultado foi verificado em todos os estados do Nordeste, tanto para a pobreza como para a extrema pobreza. Paraíba foi o único estado onde o crescimento dos rendimentos explicou inteiramente a queda do número de pobres, visto que, no período analisado, houve piora na sua distribuição de renda. Por meio dos resultados advindos da decomposição baseada numa regressão, foi possível observar que a educação e experiência foram as variáveis que mais explicaram a desigualdade de rendimentos nos estados do Nordeste, o que está em conformidade com a teoria do capital humano. Além das variáveis do capital humano, aquelas relacionadas ao gênero e a cor também contribuíram positivamente para a desigualdade de renda, sugerindo que na região pode existir algum tipo de discriminação no mercado de trabalho. Diante disso, concluiu-se que a renda domiciliar *per capita* foi de extrema importância para a queda na pobreza e que as variáveis educação e experiência foram as que mais contribuíram para o alto índice de desigualdade nos estados do Nordeste, confirmando, assim, as hipóteses feitas nesta pesquisa.

ABSTRACT

SANTOS, Vladimir Faria dos, D.Sc., Universidade Federal de viçosa, june,2011.
Effects of growth and reduction of income inequality in poverty in the Northeast of Brazil - 2003-2008 Adviser: Wilson da Cruz Vieira. Co-advisers: Steven M. Helfand and João Eustáquio de Lima.

In the Northeast, a large proportion of the population lives in absolute poverty. The high number of human beings in this situation is closely related to the high degree of inequality in the region since there is more than enough income to eradicate poverty in the country. Although the number of poor and indigent individuals is large, which has been occurring in recent years, particularly since 2003, it is its systematic reduction. This significant decrease could be related both to the growth of personal income as a less unequal distribution, in which the number of poor people there may vary by changes in economic growth and / or the concentration of income. Taking into account that there are two important factors (growth and income redistribution) which can reduce poverty and indigence, this study aimed to investigate, for all states in the Northeast, the contribution of economic growth and personal income inequality to the fall in absolute poverty and examine their main determinants. The theoretical framework is based on two groups of theories: those that focus on the characteristics of workers, and those that emphasize the structure of the labor market. It is possible to classify, within the first group, the theories of human capital and discrimination in the labor market. In the second group are the theories on compensating wage differentials and efficiency wage. As analytical models two methods of decomposition were used, both based on the Shapley value, a solution concept for cooperative games. The first function was to quantify that component (growth or redistribution) was more important in explaining the variation in poverty in the period 2003 to 2008. The second method is

based on regression (earnings equation), which aims to measure the contribution of each variable included in the econometric model for income inequality. To compute the levels of inequality three main measures were used namely the Gini index and two indexes of generalized entropy class: Theil-L and Theil-T. The results of the first decomposition showed that the income growth was the component which most explained the fall in poverty from 2003 to 2008. This result was observed in all Northeastern states, both for poverty and for extreme poverty. Paraíba is the only state where the income growth explained fully the fall in the number of poor, since in the period analyzed, there was worsening in the distribution of income. Through the results arising from the decomposition based on regression, it was clear that education and experience were the variables that explained best the income inequality in the Northeast states, which is consistent with the theory of human capital. In addition to human capital variables, those related to gender and color also contributed positively to income inequality, suggesting that the region may have some kind of discrimination in the labor market. Therefore, it was concluded that household income per capita was of utmost importance to the fall in poverty and that education and experience variables were those that contributed most to the high level of inequality in the Northeast states, thus confirming the hypothesis made in this research.

1. INTRODUÇÃO

1.1. Considerações iniciais

Em dezembro de 2010, a Declaração Universal dos Direitos Humanos completou 62 anos. Entre os direitos fundamentais do ser humano, está o direito a um padrão de vida que garanta saúde e bem-estar, o que inclui alimentação, vestuário, habitação, cuidados médicos, entre outros. No entanto, para boa parte da população mundial, esses direitos não são assegurados. De acordo com a estimativa da Organização das Nações Unidas para Agricultura e Alimentação - FAO (2010), aproximadamente 1 bilhão de pessoas passam fome no mundo. Em virtude dessa situação, não apenas o direito à alimentação é violado, mas também os direitos à saúde, a uma moradia digna, à educação de qualidade etc.

Embora haja alimentos suficientes para satisfazer toda a população do planeta, uma grande parcela dela não possui renda suficiente para adquiri-los. Segundo a FAO (2009), mais de um bilhão de pessoas vivem com menos de um dólar por dia, enquanto, para a Organização Mundial da Saúde - WHO (1999), sobreviver com menos de US\$ 2 por dia é uma realidade para, aproximadamente, a metade da população mundial.

Embora a quantidade de pessoas desnutridas tenha caído entre 2009 e 2010 (a primeira vez em 15 anos), a parcela de ser humanos extremamente pobres – isto é, que não dispõe do mínimo necessário em termos de alimentação – ainda é significativa. Conforme a FAO (2010), do total de indivíduos nessa situação, a maioria vive em países em desenvolvimento, onde eles representam cerca de 16% dos habitantes dessa região.

Ao considerar a América Latina e Caribe, verifica-se, de acordo com as estimativas da FAO (2010), que 53 milhões de pessoas passam fome. De forma geral, essa região é caracterizada pelo baixo crescimento e pobreza persistente. Para

o Banco Mundial (2006), a pobreza na América Latina é um grande obstáculo a ser superado, com um quarto da população sobrevivendo com menos de US\$ 2 por dia. Além disso, é a região com o maior nível de desigualdade no mundo.

No caso do Brasil, a pobreza é um grave problema que tem raízes históricas. Desde que era colônia, o País já apresentava um grande contingente de pessoas, entre elas, os escravos, que não tinham recursos suficientes para satisfazer suas necessidades básicas, isto é, viviam na miséria. A partir daí, o Brasil vem mantendo, em um patamar alto, esse desequilíbrio social. Segundo o Banco Mundial (2004), o Brasil é um claro *outlier* em termos de desigualdade e retém a maior parcela do número total de pobres na América Latina.

De acordo com Schwartzman (2004), qualquer que seja a análise realizada no Brasil, será observado que, ao lado de uma economia moderna, há milhões de brasileiros excluídos de seus benefícios. Isso reflete a grande desigualdade de renda que existe na economia brasileira.

Na questão da desigualdade, Langoni (1973) enfatizou que foi a partir do processo de desenvolvimento econômico – que se iniciou por volta de 1960 e que alterou a estrutura econômica do Brasil, passando de agrária para a industrial – que se agravou o problema da iniquidade distributiva no País. A justificativa, segundo esse autor, está no mercado de trabalho, visto que houve expansão, motivado pelo crescimento do setor industrial, da demanda de mão-de-obra qualificada; com isso, houve aumento dos salários relativos em favor dos mais qualificados. Os dados da Tabela 1 confirmam esses argumentos, isto é, que a partir de 1960 a desigualdade de renda se intensificou, destacando os anos de 1960 e 1970.

Tabela 1 – Proporção da renda total apropriada pelos décimos da PEA (Pessoas Economicamente Ativas). Brasil – período 1960/1990

Décimo	1960	1970	1980	1990
Primeiro	1,17	1,16	1,18	0,81
Segundo	2,32	2,05	2,03	1,80
Terceiro	3,42	3,00	2,95	2,20
Quarto	4,65	3,81	3,57	3,04
Quinto	6,15	5,02	4,41	4,06
Sexto	7,66	6,17	5,58	5,45
Sétimo	9,41	7,21	7,17	7,35
Oitavo	10,85	9,95	9,88	10,32
Nono	14,69	15,15	15,36	16,27
Décimo	39,66	46,47	47,89	48,69

Fonte: Barros e Mendonça (1995).

Pode-se observar, na Tabela 1, que, no período entre 1960 e 1990, a desigualdade de renda cresceu, na medida em que a fração de renda apropriada pelos 50% mais pobres reduziu cerca de 6 pontos percentuais e aquela apropriada pelos 20% mais ricos cresceu, aproximadamente, 11 pontos percentuais (BARROS e MENDONÇA, 1995)¹. No decorrer dos anos de 1990 e início dos anos 2000 essa discrepância permaneceu em patamares elevados. Por exemplo, no ano de 2000 a razão entre a renda dos 10% mais ricos e 40% mais pobres era de 32,9%, o que mostra o alto grau de concentração.

Embora o Brasil ainda apresente elevado índice de desigualdade, nos últimos anos tem-se verificado uma distribuição menos desigual nos rendimentos². De acordo com o Instituto de Pesquisa Econômica Aplicada – IPEA (2009b), o que se tem observado é o crescimento da renda em todos os décimos, porém, entre os mais pobres, o aumento tem sido mais significativo. A melhora na distribuição de renda no Brasil pode ser vista por meio da curva de Lorenz (Figura 1), considerando os anos de 2003 e 2008³.

¹ A fração da renda apropriada pelos 50% mais pobres foi obtida somando as parcelas de rendimentos dos cinco primeiros décimos.

² A redução nos índices de desigualdade começou ser sentida, com mais intensidade, a partir de 2001. Todavia, a partir de 2003, a pobreza, após um ligeiro crescimento entre 2002 e 2003, entra numa trajetória persistente de queda.

³ A curva de Lorenz mostra quanto cada parcela da população detém da renda total. Quanto mais próxima ela ficar da linha de 45°, menos desigual é a distribuição de rendimentos (MEDEIROS, 2006).

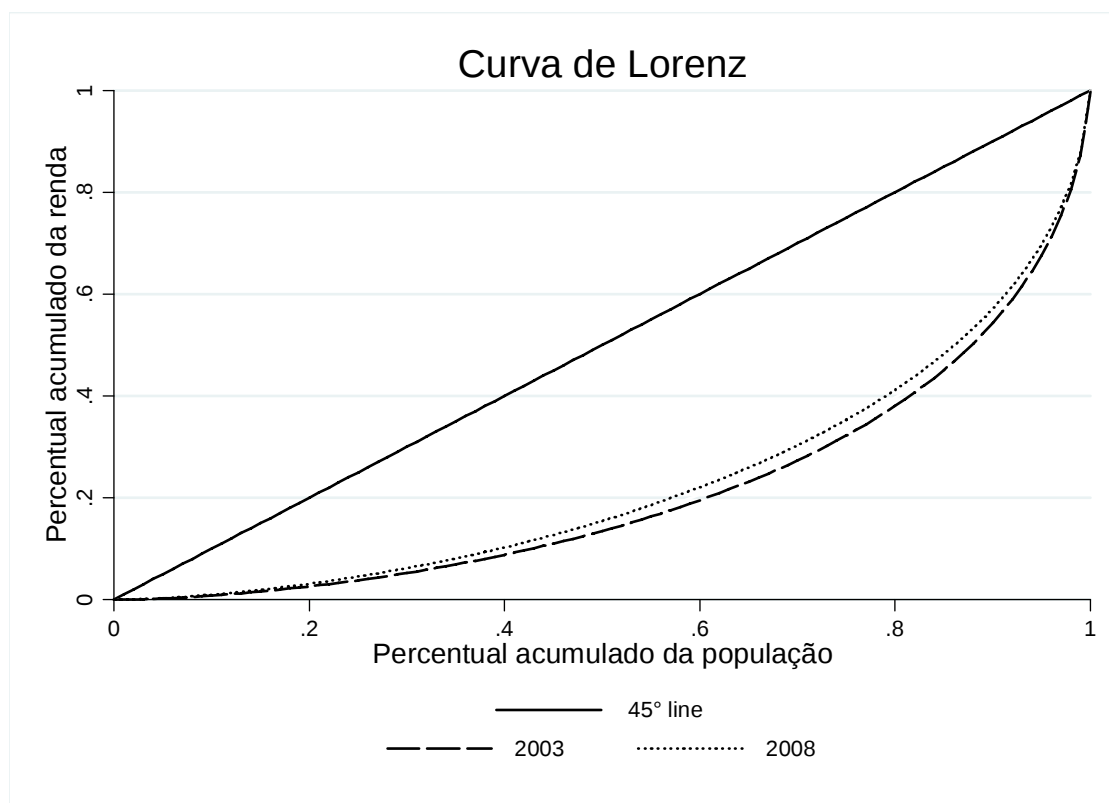


Figura 1 – Curva de Lorenz para o Brasil, 2003 e 2008.
 Fonte: Elaborado pelo autor a partir das PNADs de 2003 e 2008.

Conforme a Figura 1, observa-se a clara redução na desigualdade de rendimentos no Brasil. De acordo com o coeficiente de Gini, essa queda foi de, aproximadamente, 6% – passando de 0,578 em 2003 para 0,544 em 2008.

Assim como a desigualdade, o número de pobres no Brasil tem reduzido – sobretudo a partir de 2003 – de forma expressiva. De acordo com o Instituto de Pesquisa Econômica Aplicada - IPEA (2009b), a porcentagem de pobres em 2003 era de 35,8; em 2008, esse valor caiu para 22,6. Ao considerar a extrema pobreza, seu valor foi reduzido à metade, comparado com ano de 2003⁴.

1.2. O problema e sua importância

Segundo Rocha (2000), os países nos quais há grande número de pobres podem ser divididos em dois grupos: o primeiro é formado por nações cuja renda não é suficiente para satisfazer as necessidades mínimas da população. Nesse cenário, a

⁴ O IPEA (2009a) utilizou uma linha de extrema pobreza baseada em uma cesta de alimentos composta por produtos que proporcionam o mínimo de calorias necessárias para satisfazer uma pessoa. A linha de pobreza foi definida como o dobro da linha de indigência (extrema pobreza). O IPEA estimou diferentes valores para diversas regiões do Brasil.

renda *per capita* é tão baixa que a maioria da população vive na pobreza absoluta, independentemente de como é distribuído o produto nacional. O segundo grupo é composto pelas nações que possuem recursos mais do que suficiente para atender as necessidades mínimas de cada um de seus habitantes. A pobreza existente nesse grupo seria, então, decorrente da má distribuição da renda.

O Brasil se enquadra no segundo grupo, visto que é um dos países mais ricos do mundo em termos de Produto Interno Bruto (PIB); mas a pobreza ainda é um grande desafio a ser combatido. Levando em conta a renda familiar *per capita*, é possível observar que o volume de riqueza produzido pelo Brasil é mais do que suficiente para que não haja nenhum pobre no País. A concentração de renda, portanto, pode ser considerada como uma das principais dificuldades para a erradicação da miséria. Embora esteja em processo de queda, a desigualdade de rendimentos é uma das mais altas do planeta. Conforme Barros *et al.* (2000), aproximadamente 90% dos países do mundo possuem grau de desigualdade inferior ao encontrado no Brasil.

De acordo com Rocha (2000), a pobreza no Brasil tem caráter regional, sendo que sua incidência, independente dos indicadores utilizados, é mais elevada no Nordeste. A Figura 2 mostra que, em 2008, a proporção de pobres nessa região era de 42%; a maior entre as cinco regiões do Brasil.

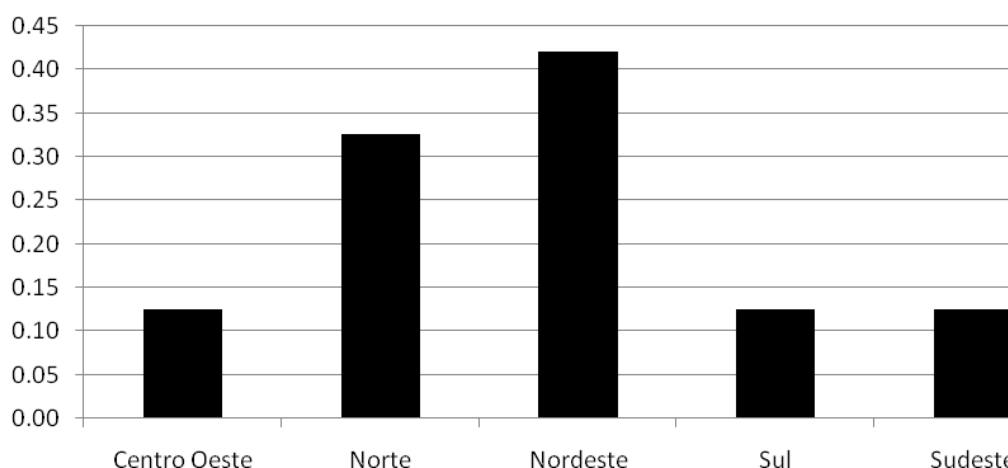


Figura 2 – Proporção de pessoas com renda domiciliar per capita inferior à linha de pobreza⁵
Fonte: Elaborado pelo autor a partir de dados do IPEADATA (2009a).

A alta proporção de pobres no Nordeste já persiste por décadas, apesar do crescimento econômico da região. Por exemplo, entre 1960 a 1969, o produto

⁵ Ver nota de rodapé número 3.

expandiu-se a uma taxa anual de 4,4%. No período compreendido entre 1970 e 1979, a taxa de crescimento foi de 9,4% ao ano. Entre 1980 e 1989, a taxa de expansão caiu para 4,3% ao ano, enquanto, no período entre 1990 e 2000, a taxa de crescimento reduziu para, aproximadamente, 2,6% (GUIMARÃES NETO, 2004). Embora a economia nordestina tenha crescido desde a década de 1960, especificamente no período 1960/2000, os indicadores sociais, sobretudo a pobreza, ainda permanecem em patamares lamentáveis. O crescimento, portanto, não proporcionou redução significativa do número de pessoas abaixo da linha de pobreza e isso pode estar relacionado ao alto nível de desigualdade existente na região⁶.

Todavia, seguindo a trajetória nacional dos últimos anos, a Região Nordeste também está apresentando uma tendência de queda em seu nível de pobreza, mas ainda permanece em nível elevado (Figura 3).

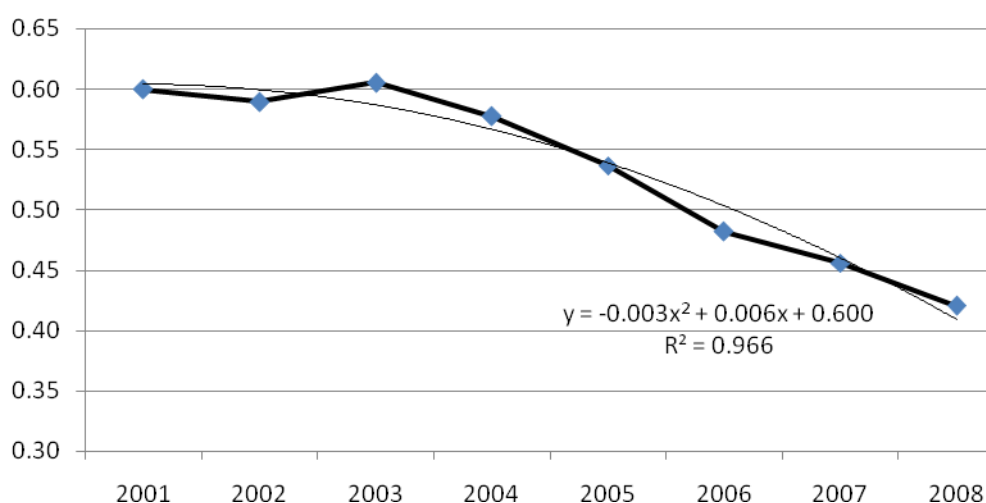


Figura 3 – Evolução da proporção de pessoas na Região Nordeste do Brasil com renda domiciliar per capita inferior à linha de pobreza, período 2001-2008⁷
 Fonte: Função estimada pelo autor a partir dos dados do IPEADATA (2009a).

Como pode ser notado, a proporção de pobres tem caído, sobretudo a partir do ano de 2003. Por meio da Figura 3 observa-se uma clara trajetória decrescente (ver a linha de tendência estimada). Comparando os anos de 2003 e 2008, a queda foi de, aproximadamente, 31%.

⁶ É importante ressaltar o impacto do crescimento populacional sobre esses indicadores. Entre 1970 e 2000, a população residente da Região Nordeste cresceu 70%, passando de 28,1 milhões para mais de 47 milhões de pessoas.

⁷ Ver nota de rodapé número 3.

Essa redução significativa pode estar relacionada a uma distribuição de renda menos desigual na Região Nordeste, visto que o número de pessoas abaixo da linha de pobreza pode variar conforme haja alterações no crescimento econômico e/ou na concentração de renda. No que tange à distribuição de rendimentos, observou-se que, na Região Nordeste, houve melhora entre os anos de 2003 e 2008. A Figura 4 reflete, por meio da curva de Lorenz, a queda na desigualdade de rendimentos na região.

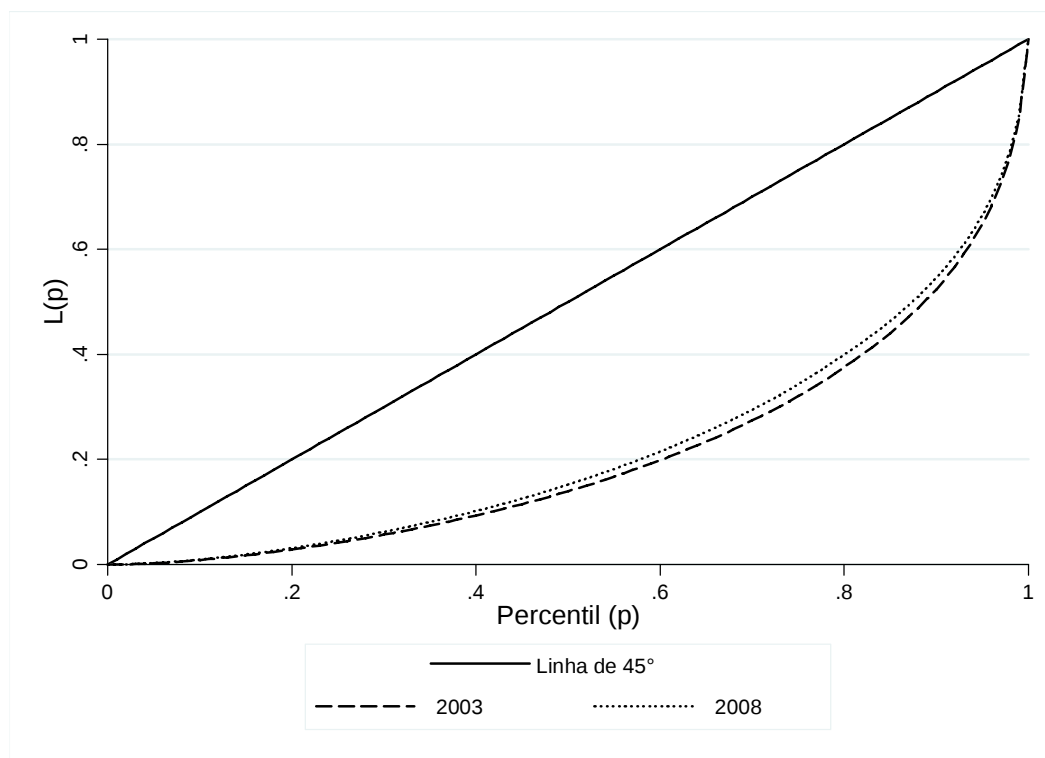


Figura 4 – Curva de Lorenz para a Região Nordeste, 2003 e 2008

Fonte: Elaborada pelo autor a partir das PNADs de 2003 e 2008.

Em termos do índice de Gini, a queda na iniquidade distributiva foi de 4,04% no período supracitado⁸. Em termos estaduais, a desigualdade seguiu a tendência da região como um todo. Em todos os estados do Nordeste, exceto Paraíba, houve queda. A Tabela 2 reporta o índice de Gini dessa região a partir de 2003.

⁸ O índice de Gini mede a extensão na qual a distribuição de renda entre os indivíduos dentro de uma economia se desvia de uma distribuição perfeitamente igualitária. Seu valor varia entre 0, quando a renda é perfeitamente distribuída, e 1, quando a renda é totalmente concentrada (BANCO MUNDIAL, 2003).

Tabela 2 – Índice de Gini da Região Nordeste, por Estados, período 2003-2008

Estados	2003	2004	2005	2006	2007	2008	$\Delta\%$ (2003-2008)
Alagoas	0,605	0,572	0,564	0,624	0,607	0,580	-4,1
Bahia	0,586	0,555	0,553	0,556	0,556	0,559	-4,6
Ceará	0,564	0,574	0,577	0,546	0,547	0,537	-4,7
Maranhão	0,573	0,609	0,519	0,596	0,555	0,522	-9,0
Paraíba	0,564	0,590	0,578	0,562	0,595	0,585	3,8
Pernambuco	0,587	0,606	0,586	0,581	0,559	0,565	-3,8
Piauí	0,596	0,588	0,589	0,599	0,593	0,573	-3,8
Rio Grande do Norte	0,561	0,569	0,595	0,557	0,560	0,550	-1,9
Sergipe	0,574	0,559	0,553	0,558	0,539	0,536	-6,6
Nordeste	0,581	0,583	0,571	0,573	0,566	0,557	-4,1

Fonte: Cálculos feitos pelo autor a partir dos dados do IPEADATA (2009a) e da PNAD 2008.

Por meio da Tabela 2, é possível observar que somente no Estado da Paraíba houve crescimento da desigualdade de renda. Em compensação, nos Estados do Maranhão e Sergipe, a queda foi superior à verificada a nível nacional⁹. Essa redução poderia ser explicada, em grande parte, pelas políticas de redistribuição de renda – por exemplo, os programas de transferência direta de renda – dos últimos anos. Há trabalhos que sugerem a eficácia desses programas em reduzir a desigualdade de renda brasileira, tais como Soares *et al.* (2006), Ferreira *et al.* (2008) e Santos *et al.* (2009). Outro fator importante no combate à desigualdade é o salário mínimo. Na última década, houve forte crescimento real de seu valor. Especificamente nos anos de 2003 a 2008, o salário mínimo se expandiu 34%. De acordo com Saboia (2007), a queda na desigualdade de rendimentos, entre 1995 e 2005, está estreitamente relacionada ao aumento do salário mínimo.

Embora uma distribuição menos desigual possa contribuir para a queda na pobreza, o crescimento nos rendimentos também é um fator importante nesse quesito. Como já ressaltado, o número de pobres pode cair tanto por meio da diminuição da desigualdade como pelo aumento da renda. Considerando os microdados das PNADs de 2003 e 2008, o crescimento do rendimento domiciliar *per capita* no Nordeste do Brasil foi superior a 43%. Esse significativo incremento de renda sugere que a queda na pobreza pode estar mais relacionada ao crescimento dos rendimentos do que à diminuição em sua concentração.

Todavia, é importante frisar – considerando a Figura 4 e a Tabela 2 – que a pobreza na Região Nordeste poderia ter caído de forma mais significativa caso o nível de desigualdade tivesse se reduzido numa proporção maior do que realmente

⁹ Entre os anos de 2003 e 2008, no Brasil, a queda do índice de Gini foi de 6%.

foi verificado. A distribuição desigual da renda é um sério problema no processo de queda da pobreza.

A concentração de rendimentos no Nordeste ainda está num nível muito elevado e, por isso, tem margem para cair de forma mais acentuada. De acordo com dados do IPEADATA (2009a), a proporção da renda apropriada pelos 50% mais pobres, segundo a renda domiciliar *per capita*, em 2007, foi de 14,68%, enquanto os 1% mais ricos se apropriaram, basicamente, da mesma proporção (14,52%). Essa distribuição é a pior entre as macrorregiões brasileiras e mostra a gravidade que é a desigualdade no Nordeste. Diante disso, a iniquidade distributiva pode ser o foco de ação do poder público para reduzir e, talvez, erradicar a pobreza. Políticas que buscam qualificar ou melhorar o nível educacional da população, por exemplo, poderiam ter uma função importante para reduzir a desigualdade e a pobreza. Em 2007, no Nordeste, aproximadamente 20% da população acima de 15 anos era totalmente analfabeta (IPEADATA, 2009a). Essa proporção é a maior do País. Portanto, tomando como base a teoria do capital humano, a educação na Região Nordeste do Brasil pode ter uma função importante na redução dos níveis de concentração de renda e pobreza¹⁰.

O papel da educação no combate à desigualdade é ainda maior quando se leva em consideração que a renda proveniente do trabalho é o principal componente dos rendimentos da população nordestina. Conforme os microdados da PNAD de 2008, 72% do rendimento domiciliar *per capita* da Região Nordeste é composto pela renda do trabalho. Nesse sentido, a qualificação do trabalhador torna-se um fator importante como redutor da desigualdade de renda, dado o impacto positivo que a instrução (educação) possui sobre os rendimentos do trabalhador.

Em virtude da grande participação dos rendimentos advindos do trabalho na renda total, é de se esperar que esse tipo de renda seja relevante para reduzir a desigualdade. Cacciamali e Camillo (2009) encontraram resultados que sugerem isso. Conforme esses autores, a renda do trabalho foi o principal componente, nos períodos de 2001 a 2004 e de 2004 a 2006, para reduzir a desigualdade nas regiões brasileiras.

O fato dos rendimentos provenientes do trabalho serem importantes componentes da renda domiciliar *per capita* e terem a capacidade de reduzir a

¹⁰ Conforme a teoria do capital humano, educação é de fundamental importância para proporcionar melhores rendimentos e, por consequência, combater a desigualdade e a pobreza.

desigualdade, torna-se de fundamental importância identificar quais os principais fatores que têm a característica de impactar tais rendimentos e, por conseguinte, a desigualdade e a pobreza.

Existem trabalhos que buscaram identificar quais são as principais variáveis no combate à desigualdade de rendimentos, porém, não há consenso a esse respeito. Salardi (2005), por exemplo, analisou, para o ano de 2002, os determinantes da desigualdade de renda para o Brasil, utilizando métodos de decomposição (por exemplo, o método de Fields). Esse autor utilizou variáveis que captam características demográficas, nível educacional, área geográfica etc. Os resultados mostraram que a desigualdade brasileira é explicada, em sua maior parte, pelas diferenças regionais e étnicas.

Ramos e Vieira (2001), por meio de métodos de decomposição (estática e dinâmica), objetivaram captar a importância relativa de algumas variáveis, tais como educação, idade, posição na ocupação, região geográfica, gênero, cor e setor de atividade. Esses autores mostraram que a variável escolaridade tem papel de destaque, comparada com as demais variáveis, para explicar a desigualdade de renda brasileira.

Berni (2007) investigou os determinantes da desigualdade da Região Nordeste como um todo. Para isso, esse autor também utilizou a decomposição de Fields. Os resultados sugeriram que a educação é o principal fator para explicar a má distribuição de renda na região. Siqueira e Siqueira (2006) também estudaram a Região Nordeste. Para isso, a desigualdade de renda foi decomposta em desigualdade intergrupo e intragrupo, utilizando dados da PNAD. Os resultados mostraram que a grande maioria da iniquidade distributiva é encontrada dentro dos estados e dos grupos de gênero e raça.

Como foi possível observar, não existe consenso a respeito de que variáveis são as mais importantes para explicar a discrepância de renda entre os mais ricos e os mais pobres. Assim, este trabalho apresenta duas contribuições. A primeira é aprofundar as discussões em torno desse assunto, principalmente no que se refere aos estados do Nordeste. Os estudos que quantificaram os principais fatores determinantes da desigualdade de renda no Nordeste analisaram a região de forma agregada, desconsiderando a heterogeneidade existente entre os estados. A segunda é a metodologia utilizada – denominada decomposição de Shapley –, que supera algumas limitações existentes nos métodos já utilizados na literatura brasileira;

especificamente, o método de Fields. A metodologia de Fields consiste em estimar uma equação de rendimentos e desagregar, quantitativamente, seus principais componentes. Entre as limitações existentes nesse método, pode-se citar o fato que a decomposição é feita considerando o logaritmo da renda e não a sua variável original, o que tende a provocar distorções nos resultados. Além disso, esse método considera somente o coeficiente de variação como medida de desigualdade, ou seja, não é possível mensurar a desigualdade por outro índice. Ao contrário, o método usado neste trabalho (Shapley) pode ser conduzido independentemente da forma funcional adotada para a equação de rendimentos e fornece a opção de se utilizar qualquer medida de desigualdade. No mais, a metodologia de Shapley também permite utilizar a variável original renda, e não o seu logaritmo.

A razão para a escolha do Nordeste é que essa região possui o maior contingente de pessoas que estão abaixo da linha de pobreza e tem o maior grau de desigualdade de rendimentos do País.

1.3. Hipótese

As hipóteses a serem testadas são: (i) A redução na pobreza, ocorrida nos últimos anos na Região Nordeste, é majoritariamente atribuída ao crescimento dos rendimentos; (ii) as variáveis relacionadas ao capital humano são as que mais explicam a desigualdade de rendimentos na Região Nordeste.

1.4. Objetivos

1.4.1. Geral

Este trabalho teve como objetivo geral investigar, para todos os estados da Região Nordeste, a contribuição de dois fenômenos – crescimento econômico e desigualdade pessoal de rendimentos – para a queda na pobreza absoluta, no período de 2003 a 2008. Em seguida, analisar os principais determinantes desses fenômenos.

1.4.2. Específicos

Especificamente, pretendeu-se:

- a) Calcular os níveis de pobreza de acordo com o rendimento domiciliar *per capita* e decompor os níveis de pobreza em dois componentes: crescimento e redistribuição (desigualdade);
- b) Identificar os determinantes dos rendimentos do trabalho das pessoas ocupadas nos estados da Região Nordeste; e
- c) Quantificar, para todos os estados do Nordeste, os principais determinantes da desigualdade de rendimento.

2. REFERENCIAL TEÓRICO

A literatura que trata da desigualdade de renda pode ser dividida em duas grandes áreas: aquela que analisa os efeitos macroeconômicos sobre a desigualdade; e aquela que foca no relacionamento entre fatores microeconômicos e desigualdade (HAMMILL, 2005).

Segundo Hammill (2005), os trabalhos que abordam a análise macroeconômica da desigualdade possuem algumas características em comum. Eles, normalmente, utilizam os agregados macroeconômicos, tais como PIB, taxa de juros, crescimento econômico, sistema tributário, entre outros. A renda, na maioria das vezes, é definida como PIB ou PIB *per capita*.

O trabalho de Kuznets (1955) foi um dos pioneiros a abordar a desigualdade num contexto macroeconômico. O autor introduziu a famosa curva do U-invertido, que estabelece que, inicialmente, a distribuição de renda torna-se mais desigual à medida que a renda cresce; porém, com o desenvolvimento do país, essa tendência se reverte e a distribuição torna-se mais igualitária.

Após o trabalho de Kuznets (1955), diversos outros (teóricos e empíricos) surgiram com o intuito de verificar a relação entre crescimento econômico e desigualdade de renda. Entre esses, citam-se Kaldor (1957), Alesina e Rodrik (1994), Person e Tabellini (1994), Li e Zou (1998), Barro (2000), Forbes (2000), García-Peñalosa e Turnovsky (2006), Jin (2009), entre outros.

As evidências empíricas e teóricas não são conclusivas. Há trabalhos que encontraram uma relação negativa entre crescimento e desigualdade. Alesina e Rodrik (1994), por exemplo, desenvolveram um modelo de crescimento econômico endógeno, utilizando trabalho e capital como fatores de produção para analisar o relacionamento entre política e crescimento econômico; considerando os conflitos distributivos existente entre os agentes. Por meio do teorema do votante mediano, os autores verificaram que a desigualdade é um fator que reduz o crescimento

econômico, visto que a má distribuição de renda tende a influenciar o votante mediano a preferir maior tributação sobre o capital que, por sua vez, reduz o crescimento da renda.

Seguindo a mesma linha de Alesina e Rodrik (1994), Persson e Tabellini (1994) desenvolveram um modelo teórico para relacionar crescimento econômico, desigualdade de renda e instituições políticas. De forma geral, o resultado dos autores evidenciou uma relação negativa entre desigualdade e crescimento. A explicação para isso estaria no fato de que a má distribuição da renda tende a gerar políticas que não protegem os direitos de propriedade e não permitem a completa apropriação privada dos retornos do investimento. Nesse cenário, portanto, os incentivos ao investimento são baixos e, por consequência, a taxa de crescimento econômico fica aquém do que poderia acontecer.

Alesina e Perotti (1996) argumentam que quanto mais desigual é a distribuição de riqueza, maior tende a ser a instabilidade política, econômica e social do país; o que incentiva os indivíduos a se envolverem em atividades ilegais. Em um ambiente de insegurança e incerteza, a acumulação de capital e, por consequência, os investimentos, são menores. Assim, a desigualdade levaria a uma taxa de crescimento futuro menor.

Por outro lado, há estudos que não encontraram relação negativa entre a má distribuição de riqueza e o crescimento da economia. Como exemplo, pode-se citar García-Peñalosa e Turnovsky (2006). Esses autores analisaram a relação entre crescimento e desigualdade de riqueza baseado na pressuposição de que a taxa de crescimento da economia e a distribuição de renda são ambas endógenas e influenciadas por mudanças estruturais e políticas macroeconômicas. Seus resultados evidenciaram que taxas de crescimento maiores estão associadas à distribuição mais desigual da riqueza.

Li e Zou (1998) também investigaram, teórica e empiricamente, o relacionamento entre desigualdade e crescimento. Assim como García-Peñalosa e Turnovsky (2006), os autores observaram, tanto teórica como empiricamente, que, ao incorporar o consumo público na função de utilidade, a má distribuição de renda pode proporcionar taxas de crescimento maiores. Esse mesmo resultado foi obtido por Forbes (2000), que encontrou uma relação positiva no curto e médio prazo entre crescimento e desigualdade.

É importante enfatizar que há também trabalhos que verificaram relação ambígua em seus resultados. Entre esses, pode-se citar Barro (2000) e Jin (2009). Especificamente, Barro (2000) encontrou uma relação negativa entre crescimento e desigualdade em países pobres, mas uma relação positiva quando foram considerados os países mais ricos.

Todos esses trabalhos supracitados utilizam um enfoque macroeconômico da desigualdade que, de acordo com Hammill (2005), omite algumas considerações que existem a respeito dos efeitos microeconômicos. Segundo esse autor, dada a probabilidade que a combinação dos efeitos micro e macroeconômico afetem a desigualdade, conclusões a respeito das suas causas – baseado somente nos fatores macroeconômicos – podem ser questionáveis.

