JOÃO RICARDO FERREIRA DE LIMA

EFEITOS DA PLURIATIVIDADE E RENDAS NÃO-AGRÍCOLAS SOBRE A POBREZA E DESIGUALDADE RURAL NA REGIÃO NORDESTE

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
2008

Ficha catalográfica preparada pela Seção de Catalogação e Classificação da Biblioteca Central da UFV

T
L732e
2008

Lima, João Ricardo Ferreira de, 1974-
Efeitos da pluriatividade e rendas não-agrícolas sobre a pobreza e desigualdade rural na região Nordeste \ João Ricardo Ferreira de Lima – Viçosa, MG, 2008.
xii, 157f.: il. (algumas col.) ; 29cm.

Inclui anexos.
Orientador: João Eustáquio de Lima.
Tese (doutorado) - Universidade Federal de Viçosa.
Referências bibliográficas: f. 117-125

1. Brasil, Nordeste - Condições econômicas. 2. Pobreza. 3. Renda - Distribuição. 4. Agricultura familiar. 5. Agricultura em tempo parcial. 6. Modelos econométricos. 7. Amostragem (Estatística). I. Universidade Federal de Viçosa. II.Título.

CDD 22.ed. 330.9813

JOÃO RICARDO FERREIRA DE LIMA

EFEITOS DA PLURIATIVIDADE E RENDAS NÃO-AGRÍCOLAS SOBRE A POBREZA E DESIGUALDADE RURAL NA REGIÃO NORDESTE

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: 07 de novembro de 2008.

Marília Fernandes Maciel Gomes

Aldenôr Gomes da Silva

Henrique Dantas Neder

Marcelo José Braga
(Co-orientador)

João Eustáquio de Lima
(Orientador)

Dedico a minha Avó Edwiges (Dudu) (in memorian), certamente a pessoa que mais torceu para que eu conseguisse fazer este Curso. Dedico também aos dois grandes amores da minha vida: minha Esposa Paloma e meu Filho João Gabriel.

AGRADECIMENTOS

À Deus por tudo que tenho e pela alegria de ter conseguido alcançar mais um objetivo na vida.

Mais do que agradecimentos, não existem palavras para descrever perfeitamente o que sinto pela minha Esposa Paloma, por tudo que ela fez por nossa família neste período em que cursei o doutoramento. Se consegui chegar ao final, foi por esforço e pelo suporte fundamental dela nos inúmeros momentos em que eu pensei que não teria forças para chegar ao final: uma palavra de estímulo, um carinho, um sorriso, um café gostoso para quem ia enfrentar a madrugada trabalhando na tese, a companhia, uma refeição para aquele que muitas vezes nem se lembrava da última vez que tinha saído da frente do computador, fizeram toda a diferença.

Um Filho definitivamente muda a forma com que o Pai enxerga o mundo. Tudo que fiz desde o nascimento do João Gabriel e ainda farei é tentando dar um melhor futuro para ele. Espero que um dia o mesmo entenda isso e me perdoe pelas, talvez, centenas de vezes em que me pediu atenção, para ir brincar com ele, para jogar bola, brincarmos de carrinho, soltar pipa na rua e eu não pude atender por estar trabalhando na tese ou estudando para as disciplinas ou preparando as monitorias de econometria. É difícil demais negar um pedido destes para uma pessoa em que eu daria minha vida sem pensar duas vezes. Hoje, felizmente, eu consigo terminar e não terei mais que escutar a pergunta que se tornou recorrente nos últimos meses: “Pai, todo dia você tem que trabalhar na sua tese e ela não acaba nunca?”

Outra pessoa mais do que importante é a minha avó Edwiges, uma das mais felizes no dia em que recebi a notícia da aprovação no curso em Viçosa e que de forma

inexplicável não estará no portão de sua casa me esperando com os braços abertos no dia em que retornarmos. Espero que onde a senhora esteja tenha sentido orgulho do neto.

Agradeço muito aos meus pais João Antonio e Waldinete, exemplos de honestidade, dignidade e força de vontade. Foram praticamente quatro anos em que estivemos distantes fisicamente, mas sempre levo vocês comigo nas lembranças e no coração. Da mesma forma é o sentimento por Daniele e Ariane, minhas irmãs e por João Victor e Aisy, sobrinhos. Apesar da distância e do contato mais reduzido neste período, amo vocês demais. Ao meu sogro Sr. Roberto e minha sogra D. Rita, que acreditaram e apoiaram muito neste período em Viçosa.

À Djail e Silvanda, que me apoiaram desde o dia em que decidi fazer a seleção de doutorado em Viçosa. Esses dois eu nem sei como vou poder retribuir o muito que fizeram por mim enquanto estive na UFV. Espero não decepcioná-los jamais.

À Universidade Federal de Viçosa, que pelo Departamento de Economia Rural me proporcionou todas as condições para concluir o curso; à Universidade Federal da Paraíba, pelo Departamento de Ciências Fundamentais e Sociais do Centro de Ciências Agrárias, pela liberação neste período para me qualificar e à CAPES pela concessão da bolsa de estudos.

Ao professor João Eustáquio, a que tenho um profundo respeito e admiração, por tudo que me ensinou e pela orientação valiosa na tese. O senhor é para mim um exemplo de pessoa íntegra que busco me espelhar. Poder trabalhar com o senhor é algo de que tenho muito orgulho e satisfação.

Aos professores que participaram da defesa: Henrique Neder, Aldenôr Gomes, Fátima Marília e Marcelo Braga pela leitura criteriosa que fizeram do trabalho e que em muito contribuíram com sugestões para o seu melhoramento.

Ao prof. Henrique Neder, pela paciência e receptividade que teve comigo no período em que estive em Uberlândia para aprender a trabalhar com os dados da PNAD, além das informações e sugestões fundamentais dadas e que em muito contribuíram para que eu pudesse concluir a tese.

Ao prof. Aldenôr, que me acompanha desde o mestrado, que prontamente se dispôs a vir à Viçosa participar da defesa, o que me deixou muito feliz. Desejo também ao senhor muita alegria e paz na nova fase da sua vida com a Edna.

No DER é até difícil colocar os nomes de todos aqueles que desejo agradecer. Todos foram maravilhosos comigo e são pessoas que levo comigo no coração. Porém, não tem como não agradecer em especial aos professores Marcelo Braga e Erly Teixeira, por acreditarem em mim, terem apoiado e sempre me recebido muito bem quando eu precisei.

Assim como é impossível não agradecer de forma especial à Carminha e Dona Cida. Para estas duas, igualmente não tenho palavras para explicar o tamanho da gratidão. Sempre tão prestativas, me ajudaram demais sempre que precisei. Vai ser dureza ir embora sem poder levá-las.

Uma das coisas que mais me orgulha ao final desta etapa é a quantidade de amigos de verdade que fizemos. Um casal especial e que vai conosco no coração é Aureliano e Valéria. Obrigado pela força que deram, pela confiança e pela amizade. Esperamos que continuem para sempre tão felizes e apaixonados e que curtam bastante a chegada do novo membro da família.

Outro casal especial é Carlos e Cláudia Piacenti. Compartilhar da amizade de vocês é algo muito gratificante. Para mim e minha família vocês são mais do que amigos, são exemplos mesmo de família feliz.

Ao trio parada dura de paraibanos: Adriano Paixão, Jean e Herbert, pelas boas risadas e ótima convivência. Vocês são mais do que amigos, são irmãos mesmo para mim.

Com carinho, agradecimentos ao Joelsio e Lisandra, Renata e Daniel, Elaine e Guilherme, Marquinhos e Rita e Mychelle e Raul.

Aos demais colegas da pós-graduação: Tozé, Rubicleis, Cida, Marco Aurélio, Adelson, Eduardo, Patrick, Cristiane, Bia, Sara, Aline, Norberto, Alessandro "Bombom", Rosângela, Giovanna Rossi, Jair, Eliane, Lúcia, Aracy, Denis, Luisa e Vanessa.

Ao pessoal radical do sk8 e da bike: Zé Pedro (Mineiro BMX) e Eliseu (Colisão), pelos momentos de descontração.

BIOGRAFIA

JOÃO RICARDO FERREIRA DE LIMA, filho de Waldinete Ferreira Xavier de Lima e João Antônio de Lima, nasceu em São Paulo, SP, em 24 de abril de 1974.

Em setembro de 1999, bacharelou-se em Ciências Econômicas pela Universidade Federal da Paraíba, em Campina Grande, PB.

Em março de 2000, ingressou no Curso de Mestrado em Economia Rural da Universidade Federal da Paraíba, em Campina Grande, PB, tendo defendido a dissertação em abril de 2002.

Entre março de 2002 e abril de 2003 foi professor Assistente no Departamento de Economia da FACEM/UERN. Em maio de 2003 tomou posse no cargo de professor Assistente no Departamento de Ciências Fundamentais e Sociais do Centro de Ciências Agrárias da Universidade Federal da Paraíba, em Areia, PB.

Ingressou no programa de Doutorado em Economia Aplicada da Universidade Federal de Viçosa (UFV), MG, em março de 2005, concluindo o mesmo, com muito orgulho, em novembro de 2008.

SUMÁRIO

Página

RESUMO.....ix
ABSTRACT.....xi
1. INTRODUÇÃO.....1
1.1. Considerações Iniciais.....1
1.2. Pluriatividade, Pobreza e Seca: conceitos fundamentais.....6
1.3. O problema e sua importância.....12
1.4. Hipóteses.....20
1.5. Objetivos.....20
1.5.1. Geral.....20
1.5.2 Específicos.....20
2. REFERÊNCIAL TEÓRICO.....22
2.1. A abordagem da utilidade conjunta.....23
2.1.1. O caso das famílias exclusivamente agrícolas.....26
2.1.2. O caso das famílias exclusivamente não-agrícolas.....27
2.1.3. O caso das famílias pluriativas.....27
2.1.4. O caso das famílias não-ocupadas.....28
2.2. Crescimento econômico, Pobreza e Distribuição de renda.....31
3. METODOLOGIA.....36
3.1. Determinantes da escolha entre alternativas de ocupação.....37
3.2. Efeitos da renda não-agrícola sobre a pobreza.....41
3.3. Efeitos da renda não-agrícola sobre a concentração.....47
3.4. Fontes de dados.....52
3.4.1. Características básicas do plano amostral da Pnad.....53

3.4.2. Outros conceitos e procedimentos utilizados.....58
4. RESULTADOS E DISCUSSÕES.....61
4.1. Estatísticas descritivas das variáveis.....61
4.2. Determinantes da escolha de alternativas de ocupação nos anos de 2003 e 2005.....65
4.3. Efeitos da renda não-agrícola sobre a pobreza e a concentração.....79
4.3.1. Simulação considerando todas as famílias agrícolas ou pluriativas.....79
4.3.2. Análise das simulações por tipos de famílias.....83
4.3.3. Simulação considerando apenas as famílias nos estados do Maranhão, Piauí e Rio Grande do Norte.....99
4.4. Elasticidades da pobreza em relação ao crescimento da renda familiar agrícola observada e simulada e ao índice de Gini.....108
5. RESUMO E CONCLUSÕES.....112
REFERÊNCIAS BILIOGRÁFICAS.....117
ANEXOS.....126

RESUMO

LIMA, João Ricardo F., D.Sc., Universidade Federal de Viçosa, Novembro de 2008. **Efeitos da Pluriatividade e Rendas Não-Agrícolas Sobre a Pobreza e Desigualdade Rural na Região Nordeste**. Orientador: João Eustáquio de Lima. Co-orientadores: Marcelo José Braga e Wilson da Cruz Vieira.

Na região Nordeste existe uma grande quantidade de famílias rurais com pelo menos um membro ocupado em atividades não-agrícolas. Porém, é também onde se concentra o maior percentual de famílias pobres. A pobreza se reflete nos indicadores sociais, como a maior taxa de mortalidade infantil entre as regiões brasileiras, menor esperança de vida ao nascer e taxa de analfabetismo mais elevada. O desenvolvimento do meio rural nordestino é, então, parte fundamental de uma estratégia de desenvolvimento nacional. Dado que diversas agências de pesquisa e financiamento internacionais têm considerado que o estímulo ao não-agrícola pode contribuir para melhorar as condições de vida das famílias rurais, este trabalho busca analisar os efeitos da pluriatividade e rendas não-agrícolas sobre a pobreza e desigualdade no Nordeste brasileiro. O referencial teórico está relacionado à oferta de mão de obra rural, focando a possibilidade dos membros da família se alocar em múltiplas fontes de ocupação. Segundo esta abordagem, a família compara as opções de trabalho e aloca seu tempo total disponível de forma a maximizar sua função de utilidade. É utilizado o modelo de seleção amostral com logit multinomial tanto para analisar os determinantes da escolha da família entre os tipos de ocupação quanto para estimar a renda média considerando a possibilidade de viés de seleção. A partir das rendas estimadas são feitas simulações visando prever qual a renda da família agrícola se fosse pluriativa ou não-agrícola; a renda da família pluriativa, caso se tornasse exclusivamente agrícola ou não-agrícola; finalmente, a renda da família não-agrícola se passasse para agrícola ou pluriativa.

Com as rendas observadas e estimadas são calculados os índices de pobreza FGT (Foster-Greer-Thorbecke), o índice de concentração de Gini e as elasticidades crescimento-renda e Gini da pobreza. A fonte de dados é a Pnad de 2003 e 2005, visando captar diferenças nos resultados considerando um ano sem e outro com chuvas regulares. Com relação aos determinantes de ocupação, independente da condição climática, anos de estudo, número de componentes da família, ser do tipo conta-própria ou empregados e residir no estado do Piauí eleva a chance da família ser pluriativa. Residir no rural mais distante do urbano, possuir uma razão de dependência mais elevada e residir em Alagoas, Sergipe ou na Bahia, reduz esta chance. A primeira simulação realizada com todas as famílias agrícolas ou todas pluriativas indicou que o não-agrícola é importante para redução da pobreza e da concentração. Em simulação posterior, considerando os diferentes tipos de família, os resultados indicam que o não-agrícola é importante para redução da pobreza. Com relação à concentração, a desigualdade não se reduz se a família agrícola passar para pluriativa. A parcela agrícola da renda é tão concentrada que o incremento com a parte não-agrícola não tem efeito na redução das disparidades. Desagregando as informações para os estados do Maranhão, Piauí e Rio Grande do Norte, que são mais "propensos" às atividades não-agrícolas, os resultados confirmam o efeito positivo sobre a redução da pobreza e demonstram que as rendas não-agrícolas também podem contribuir para redução da concentração. O efeito não é tão forte quanto na redução da pobreza, mas pode ser utilizado em conjunto com instrumentos específicos que visem à redução da concentração da renda na região. A análise das elasticidades demonstra que o crescimento da renda das famílias pluriativas tem maior efeito na redução da pobreza, relativo às famílias agrícolas, sendo a redução da desigualdade um fator bastante relevante para "puxar" as famílias para a parte de cima da linha da pobreza. Conclui-se que a pluriatividade e as rendas não-agrícolas são importantes para reduzir a pobreza e a concentração no rural nordestino, confirmando as hipóteses testadas na pesquisa.

ABSTRACT

LIMA, João Ricardo F., D.Sc., Universidade Federal de Viçosa, November, 2008. **Effects of Pluriactivity and Non-Farming Income on Poverty and Inequality in the Brazilian Northeast**. Adviser: João Eustáquio de Lima. Co-advisers: Marcelo José Braga and Wilson da Cruz Vieira.

Many rural family households in the Brazilian Northeast have at least a member involved in non-farming activities. However, this region also concentrates the highest percentage of poor families. Poverty is reflected by social indicators such as higher infant mortality rate between the Brazilian regions, lower life expectancy at birth and higher illiteracy rate. Thus, development of the northeastern rural conditions must be a fundamental part of a national development strategy. Considering that several research and funding agencies have regarded stimulus to non-farming activities as a tool to improve the living conditions of rural families, this work aims to analyze the effects of pluriactivity and non-farming income on poverty and inequality in the Brazilian Northeast. The theoretical referential is related to rural labor offer, focusing on the likelihood of the family household members to apply to multiple sources of employment. Based on this approach, the family compares the work options, allocating their total available time so as to maximize their utility function. The selection bias correction based on the multinomial logit model was applied to analyze the determinants of the family's choices from the types of jobs available as well as to estimate the mean income, considering the possibility of sample selection bias. Based on the estimated incomes, simulations are made aiming at predicting the farming family's income in the case of a pluriactive or non-farming family; the pluriactive family's income in case it became exclusively agricultural or non-farming; and, finally, the non-farming family's income, if it became a agricultural or pluriactive

family. Based on the observed and estimated incomes, the FGT (Foster-Greer-Thorbecke) poverty index, the Gini income concentration index and the growth-income and Gini poverty elasticities are calculated. The data source is Pnad of 2003 and 2005, aiming at capturing differences in the results, considering a year without and another with regular rainfall. Regardless of the climatic conditions, job determinants, such as years of schooling, number of family members, autonomous or employed type of worker, and being a resident in the state of Piauí increase the chance of the family household being pluriactive. Residing in the rural area farther from the urban area, having a higher dependence and being a resident in the states of Alagoas, Sergipe or Bahia, decrease this chance. The first simulation carried out considering all the family households as farming or as pluriactive indicated that the non-farming activity is important to reduce poverty and income concentration. The results obtained in a posterior simulation, considering the different types of family, showed that non-farming job is important to reduce poverty. With regard to income concentration, inequality does not decrease if the agricultural family becomes pluriactive. The farming part of the income is so concentrated that the non-farming increase has no effect on the reduction of the disparities. As for the states of Maranhão, Piauí and Rio Grande do Norte, which are more "inclined" to non-farming activities, the results confirm the positive effect of this tendency on poverty reduction and show that the non-farming incomes can also contribute to reduce income concentration. The effect is not as strong as on poverty reduction, but can be used in combination with specific tools aiming at reducing income concentration in the region. The analysis of elasticities shows that the income growth of the pluriactive family households has a greater effect on poverty reduction, compared to the agricultural family households, with inequality reduction being a rather relevant factor in helping families move up the poverty line. The conclusion is that the pluriactivity and the non-agricultural incomes are important to reduce the poverty and the concentration in the rural of Brazilian Northeast, confirming the hypotheses tested in the research.

1. INTRODUÇÃO

1.1 Considerações Iniciais

O termo pluriatividade surge no cenário internacional na década de 70, em referência à combinação do exercício de atividades agrícolas e não-agrícolas[1] pelos membros de uma mesma família. O debate no período era distinguir os "verdadeiros agricultores", ou seja, os exclusivamente dedicados às atividades agrícolas, daqueles que estavam engajados em outros setores da economia. O desejo do movimento sindical era de que apenas os primeiros pudessem se beneficiar dos recursos da política agrícola. As famílias que iam se tornando pluriativas buscavam, porém, simplesmente diversificar suas fontes de renda para conseguir melhorar a condição de vida e se manter no local de origem. Atualmente, as próprias organizações sindicais consideram a pluriatividade um fator positivo para o desenvolvimento rural (CARNEIRO, 2005).

No Brasil, as pesquisas relacionadas ao tema aparecem na década de 90, principalmente nos trabalhos do Projeto Rurbano do Instituto de Economia da Unicamp. Para Graziano da Silva (1999), ocorre um processo semelhante ao dos países desenvolvidos, em que, dado o elevado nível tecnológico dos meios de produção, o

[1] Na literatura internacional, são utilizados os conceitos dicotômicos de "farm work" e "off farm work", sempre relacionados ao trabalho dentro e fora da propriedade, independentemente de ser agrícola ou não-agrícola. Neste trabalho, é feita uma adaptação e se trabalha com os conceitos de agrícola e não-agrícola, mesmo que a atividade não-agrícola seja exercida dentro do domicílio, como, por exemplo, fábrica caseira de bolos ou doces, artesanatos e redes. Para o IBGE, consideram-se atividades agrícolas aquelas relacionadas à agricultura, silvicultura, pecuária, extração vegetal, pesca e piscicultura. Todo beneficiamento de produtos agropecuários, realizado em estabelecimentos rurais rudimentares e não-licenciados é considerado "indústria rural" dentro das atividades agropecuárias.

agricultor consegue ter tempo livre (se torna um *part-time farm*) para se dedicar a outras atividades não-agrícolas. Diversos estudos posteriores, de Alves (2002), Nascimento (2005), Carneiro (2005) e Lima (2001, 2006), enfatizam que a pluriatividade na agricultura brasileira é, na realidade, uma estratégia dos agricultores para buscar aumentar sua renda familiar. Isto é muito importante nas regiões menos favorecidas, como a Nordeste, em que a maior parte dos agricultores não se modernizou, não está integrada e ainda convive com maiores adversidades climáticas. Nestas condições, os riscos de sobreviver exclusivamente de atividades agrícolas aumentam e as atividades/rendas não-agrícolas são importantes para dar estabilidade na renda das famílias (LIMA, 2002).

A Tabela 1 trata da distribuição do total das famílias brasileiras e das residentes na região Nordeste, que moram no meio rural, segundo o tipo de atividade ocupada[2], em 2003. Percebe-se que, no Brasil, a maior parte, mais de 3 milhões de famílias, exerce exclusivamente atividades agrícolas (52,90%). Porém, mais de 2 milhões de famílias (pluriativas + não-agrícolas) rurais possuem pelo menos um membro ocupado em atividades não-agrícolas[3]. Cerca de 22% do total são pluriativas, ou seja, combinam atividades agrícolas e não-agrícolas e 14,50% são exclusivamente não-agrícolas.

Com relação ao Nordeste, a Tabela 1 demonstra que a maior parte das famílias está ocupada exclusivamente em atividades agrícolas (52%), mas o percentual de famílias pluriativas relativamente ao total da região é de cerca de 26%, valor maior do que em nível nacional (21,80%). Aproximadamente 10% das famílias rurais nordestinas estão ocupadas exclusivamente em atividades não-agrícolas e mais de 11% estão desocupadas. A sobrevivência deste elevado percentual de desocupados

[2] O conceito de ocupação é o do IBGE, ou seja, "ocupadas" são as pessoas de 10 anos ou mais de idade que realizavam algum trabalho (remunerado ou não) durante pelo menos 1 hora na semana de referência (período de coleta de informações (DEL GROSSI e GRAZIANO DA SILVA, 2002).

[3] Dentre os 10 ramos de atividades não-agrícolas exercidos pela população rural ocupada e residente no campo mais importantes, em apenas 5 estão concentradas cerca de 84% do total das Ocupações Rurais Não-Agrícolas (Orna). Destacam-se os ramos de Prestação de Serviços, Indústria de Transformação, Indústria da Construção, Serviços Sociais e Comércio de Mercadorias. Com relação aos setores de atividades da PEA restrita rural ocupada em Orna, predominam os Empregados, Ajudantes de Serviços Gerais, Pedreiros somados com os Ajudantes, Serviços Conta-Própria, Balconistas e Atendentes, Motoristas, Professoras Primárias e faxineiros. A grande maioria das profissões não exige grandes qualificações da mão-de-obra (especializações e formação escolar).

residindo no meio rural, possivelmente, é explicada pelas rendas não-agrícolas advindas, principalmente, de transferências públicas, como as aposentadorias e o bolsa-família, além de transferências privadas.

Tabela 1. Distribuição das famílias, Brasil e Nordeste, com domicílio nas áreas rurais não-metropolitanas segundo o tipo de atividade ocupada, 2003

Tipos de Famílias	BRASIL		NORDESTE	
	Total (em 1.000 famílias)	%	Total (em 1.000 famílias)	%
Agrícola	3.112	52,90	1.538	52,00
Pluriativa	1.281	21,80	774	26,20
Não-agrícola	854	14,50	303	10,30
Não-ocupada	639	10,80	340	11,50
Total de Famílias	5.886	100,00	2.955	100,00

Fonte: Tabulações especiais da Pnad, Projeto Rurbano, junho de 2003. In: Schneider, 2005.

A Tabela 2 compara as regiões Nordeste e Sul, com relação ao percentual de famílias pobres. Estas duas regiões são importantes por alocarem cerca de 80% da agricultura familiar brasileira[4] (NASCIMENTO, 2005; NASCIMENTO e CARDOZO, 2007). A linha da pobreza[5] foi definida por Nascimento (2005) em termos de renda média familiar *per capita* de ½ salário mínimo (s.m.) [6], em R$ de setembro de 1999.

Na região Nordeste, com exceção das famílias do tipo conta própria[7] não-agrícola, em todos os demais predominam as famílias pobres sobre o total de famílias. A proporção de pobres nas famílias de conta própria agrícola é de cerca de 60%. Na região Sul, os percentuais de famílias pobres sobre o total indicam que a maior proporção é de "não-pobres", ou seja, percebem renda média familiar superior a ½ s.m.. Os maiores percentuais de proporção de famílias pobres são relacionados às famílias exclusivamente agrícolas (Tabela 2).

[4] O conceito de agricultura familiar considerado neste trabalho é "uma unidade de produção agrícola onde propriedade e trabalho estão intimamente ligados à família" (LAMARCHE, 1993, p. 15).
[5] A definição de linha de pobreza é mais bem trabalhada na segunda parte do trabalho.
[6] O salário mínimo no período era de R$ 136,00.
[7] Neste trabalho, as famílias foram classificadas da seguinte forma: se um dos membros da família declarou ser empregador, a família foi classificada como tal independentemente da posição na ocupação dos demais membros. Na falta de empregador, a família foi classificada como de conta-própria se um dos membros ativos se declarou como tal, independentemente da posição na ocupação dos outros membros. Não havendo na família membros que se declarassem como empregadores ou conta-própria, a família foi classificada como de empregados, se pelo menos um membro declarou exercer uma atividade agrícola ou não-agrícola na semana de referência. Se ninguém estivesse ocupado, a família seria de "não-ocupados" (NASCIMENTO, 2005)

Tabela 2. Número de famílias pobres e percentual sobre o total das famílias de conta-própria nas regiões Nordeste e Sul, em 1999

Local de Domicílio/Tipo de Família	Nordeste		Sul	
	Mil fam. pobres	% fam. pobres s/total	Mil fam. pobres	% fam. pobres s/total
Não metropolitano				
Rural Agropecuário				
Conta Própria	**1.071**	**59,60**	**178**	**26,60**
Agrícola	616	60,70	151	32,40
Pluriativo	392	60,10	18	13,10
Não-agrícola	63	48,40	9	13,70

Fonte: Pnad – Tabulações Especiais do Projeto Rurbano. In: Nascimento, 2005.