As características microeconômicas, na maioria das vezes, são aquelas relacionadas à demografia, ao mercado de trabalho e à educação (HAMMILL, 2005). Segundo Garner e Terrell (2001), entre os mecanismos de transmissão, a composição da família (característica demográfica) pode ter uma função importante na determinação da desigualdade. Esses autores analisaram a mudança na desigualdade de renda na Eslováquia no período de 1988 a 1996 e seus resultados sugerem que as alterações nas características demográficas têm contribuído para aumentar a desigualdade do país.

Da mesma forma, o trabalho de Cowell e Jenkins (1995), que analisaram a economia norte-americana, sugerem que entre 20% e 30% (dependendo do modelo utilizado) da desigualdade de renda é explicada por características demográficas (sexo, raça e idade).

Leibbrandt e Woolard (2001) analisaram a relação entre mercado de trabalho e desigualdade de renda familiar na África do Sul. Os resultados demonstram que a desigualdade de renda é extremamente relacionada ao acesso ao mercado de trabalho e às variações nos salários de mercado. Por meio da decomposição do índice de Gini, esses autores mostraram que a renda do trabalho (salário) é a principal causa da desigualdade de renda na África do Sul. Fatores relacionados à demografia também se mostraram importante para explicar a desigualdade.

Fields *et al.* (2003) estudaram a dinâmica da renda *per capita* das famílias usando dados da Indonésia, África do Sul, Espanha e Venezuela. Esses autores mostraram que mudanças no mercado de trabalho que afetem o chefe da família é o fator mais importante para mudanças na renda familiar.

Outro fator importante para explicar a desigualdade, que é amplamente reconhecido na literatura, é a educação. Diversos trabalhos mostram que a educação e a experiência estão intimamente relacionadas com a desigualdade de renda ¹¹. De uma perspectiva teórica, quanto maior a disparidade educacional, maior é a desigualdade de renda, visto que o primeiro tende a ampliar a diferença de rendas (DUMAN, 2008). De acordo com Langoni (1973), o crescimento da desigualdade de renda brasileira, a partir da década de 1960, é decorrente, basicamente, do diferencial de salários entre os trabalhadores mais qualificados e menos qualificados. Em outras palavras, para Langoni, o crescimento da taxa de retorno da escolaridade foi o principal fator para o agravamento da distribuição de renda.

Cabe mencionar, segundo Alejos (2003), que na literatura é possível encontrar várias teorias que têm como objetivo esclarecer o fenômeno da desigualdade de renda, podendo dividi-las em dois grupos: aquelas que focam as características dos trabalhadores; e aquelas que enfatizam a estrutura do mercado de trabalho. Podem-se classificar, dentro do primeiro grupo, as teorias do capital humano e da discriminação no mercado de trabalho. No segundo grupo, estão as teorias dos diferenciais compensatórios e salário-eficiência e a teoria do mercado de trabalho dual. A seguir, serão detalhadas essas teorias.

2.1. Teoria do capital humano

O departamento de assuntos econômicos das Nações Unidas definiu investimento em capital humano como investimento que tem como objetivo aumentar a produtividade do fator trabalho. Assim, a produção futura de um país pode se desenvolver não somente pelo crescimento do capital físico, mas também através de investimento em educação, em treinamento da mão de obra, na aquisição de conhecimento, na melhoria da saúde e no padrão de vida dos trabalhadores, entre outros (UNITED NATIONS, 1953 *apud* DI BARTOLO, 1999). De acordo com Becker (1962), todos esses investimentos têm a capacidade de melhorar as habilidades físicas e mentais das pessoas e, portanto, proporcionar maiores rendimentos.

De forma geral, conforme Schultz (1960), a educação pode ser tratada como um investimento no ser humano e pode-se considerar suas consequências como uma

¹¹ Ver Morley (2001).

forma de capital. Dado que a educação torna-se parte integral da pessoa que a recebe, ela pode ser entendida como capital humano.

A teoria do capital humano passou a ter um papel importante na teoria econômica a partir dos trabalhos de Mincer (1958), Becker (1962) e Schultz (1960). Até então, o pensamento econômico se desenvolveu sem dar muita importância às variáveis relacionadas ao capital humano¹². Entre as variáveis mais importantes, conforme Becker (1993), pode-se citar a educação e o treinamento realizado no local de trabalho (*on-the-job training*). Essa importância está no fato de que, ao adquirir educação (formal) e treinamento, os trabalhadores obtêm mais conhecimentos, proporcionando-lhes a capacidade de analisar e resolver, de forma mais eficiente, os problemas que podem surgir durante o trabalho.

Ao se deter somente na questão da escolaridade formal, é possível exemplificar, conforme Mincer (1974), a aplicação da teoria do capital humano à decisão dos agentes econômicos em investir em escolaridade e como essa decisão afeta seus rendimentos. No modelo de Mincer, o agente tem duas opções: (i) frequentar a escola durante s anos e obter uma renda de Y_s após terminar seus estudos (a renda do trabalhador é constante ao longo de sua vida e depende somente da quantidade de educação) ou (ii) ingressar no mercado de trabalho sem adquirir educação (ou possuir um nível que é comum a todos os agentes) e obter uma renda de Y_0 (ver Figura 5). Enquanto está na escola o agente deixa de ganhar Y_0 . Durante esse período, esse agente incorre em despesas relacionadas à educação (por exemplo, livros, taxas de matrícula etc.), totalizando um montante equivalente a K . Todavia, quando o indivíduo terminar seus estudos, ele irá obter um benefício, isto é, o agente ganhará, durante toda a vida de trabalhador, uma renda de Y_s , que é maior que Y_0 .

¹² De acordo com Fernandes (2002), os economistas Adam Smith e Alfred Marsall já tinham abordado o conceito de capital humano; porém, foi a partir da década de 1960 que essa teoria se desenvolveu.

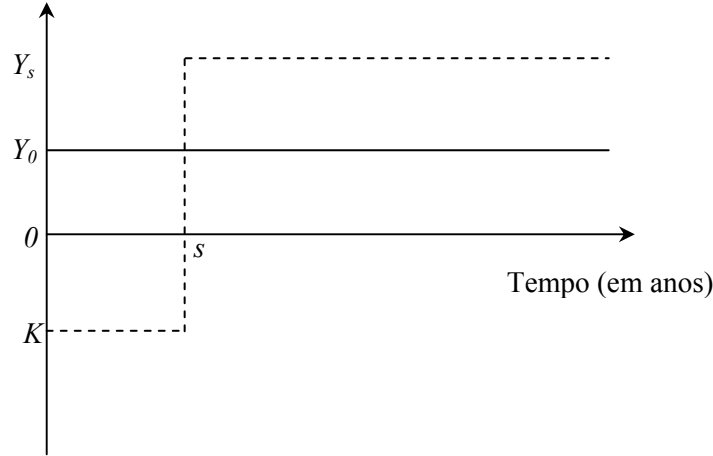


Figura 5 – Educação como investimento
Fonte: Hartog e Oosterbeek (2007).

Matematicamente, o modelo de Mincer (1974) pode ser derivado, considerando um indivíduo que adquire s anos de estudo, da seguinte maneira:

$$V_s = Y_s \int_s^n e^{-rt} dt = \frac{Y_s (e^{-rs} - e^{-rn})}{r} \quad (1)$$

em que:

Y_s = Renda do agente, cujo período investido em educação (em anos) é s ;

V_s = Valor presente do fluxo de renda do agente, obtido nos $n-s$ períodos;

n = Número máximo de períodos que o indivíduo dedica-se ao trabalho, cujo valor é igual para todos os agentes econômicos;

r = Taxa de desconto; e

t = Anos de estudos, que varia de 0 a n .

De forma similar, pode-se derivar o valor presente do fluxo de renda para um indivíduo que investiu uma quantidade de educação inferior a s . Suponha, então, que essa pessoa tenha estudado d anos menos que aquela que frequentou a escola por s anos. Assim, tem-se:

$$V_{s-d} = \frac{Y_{s-d}}{r} (e^{-r(s-d)} - e^{-rn}) \quad (2)$$

em que V_{s-d} é o valor presente do fluxo de renda do agente que adquiriu $s-d$ anos de estudo.

No intuito de obter a razão, $K_{s/s-d}$, entre a renda do indivíduo que adquire s anos de estudos e da pessoa que estuda $s-d$ anos, iguala-se (1) com (2). Assim, obtém-se:

$$K_{s/s-d} = \frac{Y_s}{Y_{s-d}} = \frac{e^{-r(s-d)} - e^{-rn}}{e^{-rs} - e^{-rn}} = \frac{e^{r(n+d-s)} - 1}{e^{r(n-s)} - 1}. \quad (3)$$

Conforme Mincer (1974), é possível observar (3) e retirar as seguintes afirmações a respeito de $K_{s/s-d}$: (i) seu valor é maior que a unidade; (ii) K é positivamente relacionada com r ; e (iii) K é negativamente relacionada com n . A primeira afirmativa diz que o indivíduo que adquiriu mais educação possui maior renda. A segunda estabelece a relação entre K e r , isto é, quanto maior a taxa intertemporal de desconto, maior é a diferença entre os rendimentos dos indivíduos, considerando os mais educados e os menos educados. Por fim, a última afirmativa estabelece que a diferença entre os rendimentos é maior quanto menor é o ciclo de vida do indivíduo (n), visto que o custo da escolaridade tende a ser recuperado ao longo de um relativo curto período de tempo.

É importante frisar que a renda pode ser melhorada, de acordo com a equação (3), por meio da educação, porém existem custos para se adquirir instrução. Caso não existissem, a demanda por educação seria ilimitada. Diante disso, torna-se relevante conhecer a quantidade de educação que maximize o fluxo de renda do agente econômico. Para isso, basta derivar (1) com respeito a s (variável de controle) e igualar o resultado a zero. Assim, tem-se: $r = \frac{Y'_s}{Y_s}$.

Note que $\frac{Y'_s}{Y_s}$ é a taxa de retorno de uma unidade adicional na escolaridade para o agente que tem s anos de escolaridade. Dessa forma, o agente irá adquirir escolaridade até o ponto em que a taxa interna de retorno seja igual a taxa de desconto.

Entre os motivos que fazem os agentes adquirirem quantidades distintas de educação, tomando como base esse modelo, pode-se citar a taxa intertemporal de desconto. Quanto maior essa taxa (as pessoas são mais impacientes), menos educação é adquirida. Outro ponto que deve ser considerado é o mercado de crédito. Neste modelo, o agente não tem dificuldade para financiar seus estudos, ou seja, o mercado de crédito é perfeito. Todavia, no mundo real, isso não ocorre. Pode ocorrer que algum agente não obtenha crédito para custear seus estudos, influenciando no montante ótimo de educação que, por sua vez, afeta os rendimentos.

Além disso, é importante destacar o papel da família no montante que seus integrantes adquirem de educação. De acordo com Becker (1993), os pais possuem uma grande influência na educação, na estabilidade mental e em muitas outras dimensões da vida de suas crianças. Considerando essa lógica, a situação financeira dos pais tem certa influência no rendimento dos seus filhos. Conforme Becker (1993), os rendimentos dos pais e dos filhos são positivamente relacionados, porém a relação não é forte. No que tange às famílias pobres, essa relação é mais acentuada. Em outras palavras, as famílias com boa condição financeira podem financiar os estudos de seus filhos, incluindo os rendimentos que eles deixam de ganhar quando estão estudando. Por outro lado, as famílias pobres não têm condições de pagar os estudos (ou treinamentos) de suas crianças. Uma solução para isso é o governo financiar os estudos dessas pessoas.

O número de crianças que os pais têm e os gastos por crianças tende a ser negativamente relacionados. Esse fato, também, tem um papel importante no volume de educação adquirido. Famílias grandes tendem a gastar bem menos com educação e treinamento do que as pequenas (BECKER, 1993).

Como já enfatizado, não somente as variáveis relacionadas ao capital humano podem estar associadas aos diferenciais de rendimento, como também a discriminação no mercado de trabalho pode influenciar nesse quesito. A seguir, será apresentada a teoria da discriminação no mercado de trabalho.

2.2. Teoria da discriminação no mercado de trabalho

Discriminação no mercado de trabalho pode ser entendida como uma situação na qual indivíduos ou grupos com habilidades produtivas (produtividades) iguais, na média, não recebem remuneração, na média, iguais (LUNDBERG; STARTZ, 1983).

Cabe mencionar que pessoas cuja produtividade seja relativamente baixa tendem também a ter salários relativamente baixos; isso, portanto, não pode ser considerado um tipo de discriminação. Por outro lado, pessoas que têm seus salários extremamente baixos, comparados à sua produtividade, podem estar sofrendo discriminação. Então, de acordo com Fernandes (2002), haverá discriminação no mercado de trabalho se pessoas ou grupos de indivíduos são diferenciados pelo mercado com base em características não-produtivas.

De acordo com Lundberg e Startz (1983), há dois tipos de modelos quando se trata de discriminação no mercado de trabalho: preferência por discriminação e discriminação estatística.

2.2.1. Preferência por discriminação

Conforme o modelo de Becker (1957), a preferência por discriminação produz diferencial de rendimento baseado nas preferências dos consumidores, dos trabalhadores e dos empregadores¹³.

2.2.1.1. Discriminação pelos empregadores

No modelo de Becker (1957), assume-se um ambiente de concorrência perfeita e função de produção com retornos constantes à escala. Nesse cenário, considera-se dois grupos de trabalhadores, W e N , que possuem as mesmas produtividades. Na ausência de qualquer tipo de discriminação e nepotismo, a taxa de salário de equilíbrio do grupo W seria o mesmo do grupo N . Todavia, assume-se que os empregadores do grupo W discriminam os indivíduos do grupo N . Becker (1957) definiu um coeficiente (d) que mensura a intensidade da discriminação. Se esse coeficiente fosse menor que zero, ter-se-ia uma situação de nepotismo; caso contrário ($d > 0$), a discriminação estaria presente no mercado de trabalho. Diante de uma taxa de salário real w , o empregador preconceituoso age como se o salário fosse $w(1 + d)$.

Conforme Becker (1957), cada empregador compara a intensidade de sua preferência por discriminar com o custo decorrente da discriminação. Assim, o processo de contratação do empregador discriminador segue a seguinte lógica:

- i. $w_W < w_N(1 + d)$: somente indivíduos do grupo W são contratados, dado que a intensidade da preferência por discriminar seja maior que os custos;
- ii. $w_W > w_N(1 + d)$: somente indivíduos do grupo N são contratados, dado que a intensidade da preferência por discriminar seja menor que os custos; e

¹³ Becker (1957) foi o primeiro autor a introduzir o modelo econômico de discriminação.

- iii. $w_W = w_N(1 + d)$: tantos os trabalhadores do grupo W como do grupo N são contratados, dado que a intensidade da preferência por discriminar seja igual ao custo.

Formalmente, pode-se tratar dessa questão da seguinte maneira. Seja $y = f(n)$ uma função de produção, em que n é o número total de trabalhadores ($n = n_W + n_N$). Desse modo, o problema do empregador discriminador é maximizar sua função de lucro:

$$\pi = pf(n) - w_W n_W - w_N n_N - dw_N \quad (4)$$

Considerando o preço, p , igual à unidade, têm-se as seguintes condições de primeira ordem: $\frac{\partial f(n)}{\partial n_W} = w_W$ e $\frac{\partial f(n)}{\partial n_N} = w_N + d$. Dividindo um pelo outro, obtém-se:

$$\frac{Pmg_W}{Pmg_N} = \frac{w_W}{(w_N + d)} \quad (5)$$

em que Pmg_W e Pmg_N são, respectivamente, as produtividades marginais do grupo W e do grupo N . Num cenário onde há discriminação, a demanda por trabalhadores do grupo N é menor e o custo unitário de produção será maior (comparado à situação onde não houvesse discriminação), visto que a demanda por W é maior. Portanto, se o empregador tem preferência por discriminar, ele tem que estar disposto a pagar por esse preconceito.

Conforme Fernandes (2002), há a possibilidade de que os empregadores tenham diferentes níveis de discriminação. Então, considera-se, de acordo com essa autora, dois empregadores com diferentes custos de discriminação, d_1 e d_2 , em que $d_1 < d_2$. Em um ambiente competitivo, os empregadores com maiores níveis de discriminação tendem a desaparecer, visto que seus custos são maiores. Assim, em um ambiente competitivo, não haveria discriminação.

Outra possibilidade de discriminação no mercado de trabalho surge entre os consumidores. A seguir, é formalizado esse tipo de discriminação.

2.2.1.2. Discriminação pelos consumidores

Considere dois grupos, W e N , formados tanto por trabalhadores como por consumidores. Os indivíduos que pertencem à W discriminam aqueles que pertencem

à N , porém o inverso não ocorre. Seja P_n o preço do produto produzido ou vendido pelos trabalhadores inseridos em N . Neste caso, o consumidor oriundo do grupo W irá agir como se $P_n(1+d)$ fosse o preço líquido, em que d é o coeficiente de discriminação desse consumidor.

Segundo Becker (1957), na ausência de discriminação, os trabalhadores dos dois grupos (W e N) – que possuem a mesma produtividade – receberiam o mesmo salário de equilíbrio, porém numa situação em que há discriminação por parte do consumidor, a taxa de salário dos integrantes do grupo N seria relativamente menor do que dos membros do grupo W .

Conforme Cain (1986), a discriminação por parte dos consumidores não tem grande relevância para explicar as diferenças, na média, dos salários recebidos pelos diferentes grupos. Essa constatação, de acordo com esse autor, pode ser explicada da seguinte forma. Assume-se que os trabalhadores do grupo N possuem as mesmas habilidades que os trabalhadores do grupo W e que os consumidores, cuja maioria estão inseridos no grupo W , estão dispostos a pagar um preço, p , para um bem produzido pelos trabalhadores de W . Caso haja consumidores (preconceituosos) que tenham contato com os produtores, eles tendem a considerar o preço efetivo do bem produzido pelos trabalhadores N como sendo $p + d$, em que p é o custo de produção e d já foi definido anteriormente. Entretanto, é importante ressaltar que há diversos postos de trabalho onde os consumidores não têm contato direto com o processo de produção e, conseqüentemente, com os trabalhadores. Desse modo, não haveria como o consumidor diferenciar o trabalhador do grupo W do grupo N , impedindo que houvesse diferenças salariais entre os grupos.

Neste cenário, os trabalhadores do grupo N iriam se especializar na produção de bens cujo contato com os consumidores não existisse, evitando, desse modo, receber um salário que seja relativamente mais baixo que o recebido pelos trabalhadores, de igual produtividade, do grupo W . Em última instância, não haveria discriminação, mas somente segregação (CAIN, 1986).

2.2.1.3. Discriminação pelos trabalhadores

Ao invés dos consumidores, são os trabalhadores que podem praticar ou sofrer a discriminação. Neste caso, então, d irá refletir um valor que retrata a preferência por discriminação de um trabalhador individual.

Da mesma forma que na discriminação pelos consumidores, suponha que há dois grupos: W e N . O primeiro é formado por trabalhadores que discriminam os pertencentes ao segundo grupo. Assim, conforme Becker (1957), se o salário bruto dos trabalhadores do grupo W for w_w , o salário líquido seria $w_w(1-d)$. O salário líquido incorpora a desutilidade do trabalhador do grupo W por compartilhar o mesmo ambiente com o indivíduo do grupo N .

De acordo com Becker (1957), o custo unitário (c) da discriminação pode ser escrito como:

$$c = \frac{w_N - w_W}{w_N} \quad (6)$$

Caso o custo unitário da discriminação seja maior que d , o trabalhador que pertence ao grupo W suporta seu preconceito e aceita trabalhar com algum trabalhador do grupo N . Por outro lado, caso c seja menor que d , o trabalhador não aceita trabalhar com algum indivíduo do grupo N . Se $c=d$, há indiferença, por parte do trabalhador do grupo W , em trabalhar ou não com algum trabalhador do grupo N .

A discriminação pode ser tratada utilizando outro enfoque: a discriminação estatística. A seguir, é apresentada essa abordagem.

2.2.2. Discriminação estatística

Os modelos de discriminação estatísticos, segundo Lundberg e Startz (1983), demonstram que a discriminação no mercado de trabalho pode ser uma resposta racional por parte das firmas frente à incerteza sobre a produtividade dos trabalhadores.

Entre os modelos que se enquadram nessa linha, pode-se citar o modelo de Spence (1973). No modelo de Spence, a educação é usada para diferenciar os trabalhadores que têm alta produtividade daqueles que possuem baixa produtividade. Spence (1973) propôs um exemplo numérico com o intuito de discutir a existência e as propriedades de seu modelo.

Supõe-se que no mercado de trabalho existam dois tipos de trabalhadores: aqueles com alta produtividade, igual a 2; e aqueles com baixa produtividade, igual a 1. Assume-se que as firmas tenham a crença de que o trabalhador com alto rendimento produtivo tem um nível de educação igual a $y \geq y^*$, cujo salário seria igual a 2; e os que possuem baixa produtividade igual a $y < y^*$, cujo salário seria

igual a 1. Pressupõe-se também que o custo da educação para os trabalhadores bons (alta produtividade) seja de $y/2$ e para os trabalhadores ruins (baixa produtividade) seja de y .

Dados os salários vigentes no mercado, cada tipo de trabalhador irá escolher o nível de educação que maximize sua utilidade. Dessa forma, os trabalhadores irão escolher $y=0$ ou $y=y^*$. Isso é justificado pelo fato de que o salário recebido pelos trabalhadores ruins decorrente de qualquer $y \in (0, y^*)$ é o mesmo que o recebido se $y=0$, porém o custo é maior; por analogia, não é racional escolher $y > y^*$.

A crença das firmas de que o trabalhador bom irá escolher $y=y^*$ e o trabalhador ruim $y=0$ somente será confirmada se:

$$\begin{cases} 2 - \frac{y^*}{2} \geq 1 - 0 & \text{para o trabalhador bom} \\ 1 - 0 \geq 2 - y^* & \text{para o trabalhador ruim} \end{cases}$$

ou seja, $1 \leq y^* \leq 2$.

Desde que algum nível de educação no intervalo $[1, 2]$ seja possível, haverá um equilíbrio em que cada tipo de trabalhador irá escolher um diferente tipo de sinalização e, conseqüentemente, terá rendimentos diferentes.

2.3. Teorias dos diferenciais compensatórios e salário-eficiência

2.3.1. Teoria dos diferenciais compensatórios¹⁴

Segundo Alejos (2003), a teoria dos diferenciais compensatórios justificaria as diferenças de rendimento existentes entre postos de trabalho, considerando-as como uma consequência dos distintos ambientes de trabalho. Por exemplo, trabalhos que envolvem potenciais danos à saúde necessitariam oferecer compensação aos trabalhadores de forma a atraí-los ao posto de trabalho¹⁵.

De forma a facilitar a exposição do modelo de diferenciais compensatórios, considere um mercado composto por trabalhadores igualmente produtivos (mesma produtividade). Suponha que existam somente duas categorias de emprego. Seja D um índice que identifica o tipo de trabalho, com $D = 0$ ou 1. Trabalho do tipo 1 está

¹⁴ Essa seção baseia-se em Rosen (1986).

¹⁵ A teoria dos diferenciais compensatórios é baseada na pressuposição de que os agentes têm informações perfeitas a respeito do mercado.

associado a empregos que envolvem algum risco de acidentes. De modo contrário, trabalho do tipo 0 envolve empregos sem riscos.

Admite-se que a preferência do trabalhador é definida com base na sua função de utilidade $U = U(C, D)$, em que U é a utilidade do indivíduo, C é um bem de consumo (que pode ser interpretado como o salário real) e D é o índice que mensura a probabilidade de acidentes. Pressupõe-se que a utilidade marginal com respeito a C é positiva e, com relação a D , negativa; ou seja, $\frac{\partial U}{\partial C} > 0$ e $\frac{\partial U}{\partial D} < 0$. Para um dado C , é natural assumir que $U(C, 0) \geq U(C, 1)$. Dessa forma, o trabalhador prefere ambientes de trabalho que não exista possibilidade de acidentes, isto é, $D = 0$. Seja C_0 o consumo de mercado (ou salário real) quando $D = 0$. Assim, dado C_0 , é possível definir um C^* , que seria o consumo necessário para que os trabalhadores do emprego tipo 1 alcançasse o nível de utilidade que C_0 garante às pessoas que estão inseridas no trabalho tipo 0. Em outras palavras, é possível definir um C^* de modo que $U(C^*, 1) = U(C_0, 0)$. Uma vez que $D = 1$ não é preferido – considerando o mesmo nível de consumo – à $D = 0$, tudo o mais mantido constante, tem-se $C^* \geq C_0$. Diante disso, $Z = C^* - C_0$ seria a compensação necessária para tornar o trabalhador indiferente entre os dois tipos de empregos.

Seja $\Delta W = W_1 - W_0$ o diferencial compensatório oferecido pelo mercado. Assim, o emprego tipo 1 proporciona ΔW unidades adicionais de C para o agente que escolher $D = 1$. Todavia, o trabalhador irá escolher o tipo de trabalho que maximiza sua utilidade. Então, a regra de decisão seria: escolher $D = 1$ se $U(\Delta W + C_0, 1) > U(C_0, 0)$ e $D = 0$ se $U(\Delta W + C_0, 1) < U(C_0, 0)$. Numa situação onde as utilidades são iguais, o trabalhador é indiferente entre o emprego do tipo 1 e o do tipo 0.

Há outro modo de analisar a regra de escolha do trabalhador sem ter que utilizar, explicitamente, a função utilidade, isto é: $D = 1$ será escolhido se $\Delta W > Z$ e $D = 0$ será a escolha se $\Delta W < Z$. É importante enfatizar que ΔW é o diferencial compensatório oferecido pelo mercado e que Z é a compensação exigida pelo trabalhador (via função utilidade) para torná-lo indiferente entre $D = 0$ e $D = 1$.

Para encontrar o equilíbrio nesse modelo, é necessário analisar a oferta e a demanda de trabalho. A seguir, é feita essa análise.

2.3.1.1. Oferta de trabalho

Em virtude do número restrito de opções disponíveis para os trabalhadores em termos de emprego, $D = 0$ e $D = 1$, as condições de sua oferta são determinadas pelos trabalhadores cujo diferencial compensatório oferecido pelo mercado (ΔW) seja maior que a compensação exigida pelo trabalhador (Z) e por aqueles em que ΔW seja menor que Z . Embora ΔW seja o mesmo para todos os empregados, Z é uma variável que depende das preferências inerentes ao trabalhador, ou seja, Z varia de pessoa para pessoa. Diante disso, é conveniente retratar as diferenças, em termos de preferências, entre os trabalhadores de forma estocástica. Assim, considere $g(Z)$ como a função densidade de preferências na população de trabalhadores e $G(Z)$ como a função de distribuição acumulada. Para um dado valor de ΔW , todos aqueles que aceitam o emprego tipo 1 satisfazem a condição $\Delta W > Z$. Algebricamente, tem-se:

$$N_1^s = \int_0^{\Delta W} g(Z) dz = G(\Delta W) \quad (7)$$

em que N_1^s é a fração de trabalhadores que aceita o tipo de emprego $D = 1$.

Por outro lado, a fração de trabalhadores que aceita o emprego tipo 0 pode ser representado por:

$$N_0^s = \int_{\Delta W}^{\infty} g(Z) dz = 1 - G(\Delta W). \quad (8)$$

De forma análoga à oferta, o lado da demanda pode ser desenvolvido algebricamente, o que será trabalhado a seguir.

2.3.1.2. Demanda de trabalho

Assim como o trabalhador, as firmas têm uma escolha binária: oferecer um ambiente de trabalho do tipo 0 (seguro) ou 1 (não seguro). Essa escolha é definida com base no custo/benefício de se ofertar um local de trabalho sem risco. Em outras palavras, a firma deve comparar os custos de prover um emprego do tipo 0 com os benefícios que isso poderia gerar. Lembre-se que o salário em um ambiente seguro ($D = 0$) é menor do que em um local com risco de acidentes. Numa situação onde o benefício exceda os custos, a firma irá escolher oferecer o trabalho $D = 0$. Por outro lado, se o custo for maior do que o benefício, a firma oferecerá o emprego $D = 1$.

Formalmente, considere B como o custo (por trabalhador) necessário para que o local de trabalho seja seguro, isto é, com boas condições de trabalho. Dessa forma, a regra de decisão da firma será: escolhe $D = 1$ se $B > \Delta W$ e $D = 0$ se $B < \Delta W$.

Admitindo que B varie entre as empresas, pode-se definir $f(B)$ como a função de densidade de B e $F(B)$ sua função de distribuição acumulada. Dado que a firma oferece $D = 1$ ou $D = 0$, a fração de trabalho $D = 1$ ofertado no mercado é representada pela seguinte função.

$$N_1^d = \int_0^{\Delta W} f(B)db = 1 - F(\Delta W). \quad (9)$$

A fração remanescente de empregos do tipo $D = 1$ é representada por:

$$N_0^d = \int_{\Delta W}^{\infty} f(B)db = F(\Delta W). \quad (10)$$

Finalmente, o equilíbrio ocorre quando a oferta e a demanda por trabalhadores forem iguais nos dois tipos de empregos, ou seja, quando $N_1^d = N_1^s$ e $N_0^d = N_0^s$.

2.3.2. Teoria salário-eficiência

A teoria do salário-eficiência baseia-se na ideia de que salários que se situam acima do verificado no mercado podem aumentar a produtividade dos trabalhadores e/ou reduzir os custos para as firmas (ALEJOS, 2003). Em outras palavras, ao pagar salários altos, comparado ao verificado em outras empresas, a firma estaria motivando seus empregados e, por consequência, aumentando sua produtividade. Todo esse processo seria resultado do seu comportamento maximizador de lucro (RIVEROS; BOUTON, 1991).

De acordo com Yellen (1984), a relação entre salário e eficiência pode ser mostrada por meio de um modelo simples. Assim, considera-se uma economia formada por firmas idênticas que atuam em um ambiente competitivo, sendo que cada firma tem a seguinte função de produção:

$$Q = F(e(w) \cdot N), \quad (11)$$

em que w é o salário real, e é o nível de esforço dos trabalhadores, cuja derivada em relação a w é positiva e N é o número de trabalhadores. As firmas podem contratar

toda a mão de obra que desejarem ao salário que escolherem oferecer. O salário ótimo, w^* , que as firmas escolherão deverá satisfazer a seguinte condição:

$$e'(w) \frac{w}{e(w)} = 1. \quad (12)$$

Em outras palavras, o salário ótimo deve satisfazer a condição de que a elasticidade do esforço do trabalho em relação ao salário real seja igual à unidade. O salário ótimo w^* , portanto, é conhecido como salário-eficiência e é independente das condições prevalentes no mercado de trabalho, ou seja, alteração nos salários de outras firmas não irá levar a firma a mudar a sua taxa de salário, visto que ela é determinada pela relação expressa na equação (12).

A teoria do mercado de trabalho dual – que será discutida na próxima seção – pode ser explicada pela pressuposição que o nexos entre salário e produtividade é importante em alguns setores da economia, porém não em outros (YELLEN, 1984).

Para o setor institucional (moderno ou formal), onde a hipótese de salário-eficiência é relevante, encontram-se firmas que pagam salários acima do verificado no mercado. Por outro lado, no setor informal (atrasado), onde a relação entre salário e produtividade é fraca ou inexistente, deve-se observar um comportamento completamente neoclássico (YELLEN, 1984).

2.4. Teoria do mercado de trabalho dual

Conforme a teoria do mercado de trabalho dual – ou teoria da segmentação do mercado de trabalho –, o mercado possui dois segmentos. Os trabalhadores que pertencem ao primeiro – institucional, moderno ou formal – recebem salários relativamente altos, possuem sistema de promoção na carreira e têm estabilidade no emprego. Por outro lado, os trabalhadores que pertencem ao segundo segmento – atrasado, tradicional ou informal – recebem baixos salários, não possuem estabilidade e nem sistema de promoção na carreira e mudam de emprego com maior frequência.

Os economistas Doeringer e Piore são mencionados como os pioneiros da teoria do mercado de trabalho dual¹⁶. Foram eles que introduziram o conceito de Mercado Interno de Trabalho (MIT), correspondente ao setor institucional. Define-se MIT como uma unidade administrativa onde há um conjunto de regras e

¹⁶ Ver Doeringer e Piore (1971).

procedimentos que governam a contratação e a remuneração da mão de obra. A diferença com relação ao Mercado Externo de Trabalho, que corresponde ao setor informal (atrasado), é que o último é governado pelas forças de mercado e não por regras administrativas; as contratações e remunerações da mão de obra são governadas diretamente por variáveis econômicas (DOERINGER; PIORE, 1971).

Os trabalhadores do mercado interno possuem algumas vantagens e direitos – por exemplo, promoções, oportunidade de treinamento, altos salários, estabilidade etc. – que não estão disponíveis aos trabalhadores do mercado externo. Diante das condições mais favoráveis, os trabalhadores do mercado interno ficam motivados a se aperfeiçoar ainda mais e, por extensão, reduzem a mobilidade externa.

Conforme Doeringer e Piore (1971), as firmas no mercado interno investem em ativos específicos, o que requer, por conseguinte, trabalhadores mais qualificados. Porém, tal qualificação não depende somente da educação formal, mas também do treinamento recebido no trabalho e na experiência adquirida com o tempo. Quanto mais específico for o trabalho, maior será o custo de treinamento, o que gera maior estabilidade e uma menor rotatividade da mão de obra. Diante disso, o empregador oferece altos salários para manter o trabalhador em seu posto, visto que não é lucrativo, dado o alto custo de treinamento, ter alta rotatividade na firma.

Diante do exposto, um dos motivos que levam a economia a ter grandes diferenciais de rendimentos seria essa divisão no mercado de trabalho. Indivíduos que trabalham no mercado externo – informal ou tradicional – possuem pouca probabilidade de ter acesso ao setor que proporciona maiores rendimentos.

É importante fazer uma ressalva a respeito das teorias que foram apresentadas. Todas elas se complementam para explicar o fenômeno da desigualdade de renda, visto que não é possível atrelar somente uma teoria à iniquidade distributiva. Assim, a ligação que há entre elas vem do fato de que todas buscam explicar, a partir de diferentes abordagens, o diferencial de rendimentos.

2.5. Pobreza, crescimento e desigualdade

A redução na pobreza é determinada pela taxa de crescimento da renda média da população e pela sua redistribuição (queda na desigualdade). Esse fato, conforme Barros e Mendonça (1997), é uma das grandes razões pelas quais os objetivos das políticas econômicas focam na busca do crescimento econômico e da equidade.

Entretanto, dependendo do país, essas políticas são mais voltadas para o crescimento ou para a redistribuição. No intuito de escolher as políticas ótimas – as mais eficazes – que devem ser adotadas, é necessário conhecer o impacto de ambas as variáveis, crescimento e redistribuição, sobre a pobreza. Essa escolha depende do conhecimento de dois parâmetros (Figura 6): (i) o conjunto de políticas disponíveis no país (parte sombreada da Figura 1), sendo que todas elas são em termos de efeitos sobre o crescimento e o nível de desigualdade; e (ii) estimativas da relevância relativa do crescimento econômico e de reduções no nível de concentração sobre a pobreza. Essas estimativas servem para informar sobre o que é desejável no combate à pobreza.

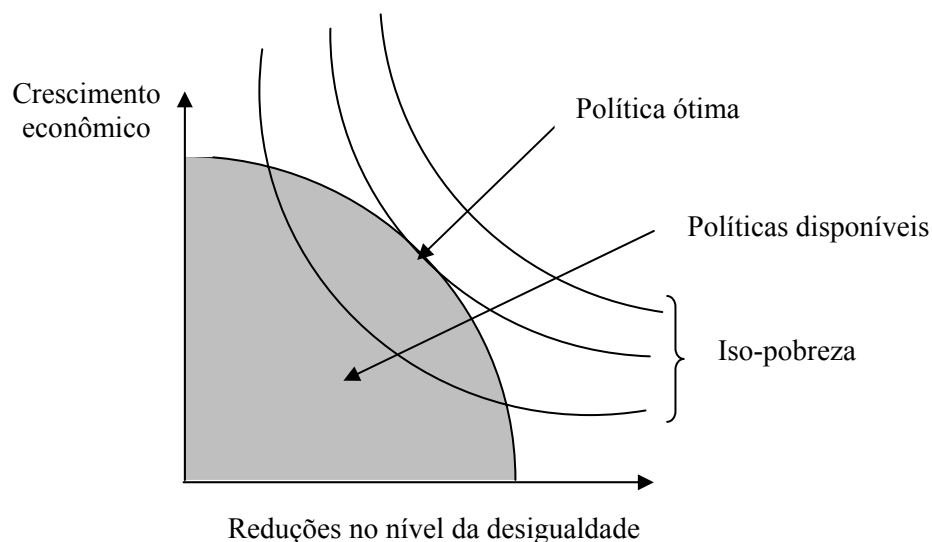


Figura 6 – Política ótima no combate à pobreza

Fonte: Barros e Mendonça (1997).

Conforme Barros e Mendonça (1997), a obtenção desses dois parâmetros não é uma tarefa fácil. Neste trabalho, então, o foco ficou no segundo, isto é, na obtenção das estimativas da importância do crescimento e da redistribuição sobre a pobreza.