A pobreza rural na região Nordeste é um problema crônico e antigo. Do início dos anos 70 até o final dos anos 90, mudanças importantes ocorreram. No período entre as décadas de 1970 e 1980, houve aumento na desigualdade entre as regiões, com prejuízo para o Nordeste, menos desenvolvido. Neste período, a região aumentou sua participação na pobreza rural nacional de 46% para 61%. Entre 1981-1990, embora a região permaneça com uma elevada proporção de pobres, apresenta o melhor desempenho regional na redução da pobreza nacional. Apesar de não conseguir alterar significativamente sua posição no contexto nacional, principalmente após a metade da década de 90, o Nordeste passa a reduzir tanto a proporção de pobres quanto a intensidade da pobreza e a severidade da pobreza[8], devido à estabilidade econômica com o plano real e o impacto da previdência social nas áreas rurais (ROCHA, 2006).

Infelizmente, a redução da pobreza intra-regional não pode ser comemorada devido à comparação inter-regional. Segundo Rocha (2006), após analisar o indicador de pobreza "proporção de pobres",

> *a participação do Nordeste na população brasileira declina de 30,4%, em 1970, para 28,9%, em 1999. Nesse sentido, o progresso realizado na redução das desigualdades regionais no longo prazo foi nulo quando se considera esse indicador de pobreza. É importante destacar ainda o forte crescimento da participação do Nordeste na pobreza rural brasileira, que passa de 45,8%, em 1970, para 68,17%,*

[8] Estes três índices, proporção de pobres, intensidade da pobreza e severidade da pobreza, compõem o que se denomina de índices FGT (Foster-Greer-Thorbecke) e são demonstrados no tópico 3.2.

em 1999. Assim, a pobreza rural nordestina se apresenta, em 1999, como o núcleo 'duro' da pobreza no Brasil (ROCHA, 2006, p. 89).

A pobreza por insuficiência de renda é acompanhada por outros indicadores que também representam e destacam outras dimensões importantes da pobreza, que é um fenômeno multidimensional. Como pode ser observada na Tabela 3, para o ano de 2005, a taxa bruta de mortalidade[9] no Nordeste é de 6,90%, enquanto a nacional é de 6,30%; a taxa de mortalidade infantil[10] é de 39,50%, valor bem superior à média nacional, de 26,60%. A esperança de vida ao nascer é inferior à média nacional, 68,6 anos contra 71,7; e as taxas mais elevadas de analfabetismo se encontram na região, cerca de 22,40%.

Tabela 3. Comparação entre indicadores sociais em nível de Brasil e da região Nordeste, 2004

	Taxa bruta de Mortalidade (%)	Taxa de Mortalidade Infantil (%)	Esperança de vida ao nascer	Taxa de analfabetismo (%)	Taxa de analfabetos funcionais (%)		Serviço Público de Saneamento (%)
					Total	Rural	
Brasil	6,30	26,60	71,7	11,40	24,40	47,50	60,30
Nordeste	6,90	39,50	68,6	22,40	37,60	59,40	34,90

Fonte: IBGE (2006).

Sobre os analfabetos funcionais, ou seja, as pessoas com 15 anos ou mais de idade com menos de quatro anos completos de estudo, o percentual na região Nordeste é de 37,60%. No meio rural, este percentual sobe para 59,40%. Quanto à condição do domicílio, apenas 34,90% dos domicílios urbanos possuem serviço público de saneamento, enquanto a média nacional é de 60,30%. Apenas para comparação das disparidades regionais no Brasil, na região Sudeste esse índice chega a 82,80%.

Qualquer estratégia visando ao desenvolvimento nacional precisa equacionar o problema do subdesenvolvimento do meio rural nordestino. É necessário criar diversos tipos de políticas que consigam elevar a renda das famílias e, assim, "puxá-las" para a

[9] Quociente entre o número de óbitos ocorridos durante um ano civil e a população total ao meio do ano civil. Representa a freqüência com que ocorrem os óbitos em uma população (IBGE, 2006).
[10] Freqüência com que ocorrem os óbitos infantis (menores de um ano de idade) em uma população em relação ao número de nascidos vivos em determinado ano civil (IBGE, 2006).

parte superior da linha da pobreza, possibilitando a melhora nos indicadores sociais regionais e aproximando estas famílias dos níveis baixos de pobreza encontrados na região Sul do Brasil.

1.2. Pluriatividade, Pobreza e Seca: conceitos fundamentais

Esta pesquisa parte do pressuposto de que rural não é sinônimo de agrícola. A grande maioria dos autores, na história do pensamento econômico, desde os fisiocratas[11] até os teóricos denominados de "orientação produtivista" [12], ao analisar a importância do meio rural para o desenvolvimento nacional, confunde desenvolvimento agrícola com desenvolvimento rural. Outro pressuposto considerado é que o desenvolvimento agrícola não conduz, por si só, ao desenvolvimento rural.

O desenvolvimento agrícola, segundo Navarro (2001), está relacionado às condições/características da produção agropecuária, se restringe ao sentido produtivo, identificando suas tendências em um período de tempo dado. Refere-se a aumento na produtividade, área plantada, redução de custos médios, economias de escala, por exemplo. Por outro lado, desenvolvimento rural é um processo resultante de ações articuladas, com o objetivo de alterar a situação socioeconômica e ambiental da população rural, melhorando sua renda[13], qualidade de vida e bem-estar (SCHNEIDER, 2003).

Esta noção se fortalece recentemente pela valorização que ocorre dos espaços rurais. Não faz muito tempo que rural era considerado sinônimo de atraso e urbano de

[11] Para Turgot, Petit e, principalmente, Quesnay, a agricultura possui um papel fundamental no desenvolvimento, contrariando o pensamento clássico.

[12] São assim denominados os modelos de Schultz e Hayami-Ruttan, que possuem a característica de enfatizar a relevância da mudança tecnológica e do aumento do fator capital em relação ao trabalho como forma de elevar a produtividade e a produção agropecuária. Estes modelos inspiram o que ficou conhecido como "Revolução Verde". Consideram que se pode transferir diretamente a tecnologia – por meio de "pacotes tecnológicos" - e os modelos de extensão agropecuária dos países desenvolvidos (principalmente o norte-americano) para os em desenvolvimento. Basicamente se busca elevar a produtividade com a utilização massiva de insumos químicos, como fertilizantes e defensivos, além da mecanização dos processos de produção e colheita.

[13] A questão central, principalmente na região Nordeste, continua a mesma, busca por renda. A real possibilidade de reorganizar a economia da região e melhorar a situação de vida da população rural, passa por "dar maior estabilidade à renda real do trabalhador, e não à oferta de alimentos. Se durante a seca o trabalhador compra alimentos no mercado é porque o governo, embora artificialmente, cria o poder de compra necessário. Por mais estável que seja a oferta de alimentos, se a seca continua a reduzir drasticamente a renda do trabalhador, seus efeitos econômicos continuarão a ser os mesmos" (GTDN, 1978, p.72).

modernidade. No meio rural ganham importância muitas atividades não-agrícolas, fundamentais para a manutenção da família na propriedade, por proporcionarem elevação e relativa estabilidade da renda. As famílias rurais estão cada vez mais se tornando pluriativas ou exclusivamente não-agrícolas. Porém, vale ressaltar que isto não significa o fim da agricultura.

O conceito de pluriatividade pode ser definido como a combinação de atividades – por indivíduos e famílias – em diferentes setores. No entanto, neste caso, restringe-se o "conceito" para que um desses setores seja a agricultura (KAGEYAMA, 1998). Utiliza-se este conceito levando também em consideração a existência de pluriatividade quando pelo menos um dos membros da família exerce atividade agrícola, tanto na ocupação principal quanto na secundária, enquanto os demais exercem atividades não-agrícolas.

A pluriatividade pode se constituir numa estratégia de fortalecimento da economia familiar, pois aumenta a renda da família e possibilita a expansão da propriedade e a aquisição de ativos produtivos. O estabelecimento da pluriatividade depende de ser gerado excesso de capacidade produtiva da força de trabalho dentro da unidade produtiva familiar: seja por desenvolvimento tecnológico, seja por escassez de fator de produção terra, por exemplo.

A pluriatividade representa uma alternativa de diversificação produtiva, ocupacional e de geração de fontes de renda para as famílias de agricultores, tanto para aqueles que se encontram em regiões de agricultura moderna quanto os que residem em regiões mais desfavorecidas. Nestas regiões, a agricultura praticamente é inviabilizada por razões estruturais, tornando os agricultores extremamente vulneráveis do ponto de vista social, econômico e ambiental.

Um ponto a ser analisado, fundamental no estudo da pluriatividade, é o mercado de trabalho. A manutenção de vínculo com a propriedade rural implica que se trabalhe em lugares não muito distantes, principalmente se a pessoa continua morando na propriedade. Essas atividades podem estar especificamente no rural ou em áreas urbanas próximas. O que importa é poderem ser exercidas por membros das famílias

rurais pluriativas. Isto vai depender das características tanto econômicas, quanto sociais e culturais das "economias locais" [14].

Uma família rural pluriativa, que tenha membros em idade de trabalhar e que, além disso, estejam capacitados (qualificados) para exercer algumas atividades não-agrícolas, não é suficiente para considerar que um processo de pluriatividade será desencadeado. Isto porque se faz necessário que a localidade crie um mercado de trabalho apto a absorver esta força de trabalho disponível.

Segundo Mattei (1999), podem-se elencar os fatores condicionantes da pluriatividade da seguinte forma:

1. natureza econômica: tamanho das unidades de exploração, renda, grau de inserção nas economias locais;
2. natureza social: nível de escolaridade das famílias, sexo, ciclo familiar, desejo de obter melhores condições de vida;
3. natureza pessoal: melhorar/aumentar padrões de consumo, membros que buscam trabalhar em outra atividade, aprender novas profissões;
4. natureza contextual: desenvolvimento da economia local, políticas de desenvolvimento rural que favoreçam o crescimento de Orna (ocupação rural não agrícola) no rural; e
5. natureza de lazer: sítios, chácaras de final de semana, turismo rural, áreas de preservação ambiental, pesque-pague, rodeios (no Nordeste, as vaquejadas).

Dado que o conceito de pobreza é diversas vezes utilizado sem maiores detalhamentos, assim como sua relação com a concentração de renda, faz-se necessário tecer considerações mais apuradas sobre este complexo fenômeno. O tipo de pobreza que se analisa nesta pesquisa é a absoluta, ou seja, o não-atendimento das necessidades mínimas para manutenção da família (sobrevivência física). Na verdade, este é o conceito operacionalmente relevante no Brasil, em função do o grande número

[14] Segundo Saraceno (*apud* KAGEYAMA, 1998, p. 542), "o enfoque da economia local é baseado num espaço mais precisamente definido, não necessariamente coincidente com uma região administrativa. Idealmente, ele seria delimitado pela maior parte contígua em que tenha lugar o fenômeno da comutação e/ou empresas troquem quantidades relevantes de bens e serviços segundo um padrão consistente".

de famílias que não conseguem fazer três refeições diárias. Em economias monetizadas e com incidência de pobreza absoluta, é comum atrelar a renda como critério de pobreza. Inicialmente, é estabelecido um valor monetário relacionado ao custo do atendimento das necessidades médias das pessoas dentro de um determinado contexto, normalmente denominado de linha da pobreza. Com isto, a família é considerada pobre se sua renda *per capita* for inferior ou igual a uma determinada linha de pobreza. A definição de pobreza, então, está relacionada à insuficiência de renda, considerando esta como *proxy* de bem-estar (ROCHA, 2006).

A definição da linha da pobreza é extremamente importante para definir o tamanho da pobreza. Um país pobre não pode ter a mesma linha de pobreza de um país desenvolvido, já que nos países com maior PIB *per capita* não se tem mais a preocupação com o atendimento das necessidades de sobrevivência. A preocupação está mais relacionada com a busca de igualdade de padrões de consumo. Dentro de um país heterogêneo regionalmente como o Brasil, inclusive em questões culturais, com diferenças de produtos típicos de cada região que fazem parte da cesta de consumo, o ideal é construir diferentes linhas de pobreza adaptadas às distintas realidades[15].

Nesta pesquisa, a linha de pobreza definida é classificada como arbitrária. É o caso de definir a linha via um múltiplo do salário mínimo ou a forma do Banco Mundial quando considera o valor de US$ 1,06 ao dia, por exemplo. Existem desvantagens nessa escolha, já que em nenhum momento o salário mínimo foi estabelecido fielmente de acordo com o custo de uma cesta de bens que atendam às necessidades básicas das famílias. Por outro lado, é uma das formas mais utilizadas no Brasil nos estudos sobre pobreza (ROCHA, 2006).