É importante enfatizar que não há uma única estrutura teórica que engloba, simultaneamente, os diferentes componentes do triângulo pobreza, desigualdade de renda e crescimento econômico. Talvez a justificativa para isso esteja na dificuldade de relacionar crescimento econômico, que tem uma conotação macroeconômica, e pobreza, que possui um enfoque mais microeconômico. A relação entre os três componentes é feita, na literatura, basicamente em nível empírico, utilizando rendimento como *proxy* para crescimento econômico. Embora não haja uma estrutura teórica, é possível retirar, por meio de trabalhos empíricos, algumas conclusões a respeito da inter-relação entre esses três fenômenos econômicos. Kraay (2006), por exemplo, decompôs as mudanças (variações) na pobreza em três componentes (alta taxa de crescimento da renda média, alta sensibilidade da pobreza para o crescimento da renda média e um padrão de crescimento das rendas relativas que reduz a pobreza), utilizando uma amostra contendo países em desenvolvimento durante os anos de 1980 e 1990. De acordo com seus resultados, no médio e longo prazo, a maior parte das mudanças na pobreza é atribuída às variações nas rendas médias, o que sugere que políticas e instituições que busquem promover o crescimento econômico deveria ser o foco central dos gestores públicos. O restante da variação na pobreza seria explicado por alterações na renda relativa.

Embora o crescimento econômico seja importante para a queda da pobreza (KRAAY, 2006), não é possível interpretá-lo como o único fator associado à redução dos números de pobres. Heltberg (2002) enfatizou que o crescimento é uma condição necessária para o alívio da pobreza, porém o grau de desigualdade de renda também possui uma função importante nesse quesito. Crescimento econômico e iniquidade distributiva estão inter-relacionados, sendo que a eficiência com que o primeiro afeta a pobreza depende de como a renda é distribuída. Conforme Heltberg (2002), uma distribuição de renda desigual é um sério impedimento para que haja redução do número de pobres. De forma geral, o efeito do crescimento da renda sobre a pobreza é, em si, uma função do grau de desigualdade. Isto é, quanto menos desigual for a distribuição de rendimentos maior tende a ser o impacto gerado pelo progresso econômico (elevação da renda).

Ravallion (1997) também forneceu uma importante contribuição nesse sentido. As perguntas que esse autor buscou responder foram as seguintes: as pessoas pobres que moram em países cuja desigualdade de renda é elevada possuem as mesmas perspectivas de escapar da pobreza como aquelas que vivem em países com baixa desigualdade? Há a possibilidade de que a desigualdade, em um patamar muito elevado, impeça a redução da pobreza absoluta, mesmo havendo condições iniciais e políticas que são favoráveis ao crescimento econômico? Em última instância, essas perguntas buscam relacionar os três componentes supracitados (pobreza, desigualdade e crescimento econômico). Por meio de um modelo empírico, Ravallion (1997) observou que a alta desigualdade tende a reduzir o impacto favorável do crescimento sobre a pobreza. Conforme os resultados desse autor, a elasticidade com respeito ao crescimento econômico reduz-se, fortemente, com o aumento da concentração de renda. Em outras palavras, o efeito do crescimento da renda sobre a pobreza é menor em países com alta desigualdade do que em países com baixa iniquidade distributiva. Nesse cenário, políticas que busquem conciliar crescimento e redistribuição seria a melhor forma de combater a pobreza, como foi enfatizado por Barros e Mendonça (1997). No entanto, esses autores ressaltaram que, no caso do Brasil, o alto índice de desigualdade permitiria que a pobreza fosse diminuída, de forma significativa, sem que houvesse crescimento econômico, necessitando, entretanto, a redução dos níveis de concentração de renda a patamares similares aos observados nos países da América Latina.

Diante do exposto, pode-se dizer que os objetivos da política governamental não podem focar somente no crescimento econômico ou na redução da desigualdade de renda, sendo necessário um ponto de equilíbrio. Em virtude disso é que se deve identificar que fatores são os mais importantes para influenciar o crescimento da renda e a redução de sua desigualdade.

3. METODOLOGIA

Para alcançar os objetivos propostos, foram utilizados dois métodos de decomposição, ambos baseados no valor de Shapley. O primeiro divide uma dada variação na pobreza em dois componentes, crescimento e redistribuição (desigualdade). Por meio dessa metodologia, foi possível trabalhar a segunda hipótese desta pesquisa. O segundo método é aplicado em um modelo econométrico (regressão) que permite determinar os rendimentos das pessoas ocupadas nos estados da Região Nordeste, possibilitando, por conseguinte, testar a veracidade da segunda hipótese.

Como é possível observar, o valor de Shapley é de extrema importância para as decomposições que foram realizadas. Dessa forma, primeiramente, apresenta-se uma breve discussão a respeito desse ramo da teoria dos jogos.

3.1. Valor de Shapley¹⁷

O valor de Shapley é um conceito de solução para jogos cooperativos que permite ou sugere a divisão de lucros ou custos comuns advindos de uma grande coalizão. Seja $N = \{1, 2, \dots, m\}$ um conjunto finito de jogadores. Então, um subconjunto não-vazio de N é chamado de coalizão.

Para cada coalizão $S \subseteq N$, $v(S)$ é um valor, positivo ou negativo, que está disponível para a divisão entre os membros da coalizão S . Essa divisão é realizada com base na contribuição marginal média de cada jogador à coalizão.

Conforme Osborne e Rubinstein (1994), a contribuição marginal do jogador i para qualquer coalizão S , com $i \notin S$, em um dado jogo, pode ser mensurada da seguinte maneira:

¹⁷ Para mais detalhes sobre valor de Shapley, ver Osborne e Rubinstein (1994).

$$\Delta_i(S) = v(S \cup \{i\}) - v(S). \quad (13)$$

Assim, o valor de Shapley do jogador i , denotado por $\phi_i^S(v)$, é dado por:

$$\phi_i^S(v) = \frac{1}{|N|!} \sum_{R \in \mathfrak{R}} \Delta_i(S_i(R)), \text{ para cada } i \in N, \quad (14)$$

em que $\mathfrak{R}$ é o conjunto de todas ordenações de $|N|!$ e $S_i(R)$ é o conjunto dos jogadores antecedentes a i na ordenação $\mathfrak{R}$. Por conversão, $0! = 1$ e $v(\emptyset) = 0$.

3.2. Relação entre nível de pobreza, crescimento e desigualdade¹⁸

Como ressaltado, nos últimos anos a desigualdade social no Brasil tem caído de forma expressiva. As políticas de transferência direta de renda e a estabilidade econômica têm importante função nesse processo. A queda nos índices de concentração pode, entretanto, ser acompanhado por crescimento econômico ou por recessão, o que tende a reforçar ou a reduzir o impacto das políticas de redistribuição. Neste contexto, são apresentados alguns métodos que têm como objetivo separar os efeitos crescimento e redistribuição.

3.2.1. Decomposição de Datt e Ravallion

Datt e Ravallion (1992) propuseram um método que permite mensurar a importância relativa do crescimento e da redistribuição (desigualdade) de renda na variação da pobreza, considerando dois pontos no tempo, t e $t+n$. Esta técnica de decomposição tem sido amplamente utilizada na literatura, entre eles citam-se: Helfand *et al.* (2009), Ferreira *et al.* (2008), Guimarães (2007), Chen e Wang (2001), McKay (1997), entre outros.

De forma geral, a metodologia desenvolvida por Datt e Ravallion (1992) decompõe uma dada variação na pobreza em um componente crescimento, denotado por $G(t, t+n; r)$, um componente redistribuição, denotado por $D(t, t+n; r)$ e um resíduo $R(t, t+n; r)$, em que r é um período de referência ou *benchmark* (que pode ser t ou $t+n$); t é o período inicial; e $t+n$, o período final.

O componente crescimento mensura o impacto na pobreza decorrente da mudança na renda média, mantendo a curva de Lorenz constante ao nível de referência L_r . Por sua vez, o componente redistribuição reflete o efeito de uma

¹⁸ Parte dessa seção é baseada em Baye (2006).

variação na curva de Lorenz sobre a pobreza agregada, fixando a renda média ao nível de referência μ_r . O resíduo $R(\cdot)$ surge quando a medida de pobreza não é separável aditivamente entre μ e L .

Diante do exposto, pode-se considerar uma medida de pobreza $P_t = P\left(\frac{Z}{\mu_t}, L_t\right)$ – em que μ é a renda média; L é um vetor de parâmetros que define a curva de Lorenz; e Z é a linha de pobreza – e decompor uma dada mudança em P_t da seguinte maneira:

$$P_{t+n} - P_t = G(t, t+n; r) + D(t, t+n; r) + R(t, t+n; r), \quad (15)$$

em que:

$$G(t, t+n; r) = P\left(\frac{Z}{\mu_{t+n}}, L_r\right) - P\left(\frac{Z}{\mu_t}, L_r\right),$$

$$D(t, t+n; r) = P\left(\frac{Z}{\mu_r}, L_{t+n}\right) - P\left(\frac{Z}{\mu_r}, L_r\right),$$

$$R(t, t+n; r) = G(t, t+n; t+n) - G(t, t+n; t) = D(t, t+n; t+n) - D(t, t+n; t)$$

O resíduo pode ser relativamente grande sempre que os efeitos marginais sobre a pobreza decorrente de mudanças na renda média (curva de Lorenz) dependerem fortemente do formato da curva de Lorenz (média da distribuição).

A decomposição de Datt e Ravallion (1992) apresenta algumas limitações. Esse fato propiciou o surgimento de outras técnicas, apresentadas a seguir.

3.2.2. Decomposição de Kakwani

Pelo que foi exposto, a metodologia de Datt e Ravallion (1992) inclui um resíduo e um período base (*benchmark*) para decompor a variação na pobreza em um componente crescimento e redistribuição. Isso gera um problema de assimetria, no que diz respeito ao período inicial e final. Nesse sentido, Kakwani (1997) desenvolveu um método axiomático para eliminar tanto o resíduo como a assimetria¹⁹. Em outras palavras, por meio da técnica de Kakwani (1997), a variação na pobreza, entre dois pontos quaisquer no tempo (t e $t+n$), é inteiramente explicada

¹⁹ Kakwani e Subbaro (1990) *apud* Araújo (2007) desenvolveram uma técnica de decomposição que dá a falsa impressão de que o método é exato. Os autores apenas alocaram o resíduo no componente redistribuição. Em outras palavras, os autores – logo após obter o componente crescimento – calcularam o efeito redistribuição subtraindo da variação da pobreza o componente crescimento. Em última instância, o resíduo está somado ao componente redistribuição.

pelos componentes crescimento, $G_{t, t+n}$, e redistribuição, $D_{t, t+n}$. Os axiomas propostos por Kakwani podem ser resumidos da seguinte forma:

- a) Se $D_{t, t+n} = 0$, tem-se $\Delta P_{t, t+n} = G_{t, t+n}$; e se $G_{t, t+n} = 0$, tem-se $\Delta P_{t, t+n} = D_{t, t+n}$;
- b) Se $G_{t, t+n} \leq 0$ e $D_{t, t+n} \leq 0$, tem-se $\Delta P_{t, t+n} \leq 0$; e se $G_{t, t+n} \geq 0$ e $D_{t, t+n} \geq 0$, tem-se $\Delta P_{t, t+n} \geq 0$; e
- c) Se $G_{t, t+n} = -G_{t, t+n}$ e $D_{t, t+n} = -D_{t, t+n}$, então a simetria é garantida.

O primeiro axioma estabelece que, numa situação em que o efeito crescimento seja zero, a variação na pobreza é completamente explicada pela mudança na desigualdade, isto é, pela concentração ou desconcentração de renda. Por outro lado, se o efeito redistribuição for igual a zero, a mudança na pobreza é explicada somente pelo componente crescimento. Por fim, caso os efeitos crescimento e redistribuição sejam ambos iguais a zero, a pobreza, entre o período t e $t+n$, permanece inalterada.

O segundo axioma diz que numa situação em que ambos os efeitos (crescimento e redistribuição) assumem o mesmo sinal algébrico, a variação na pobreza também apresentará o mesmo sinal. Finalmente, o terceiro axioma assegura a simetria entre os períodos inicial e final. Em outras palavras, a única diferença quando se move de $t+n$ para t é o sinal, isto é, os efeitos são os mesmos; todavia, com o sinal contrário.

De forma geral, o modelo de Kakwani (1997) mensura a contribuição da redistribuição como a média de dois efeitos:

- a) Efeito redistribuição quando o crescimento no período inicial é mantido constante; e
- b) Efeito redistribuição quando o crescimento no período final é mantido constante.

A contribuição do crescimento é mensurada seguindo essa mesma lógica.

Diante da falta de uma estrutura mais geral para as técnicas de decomposição, Shorrocks (1999) propôs um método baseado na teoria dos jogos, especificamente nos jogos cooperativos. Shorrocks (1999) utilizou um conceito de solução, denominado de valor de Shapley, para elaborar a sua metodologia de decomposição.

3.2.3. Decomposição de Shapley

Para demonstrar a decomposição é necessário, primeiramente, definir uma linha de pobreza que, neste caso, será denotada por z . Dessa forma, o nível de pobreza no tempo t pode ser expresso como uma função, $P(\mu_t, L_t)$, da renda média (μ_t) e da curva de Lorenz (L_t). O fator de crescimento na variação da pobreza, entre o período t e $t+n$, é denotado por $G = \frac{\mu_{t+n} - \mu_t}{\mu_t} = \frac{\mu_{t+n}}{\mu_t} - 1$, enquanto o fator redistribuição, que é relacionado com a curva de Lorenz, é dado por $D = L_{t+n} - L_t$.

Neste trabalho, foram adotadas as medidas de pobreza P_α , desenvolvidas por Foster *et al.* (1984). Essas medidas são amplamente utilizadas na literatura que trata da pobreza. Então, uma variação na classe P_α é dada pela seguinte equação:

$$\Delta P_\alpha = P_\alpha(\mu_{t+n}, L_{t+n}) - P_\alpha(\mu_t, L_t) \quad (16)$$

Substituindo o fator crescimento, $G = \frac{\mu_{t+n} - \mu_t}{\mu_t}$, e o fator redistribuição,

$D = L_{t+n} - L_t$, na equação (16), obtém-se:

$$\begin{aligned} \Delta P_\alpha &= P_\alpha(\mu_t(1+G), L_t + D) - P_\alpha(\mu_t, L_t) \\ &= v_\alpha(G, D) \end{aligned} \quad (17)$$

A expressão do lado direito de (17), portanto, é a função v que fornece a contribuição, conjuntamente, de G e D para a variação de P_α . O próximo passo, então, é obter, individualmente, as parcelas de G e D .

Dado que há somente dois fatores (G e D), as ordenações possíveis são $N! = 2! = 2$, ou seja: $\{G, D\}$ e $\{D, G\}$. As contribuições marginais do crescimento e da redistribuição para uma dada mudança da pobreza (ΔP) são dadas na Tabela 4:

Tabela 3 – Decomposição do valor de Shapley

Contribuições Marginais	Ordenação de N!	Cálculos das contribuições marginais
Contribuição marginal de G	{G, D}	$v(G) - v(\emptyset) = v(G)$
	{D, G}	$v(D, G) - v(D) = v(D, G) - v(D)$
	Somatório	$v(D, G) - v(D) + v(G)$
Contribuição marginal de D	{G, D}	$v(G, D) - v(G) = v(G, D) - v(G)$
	{D, G}	$v(D) - v(\emptyset) = v(D)$
	Somatório	$v(G, D) - v(G) + v(D)$

Fonte: elaborado pelo autor.

Por meio da Tabela 4 e tomando $N! = 2$ é possível encontrar o valor de Shapley referente aos componentes crescimento (G) e redistribuição (R), respectivamente:

$$\phi_G^S(v) = \frac{1}{2} [v(G, D) - v(D) + v(G)] \text{ e} \quad (18)$$

$$\phi_D^S(v) = \frac{1}{2} [v(G, D) - v(G) + v(D)]. \quad (19)$$

Na ausência de crescimento, G toma o valor de zero, e a variação na pobreza decorrente somente da redistribuição torna-se, conforme a equação (46):

$$v(D) = P(\mu_t, L_{t+n}) - P(\mu_t, L_t). \quad (20)$$

Por outro lado, assumindo que $D = 0$, tem-se:

$$v(G) = P(\mu_{t+n}, L_t) - P(\mu_t, L_t). \quad (21)$$

Substituindo as equações (17), (20) e (21) nas equações (18) e (19) encontra-se, respectivamente, a expressão completa das contribuições do crescimento e da redistribuição sobre a mudança nas classes P_α :

$$\begin{aligned} \phi_{\alpha G}^S(v) &= \frac{1}{2} \{P_\alpha(\mu_{t+n}, L_{t+n}) - P_\alpha(\mu_t, L_t) - [P_\alpha(\mu_t, L_{t+n}) - P_\alpha(\mu_t, L_t)] \\ &\quad + [P_\alpha(\mu_{t+n}, L_t) - P_\alpha(\mu_t, L_t)]\} \\ &= \frac{1}{2} \{P_\alpha(\mu_{t+n}, L_{t+n}) - P_\alpha(\mu_t, L_{t+n}) + P_\alpha(\mu_{t+n}, L_t) - P_\alpha(\mu_t, L_t)\} \end{aligned} \quad (22)$$

$$\begin{aligned} \phi_{\alpha D}^S(v) &= \frac{1}{2} \{P_\alpha(\mu_{t+n}, L_{t+n}) - P_\alpha(\mu_t, L_t) - [P_\alpha(\mu_{t+n}, L_t) - P_\alpha(\mu_t, L_t)] \\ &\quad + [P_\alpha(\mu_t, L_{t+n}) - P_\alpha(\mu_t, L_t)]\} \\ &= \frac{1}{2} \{P_\alpha(\mu_{t+n}, L_{t+n}) - P_\alpha(\mu_{t+n}, L_t) + P_\alpha(\mu_t, L_{t+n}) - P_\alpha(\mu_t, L_t)\} \end{aligned} \quad (23)$$

Diante disso, é possível verificar que a variação total na pobreza é a soma dos componentes crescimento e redistribuição dado pelas contribuições de Shapley:

$$\Delta P_{\alpha} = \phi_{\alpha G}^s(\nu) + \phi_{\alpha D}^s(\nu) \quad (24)$$

Neste trabalho o método utilizado, dado a existência de algumas vantagens, foi a decomposição baseado no valor de Shapley.

3.3. Decomposição baseada em um modelo de regressão

Diferentemente dos métodos tradicionais de decomposição por subgrupo populacional ou por fontes de renda na análise da desigualdade de renda, a abordagem baseada em regressão tem a vantagem de permitir a inclusão de quaisquer variáveis explicativas, tais como variáveis econômicas, sociais e demográficas (GUNATILAKA; CHOTIKAPANICH, 2009). Conforme Wan e Zhou (2005), todas as abordagens baseadas em regressão se iniciam com uma função de rendimentos (*income generating function*), que pode ser escrita como:

$$\ln(Y_i) = \beta_0 + \sum_j^n \beta_j X_{ji} + \varepsilon_i, \quad (25)$$

em que Y é, nesta pesquisa, o rendimento mensal de todos os trabalhos, X_{ji} são os fatores que determinam a renda e ε_i é um termo de erro aleatório com as propriedades usuais. O uso de uma especificação semi-log é justificado, conforme Wan e Zhou (2005), pelo fato de a variável renda ter distribuição aproximadamente log-normal.

Foram consideradas as seguintes variáveis explicativas:

- a) Variável binária para gênero, que assume valor 1 para homens e 0, caso contrário;
- b) Variável binária para Cor, que assume valor 1 para cor branca e 0, caso contrário;
- c) Idade, medida em dezenas de anos. Foi considerada também idade ao quadrado, visto que o rendimento não varia linearmente com a idade;
- d) Escolaridade, em anos de estudo;
- e) Três variáveis binárias para distinguir o posto de trabalho: dirigentes, profissionais das ciências e das artes, técnicos de nível médio, trabalhadores de serviços administrativos, trabalhadores dos serviços, vendedores e prestadores de serviços do comércio, trabalhadores agrícolas, trabalhadores

da produção de bens e serviços e de reparação e manutenção, membros das forças armadas, e ocupações mal definidas ou não declaradas; e

- f) Quatro variáveis binárias para distinguir o setor em que o indivíduo trabalha: indústria, agricultura, serviço e construção.

É pertinente discutir a respeito das variáveis escolhidas. Praticamente todas foram selecionadas de acordo com a literatura e com as teorias expostas na seção anterior. Assim, as variáveis educação, idade e idade ao quadrado refletem, parcialmente, os efeitos do capital humano sobre a produtividade e, assim, sobre os rendimentos dos trabalhadores.

Os possíveis efeitos da discriminação no mercado de trabalho são capturados pelas variáveis gênero e cor. Os diferenciais nos rendimentos devido a existência de diferenciais compensatórios e salário-eficiência são capturados pelas variáveis *dummies*, que distinguem o posto de trabalho dos trabalhadores. Finalmente, a variável *dummy*, que distingue o setor em que o indivíduo trabalha, foi incluída com base na teoria do mercado de trabalho dual.

De posse da equação de rendimentos, é possível computar a contribuição de cada variável para a desigualdade de renda. O procedimento de decomposição proposto por Shorrocks (1999) segue a mesma lógica do valor de Shapley. Por isso, então, ela é denominada decomposição de Shapley.

Diferentemente dos métodos utilizados na literatura brasileira, que fizeram uso da decomposição baseada em regressão (por exemplo, a decomposição de Fields), a vantagem da decomposição de Shapley é que ela pode ser aplicada a quaisquer tipos de especificação econométrica adotadas para a equação de rendimentos (WAN, 2004). Dado que a forma funcional da equação (25) é do tipo semi-log, a equação de rendimento, em termos da variável original rendimento, é não-linear. Desse modo, não seria possível utilizar (ao menos em termos da variável rendimento), por exemplo, a decomposição de Fields, utilizada por Salardi (2005), mas sim a decomposição de Shapley²⁰. Outra vantagem desse método é que ele pode ser utilizado para decompor qualquer índice de desigualdade (Índice de Gini, Índice de Theil (L e T), Coeficiente de Variação etc.), diferentemente, por exemplo, do método de Fields, que considera somente o Coeficiente de Variação.

²⁰ Salardi (2005) utilizou o logaritmo do rendimento, e não a variável rendimento, para decompor a equação de rendimentos.

O procedimento da decomposição de Shapley pode ser entendido, conforme Wan e Zhou (2005), utilizando a seguinte função de rendimentos:

$$Y = f(X_1, \dots, X_k) \quad (26)$$

Normalmente, o vetor X é diferente para cada indivíduo (observações). Assim, substitui-se X_k por sua média amostral, eliminando qualquer diferença, em termos de X_k , entre os indivíduos. A renda resultante, chamada de Y_k , ainda difere entre os indivíduos, porém não mais em razão de X_k . Em outras palavras, a desigualdade em Y_k , denotada por $I(Y_k)$ e que foi obtida, no caso deste trabalho, por meio do índice de Gini, de Theil-T e Theil-L, é causada por diferenças em X , excluída a variável X_k . Dessa forma, a contribuição da variável X_k na desigualdade total (C_k) pode ser conseguida da seguinte forma: $C_k = I(Y) - I(Y_k)$, para $k = 1, 2, \dots, K$. O próximo passo é substituir, além de X_k , X_j pela sua média amostral. Da mesma maneira que o passo anterior, é necessário obter a renda resultante, Y_{kj} . A contribuição do segundo passo, ou “segundo round”, pode ser obtida por $C_k = I(Y_j) - I(Y_{jk})$ para $k, j = 1, 2, \dots, K$ ($k \neq j$). Por meio do mesmo processo, é obtida a contribuição do “terceiro round”, $C_k = I(Y_{ij}) - I(Y_{ijk})$ para $k, j, i = 1, 2, \dots, K$ ($k \neq j \neq i$). O procedimento continua até que todo o vetor X seja substituído pela média amostral.

Pelo fato de a equação (26) ser do tipo semi-log, é necessário transformá-la, depois de estimada, em um modelo linear para que a decomposição de Shapley seja aplicada em Y e não em $\log(Y)$.

3.3.1. Decomposição de Shapley: exemplo simplificado

Com o objetivo de fornecer uma ideia intuitiva da decomposição de Shapley, é apresentado um exemplo simplificado desse método. Para demonstrar a abordagem desenvolvida por Shorrocks (1999) é necessário, primeiramente, definir a função de rendimentos que, neste exemplo, incorpora somente três variáveis explicativas:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon, \quad (27)$$

em que Y é um vetor de rendas (rendimentos), os X 's são as variáveis independentes e ε é o termo de erro estocástico. Tanto a variável rendimentos como a forma funcional da equação (27) podem assumir qualquer especificação; porém, para

facilitar o entendimento, neste exemplo, usa-se a forma funcional linear. Dessa forma, a partir de (27), tem-se a seguinte função estimada:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3. \quad (28)$$

Para fazer a analogia com o valor de Shapley, considera-se as variáveis inseridas no modelo como jogadores. Portanto, tem-se, neste exemplo, $N = \{1, 2, 3\}$, em que N é o conjunto dos jogadores (variáveis). Assim, o número de ordenações que podem ser feitas a partir do conjunto N é 6, isto é, $N! = 6$. Essas ordenações podem ser observadas na primeira coluna da Tabela 3.

Tabela 4 – Valor de Shapley aplicada à análise de regressão

Ordenações	Contribuição marginal das variáveis		
	X_1	X_2	X_3
(1,2,3)	$\nu(1) - \nu(\emptyset)$	$\nu(1,2) - \nu(1)$	$\nu(1,2,3) - \nu(1,2)$
(1,3,2)	$\nu(1) - \nu(\emptyset)$	$\nu(1,3,2) - \nu(1,3)$	$\nu(1,3) - \nu(1)$
(2,1,3)	$\nu(2,1) - \nu(2)$	$\nu(2) - \nu(\emptyset)$	$\nu(2,1,3) - \nu(2,1)$
(2,3,1)	$\nu(2,3,1) - \nu(2,3)$	$\nu(2) - \nu(\emptyset)$	$\nu(2,3) - \nu(2)$
(3,1,2)	$\nu(3,1) - \nu(3)$	$\nu(3,1,2) - \nu(3,1)$	$\nu(3) - \nu(\emptyset)$
(3,2,1)	$\nu(3,2,1) - \nu(3,2)$	$\nu(3,2) - \nu(3)$	$\nu(3) - \nu(\emptyset)$

Fonte: Elaborado pelo autor.

De posse das ordenações provenientes do conjunto N , é possível obter a contribuição marginal de cada uma das variáveis advindas de cada coalizão (ordenação). Essas contribuições podem ser vistas nas colunas 2, 3 e 4 da Tabela 3.

Para simplificar a exposição, optou-se por analisar somente a variável X_1 . A lógica é a mesma para as demais.

Como o objetivo da decomposição é quantificar a contribuição de cada variável inserida na equação de rendimentos, o valor $\nu(\cdot)$ representa um índice qualquer de desigualdade, que pode ser o índice de Gini, de Atkinson, de Theil etc. Assim, $\nu(1)$ é uma medida de desigualdade obtida a partir de (28), porém com os valores de X_2 e X_3 substituídos por suas médias amostrais, o que, por conseguinte, elimina qualquer desigualdade existente nessas variáveis. Formalmente, por meio de (28), tem-se:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3, \quad (29)$$

em que $\bar{X}_2$ e $\bar{X}_3$ são variáveis formadas por suas respectivas médias amostrais e $\hat{Y}$ é o vetor formado pelos rendimentos previstos.

De posse da variável dependente $\hat{Y}$, equação (29), utiliza-se um índice qualquer (por exemplo, o índice de Gini) para obter o grau desigualdade de rendimentos, ou seja, o índice escolhido é aplicado sobre o vetor $\hat{Y}$. Esse procedimento fornece $\nu(1)$. Em seguida, é necessário possuir $\nu(\emptyset)$, que é a desigualdade de rendimentos obtida a partir da seguinte equação:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3. \quad (30)$$

A equação (30) fornece o rendimento previsto quando todas as variáveis são substituídas por suas respectivas médias amostrais. Assim, elimina-se toda a diferença existente entre os indivíduos da amostra. A partir de (20), então, calcula-se $\nu(\emptyset)$, cujo valor é zero, pois não há diferença entre as observações da amostra.

Considerando a terceira linha da segunda coluna da Tabela 3, observa-se que é necessário obter $\nu(2,1)$ e $\nu(2)$. Esses valores são calculados aplicando o índice de desigualdade sobre a variável dependente das seguintes equações, respectivamente:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 \bar{X}_3, \quad (31)$$

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 \bar{X}_3, \quad (32)$$

em que, como já enfatizado, a barra sobre as variáveis significa que os valores originais foram substituídos pelas respectivas médias amostrais. O procedimento para calcular os demais valores é análogo ao que já foi visto até aqui.

Então, a partir da equação (14), e sabendo que $N! = 6$, é possível mensurar o valor de shapley para o jogador 1 (variável X_1) da seguinte forma:

$$\phi_1 = \frac{1}{6} \sum_{R \in \mathfrak{R}} \Delta_i(S_i(R)), \quad (33)$$

em que $\sum(\cdot)$ é a soma da coluna 2 da Tabela 3 e ϕ_1 é a contribuição da variável X_1 à desigualdade de rendimentos. Aplicando todo o processo já visto, porém considerando as colunas 3 e 4 da Tabela 3, obtém-se, respectivamente, ϕ_2 e ϕ_3 .

Um dos requisitos para que a decomposição de Shapley seja aplicada é, como já foi enfatizado, que haja uma equação de rendimentos. Na maioria das vezes, é necessário estimá-la. Porém, no processo de estimação, deve-se levar em conta a possibilidade de que apareçam alguns problemas econométricos; entre eles, citam-se

o viés de seletividade e a endogeneidade. De forma a melhor esclarecer esses pontos, apresenta-se, a seguir, a discussão e as possíveis soluções para esses possíveis casos.

3.4. Procedimento de Heckman

De acordo com Kassouf (1994), a estimação das equações de rendimentos, por meio do método dos Mínimos Quadrados Ordinários (MQO), pode gerar coeficientes viesados. A explicação para isso está na provável seletividade amostral, uma vez que é comum utilizar, no processo de estimação, somente indivíduos que estão empregados, isto é, que possuem algum rendimento proveniente do trabalho. Aqueles trabalhadores que não recebem renda do trabalho não seriam considerados nos cálculos. Para levar em conta essa possibilidade, Heckman (1979) desenvolveu um método que permite estimativas consistentes.

Formalmente, pode-se demonstrar, segundo Kassouf (1994), o problema do viés de seletividade por meio da seguinte equação:

$$L_i^* = \gamma' Z_i + u_i, \quad (34)$$

em que L_i^* é uma variável (latente) que representa participação no mercado de trabalho; Z_i é um vetor de variáveis que explica a participação ou não do indivíduo na força de trabalho; γ é um vetor de coeficientes a ser estimado; e u_i é o termo de erro aleatório com distribuição normal.

É importante frisar que L_i^* não é observado, mas é possível identificar aqueles indivíduos que trabalham ou não, de tal forma que:

$$\begin{aligned} L_i &= 1 \text{ se } L_i^* > 0 \\ L_i &= 0 \text{ se } L_i^* \leq 0 \end{aligned}$$

Diante disso, pode-se definir Y_i como o logaritmo natural dos rendimentos recebido por cada um dos indivíduos que fazem parte da força de trabalho:

$$Y_i = \beta' X_i + \varepsilon_i, \quad (35)$$

em que X_i é um vetor de variáveis explicativas; β é um vetor de parâmetros a ser estimado; e ε_i é o termo de erro aleatório. Note que Y_i é observado somente quando $L_i^* > 0$. Por pressuposição, assume-se que u_i e ε_i possuem distribuição normal

bivariada com média zero, desvios-padrão σ_u e σ_ε , e correlação ρ . Dado isso e tomando a esperança (condicionada) de ambos os lados de (34), tem-se:

$$\begin{aligned} E(Y_i | Y_i \text{ é observado}) &= E(Y_i | L_i^* > 0) = E(Y_i | u_i > -\gamma'Z_i) \\ &= \beta'X_i + E(\varepsilon_i | u_i > -\gamma'Z_i) \\ &= \beta'X_i + \rho\sigma_\varepsilon\lambda_i(\sigma_u) \end{aligned} \quad (36)$$

em que

$$\lambda_i(\sigma_u) = \frac{\phi\left(\frac{\gamma'Z_i}{\sigma_u}\right)}{\Phi\left(\frac{\gamma'Z_i}{\sigma_u}\right)}$$

e ϕ e Φ são, respectivamente, a função densidade normal e a função distribuição normal. Na literatura, $\lambda_i(\sigma_u)$ é conhecido como a razão inversa de Mills.

Com base na equação (35), é possível observar que a estimação da equação de rendimentos – considerando somente os indivíduos que possuem alguma atividade remunerada – levaria a um viés de especificação (viés de seletividade), visto que a variável *lambda* (razão inversa de Mills) não estaria incluída no modelo. Ao inserir a variável $\lambda_i(\sigma_u)$ na equação (35), as estimativas dos parâmetros tornam-se consistentes, uma vez que a decisão do agente econômico de se inserir ou não na força de trabalho teria sido levada em conta nas estimativas.

A razão inversa de Mills é obtida por meio de um modelo *Probit* que tem como objetivo explicar os fatores que influenciam a decisão do indivíduo em entrar ou não no mercado de trabalho. Após obter a estimativa da variável $\lambda_i(\sigma_u)$, pode-se estimar a equação de rendimentos, em função de X e *lambda*, por meio do MQO (KASSOUF, 1994).

Assim, para este trabalho, a equação de seleção (modelo *Probit*) foi definida da seguinte forma:

$$Z_i = \beta_0 + \beta_1\text{Sexo}_i + \beta_2\text{Cor}_i + \beta_3\text{Id}_i + \beta_4\text{Id}_i^2 + \beta_5\text{Edu}_i + \beta_6\text{Crian}_i + \beta_7\text{Chefe}_i + \varepsilon_i \quad (37)$$

em que:

Z_i = Variável binária que reflete a condição de atividade do i -ésimo indivíduo, isto é, 1 se faz parte da força de trabalho e 0, caso contrário;

Sexo = Variável binária para gênero, que assume valor 1 para homens e 0, caso contrário;

Cor = Variável binária para Cor, que assume valor 1 para cor branca e 0, caso contrário;

Id = Idade, medida em dezenas de anos;

Id^2 = Idade ao quadrado, em dezenas de anos;

Educ = Escolaridade, em anos de estudo;

Crian = Número de crianças na família com idade de 0 a 4 anos; e

Chefe = Variável binária para chefe, que assume valor 1 para indivíduo que é o chefe da família e 0, caso contrário.

Essas variáveis foram escolhidas de acordo com a teoria e a literatura econômica.

Por meio da equação (37), obtém-se a razão inversa de Mills, que é inserida na seguinte equação de rendimentos:

$$Y_i = \beta_0 + \beta_1 Educ_i + \beta_2 Id_i + \beta_3 Id_i^2 + \beta_4 Sexo + \beta_5 Cor_i + \beta_6 Ocup1_i + \beta_7 Ocup2_i + \beta_8 Indústria_i + \beta_9 Serviço_i + \beta_{10} Construção_i + \beta_{11} \lambda_i + \varepsilon_i, \quad (38)$$

em que Y_i é o logaritmo natural do rendimento mensal de todos os trabalhos, $Ocup_j$ ($j=1$ e 2) são variáveis *dummies* que refletem o tipo de ocupação do trabalhador, sendo a categoria-base formada pelos trabalhadores agrícolas e pelos trabalhadores da produção de bens e serviços e de reparação e manutenção; $Ocup_1$, os dirigentes em geral, os profissionais das ciências e das artes e os técnicos, $Ocup_2$, trabalhadores dos serviços, vendedores e prestadores de serviço do comércio, membros das forças armadas e auxiliares e ocupações mal definidas; as variáveis *dummies* *Indústria*, *Serviço* e *construção* indicam o setor de atividade do trabalhador, sendo a categoria-base formada pelo setor agricultura; as demais variáveis já foram especificadas.

3.5. Variáveis explicativas endógenas

Assim como o viés de seletividade, o problema da endogeneidade pode gerar estimativas viesadas e inconsistentes. De acordo com Cameron e Trivedi (2009), a pressuposição fundamental para que os estimadores de MQO sejam consistentes é a

não correlação entre o termo erro e os regressores. Caso isso não ocorra, não é possível fornecer uma interpretação causal no modelo estimado. Formalmente, conforme Wooldridge (2002), tem-se o seguinte modelo populacional:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \quad (39)$$

$$E(u) = 0, \quad Cov(X_j, u) = 0, \quad j = 1, 2, \dots, k-1, \quad (40)$$

em que X_k pode ser correlacionado com o termo erro (u_i); ou seja, os regressores $X_1, X_2, \dots, X_{k-1}$ são considerados exógenos, porém X_k é possivelmente endógeno na equação (39).

A estimação de (39) por MQO produzirá resultados inconsistentes caso a covariância entre a variável X_k e o termo u_i seja diferente de zero. Todavia, existem métodos que conseguem contornar o problema da endogeneidade como, por exemplo, o método das Variáveis Instrumentais (VI) e dos Mínimos Quadrados de Dois Estágios (MQ2E). A diferença entre esses dois métodos está na utilização de uma equação identificada (VI), cujo número de variáveis instrumentais é exatamente igual à quantidade de regressores endógenos, ou uma equação superidentificada (MQ2E), onde há mais instrumentos do que variáveis explicativas endógenas.

A ideia por trás desses métodos é encontrar uma ou mais variáveis que não sejam correlacionadas com o termo erro, mas que tenham forte correlação com a (s) variável(eis) endógena(s). Com relação ao MQ2E, essas condições podem ser expressas da seguinte forma, conforme Wooldridge (2002):

Considere a equação (39), onde X_k é uma potencial variável endógena. Assume-se que há mais que um instrumento para X_k ; isto é, $Z_1, Z_2, \dots, Z_M$. Assim, tem-se:

$$Cov(Z_h, u) = 0, \quad h = 1, 2, \dots, M, \quad (41)$$

de modo que cada um dos regressores (Z) seja exógeno na equação (39). Numa situação onde essas variáveis tenham alguma correlação parcial com X_k , tem-se M diferentes estimadores VI. Além disso, cabe ressaltar que qualquer combinação linear entre $X_1, X_2, \dots, X_{k-1}, Z_1, Z_2, \dots, Z_M$ também é não-correlacionado com o termo estocástico u , tornando-as possíveis variáveis instrumentais. Nesse sentido, é necessário escolher a melhor VI. O MQ2E é um método que fornece suporte a essa escolha.

Para ilustrar o MQ2E, de acordo com Wooldridge (2002), seja Z um vetor formado por todas as variáveis exógenas, isto é, $Z \equiv (X_1, X_2, \dots, X_{k-1}, Z_1, Z_2, \dots, Z_M)$. Ao se levar em conta todas as combinações lineares de Z , o MQ2E escolhe aquela que possui a maior correlação com X_k , ou seja:

$$X_k = \delta_0 + \delta_1 X_1 + \dots + \delta_{k-1} X_{k-1} + \theta_1 Z_1 + \dots + \theta_M Z_M + \varepsilon_k, \quad (42)$$

em que

$$E(\varepsilon_k) = 0, Cov(X_1, \varepsilon_k) = 0, \dots, Cov(X_{k-1}, \varepsilon_k) = 0, Cov(Z_1, \varepsilon_k), \dots, Cov(Z_M, \varepsilon_k).$$

Assim, o melhor estimador VI é obtido pela seguinte combinação linear:

$$X_k^* \equiv \delta_0 + \delta_1 X_1 + \dots + \delta_{k-1} X_{k-1} + \theta_1 Z_1 + \dots + \theta_M Z_M. \quad (43)$$

Segundo Wooldridge (2002), X_k^* pode ser interpretado como a parte de X_k que não possui correlação com o termo erro. Se ε_k for correlacionado com u , conclui-se que X_k é endógeno.

De posse de X_k^* , é possível usá-lo como variável instrumental para X_k na equação (39).

Diante do exposto, pode-se apresentar, para uma amostra, um resumo do método MQ2E da seguinte forma:

- a) Obter os valores de $\hat{X}_k$ por meio da regressão (MQO) de X_k sobre $X_1, X_2, \dots, X_{k-1}, Z_1, Z_2, \dots, Z_M$ ²¹, e
- b) Obter as estimativas da regressão (MQO) de Y sobre $X_1, X_2, \dots, X_{k-1}, \hat{X}_k$.

No caso deste trabalho, o regressor X_k é a educação. Na literatura, é comum encontrar trabalhos que levam em consideração a possibilidade de endogeneidade dessa variável (ANGRIST; KRUEGER, 1991; UEDA; HOFFMAN, 2002; TEIXEIRA, 2006; LEIGH; RYAN, 2008).

Dado que a endogeneidade e o viés de seletividade são problemas que podem causar estimativas viesadas e inconsistentes, Wooldridge (2002) propôs um método que busca corrigir, simultaneamente, esses dois possíveis casos. Esse método é uma extensão do procedimento de Heckman que leva em consideração o potencial endógeno de algumas variáveis, tais como a educação. O modelo pode ser

²¹ $\hat{X}_k$ é o valor estimado de X_k^* .

especificado, segundo Wooldridge (2002), a partir de três equações (populacionais), isto é:

$$y_1 = Z_1\delta_1 + \alpha_1 y_2 + u_1 \quad (44)$$

$$y_2 = Z\delta_2 + v_2 \quad (45)$$

$$y_3 = 1(Z\delta_3 + v_3 > 0) \quad (46)$$

em que δ_1 , δ_2 , δ_3 e α_1 são parâmetros a serem estimados; e u_1 , v_2 , e v_3 são os termos de erro estocástico. Na primeira equação, y_1 é a variável de interesse – que, neste trabalho, representa o rendimento mensal de todos os trabalhos – e y_2 é a potencial variável endógena – educação, neste estudo. A equação (45) seria o primeiro estágio no método dos Mínimos Quadrados de Dois Estágios, que tem como objetivo obter o instrumento a ser inserido na equação de rendimentos. A terceira equação, da mesma forma que no procedimento de Heckman, é a equação de seleção. Essa equação, que normalmente é obtida por meio de um modelo *probit*, busca levar em conta o possível viés de seleção.

Dentre os casos no qual esse modelo se aplica, cita-se aquele em que y_2 é sempre observado, mas é uma variável explicativa endógena na equação (44). Essa situação se ajusta perfeitamente ao que se pretende fazer neste trabalho, visto que y_1 é o rendimento e y_2 é a variável “anos de estudos”.²² Se y_1 e y_2 fosse observado para cada variável do vetor Z , seria necessário estimar somente a equação (44) por MQ2E, considerando y_2 como endógeno.

Segundo Wooldridge (2002), esse método apresenta as seguintes pressuposições:

- a) O vetor Z e a variável y_3 são sempre observáveis, porém y_1 e y_2 são observados somente quando $y_3 = 1$ (o indivíduo participa do mercado de trabalho);
- b) u_1 e v_3 não são correlacionados com Z ;
- c) v_3 possui distribuição normal, com média zero e variância constante;
- d) $E(u_1|v_3) = \gamma_1 v_3$; e

²² Anos de estudos é sempre observado, independente se o indivíduo está inserido no mercado de trabalho ou não.

e) $E(Z' | \nu_2) = 0$.

Assumindo as pressuposições (a)-(e), a equação de rendimentos pode ser estimada, segundo Wooldridge (2002), por meio dos seguintes passos:

- a) Estime δ_3 por meio da regressão (*probit*) de y_3 sobre Z usando a amostra completa. Em seguida, calcule a razão inversa de Mills; e
- b) Utilizando a amostra cujas observações contêm somente indivíduos que trabalham ($y_3 = 1$), estime a seguinte equação:

$$y_{i1} = Z_{i1}\delta_1 + \alpha_1 y_{i2} + \gamma_1 \hat{\lambda}_{i3} + erro_i \quad (47)$$

por MQ2E utilizando os instrumentos ($Z_i, \hat{\lambda}_{i3}$).

Em virtude de se utilizar, como regressor, uma variável estimada, os erros-padrão são inválidos (WOOLDRIDGE, 2002). Dessa forma, optou-se por utilizar as técnicas de *bootstrap* para gerar os erros-padrão.

Dado a falta de boas variáveis que pode ser utilizada como instrumentos na PNAD, utilizaram-se variáveis que não estão nesse banco de dados. Assim, as variáveis que serviram como instrumentos, baseado em Teixeira (2006) e Costa (2008), foram as seguintes: oferta de escolas primárias e a lei 5.692 de 1971, que teve como objetivo ajustar o sistema educacional do Brasil.

O primeiro instrumento (número de escolas) foi inserido no modelo considerando o estado e o ano em que a pessoa nasceu. Essa é uma variável que está correlacionada com a variável educação, ou seja, o aumento do número de escolas está associado a maiores oportunidades de estudos e, por conseguinte, a mais anos de instrução. Além disso, oferta de escolas é, claramente, uma variável exógena na equação de rendimentos.

O segundo instrumento foi utilizado para refletir a mudança ocorrida após a lei nº. 5.692 de 1971. De acordo com Machado e Gonzaga (2007), o sistema educacional do Brasil era dividido, por volta da década de 1960, em duas fases: o ensino primário e o médio. O primeiro durava quatro anos, com início aos sete anos de idade. O segundo tinha como objetivo a formação dos adolescentes e também era dividido em duas fases: o ginásio, com duração de quatro anos; e o colegial, que tinha pelo menos três séries. Em 1971, a lei nº. 5.692 incorporou os quatro anos do ensino ginásial à escola primária, o que aumentou de quatro para oito anos o tempo mínimo de permanência na escola. Então, crianças entre 7 e 14 anos eram obrigadas

a frequentar a escola. Para refletir essa alteração no sistema educacional foi usada uma variável *dummy* que assume o valor 1 se o indivíduo nasceu em 1957 ou após isso; e 0, caso contrário.

Diante disso, a equação de seleção (modelo *probit*) foi estimada da seguinte forma²³:

$$Z_i = \beta_0 + \beta_1 \text{Sexo}_i + \beta_2 \text{Gen}_i + \beta_3 \text{Id}_i + \beta_4 \text{Id}_i^2 + \beta_4 \text{esc} + \beta_5 \text{lei} + \varepsilon_i, \quad (48)$$

em que *esc* é o número de escola no estado e no mês de nascimento da pessoa e *lei* é a variável *dummy* que reflete a lei nº. 5.692 de 1971. As outras variáveis já foram definidas.

A partir da equação (48), encontra-se a razão inversa de Mills e a incorpora na equação de rendimentos, que já foi especificada anteriormente.

3.6. Medidas de desigualdade de renda

Na literatura, há diversas maneiras de mensurar o grau de desigualdade, sendo que todas elas possuem algum recurso matemático ou intuitivo (LITCHFIELD, 1999). Entre as principais medidas que condensa toda a informação da distribuição de renda em um único valor, pode-se citar os mais utilizados: coeficiente de Gini, Theil-L e Theil-T. Os dois últimos fazem parte da classe de entropia generalizada.

Antes de fazer uma análise mais detalhada dos índices que foram utilizados nesta pesquisa, é importante apresentar as propriedades (axiomas) que todas as medidas de desigualdade deveriam satisfazer. Conforme Letchfield (1999), existem cinco axiomas, a saber: princípio da transferência de Pigou-Dalton, invariância de escala, princípio da população, anonimato e decomponibilidade.

O primeiro axioma requer que a medida de desigualdade aumente (ou ao menos não reduza) em resposta a uma transferência regressiva de renda – isto é, de uma pessoa mais pobre para uma mais rica – mantendo, porém, a renda média inalterada. Por outro lado, havendo transferência de uma pessoa no topo da distribuição para uma que está na base, o índice deveria registrar uma queda (ou ao menos não aumentar). Matematicamente, considere o vetor Y de rendas, $y_1, y_2, \dots, y_n$; em que n é o número de indivíduos na população. Seja $F(Y)$ a função de distribuição acumulada e $I(Y)$ a desigualdade em Y , mensurada por um

²³ É importante frisar que a variável educação não deve ser inserida como regressor no modelo *Probit*, visto que ela é a variável a ser instrumentalizada.

índice qualquer. Agora, de acordo com Letchfield (1999), suponha que há uma transferência de renda, no montante de ψ , do indivíduo y_i para o indivíduo y_j , obtendo, assim, o vetor Y' . Considere $y_j > y_i$ e $y_j + \psi > y_i - \psi$. Dessa forma, o princípio da transferência é satisfeito se $I(Y') \geq I(Y)$.

O princípio da invariância de escala requer que a medida de desigualdade sejam invariantes às mudanças proporcionais uniformes. Ou seja, numa situação onde a renda de cada indivíduo da população mude na mesma proporção, a desigualdade deveria manter-se constante. Matematicamente, tem-se que para qualquer escalar $\lambda > 0$, $I(Y) = I(\lambda Y)$.

O princípio da população exige que as medidas de desigualdade seja invariante à replicações da população. Em outras palavras, a junção de duas distribuições idênticas não deveria alterar a desigualdade. Então, para qualquer escalar $\lambda > 0$, $I(Y) = I(Y(\lambda))$, em que $Y(\lambda)$ é uma concatenação (junção) do vetor Y , λ vezes.

O princípio do anonimato ou da simetria requer que a medida de desigualdade leve em consideração somente a renda (ou outro indicador de bem-estar), desconsiderando qualquer outra característica dos indivíduos. Dessa forma, para qualquer permutação Y' de Y , tem-se $I(Y) = I(Y')$.

Por fim, o princípio da decomponibilidade exige que a medida de desigualdade global seja consistentemente relacionada com as partes que formam a distribuição, tais como os subgrupos populacionais. Por exemplo, caso a desigualdade entre subgrupos da população cresça, então, é de se esperar que a desigualdade global (total) também cresça.

A classe de entropia generalizada (especificamente o coeficiente de Theil-T e Theil-L) foi escolhida para ser utilizada na decomposição de Shapley, porque, conforme Ferreira e Litchfield (2000), seus membros satisfazem todos os axiomas desejáveis. Entretanto, utilizou-se também o coeficiente de Gini, visto que é o índice mais utilizado na literatura sobre desigualdade, o que permite, por conseguinte, maior grau de comparabilidade com outros estudos.

O coeficiente de Gini foi desenvolvido por Corrado Gini em 1914 e é calculado a partir da curva de Lorenz (PINTO, 2007). Segundo Medeiros (2006), a curva de Lorenz foi criada pelo economista Max Otto Lorenz, em 1905. Ela mostra graficamente como cada fração da população apropria-se da renda total. A sua

representação gráfica é bem simples, onde um eixo (horizontal) representa as frações acumuladas da população e, o outro eixo (vertical), as frações acumuladas da renda total (Figura 7).

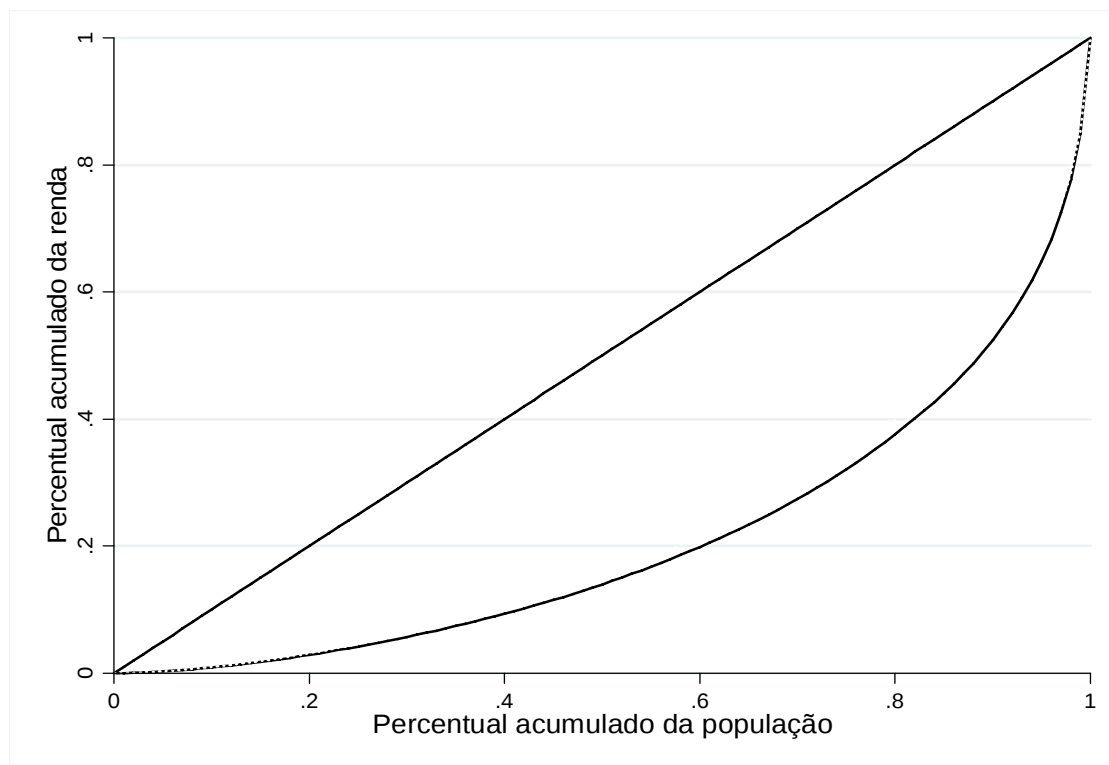


Figura 7 – Curva de Lorenz para uma distribuição contínua
Fonte: Adaptado de Medeiros (2006).

Numa situação onde há perfeita distribuição de renda, a curva de Lorenz seria representada pela linha reta de 45 graus (Figura 7). Por outro lado, num ambiente em que existe desigualdade de rendimentos, ela seria representada por uma curva (convexa) abaixo da linha de perfeita distribuição. Quanto mais distante a curva de Lorenz estiver da reta de 45 graus – isto é, quanto mais acentuado for o arco da curva – pior é a distribuição de renda.

Matematicamente, a curva de Lorenz pode ser definida, de forma contínua, pela seguinte equação (MEDEIROS, 2006):

$$p = F(x) \Rightarrow L(p) = \frac{1}{\mu} \int_0^x Xf(X)dX \text{ para } 0 \leq p \leq 1 \quad (49)$$

em que $L(P)$ é a curva de Lorenz; p são as frações acumuladas da população que são definidas pela função de distribuição acumulada $F(X)$; X representa a renda da distribuição; e μ é a renda média.

A curva de Lorenz está estreitamente relacionada com o coeficiente de Gini. É a partir do primeiro que se calcula o segundo. Então, o índice de Gini (G) pode ser definido como a razão entre a área situada entre a reta de perfeita distribuição (linha de 45 graus) e a curva de Lorenz. Denotando essa área por π , tem-se:

$$G = \frac{\pi}{0,5} = 2\pi \quad (50)$$

A área representada por π pode variar entre 0 e 0,5. Dessa forma, o coeficiente de Gini assume valores compreendidos entre 0 (perfeita distribuição de renda) e 1 (renda totalmente concentrada).

Os outros índices utilizados nesta pesquisa referem-se aos da classe de entropia generalizada, que pode ser definida, de acordo com Ferreira e Litchfield (2000), como:

$$EG(\alpha) = \frac{1}{(\alpha^2 - \alpha)} \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{y_i}{\bar{y}} \right)^\alpha - 1 \right] \quad (51)$$

em que n é número de pessoas na amostra; y_i é a renda domiciliar *per capita* para o indivíduo i ; e $\bar{y}$ é a renda média (aritmética). O parâmetro α reflete o peso dado às distâncias entre rendas de diferentes partes da distribuição, podendo assumir qualquer valor real. Quanto menor for o valor de α , mais sensível a mudanças na cauda inferior da distribuição torna-se o índice. Por outro lado, à medida que o valor de α cresce, o índice EG torna-se mais sensível a mudanças na cauda superior da distribuição. Assumindo parâmetros iguais a 0 e 1, tem-se os índices Theil-L e Theil-T, respectivamente:

$$EG(0) = \frac{1}{n} \sum_{i=1}^n \log \frac{\bar{y}}{y_i} \quad (52)$$

$$EG(1) = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{\bar{y}} \log \frac{y_i}{\bar{y}} \quad (53)$$

É importante frisar que o coeficiente de Theil-L só é aplicado a indivíduos com rendimento domiciliar *per capita* diferente de zero, visto que, por motivos matemáticos, a equação (52) não permite $y_i = 0$.

3.6.1. Medidas de pobreza

Como foi enfatizado, para aplicar a decomposição de Shapley, é necessário, primeiramente, definir as medidas de pobreza. As medidas que foram utilizadas neste trabalho pertencem à classe P_α , especificamente P_0 , P_1 e P_2 , que representam, conforme Ferreira e Litchfield (2000), a proporção de pobres; o déficit de pobreza, diferença entre a linha de pobreza e os rendimentos dos pobres; e a desigualdade entre os pobres, respectivamente. A classe P_α foi desenvolvida por Foster *et al.* (1984) e pode ser sintetizada da seguinte maneira:

$$P_\alpha = \frac{1}{n} \sum_{i=1}^q \left(\frac{z - y_i}{z} \right)^\alpha, \alpha = 0,1,2, \quad (54)$$

em que z é linha de pobreza; y_i é a renda domiciliar *per capita* do i -ésimo domicílio; e q é o número de domicílios.

Por meio da equação (54), nota-se que a linha de pobreza é de fundamental importância para obter os índices. No Brasil, não há uma linha de pobreza oficial e sua escolha é feita de forma arbitrária. Assim, optou-se por utilizar, conforme Hoffmann (2000) e Helfand *et al.* (2009), as linhas de pobreza e extrema pobreza baseadas no salário mínimo. A linha de pobreza absoluta foi fixada em meio salário mínimo (SM) de 2003, enquanto a de extrema pobreza em um quarto do SM.

3.7. Fonte de dados e tratamento dos dados

A maior parte dos dados que foi usado neste trabalho é oriunda da Pesquisa Nacional por Amostra de Domicílios (PNAD), referente aos anos de 2003 e 2008. Optou-se por iniciar a análise a partir de 2003 porque foi o ano em que a pobreza começou, ao menos no período recente, a declinar de forma mais nítida (ver Figura 3).

O foco do trabalho será a Região Nordeste e a maioria das variáveis, já especificadas anteriormente, estão contidas na PNAD. A linha de pobreza, o rendimento mensal de todos os trabalhos e a renda domiciliar *per capita* foram corrigidos para refletir os valores de setembro de 2008, utilizando o INPC do IBGE.

É importante fazer uma ressalva a respeito da PNAD. Ela representa uma amostra complexa, ou seja, não é baseada numa amostra aleatória simples. Diante

disso, foi necessário levar em conta os efeitos do plano amostral para que os resultados das estimativas de significância não sejam calculados de forma incorreta²⁴.

A variável oferta de escolas, que é um dos instrumentos usados na metodologia desenvolvida por Wooldridge (2002), foi obtida no site do IBGE (www.ibge.gov.br), especificamente no item estatísticas do século XX.

Optou-se por utilizar rendimento mensal e renda domiciliar *per capita* como *proxies* de bem-estar. Essas *proxies*, no entanto, envolvem debates na literatura empírica. Uma parte dos diferenciais de rendimento (renda) pode ter caráter transitório, ou seja, algumas atividades estão sujeitas a variações cíclicas e/ou sazonais. Essa questão transitória pode ser tratada utilizando o consumo como *proxy* de bem-estar, já que seu comportamento é mais suave (sofre menos flutuações transitórias) ao longo da vida de um indivíduo. Porém, pelo fato de haver limitação em relação à disponibilidade de dados referentes a consumo, utilizar rendimento e/ou renda domiciliar é uma prática mais usual na literatura brasileira.

O *software* utilizado para realizar a maior parte das estimativas é o Stata 10.1. As medidas de pobreza da classe P_α e a decomposição da variação da pobreza foram calculadas usando o *Distributive Analyse Stata Package* (DASP), em sua versão 2.1, desenvolvido por Araar e Duclos (2009).

Como foi possível notar, a decomposição baseada numa regressão (Shapley) envolve, dependendo do número de variáveis que são inseridas na equação de rendimentos, uma grande quantidade de cálculos, o que torna extremamente trabalhoso caso seja feito manualmente. Dessa forma, a decomposição foi implementada por um programa Java desenvolvido pelo *World Institute for Development Economics research of the United Nations University* (UNU-WIDER).

As rotinas usadas para retirar os dados da PNAD, considerando o seu plano amostral, podem ser observadas no Anexo C.

²⁴ Mais detalhes sobre o plano amostral da PNAD podem ser encontrados no Anexo A.

4. RESULTADOS E DISCUSSÃO

Para facilitar o entendimento da análise, os resultados foram divididos em duas grandes seções. Na primeira são apresentadas, para cada estado do Nordeste, as estimativas referentes às medidas de pobreza, considerando os índices da classe FGT (α), e a decomposição das variações na pobreza em dois componentes: crescimento e desigualdade de renda. Na segunda seção são expostos e analisados os determinantes dos rendimentos dos trabalhadores e os principais fatores que contribuem para a desigualdade de renda nos estados do Nordeste.

4.1. Medidas de pobreza e decomposição

As estimativas de pobreza absoluta dependem da escolha de uma linha que separa os pobres dos não pobres. Como já foi enfatizado, não há, no Brasil, uma linha de pobreza (LP) oficial²⁵. Assim, utilizou-se meio salário mínimo (SM) de setembro de 2003 para definir a linha de pobreza, e um quarto do SM para a extrema pobreza. Essas linhas foram corrigidas para refletir os valores de setembro de 2008, utilizando o INPC do IBGE²⁶. O salário mínimo em 2003 era de R\$240, o que gerou uma linha de pobreza e de extrema pobreza de R\$120 e R\$60, respectivamente. Esses valores, em 2008, corresponderam, respectivamente, a R\$154,30 e R\$77,15.

²⁵ O governo brasileiro criou uma linha de extrema pobreza no valor de R\$ 70 com intuito de definir um critério para incluir ou não as pessoas nos programas de combate à extrema pobreza.

²⁶ A justificativa por utilizar o salário mínimo de 2003, e não de 2008, para gerar as linhas de pobreza é devido ao fato de que qualquer aumento real do SM contribui para a queda na pobreza. Se fosse utilizado o SM de 2008, a linha de pobreza seria superestimada, o que subestimaria o feito do crescimento real do SM sobre a redução na pobreza e extrema pobreza.

Com base nesses valores, foram calculados os índices de pobreza da classe FGT para os anos de 2003 e 2008, que podem ser observados na Tabela 5.

Tabela 5 – Medidas de pobreza da classe FGT (α), 2003 e 2008

Estados	2003			2008		
	P_0	P_1	P_2	P_0	P_1	P_2
Maranhão	0,637 (0,027)	0,316 (0,023)	0,199 (0,019)	0,430 (0,033)	0,186 (0,023)	0,114 (0,018)
Piauí	0,597 (0,035)	0,314 (0,023)	0,210 (0,017)	0,402 (0,021)	0,177 (0,012)	0,106 (0,009)
Ceará	0,565 (0,013)	0,269 (0,010)	0,167 (0,008)	0,370 (0,017)	0,148 (0,010)	0,081 (0,006)
Rio G. do Norte	0,525 (0,020)	0,234 (0,012)	0,140 (0,009)	0,317 (0,017)	0,131 (0,006)	0,078 (0,004)
Paraíba	0,547 (0,031)	0,249 (0,017)	0,148 (0,012)	0,373 (0,022)	0,153 (0,010)	0,087 (0,006)
Pernambuco	0,572 (0,011)	0,276 (0,009)	0,175 (0,007)	0,395 (0,010)	0,166 (0,007)	0,097 (0,006)
Alagoas	0,648 (0,024)	0,320 (0,020)	0,197 (0,017)	0,472 (0,033)	0,213 (0,019)	0,128 (0,013)
Sergipe	0,494 (0,025)	0,223 (0,019)	0,135 (0,016)	0,337 (0,013)	0,135 (0,006)	0,076 (0,003)
Bahia	0,575 (0,012)	0,272 (0,010)	0,168 (0,008)	0,372 (0,011)	0,157 (0,007)	0,093 (0,005)
Nordeste	0,578 (0,007)	0,277 (0,005)	0,173 (0,004)	0,385 (0,007)	0,162 (0,004)	0,095 (0,003)

Fonte: Resultados da pesquisa.

Conforme a Tabela 5, cujos erros-padrão estão entre parêntese, nota-se o grande problema que é a pobreza na Região Nordeste. Considerando a proporção de pobres (P_0), observa-se que, em 2003, mais da metade da população nos estados do Nordeste, exceto em Sergipe, viviam abaixo da LP. Especificamente nos Estados do Maranhão e Alagoas, mais de 60% de seus habitantes eram pobres, isto é, a renda domiciliar *per capita* era inferior a R\$ 154. Por outro lado, Sergipe possuía a menor proporção de pobres da região, embora em patamares bem elevados.

Ainda que o índice P_0 seja uma boa medida para refletir a situação da pobreza, ele não leva em conta a distância dos rendimentos dos pobres em relação à LP, ou seja, P_0 não tem a capacidade de captar a profundidade (intensidade) da

pobreza. A medida P_1 (déficit de pobreza), por outro lado, fornece essa informação. De acordo com esse índice, observa-se que, em 2003, os rendimentos dos indivíduos estavam, na média, bem abaixo da LP, destacando os Estados do Maranhão, Piauí e Alagoas. Esses três estados possuíam um índice que estava bem acima do que foi verificado na região como um todo (0,277).

O índice P_2 , diferentemente de P_0 e P_1 , incorpora a desigualdade de rendimentos existentes entre os pobres, isto é, ele é capaz de mostrar quão disperso são os indivíduos abaixo da linha de pobreza. Assim, como nas outras medidas, os Estados do Maranhão, Piauí e Alagoas apresentaram, em 2003, os piores indicadores, a saber: 0,199, 0,210 e 0,197, respectivamente. Além de possuírem as maiores quantidades de pobres e os mais elevados déficits de pobreza, esses estados também exibiam uma grande dispersão entre os indivíduos pobres. Isto é, a desigualdade de renda entre as pessoas abaixo da LP era extremamente grande.

Com relação ao ano de 2008, verificou-se uma melhora significativa. Houve queda na pobreza de acordo com todas as medidas. A proporção de pobres diminuiu, em média, 19 pontos percentuais (p.p.), o que mostra que um número expressivo de pessoas ultrapassou a linha de pobreza. Rio Grande do Norte apresentou o melhor desempenho nesse quesito, o índice P_0 caiu 40%. O déficit de pobreza também melhorou de forma expressiva. Seu valor caiu, em média, 41%, indicando que aqueles indivíduos que permaneceram abaixo da LP ficaram menos pobres. O Estado do Ceará teve a maior queda, passando de 0,269 para 0,148. Por fim, a queda de 45% no índice P_2 sugere que os rendimentos entre as pessoas mais pobres da Região Nordeste ficaram menos desiguais. O Estado do Ceará obteve, assim como no déficit de pobreza, o melhor resultado em termos do índice P_2 , apresentando uma queda de 51%.

Já os resultados referentes às medidas de extrema pobreza – pessoas que possuem renda domiciliar *per capita* menor que R\$77,15 – podem ser vistos na Tabela 6.

Tabela 6– Medidas de extrema pobreza da classe FGT (α), 2003 e 2008

Estados	2003			2008		
	P_0	P_1	P_2	P_0	P_1	P_2
Maranhão	0,309 (0,032)	0,134 (0,018)	0,080 (0,013)	0,166 (0,031)	0,078 (0,016)	0,050 (0,010)
Piauí	0,329 (0,027)	0,156 (0,014)	0,095 (0,009)	0,175 (0,017)	0,071 (0,008)	0,039 (0,006)
Ceará	0,254 (0,011)	0,111 (0,008)	0,067 (0,006)	0,130 (0,013)	0,045 (0,005)	0,024 (0,003)
Rio G. do Norte	0,217 (0,014)	0,089 (0,008)	0,050 (0,007)	0,117 (0,009)	0,052 (0,003)	0,033 (0,003)
Paraíba	0,233 (0,019)	0,093 (0,010)	0,052 (0,007)	0,138 (0,011)	0,053 (0,004)	0,031 (0,003)
Pernambuco	0,263 (0,012)	0,119 (0,007)	0,075 (0,005)	0,148 (0,010)	0,062 (0,005)	0,039 (0,003)
Alagoas	0,323 (0,029)	0,127 (0,017)	0,069 (0,012)	0,209 (0,024)	0,083 (0,009)	0,050 (0,005)
Sergipe	0,220 (0,027)	0,088 (0,015)	0,049 (0,010)	0,112 (0,006)	0,045 (0,003)	0,027 (0,002)
Bahia	0,263 (0,013)	0,110 (0,007)	0,064 (0,005)	0,142 (0,008)	0,059 (0,005)	0,036 (0,004)
Nordeste	0,268 (0,007)	0,115 (0,004)	0,068 (0,003)	0,147 (0,005)	0,060 (0,003)	0,036 (0,002)

Fonte: Resultados da pesquisa.

Por extrema pobreza (*food poverty*), entende-se a insuficiência de rendimentos para atender as necessidades mínimas de alimentação (ROCHA, 1996). Nesse sentido, observa-se, na Tabela 6, que do total de pobres existentes em 2003, na Região Nordeste, quase a metade (46,4%) deles não tinha renda suficiente para adquirir uma cesta alimentar que pudesse suprir suas necessidades calóricas²⁷. Esse resultado é ainda pior nos Estados do Maranhão, Piauí e Alagoas, cujo índice P_0 é superior a 0,30, revelando que, nesses estados, mais de 30% da população, em 2003, vivia abaixo da linha de extrema pobreza. Embora esses números sejam elevados, eles estão numa trajetória descendente. Considerando a Região Nordeste, verifica-se que a proporção de pessoas abaixo da linha de indigência caiu 45%, o que mostra que uma grande parte da população nordestina passou a ter renda domiciliar *per capita* acima de R\$77,15 (valores de setembro de 2008). A mesma tendência de

²⁷ É importante frisar que a PNAD não registra a produção de autoconsumo. Esse fato pode levar a uma superestimação da pobreza e da extrema pobreza.

queda é verificada nos outros dois índices. Em média, a redução de P_1 e P_2 foi de 48% e 46%, respectivamente, revelando que as pessoas indigentes ficaram mais próximas da linha de extrema pobreza e que a desigualdade de renda entre elas tornou-se menos acentuada.

Por meio dos dados já apresentados é possível observar quanto cada estado contribui para a pobreza da região (Tabela 7).

Tabela 7 – Contribuição relativa dos estados do Nordeste para a pobreza (P_0) regional, 2003-2008

Estados	2003		2008	
	Parcela da população	Contribuição relativa	Parcela da população	Contribuição relativa
Maranhão	0,12	13,24	0,12	13,50
Piauí	0,06	6,13	0,06	6,18
Ceará	0,16	15,41	0,16	15,15
Rio G. do Norte	0,06	5,45	0,06	4,88
Paraíba	0,07	6,84	0,07	6,91
Pernambuco	0,16	16,07	0,16	16,70
Alagoas	0,06	6,76	0,06	7,36
Sergipe	0,04	3,27	0,04	3,34
Bahia	0,27	26,84	0,27	25,96
Nordeste	1,00	100,00	1,00	100,00

Fonte: Resultados da pesquisa.

Embora Alagoas tenha apresentado o maior índice P_0 , esse estado foi responsável, em 2003, por somente 6,76% da pobreza absoluta da Região Nordeste. Esse fato está relacionado à pequena participação, em termos populacionais, de Alagoas na região como um todo. Em outras palavras, a proporção de pobres do estado, comparado com Nordeste, é pequena. Mesmo mantendo a parcela de sua população constante (6%), em 2008, a contribuição relativa de Alagoas para a pobreza passou para 7,36%. A justificativa para isso está na queda mais acentuada do

número de pobres dos outros estados, ou seja, como se trata de uma contribuição relativa, quando a proporção de pobres se reduz de forma mais expressiva em uma área, a outra que apresentar redução menor tende a contribuir mais para a pobreza.

Pernambuco foi o segundo estado em termos de contribuição para a pobreza da região. Em 2003, esse estado possuía mais de 16% do número de pobres do Nordeste. A situação não se alterou em 2008, cuja proporção permaneceu na casa dos 16%.

Ainda que não tenha apresentado o maior índice de proporção de pobres, Bahia foi o estado que mais contribuiu para a pobreza na região. Do total de pessoas abaixo da linha de pobreza no Nordeste, 26,84% viviam na Bahia em 2003. Como já enfatizado, essa parcela significativa está relacionada ao tamanho da população existente no estado. Todavia, em 2008, a fatia da pobreza que corresponde à Bahia caiu para 25,96%, reflexo da redução significativa do número de pobres (35,3%).

Como foi possível notar, Maranhão, Piauí e Alagoas apresentaram os piores indicadores do Nordeste. Esse resultado pode estar associado à baixa renda *per capita*. Na Região, esses estados possuem o menor rendimento domiciliar *per capita*. Todavia, nos últimos anos observou-se forte crescimento da renda, o que influenciou diretamente na queda da pobreza e da extrema pobreza. O aumento nos rendimentos da população nordestina pode ser observada na Tabela 8, cujos erros-padrão estão entre parênteses.

Tabela 8 – Rendimento domiciliar per capita médio (por mês), em valores de setembro de 2008, dos estados da Região Nordeste, 2003 e 2008

Estados	2003	2008	Variação percentual
Maranhão	215,44 (24,65)	291,72 (24,58)	35,41
Piauí	234,25 (29,42)	360,41 (41,48)	53,86
Ceará	244,14 (9,37)	352,93 (14,77)	44,56
Rio G. do Norte	273,80 (23,27)	419,70 (31,08)	53,29
Paraíba	265,02 (34,74)	398,95 (50,15)	50,53
Pernambuco	259,14 (10,52)	364,53 (14,98)	40,67
Alagoas	231,20 (26,59)	318,28 (33,87)	37,66
Sergipe	307,20 (28,58)	390,68 (27,85)	27,17
Bahia	258,87 (11,28)	374,95 (14,14)	44,84
Nordeste	251,44 (6,20)	360,40 (7,87)	43,33

Fonte: Resultados da pesquisa.

Por meio da Tabela 8, observa-se que entre os anos de 2003 e 2008 a renda domiciliar *per capita* (rdpc), em todos os estados do Nordeste, cresceu de forma expressiva. Analisando a região de forma agregada, nota-se que a rdpc elevou-se de R\$251,44 para R\$360,40, o que ocasionou crescimento de mais de 43%. Piauí, Rio Grande do Norte e Paraíba foram os que tiveram o melhor desempenho. A renda nesses estados aumentou mais de 50%. Por outro lado, Sergipe apresentou o menor crescimento. Sua renda passou de R\$306 para R\$390, uma elevação de 27,58%. Esse resultado pode ter influenciado o fato de Sergipe ter apresentado a menor queda, entre os estados da Região Nordeste, em seu índice de pobreza P_0 . Em outras palavras, a queda do número de pobres, que foi inferior ao observado nos outros Estados, pode ter certa relação com o crescimento menor, comparado com as demais

unidades federativas do Nordeste, em sua renda. Com relação ao Rio Grande do Norte, a rdpc foi a que apresentou maior variação (53%). Ao mesmo tempo, esse estado exibiu a maior queda na proporção de pobres entre 2003 e 2008 (20,8 p.p.).

Vale frisar que o aumento na renda domiciliar *per capita* não necessariamente reduz a pobreza. É importante saber quais são as pessoas que estão sendo beneficiadas por tal crescimento. No intuito de analisar quais são os grupos que obtiveram os melhores ganhos, calculou-se a taxa de crescimento da renda média por décimo da distribuição (Tabela 9).