A pobreza e a desigualdade estão totalmente relacionadas, na medida em que a persistência da pobreza absoluta no Brasil decorre da grande desigualdade na distribuição da renda. Com um grau de concentração tão elevado, mesmo avanços em

[15] Rocha (2006) mostra melhor esta questão, além de demonstrar como se constroem linhas de pobreza adequadas a determinadas realidades.

redução da proporção de pobres podem praticamente não alterar os valores do índice de Gini[16], como o que decorre depois do Plano Real em 1994 (ROCHA, 2006).

Outro fator importante, considerado para escolher a renda como *proxy* do nível e da distribuição do bem-estar, é a família ser a unidade analítica utilizada na abordagem da insuficiência de renda, assim como é para o caso da pluriatividade. A utilização da renda de todas as pessoas da família se justifica por levar em conta a estratégia de sobrevivência normalmente adotada, recorrendo-se àqueles com renda para atender às necessidades daqueles sem rendimento na família (ROCHA, 2006). A desvantagem está relacionada à subestimação, encontrada nos dados da Pnad, das rendas das famílias, principalmente, daquelas com maior poder aquisitivo, o que leva também a uma subestimação do grau de desigualdade. Além desta, a Pnad não inclui na renda a produção para o autoconsumo, o que subestima a renda nos estratos mais baixos e pode levar à superestimação da desigualdade e da pobreza (ROCHA, 2006; GUIMARÃES, 2006 e 2007).

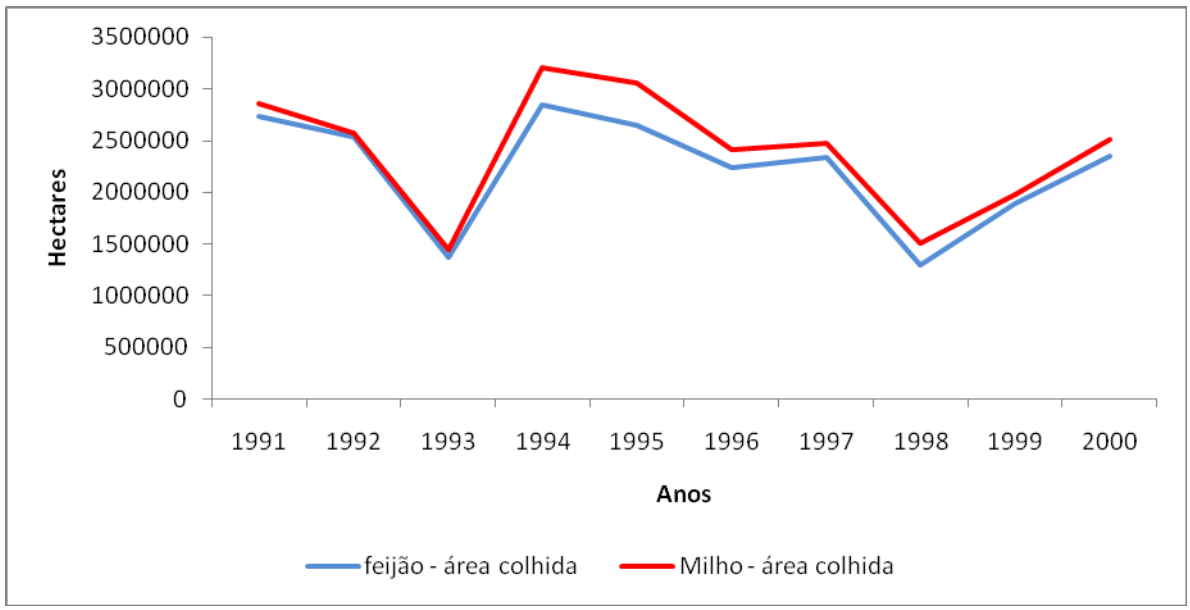
Todas as análises realizadas para a região Nordeste, precisam necessariamente considerar se o ano de estudo é (ou não) de chuvas regulares. Isto porque uma característica diferenciadora do Nordeste das demais regiões do Brasil é justamente a ocorrência periódica da seca. Segundo Molion (1986, p. 40), "há registros históricos da ocorrência de secas no Nordeste desde os primórdios da colonização". A região Nordeste é diferenciada das demais por apresentar um clima semi-árido decorrente de grande variabilidade (espacial e temporal) da precipitação pluviométrica. Enquanto no litoral a média de precipitação supera os 2000 mm anuais, no interior ela é inferior a 400 mm por ano. A área mais seca do Brasil, conhecida como "o triângulo mais seco" do país, é formada pelas cidades de Cabaceiras e Barra de Santa Rosa na Paraíba e Acari no Rio Grande do Norte, apresentando índices inferiores a 300 mm (RODRIGUEZ, 2002).

Analisando tecnicamente o fenômeno, Molion (1986, p.46) descreve que "a semi-aridez do Nordeste é determinada primordialmente pela circulação geral da atmosfera, ou seja, por um fenômeno externo à região, estabelecido provavelmente há

[16] Índice de concentração proposto pelo matemático italiano Conrado Gini, em 1914. Este índice é detalhado na parte da metodologia desta pesquisa.

cerca de 20.000 anos, no fim da era glacial". Porém, existem fatores locais como a topografia e a alta refletividade da crosta[17] que contribuem para a ocorrência das secas.

A Figura 1 reporta informações sobre a área colhida na região Nordeste, das culturas de milho e feijão, que constituem a base da produção da maior parte dos agricultores da região, entre os anos de 1991 a 2000. Os anos de 1993 e 1998 são considerados com ausência de chuvas regulares, ficando isto evidente pela redução drástica que causa na área colhida.



Fonte: IPEADATA, 2008.

Figura 1 – Área colhida de Feijão e Milho na Região Nordeste nos anos de 1991 a 2000.

Do ponto de vista socioeconômico, pela situação de miséria em que vive grande parte das famílias no rural do Nordeste, "a seca provoca, sobretudo, uma crise da agricultura de subsistência. Daí suas características de calamidade social" (GTDN, 1978, p. 65). O problema é que a seca atinge toda a população, contudo, a mais afetada e que mais sofre as consequências é justamente a menos preparada para enfrentá-la, a camada que sobrevive da agricultura de subsistência.

[17] Isto reduz a absorção de calor, tornando a região relativamente mais fria que as circunvizinhas, refletindo-se nas massas de ar e na formação das nuves de chuva.

1.3 O Problema e sua importância

Na região Nordeste existe uma grande quantidade de famílias rurais com pelo menos um membro ocupado em atividades não-agrícolas, sendo esta a região que se concentra o maior percentual de famílias pobres. A pobreza se reflete nos indicadores sociais regionais. O desenvolvimento do meio rural nordestino constitui-se parte fundamental de uma estratégia de desenvolvimento nacional. Diversas agências de pesquisa e financiamento internacionais têm considerado que o estímulo ao não-agrícola pode contribuir para melhorar as condições de vida das famílias rurais. Neste sentido, este trabalho busca analisar os determinantes da pluriatividade e os efeitos das rendas não-agrícolas sobre a pobreza e desigualdade no Nordeste brasileiro, buscando analisar se o estímulo às ocupações não-agrícolas pode se constituir numa nova política pública visando ao desenvolvimento e a se agregar com as existentes.

Em décadas passadas, muitos habitantes da região Nordeste, face a todas as dificuldades com que se defrontavam, decidiam migrar para outras cidades ou mesmo para outras regiões em busca de melhores condições de vida (GARCIA JR., 1989). Para exemplificar, considerando a evolução da população brasileira de 10 anos ou mais na década de 80 e de 90, constata-se um grande crescimento da população total (passando de 88,9 milhões de pessoas em 1981 para 130,1 milhões em 1999). Este crescimento ocorre, principalmente, em áreas urbanas, onde a população salta de 64,7 milhões de pessoas para 104,7 milhões (aumento de 40 milhões de pessoas). Nas áreas rurais, ocorre uma pequena redução da população com 10 anos ou mais, entre 1981 e 1992, demonstrando o êxodo rural (LIMA, 2002).

Porém, devido à falta de qualificação profissional e da grande concorrência no mercado de trabalho, as possibilidades de o migrante sertanejo conseguir ocupação e renda para viver nas grandes regiões metropolitanas são pequenas. A reversão (ou pelo menos um arrefecimento) do fenômeno da migração, em dados agregados para o Brasil, acaba ocorrendo nos anos 90, com a população rural crescendo a uma taxa de 0,90% a.a. e tendo um incremento, entre 1992 e 1999, de aproximados 1,4 milhões de pessoas (LIMA, 2002).

Embora a população rural ocupada volte a crescer nos anos 90, os dados da população rural ocupada em atividades agrícolas demonstram uma queda de -1,70% a.a., entre 1992 e 1999. Isto reflete "um crescimento da mecanização dos cultivos e da colheita de grãos, especialmente nas regiões de expansão da fronteira agrícola. Esse avanço tecnológico em nossa agricultura tem gerado uma tendência de desemprego estrutural desde os anos 70" (GRAZIANO DA SILVA e DEL GROSSI, 2001, p. 4). Uma das explicações para o fato de a população rural ocupada estar aumentando mesmo quando cai o emprego agrícola é o crescimento da geração de ocupação não-agrícola em áreas rurais. O crescimento dos inativos residentes em áreas rurais também contribui para o aumento da população. Como exemplo, há os aposentados e os desempregados que continuam residindo no meio rural. Isto indica que "o meio rural brasileiro já se converteu também num lugar de residência dissociado do local de trabalho. (...) as pessoas residentes no meio rural não estão necessariamente ocupadas, nem muito menos ocupadas em atividades agrícolas" (GRAZIANO DA SILVA e DEL GROSSI, 2001, p. 5-6).

A Tabela 4 mostra a distribuição das famílias e as taxas de crescimento entre 2001 e 2005 para a região Nordeste. Observa-se, dentre as famílias do tipo conta-própria, uma redução das exclusivamente agrícolas (-3,10%) e um crescimento acentuado entre as exclusivamente não-agrícolas (8%). Com relação aos empregados, observa-se um crescimento tanto na parte agrícola (3,6%) como na pluriativa (5,10%). Segundo Nascimento e Cardozo (2007), isto possivelmente indica que as famílias rurais do Nordeste estão cada vez mais se proletarizando, ou seja, dependendo da renda proveniente de salários.

No ano de 2005, é importante salientar a grande quantidade de famílias de não-ocupados residindo no meio rural, como também demonstra a Tabela 4. O valor de 362 mil famílias de não-ocupados supera o número de famílias pluriativas (122 mil) somado ao de famílias não-agrícolas (211 mil) do tipo empregados. Outra questão importante é que demonstra uma taxa de crescimento da população residente no meio rural de 1,3%, invertendo a forma de migração existente anteriormente.

Tabela 4 – Distribuição e Taxas de crescimento dos tipos de famílias, segundo o local de domicílio, Nordeste, 2001 a 2005

Local Domicílio/Tipo de Família	Distribuição 2005	Taxa de Crescimento (%) 2001-2005
Não Metropolitano		
Rural Agropecuário		
Conta Própria	1.513	-0,70
Agrícola	771	-3,10*
Pluriativo	603	1,30
Não-agrícola	139	8,00*
Empregados	952	4,20***
Agrícola	620	3,60**
Pluriativo	122	5,10***
Não-agrícola	211	5,40
Não ocupados	362	2,80
Total	2.934	1,30***

Fonte: Pnad – Tabulações Especiais do Projeto Rurbano. In: Nascimento e Cardozo (2007).

***, ** e * indicam, respectivamente, 5%, 10% e 20%.

A Tabela 5 trata da distribuição e taxa de crescimento, entre 2002 e 2004, das principais ocupações não-agrícolas exercidas pelas famílias residentes no meio rural do Nordeste. O que se observa é que as ocupações são tipicamente urbanas, mas exercidas por pessoas que declaram sua residência no meio rural. A melhoria do sistema de transportes, com mais estradas pavimentadas e aumento do número de meios de transporte, por exemplo, facilita o acesso do residente rural à cidade para trabalhar (KAGEYAMA, 1998; TEIXEIRA, 2004; GRAZIANO DA SILVA *et al.*, 2005).

Assim, podem-se explicar as elevadas taxas de crescimento de ocupações como: vendedores e demonstradores em lojas ou mercados (10,5%), trabalhadores na manutenção e conservação de edifícios (14,8%) e trabalhadores nos serviços de administração de edifícios (10,2%). Já com relação a professores de ensino fundamental com formação de nível médio ocorre uma redução de -2,1% a.a.. Isto pode ser algo importante se estiverem sendo substituídos por profissionais com maior

formação. Faz-se necessária uma maior investigação para entender esta redução, assim como a dos agentes de saúde (-5,2%), como indicado na Tabela 5.

Tabela 5 - Distribuição (2004) e Taxas de Crescimento (2002-2004) das principais ocupações das pes

OCUPAÇÕES NÃO-AGRÍCOLAS
Vendedores e demonstradores em lojas ou mercados
Trabalhadores na manutenção e conservação de edifícios
Professores (formação nível médio) - ensino fundamental
Extrativistas florestais
Agentes da saúde e do meio ambiente
Trabalhadores nos serviços de administração de edifícios
Gerentes de produção e operações
Trabalhadores de montagem de estruturas de madeira e metal
Trabalhadores de instalações elétricas

Fonte: Projeto Rurbano, processamento das Pnad.
***, ** e * indicam, respectivamente, 5%, 10% e 20%.