Tabela 9 – Taxa de crescimento da renda média por décimo da distribuição (%), 2003-2008

Décimos	MA	PI	CE	RN	PB	PE	AL	SE	BA	NE
1	56,73	85,35	22,57	48,45	37,79	59,74	39,20	56,95	52,11	57,62
2	70,48	80,87	18,73	62,69	49,41	52,39	38,82	59,60	57,15	57,54
3	66,32	85,07	31,84	56,34	46,63	51,12	47,04	48,94	58,79	57,28
4	64,09	72,05	34,60	62,11	46,21	48,60	55,00	42,41	60,15	56,58
5	63,86	61,49	33,58	56,38	53,31	43,72	54,47	40,01	58,30	55,31
6	60,60	58,14	34,52	61,67	46,82	43,11	60,13	42,67	57,15	53,70
7	52,12	57,20	33,34	66,90	39,10	40,46	56,48	35,85	54,51	49,70
8	46,69	53,03	31,82	56,15	36,99	36,69	55,65	31,27	49,26	44,52
9	37,21	50,75	27,08	44,96	32,99	32,86	47,53	22,65	46,61	40,09
10	21,78	49,60	28,46	50,86	63,32	34,64	28,70	19,18	34,94	36,36

Fonte: Resultados da pesquisa.

De acordo com a Tabela 9, em todos os décimos da distribuição houve crescimento significativo da renda média no período 2003 a 2008, embora nem todos tenham se beneficiado de forma homogênea. Na maioria dos casos, a parcela da população que forma a base da distribuição teve o melhor desempenho. Considerando a região como um todo, a renda média dos trinta por cento mais pobres cresceu mais de 57% no período; fator esse que contribuiu, de forma expressiva, para a queda do número de pobres. Quando se compara esses resultados com o encontrado em Salardi (2008), verifica-se que, no período de 2002 a 2003, também houve crescimento, em todos os décimos, da renda média.

Ainda que o topo da distribuição tenha, na maioria dos estados, um resultado inferior à base, o crescimento da renda nesse segmento foi extremamente elevado. Como exemplo, pode-se citar Rio Grande do Norte e Paraíba. Especificamente nesse último estado, nota-se que, diferentemente do verificado em outras unidades

federativas do Nordeste, os grupos mais pobres não apresentaram resultados superiores aos observados nas classes mais altas. Como exemplo, pode-se citar o aumento na renda dos 10% mais ricos da Paraíba. Entre 2003 e 2008, os rendimentos desse grupo cresceram mais de 63%; resultado esse que pode estar relacionado ao aumento observado na iniquidade distributiva do estado. O coeficiente de Gini, no período analisado, cresceu 3,8%. Paraíba foi o único estado que teve sua desigualdade de rendimentos afetada negativamente, isto é, aumentou. A partir desse fato, pode-se concluir que a queda do número de pobres foi inteiramente explicada pelo crescimento da renda, a desigualdade não contribuiu com nada para essa redução; o que mostra que a pobreza poderia ter caído ainda mais caso o crescimento nos rendimentos tivesse sido acompanhada pela redução na desigualdade.

Pelo que já foi discutido, a queda na pobreza pode ser explicada tanto pelo crescimento da renda como por sua redistribuição. De acordo com o que foi apresentado até o momento, o Nordeste teve ganhos expressivos em termos de rendimentos; porém, é necessário considerar os benefícios advindos da redistribuição de renda. A questão da desigualdade, então, pode ser melhor observada quando se considera a parcela da renda total apropriada pelos décimos da distribuição (Tabela 10) e não a variação absoluta dos rendimentos (Tabela 9).

Conforme a Tabela 10, verifica-se que os 10% mais ricos, em todas as unidades federativas do Nordeste, continuam com a maior fatia da renda total, detendo mais de 40% da renda gerada. Por outro lado, o primeiro décimo não possui, na maioria dos estados, nem 1% dessa renda, revelando o grande problema distributivo na região. Embora a desigualdade de rendimentos ainda seja alta, o cenário melhorou de 2003 para 2008. No Nordeste como um todo, o topo da distribuição perdeu participação na renda total, passando de 47,5% para 45,4% (uma queda de 4,4%), enquanto o primeiro décimo aumentou sua fração, passando de 0,009 para 0,01 (crescimento de 11%). É bem provável que foi isso que levou o índice de Gini a cair 4% na Região. Em termos estaduais, os ganhos e as perdas foram mais desiguais. No Ceará, os mais pobres (primeiro décimo) foram os que mais ganharam em termos de participação na renda gerada, obtendo um acréscimo de 41%. Por outro lado, o estado onde o último décimo da distribuição perdeu mais foi Maranhão, cuja fatia da renda foi reduzida para 41,5% – uma redução de 11,5%. Esse fato impactou diretamente no coeficiente de Gini, que reduziu 9% nesse estado.

Nos Estados do Rio Grande do Norte, Paraíba e Alagoas o crescimento da renda média do grupo de pessoas inseridas no primeiro décimo pode ser explicado inteiramente pelo crescimento econômico, visto que a fatia do bolo apropriada por esse grupo caiu no Rio Grande do Norte e Alagoas e permaneceu constante na Paraíba. Nesse último estado, somente os décimos 5 e 10 aumentaram suas frações. Ainda que o fator redistribuição não tenha proporcionado crescimento da renda média (primeiro décimo) nos estados supracitados, ele contribuiu, no caso de AL e RN, para a queda na pobreza, já que existem décimos, que se beneficiaram com a queda na desigualdade (aumento da participação na renda total), que estão abaixo da linha de pobreza.

Tabela 10 – Parcela dos rendimentos apropriados pelos décimos da distribuição, 2003 e 2008

Décimos	MA		PI		CE		RN		PB		PE		AL		SE		BA		NE	
	2003	2008	2003	2008	2003	2008	2003	2008	2003	2008	2003	2008	2003	2008	2003	2008	2003	2008	2003	2008
1	0,009	0,009	0,007	0,009	0,009	0,012	0,010	0,009	0,010	0,010	0,007	0,009	0,010	0,009	0,009	0,011	0,009	0,009	0,009	0,010
2	0,021	0,024	0,016	0,019	0,021	0,024	0,021	0,022	0,021	0,021	0,019	0,022	0,020	0,020	0,019	0,023	0,020	0,021	0,020	0,022
3	0,030	0,035	0,025	0,029	0,030	0,032	0,030	0,031	0,030	0,029	0,028	0,030	0,027	0,028	0,027	0,032	0,028	0,030	0,028	0,031
4	0,038	0,045	0,034	0,038	0,039	0,042	0,038	0,040	0,038	0,037	0,036	0,038	0,034	0,037	0,036	0,041	0,036	0,039	0,037	0,040
5	0,047	0,056	0,046	0,048	0,049	0,053	0,050	0,052	0,047	0,047	0,045	0,049	0,043	0,046	0,047	0,052	0,045	0,050	0,046	0,051
6	0,059	0,070	0,060	0,062	0,063	0,065	0,060	0,065	0,060	0,058	0,058	0,062	0,053	0,060	0,058	0,066	0,058	0,062	0,059	0,063
7	0,077	0,086	0,076	0,078	0,078	0,081	0,076	0,082	0,080	0,074	0,076	0,078	0,069	0,078	0,078	0,084	0,074	0,079	0,076	0,080
8	0,102	0,110	0,103	0,102	0,102	0,106	0,104	0,106	0,106	0,096	0,102	0,104	0,092	0,102	0,106	0,110	0,101	0,105	0,102	0,105
9	0,149	0,150	0,151	0,147	0,148	0,146	0,159	0,149	0,151	0,133	0,150	0,146	0,135	0,143	0,161	0,155	0,146	0,149	0,149	0,146
10	0,469	0,415	0,482	0,469	0,461	0,439	0,451	0,444	0,457	0,496	0,479	0,462	0,517	0,477	0,459	0,427	0,483	0,454	0,475	0,454

Fonte: Resultados da pesquisa.

Por meio da Tabela 10 não é possível saber, *a priori*, com exceção da Paraíba, qual fator foi mais importante para explicar a queda na pobreza. Assim, utilizou-se a decomposição baseada no valor de Shapley para encontrar a contribuição exata de cada fator. Os resultados podem ser vistos na Tabela 11.

Tabela 11 – Decomposição da pobreza e extrema pobreza (P_0) em dois componentes: crescimento (ϕ_G^S) e redistribuição (ϕ_D^S), 2003-2008²⁸

Estados	Pobreza			Extrema pobreza		
	$\phi_G^S(\nu)$	$\phi_D^S(\nu)$	ΔP_0	$\phi_G^S(\nu)$	$\phi_D^S(\nu)$	ΔP_0
Maranhão	-0,1387	-0,0679	-0,2066	-0,0921	-0,0512	-0,1434
Piauí	-0,1773	-0,0175	-0,1948	-0,1105	-0,0433	-0,1538
Ceará	-0,1635	-0,0323	-0,1958	-0,1053	-0,0180	-0,1233
Rio G. do Norte	-0,1838	-0,0247	-0,2086	-0,0958	-0,0043	-0,1001
Paraíba	-0,1925	0,0184	-0,1741	-0,1155	0,0201	-0,0953
Pernambuco	-0,1456	-0,0318	-0,1774	-0,0982	-0,0175	-0,1157
Alagoas	-0,1289	-0,0469	-0,1758	-0,1110	-0,0029	-0,1138
Sergipe	-0,1003	-0,0570	-0,1573	-0,0598	-0,0483	-0,1080
Bahia	-0,1601	-0,0427	-0,2028	-0,1036	-0,0168	-0,1205
Nordeste	-0,1557	-0,0366	-0,1924	-0,1030	-0,0179	-0,1209

Fonte: Resultados da pesquisa.

De acordo com a Tabela 11, é possível, considerando o Nordeste como um todo, notar que, numa situação onde a distribuição de renda permaneça constante – isto é, ao seu nível de 2003 – o crescimento da renda domiciliar *per capita* média teria sido responsável pela redução de, aproximadamente, 15 p.p. na proporção de pobres. Em termos de extrema pobreza, o componente crescimento teria reduzido o nível de indigência, dado a desigualdade constante, em 10 p.p. O fator redistribuição reforçou, embora em uma proporção menor, o efeito crescimento. Em um cenário onde o crescimento econômico tivesse permanecido inalterado, entre 2003 e 2008, a queda da desigualdade verificada no período teria reduzido a pobreza em 3,6 p.p. e a extrema pobreza em 1,7 p.p. Diante disso, tem-se que o crescimento da rdpc explicou, aproximadamente, 81% da redução da pobreza e 85% da extrema pobreza na Região Nordeste. Marinho e Soares (2003), utilizando outra metodologia e um

²⁸ Os resultados referentes à P_1 e P_2 encontram-se no Anexo B (Tabelas B.1 e B.2). Optou-se por não discutir os resultados referentes a esses índices por que foi encontrada a mesma relação da verificada na Tabela 12.

período distinto do que é tratado neste estudo (1985 a 1999), decompôs a variação da pobreza nos componentes crescimento e concentração. De acordo com seus resultados, a queda na pobreza foi, no Nordeste, explicada inteiramente pelo fator crescimento (-55,73%). A desigualdade, diferentemente do que foi observado nesta pesquisa, também teve papel importante (44,27%), porém no sentido de aumentar a proporção de pobres. Todavia, de acordo com esses autores, a pobreza é mais sensível a alterações da desigualdade do que do crescimento econômico.

O Estado do Rio Grande do Norte foi o que apresentou o melhor resultado em termos de queda no número de pobres (20,86 p.p.). A maior parte dessa redução foi consequência do crescimento dos rendimentos, ou seja, 88% da mudança na proporção de pobres foi devido ao fator renda. A redistribuição, por outro lado, contribuiu de forma menos significativa, tendo participado com somente 12% da variação total. Esse resultado está fortemente associado à pequena redução da desigualdade de rendimentos observada no estado. O coeficiente de Gini, entre 2003 e 2008, caiu apenas 1,9%. No caso da extrema pobreza, encontrou-se a mesma situação, isto é, o componente crescimento explicando a maior parcela da variação da pobreza (96%).

Maranhão foi o segundo estado a apresentar a maior queda na pobreza (20,66 p.p.). A decomposição de Shapley mostrou que o crescimento na renda foi o principal responsável por essa redução. Em outras palavras, se a desigualdade tivesse permanecido constante, a pobreza teria caído – em decorrência da elevação dos rendimentos – 13,87 p.p., o que representa 67% da queda verificada no índice P_0 . A contribuição do fator redistributivo foi menor, isto é, 33% da redução na proporção de pobres foi resultado da diminuição no grau de desigualdade. Com relação à extrema pobreza, a redistribuição da renda teve um desempenho um pouco melhor, cabendo a ela 36% da variação na pobreza, enquanto o fator crescimento contribuiu com 64%.

Orair (2006) quantificou, por meio de uma metodologia distinta da utilizada neste trabalho, a contribuição do crescimento econômico (rendimentos) e da desigualdade de renda na variação da pobreza. Esse autor usou, para todos os estados do Brasil, duas linhas de pobreza e diversos subperíodos, abrangendo os anos de 1992 a 2004. Considerando o Maranhão e o período 1993/1995, cuja variação na pobreza se assemelha ao verificado no período 2003/2008 (-0,20128), Orair (2006) encontrou estimativas semelhantes ao encontrado neste trabalho, o crescimento da

renda explicou 64% da mudança na proporção de pobres, enquanto a desigualdade, 36%.

Sergipe teve a segunda maior queda no índice de desigualdade, de acordo com o coeficiente de Gini (6,5%). Essa queda levou o componente redistribuição a ter maior participação na explicação do declínio de 15,73 p.p. na proporção de pobres. Em outras palavras, a redução na iniquidade distributiva contribuiu com 36,2% da variação da pobreza. A importância da diminuição do grau de desigualdade é maior para a extrema pobreza. Aproximadamente 45% da variação na extrema pobreza foi explicada pela melhora na distribuição de renda. A contribuição do fator crescimento é relativamente menor, comparado com os restantes dos estados. Uma das justificativas para isso seria o crescimento menor dos rendimentos (27%), considerando, naturalmente, as taxas verificadas nas outras unidades federativas do Nordeste, cuja média foi de 43,3%.

Outro ponto que deve ser mencionado refere-se à Paraíba. Conforme a Tabela 11, a pobreza poderia ter caído ainda mais caso tivesse havido alguma redistribuição de renda; ou seja, numa situação onde não houvesse alteração na desigualdade, o número de pobres iria cair 19,25 p.p. Por outro lado, caso os rendimentos permanecesse constante, a pobreza teria aumentado – em decorrência do aumento na desigualdade – 1,8 p.p. Em termos de indigência (extrema pobreza), o resultado se repete, cabendo somente ao fator crescimento a queda do número de pessoas abaixo da linha de extrema pobreza.

De acordo com os resultados encontrados (Tabela 11), o fator crescimento foi o principal responsável pela queda do número de pobres e de indigentes, o que confirma a primeira hipótese deste trabalho. Porém, não é possível, com base nas estimativas feitas, inferir sobre o grau de importância da desigualdade e do crescimento econômico (renda); ou seja, não há informações suficientes para definir quem tem a maior capacidade de reduzir a pobreza. Por meio dos resultados é possível dizer apenas que, no período 2003/2008, o componente renda teve peso maior do que a redistribuição. Até porque há trabalhos que obtiveram o contrário, ou seja, a desigualdade foi mais importante. Helfand *et al.* (2009), por exemplo, utilizou a metodologia de Datt e Ravallion (1992) para decompor a variação da pobreza rural nos componentes crescimento e desigualdade. Dentre os seus resultados, eles encontraram que, no Brasil, no período 1998/2008, a queda da desigualdade teve

papel mais importante para explicar a redução do número de pobres do que o crescimento.

Como foi visto, a renda domiciliar *per capita* foi de extrema relevância para que a pobreza e a indigência caíssem de forma significativa na Região Nordeste. Em virtude disso, buscou-se desagregar a rdpc em suas diversas categorias, a saber: renda do trabalho, previdência, aluguéis e doações, investimento e outras rendas. O objetivo foi quantificar a contribuição de cada um dos tipos de renda na variação da rdpc. A Tabela 12 mostra os resultados.

Tabela 12 – Composição da renda domiciliar *per capita* (rdpc) nos estados do Nordeste do Brasil, 2003 e 2008

Rendas	MA		PI		CE		RN		PB		PE		AL		SE		BA		NE	
	2003	2008	2003	2008	2003	2008	2003	2008	2003	2008	2003	2008	2003	2008	2003	2008	2003	2008	2003	2008
Rdpc																				
Média	215,4	291,7	234,3	360,4	244,1	352,9	273,8	419,7	265,0	398,9	259,1	364,5	231,2	318,3	307,2	390,7	258,9	374,9	251,4	360,4
Renda do trabalho																				
Média	166,8	222,9	157,6	233,7	177,5	249,4	193,7	305,3	182,1	272,2	181,2	256,3	165,2	232,1	231,2	302,9	195,9	274,0	183,2	258,9
Participação no total	0,77	0,76	0,67	0,65	0,73	0,71	0,71	0,73	0,69	0,68	0,70	0,70	0,71	0,73	0,75	0,78	0,76	0,73	0,73	0,72
Previdência																				
Média	38,74	40,25	53,73	79,51	45,76	61,12	52,70	68,88	55,36	77,77	49,77	60,67	41,05	48,16	51,35	52,38	44,03	60,51	46,6	60,0
Participação no total	0,18	0,14	0,23	0,22	0,19	0,17	0,19	0,16	0,21	0,19	0,19	0,17	0,18	0,15	0,17	0,13	0,17	0,16	18,5	16,6
Aluguéis e doações																				
Média	2,27	3,79	5,96	10,66	5,37	7,66	7,34	8,88	7,61	8,09	7,09	7,06	7,57	5,85	8,34	4,76	5,35	7,9	5,8	7,2
Participação no total	0,01	0,01	0,03	0,03	0,02	0,02	0,03	0,02	0,03	0,02	0,03	0,02	0,03	0,02	0,03	0,01	0,02	0,02	0,02	0,02
Juros e bolsa família																				
Média	1,86	12,15	6,03	13,89	4,25	14,06	3,69	13,50	5,66	17,00	4,58	15,73	3,07	14,07	3,03	8,82	4,04	13,17	4,02	13,8
Participação no total	0,01	0,04	0,03	0,04	0,02	0,04	0,01	0,03	0,02	0,04	0,02	0,04	0,01	0,04	0,01	0,02	0,02	0,04	0,02	0,04
Outras rendas																				
Média	5,79	12,65	10,96	22,66	11,29	20,64	16,37	23,16	14,29	23,86	16,51	24,80	14,34	18,11	13,32	21,86	9,51	19,39	11,8	20,4
Participação no total	0,03	0,04	0,05	0,06	0,05	0,06	0,06	0,06	0,05	0,06	0,06	0,07	0,06	0,06	0,04	0,06	0,04	0,05	0,05	0,06

Fonte: Resultados da pesquisa.

Considerando a Tabela 12, observa-se que a renda domiciliar *per capita* depende fundamentalmente dos rendimentos provenientes do trabalho, cuja participação na rdpc ficou, nos estados do Nordeste, em torno de 72%. Especificamente em Sergipe, esse valor aproximou-se, em 2008, dos 80%, mostrando a importância do trabalho na composição da renda domiciliar. Entre 2003 e 2008, os rendimentos do trabalho tiveram forte crescimento em todos os estados, o que sugere que na região a quantidade de empregos cresceu. De acordo com o Ministério do Trabalho e do Emprego – Anuário Estatístico RAIS (2008), o número de postos de trabalho aumentou, entre 2003 e 2008, cerca de 36%.

Com relação à contribuição para a variação na rdpc, a renda do trabalho explicou, na Região Nordeste, 70% do aumento da renda domiciliar, cujo crescimento tinha sido de 43,33%. Em termos estaduais, destaca-se, novamente, Sergipe. A variação na renda domiciliar do estado (27,17%) foi fortemente influenciada pelos rendimentos do trabalho, do qual contribuiu com 86% da mudança total.

O rendimento advindo da previdência foi o segundo tipo de renda com maior participação na rdpc. Considerando o Nordeste de forma agregada, observa-se que, em 2003, essa categoria correspondeu a 18,5% da renda domiciliar *per capita*. Em 2008 houve ligeira queda, porém se manteve como um componente importante na composição da rdpc. Resultado semelhante foi encontrado em Helfand *et al.* (2009), porém para as áreas rurais do Brasil. De acordo com esse trabalho, as aposentadorias e pensões contribuem de forma significativa nos rendimentos das famílias rurais, perdendo, somente, para a renda do trabalho.

Em termos de contribuição para a variação na rdpc, a variável previdência explicou, no Nordeste, aproximadamente 12% do crescimento nos rendimentos, o que reflete a relevância das aposentadorias na composição da renda domiciliar. Essa importância está diretamente relacionada às mudanças inseridas na constituição de 1988, cujas alterações, de acordo com Ferreira (2006), levaram a um aumento significativo na quantidade de benefícios pagos e no seu valor médio, sobretudo nas áreas rurais. Segundo Brandt (2001) *apud* Ferreira (2006), no período de 1988 a 2000 houve crescimento de 71,6% no número de benefícios pagos pela previdência.

Um ponto que merece destaque refere-se ao investimento. Como pode ser observado (Tabela 12), entre 2003 e 2008, esse tipo de renda aumentou expressivamente. Na Região Nordeste, houve crescimento de, aproximadamente,

244%. A explicação para essa forte elevação pode estar em sua composição. A variável “investimento” inclui os juros advindos da poupança e de outras aplicações financeiras, os dividendos, os rendimentos de programas sociais (Bolsa Família, por exemplo) e outros rendimentos. A partir de 2002, os programas sociais que tinham como objetivo transferir renda para os mais pobres se expandiram de forma considerável. Diante disso, o aumento significativo da variável investimento pode ter ocorrido em virtude da expansão dos programas de transferência direta de renda. De acordo com os microdados da PNAD de 2003 e 2008, os rendimentos inseridos no componente investimento cresceram, no Nordeste, 442,5% no primeiro percentil (1% mais pobres), enquanto no último percentil (1% mais ricos) o aumento foi de apenas 22,7%. Isso sugere que o aumento significativo da variável investimento pode estar relacionado, sobretudo, ao crescimento dos rendimentos advindos de programas sociais. Todavia, não é possível ter certeza a esse respeito, visto que a PNAD de 2003 e 2008 não desagrega a variável investimento em seus vários componentes.

Como foi possível notar até o momento, a renda do trabalho foi de extrema importância para explicar a variação no rendimento domiciliar que, por sua vez, foi o principal determinante da queda da pobreza e extrema pobreza verificada no período 2003/2008. Assim sendo, na próxima seção foram obtidos os principais determinantes dos rendimentos provenientes do trabalho.

4.2. Decomposição baseada numa regressão

O viés de seletividade e a endogeneidade da variável educação são dois possíveis problemas que podem ocorrer no processo de estimação das equações de rendimentos. No intuito de resolver esses possíveis casos, procederam-se os seguintes passos: primeiro foram estimadas, por meio do procedimento de Heckman, as equações de rendimentos considerando somente o provável viés de seletividade. Em segundo lugar, estimou-se, por meio da metodologia desenvolvida por Wooldridge (2002), as equações de rendimentos considerando, simultaneamente, as possibilidades de viés de seletividade e de endogeneidade da variável educação.

4.2.1. Equação de rendimentos estimada pelo procedimento de Heckman

A primeira etapa para se obter a equação de rendimentos, por meio do procedimento de Heckman, é estimar a equação de participação, que busca analisar as variáveis que influenciam a probabilidade de uma pessoa participar ou não da força de trabalho. Essa equação pode ser vista na Tabela 13.

Tabela 13 – Equações de participação (modelo *Probit*) no mercado de trabalho nordestino, 2008

Variáveis	MA	PI	CE	RN	PB	PE	AL	SE	BA	NE
Cons.	-4,4269 (0,1839)***	-4,1358 (0,1925)***	-3,5940 (0,0332)***	-3,9817 (0,1730)***	-4,1205 (0,0802)***	-4,1933 (0,1016)***	-4,1580 (0,1984)***	-3,9905 (0,2276)***	-3,9024 (0,0874)***	-3,9719 (0,0467)***
Educ.	0,0734 (0,0066)***	0,0462 (0,0067)***	0,0637 (0,0034)***	0,0659 (0,0060)***	0,0819 (0,0062)***	0,0769 (0,0040)***	0,0612 (0,0049)***	0,0561 (0,0076)***	0,0693 (0,0032)***	0,0689 (0,0016)***
Sexo	0,978 (0,0595)***	0,9284 (0,0563)***	0,7417 (0,0332)***	0,8943 (0,0559)***	0,9254 (0,0615)***	0,9118 (0,0339)***	0,8397 (0,0602)***	0,7362 (0,0697)***	0,8801 (0,0246)***	0,8737 (0,0151)***
Cor	-0,060 (0,0553)	-0,0654 (0,6165)	-0,0528 (0,0305)*	-0,0438 (0,0444)	0,0309 (0,0375)	-0,0215 (0,0312)	0,1308 (0,0409)***	0,0568 (0,0493)	-0,0765 (0,0308)**	-0,0476 (0,0133)***
Idade	2,0123 (0,1105)***	2,0556 (0,0876)***	1,6515 (0,0506)***	1,8421 (0,1063)***	1,7657 (0,0483)***	1,8639 (0,0489)***	1,8286 (0,1167)***	1,9355 (0,1151)***	1,8117 (0,0497)***	1,8083 (0,0254)***
Idade ²	-0,2421 (0,0141)***	-0,2396 (0,0098)***	-0,1945 (0,0674)***	-0,0220 (0,0141)***	-0,2158 (0,0076)***	-0,2275 (0,0063)***	-0,2252 (0,0162)***	-0,2400 (0,0175)***	-0,2173 (0,0067)***	-0,2170 (0,0034)***
Criança	0,1158 (0,0438)**	0,0468 (0,0645)	-0,0089 (0,0314)	-0,0764 (0,0639)	0,0688 (0,0436)	0,0518 (0,0278)*	-0,0131 (0,0469)	-0,0543 (0,0627)	0,0039 (0,0281)	0,0245 (0,0131)*
Chefe	0,5633 (0,0730)***	0,5161 (0,0811)***	0,6020 (0,0278)***	0,4287 (0,0936)***	0,6252 (0,0541)***	0,4818 (0,0336)***	0,7162 (0,0664)***	0,5877 (0,0608)***	0,4867 (0,0343)***	0,5391 (0,0171)***
Observações	4814	3881	18095	4809	5534	18157	3884	4521	24635	88330
População	4228031	2130021	6028229	2267435	2759061	6064189	2177670	1487961	9885248	37027845
Teste F	31,00***	100,58***	216,19***	43,67***	25,44***	144,10***	38,57***	53,73***	159,32***	448,70***

Fonte: Resultados da pesquisa.

Nota: Educ.= Anos de estudos

*** significativo a 1%

*significativo a 5%

*significativo a 10%

Os erros-padrão (linearizado) estão entre parênteses

Os valores não assinalados indicam que não são significativos.

Os resultados do modelo *Probit* mostram, como já era esperado, um coeficiente positivo para a variável educação, sugerindo que, para todos os estados do Nordeste, os anos de estudos estão relacionados, positivamente, com a probabilidade de um indivíduo estar empregado; quanto mais anos se dedica aos estudos, maior é essa probabilidade. Em todas as estimativas, os coeficientes foram significativos ao nível de 1%.

Da mesma forma que a educação, a variável sexo obteve, em todos os estados, o sinal esperado. Conforme as estimativas, indivíduos que são do sexo masculino têm maior chance de estar empregado, sugerindo que na Região Nordeste pode haver algum tipo de discriminação entre gênero. Resultado semelhante foi encontrado por Santos (2009), que também estimou equações de rendimentos por meio do procedimento de Heckman, encontrando – na equação de participação – uma relação positiva entre o sexo masculino e maior probabilidade de estar empregado.

A variável Cor foi inserida no modelo para refletir algum tipo de discriminação racial. Todavia, o sinal encontrado não foi o esperado. Na maioria dos estados (exceto Paraíba, Alagoas e Sergipe) encontrou-se um coeficiente negativo, sugerindo que o indivíduo que não é da cor preta teria menos probabilidade de estar empregado. É importante frisar também que na maior parte desses estados o coeficiente não foi significativo a 10%. Embora o sinal tenha sido o não previsto, os trabalhos de Santos (2009), Hoffmann e Kassouf (2005), Kassouf (1994; 1997) encontraram o mesmo fenômeno, ou seja, uma relação positiva entre a cor preta e a probabilidade de estar no mercado de trabalho.

Os coeficientes relacionados às variáveis idade e idade ao quadrado, que foram incorporadas na análise para levar em consideração a experiência do indivíduo, apresentaram o sinal esperado. Em outras palavras, a experiência adquirida no decorrer do tempo tem a capacidade de aumentar a probabilidade do indivíduo estar empregado, porém após certo ponto, essa experiência tem um efeito negativo. Isso pode ser justificado, conforme Kassouf (1997), pelo fato de que o trabalhador envelhece e perde parte de suas habilidades.

Com relação à variável criança na família, a maioria dos coeficientes não apresentou significância estatística. Entretanto, aqueles que foram significativos ao nível de, no máximo, 10% tiveram coeficientes positivos, sugerindo que indivíduos que possuíam crianças menores de 4 anos tinham maior probabilidade de estar empregado. Esse resultado está associado às despesas que surgem quando existem

crianças na família. Numa situação como essa, o compromisso de melhorar a condição financeira da unidade familiar aumenta.

Por fim, o coeficiente relacionado à variável chefe apresentou o sinal esperado, mostrando que nos estados do Nordeste a participação no mercado de trabalho aumenta quando o indivíduo assume o papel de chefe da família. Esse resultado pode ser explicado pelo motivo de que o chefe da unidade familiar tem maior responsabilidade de sustentar a casa.

O segundo passo no procedimento de Heckman é obter, por meio das equações de participação (Tabela 13), a razão inversa de Mills, que é inserida como um dos regressores na equação de rendimentos. É importante ressaltar que nesse processo de estimação há a possibilidade de que a hipótese de homoscedasticidade seja violada. A grande consequência da heterocedasticidade é em relação aos erros-padrão, que tende a ser subestimado. Em virtude de se estar levando em conta o fato da PNAD ser uma amostra complexa e que os erros-padrão foram obtidos por linearização de Taylor, não é necessário se preocupar com o problema da heteroscedasticidade. Além disso, o objetivo maior são as estimativas pontuais, visto que são elas as utilizadas na decomposição de Shapley. Assim sendo, os resultados referentes às equações de rendimentos dos estados do Nordeste são mostrados na Tabela 14.

Tabela 14 – Equações de rendimentos para os estados da Região Nordeste, procedimento de Heckman, 2008

Variáveis	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)***
Idade	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)***	0,3270 (0,1872)*	0,7184 (0,0536)***	0,5918 (0,0397)***
Idade 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)***
Ocup1	0,5427 (0,0729)***	0,8395 (0,0939)***	0,6224 (0,0635)***	0,5434 (0,1013)***	0,5233 (0,0826)***	0,4970 (0,0536)***	0,6046 (0,0785)***	0,5652 (0,0887)***	0,4754 (0,0485)***	0,5529 (0,0259)***
Ocup2	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 (0,0670)	0,0782 (0,0840)	-0,0375 (0,0272)	0,0315 (0,0938)*
Indústria	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)***
Serviço	0,3665 (0,0907)***	0,7038 (0,1096)***	0,3783 (0,0522)***	0,4587 (0,0764)***	0,4482 (0,1151)***	0,2569 (0,0541)***	0,2300 (0,1189)*	0,2122 (0,0974)**	0,2254 (0,0482)***	0,3373 (0,0275)***
Construção	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)***

Fonte: Resultados da pesquisa.

Nota: Educ.= Anos de estudos, Ocup1 = são os profissionais/técnicos; Ocup2 = trabalhadores de nível médio; Ocup 3 (base) = Operários (*blue collars*).

* significativo a 10% **significativo a 5% ***significativo a 1%

Os erros-padrão (linearizado) estão entre parênteses. Os valores não assinalados indicam que não são significativos.

De acordo com a Tabela 14, observa-se que a razão inversa de Mills (λ) tem significância estatística (ao nível de 1%) em todos os estados do Nordeste, indicando que sua inclusão era necessária para evitar o viés de seletividade.

Na maioria dos casos, os coeficientes das equações são estatisticamente significativos e possuem os sinais esperados, salvo algumas poucas exceções.

Conforme os resultados, nota-se que a variável anos de estudos (*educ*), cujas estimativas são altamente significativas, tem a capacidade de impactar, positivamente, os rendimentos provenientes do trabalho. Essa constatação é válida para todos os estados do Nordeste; o que confirma os efeitos benéficos que a educação proporciona à renda dos indivíduos. As estimativas da variável *Educ* também podem ser interpretadas como a taxa de retorno da educação. Assim, por cada ano adicional de estudo, tem-se, no Nordeste como um todo, um aumento de 8,58% nos rendimentos²⁹. Em termos estaduais, Ceará e Bahia tiveram as melhores taxas de retorno, obtendo valores acima dos 8,8%. No que se refere à Bahia, o coeficiente encontrado é bem semelhante ao verificado em Lacerda (2008), cujo valor foi de 0,0873. Entretanto, essa autora, ao estimar sua equação de rendimentos para ano de 2005, não levou em conta a possibilidade de viés de seletividade. Maranhão foi o estado onde a taxa de retorno da escolaridade apresentou seu menor nível, mostrando que o incremento nos rendimentos, proporcionado por cada ano adicional de estudos, é menor do que aquele observado nas outras unidades federativas do Nordeste.

No que tange à idade, *proxy* para experiência, verifica-se que o coeficiente associado à variável idade é positivo, porém o que é relacionado com a idade ao quadrado é negativo. Isso mostra que a relação entre experiência e rendimentos tem o comportamento do U-invertido, ou seja, à medida que o indivíduo ganha mais experiência em seu ambiente de trabalho, os rendimentos tendem a crescer; entretanto, ao atingir certo ponto, a renda passa a cair. Considerando o Nordeste de forma agregada, a idade que maximizaria os rendimentos estaria em torno de 55 anos³⁰. No caso do Brasil, essa idade seria, de acordo com Kassouf (1994), 50 anos para homens e, aproximadamente, 46 anos para as mulheres. Conforme as

²⁹ Para evitar qualquer tipo de imprecisão, Wooldridge (2006) sugeriu utilizar a fórmula $100[\exp(x)-1]$ para calcular o acréscimo percentual nos rendimentos gerado por cada ano a mais de estudo, em que x é o coeficiente estimado da variável educação.

³⁰ Esse resultado foi obtido simplesmente derivando a equação de rendimentos com respeito à idade e igualando-a a zero. Como a idade está em dezenas de anos, multiplicou-se o valor encontrado por dez.

estimativas de Hoffmann (2000), para o ano de 1997, a idade que maximizaria a renda dos brasileiros estaria em torno de 50 anos.

As estimativas dos coeficientes educação, idade e idade ao quadrado estão de acordo com a teoria clássica do capital humano, que estabelece uma relação positiva entre educação e rendimentos e uma forma parabólica para a variável experiência.

Segundo Becker (1993), a teoria do capital humano assume que os anos de estudos aumentam os rendimentos e a produtividade por meio do efeito que a educação tem em melhorar as habilidades das pessoas e o modo como elas conseguem resolver seus problemas. Além disso, esse autor afirmou que as rendas dos indivíduos mais educados estão, quase sempre, acima da média, porém os ganhos são, normalmente, maiores em países menos desenvolvidos. No caso do Nordeste, região mais pobre e com maior número de pessoas analfabetas do Brasil, já era esperado que a educação tivesse papel importante para explicar os rendimentos.

A variável sexo, assim como anos de estudos, apresentou o sinal esperado, isto é, positivo. De acordo com os resultados, em todos os estados do Nordeste o sexo masculino possuía rendimentos maiores do que o feminino. Considerando a região como um todo, homens recebiam, em média, 62,2% a mais do que as mulheres, descontados os efeitos oriundos dos outros regressores. Por outro lado, Alagoas possuía um diferencial bem menor (embora elevado), sendo que a diferença no estado ficou em torno de 33%. Ueda e Hoffmann (2002) também encontraram, para o Brasil, estimativas elevadas para o coeficiente de sexo. Os valores obtidos por esses autores, referente ao ano de 2002, variaram de 0,4415 a 0,4506, dependendo do modelo. Isso sugere que, no Brasil, a diferença de rendimentos entre homens e mulheres também é elevada. No meio rural do Brasil, Ney e Hoffmann (2009) verificaram que os homens recebiam uma renda, aproximadamente, 54% superior à renda das mulheres.

Quanto à variável cor, o sinal apresentado está de acordo com o que era esperado e, na maioria dos estados, os coeficientes foram estatisticamente significativos ao nível de 1%. Então, conforme a Tabela 15, pode-se observar que a diferença de rendimentos entre as pessoas brancas e as demais (preta, parda, indígena e amarela) não é tão grande, comparado ao que foi encontrado na variável sexo. Os resultados sugerem que no Maranhão a discrepância de renda, em termos raciais, é a maior da região. Nesse estado, as pessoas da cor branca recebem, em média, 17,34%

a mais do que os indivíduos da categoria base. No Nordeste como um todo, essa diferença foi de 13%.

Segundo Lundberg e Startz (1983), a discriminação no mercado de trabalho ocorre quando indivíduos, ou grupos, com qualificações (habilidades) iguais, na média, não recebem os mesmos rendimentos, na média. Diante disso, as desigualdades de rendimentos verificadas entre gêneros e entre raças, após ter controlado as demais variáveis, sugere que há algum tipo de discriminação no mercado de trabalho do Nordeste.

As variáveis Ocup1, Ocup2 e Ocup3 (base) foram inseridas no modelo para que se possa analisar os efeitos do tipo de ocupação do trabalhador sobre os rendimentos. De acordo com os resultados, a variável Ocup1 – que representa os dirigentes em geral, os profissionais das ciências e das artes e os técnicos – apresentou o sinal esperado. Em outras palavras, os trabalhadores desse grupo possuem rendimentos bem acima daqueles que formam o grupo-base, os operários (*blue collars*). Para o Nordeste como um todo, o grupo Ocup1 recebe, em média, 73,8% a mais do que os trabalhadores braçais.

Os rendimentos dos trabalhadores de nível médio – representados pelos trabalhadores de serviços administrativos, trabalhadores dos serviços, vendedores e prestadores de serviço do comércio, membros das forças armadas e auxiliares e ocupações mal definidas – da mesma forma que em Ocup1, são maiores, exceto para Pernambuco e Bahia, do que aqueles observados em Ocup3. Piauí apresentou o maior coeficiente, significativo a 1%, relacionado à variável Ocup2, sugerindo que os trabalhadores que ocupam cargos intermediários têm rendimentos, na média, 20,3% acima do que é verificado na categoria base.

De acordo com as teorias dos diferenciais compensatórios e salário-eficiência, discutido no referencial teórico, a desigualdade de rendimentos pode ser explicada, parcialmente, pelas características próprias de cada atividade econômica. Em virtude de se ter controlado – nas equações de rendimentos – os efeitos oriundos das variáveis educação, idade, sexo, cor e setor de atividade, é possível que as estimativas das variáveis Ocup1 e Ocup2 estejam refletindo o que as teorias supracitadas estabelecem. Em outras palavras, as vantagens ou desvantagens existentes nos postos de trabalho (teoria dos diferenciais compensatórios), assim como as políticas internas adotadas por algumas firmas de fixar o salário acima do

verificado no mercado de trabalho (teoria do salário-eficiência) podem ser importantes para determinar a remuneração dos trabalhadores.

Os regressores Indústria, Serviço, Construção e Agricultura (base) têm como objetivo analisar os impactos de se trabalhar nos respectivos setores sobre os rendimentos. Todas essas variáveis tiveram os sinais de acordo com o esperado e foram estatisticamente significativas, exceto a variável indústria no estado da Bahia. Conforme as estimativas, as ocupações na indústria, serviço e construção geram rendimentos bem acima do verificado na categoria base (agricultura). Diante disso e tomando como base a teoria do mercado de trabalho dual – que assume a existência de dois segmentos no mercado: o moderno e o atrasado –, pode-se dizer que a agricultura no Nordeste é o setor que melhor se enquadra no segmento atrasado. Esse segmento, de acordo com a teoria do mercado de trabalho dual, caracteriza-se por oferecer baixos salários, por não possuir estabilidade, por não haver promoção na carreira, por maior rotatividade da mão de obra etc. Resultado semelhante foi obtido por Cirino (2008). O autor estimou, entre outras coisas, equações de rendimentos por gênero para o Brasil. De acordo com seus resultados, dentre os setores de atividades analisados (Indústria, Comércio, Construção, Agricultura e Serviços), Agricultura é o que proporciona os menores rendimentos, tanto para homens como para mulheres.

Embora seja importante identificar os principais determinantes dos rendimentos das pessoas, saber quais fatores explicam a desigualdade de renda também é de grande relevância. Assim, por meio das equações de rendimentos (Tabela 14) foram calculadas as contribuições de cada uma das variáveis para a desigualdade de rendimentos. Para computar os níveis de iniquidade foram usadas as três principais medidas, a saber, índice de Gini e os dois índices da classe de entropia generalizada: Theil-L e Theil-T. Da mesma forma que Gunatilaka e Ghotikapanich (2009), optou-se por trabalhar com algumas variáveis agregadas. Isso tende a facilitar as associações com as teorias apresentadas anteriormente. Assim, a desigualdade foi decomposta em seis componentes: Educação, Idade (idade e idade ao quadrado), Sexo, Cor, Ocupação (Ocup1 e Ocup2) e Setor (Indústria, Serviço e Construção). Os resultados podem ser vistos na Tabela 15.

Tabela 15 – Decomposição (Shapley) baseada na equação de rendimentos estimada pelo procedimento de Heckman (%), 2008

Variáveis	Coeficiente de Gini									
	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
Idade	18,86	18,78	21,78	14,66	19,26	22,33	19,33	15,87	24,99	22,24
Sexo	6,50	1,67	8,01	4,15	2,95	5,19	1,72	6,13	6,76	6,12
Cor	2,07	1,43	1,74	1,70	0,31	2,21	1,47	1,12	1,23	1,53
Ocupação	10,40	12,94	11,58	11,39	10,17	10,60	13,32	13,02	8,72	10,55
Setor	9,70	5,03	7,14	9,76	9,00	4,30	5,18	5,28	4,26	6,55
Resíduo	34,88	37,79	29,21	35,69	36,77	35,69	35,98	33,50	33,99	32,51
Total	100	100	100	100	100	100	100	100	100	100
Variáveis	Coeficiente de Theil - L									
	MA	PI	CE	RN	PB	PE	AL	SE	BA	NE
Educ	12,31	12,59	16,07	15,36	14,99	14,44	15,74	17,74	14,45	15,28
Idade	6,25	4,95	7,50	4,67	6,00	7,34	6,18	5,50	10,11	7,97
Sexo	3,35	-0,15	4,83	1,09	0,49	2,45	-0,18	2,58	3,88	3,11
Cor	1,04	0,58	0,94	0,75	0,15	1,20	0,73	0,44	0,68	0,80
Ocupação	7,73	7,85	9,17	7,98	7,36	8,29	9,81	9,13	7,00	8,37
Setor	5,89	2,09	3,69	5,20	4,78	2,10	2,62	2,95	2,00	3,60
Resíduo	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
Variáveis	Coeficiente de Theil - T									
	MA	PI	CE	RN	PB	PE	AL	SE	BA	NE
Educ	10,77	11,66	13,33	14,06	12,54	10,74	13,51	15,85	11,08	12,49
Idade	6,76	5,52	8,54	5,29	6,30	6,50	6,39	5,76	9,95	8,66
Sexo	2,96	-0,02	4,45	1,08	0,64	2,14	-0,02	2,25	3,51	2,85
Cor	1,02	0,57	0,94	0,73	0,13	1,06	0,66	0,46	0,69	0,78
Ocupação	7,47	8,05	8,82	8,27	6,85	7,28	9,33	9,00	6,59	7,96
Setor	4,54	1,69	2,44	4,12	3,35	1,22	1,96	2,34	1,09	2,33
Resíduo	66,49	72,53	61,48	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

Fonte: Resultados da pesquisa.

De acordo com os resultados, verifica-se – quando o coeficiente de Gini é utilizado – que os modelos explicam entre 62,2 (Piauí) e 70,8% (Ceará) do total da desigualdade, o que mostra o bom poder explicativo do método de decomposição. Segundo Wan e Zhou (2005), modelos que explicam somente 30 a 40% da iniquidade, deixando o restante para o resíduo, não podem ter estimativas confiáveis. O trabalho de Gunatilaka e Ghotikapanich (2009) – que teve como objetivo analisar a

evolução da desigualdade em Sri Lanka e usar técnicas de decomposição, entre elas a de Shapley, para encontrar os principais fatores causadores da má distribuição – não conseguiu explicar, por meio dos modelos utilizados, nem 50% do total da desigualdade. Houve estimativas em que o poder de explicação foi de somente 33,12%.

Focando nos resultados obtidos por meio do índice de Gini, nota-se que, em cinco estados (Piauí, Rio Grande do Norte, Paraíba, Alagoas e Sergipe), a educação foi a principal responsável pela desigualdade de rendimentos, explicando, no mínimo, 21% dela. Na região como um todo, o coeficiente de Gini, em 2008, foi de, aproximadamente, 0,55. Desse total, a educação contribuiu com 20,49%.

Em virtude do efeito positivo que a educação tem sobre os rendimentos (ver Tabela 15) e ser um fator que tende a concentrar a renda – dado que pessoas mais bem educadas, isto é, com mais anos de estudos, tendem a ter rendimentos mais elevados –, Wan (2004) enfatizou que é esperado que o coeficiente relacionado à educação seja positivo. Gunatilaka e Ghotikapanich (2009), utilizando tanto a metodologia baseada no valor de Shapley como o método de Fields, também encontraram um valor positivo e elevado para a variável educação. Nesse trabalho, os autores observaram que a situação educacional em Sri Lanka foi o segundo fator que mais explicou a desigualdade de renda do país.

Maranhão foi a unidade federativa cuja contribuição da educação para a desigualdade apresentou o menor valor. É possível que esse resultado esteja relacionado à taxa de retorno da escolaridade desse estado. De acordo com a Tabela 14, Maranhão possui, na Região Nordeste, o menor coeficiente para a variável anos de estudos, o que gerou a menor taxa de retorno de uma unidade adicional de estudo. Conforme Ney (2006), uma das razões para o poder explicativo da educação na desigualdade de renda é a taxa de retorno da escolaridade.

Outra variável que foi importante para explicar a desigualdade no Nordeste é a idade, *proxy* para experiência. De acordo com a Tabela 16, a contribuição da idade, utilizando o índice de Gini, variou de 14,7 (Rio Grande do Norte) a 25% (Bahia). No Nordeste como um todo, a experiência foi o principal responsável pelo elevado índice de Gini, contribuindo com, aproximadamente, 22%. Esse resultado difere do encontrado em Berni (2007), que utilizou a metodologia de Fields para quantificar a importância de cada variável inserida na equação de rendimentos sobre a desigualdade. De acordo com seus resultados, a idade não foi, em 1995 e 2005, um

fator tão importante para explicar a desigualdade no Nordeste. Entretanto, devem-se fazer algumas ressalvas a respeito das comparações, visto que o autor usou outra metodologia – que possui alguns problemas, já mencionados anteriormente – e as variáveis não correspondem exatamente às utilizadas neste trabalho. Além disso, o período não é o mesmo.

De forma geral, os resultados referentes à educação e à idade estão de acordo com a teoria do capital humano, que estabelece a ligação entre a desigualdade de rendimentos e a qualificação dos indivíduos. Ao mesmo tempo que a educação afeta o nível de renda, ela também pode causar a sua má distribuição, considerando que, dependendo de onde se vive e da condição financeira, nem todas as pessoas têm as mesmas oportunidades educacionais.

As variáveis que caracterizam os postos de trabalho, agregada em Ocupação (Tabela 15), foram inseridas na análise para captar os prováveis efeitos estabelecidos pelas teorias dos diferenciais compensatórios e salário-eficiência. De acordo com essas teorias, o diferencial de rendimentos pode ser explicado tanto pelas vantagens ou desvantagens presentes em cada posto de trabalho, no caso dos diferenciais compensatórios; quanto pelos salários que as firmas pagam aos seus empregados acima do praticado no mercado (teoria do salário-eficiência). Levando em conta essas teorias, as diferenças existentes entre postos de trabalho foram de extrema importância para explicar a desigualdade de rendimentos no Nordeste do País. De acordo com os resultados, essas variáveis foram responsáveis, na média, por 10,55% do índice de Gini, ou seja, foi o terceiro grupo que mais contribuiu para a iniquidade na região. Em Alagoas, o grupo Ocupação explicou 13,32% do coeficiente de Gini. Alejos (2003) identificou, pelo método de Fields, os principais determinantes da desigualdade de renda na Guatemala. Seus resultados mostram que as diferenças nos postos de trabalho também têm um peso grande para explicar a má distribuição dos rendimentos, ficando à frente das variáveis relacionadas ao capital humano.

As variáveis relativas ao setor de atividade do trabalhador, que foi inserida para levar em conta a teoria do mercado de trabalho dual, tiveram a capacidade de explicar 6,55% do índice de Gini, que é um valor relativamente significativo. Esses resultados estão, de certa forma, de acordo com a teoria da segmentação do mercado de trabalho, sendo a agricultura o setor que se enquadra melhor no perfil do segmento atrasado. É a diferença existente entre os setores que explica parte da desigualdade de renda existente na Região Nordeste.

Por meio das variáveis sexo e cor observa-se que a discriminação é um fator importante na desigualdade de renda. Na Região Nordeste, essas duas variáveis são responsáveis por 7,65% da iniquidade distributiva, mensurada pelo coeficiente de Gini. Porém, a discriminação por gênero é mais acentuada, cabendo a ela 6,12% do total da desigualdade. Embora no Piauí o rendimento dos homens fosse, aproximadamente, 87% superior à renda das mulheres, a contribuição da variável sexo para o coeficiente de Gini não foi elevado (1,67%), comparado aos demais estados. Por outro lado, Alagoas apresentou a menor diferença de renda entre os sexos (Tabela 14) e esse diferencial não foi um fator tão relevante para explicar a iniquidade de rendimentos. Fazendo analogia com a teoria da discriminação no mercado de trabalho, teríamos claramente a discriminação por parte do empregador, sendo o sexo feminino o principal alvo do preconceito. Embora a mulher pudesse ter a mesma produtividade que o homem, seus rendimentos não seriam iguais.

A decomposição também foi conduzida por meio dos índices Theil-L e Theil-T (Tabela 15). Como pode ser visto, a parcela da desigualdade que não foi explicada pelas variáveis inseridas na equação de rendimentos – ou seja, o resíduo – foi bem maior do que aquela observada quando se utilizou o coeficiente de Gini. Há casos onde as variáveis explicam somente 27,9% dos coeficientes (Piauí). Essa diferença nos resíduos também foi encontrada em Wan e Zhou (2005). Esses autores usaram os coeficientes de Gini e de Theil-L no processo de decomposição. Em todas as suas estimativas, o resíduo gerado, quando o coeficiente Theil-L foi utilizado, ficou bem acima daquele verificado quando o índice de Gini foi a medida de desigualdade usada.

Embora os resíduos sejam grandes quando os coeficientes de Theil-L e Theil-T são utilizados, as conclusões a respeito de qual variável (is) contribuíram mais para explicar a iniquidade não se alteraram, ou seja, as variáveis relacionadas ao capital humano continuam sendo as principais causadoras da desigualdade de rendimentos. As diferenças, em termos de valores, observadas quando distintas medidas são usadas pode ser explicada, conforme Wan e Zhou (2005), pela sensibilidade que cada índice possui nos diferentes segmentos da curva de Lorenz e pela utilização de distintas funções de bem-estar. Por exemplo, conforme Hoffmann (1998), o coeficiente Theil-T é mais sensível a mudanças nas rendas situadas no extremo direito da distribuição, ou seja, os relativamente mais ricos. Por outro lado, o índice Theil-L é mais sensível a alterações no extremo esquerdo da distribuição (os mais

pobres). Por fim, o coeficiente de Gini é mais sensível a modificações nas rendas situadas no intervalo com maior densidade de frequência, isto é, em volta da moda ou mediana da distribuição.

Diante do que foi encontrado pelo método de decomposição, é possível aceitar a segunda hipótese feita neste trabalho, isto é, a de que as variáveis relacionadas ao capital humano são as principais causadoras da desigualdade de rendimentos nos estados do Nordeste do Brasil. Contudo, deve-se destacar o efeito da discriminação no mercado de trabalho e os diferenciais nos postos de trabalho para explicar a desigualdade.

Como já foi enfatizado, as equações de rendimentos podem apresentar o problema da endogeneidade na variável educação. Para contornar esse possível caso, utilizou-se a metodologia desenvolvida por Wooldridge (2002), que busca controlar, simultaneamente, o viés de seletividade e a endogeneidade.

4.2.2. Viés de seletividade e endogeneidade da variável educação

As equações de rendimentos dos estados do Nordeste estimadas pelo método descrito em Wooldridge (2002) podem ser vistas na Tabela 16. Cabe destacar que os erros-padrão foram obtidos por meio de *bootstrap*, dado o problema que pode surgir ao utilizar uma variável estimada como regressor³¹.

³¹ Ver Wooldridge (2002).

Tabela 16 – Equações de rendimentos para os estados da Região Nordeste, metodologia de Wooldridge, 2008³²

Variáveis	MA	PI	CE	RN	PB	PE	AL	SE	BA	NE
Cons.	4,241 (1,108)***	3,138 (1,079)***	3,262 (1,149)***	3,347 (1,005)***	4,375 (1,522)***	4,777 (1,554)***	3,886 (1,679)**	3,783 (2,293)*	3,291 (0,666)***	2,941 (0,593)***
Educ.	0,012 (0,093)	0,117 (0,079)	0,151 (0,068)**	0,085 (0,073)	0,052 (0,136)	0,015 (0,210)	0,088 (0,151)	0,119 (0,191)	0,121 (0,047)**	0,145 (0,038)***
Idade	0,423 (0,278)	0,389 (0,293)	0,429 (0,310)	0,545 (0,268)**	0,248 (0,338)	0,234 (0,261)	0,469 (0,434)	0,343 (0,568)	0,586 (0,175)***	0,628 (0,159)***
Idade ²	-0,041 (0,029)	-0,027 (0,029)	-0,029 (0,032)	-0,045 (0,029)	-0,011 (0,034)	-0,015 (0,029)	-0,034 (0,048)	-0,015 (0,057)	-0,046 (0,019)**	-0,052 (0,016)***
Sexo	0,455 (0,140)***	0,569 (0,137)***	0,455 (0,129)***	0,509 (0,136)***	0,225 (0,213)	0,241 (0,151)	0,275 (0,207)	0,483 (0,225)**	0,455 (0,075)***	0,501 (0,067)***
Cor	0,225 (0,098)**	0,050 (0,089)	0,079 (0,662)	0,120 (0,094)	0,029 (0,124)	0,219 (0,164)	0,086 (0,099)	0,059 (0,120)	0,141 (0,041)***	0,083 (0,033)**
Ocup1	0,845 (0,458)*	0,599 (0,471)	0,361 (0,303)	0,588 (0,367)	0,744 (0,775)	0,925 (1,060)	0,597 (0,919)	0,425 (1,039)	0,362 (0,223)*	0,284 (0,188)
Ocup2	0,127 (0,093)	0,134 (0,110)	0,062 (0,029)**	0,019 (0,072)	0,061 (0,157)	0,037 (0,229)	0,017 (0,168)	0,043 (0,300)	-0,064 (0,043)	-0,028 (0,032)
Indústria	0,504 (0,292)*	0,485 (0,230)**	0,107 (0,219)	0,425 (0,262)*	0,409 (0,354)	0,623 (0,847)	0,319 (0,435)	-0,152 (0,646)	0,181 (0,195)	0,082 (0,134)
Serviço	0,576 (0,310)*	0,571 (0,273)**	0,196 (0,265)	0,497 (0,263)**	0,619 (0,429)	0,580 (0,768)	0,217 (0,509)	0,137 (0,661)	0,161 (0,182)	0,152 (0,137)
Construção	0,474 (0,208)**	0,606 (0,166)***	0,393 (0,123)***	0,442 (0,119)***	0,409 (0,184)***	0,415 (0,500)	0,138 (0,360)	0,244 (0,462)	0,190 (0,125)	0,239 (0,083)***
Lambda	-0,537 (0,288)*	-0,880 (0,342)***	-0,623 (0,309)**	-0,399 (0,292)	-0,802 (0,360)**	-0,606 (0,278)**	-0,472 (0,468)	-0,424 (0,608)	-0,406 (0,206)**	-0,316 (0,168)*
DWH (prob.)	0,447	0,567	0,326	0,830	0,629	0,402	0,967	0,737	0,551	0,133

Fonte: Resultados da pesquisa.

*** significativo a 1% **significativo a 5% *significativo a 10%. DWH= Durbin-Wu-Hausman. Os valores não assinalados indicam que não são significativos. Entre parênteses estão os erros-padrão obtidos por meio de bootstrap (300 replicações).

³² Por meio do teste de Sargan observou-se que todos os instrumentos são válidos.

No intuito de verificar a possível existência de endogeneidade na variável educação, foi executado o teste de Durbin-Wu-Hausman para exogeneidade. Assim, de acordo com a Tabela 16, não é possível rejeitar a hipótese de exogeneidade em nenhum dos estados do Nordeste. Em outras palavras, nas equações de rendimentos não há indícios de que a variável educação seja correlacionada com o resíduo, o que sugere que nestes modelos a variável possa ser considerada exógena e que o modelo mais apropriado seria aquele cujos resultados foram gerados a partir do procedimento de Heckman (Tabela 14). Embora na literatura seja mais comum que os testes de exogeneidade acusem a presença de endogeneidade para a variável educação, há também estimativas que sugerem a possível não correlação entre anos de estudos e o termo erro. Como exemplo, pode-se citar Lei (2005), que teve como objetivo calcular as taxas de retorno da educação no Canadá, considerando, separadamente, o sexo feminino e o masculino. O retorno para as mulheres variou de 3,3 a 4,9%, enquanto para os homens ficou em torno de 3,9%. No quesito endogeneidade, o teste de Durbin-Wu-Hausman sinalizou para exogeneidade (não foi possível rejeitar a hipótese de que a educação é exógena) no caso do sexo feminino, e para endogeneidade para as mulheres. Assim como Lei (2005), Khitarishvili (2010) também estimou as taxas de retorno da educação, porém para a república da Geórgia. Conforme suas estimativas, a educação também pode ser considerada exógena.

Apesar de o teste ter sugerido que a variável “anos de estudos” seja exógena, optou-se por continuar a análise a fim de fazer comparações. Em outras palavras, seria interessante saber se os resultados da decomposição são coerentes com os encontrados na Tabela 15.

De acordo com Tabela 16, observa-se que a educação continua com o sinal esperado, revelando, como sugere a teoria do capital humano, o impacto positivo dos anos de estudos sobre os rendimentos do trabalho. As variáveis idade e idade ao quadrado também apresentaram o sinal previsto; isto é, positivo e negativo, respectivamente.

As variáveis que buscam captar os efeitos da discriminação no mercado de trabalho estão de acordo com o que foi encontrado nas estimativas contidas na Tabela 14, que não considera a possibilidade de endogeneidade. Em outras palavras, é possível que na Região Nordeste o preconceito de gênero e de raça exista, sendo que a diferença de renda entre mulheres e homens é bem maior do que o verificado entre raças. Na região como um todo, o rendimento do sexo feminino foi, em média,

65% menor que o recebido pelos homens; estimativa semelhante (62,2%) ao encontrado quando foi utilizado apenas o procedimento de Heckman para obter a equação de rendimentos do Nordeste. Por outro lado, o diferencial de renda entre as pessoas brancas e as demais é bem menor. Não possuir cor preta, no Nordeste, proporcionava, em média, rendimentos 8,7% acima daqueles que possuem.

Quanto à influência dos postos de trabalho sobre os rendimentos, nota-se, assim como encontrado na Tabela 14, que os dirigentes, os profissionais das ciências e das artes e os técnicos possuem rendimentos bem acima daqueles trabalhadores que formam a categoria-base (*blue collars*). Os coeficientes relacionados à variável Ocup2 apresentaram, para a maioria dos estados, o sinal que era esperado – positivo.

No que se refere ao impacto de se trabalhar em um dado setor sobre a renda, verifica-se que, na maioria das estimativas, os rendimentos provenientes da indústria são bem maiores que o da agricultura (categoria-base). De forma análoga ao ocorrido nas estimativas da Tabela 14, o sinal do coeficiente da variável indústria, no Estado de Sergipe, não foi o esperado (positivo). Os coeficientes das variáveis Serviço e Construção apresentaram, em todos os estados, o sinal que era previsto – positivo. Em média, os trabalhadores desses setores possuem rendimentos bem acima do que aqueles inseridos na agricultura.

É importante frisar que a razão inversa de Mills (Lambda) foi estatisticamente significativa na maioria das equações de rendimentos, revelando que a sua inclusão era necessária para evitar o viés de seletividade. Somente nos Estados do Rio Grande do Norte, Alagoas e Sergipe a variável não foi significativa ao nível de 10%. Essa não significância pode estar relacionada ao fato de que o modelo que seria o mais adequado é o apresentado anteriormente (Tabela 14), visto que, pelo teste de exogeneidade, verificou-se que a variável educação não é correlacionada com o resíduo.

A partir das equações de rendimento, aplicou-se a decomposição de Shapley para quantificar a contribuição de cada variável para a desigualdade de rendimentos. Os resultados podem ser vistos na Tabela 17.

Tabela 17 – Decomposição (Shapley) baseada na equação de rendimentos estimada pelo método de Wooldridge (%), 2008

Variáveis	Coeficiente de Gini									
	MA	PI	CE	RN	PB	PE	AL	SE	BA	NE
Educ	2,30	35,53	41,68	21,83	20,58	3,30	27,49	36,23	29,99	37,76
Idade	17,05	16,75	20,06	21,67	20,43	10,88	21,11	17,61	24,27	26,03
Sexo	5,77	6,24	6,09	5,18	4,89	2,72	2,35	7,14	6,09	6,57
Cor	3,14	0,46	0,96	1,62	1,53	4,07	1,16	0,68	1,46	0,98
Ocupação	17,58	11,61	6,70	11,22	10,58	26,22	12,50	9,05	7,32	5,09
Setor	15,80	15,35	4,17	9,51	8,97	12,44	5,82	4,21	3,02	2,85
Resíduo	38,36	14,07	20,36	28,97	33,02	40,37	29,57	25,08	27,85	20,71
Total	100	100	100	100	100	100	100	100	100	100
Variáveis	Coeficiente de Theil - L									
	MA	PI	CE	RN	PB	PE	AL	SE	BA	NE
Educ	1,41	29,02	38,53	17,43	15,55	1,99	20,25	30,03	25,57	36,70
Idade	4,62	3,90	5,82	7,98	7,12	2,95	7,65	6,74	8,88	9,40
Sexo	2,45	2,56	2,37	2,07	1,84	0,43	0,30	3,17	2,91	2,94
Cor	1,57	0,22	0,53	0,81	0,72	1,89	0,59	0,27	0,88	0,56
Ocupação	12,81	9,76	5,76	9,44	8,43	17,13	10,22	7,32	6,08	4,81
Setor	9,83	11,68	2,10	5,83	5,21	6,32	3,62	2,37	1,70	1,69
Resíduo	67,33	42,86	44,90	56,44	61,13	69,29	57,35	50,10	53,98	43,91
Total	100	100	100	100	100	100	100	100	100	100
Variáveis	Coeficiente de Theil - T									
	MA	PI	CE	RN	PB	PE	AL	SE	BA	NE
Educ	1,28	26,57	31,94	15,52	15,18	1,67	18,02	26,18	20,64	29,65
Idade	4,75	4,97	6,93	8,78	8,59	2,43	7,55	7,07	8,55	9,85
Sexo	2,13	2,38	2,29	2,07	2,02	0,40	0,30	2,72	2,61	2,72
Cor	1,50	0,21	0,50	0,77	0,76	1,60	0,56	0,29	0,89	0,54
Ocupação	12,45	9,66	5,52	9,63	9,42	15,26	10,14	7,17	5,89	4,67
Setor	7,92	9,18	1,48	4,31	4,22	4,46	2,82	1,84	1,07	1,05
Resíduo	69,96	47,04	51,32	58,92	59,81	74,18	60,62	54,72	60,36	51,52
Total	100	100	100	100	100	100	100	100	100	100

Fonte: Resultados da pesquisa.

As conclusões a respeito de quais variáveis foram mais importantes para explicar a desigualdade de renda, feitas anteriormente, não se alteraram ao se utilizar as equações de rendimentos obtidas pelo método descrito em Wooldridge (2002). É importante ressaltar que as variáveis inseridas na decomposição explicam a maior parte da desigualdade quando mensurada pelo índice de Gini. No caso do Piauí, o resíduo é de somente 14%, o que demonstra o bom poder explicativo dos regressores.

Embora o teste de Durbin-Wu-Hausman tenha sugerido que a variável educação não é endógena e que o modelo mais apropriado seria aquele cujos resultados são apresentados na Tabela 14, o método de decomposição continua sugerindo que as variáveis relacionadas ao capital humano são as principais causas da desigualdade de renda na Região Nordeste. A exceção é o Estado de Pernambuco, onde a variável Ocupação foi a principal responsável pela desigualdade. Todavia, as variáveis anos de estudos e idade continuam sendo importantes para explicar a desigualdade do Estado.

O fato de o capital humano ter sido um importante responsável pela desigualdade verificada na região em 2008, é indício de que a educação pode melhorar o panorama da distribuição de renda no longo prazo, especialmente no Nordeste, onde grande proporção da população é analfabeta. Portanto, é de extrema relevância que as autoridades governamentais aumentem seus esforços para melhorar o nível educacional da Região Nordeste.

A discriminação no mercado de trabalho continua tendo papel destacado para explicar a desigualdade. Considerando o índice de Gini, as variáveis relacionadas ao gênero e à raça contribuíram com 7,55% para a iniquidade de renda no Nordeste, destacando-se, porém, o Sexo que explicou a maior parcela (6,57%). Logo em seguida, fica a variável Ocupação, que contribuiu com 5,09% do coeficiente de Gini, o que dá respaldo à teoria do diferencial compensatório e do salário-eficiência.

Quanto à decomposição feita a partir dos coeficientes de Theil – L e Theil – T, observa-se que o resíduo, assim como na Tabela 15, apresentou valor alto. Mesmo assim, as variáveis relacionadas ao capital humano (anos de estudos e idade) continuam, na maioria dos casos, sendo os principais fatores para explicar a desigualdade de rendimentos.

5. CONCLUSÕES

A pobreza e a extrema pobreza são grandes problemas que já perduram a um longo tempo no Brasil. Os números, que já são extremamente desfavoráveis no País, pioram quando se trata da Região Nordeste. É possível dizer que essa grande quantidade de pessoas abaixo da linha de pobreza está estreitamente relacionada à má distribuição de renda, visto que o País tem renda suficiente para que a pobreza e a miséria sejam erradicadas. Todavia, o que vem ocorrendo nos últimos anos é a queda persistente tanto do número de pobres como da desigualdade de rendimentos; fato este que está acontecendo tanto em nível nacional como regional. Diante desse cenário, este trabalho teve como objetivo analisar quais efeitos, crescimento da renda (*Proxy* para crescimento econômico) ou distribuição de renda foram mais importantes para explicar a queda da pobreza nos estados do Nordeste no período de 2003 a 2008. Assim, utilizou-se a decomposição de Shapley para quantificar qual componente foi mais importante para a variação na pobreza.

Os resultados da decomposição mostraram que o crescimento econômico (rendimento) foi o principal componente para explicar a queda na pobreza no período 2003 a 2008. Esse resultado foi observado em todos os estados do Nordeste, tanto para a pobreza quanto para a extrema pobreza. Paraíba foi o único estado onde o crescimento dos rendimentos explicou inteiramente a queda do número de pobres, visto que, no período analisado, houve piora na sua distribuição de renda.

Embora o fator crescimento tenha sido o principal responsável pela queda na pobreza, na maioria dos estados (exceto Paraíba, como já mencionado) a redistribuição de renda ocorrida entre os anos de 2003 a 2008 reforçou o efeito do componente crescimento. Em outras palavras, caso a desigualdade de rendimentos tivesse permanecido constante, a queda no número de pobres e de indigentes teria sido menor. Entre os estados da Região Nordeste, Sergipe foi a unidade federativa

onde o fator redistribuição obteve a maior participação para explicar a pobreza e a extrema pobreza.

Como foi possível observar, o crescimento da renda domiciliar *per capita* (rdpc) foi de fundamental importância para a redução da pobreza na Região do Nordeste. Diante desse fato, a rdpc foi desagregada em suas diversas categorias no intuito de conhecer que tipo de renda foi a principal responsável pelo seu aumento. Por meio dos resultados, concluiu-se que o rendimento do trabalho foi a categoria que mais contribuiu com a variação na renda da população do Nordeste. Em última instância, é possível dizer que os rendimentos provenientes do trabalho tiveram papel relevante na redução da pobreza. Em Sergipe, o peso da renda do trabalho na composição da renda domiciliar *per capita* foi o maior entre os nove estados do Nordeste (78% em 2008), o que mostrou a importância dos rendimentos oriundos do trabalho para a população sergipana.

Embora o crescimento dos rendimentos tenha sido fundamental, no Nordeste, para reduzir o número de pessoas abaixo da linha de pobreza e de indigência, a redistribuição de renda é um fator que deve ser perseguida, dada a longa distância que separa os mais ricos dos mais pobres. A diferença que há, em termos de iniquidade de renda, entre a Região Nordeste e as demais regiões do Brasil é extremamente elevada, o que sugere que a desigualdade tem grande margem para cair. Então, dada a relação negativa entre redistribuição de renda e pobreza, é possível que o número de pobres e de indigentes reduza-se de forma mais intensa.

Diante de um cenário onde a desigualdade é um grande obstáculo à queda na pobreza, é de fundamental importância identificar quais as variáveis seriam as mais relevantes no combate à concentração de renda. Para isso, utilizou-se uma metodologia que permite, a partir de uma equação de rendimentos, quantificar a contribuição de cada variável para a desigualdade de renda.

Por meio dos resultados provenientes da decomposição baseada numa regressão, foi possível concluir que o problema principal da desigualdade na Região Nordeste está relacionado às variáveis do capital humano. Em outras palavras, a educação e a experiência foram os fatores que mais explicaram a grande desproporção de renda entre a camada superior e a camada inferior da distribuição. A falta de qualificação e o baixo nível do ensino na região são os principais fatores que levaram os estados nordestinos a terem a pior distribuição de rendimentos do País.

Em um cenário onde se espera que o Brasil continue crescendo, o investimento na ampliação e na melhoria do ensino vai ser crucial para que a distância entre os que estão na base e no topo da distribuição reduza-se de forma significativa. Caso isso não ocorra, é possível que aconteça o que Langoni (1973) concluiu a respeito da década de 1960, período onde houve crescimento econômico e aumento na desigualdade. Naquela época, houve grande procura por mão de obra qualificada, porém não havia oferta suficiente. Assim sendo, aqueles que eram instruídos (maior nível de escolaridade) passaram a auferir grandes rendas, causando, por conseguinte, a piora na distribuição dos rendimentos. Atualmente, há indícios que isso possa vir a acontecer. De acordo com estudo do IPEA (2010), há setores econômicos que podem registrar escassez de mão de obra qualificada e com experiência. Então, políticas governamentais que tenham como objetivo treinar, instruir e qualificar a população do Nordeste pode ter a capacidade de aumentar os rendimentos dos trabalhadores mais pobres, possibilitando que a desigualdade e, por conseguinte, a pobreza caía de forma significativa. Considerando que o Nordeste é a região com o maior índice de analfabetismo no Brasil, as políticas com prioridade máxima são aquelas que busquem alfabetizar os jovens e adultos. Portanto, o primeiro passo a ser feito é levar as crianças e os adultos, que não tiveram oportunidade de estudar, para a escola e lhes fornecer condições para que eles completem seus estudos. O segundo passo refere-se à qualidade do ensino, focando na capacitação dos professores e na melhoria da infraestrutura das escolas.

É importante frisar que o benefício gerado pela diminuição da concentração de renda não se resume apenas ao seu efeito direto sobre a pobreza, mas também ao seu efeito indireto. Como observado anteriormente, há autores que encontram evidências que mostraram que o impacto do crescimento econômico sobre a pobreza depende do grau de desigualdade inicial; sendo este menor e, considerando um dado crescimento do produto (rendimentos), a queda do número de pobres tende a ser mais intensa. Os expressivos ganhos de rendimentos, observados entre os anos de 2003 e 2008, poderiam ter reduzido a pobreza num montante maior caso a renda da população do Nordeste fosse mais bem distribuída.

Além das variáveis do capital humano, aquelas relacionadas ao gênero e à cor também contribuíram positivamente para a desigualdade de renda, confirmando o que a teoria da discriminação no mercado de trabalho estabelece. Diante desse fato, pode-se concluir que na Região Nordeste pode haver alguma forma de discriminação

(por parte dos empregadores), sobretudo contra as mulheres. A diferença, em termos de renda, entre homens e mulheres é substancial – bem mais acentuada do que os diferenciais de rendimentos existentes entre raças. No Ceará, esse problema é mais acentuado. A discriminação foi um dos fatores que mais explicaram a desigualdade de rendimentos no estado. Nesse sentido, seria de fundamental importância políticas ou programas escolares (curriculares) que tenham como objetivo trabalhar a questão do preconceito com as crianças, prevenindo o aparecimento de algum tipo de discriminação nas gerações futuras. Além disso, políticas públicas – ou leis que prevêm penalidades mais rígidas – de combate à discriminação seria também de extrema relevância para reduzir a concentração de renda nos estados do Nordeste.

Outro ponto que deve ser mencionado refere-se às especificidades existentes em cada atividade econômica. De acordo com as teorias dos diferenciais compensatórios e salário-eficiência, as diferenças existentes entre as firmas e entre os postos de trabalho podem contribuir com a desigualdade de renda. Os resultados encontrados sugerem que a relação proposta por essas teorias poderia ser verdadeira, visto que as características próprias de cada posto de trabalho foram significativas para explicar a desigualdade na Região Nordeste. Os estados onde essas características foram mais acentuadas foram: Piauí, Alagoas e Sergipe. Grande parte da desigualdade de rendimentos nessas áreas é devido às diferenças entre os postos de trabalho.

Assim como qualquer pesquisa acadêmica, este trabalho apresenta algumas limitações, que podem ser encaradas como oportunidades para futuras pesquisas. Uma delas refere-se aos cálculos das elasticidades com respeito ao crescimento e à desigualdade. Neste trabalho, calculou-se somente a contribuição do crescimento da renda e da desigualdade para a variação na pobreza, o que não revela a sensibilidade com que esses componentes afetam a pobreza. A outra limitação está relacionada ao banco de dados. Utilizou-se, para estimar as equações de rendimentos, apenas o ano de 2008. É possível utilizar dados empilhados (diversos anos) para aumentar a variabilidade da amostra e obter resultados mais acurados.

REFERÊNCIAS BIBLIOGRÁFICAS

ALEJOS, L. **Contribution of the determinants of income inequality in Guatemala**. Instituto de Investigaciones Económicas y Sociales, Universidad Rafael Kandívar: Guatemala, 2003.

ALESINA, A.; RODRIK, D. Distributive politics and economic growth. **NBER Working Paper**, n. 3668, 1991.

ALESINA, A.; PEROTTI, R. Income distribution, political instability, and investment. **European Economic Review**, v.4, p.1203-1228, 1996.

ANGRIST, J. D.; KRUEGER, A. B. Does compulsory school attendance affect schooling and earnings? **The Quarterly Journal of Economics**, v.106, n.4, p.979-1014, 1991.