No Brasil, o setor mais dinâmico de geração de Orna e que contribui para redução da pobreza rural é o de serviços domésticos (NASCIMENTO, 2004). É também uma questão de gênero, que proporciona às famílias com mulheres em condições de exercer esta atividade maior probabilidade de serem pluriativas[18]. Para o ano de 1999, segundo Leone (2004), 84,90% das famílias rurais pluriativas do Nordeste possuíam pelo menos uma mulher ocupada entre seus componentes, sugerindo o peso da mulher para a família ser pluriativa.

Por outro lado, em seu trabalho de doutoramento, Nascimento (2005) chega a conclusões interessantes e intrigantes. Para este autor, a pluriatividade é mais característica em regiões pobres[19], como a Nordeste, sendo justamente nesta região

[18] Segundo Ferreira e Lanjouw (2001), dos membros das famílias, as mulheres possuem a maior probabilidade de participar das Orna. De Janvry e Saudolet (2001) encontraram na pesquisa para o México que, apenas para o sexo feminino, morar mais próximo de centro urbano eleva a possibilidade de se ocuparem em atividades não-agrícolas.

[19] Esta afirmação não é compartilhada por Schneider (2006, p. 230), que considera que "a combinação de atividades agrícolas com outras não agrícolas pelos membros de uma mesma família parece ser um fenômeno mais acentuado no Sul do Brasil, onde se verifica um aumento mais expressivo, passando de 142 mil famílias em 2001 para 178 mil em 2004".

onde predomina a quantidade de famílias pobres sobre o total de famílias. No Nordeste, as famílias enfrentam a pobreza diversificando suas fontes de ocupação e renda, porém em setores que exigem baixa qualificação da mão-de-obra e pagam baixos salários, o que apenas contribui para sua reprodução e não para gerar um processo que induza à melhoria do bem-estar. Segundo Nascimento (2004), a pluriatividade no Nordeste só cresce devido ao emprego doméstico, pois o Nordeste não tem indústrias rurais e a parca agroindústria ligada ao setor sucroalcooleiro constantemente passa por instabilidades.

Na análise da Tabela 6, percebe-se que tanto para as famílias do tipo conta-própria quanto para aquelas de empregados, as exclusivamente agrícolas percebem menores rendimentos médios (R$ 381,65 para conta-própria e R$ 344,57 para empregados). Para os dois tipos de famílias (conta-própria e empregados), as famílias exclusivamente agrícolas são mais dependentes de rendas de aposentadorias/pensões e outras fontes, comparativamente às famílias pluriativas. As famílias do tipo pluriativas, ou seja, as que combinam atividades agrícolas e não-agrícolas são as que possuem maiores valores de renda média, respectivamente R$ 545,00 para as conta-própria e R$ 559,10 para os empregados.

Como se pode observar ainda na Tabela 6, as famílias que sobrevivem apenas de atividades agrícolas vivem em situação de maior dificuldade, haja vista receberem as menores rendas. As pluriativas possuem rendas superiores à média do total dos conta-própria e empregados. Com isto, fica clara a grande representatividade das famílias pluriativas e não-agrícolas nordestinas, além de a pluriatividade ser importante e estar aumentando por existirem possibilidades de ampliar e diversificar as atividades no meio rural.

Tabela 6: Composição da Renda Média das famílias extensas[20], Nordeste, 2004 (R$ set. 2004)

LOCAL DOMICÍLIO / TIPO	AGRÍCO-	NÃO-	APOS/PENS	OUTRAS	2004

[20] Inclui além da família nuclear, os parentes e agregados que vivem no mesmo domicílio. Foram excluídos desta unidade de análise os empregados domésticos e seus parentes e pensionistas que pagam pensão ao chefe do domicílio (GRAZIANO DA SILVA e DEL GROSSI, 2001).

DE FAMÍLIA	LA %	AGRIC %	%	%	R$
Rural Agropecuário	42,4	18,8	30,1	8,7	452,78
Conta-Própria	44,3	18,3	28,3	9,1	447,84
Agrícola	52,3	0,0	37,6	10,1	381,65
Pluriativo	44,6	28,3	19,2	7,9	545,00
Não-agrícola	0,0	71,4	19,3	9,3	505,00
Empregados	48,6	27,8	15,7	7,9	409,40
Agrícola	74,4	0,0	16,7	9,0	344,57
Pluriativo	41,7	42,6	8,9	6,8	559,10
Não-agrícola	0,0	73,8	19,5	6,7	506,10
Não-ocupado na semana	0,0	0,0	86,5	13,5	364,29

Fonte: Projeto Rurbano, processamento das Pnad.

As rendas das famílias pluriativas e exclusivamente não-agrícolas, que moram no meio rural, são mais elevadas do que as rendas das famílias exclusivamente agrícolas (LIMA, 2002; GRAZIANO DA SILVA e DEL GROSSI, 2001; ADAMS, 2001; DE JANVRY *et al.*, 2005), e muitos trabalhos de organismos internacionais têm destacado a importância do não-agrícola nas estratégias de políticas para redução da pobreza no meio rural[21] (VEIGA, 2001; VALDÉS e MISTIAEN, 2001; OIT, 2005; WORD BANK, 2005, NASCIMENTO e CARDOSO, 2007).

Considerando a literatura nacional, Silva Júnior (2006), trabalhando de forma mais desagregada, demonstra que o estado da região Nordeste em que o trabalhador reside tem influência na probabilidade de estar acima ou abaixo da linha de pobreza. Nos estados da Paraíba e Piauí, a probabilidade de o indivíduo estar abaixo da linha é maior do que se residissem na Bahia ou no Rio Grande do Norte.

Veras (2004), em pesquisa de campo no estado de Alagoas, identifica que as pessoas com baixa ou nenhuma escolaridade ou trabalham na agricultura ou estão não-ocupadas. Por outro lado, o acesso à pluriatividade ou às atividades não-agrícolas está relacionado ao fato de possuírem o segundo grau (ensino médio). De forma

[21] Atualmente, entende-se que desenvolvimento rural (preocupação com as condições de vida da população, a qualidade dos produtos, os níveis de renda, o acesso a terra, as relações de trabalho e a conservação dos recursos naturais, por exemplo) é um conceito bem mais amplo do que desenvolvimento agrícola, em que a grande preocupação são aspectos meramente produtivos. Conseqüentemente, pode-se considerar que a reestruturação das bases econômicas do meio rural, necessária para o combate à fome e à redução da pobreza, deve ocorrer também com um maior estímulo às chamadas Orna, de acordo com as potencialidades de cada localidade.

semelhante a Veras (2004) e Ferreira e Lanjouw (2001), Silva Júnior (2006) encontra na educação o fator mais relevante para explicar a pobreza, considerando que o foco principal de políticas públicas que buscam reduzir a pobreza deve ser investir na educação de qualidade e incentivar as atividades não-agrícolas no meio rural.

Ainda para o Brasil, segundo Neder (2003a) e Mariano e Neder (2004), a renda não-agrícola é importante para explicar a redução do número de famílias que estão abaixo da linha de pobreza, porém, a renda gerada nas atividades não-agrícolas contribui para aumentar a desigualdade de renda entre as famílias. Para o ano de 2001, demonstram que o índice de Gini das famílias que receberam renda não-agrícola é de 0,83 e para as famílias agrícolas, 0,62 (MARIANO e NEDER, 2004). Esta análise é relevante, haja vista que se busca um desenvolvimento do meio rural maior, mais equilibrado e equitativo.

Existem diversos trabalhos na literatura internacional, como De Janvry *et al* (2005) e Zhu e Luo (2006), analisando o caso da China; Araújo (2004), no caso do México; Adams e He (1995) e Adams (1999), para o Paquistão e Egito, respectivamente; e Escobal (2001) para o Peru, analisando o papel das rendas e atividades não-agrícolas na redução da pobreza no meio rural. Na literatura nacional, entre outros, citam-se os trabalhos de Silva Júnior (2006), Nascimento (2004, 2005), Mariano e Neder (2004), Ferreira e Lanjouw (2001), Schneider e Fialho (2000), relacionando, para as diversas partes do país, a importância das rendas e Orna para reduzir a pobreza no meio rural.

Analisando na literatura internacional a interação entre pobreza e Orna, Araújo (2004) encontra uma correlação negativa entre pobreza e Orna, principalmente no setor de serviços, dos municípios rurais. Os resultados obtidos por De Janvry *et al.* (2005) demonstram que a participação em atividades não-agrícolas reduz a pobreza rural. Para este autor, quantidade de terras *per capita* e local de moradia estão negativamente correlacionados com a participação em Orna e existe uma relação quadrática com retornos decrescentes entre o número de anos de educação e participação em Orna.

Com relação à equação da renda, número de dependentes, de trabalhadores e local de moradia, eles estão positivamente relacionados com a renda do domicílio. Nível educacional, porém, não tem influência na renda. Analisando as classes decompostas pelo índice de Foster-Greer-Thorbecke (FGT) (FOSTER *et al.*, 1984), demonstra-se que na classe "incidência de pobreza" a renda não-agrícola conduz à redução de pobreza de 68,90% para 10,90%, e na de pobreza severa de 26,30% para 1,20%.

Sobre a desigualdade, De Janvry *et al.* (2005), após simularem a renda de domicílios pluriativos sobre a condição de que participassem apenas de atividades agrícolas e analisar os cálculos dos índices de Gini, concluíram que o índice é menor na presença de renda não-agrícola. Em outras palavras, a pluriatividade, no caso da China, reduz a desigualdade de renda, resultados confirmados no trabalho de Zhu e Luo (2006). Para estes autores, a pluriatividade reduz a desigualdade de renda rural na medida em que alarga o leque de possibilidade de escolhas de ocupação, desproporcionalmente favorecendo os domicílios mais pobres.

Segundo Adams (1999), utilizando análise de decomposição das rendas para os dados do Egito, a renda agrícola é uma fonte de aumento da desigualdade de renda por estar relacionada com a posse da terra, que também é muito concentrada. A renda não-agrícola, por outro lado, não está vinculada à posse de terra, ou seja, não existe correlação positiva entre ambos, sendo mais importante para os pobres.

É necessário, então, avançar os conhecimentos sobre esta temática, analisando o efeito da pluriatividade (combinação de atividades agrícolas e não-agrícolas pelos membros da família) e das rendas não-agrícolas (tanto advindas de Orna quanto de transferências públicas e privadas) sobre a pobreza e a desigualdade, especificamente no meio rural da região Nordeste. Esta pesquisa apresenta-se como uma contribuição adicional aos estudos das transformações recentes no meio rural brasileiro, em especial no caso do Nordeste, no sentido de aprofundar o entendimento da importância da diversificação das atividades e rendas para a melhoria do bem-estar da população. Neste trabalho será mais bem compreendido como a diversificação de atividades e a entrada de outras rendas afetam a pobreza e a desigualdade, esperando-se que as

evidências empíricas encontradas contribuam para os formuladores de políticas públicas orientarem ações diferentes das voltadas exclusivamente para as atividades agropecuárias, com o intuito de desenvolver o meio rural nordestino.

1.4. Hipóteses

As hipóteses a serem testadas são: a) na região Nordeste, a pluriatividade contribui para redução da pobreza no meio rural; b) as rendas não-agrícolas colaboram para reduzir a concentração de renda no meio rural da região.

1.5. Objetivos

1.5.1 Geral

Neste trabalho, objetiva-se analisar os efeitos da pluriatividade e rendas não-agrícolas sobre a pobreza e desigualdade rural na região Nordeste nos anos de 2003 e 2005. O ano de 2003 pode ser considerado com ausência de chuvas regulares na maior parte da região (ano de seca). O ano de 2005, por outro lado, um ano de chuvas regulares. A *priori*, imagina-se que a escolha de apenas um ano pode viesar os resultados, favorecendo a parcela de renda não-agrícola se o ano for de seca, ou favorecendo a renda agrícola no ano de chuva. Por isto, é necessário estudar as duas situações.

1.5.2 Específicos

Em termos específicos, a pesquisa tem os seguintes objetivos:

a) Identificar os determinantes da escolha do indivíduo com relação às atividades agrícola, não-agrícola, ser pluriativo ou não-ocupado;

b) Comparar o impacto na renda total das atividades não-agrícolas tanto para as famílias exclusivamente agrícolas quanto pluriativas, avaliando sua importância na redução da pobreza;
c) Analisar o impacto das rendas não-agrícolas sobre a desigualdade no meio rural; e
d) Analisar comparativamente os diferentes impactos da questão climática (ano com e sem chuvas regulares) sobre os determinantes da escolha, redução da pobreza e desigualdade.

e)

2. REFERENCIAL TEÓRICO

O referencial teórico[22] utilizado nesta pesquisa está relacionado à oferta de mão-de-obra rural, focando a possibilidade de a família poder se engajar em múltiplas fontes de emprego (agrícola e/ou não-agrícola). Segundo esta abordagem, a família compara as opções de trabalho e aloca seu tempo total disponível de forma a maximizar sua função de utilidade. Busca-se, assim, modelar o domicílio de forma a entender as decisões da família sobre consumo, produção e a alocação de tempo. Quanto mais tempo é alocado em um tipo de trabalho, menos resta para os demais, ou seja, quanto mais tempo é alocado para o emprego em atividades não-agrícolas, menos tempo se dispõe para as atividades agrícolas.