ARAAR, A.; DUCLOS, J. Y. **DASP: Distributive Analysis Stata Package**. PEP & CIRPÉE e World Bank, Laval University, 2009.

ARAÚJO, T. F. **As inter-relações entre pobreza, desigualdade e crescimento nas mesorregiões mineiras, 1970-2000**. 193f. Dissertação (Mestrado em Economia) – CEDEPLAR, Universidade Federal de Minas Gerais, Belo Horizonte, 2007.

BANCO MUNDIAL. **Inequality and economic development in Brazil**. Washington, D.C., 2004.

BANCO MUNDIAL. **Poverty reduction and growth: virtuous and vicious circles**. Washington, DC, 2006.

BANCO MUNDIAL. **World development report 2003: sustainable development in a dynamic world**. Washington, DC: World Bank, 2003.

BARRO, R. J. inequality and growth in a panel of countries. **Journal of Economic Growth**, v.5, n.1, p.5-32, 2000.

BARROS, R. P.; HENRIQUES, R.; MENDONÇA, R. Desigualdade e pobreza no Brasil: retrato de uma estabilidade inaceitável. **Revista Brasileira de Ciências Sociais**, v.15, n.42, p.123-142, 2000.

BARROS, R. P.; MENDONÇA, R. S. P. A evolução do bem-estar, pobreza e desigualdade no Brasil desde 1960. **Pesquisa e Planejamento Econômico**, v.25, n.1, p.115-164, 1995.

BARROS, R. P.; MENDONÇA, R. S. P. **O impacto do crescimento econômico e de reduções no grau de desigualdade sobre a pobreza**. Texto para Discussão,n.528, Instituto de Pesquisa Econômica Aplicada, 1997.

BAYE, F. M. Growth, redistribution and poverty changes in Cameroon: a Shapley decomposition analysis. **Journal of African Economics**, v.15, n.4, p.543-570, 2006.

BECKER, G. S. **The economics of discrimination**. Chicago: The University of Chicago Press, 1957.

BECKER, G. S. Investment in human capital: a theoretical analysis. **The Journal of Political Economy**, v.70, n.5, p.9-49, 1962.

BECKER, G. S. **Human Capital**: a theoretical and empirical analysis, with special reference to education. Chicago: The University of Chicago Press, 1993.

BERNI, H. A. **Evolução dos determinantes da desigualdade de renda salarial no Nordeste**. 48f. Dissertação (Mestrado em Economia) – CAEN, Universidade Federal do Ceará, Fortaleza, 2007.

CAIN, G. G. The economic analysis of labor market discrimination: a survey. In: ASHENFELTER, O.; LAYARD, R (Org). **Handbook of Labor Economics**. North Holland, 1986.

CAMERON, A. C.; TRIVEDI, P. K. **Microeconometrics using stata**. College Station, Texas: Stata Press Publication, 2009. 692 p.

CHEN, S.; WANG, Y. China's **growth and poverty reduction**: Trends between 1990 and 1999. Washington, DC: World Bank (Policy research working paper), 2001.

CACCIAMALI, M. C.; CAMILLO, V. S. Redução da desigualdade da distribuição de renda entre 2001 e 2006 nas macrorregiões brasileiras: tendência ou fenômeno transitório? **Economia e Sociedade**, v.18, n.2, p.287-315, 2009.

CIRINO, J. F. **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) – Departamento de Economia Rural, Universidade Federal de Viçosa, Minas Gerais, 2008.

COSTA, R. R. F. **O efeito da educação sobre estado de saúde individual no Brasil**. 91f. Dissertação (Mestrado em Economia) – CEDEPLAR, Universidade Federal de Minas Gerais, 2008.

COWELL, F. A.; JENKINS, S. P. How much inequality can we explain? A methodology and an application to the United States. **The Economic Journal**, v.105, n.429, p.421-430, 1995.

DATT, G.; RAVALLION, M. Growth and redistribution components of changes in poverty measures: a decomposition with application to Brazil and India in the 1980s'. **Journal of Development Economics**, v. 38, p.275-295, 1992.

Di BARTOLO, A. **Modern human capital analysis**: estimation of US, Canada and Italy earning functions. Working Paper, n. 212. Dipartimento di Statistica, Università degli studi Milano-Bicocca, Via Bicocca degli Arcimboldi, 1999.

DOERINGER, P. B.; PIORE, M. J. **Internal labor markets and manpower analysis**. Lexington, Mass: D.C. Heath – Lexington Books, 1971. 344 p.

DUMAN, A. ¿Education and income inequality in turkey: does schooling matter?. **Financial Theory and Practice**, v.32, n.3, p.369-385, 2008.

FARIAS, A. L. C. **Aplicação do teste de elegibilidade multidimensional na definição do público alvo beneficiário de políticas públicas**. 187f. Dissertação (Mestrado em Estudos Populacionais e Pesquisas Sociais) – ENCE, Escola Nacional de Ciência Estatística, 2006.

FERNANDES, R. Desigualdade salarial: aspectos teóricos. In: CORSEUIL, C. H. (Org.). **Estrutura salarial**: aspectos conceituais e novos resultados para o Brasil. Rio de Janeiro: IPEA, p. 1-51, 2002.

FERREIRA, C. R. Aposentadorias e distribuição da renda no Brasil: uma nota sobre o período 1991 a 2001. **Revista Brasileira de Economia**, v.60, n.3, p.247-260, 2006.

FERREIRA, F. H. G.; LITCHFIELD, J. A. Desigualdade, pobreza e bem-estar social no Brasil – 1981/95. In: HENRIQUES, R. (Org.). **Desigualdade e pobreza no Brasil**. Rio de Janeiro: IPEA, 2000. Cap. 2.

FERREIRA, F., LEITE, P., LITCHFIELD, J. The rise and fall of Brazilian inequality: 1981–2004. **Macroeconomic Dynamics**, v.12, p.199-230, 2008.

FIELDS, G. S.; CICHELO, P. L.; FREIJE, S.; MENÉNDEZ, M.; NEWHOUSE, D. Household income dynamics: a four-country story. **The Journal of Development Studies**, v.40, n.2, p.30-54, 2003.

FOOD AND AGRICULTURE ORGANIZATION – FAO. **The State of Food Insecurity in the World: Addressing food insecurity in protracted crises**. Disponível em <<http://www.fao.org/docrep/013/i1683e/i1683e.pdf>> acesso em: 18 de jul. 2011

FOOD AND AGRICULTURE ORGANIZATION – FAO. **Poverty**. Disponível em: <<http://www.fao.org/kids/en/poverty.html>>. Acesso em: 16 out. 2009.

FORBES, K. J. A reassessment of the relationship between inequality and growth. The **American Economic Review**, v.90, n.4, p.869-887, 2000.

FOSTER, J.; GREER, J.; THORBECKE, E. A class of decomposable poverty measures. **Econometrica**, v.52, n.3, p.761-766, 1984.

GARCÍA-PENALOSA, C.; TURNOVSKY, S. J. Growth and income inequality: a canonical model. **Economic Theory**, v.28, n.1, p.25-49, 2006.

GARNER, T. I; TERRELL, K. **Some explanations for changes in the distribution of household income in Slovakia: 1988 and 1996**. Working Paper, n.345, U.S. Department of Labour, Bureau of Labour Statistics, Office of Prices and Living Conditions, Washington, D.C. 2001.

GUIMARÃES NETO, L. O Nordeste, o planejamento regional e as armadilhas da Macroeconomia. In: **Desigualdades Regionais**. Salvador: Superintendência de Estudos Econômicos e Sociais da Bahia, p.153-175, 2004.

GUIMARÃES, P. W. **Variação de renda familiar, desigualdade e pobreza no Brasil**. 177f. Tese (Doutorado em Economia Aplicada) – Departamento de Economia Rural, Universidade Federal de Viçosa, Minas Gerais, 2007.

GUNATILAKA, R.; CHOTIKAPANICH, D. Accounting for Sri Lanka's expenditure inequality 1980-2002: regression-based decomposition approaches. **Review of Income and Wealth**, v.55, n.4, p.882-905, 2009.

HAMMILL, M. Income inequality in Central America, Dominican Republic and Mexico: assessing the importance of individual and household characteristics. **Serie Estudios y Perspectivas**, n.43, Mexico, 2005.

HARTOG, J.; OOSTERBEEK, H. What should you know about the private returns to education? In: HARTOG, J.; VAN DEN BRINK, H. M. **Human Capital: Theory and Evidence**. Cambridge University Press: Cambridge, UK, 2007.

HECKMAN, J. J. Sample selection bias as a specification error. **Econometrica**, v.47, n.1, p.153-161, 1979.

HELFAND, S. M.; ROCHA, R. VINHAIS, H. E. F. Pobreza e desigualdade de renda no Brasil rural: uma análise da queda recente. **Pesquisa e Planejamento Econômico**, v.39, n.1, p.67-88, 2009.

HELTBERG, R. **The poverty elasticity of growth**. United Nations University, World Institute for Development Economics Research – WIDER, 2002.

HOFFMANN, R. **Distribuição de renda, medidas de desigualdade e pobreza**. São Paulo: Edusp, 1998.

HOFFMANN, R. Mensuração da desigualdade e da pobreza no Brasil. In: HENRIQUES, R. (Org.). **Desigualdade e pobreza no Brasil**. Rio de Janeiro: IPEA, 2000. Cap. 3.

HOFFMANN, R.; SIMÃO, R. C. Determinantes do rendimento das pessoas ocupadas em Minas Gerais em 2000: o limiar no efeito da escolaridade e as diferenças entre mesorregiões. **Nova Economia**, v.15, n.2, p.35-62, 2005.

HOFFMANN, R. KASSOUF, A. L. Deriving conditional and unconditional marginal effects in log earnings equations estimated by Heckman's procedure. **Applied Economics**, v.37, n.11, p.1303-1311, 2005.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA – IPEA. **Comunicado da presidência n. 30**. PNAD 2008: primeiras análises. Disponível em <<http://www.ipea.gov.br>>. Acesso em: 14 out. 2009a.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA – IPEADATA. **Banco de Dados**. Disponível em: <<http://www.ipeadata.gov.br>>. Acesso em: 18 out. 2009b.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA – IPEA. **Emprego e oferta qualificada de mão de obra no Brasil**: impactos do crescimento econômico pós-crise. Comunicados do IPEA, n.41, 2010.

JIN, Y. A note on inflation, economic growth, and income inequality. **Macroeconomic Dynamics**, v.13, n.1, p.138-147, 2009.

KAKWANI, N. **On measuring growth and inequality components of poverty with application to Thailand**. Texto para discussão: School of economics, The University of New South Wales, 1997.

KALDOR, N. A model of economic growth. **Economic Journal**, v.67, n.268, p.591-624, 1957.

KASSOUF, A. L. The wage rate estimation using the Heckman procedure. **Revista de Econometria**, v.14, n.1, p.89-107, 1994.

KASSOUF, A. L. Retorno à escolaridade e ao treinamento nos setores urbanos e rural do Brasil. **Revista de Economia e Sociologia Rural**, v.35, n.2, p.59-76, 1997.

KHITARISHVILI, T. **Assessing the Returns to Education in Georgia**. Levy Economics Institute Working paper, 2010.

KRAAY, A. When is growth pro-poor? Evidence from a panel of countries. **Journal of Development Economics**, v.80, p.198-227, 2006.

KUZNETS, S. Economic growth and income inequality. **American Economic Review**, v.45, n.1, p.1-28, 1955.

LACERDA, F. C. C. Desigualdade de rendimentos na Bahia: estimação de equações de rendimentos com base nos microdados da PNAD 2005. IV Encontro de Economia Bahiana, set de 2008, Bahia. **Anais...** Salvador, BA, 2008.

LANGONI, C. G. **Distribuição de renda e desenvolvimento econômico do Brasil**. Rio de Janeiro: Editora Expressão e Cultura, 1973.

LEIBBRANDT, M.; WOOLARD, I. The labour market and household income inequality in South Africa: Existing evidence and new panel data. **Journal of International Development**, v.13, n.6, p.671-689, 2001.

LEIGH, A.; RYAN, C. Estimating returns to education using different natural experiment techniques. **Economics of Education Review**, v.27, p.149-160, 2008.

LI, H. Y.; ZOU, H.F. Income inequality is not harmful to growth, theory and evidence. **Review of Development Economics**, v.2, n.3, p.318-334, 1998.

LEI, J. J. **Parametric and semiparametric estimations of the return to schooling of wage workers in Canada**. 51f. Dissertação (Mestrado em Economia) – Departamento de Economia da Universidade do Colorado, Estados Unidos da América, 2005.

LIMA, J. R. F. **Efeitos da pluriatividade e rendas não-agrícolas sobre a pobreza e desigualdade rural na Região Nordeste**. 158f. Tese (Doutorado em Economia Aplicada) – Departamento de Economia Rural, Universidade Federal de Viçosa, Minas Gerais, 2008.

LITCHFIELD, J. A. **Inequality: methods and tools**. Washington D.C.: World Bank, 1999.

LUNDBERG, S. J.; STARTZ, R. Private discrimination and social intervention in competitive labor market. **The American Economic Review**, v.73, n.3, p.340-347, 1983.

MACHADO, D. C.; GONZAGA, G. O impacto dos fatores familiares sobre a defasagem idade-série de crianças no Brasil. **Revista Brasileira de Economia**, v.61, n.4, p.449-476, 2007.

MACKAY, A. Poverty reduction through economic growth: some issues. **Journal of International Development**, v.9, n.4, p.665-673, 1997.

MARINHO, E.; SOARES, F. Impacto do crescimento econômico e da concentração de renda sobre a redução da pobreza nos estados brasileiros. In: XXXI ENCONTRO NACIONAL DE ECONOMIA, 31., 2003, Porto Seguro. **Anais...** ANPEC, 2003.

MEDEIROS, M. **Uma introdução às representações gráficas da desigualdade de renda**. Texto para Discussão, n.1202, Instituto de Pesquisa Econômica Aplicada, 2006.

MINCER, J. **Schooling, experience, and earnings**. New York: Columbia University Press, 1974.

MINCER, J. Investment in human capital and personal income distribution. **The Journal of Political Economy**, v.66, n.4, p.281-302, 1958.

MINISTÉRIO DO TRABALHO E EMPREGO – Anuário Estatístico RAIS – **Relatório Anual de Informações Sociais** (2008). Disponível em <<http://anuariorais.caged.gov.br/>>. Acesso em: 02 fev. 2011.

MORLEY, S. **The income distribution problem in Latin America and the Caribbean**. Economic Commission of Latin America and the Caribbean (ECLAC), Libros de CEPAL, n. 65, Santiago: Chile, 2001.

NASCIMENTO SILVA, P. L.; PESSOA, D. G. C.; LILA, M. F. Análise de dados da PNAD: incorporando a estrutura do plano amostral. **Ciência e Saúde Coletiva**, v.7, n.4, p.659-670, 2002.

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

NEY, M. G.; HOFFMANN, R. Educação, concentração fundiária e desigualdade de rendimentos no meio rural brasileiro. **Revista de Economia e Sociologia Rural**, v.47, n.1, p.147-181, 2009.

ORAIR, R. O. **Como crescimento e desigualdade afetam a pobreza?** 125f. Dissertação (Mestrado em Economia) – Instituto de economia da UNICAMP, São Paulo, 2006.

ORGANIZAÇÃO MUNDIAL DA SAÚDE – WHO. **Poverty and health**. Genebra, Organização Mundial da Saúde, 14 de setembro de 1999. Disponível em: <http://apps.who.int/gb/archive/pdf_files/EB105/ee5.pdf>. Acesso em: 16 out. 2009.

OSBORNE, M. J.; RUBINSTEIN, A. **A course in game theory**. Cambridge, MA: MIT Press, 1994. 352 p.

PERSSON, T.; TABELLINI, G. Is inequality harmful for growth? **American Economic Review**, v.84, n.3, p.600-621, 1994.

PINTO, J. M. **Crescimento econômico e desigualdade de renda no estado de São Paulo: uma análise das disparidades regionais**. 101f. Dissertação (Mestrado em Economia Aplicada) – Departamento de Economia Administração e Contabilidade, Universidade de São Paulo: Ribeirão Preto, 2007.

RAMOS, L.; VIEIRA, M. L. **Desigualdade de rendimentos no Brasil nas décadas de 80 e 90: evolução e principais determinantes**. Texto para Discussão, n.803, Instituto de Pesquisa Econômica Aplicada, 2001.

RAVALLION, M. Can high-inequality developing countries escape absolute poverty? **Economics Letters**. v.56, p.51-57, 1997.

RIVEROS, L. A.; BOUTON, L. **Efficiency wage theory, labor markets, and adjustment**. PRE Working Paper, n.731, The World Bank, 1991.

ROCHA, S. **Renda e pobreza: os impactos do plano real**. Texto para Discussão, n.439, Instituto de Pesquisa Econômica Aplicada, 1996.

ROCHA, S. **Pobreza e desigualdade no Brasil: o esgotamento dos efeitos distributivos do Plano Real**. Texto para Discussão, n.721, Instituto de Pesquisa Econômica Aplicada, 2000.

ROSEN, S. The theory of equalizing differences. In: ASHENFELTER, O; LAYARD, R. (Orgs.). **Handbook of labor economics**. North-Holand, 1986.

SALARDI, P. **How much of Brazilian inequality can explain?** An attempt of income differentials using PNAD 2002. 2005. Disponível em: <<http://www-3.unipv.it/webdept/prin/workpv05.pdf>>. Acesso em: 06 dez. 2009.

SALARDI, P. **Brazilian poverty between and within groups: decomposition by geographical, group-specific poverty lines**. 2008 Disponível em: <<http://www.sussex.ac.uk/Units/PRU/wps/wp41.pdf>>. Acesso em: 10 dez. 2010.

SABOIA, J. Efeitos do salário mínimo sobre a distribuição de renda no Brasil no período 1995/2005 – resultados de simulações. **Econômica**. v. 9, n.2, p.270-295, 2007.

SANTOS, G. C. **Inserção e rendimentos no mercado de trabalho brasileiro**. 114f. Dissertação (Mestrado em Economia) – Departamento de Economia, Universidade Federal de Viçosa, Minas Gerais, 2009.

SANTOS, V. F.; VIEIRA, W. C.; REIS, B. S. Effects of alternative policies on income redistribution: Evidence from Brazil. **Development Policy Review**, v.27, n.5, p.601-616, 2009.

SCHULTZ, T. W. Capital formation by education. **The Journal of Political Economy**, v.68, n.6, p.571-583, 1960.

SCHWARTZMAN, S. **As causas da pobreza**. Rio de Janeiro: FGV, 2004. 208 p.

SHORROCKS, A. F. **Decomposition procedures for distributional analysis: a unified framework based on the Shapley value**. University of Essex, Mimeogr. 1999.

SIQUEIRA, M. L.; SIQUEIRA, M. L. Desigualdade de renda no Nordeste brasileiro: uma análise de decomposição. **XI Encontro Regional de Economia – ANPEC/Banco do Nordeste**, 2006.

SOARES, F. V.; SOARES, S.; MEDEIROS, M.; OSÓRIO, R. G. **Programas de transferências de renda no Brasil: impactos sobre a desigualdade**. Texto para Discussão, n.1228. Instituto de Pesquisa Econômica Aplicada, 2006.

SPENCE, M. Job market signaling. **The Quarterly Journal of Economics**, v.87, n.3, p.355-374, 1973.

TEIXEIRA, W. M. **Equações de rendimentos e a utilização de instrumentos para o problema de endogeneidade da educação**. 141f. Tese (Doutorado em Economia) – Departamento de Economia Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2006.

UEDA, E. M.; HOFFMANN, R. Estimando o retorno da educação no Brasil. **Revista de Economia Aplicada**, v.6, n.2, p.209-238, 2002.

WAN, G. Accounting for income inequality in rural China: a regression-based approach. **Journal of Comparative Economics**, v.32, n.2, p.348-363, 2004.

WAN, G.; ZHOU, Z. Income inequality in rural China: regression-based decomposition using household data. **Review of Development Economics**, v.9, n.1, p.107-120, 2005.

WOOLDRIDGE, J. M. **Econometric analysis of cross section and panel data**. Cambridge, Massachusetts: MIT Press, 2002. 752 p.

WOOLDRIDGE, J. M. **Introdução à econometria: uma abordagem moderna**. São Paulo: Thompson Learning, 2006. 684 p.

YELLEN, J. L. Efficiency wage models of unemployment. **The American Economic Review**, v.74, n.2, p.200-205, 1984.

ANEXOS

ANEXO A

Plano amostral da PNAD

A Pesquisa Nacional por Amostra de Domicílios – PNAD é uma pesquisa anual feita pelo Instituto Brasileiro de Geografia e Estatística (IBGE) com objetivo de produzir informações básicas a respeito do panorama socioeconômico do Brasil³³. A PNAD tem propósitos múltiplos, entre os quais, busca investigar as diversas características socioeconômicas da população como, por exemplo, educação, trabalho, rendimento e habitação.

É importante ressaltar que a PNAD é caracterizada como uma amostra complexa que envolve estratificação e conglomeração em um, dois ou três estágios de seleção, dependendo do estrato (NASCIMENTO SILVA *et al.*, 2002). Tratar a PNAD como IID (Independente e Identicamente Distribuídos) pode levar a estimativas incorretas de significância ou de intervalos de confiança. Portanto, é necessário considerar o plano amostral da PNAD para que as estimativas não sejam calculadas de forma incorreta. Segundo Lima (2008), é de fundamental importância que se conheça em qual estrato e em qual unidade primária amostral (PSU) está localizado o domicílio da amostra para que seja possível utilizar os pesos corretamente e, assim, usar as informações do desenho amostral para realizar inferências.

De acordo com Nascimento Silva *et al.* (2002), a PNAD é estratificada em duas etapas. Primeiramente, realiza-se uma estratificação que divide o Brasil em 36 estratos “naturais”, sendo que 18 deles são estratos independentes. As demais unidades da federação (Pará, Ceará, Pernambuco, Bahia, Minas Gerais, Rio de Janeiro, São Paulo, Paraná e Rio Grande do Sul) dão origem aos outros 18 estratos, visto que cada uma delas é dividida em dois estratos, a saber: Região Metropolitana e os demais municípios do estado. Essas estratificações podem ser observadas na Tabela 5.

³³ A PNAD não é realizada nos anos de Censo Demográfico.

Tabela A. 1 – Estratificação geográfica da PNAD

Estratos	
1	Distrito Federal
2	Rondônia
3	Acre
4	Amazonas
5	Roraima
6	Amapá
7	Tocantins
8	Sergipe
9	Mato Grosso do Sul
10	Mato Grosso
11	Goiás
12	Piauí
13	Rio Grande do Norte
14	Paraíba
15	Alagoas
16	Espírito Santo
17	Santa Catarina
18	Maranhão
19	Pará
20	Região metropolitana de Belém
21	Ceará
22	Região metropolitana de Fortaleza
23	Pernambuco
24	Região metropolitana de Recife
25	Bahia
26	Região metropolitana de Salvador
27	Minas Gerais
28	Região metropolitana de Belo Horizonte
29	Rio de Janeiro
30	Região metropolitana do Rio de Janeiro
31	São Paulo
32	Região metropolitana de São Paulo
33	Paraná
34	Região metropolitana de Curitiba
35	Rio Grande do Sul
36	Região metropolitana de Porto Alegre

Fonte: Farias (2006).

Conforme Nascimento Silva *et al.* (2002), nas regiões metropolitanas a amostra é estratificada por municípios e amostragem conglomerada em dois estágios, em que as unidades primárias de amostragens são os setores censitários (UPA) e as unidades secundárias (USA), os domicílios. Nos municípios (estrato) a seleção das UPAs é conduzida por meio de amostragem sistemática com probabilidade proporcional ao tamanho, cuja medida é o número de domicílios obtida do censo demográfico.

Nos 27 estratos naturais que estão fora da região metropolitana existem municípios considerados, em termos populacionais, grandes, sendo, portanto, incluídos na amostra com certeza. Esses municípios são chamados de autorrepresentativos e seu plano amostral é igual ao usado nas regiões metropolitanas, isto é, a amostragem é conglomerada em dois estágios, onde os municípios são os estratos, as UPAs são os setores censitários e as USAs, os domicílios (FARIA, 2006; NASCIMENTO SILVA *et al.*, 2002).

Por outro lado, nos municípios não autorrepresentativos, a estratificação é feita por tamanho e proximidade geográfica, onde os municípios são selecionados

por meio de amostragem sistemática, com probabilidades proporcionais ao tamanho da população (censo demográfico). No segundo estágio de seleção, os setores censitários são obtidos dentro dos municípios selecionados no processo de estratificação (primeiro estágio). Diante disso, tem-se um plano amostral conglomerado em três estágios, em que os estratos são o grupo de municípios, as UPAs são os municípios, as USAs são os setores censitários e as unidades terciárias de amostragem (UTAs) são os domicílios.

No último passo, é utilizado o cadastro de projetos de novas construções para completar a amostra. Conforme Nascimento Silva *et al.* (2002), esse cadastro possui projetos habitacionais com mais de 30 domicílios que foram formados após o censo anterior. Assim, as novas construções são estratificadas por municípios, sendo o plano amostral, dos estratos, conglomerado em um estágio. O desenho amostral da PNAD pode ser resumido por meio da Figura 6.

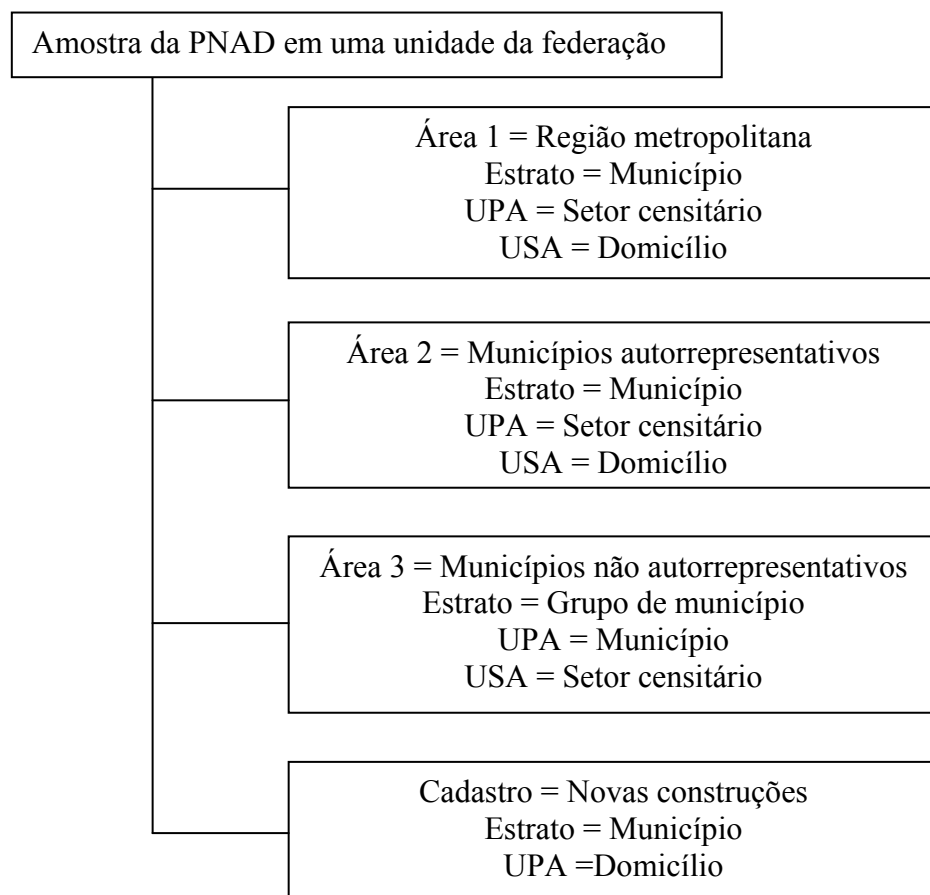


Figura A. 1 – Ilustração do plano amostral da PNAD.

Fonte: Nascimento Silva *et al.* (2002).

Para mais detalhes sobre o plano amostral da PNAD, ver Guimarães (2007) e Lima (2008).

ANEXO B

Tabela B. 1 – Decomposição da pobreza e extrema pobreza (P1) em dois componentes: crescimento (ϕ_G^S) e redistribuição (ϕ_D^S), 2003 e 2008

Estados	Pobreza			Extrema pobreza		
	$\phi_G^S(\nu)$	$\phi_D^S(\nu)$	ΔP_0	$\phi_G^S(\nu)$	$\phi_D^S(\nu)$	ΔP_0
Maranhão	-0,085	-0,045	-0,130	-0,0394	-0,0174	-0,0568
Piauí	-0,109	-0,028	-0,137	-0,0602	-0,0253	-0,0855
Ceará	-0,096	-0,025	-0,120	-0,0435	-0,0228	-0,0663
Rio G. do Norte	-0,099	-0,004	-0,103	-0,0407	0,0039	-0,0368
Paraíba	-0,105	0,009	-0,096	-0,0462	0,0071	-0,0390
Pernambuco	-0,087	-0,023	-0,111	-0,0394	-0,0174	-0,0568
Alagoas	-0,092	-0,015	-0,107	-0,0503	0,0061	-0,0442
Sergipe	-0,056	-0,033	-0,089	-0,0245	-0,0182	-0,0428
Bahia	-0,094	-0,021	-0,115	-0,0440	-0,0071	-0,0512
Nordeste	-0,093	-0,022	-0,115	-0,0435	-0,0112	-0,0547

Fonte: Resultados da pesquisa.

Tabela B. 2 – Decomposição da pobreza e extrema pobreza (P2) em dois componentes: crescimento (ϕ_G^S) e redistribuição (ϕ_D^S), 2003 e 2008

Estados	Pobreza			Extrema pobreza		
	$\phi_G^S(\nu)$	$\phi_D^S(\nu)$	ΔP_0	$\phi_G^S(\nu)$	$\phi_D^S(\nu)$	ΔP_0
Maranhão	-0,058	-0,028	-0,086	-0,024	-0,006	-0,029
Piauí	-0,077	-0,026	-0,104	-0,039	-0,017	-0,056
Ceará	-0,063	-0,023	-0,086	-0,024	-0,019	-0,043
Rio G. do Norte	-0,063	0,000	-0,062	-0,023	0,007	-0,016
Paraíba	-0,068	0,007	-0,061	-0,025	0,004	-0,021
Pernambuco	-0,058	-0,020	-0,078	-0,022	-0,014	-0,036
Alagoas	-0,066	-0,002	-0,068	-0,028	0,010	-0,019
Sergipe	-0,036	-0,024	-0,060	-0,013	-0,009	-0,022
Bahia	-0,063	-0,013	-0,075	-0,024	-0,004	-0,028
Nordeste	-0,062	-0,015	-0,077	-0,024	-0,007	-0,031

Fonte: Resultados da pesquisa.

ANEXO C

As rotinas apresentadas logo abaixo foram utilizadas para retirar as variáveis do banco de dados da PNAD. Essas rotinas foram elaboradas com base no trabalho de Lima (2008) e são feitas para serem usadas no *software* STATA 10.1. No decorrer da rotina são inseridos comentários a respeito do que foi feito. Qualquer expressão que tenha asterisco na frente o modelo reconhece como comentário.

```
*****
*DO-FILE PARA PREPARAÇÃO DE DADOS (Nordeste) PNAD 2003
*Data: 17/05/2010
*Versão: 1.0
*Autor: Vladimir Faria dos Santos
*****

clear
cap log close
log using algoritmoNE_2003, replace
set more off
set memory 300m

*Leitura das informações do desenho da amostra no arquivo de
domicílios

#delimit;
infix ano 1-4 uf 5-6 controle 5-12 serie 13-15 tipo 16-17
espdom 22-22 strat 169-175 psu 176-182 sitcen 88-88 areacen
90-90 ren_dom 154-165 size 18-19
using "c:\PNAD\PNAD_2003\dados_temp\DOM2003.txt", clear;
#delimit cr

* Dados para o meio urbano e rural do Nordeste

#delimit;
sort controle serie, stable;
format controle %15.0g;
format serie %15.0g;
replace controle = float(controle);
replace serie = float(serie);
keep if uf >= 21 & uf <= 29;
keep if espdom == 1;
#delimit cr

save dom_NE_2003, replace

sum uf
tab sitcen

* Leitura dos dados das pessoas 2003

#delimit;
infix ano 1-4 uf 5-6 controle 5-12 serie 13-15 ordem 16-17
```

```

sexo 18-18 rend_fal 736-747 areacen 766-766 sitcen 767-767
pesopes 768-772 rend_tra 700-711 apos1 547-558 apos2 519-530
pens1 533-544 pens2 561-572 abon 575-586 doac 603-614 invest
617-628 alug 589-600 pesofam 773-777 tipo_ocup 676-677 idade
27-29 educ 666-667
using "c:\PNAD\PNAD_2003\dados_temp\PES2003.txt", clear;
#delimit cr

#delimit;
keep if uf >= 21 & uf<= 29;
#delimit cr

* JUNCAO DAS INFORMACOES DE DESENHO DA AMOSTRA
* AO ARQUIVO DE PESSOAS DA PNAD2003

#delimit ;
sort controle serie, stable;
merge controle serie using dom_NE_2003;
#delimit cr

#delimit ;
keep if _merge == 3;
drop _merge;
#delimit cr

save pes_NE_2003, replace

*criando a variável renda de previdência

recode apos1 pens1 apos2 pens2 abon doac invest alug (-1 99=0)
recode apos1 pens1 apos2 pens2 abon doac invest alug (.=0)
recode rend_tra (.=0.00000001)
gen income=rend_tra+apos1+pens1+apos2+pens2+abon+doac+ invest+
alug
gen previ=apos1+apos2
recode income(0.00000001=.)

recode rend_tra(0.00000001=.)
recode previ(0=.)

*criando a variável aluguéis e doações
gen alu_doa=alug+doac

recode alu_doa(0=.)

*criando a variável outras rendas
gen outr_rend= pens1+pens2+abon+invest

recode outr_rend(0=.)

recode apos1 pens1 apos2 pens2 abon doac invest alug (0=.)

save pes_NE_2003, replace

* COLOCACAO DOS NOVOS PESOS NO ARQUIVO DE PESSOAS

```

* LEITURA DO ARQUIVO DE NOVOS PESOS

```
clear
#delimit ;
infix v0101 1-4 v0102 5-12 v0103 13-15
v0301 16-17 pesosnew 18-22 pesofamnew 23-27
using "c:\PNAD\PNAD_2003\dados_temp\PesoPes2003.txt";
rename v0102 controle;
rename v0103 serie;
rename v0301 ordem;
sort controle serie ordem, stable;
#delimit cr
```

save newpesos_NE_2003, replace

* JUNCAO DO ARQUIVO PNAD 2003 COM O ARQUIVO DE NOVOS PESOS

use pes_NE_2003, clear

```
#delimit;
format controle %15.0g;
format serie %15.0g;
format ordem %15.0g;
replace controle = float(controle);
replace serie = float(serie);
replace ordem = float(ordem);
sort controle serie ordem, stable;
merge controle serie ordem
using newpesos_NE_2003;
#delimit cr
```

```
#delimit;
keep if _merge == 3;
drop _merge;
#delimit cr
```

save pes_NE_2003, replace

*Eliminando as rendas não-declaradas

```
format rend_fal %12.0g
drop if rend_fal > 990000000000
drop if ren_dom > 990000000000
```

```
count if previ >500000 & previ !=.
drop if previ >500000 & previ !=.
count if rend_tra >500000 & rend_tra !=.
drop if rend_tra >500000 & rend_tra !=.
count if alu_doa>500000 & alu_doa !=.
drop if alu_doa>500000 & alu_doa !=.
count if outr_rend>500000 & outr_rend !=.
drop if outr_rend>500000 & outr_rend !=.
```

*Geração da renda domiciliar per capita

generate rdpc=ren_dom/size

```
drop if rdpc>5000000
```

```
*converter a rdpc para reais de setembro de 2008  
*utilizando o INPC baseado em Couseuil et al. (2002)
```

```
replace rdpc = 1.2858*rdpc  
replace rend_tra = 1.2858*rend_tra  
replace previ = 1.2858*previ  
replace alu_doa = 1.2858*alu_doa  
replace outr_rend = 1.2858*outr_rend  
replace invest = 1.2858*invest  
replace pens1 = 1.2858*pens1  
replace pens2 = 1.2858*pens2  
replace abon = 1.2858*abon
```

```
save pes_NE_2003, replace
```

```
save pes_NE_2003, replace
```

```
*Declarando o conjunto de dados como sendo de amostra complexa
```

```
use pes_NE_2003, clear
```

```
svyset psu [pweight= pesopesnew], strata(strat)  
vce(linearized)singleunit(missing)|| _n
```

```
svydes, single
```

```
* ROTINA DE ALOCACAO DE ESTRATOS COM UM UNICO PSU EM ESTRATOS  
COM MAIOR NUMERO
```

```
* DE OBSERVACOES UTILIZANDO O DO.FILE idonepsu - ANO DE 2003
```

```
use pes_NE_2003, clear  
keep if uf < 21  
gene novo_str=.  
gene novo_psu =.  
format novo_psu %12.0g  
save acum, replace
```

```
capture program drop prog2  
program define prog2  
use pes_NE_2003, clear  
keep if uf ==estado  
idonepsu, strata(strat) psu(psu) generate(novo_)  
quietly {  
save transf, replace  
use acum, replace  
append using transf  
save acum, replace  
}  
end
```

```
scalar estado =.  
foreach i in 21 22 23 24 25 26 27 28 29 {
```

```

        scalar estado = `i'
        prog2
    }
    drop strat psu
    rename novo_str strat
    rename novo_psu psu

    svyset psu[pweight=pesopesnew], strata(strat)
    vce(linearized)singleunit(missing)|| _n
    svydes, single

    save pes_NE_2003, replace

    * ROTINA DE ALOCAÇÃO DE ESTRATOS COM UM ÚNICO PSU EM ESTRATOS
    COM MAIOR NÚMERO
    * DE OBSERVAÇÕES UTILIZANDO O DO.FILE idonepsu invest - ANO
    DE 2008

    use pes_NE_2003, clear
    keep if uf < 21
    gene novo_str = .
    gene novo_psu = .
    format novo_psu %12.0g
    save acum.dta, replace

    capture program drop prog1
    program define prog1
    use pes_NE_2003, clear
    keep if uf == estado
    gene novo_str = strat
    gene novo_psu = psu
    quietly {
    save transf.dta, replace
    use acum.dta, clear
    append using transf.dta
    save acum.dta, replace
    }
    end

    capture program drop prog2
    program define prog2
    use pes_NE_2003, clear
    keep if uf == estado
    idonepsu invest, strata(strat) psu(psu) generate(novo_)
    quietly {
    save transf.dta, replace
    use acum.dta, clear
    append using transf.dta
    save acum.dta, replace
    }
    end

    scalar estado = .

    foreach i in 22 24 25 27 28 {
    scalar estado = `i'