Este referencial é semelhante ao apresentado em Lee (1998), Andrade (2003) e Zeng (2005), com pequenas modificações, buscando adaptá-lo à realidade brasileira. Como a unidade analítica utilizada na pesquisa é a família, o modelo escolhido não pode considerar apenas a função de utilidade do chefe da família, como ocorre em Andrade (2003). É necessário um modelo que capte a interdependência das decisões de oferta de trabalho. A decisão de um membro da família de participar em atividades não-agrícolas deve ter uma relação de interdependência com a decisão de outro membro. Segundo Lee (1998), analisar a decisão familiar de oferta de trabalho envolve

[22] Existem outras formas de abordar a questão, como o método de análise de sistemas agrários. Segundo este enfoque, analisa-se a família como um núcleo que busca reprodução social, sendo a pluriatividade importante para preencher o tempo de ociosidade (não-trabalho). Para maiores informações, vide Abramovay (1998) e Romero (1998). Neste trabalho, opta-se pelo referencial teórico que considera a família como maximizadora de utilidade, possibilitando fazer uso do modelo econométrico logit multinomial, conforme descrito no item 3.1.

a questão de como modelar a maneira pela qual os membros tomam suas decisões econômicas. Basicamente existem três grupos de modelos[23], mas por facilidade de interpretação esta pesquisa utiliza o modelo de "utilidade conjunta" (*joint utility*).

2.1 A abordagem da utilidade conjunta

Segundo Lee (1998), a abordagem da utilidade conjunta assume que os membros da família maximizam uma função de utilidade, comum a todos. Esta função de utilidade possui os atributos e o comportamento econômico dos membros com argumentos separados (porém agregáveis), assim como as propriedades usuais da função de utilidade da teoria do consumidor individual. Esta abordagem é usada para modelar domicílios parcialmente integrados aos mercados. A família, internamente, decide o que consumir e produzir, bem como quanto ofertar de trabalho nas atividades agrícolas e não-agrícolas. Uma hipótese importante do modelo é a total cooperação entre os membros da família. Para facilidade de exposição, é apresentado o caso de uma família composta de duas pessoas (esposo e esposa - o caso mais simples de ser demonstrado) e que pode ser estendido para o caso de uma família com *n* componentes. A família, então, busca maximizar a seguinte função de utilidade,

$$\max_{T_{d1},T_{d2},C,T_{agr1},T_{agr2},T_{nag1},T_{nag2}} U(T_{d1},T_{d2},C;f) \qquad (1)$$

sujeito a

$$C = f(p;T_{agr1},T_{agr2};H,Z_{agr}) + g(T_{nag1},T_{nag2};H,Z_{nag}) + RNT \qquad (2)$$

$$T_i = T_{di} + T_{agri} + T_{nagi}, \text{ com i=1,2} \qquad (3)$$

$$T_{agri},T_{nagi} \geq 0, \text{ com i=1,2} \qquad (4)$$

em que ,

[23] Os outros dois modelos são o da "família tradicional" e o de "barganha". Para maiores informações, vide Lee (1998).

T_d= tempo alocado nas atividades do domicílio, envolvendo os cuidados com os membros familiares e atividades afins, podendo também ser considerado como lazer;
C= consumo de bens;
J= características da família que afetam suas preferências;
f= indica que o consumo é uma função f da renda agrícola;
g= indica que o consumo é uma função g da renda não-agrícola;
p= vetor de preços dos produtos agrícolas e insumos, menos o trabalho no próprio domicílio;
T_{agr}= tempo de trabalho em atividades agrícolas na propriedade;
Z_{agr}= insumos fixos da propriedade;
T_{nag}= tempo de trabalho em atividades não-agrícolas dentro ou fora da propriedade;
H= Capital humano que influencia no nível de renda agrícola e não-agrícola;
Z_{nag}= Outras variáveis que influenciam o nível salarial;
RNT= Renda do "não-trabalho" ou transferências (aposentadorias, pensões, bolsas do governo, auxílio enviado por um parente que migrou etc.);
T_i= Trabalho total; e
i= 1 ou 2, se refere ao membro da família. Ex.: marido e a esposa.

A utilidade da família é determinada pelo tempo gasto com as atividades do domicílio/lazer (T_d) e com o consumo conjunto de bens (C). É permitido que a função de utilidade varie de acordo com as características das famílias (J). Para maximizar sua função, a família se defronta basicamente com duas restrições: a) orçamentária, ou seja, o nível de consumo depende da soma das rendas agrícola, não-agrícola e de transferências; e b) temporal, ou seja, existe um montante fixo de tempo, que deve ser alocado em atividades agrícolas, não-agrícolas e lazer tanto no caso da mulher quanto do marido. O modelo assume ainda que tanto o tempo de trabalho agrícola quanto o não-agrícola (ou ambos) podem ser zero[24] (restrição de não-negatividade) e que a taxa de salário é constante[25]. Como na maximização deste problema pode-se encontrar uma solução interior ($T_{agr}>0$ e $T_{nag}>0$), mas também uma de canto (quando $T_{agr}=0$ ou $T_{nag}=0$ ou ambas), para solução do Lagrangiano é utilizado o método de Kuhn-Tucker.

O Lagrangeano da função de utilidade (1), sujeito às restrições (2), (3) e (4), é dado por

[24] Esta adaptação é importante, haja vista que a família pode se dedicar exclusivamente a atividades agrícolas ($T_{agr}>0$ e $T_{nag}=0$), ou exclusivamente a atividades não-agrícolas ($T_{agr}=0$ e $T_{nag}>0$), ou ainda ser pluriativa ($T_{agr}>0$ e $T_{nag}>0$) ou não-ocupada, vivendo exclusivamente de transferências ($T_{agr}=0$ e $T_{nag}=0$).
[25] Segundo Lee (1998), esta hipótese pode ser considerada razoável, já que não se estará analisando o número de horas dedicadas às atividades não-agrícolas/agrícolas, e sim na decisão de participar ou não em atividades não-agrícolas.

$$L = U(T_{d1}, T_{d2}, C; f) + \tau(T_1 - T_{d1} - T_{agr1} - T_{nag1}) + \varphi(T_2 - T_{d2} - T_{agr2} - T_{nag2}) + \lambda[f(p; T_{agr1}, T_{agr2}; Z_{agr}) + g(T_{nag1}, T_{nag2}; H, Z_{nag}) + RNT - C] \quad (5)$$

Tomando as derivadas primeiras e considerando as restrições de não-negatividade, têm-se

$$\frac{\partial L}{\partial T_{d1}} = \frac{\partial U}{\partial T_{d1}} - \tau = U_1 - \tau = 0 \quad (6)$$

$$\frac{\partial L}{\partial T_{d2}} = \frac{\partial U}{\partial T_{d2}} - \varphi = U_2 - \varphi = 0 \quad (7)$$

$$\frac{\partial L}{\partial C} = \frac{\partial U}{\partial C} - \lambda = U_3 - \lambda = 0 \quad (8)$$

$$\frac{\partial L}{\partial T_{agr1}} = \lambda \frac{\partial f}{\partial T_{agr1}} - \tau \leq 0, T_{agr1} \geq 0, (\lambda f_1 - \tau) T_{agr1} = 0 \quad (9)$$

$$\frac{\partial L}{\partial T_{agr2}} = \lambda \frac{\partial f}{\partial T_{agr2}} - \varphi \leq 0, T_{agr2} \geq 0, (\lambda f_2 - \varphi) T_{agr2} = 0 \quad (10)$$

$$\frac{\partial g}{\partial T_{nag1}} = \lambda \frac{\partial L}{\partial T_{nag1}} - \tau \leq 0, T_{nag1} \geq 0, (\lambda g_1 - \tau) T_{nag1} = 0 \quad (11)$$

$$\frac{\partial L}{\partial T_{nag2}} = \lambda \frac{\partial L}{\partial T_{nag2}} - \varphi \leq 0, T_{nag2} \geq 0, (\lambda g_2 - \tau) T_{nag2} = 0 \quad (12)$$

Desta forma, as condições de otimização do Lagrangeano dependem da presença ou não de trabalho fora da propriedade. Os tópicos seguintes buscam demonstrar que a decisão de alocar ou não trabalho fora da propriedade depende de a renda obtida da atividade não-agrícola ser ou não maior do que o preço sombra[26] do tempo de ser exclusivamente agrícola.

[26] O preço-sombra tem a seguinte interpretação: é o benefício encontrado (alteração) na função objetivo (a qual vai ser maximizada) quando se acrescenta uma unidade ao recurso escasso/restrição (tempo, neste caso) (LINS e CALÔBA, 2006). Então, para ofertar trabalho fora da propriedade, a renda extra obtida da atividade não-agrícola deve proporcionar uma elevação no nível utilidade/satisfação da família maior do que a do preço-sombra do tempo de ser exlusivamente agrícola.

2.1.1 O caso das famílias exclusivamente agrícolas

No caso das famílias exclusivamente agrícolas (T_{nag}=0), definindo $W_{1agr} = \frac{\tau}{\lambda}$ e $W_{2agr} = \frac{\varphi}{\lambda}$, as condições de otimalidade podem ser simplificadas para:

$$\lambda f_1 - \tau = 0 \equiv f_1 = W_1 \quad (13)$$

$$\lambda f_2 - \varphi = 0 \equiv f_2 = W_2 \quad (14)$$

$$W_1 = \frac{\tau}{\lambda} = \frac{U_1}{U_2} \quad (15)$$

$$W_2 = \frac{\varphi}{\lambda} = \frac{U_2}{U_2} \quad (16)$$

$$T_i = T_{di} + T_{agri}, \text{com } i = 1, 2 \quad (17)$$

$$C + W_{1agr}T_{d1} + W_{2agr}T_{d2} = W_{1agr}T_1 + W_{2agr}T_2 + [f(p; T_{agr1}, T_{agr2}; Z_{agr}) - (W_{1agr}T_{agr1} + W_{2agr}T_{agr2})] + RNT \quad (18)$$

Denominando de lucro da atividade agrícola (π) a parcela $[f(p; T_{agr1}, T_{agr2}; Z_{agr}) - (W_{1agr}T_{agr1} + W_{2agr}T_{agr2})]$, pode-se simplificar a equação (18)

$$C + W_{1agr}T_{d1} + W_{2agr}T_{d2} = W_{1agr}T_1 + W_{2agr}T_2 + \pi + RNT \quad (19)$$

mais as equações (9) e (10) derivadas acima.

Estas condições indicam que uma vez determinado o preço do tempo de cada membro da família (W_1 e W_2), a decisão de alocar o tempo de trabalho apenas em atividades agrícolas pode ser explicada em termos de comportamento maximizador de lucro, ou seja, se o ganho "marginal" com a atividade não-agrícola é menor do que o preço-sombra do tempo de ser exclusivamente agrícola (W_1 e W_2). O consumo e o tempo dedicado ao "lazer" dependem da soma do valor imputado à dotação de tempo

do casal ($W_1T_1 + W_2T_2$), do lucro da atividade agrícola (π) e da renda do não-trabalho (RNT).

2.1.2 O caso das famílias exclusivamente não-agrícolas

No caso das famílias exclusivamente não-agrícolas (T_{agr}=0), definindo $W_{1nag} = \frac{\tau}{\lambda}$ e $W_{2nag} = \frac{\varphi}{\lambda}$, as condições de otimalidade podem ser simplificadas para:

$$\lambda g_1 - \tau = 0 \equiv g_1 = W_{1nag}$$

(20)

$$\lambda g_2 - \varphi = 0 \equiv g_2 = W_{2nag} \quad (21)$$

$$T_i = T_{di} + T_{nagi}, \text{com } i = 1,2 \quad (22)$$

$$C + W_{1nag}T_{d1} + W_{2nag}T_{d2} = g(T_{nag1}, T_{nag2}; H, Z_{nag}) + RNT \quad (23)$$

mais as equações (11), (12), (16) e (17), derivadas acima.

Estas condições também indicam que uma vez determinado o preço do tempo de cada membro da família (W_{1nag} e W_{2nag}), a decisão de alocar o tempo de trabalho apenas em atividades não-agrícolas depende de o preço-sombra do trabalho não-agrícola (W_{1nag} e W_{2nag}) ser maior do que a renda "marginal" obtida da atividade agrícola. O consumo e o tempo dedicado ao "lazer" dependem dos ganhos obtidos das Orna e da renda do não-trabalho (RNT).

2.1.3 O caso das famílias pluriativas

No caso das famílias pluriativas (T_{agr},$T_{nag} \geq 0$), novamente definindo $W_1 = \frac{\tau}{\lambda}$ e $W_2 = \frac{\varphi}{\lambda}$ as condições de otimalidade podem ser demonstradas pelas equações (3), (9), (10), (11), (12), (13), (14), (15), (16), (20), (21). Para a equação de consumo, tem-se:

$$C + W_{1nag}T_{d1} + W_{2nag}T_{d2} + W_{1agr}T_{d1} + W_{2agr}T_{d2} = W_{1nag}T_1 + W_{2nag}T_2 + W_{1agr}T_1 + W_{2agr}T_2 + [f(p; T_{agr1}, T_{agr2}; Z_{agr}) - (W_{1agr}T_{agr1} + W_{2agr}T_{agr2}) - (W_{1nag}T_{agr1} + W_{2nag}T_{agr2})] + g(T_{nag1}, T_{nag2}; H, Z_{nag}) + RNT$$

a qual pode ser resumida para

$$C + L = VT + \pi^* + S + RNT \quad (24)$$

em que,

L = $W_{1nag}T_{d1} + W_{2nag}T_{d2} + W_{1agr}T_{d1} + W_{2agr}T_{d2}$;

VT = $W_{1nag}T_1 + W_{2nag}T_2 + W_{1agr}T_1 + W_{2agr}T_2$;

π = $[f(p; T_{agr1}, T_{agr2}; Z_{agr}) - (W_{1agr}T_{agr1} + W_{2agr}T_{agr2}) - (W_{1nag}T_{agr1} + W_{2nag}T_{agr2})]$;

S= $g(T_{nag1}, T_{nag2}; H, Z_{nag})$

L pode ser considerado o custo do lazer avaliado tanto em função do salário dentro quanto fora da propriedade; VT é o valor do tempo endógeno avaliado tanto pela taxa salarial de mercado fora quanto dentro da propriedade; π é o lucro agrícola, obtido da redução das receitas obtidas do trabalho agrícola ($f(p; T_{agr1}, T_{agr2}; Z_{agr})$) e os custos $(W_{1agr}T_{agr1} + W_{2agr}T_{agr2})$, além do custo de oportunidade $(W_{1nag}T_{agr1} + W_{2nag}T_{agr2})$, ou seja, quanto se estaria ganhando caso o membro tivesse ocupado seu tempo de trabalho em atividade não-agrícola; e S é o salário da atividade não-agrícola.

2.1.4 O caso das famílias não-ocupadas

No caso das famílias não-ocupadas (T_{agr}, T_{nag} = 0), a equação de consumo e as condições de otimalidade se resumem a:

$$C = RNT \tag{25}$$

$$T_i = T_{di}\ ,\text{com}\ i = 1,2 \tag{26}$$

$$\frac{\partial L}{\partial T_{agr1}} = \lambda f_1 - \tau \leq 0, T_{agr1} = 0, (\lambda f_1 - \tau) T_{agr1} = 0 \tag{27}$$

$$\frac{\partial L}{\partial T_{agr2}} = \lambda f_2 - \varphi \leq 0, T_{agr2} = 0, (\lambda f_2 - \varphi) T_{agr2} = 0 \tag{28}$$

$$\frac{\partial L}{\partial T_{nag1}} = \lambda g_1 - \tau \leq 0, T_{nag1} = 0, (\lambda g_1 - \tau) T_{nag1} = 0 \tag{29}$$

$$\frac{\partial L}{\partial T_{nag2}} = \lambda g_2 - \varphi \leq 0, T_{nag2} = 0, (\lambda g_2 - \tau) T_{nag2} = 0 \tag{30}$$

Neste caso, não se considera que a família maximiza sua função de utilidade com zero hora trabalhada em atividades agrícolas e/ou não-agrícolas. Também não se considera que estas famílias prefiram lazer a consumo. A situação aqui descrita é o caso de nenhum membro da família conseguir emprego em atividades agrícolas e não-agrícolas. Esta situação é relevante, pois muitas famílias que moram no meio rural se enquadram nesta categoria. Estas famílias preferem não migrar, porque não acreditam que em outras regiões ou estados conseguirão viver em melhores condições, dado que não são qualificadas, não possuindo grau elevado de escolaridade. Então, ficam na propriedade sobrevivendo das aposentadorias dos pais ou parentes e dos programas de assistência social do governo.

Assim, ficam demonstradas as diferenças de comportamento econômico das famílias agrícolas, não-agrícolas, pluriativas e de não-ocupados. Resta identificar como a decisão de participação no trabalho agrícola, não-agrícola ou ambos é tomada. Para a família decidir ser exclusivamente agrícola (T_{nagi}=0), a renda obtida de emprego não-agrícola (g_1 e g_2) deve ser menor que a razão do preço-sombra do tempo doméstico (lazer) pelo preço-sombra da renda total ($\tau/\lambda = W_{1agr}$ e $\varphi/\lambda = W_{2agr}$)

$$\lambda g_1 - \tau \leq 0, T_{nag1} = 0, (\lambda g_1 - \tau) T_{nag1} = 0$$
$$\lambda g_1 \leq \tau \equiv g_1 \leq W_{1agr} \quad (31)$$

e

$$\lambda g_2 - \varphi \leq 0, T_{nag2} = 0, (\lambda g_2 - \tau) T_{nag2} = 0$$
$$\lambda g_2 \leq \varphi \equiv g_2 \leq W_{2agr} \quad (32)$$

Enquanto o acréscimo na utilidade causada por aumento do tempo de lazer for maior do que o acréscimo na utilidade gerado pela elevação do tempo alocado em atividades não-agrícolas, T_{nag} será igual a zero. Se o inverso ocorre, ou seja, $g_i \geq W_{iagr}$, então, necessariamente uma quantidade positiva de T_{nag} será desprendida. Para analisar se a família será exclusivamente agrícola ou pluriativa, deve-se, conseqüentemente, analisar tanto g_i quanto W_{iagr}.

Já com relação às famílias exclusivamente não-agrícolas, a análise é semelhante. A família alocará tempo em atividade agrícola (T_{agr}) apenas quando o lucro (π)[27] advindo da atividade for maior do que o preço-sombra do tempo de lazer.

$$\lambda f_1 - \tau \leq 0, T_{agr1} = 0, (\lambda f_1 - \tau) T_{agr1} = 0$$
$$\lambda f_1 \leq \tau \equiv f_1 \leq W_{1nag} \quad (33)$$

$$\lambda f_2 - \varphi \leq 0, T_{agr2} = 0, (\lambda f_2 - \varphi) T_{agr2} = 0$$
$$\lambda f_2 \leq \varphi \equiv f_2 \leq W_{2nag} \quad (34)$$

Sumarizando as funções de participação no mercado de trabalho (L_i^*), têm-se:

a-) $T_{nag} = 0$, famílias exclusivamente agrícolas:

[27] Já que como demonstrado, $\pi = [f(F_1 T_{agr1}, T_{agr1}, Z_{agr}) - (W_{1agr} T_{agr1} + W_{1agr} T_{agr1}) - (W_{1nag} T_{agr1} + W_{1nag} T_{agr1})]$;

$$L_i^* (H, Z_{nag}, Z_{agr}, p, T, RNT, J) \equiv g_i (H, Z_{nag}) - W_{iagr} (H, Z_{agr}, p, T, RNT, J) \leq 0 \qquad (35)$$

b-) $T_{agr} = 0$, famílias exclusivamente não-agrícolas:

$$L_i^* (H, Z_{nag}, Z_{agr}, p, T, RNT, J) \equiv f_i (H, Z_{nag}) - W_{nag} (H, Z_{agr}, p, T, RNT, J) \leq 0 \qquad (36)$$

c-) $T_{agr}, T_{nag} > 0$, famílias pluriativas:

$$L_i^* (H, Z_{nag}, Z_{agr}, p, T, RNT, J) \equiv g_i (H, Z_{nag}) - W_{iagr} (H, Z_{agr}, p, T, RNT, J) > 0 \qquad (37)$$

e

$$L_i^* (H, Z_{nag}, Z_{agr}, p, T, RNT, J) \equiv f_i (H, Z_{nag}) - W_{nag} (H, Z_{agr}, p, T, RNT, J) > 0 \qquad (38)$$

Seguindo o raciocínio desenvolvido por Andrade (2003), se for definido "δ" como a diferença entre $g_i - W_{iagr}$ e entre $f_i - W_{nag}$, pode-se argumentar que uma elevação nas variáveis que aumentam g_i e f_i ou reduzem W_{iagr} ou W_{nag} contribuem para acrescer "δ". Assim, para todos os tipos de família se espera que as variáveis relacionadas com capital humano possuam sinal positivo e que exerçam influência na decisão de participar do mercado de trabalho na mesma direção de g_i e f_i e na direção contrária de W_{iagr} e W_{nag}. Adicionalmente, os impactos das variáveis p, T e J na decisão de participação ocorrem sempre na direção contrária de W_{iagr}. Para as famílias em que T_{agr} é positivo, outras variáveis que afetam o mercado de trabalho não-agrícola (Z_{nag}) possuem sinal positivo e vão na mesma direção de g_i. Entretanto, com Z_{agr} ocorre o inverso. Por outro lado, nas famílias em que T_{nag} é positivo, são as outras variáveis que afetam o mercado de trabalho agrícola (Z_{agr}), que possuem sinal positivo e seguem a mesma direção de f_i.

2.2 Crescimento econômico, Pobreza e Distribuição da renda

Após modelar o domicílio, buscando entender a decisão da família em se engajar nas possíveis fontes de emprego, com o objetivo de maximizar a função de utilidade familiar, é analisado o efeito dessa decisão sobre a pobreza e a concentração da renda. A relação entre crescimento econômico[28], pobreza e concentração não está formalizada num modelo econômico[29]. No atual estado da arte, também não existe consenso na literatura disponível sobre as reais relações entre os três fenômenos.

Uma forma[30] de investigar as relações entre estas variáveis, é analisar o comportamento de pares, ou seja, após eleger uma variável (crescimento, neste caso) estabelece-se a relação com outra (pobreza ou desigualdade). Nesta abordagem, a direção na condução da análise faz diferença. Pode-se buscar entender o efeito da redução da desigualdade da distribuição da renda no crescimento econômico, o que não é o mesmo que estudar o impacto do crescimento na distribuição da renda. Neste trabalho, o sentido de causalidade em que se conduz a análise, possibilita identificar dois efeitos: a) aquele relativo à mudança na renda média familiar sobre a pobreza; e b) o relativo à variação desta renda sobre a concentração.

Considerando, inicialmente, a relação entre crescimento econômico e pobreza, o entendimento primordial é sobre a necessidade de estabelecer políticas econômicas que compatibilizem crescimento com redução da pobreza. Se a camada da população menos favorecida consegue se apropriar mais dos benefícios do crescimento do que a mais rica, tem-se uma das definições de "*crescimento pró-pobre*". Segundo esta definição, para ser pró-pobre o crescimento deve resultar em redução da pobreza e desigualdade. Não se considera pró-pobre o caso de redução da pobreza sem redução de desigualdade (RAVALLION, 2004). Esta abordagem possui limitações, pois, seguindo esta forma de pensar o crescimento pró-pobre, uma sociedade seria favorável

[28] Normalmente, os estudos sobre esta temática utilizam como *proxy* para o crescimento econômico o PIB *per capita* ou a renda média familiar (GUIMARÃES, 2007). Nesta pesquisa, entende-se como crescimento a variação no tempo do nível de renda média.

[29] Contudo, possui um vasto arcabouço teórico relacionado à distribuição funcional da renda, ou seja, sua divisão entre as classes sociais, objeto de estudo desde os Fisiocratas com o "*Tableau Economique*", passando por Marx e Marshall, até mais recentemente com Kaldor. Uma revisão interessante pode ser encontrada em Alvarez (1996).

[30] Outra possibilidade é a análise conjunta das três variáveis. Para informações deste caso, vide Guimarães (2007).

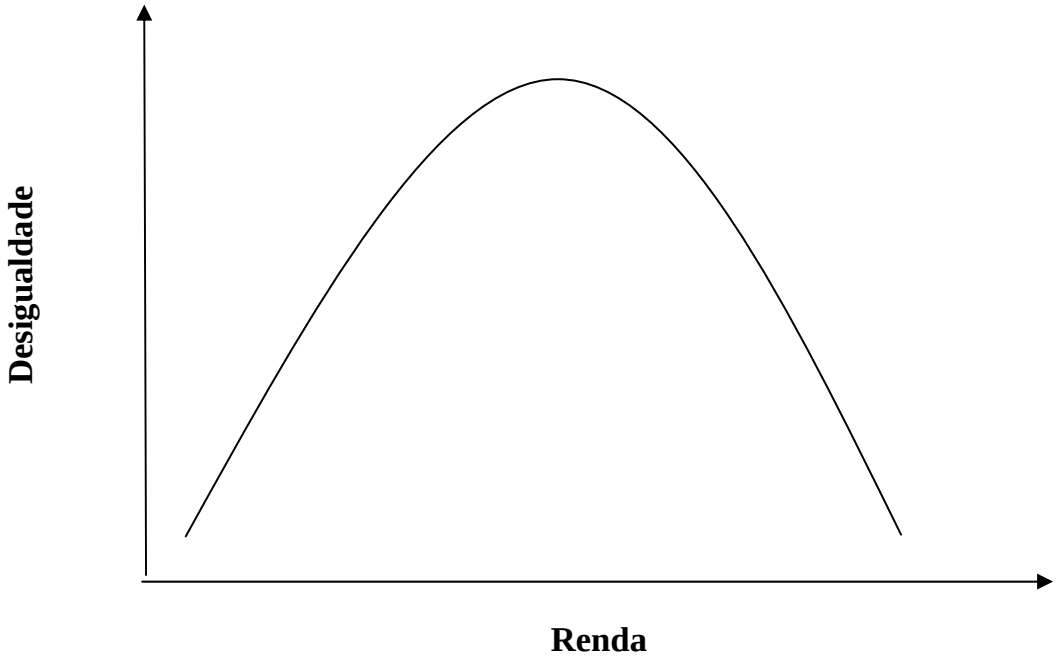
a um crescimento da renda média de 2% com aumento da renda dos mais pobres de 3% a outro conjunto de políticas que levasse ao crescimento da renda média em 6%, com elevação da renda dos mais pobres de 4% (LOPEZ, 2004).