```

```

prog1
}

foreach i in 21 23 26 29{
scalar estado = `i'
prog2
}

drop strat psu
rename novo_str strat
rename novo_psu psu

svyset psu [pweight= pesosnew], strata(strat)
vce(linearized) singleunit(missing)
svydes invest, single
save pes_NE_2003, replace

log close
view algoritmoNE_2003.smcl
exit

*****
*DO-FILE PARA PREPARAÇÃO DE DADOS (Nordeste) PNAD 2008
*Data: 06/07/2010
*Versão: 1.0
*Autor: Vladimir Faria dos Santos
*****

cd c:\tese

clear
cap log close
log using funcaoNE_2008, replace
set more off
set memory 300m

*Leitura das informações do desenho da amostra no arquivo de
domicílios

#delimit;
infix ano 1-4 uf 5-6 controle 5-12 serie 13-15 tipo 16-17
espdom 22-22 strat 159-165 psu 166-172 sitcen 81-81 areacen
83-83
probmun 94-105 probsetor 109-120
using "c:\PNAD\PNAD_2008\dados_temp\DOM2008.txt", clear;
#delimit cr

* Dados para o meio urbano e rural do Nordeste

#delimit;
sort controle serie, stable;
format controle %15.0g;
format serie %15.0g;
replace controle = float(controle);
replace serie = float(serie);

```

```

keep if uf >= 21 & uf <= 29;
keep if espdom == 1;
#delimit cr

save dom_pnad_2008, replace

sum uf
tab sitcen

* Leitura dos dados das pessoas 2008

#delimit;
infix ano 1-4 uf 5-6 controle 5-12 serie 13-15 sexo 18-18
idade 27-29 cond_dom 30-30 cond_fam 31-31 cor 33-33 horas_pr
345-346 horas_sec 439-440 cond_ativ 656-656 rend_fal 724-735
horas_out 471-472 edu 654-655 horas_cat 660-660 cond_ocup 657-
657 tipo_ocup 664-665 tipo_setor 662-663 rend_dom 712-723
rend_tot 688-699 rend_fon 700-711 rend_pr 676-687 size 738-739
areacen 740-740 rdpc 757-768 sitcen 741-741 pesopes 742-746
pesofam 747-751 num_fam 32-32 apos1 506-517 pens 520-531 apos2
534-545 pens2 548-559 abon 562-573 doac 590-601 invest 604-615
alug 576-587 bith 23-26
using "c:\PNAD\PNAD_2008\dados_temp\PES2008.txt", clear;
#delimit cr

#delimit;
keep if uf >= 21 & uf<= 29;
#delimit cr

* JUNCAO DAS INFORMACOES DE DESENHO DA AMOSTRA
* AO ARQUIVO DE PESSOAS DA PNAD2008

#delimit ;
sort controle serie, stable;
merge controle serie using dom_pnad_2008;
#delimit cr

#delimit ;
keep if _merge == 3;
drop _merge;
#delimit cr

save pes_pnad_2008, replace

*Geração da variável condição de atividade

recode cond_ativ (1=1)(2=0)

save pes_pnad_2008, replace

*Geração da variável sexo

recode sexo (2=1)(4=0)

```

*geração da variável cor:1-branco e 0-pretos, pardos, amarelo e índio

```
drop if cor ==9
recode cor(2=1)
recode cor(4 5 6 8 0=0)
```

*geração da variável tipo de ocupação do trabalhador.
*ocup1 = profissionais e técnicos
*ocup2= intermediário
*ocup3=blue collar

```
#delimit;
gen ocup1= tipo_ocup;
recode ocup1(1 2 3=1);
recode ocup1(4 5 6 7 8 9 10=0);
#delimit cr
```

```
#delimit;
gen ocup2= tipo_ocup;
recode ocup2(1 2 3 7 8=0);
recode ocup2(4 5 6 9 10=1);
#delimit cr
```

```
#delimit;
gen ocup3= tipo_ocup;
recode ocup3(1 2 3 4 5 6 9 10=0);
recode ocup3(7 8 =1);
#delimit cr
```

save pes_pnad_2008, replace

*geração da variável setor de atividade

*geração da variável setor de atividade

drop if tipo_setor==13

*sec 1 = indústria

```
#delimit;
gen sec1= tipo_setor;
recode sec1(1 4 5 6 7 8 9 10 11 12=0);
recode sec1(2 3=1);
#delimit cr
```

*sec2 = agricultura

```
#delimit;
gen sec2=tipo_setor;
recode sec2(1=1);
recode sec2(2 3 4 5 6 7 8 9 10 11 12=0);
#delimit cr
```

*sec3 = construção

```

#delimit;
gen sec3=tipo_setor;
recode sec3(1 2 3 5 6 7 8 9 10 11 12=0);
recode sec3(4=1);
#delimit cr

*sec4=serviços

#delimit;
gen sec4=tipo_setor;
recode sec4(1 2 3 4=0);
recode sec4(5 6 7 8 9 10 11 12=1);
#delimit cr

*geração da variável total de horas trabalhada (trabalho
total)

*é necessário transformar os missing value para zero para poder
somar
recode horas_pr horas_sec horas_out (-1 99=0)
recode horas_pr horas_sec horas_out (.=0)

#delimit;
gen horas_tot= horas_pr+horas_sec+horas_out;
gen horas_totmes=4*(horas_tot);
#delimit cr

save pes_pnad_2008, replace

*gerando renda de não trabalho

recode apos1 pens apos2 pens2 abon doac invest alug (-1 99=0)
recode apos1 pens apos2 pens2 abon doac invest alug (.=0)

gen rend_nt= apos1+pens+apos2+pens2+abon+doac+invest+alug

recode rend_nt (0=.)

save pes_pnad_2008, replace

*retirando outliers das variáveis renda

format rend_tot %15.0g
count if rend_tot >500000 & rend_tot!=.
drop if rend_tot >500000 & rend_tot!=.
count if rend_fon>500000 & rend_fon!=.
drop if rend_fon>500000 & rend_fon!=.
count if rend_fal>500000 & rend_fal!=.
drop if rend_fal>500000 & rend_fal!=.
count if rend_nt >500000 & rend_nt!=.
drop if rend_nt >500000 & rend_nt!=.
count if rdpc >500000 & rdpc!=.
drop if rdpc >500000 & rdpc!=.

```

```

format rend_pr %15.0g
count if rend_pr >500000 & rend_pr!=.
drop if rend_pr >500000 & rend_pr!=.

*gerando o logaritmo da razão entre renda de todos os
trabalhos
* e horas trabalhadas no mes

gen ln_rend_tot = ln(rend_tot)
drop if ln(rend_tot)<1
gen ln_horas_tot=ln(horas_totmes)
gen ln_rend_horas=(ln_rend_tot/ln_horas_tot)
save pes_pnad_2008, replace

*gerando idade em dezenas de anos
recode idade (999=.)
gen idade10=(idade/10)

*gerando idade ao quadrado
gen idade2=idade10^2

*excluindo as observações de anos de estudo que não foram
declarados
recode edu (17=.)

*gerando a variável chefe da família

gen chefe=.
replace chefe=1 if cond_fam==1
replace chefe=0 if chefe==.

*gerando a variável número de criança (0 a 4 anos) na família

gen crianc0a4=.
replace crianc0a4 = 1 if cond_fam ==3 & idade<4
replace crianc0a4 = 0 if crianc0a4==.

egen float fam = group(uf controle serie num_fam)
bysort fam: egen float famcrianc0a4 = mean(crianc0a4)
bysort fam: egen float famcriancnum0a4 = sum(famcrianc0a4)

gen crianca0a4=.
replace crianca0a4 = 1 if famcrianc0a4~=0
replace crianca0a4 = 0 if famcrianc0a4==0
sort fam

*gerando a variável número de criança ( 5 a 14 anos) na
família

gen crianc5a14=.
replace crianc5a14 = 1 if cond_fam ==3 & idade>=5 & idade<=14
replace crianc5a14 = 0 if crianc5a14==.

*para entender egen group() ver:
http://www.stata.com/help.cgi?egen,+group()

```

```

bysort fam: egen float famcrianc5a14 = mean(crianc5a14)
bysort fam: egen float famcriancnum5a14 = sum(famcrianc5a14)

gen crianca5a14=.
replace crianca5a14 = 1 if famcrianc5a14~=0
replace crianca5a14 = 0 if famcrianc5a14==0
sort fam

save pes_pnad_2008, replace

*Declarando o conjunto de dados como sendo de amostra complexa

use pes_pnad_2008, clear

svyset psu [pweight= pesopes], strata(strat) vce(linearized)
singleunit(missing)

svydes, single

* ROTINA DE ALOCAÇÃO DE ESTRATOS COM UM ÚNICO PSU EM ESTRATOS
COM MAIOR NÚMERO
* DE OBSERVAÇÕES UTILIZANDO O DO.FILE idonepsu - ANO DE 2003
use pes_pnad_2008, clear
keep if uf < 21
gene novo_str = .
gene novo_psu = .
format novo_psu %12.0g
save acum.dta,replace

capture program drop prog1
program define prog1
use pes_pnad_2008, clear
keep if uf == estado
gene novo_str = strat
gene novo_psu = psu
quietly {
save transf.dta,replace
use acum.dta, clear
append using transf.dta
save acum.dta,replace
}
end

capture program drop prog2
program define prog2
use pes_pnad_2008, clear
keep if uf == estado
idonepsu , strata(strat) psu(psu) generate(novo_)
quietly {
save transf.dta, replace
use acum.dta, clear
append using transf.dta
save acum.dta,replace
}

```

```

end

*prog1 é feito para estados sem PSU único

scalar estado = .

foreach i in 28 {
  scalar estado = `i'
  prog1
}

foreach i in 21 22 23 24 25 26 27 29{
  scalar estado = `i'
  prog2
}

drop strat psu
rename novo_str strat
rename novo_psu psu

svyset psu [pweight= pesopes], strata(strat) vce(linearized)
singleunit(missing)
svydes, single
save pes_pnad_2008, replace

* ROTINA DE ALOCAÇÃO DE ESTRATOS COM UM ÚNICO PSU EM ESTRATOS
COM MAIOR NÚMERO
* DE OBSERVAÇÕES UTILIZANDO O DO.FILE idonepsu rend_pr - ANO
DE 2008

*Ao utilizar "svydes rend_pr, single" verificou-se que há PSU
nessa variável
*Além disso, ao pedir sua média, svy: mean rend_pr, não
aparecia o desvio padrão
*Assim, teve que fazer novamente a alocação de estratos
considerando
*a variável rend_pr

use pes_pnad_2008, clear
keep if uf < 21
gene novo_str = .
gene novo_psu = .
format novo_psu %12.0g
save acum.dta, replace

capture program drop prog1
program define prog1
use pes_pnad_2008, clear
keep if uf == estado
gene novo_str = strat
gene novo_psu = psu
quietly {
save transf.dta, replace
use acum.dta, clear

```

```

append using transf.dta
save acum.dta,replace
}
end

capture program drop prog2
program define prog2
use pes_pnad_2008, clear
keep if uf == estado
idonepsu rend_pr, strata(strat) psu(psu) generate(novo_)
quietly {
save transf.dta, replace
use acum.dta, clear
append using transf.dta
save acum.dta,replace
}
end

scalar estado = .

foreach i in 21 22 23 24 25 27 {
scalar estado = `i'
prog1
}

foreach i in 26 28 29{
scalar estado = `i'
prog2
}

drop strat psu
rename novo_str strat
rename novo_psu psu

svyset psu [pweight= pesopes], strata(strat) vce(linearized)
singleunit(missing)
svydes rend_pr, single
save pes_pnad_2008, replace

* ROTINA DE ALOCAÇÃO DE ESTRATOS COM UM UNICO PSU EM ESTRATOS
COM MAIOR NUMERO
* DE OBSERVAÇÕES UTILIZANDO O DO.FILE idonepsu rend_nt - ANO
DE 2008

use pes_pnad_2008, clear
keep if uf < 21
gene novo_str = .
gene novo_psu = .
format novo_psu %12.0g
save acum.dta,replace

capture program drop prog1
program define prog1
use pes_pnad_2008, clear
keep if uf == estado

```

```

gene novo_str = strat
gene novo_psu = psu
quietly {
  save transf.dta, replace
  use acum.dta, clear
  append using transf.dta
  save acum.dta, replace
}
end

capture program drop prog2
program define prog2
  use pes_pnad_2008, clear
  keep if uf == estado
  idonepsu rend_nt, strata(strat) psu(psu) generate(novo_)
  quietly {
    save transf.dta, replace
    use acum.dta, clear
    append using transf.dta
    save acum.dta, replace
  }
end

scalar estado = .

foreach i in 24 27 28 {
  scalar estado = `i'
  prog1
}

foreach i in 21 22 23 25 26 29{
  scalar estado = `i'
  prog2
}

drop strat psu
rename novo_str strat
rename novo_psu psu

svyset psu [pweight= pesopes], strata(strat) vce(linearized)
singleunit(missing)
svydes rend_nt, single
save pes_pnad_2008, replace

*svydes, finalstage /*comando para detectar a existencia de
obs. unica em psu*/

log close
view funcaoNE_2008.smcl
exit

```

```
*****
*****
*DO-FILE PARA PREPARAÇÃO DE DADOS, Metodologia de Wooldridge
(2002) (Nordeste) PNAD 2008
*Data: 06/12/2010
*Versão: 1.0
*Autor: Vladimir Faria dos Santos
*****
*****
```

```
cd c:\endogeneidade
```

```
clear
cap log close
log using funcaoNE2_2008, replace
set more off
set memory 300m
```

```
*Leitura das informações do desenho da amostra no arquivo de
domicílios
```

```
#delimit;
infix ano 1-4 uf 5-6 controle 5-12 serie 13-15 tipo 16-17
espdom 22-22 strat 159-165 psu 166-172 sitcen 81-81 areacen
83-83 probmun 94-105 probsetor 109-120
using "c:\PNAD\PNAD_2008\dados_temp\DOM2008.txt", clear;
#delimit cr
```

```
* Dados para o meio urbano e rural do Nordeste
```

```
#delimit;
sort controle serie, stable;
format controle %15.0g;
format serie %15.0g;
replace controle = float(controle);
replace serie = float(serie);
keep if uf >= 21 & uf <= 29;
keep if espdom == 1;
#delimit cr
```

```
save dom_pnad2_2008, replace
```

```
sum uf
tab sitcen
```

```
* Leitura dos dados das pessoas 2008
```

```
#delimit;
infix ano 1-4 uf 5-6 controle 5-12 serie 13-15 sexo 18-18
idade 27-29 cond_dom 30-30 cond_fam 31-31 cor 33-33
horas_pr 345-346 horas_sec 439-440 cond_ativ 656-656 rend_fal
724-735 horas_out 471-472 edu 654-655 horas_cat 660-660
cond_ocup 657-657 tipo_ocup 664-665 tipo_setor 662-663
```

```

rend_dom 712-723 rend_tot 688-699 rend_fon 700-711 rend_pr
676-687 size 738-739 areacen 740-740 rdpc 757-768 sitcen 741-
741 pesopes 742-746 pesofam 747-751 num_fam 32-32 apos1 506-
517 pens 520-531 apos2 534-545 pens2 548-559 abon 562-573 doac
590-601 invest 604-615 alug 576-587 bith 23-26 mes 21-22
using "c:\PNAD\PNAD_2008\dados_temp\PES2008.txt", clear;
#delimit cr

#delimit;
keep if uf >= 21 & uf<= 29;
#delimit cr

* JUNCAO DAS INFORMACOES DE DESENHO DA AMOSTRA
* AO ARQUIVO DE PESSOAS DA PNAD2008

#delimit ;
sort controle serie, stable;
merge controle serie using dom_pnad2_2008;
#delimit cr

#delimit ;
keep if _merge == 3;
drop _merge;
#delimit cr

save pes_pnad2_2008, replace

*Geração da variável condição de atividade

recode cond_ativ (1=1)(2=0)

save pes_pnad2_2008, replace

*Geração da variável sexo

recode sexo (2=1)(4=0)

*geração da variável cor:1-branco e 0-pretos, pardos, amarelo
e índio

drop if cor ==9
recode cor(2=1)
recode cor(4 5 6 8 0=0)

*geração da variável tipo de ocupação do trabalhador.
*ocup1 = profissionais e técnicos
*ocup2= intermediário
*ocup3=blue collar

#delimit;
gen ocup1= tipo_ocup;
recode ocup1(1 2 3=1);
recode ocup1(4 5 6 7 8 9 10=0);
#delimit cr

#delimit;

```

```

gen ocup2= tipo_ocup;
recode ocup2(1 2 3 7 8=0);
recode ocup2(4 5 6 9 10=1);
#delimit cr

#delimit;
gen ocup3= tipo_ocup;
recode ocup3(1 2 3 4 5 6 9 10=0);
recode ocup3(7 8 =1);
#delimit cr

save pes_pnad2_2008, replace

*geração da variável setor de atividade

drop if tipo_setor==13

*sec 1 =  indústria

#delimit;
gen sec1= tipo_setor;
recode sec1(1 4 5 6 7 8 9 10 11 12=0);
recode sec1(2 3=1);
#delimit cr

*sec2 = agricultura

#delimit;
gen sec2=tipo_setor;
recode sec2(1=1);
recode sec2(2 3 4 5 6 7 8 9 10 11 12=0);
#delimit cr

*sec3 = construção

#delimit;
gen sec3=tipo_setor;
recode sec3(1 2 3 5 6 7 8 9 10 11 12=0);
recode sec3(4=1);
#delimit cr

*sec4=serviços

#delimit;
gen sec4=tipo_setor;
recode sec4(1 2 3 4=0);
recode sec4(5 6 7 8 9 10 11 12=1);
#delimit cr

*geração da variável total de horas trabalhada (trabalho
total)

```

```

*é necessário transformar os missing value para zero para poder
*somar
recode horas_pr horas_sec horas_out (-1 99=0)
recode horas_pr horas_sec horas_out (.=0)

#delimit;
gen horas_tot= horas_pr+horas_sec+horas_out;
gen horas_totmes=4*(horas_tot);
#delimit cr

save pes_pnad2_2008, replace

*gerando renda de não trabalho

recode apos1 pens apos2 pens2 abon doac invest alug (-1 99=0)
recode apos1 pens apos2 pens2 abon doac invest alug (.=0)

gen rend_nt= apos1+pens+apos2+pens2+abon+doac+invest+alug

recode rend_nt (0=.)

save pes_pnad2_2008, replace

*retirando outliers das variáveis renda

format rend_tot %15.0g
count if rend_tot >500000 & rend_tot!=.
drop if rend_tot >500000 & rend_tot!=.
count if rend_fon>500000 & rend_fon!=.
drop if rend_fon>500000 & rend_fon!=.
count if rend_fal>500000 & rend_fal!=.
drop if rend_fal>500000 & rend_fal!=.
count if rend_nt >500000 & rend_nt!=.
drop if rend_nt >500000 & rend_nt!=.
count if rdpc >500000 & rdpc!=.
drop if rdpc >500000 & rdpc!=.

format rend_pr %15.0g
count if rend_pr >500000 & rend_pr!=.
drop if rend_pr >500000 & rend_pr!=.

*gerando o logaritmo da razão entre renda de todos os
trabalhos
* e horas trabalhadas no mes

gen ln_rend_tot = ln(rend_tot)
drop if ln(rend_tot)<1
gen ln_horas_tot=ln(horas_totmes)
gen ln_rend_horas=(ln_rend_tot/ln_horas_tot)
save pes_pnad2_2008, replace

*gerando idade em dezenas de anos
recode idade (999=.)
gen idade10=(idade/10)

*gerando idade ao quadrado

```

```

gen idade2=idade10^2

*excluindo as observações de anos de estudo que não foram
declarados
recode edu (17=.)

*gerando a variável chefe da família

gen chefe=.
replace chefe=1 if cond_fam==1
replace chefe=0 if chefe==.

*gerando a variável número de criança (0 a 4 anos) na família

gen crianc0a4=.
replace crianc0a4 = 1 if cond_fam ==3 & idade<4
replace crianc0a4 = 0 if crianc0a4==.

egen float fam = group(uf controle serie num_fam)
bysort fam: egen float famcrianc0a4 = mean(crianc0a4)
bysort fam: egen float famcriancnum0a4 = sum(famcrianc0a4)

gen crianca0a4=.
replace crianca0a4 = 1 if famcrianc0a4~=0
replace crianca0a4 = 0 if famcrianc0a4==0
sort fam

*gerando a variável número de criança ( 5 a 14 anos) na
família

gen crianc5a14=.
replace crianc5a14 = 1 if cond_fam ==3 & idade>=5 & idade<=14
replace crianc5a14 = 0 if crianc5a14==.

bysort fam: egen float famcrianc5a14 = mean(crianc5a14)
bysort fam: egen float famcriancnum5a14 = sum(famcrianc5a14)

gen crianca5a14=.
replace crianca5a14 = 1 if famcrianc5a14~=0
replace crianca5a14 = 0 if famcrianc5a14==0
sort fam

save pes_pnad2_2008, replace

*geração da variável oferta de escolas

gen escola=bith
drop if escola<1935
drop if escola >2000

```

* Maranhão

```
#delimit;  
recode escola (2000=13307) (1999=13312) (1998=13581)  
(1997=13406) (1996=13014) (1995=13045) (1994=12878)  
(1993=12784) (1992=12010) (1991=11965) (1990=11992)  
(1989=12019) (1988=11161) (1987=9888) (1986=9442) (1985=8766)  
(1984=8723) (1983=9502) (1982=9504) (1981=9595)  
(1980=8928) (1979=8704) (1978=8654) (1977=8224) (1976=7576)  
(1975=8367) (1974=7456) (1973=6955) (1972=6043) (1971=4995)  
(1970=5046) (1969=4623) (1968=4698) (1967=4083) (1966=3795)  
(1965=3139) (1964=3001) (1963=2755) (1962=2373) (1961=2016)  
(1960=1784) (1959=1868) (1958=1900) (1957=1788)  
(1956=1719) (1955=1690) (1954=1581) (1953=1274) (1952=1005)  
(1951=1077) (1950=1041) (1949=822) (1948=906) (1947=737)  
(1946=764) (1945=747) (1944=810) (1943=884) (1942=893)  
(1941=816) (1940=925) (1939=672) (1938=544) (1937=412)  
(1936=409) (1935=390) if uf==21;  
#delimit cr
```

*Piauí

```
#delimit;  
recode escola (2000=7797) (1999=7864) (1998=7964) (1997=8234)  
(1996=8259) (1995=8446) (1994=8514) (1993=8260) (1992=8382)  
(1991=8135) (1990=8179) (1989=8223) (1988=7988) (1987=7615)  
(1986=7254) (1985=7170) (1984=6956) (1983=7048) (1982=6903)  
(1981=6805) (1980=7888) (1979=6458) (1978=6825) (1977=6665)  
(1976=5111) (1975=5313) (1974=4831) (1973=5237) (1972=4296) (1971=3  
608) (1970=2613) (1969=2699) (1968=2999)  
(1967=2274) (1966=2112)  
(1965=2243) (1964=2604) (1963=2163) (1962=1403) (1961=1311)  
(1960=1175) (1959=1327) (1958=1373) (1957=1411) (1956=1563)  
(1955=1152) (1954=1052) (1953=821) (1952=808) (1951=753) (1950=781)  
(1949=665) (1948=691) (1947=638) (1946=519) (1945=600)  
(1944=610) (1943=602) (1942=597) (1941=485) (1940=462)  
(1939=449) (1938=444) (1937=407) (1936=340) (1935=266) if  
uf==22;  
#delimit cr
```

*Ceará

```
#delimit;  
recode escola (2000=11444) (1999=11720) (1998=12064)  
(1997=13664) (1996=14058) (1995=13517) (1994=13682)  
(1993=13539) (1992=14017) (1991=14254) (1990=13664)  
(1989=13074) (1988=14303) (1987=13635) (1986=13373)  
(1985=12779) (1984=16645) (1983=16263) (1982=17476)  
(1981=17483) (1980=21875) (1979=19809) (1978=18046)  
(1977=16759) (1976=15610) (1975=16980) (1974=16817)  
(1973=14240) (1972=14059) (1971=11653) (1970=9698) (1969=9815)  
(1968=9680) (1967=6461) (1966=6893) (1965=8353) (1964=8179)  
(1963=8161) (1962=6626) (1961=7484) (1960=7040) (1959=6923)  
(1958=5584) (1957=5988) (1956=6174) (1955=5858) (1954=5287)  
(1953=4305) (1952=3873) (1951=3574) (1950=3724) (1949=2637)
```

```
(1948=3007) (1947=2266) (1946=2036) (1945=1933) (1944=1899)
(1943=1848) (1942=1850) (1941=1715) (1940=1629) (1939=1623)
(1938=1549) (1937=1191) (1936=954) (1935=867) if uf==23;
#delimit cr
```

*Rio Grande do Norte

```
#delimit;
recode escola (2000=4082) (1999=4099) (1998=4129) (1997=4229)
(1996=4312) (1995=4321) (1994=4368) (1993=4480) (1992=4526)
(1991=4515) (1990=4587) (1989=4659) (1988=4541) (1987=4553)
(1986=4584) (1985=4504) (1984=4362) (1983=4668) (1982=4556)
(1981=3756) (1980=5486) (1979=4733) (1978=4520) (1977=4457)
(1976=4293) (1975=4236) (1974=4128) (1973=4166) (1972=4168)
(1971=3857) (1970=3795) (1969=3615) (1968=3548) (1967=3319)
(1966=3293) (1965=3401) (1964=3165) (1963=3046) (1962=2801)
(1961=2657) (1960=2456) (1959=2293) (1958=2215) (1957=1903)
(1956=1722) (1955=1718) (1954=1545) (1953=1307) (1952=1226)
(1951=1212) (1950=1219) (1949=953) (1948=972) (1947=934)
(1946=868) (1945=784) (1944=757) (1943=772) (1942=730)
(1941=660) (1940=601) (1939=612) (1938=575) (1937=519)
(1936=432) (1935=415) if uf==24;
#delimit cr
```

*Paraíba

```
#delimit;
recode escola (2000=6922) (1999=6957) (1998=7074) (1997=7765)
(1996=7622) (1995=7701) (1994=7699) (1993=7777) (1992=7749)
(1991=7780) (1990=7867) (1989=7955) (1988=8079) (1987=8349)
(1986=8534) (1985=8004) (1984=8355) (1983=9264) (1982=9046)
(1981=8612) (1980=9410) (1979=8018) (1978=8204) (1977=7904)
(1976=5906) (1975=7917) (1974=7533) (1973=7281)
(1972=6445) (1971=5912) (1970=5446) (1969=5436) (1968=4790)
(1967=4491) (1966=4596) (1965=4607) (1964=4234) (1963=3861)
(1962=3258) (1961=3110) (1960=2645) (1959=2645) (1958=2590)
(1957=1948) (1956=1909) (1955=1920) (1954=1830) (1953=1841)
(1952=1852) (1951=1574) (1950=1713) (1949=1149) (1948=1318)
(1947=979) (1946=1003) (1945=1021) (1944=991) (1943=1041)
(1942=1079) (1941=875) (1940=983) (1939=1003) (1938=983)
(1937=905) (1936=770) (1935=649) if uf==25;
#delimit cr
```

*Pernambuco

```
#delimit;
recode escola (2000=10889) (1999=10910) (1998=10848)
(1997=11126) (1996=10990) (1995=10989) (1994=10922)
(1993=10806) (1992=10997) (1991=10817) (1990=10805)
(1989=10793) (1988=10740) (1987=10792) (1986=10757)
(1985=10510) (1984=10808) (1983=10668) (1982=10675)
(1981=10481) (1980=12691) (1979=10551) (1978=10211) (1977=9638)
(1976=9695) (1975=11560) (1974=10861) (1973=10623)
(1972=10371) (1971=9709) (1970=8885) (1969=9397) (1968=8687)
(1967=7574) (1966=7623) (1965=7387) (1964=8173) (1963=7828)
(1962=6536) (1961=6450) (1960=6602) (1959=6506) (1958=6611)
```

```

(1957=6417) (1956=5371)(1955=5371) (1954=5094) (1953=4505)
(1952=4430) (1951=4330) (1950=4380) (1949=2648) (1948=2912)
(1947=2383) (1946=2373) (1945=2506) (1944=2461) (1943=2515)
(1942=2522) (1941=2340) (1940=2303) (1939=2195) (1938=2097)
(1937=2027) (1936=1869) (1935=1860) if uf==26;
#delimit cr

```

*Alagoas

```

#delimit;
recode escola (2000=3774) (1999=3725) (1998=3823) (1997=3909)
(1996=3836) (1995=3826)(1994=3803) (1993=3725) (1992=3672)
(1991=3670) (1990=3479) (1989=3288) (1988=3446)(1987=3526)
(1986=3518) (1985=3469) (1984=3360) (1983=3239) (1982=3180)
(1981=3280) (1980=3703)(1979=3020) (1978=2940) (1977=2844)
(1976=2653) (1975=2852) (1974=2735) (1973=2760)
(1972=2792)(1971=2785) (1970=2605) (1969=2576) (1968=2544)
(1967=2393) (1966=2249) (1965=2248) (1964=2246) (1963=2032)
(1962=1811) (1961=1553) (1960=1578) (1959=1479) (1958=1511)
(1957=1399) (1956=1335)(1955=1248) (1954=1230) (1953=1148)
(1952=1056) (1951=1013) (1950=1035) (1949=697) (1948=727)
(1947=667) (1946=656) (1945=736) (1944=775) (1943=799)
(1942=753) (1941=683) (1940=678) (1939=673) (1938=614)
(1937=683) (1936=550) (1935=621) if uf==27;
#delimit cr

```

*Sergipe

```

#delimit;
recode escola (2000=2287) (1999=2287) (1998=2277) (1997=2211)
(1996=2160) (1995=2126)(1994=2094) (1993=2079) (1992=2062)
(1991=2047) (1990=2042) (1989=2038) (1988=1924)(1987=1954)
(1986=1971) (1985=1948) (1984=1956) (1983=1925) (1982=2068)
(1981=2033) (1980=2460)(1979=2102) (1978=2156) (1977=1963)
(1976=1885) (1975=2148) (1974=2112) (1973=2112)
(1972=1931)(1971=1887) (1970=1753) (1969=1566) (1968=1521)
(1967=1424) (1966=1511) (1965=1716) (1964=1411) (1963=1315)
(1962=1248) (1961=1254) (1960=1241) (1959=1217) (1958=1102)
(1957=1069) (1956=1020)(1955=901) (1954=885) (1953=814)
(1952=752) (1951=760) (1950=756) (1949=710) (1948=747)
(1947=672) (1946=662) (1945=741) (1944=719) (1943=727)
(1942=732) (1941=655) (1940=641) (1939=619) (1938=588)
(1937=422) (1936=387) (1935=357) if uf==28;
#delimit cr

```

*Bahia

```

#delimit;
recode escola (2000=26853) (1999=27172) (1998=27398)
(1997=27778) (1996=26854) (1995=27194)(1994=25416)
(1993=25416) (1992=25525) (1991=25631) (1990=25300)
(1989=24969) (1988=24760)(1987=24351) (1986=24533)
(1985=23827) (1984=23516) (1983=24490) (1982=25129)
(1981=24075) (1980=24668)(1979=23238) (1978=20396)
(1977=19750) (1976=17394) (1975=20205) (1974=18836)

```

```

(1973=17744)          (1972=17878)(1971=15230)          (1970=13933)
(1969=14420)          (1968=13001)          (1967=10793)          (1966=11170)
(1965=11378) (1964=12483) (1963=10552) (1962=8621) (1961=8044)
(1960=7118)          (1959=6665)          (1958=7265)          (1957=6746)
(1956=6287)(1955=5314) (1954=5007) (1953=4665) (1952=4355)
(1951=4156) (1950=4256) (1949=2388) (1948=2450) (1947=2326)
(1946=2045) (1945=2107) (1944=2208) (1943=2211) (1942=2332)
(1941=1996) (1940=1852) (1939=1890) (1938=1858) (1937=1590)
(1936=1550) (1935=1694) if uf==29;
#delimit cr

```

```

save pes_pnad2_2008, replace

```

```

*Geração da variável lei

```

```

gen lei=bith
replace lei=0 if lei<1957
replace lei=1 if lei>=1957

```

```

save pes_pnad2_2008, replace

```

```

*Declarando o conjunto de dados como sendo de amostra complexa

```

```

use pes_pnad2_2008, clear

```

```

svyset psu [pweight= pesos], strata(strat) vce(linearized)
singleunit(missing)

```

```

svydes, single

```

```

* ROTINA DE ALOCACAO DE ESTRATOS COM UM UNICO PSU EM ESTRATOS
COM MAIOR NUMERO

```

```

* DE OBSERVACOES UTILIZANDO O DO.FILE idonepsu - ANO DE 2003

```

```

use pes_pnad2_2008, clear
keep if uf < 21
gene novo_str = .
gene novo_psu = .
format novo_psu %12.0g
save acum.dta,replace

```

```

capture program drop prog1
program define prog1
use pes_pnad2_2008, clear
keep if uf == estado
gene novo_str = strat
gene novo_psu = psu
quietly {
save transf.dta,replace
use acum.dta, clear
append using transf.dta
save acum.dta,replace
}
end

```

```

capture program drop prog2
program define prog2
use pes_pnad2_2008, clear
keep if uf == estado
idonepsu , strata(strat) psu(psu) generate(novo_)
quietly {
save transf.dta, replace
use acum.dta, clear
append using transf.dta
save acum.dta,replace
}
end

*prog1 é feito para estados sem PSU único

scalar estado = .

foreach i in 28 {
scalar estado = `i'
prog1
}

foreach i in 21 22 23 24 25 26 27 29{
scalar estado = `i'
prog2
}

drop strat psu
rename novo_str strat
rename novo_psu psu

svyset psu [pweight= pesos], strata(strat) vce(linearized)
singleunit(missing)
svydes, single
save pes_pnad2_2008, replace

* ROTINA DE ALOCAÇÃO DE ESTRATOS COM UM ÚNICO PSU EM ESTRATOS
COM MAIOR NÚMERO
* DE OBSERVAÇÕES UTILIZANDO O DO.FILE idonepsu rend_pr - ANO
DE 2008

use pes_pnad2_2008, clear
keep if uf < 21
gene novo_str = .
gene novo_psu = .
format novo_psu %12.0g
save acum.dta,replace

capture program drop prog1
program define prog1
use pes_pnad2_2008, clear
keep if uf == estado

```

```

gene novo_str = strat
gene novo_psu = psu
quietly {
  save transf.dta,replace
  use acum.dta, clear
  append using transf.dta
  save acum.dta,replace
}
end

capture program drop prog2
program define prog2
  use pes_pnad2_2008, clear
  keep if uf == estado
  idonepsu rend_pr, strata(strat) psu(psu) generate(novo_)
  quietly {
    save transf.dta, replace
    use acum.dta, clear
    append using transf.dta
    save acum.dta,replace
  }
end

scalar estado = .

foreach i in 21 22 23 24 25 27 {
  scalar estado = `i'
  prog1
}

foreach i in 26 28 29{
  scalar estado = `i'
  prog2
}

drop strat psu
rename novo_str strat
rename novo_psu psu

svyset psu [pweight= pesopes], strata(strat) vce(linearized)
singleunit(missing)
svydes rend_pr, single
save pes_pnad2_2008, replace

* ROTINA DE ALOCAÇÃO DE ESTRATOS COM UM ÚNICO PSU EM ESTRATOS
COM MAIOR NÚMERO
* DE OBSERVAÇÕES UTILIZANDO O DO.FILE idonepsu rend_nt - ANO
DE 2008

use pes_pnad2_2008, clear
keep if uf < 21
gene novo_str = .
gene novo_psu = .
format novo_psu %12.0g
save acum.dta,replace

```

```

capture program drop prog1
program define prog1
use pes_pnad2_2008, clear
keep if uf == estado
gene novo_str = strat
gene novo_psu = psu
quietly {
save transf.dta, replace
use acum.dta, clear
append using transf.dta
save acum.dta, replace
}
end

capture program drop prog2
program define prog2
use pes_pnad2_2008, clear
keep if uf == estado
idonepsu rend_nt, strata(strat) psu(psu) generate(novo_)
quietly {
save transf.dta, replace
use acum.dta, clear
append using transf.dta
save acum.dta, replace
}
end

scalar estado = .

foreach i in 24 27 {
scalar estado = `i'
prog1
}

foreach i in 21 22 23 25 26 28 29{
scalar estado = `i'
prog2
}

drop strat psu
rename novo_str strat
rename novo_psu psu

svyset psu [pweight= pesopes], strata(strat) vce(linearized)
singleunit(missing)
svydes rend_nt, single
save pes_pnad2_2008, replace

*svydes, finalstage /*comando para detectar a existencia de
obs. unica em psu*/

log close
view funcaoNE2_2008.smcl
exit

```