Uma segunda definição para crescimento pró-pobre é a que considera ser necessário apenas que uma medida de pobreza de interesse (proporção de pobres, hiato da pobreza ou outra a ser definida) se reduza (RAVALLION, 2004; KRAAY, 2004). Esta medida também possui deficiências, pois supondo um crescimento de 6%, mesmo que a redução da proporção de pobres seja de apenas 0,1%, este é considerado pró-pobre, desconsiderando todas as implicações sociais da elevação da desigualdade na distribuição (LOPEZ, 2004). Alinhado com a segunda definição, Kraay (2004) identifica três potenciais fontes de crescimento pró-pobre: i) alta taxa de crescimento; ii) elevada elasticidade da pobreza[31] em relação ao crescimento da renda média; e iii) padrão de redução da pobreza via crescimento da renda relativa. Em seus resultados, o autor demonstra que no médio e longo prazo, a maior parte da variação na proporção de pobres decorre de crescimento na renda média. Para os outros índices de pobreza, como a incidência e a severidade, indica que a mudança na distribuição da renda é mais relevante do que o crescimento.

Quando se trata da relação entre crescimento e desigualdade, o trabalho clássico sobre a questão é o de Kuznets (1955). Este autor, na verdade, descreve uma relação entre estas variáveis, conhecida na literatura como Hipótese do U-invertido de Kuznets[32] (Figura 2). Segundo esta hipótese, a distribuição de renda se torna mais desigual no estágio inicial do desenvolvimento (país eminentemente agrícola que vai se industrializando) e vai reduzindo a desigualdade conforme o caminhar do processo de industrialização, devido à propagação dos ganhos de produtividade do setor industrial e à redução do peso relativo do setor agropecuário na economia.

[31] Uma das maneiras mais utilizadas para avaliar o crescimento pró-pobre é o cálculo das elasticidades do crescimento da pobreza em relação à pobreza. O conceito desta elasticidade e a forma de calcular estão definidos no item 3.2.

[32] Atualmente, muitas pesquisas estão sendo elaboradas com uma versão modificada da Curva de Kuznets original. Esta vem sendo utilizada para relacionar "poluição" e crescimento, sendo denominada de Curva Ambiental de Kuznets. Como exemplo, vide Arraes *et al* (2006).



Fonte: Guimarães, 2007.

Figura 2 – A hipótese do U-Invertido de Kuznets.

Kuznets (1955) utiliza informações de séries temporais para os Estados Unidos, Inglaterra e Alemanha. Diversos autores posteriormente buscaram comprovar a hipótese do U-invertido, modelando a relação entre desigualdade e crescimento da seguinte forma:

$$I = \hat{\beta}_0 + \hat{\beta}_1 Y + \hat{\beta}_2 Y^2 + \varepsilon$$

em que os β's são os parâmetros estimados; ε é o termo de erro estocástico; I é algum índice de desigualdade; e Y a renda *per capita*. O teste a ser feito é com relação à negatividade do sinal de $\hat{\beta}_2$ (LOPEZ, 2004).

Existem diversas críticas à hipótese de Kuznets e também diversos trabalhos que lhe dão suporte. Stewart (2000) sumariza da seguinte forma as explicações que se apresentam para a Curva de Kuznets:

a) Uma economia com um setor agrícola com renda menos concentrada e baixa produtividade, em conjunto com um setor industrial mais produtivo e com renda mais concentrada. A desigualdade aumenta, então, por dois fatores: as diferenças nas rendas médias entre setores e devido à mobilidade intra-setorial da mão-de-obra. Quando a economia se moderniza, a desigualdade desaparece inter e intra-setorial.
b) Outra explicação, derivada do modelo de crescimento de Lewis, é que a produtividade e o crescimento da renda estão restritos ao setor industrial, em que a participação nos lucros cresce, enquanto no setor agropecuário a renda permanece baixa, levando ao aumento da desigualdade entre os setores.
c) Uma última explicação considera que uma distribuição inicial desigual de ativos contribui para a crescente desigualdade, pois as pessoas com mais ativos também acumulam mais. A taxa de retorno do capital pode até se reduzir, mas não os efeitos da desigualdade.

Após demonstrar diversos casos de países em que o crescimento econômico leva a um aumento da desigualdade e à sua redução, Stewart (2000, p. 5) conclui que "o crescimento é neutro em relação à distribuição, ou seja, ele não leva necessariamente a uma piora ou melhora da distribuição de renda e pode ser consistente com ambos. Fatores estruturais e medidas de políticas determinam as experiências dos países".

3. METODOLOGIA

Esta pesquisa basicamente se divide em três partes. A primeira busca entender os determinantes da escolha da família em participar apenas de atividades agrícolas, não-agrícolas, ou ser pluriativa. Após identificar as variáveis que contribuem para a probabilidade de as famílias escolherem entre os tipos de ocupação, a segunda parte procura demonstrar se a pluriatividade favorece a redução da pobreza na família. Se a hipótese for confirmada, os formuladores de políticas públicas possuem mais uma frente a ser estimulada. A análise da terceira parte está centrada na relação entre as atividades e rendas não-agrícolas com a concentração de renda no meio rural. Esta última preocupação é relevante, ainda mais se não for confirmada a hipótese de que as rendas não-agrícolas contribuem para reduzir a desigualdade, haja vista que o desenvolvimento rural também requer maior igualdade.

Por meio do modelo de seleção amostral com logit multinomial, estuda-se a importância da pluriatividade, identificando os determinantes da escolha da família em participar apenas de atividades agrícolas, não-agrícolas, ou ser pluriativa. Por meio de simulações nas rendas médias estimadas, busca-se demonstrar a importância das rendas não-agrícolas para redução da pobreza no rural da região Nordeste pelo estudo dos índices FGT (Foster-Greer-Thorbecke) e elasticidade renda da pobreza. O impacto das rendas não-agrícolas sobre a desigualdade no meio rural é feito pela análise dos índices de Gini e pelo cálculo da elasticidade Gini da pobreza. A demonstração de que Orna difunde efeitos positivos na produção agropecuária dos estabelecimentos rurais

requer a comparação entre a renda observada e a simulada das famílias exclusivamente agrícolas, pluriativas e não-agrícolas.

3.1 Determinantes da escolha entre alternativas de ocupação

No interior da família, que busca a melhor estratégia de sobrevivência, os membros que a constituem podem se ocupar apenas em atividades agrícolas, não-agrícolas, ser pluriativos ou estar não-ocupados. No geral, ocupações são sempre consideradas variáveis nominais ou cujas categorias não possuem uma ordem natural, e tampouco podem efetivamente ser ranqueadas (HOFFMANN, 2004).

A cada uma das alternativas/categorias se associa uma utilidade, que a família obtém da decisão tomada. Esta utilidade é modelada como uma utilidade aleatória, ou seja, constituída de um componente determinístico e de um estocástico. Em outras palavras, a escolha de uma alternativa *j*, em que *j=1,...,J*, para uma família *i*, com *i=1,...,I*, visando a maximizar a utilidade U_{ij} é dada por:

$$U_{ij} = \overline{U}_{ij} + u_{ij} \qquad j=1,...,J,\ i=1,...,I; \tag{39}$$

em que $\overline{U}_{ij}$ (utilidade da família i de uma alternativa j) é o componente determinístico da função de utilidade e pode ser interpretado como uma função de utilidade indireta, sendo idêntico para todas as famílias, com o termo u_{ij} sendo o componente estocástico.

A família *i* escolhe a alternativa *j* em relação à *k*, se $U_j > U_{k,}$. O modelo estatístico que reflete esta maior utilidade evidencia-se pela probabilidade de escolha da alternativa *j,* que ocorre quando

$$Prob(U_{ij} > U_{ik,}), \qquad k \neq j \tag{40}$$

O modelo se torna operacionalizável ao escolher a distribuição do erro, sendo a distribuição logística mais utilizada (GREENE, 2003).

A função de utilidade indireta da família pode ser definida como a função linear[33]:

$$\overline{U_{ij}} = \alpha_j + \beta'_j x_i \qquad j=1,...,J,\ i=1,...,I \qquad (41)$$

em que α_j é um termo constante para cada alternativa, x'_i é o vetor de variáveis relacionadas às características da família *i* e tudo mais relevante que afete a escolha.

O processo de escolha por parte das famílias é simultâneo e considera as alternativas mutuamente exclusivas, ou seja, a família escolhe apenas uma das opções. Se os *j* termos de perturbação (u_{ij}) são IID, ou seja, independentes e identicamente distribuídos, com uma distribuição Weibull, tem-se o modelo logit multinomial[34],

$$Prob(Y_i = j|x_i) = \frac{e^{\beta' x_{ij}}}{\sum_{k=0}^{J} e^{\beta' x_{ik}}} \qquad j = 0,\ldots,J, i = 1,\ldots,I \qquad (42)$$

O logit multinomial descreve a probabilidade de certo evento *k* ocorrer, determinada por um conjunto de características incluídas no vetor x_j. A partir da estimação da equação (42), é gerado um conjunto de probabilidades para as J escolhas para a família *i*, e surge uma indeterminação. Para resolver este problema, é feita uma normalização, em que uma alternativa é definida como referência. Supondo, por exemplo, $\beta_1 = 0$, obtém-se:

$$Prob(Y_i = j|x_i) = \frac{e^{\beta' x_{ij}}}{1 + \sum_{k=1}^{J} e^{\beta' x_{ik}}} \qquad j = 0,2,\ldots,J; i = 1,\ldots,I, \beta_1 = 0 \qquad (43)$$

[33] McFadden *apud* (DOW, 1999) tem usado este tipo de especificação para ilustrar um possível conjunto de condições suficientes para consistência com as hipóteses do modelo de utilidade aleatória.

[34] É possível também a estimação de um modelo probit multinomial. O *software* Stata, por exemplo, possui essa opção. Porém, o custo computacional de utilizar o probit multinomial é muito elevado, sendo o logit multinomial muito mais utilizado. Pode-se encontrar uma breve discussão sobre as dificuldades de utilizar o probit multinomial em Greene (2003).

em que a estimação de (43) é feita por meio do método de máxima verossimilhança[35] e o número de parâmetros a ser estimado é igual ao número de características individuais (idade, escolaridade etc.) multiplicado por *j-1*, em que *j* é o número total de ocupações (MADDALA, 1986). O modelo logit multinomial pode ser pensado como um conjunto de modelos logits binários estimados simultaneamente[36] para comparação entre todas as alternativas (LONG e FREESE, 2006).

Além da possibilidade de estimar a probabilidade de o indivíduo "*i*" escolher a alternativa "*k*", se podem ser também analisados, por variável explicativa, os efeitos parciais, ou seja, dada uma variação (discreta) na variável explicativa, quanto isto eleva ou reduz a probabilidade de o indivíduo "*i*" escolher a alternativa "*k*".

A interpretação dos coeficientes estimados nos modelos do tipo logit ou logit multinomial não é usual por ser mais difícil sua interpretação. Os sinais dos coeficientes estimados são utilizados para indicar as direções das chances ou probabilidades, quando ocorrem modificações nas variáveis explicativas, apesar de existirem outros fatores que também afetam estas direções (WOOLDRIDGE, 2002).

Normalmente se analisam as razões de chances, que no logit multinomial é denominado de razão relativa de risco (RRR) [37] e as probabilidades. A razão relativa de risco, por outro lado, retrata a variação na razão de probabilidades de escolha entre as diversas alternativas, dada uma alteração na variável explicativa (tudo mais constante):

[35] Na realidade, como pode ser observado no Anexo A, quando se trata de Amostra Complexa, as estimações são feitas pelo método da máxima pseudo-verossimilhança.

[36] O primeiro problema da estimação de vários logits binários é a quantidade de modelos que devem ser estimados quando se possui um grande número de categorias e se querem comparações entre elas. Uma segunda questão que também demonstra a ineficiência desta opção é que os erros-padrão tendem a ser maiores, pois existirão diferentes tamanhos de amostras e uma variedade de distribuições binomiais. Porém, apesar de esta abordagem ser ineficiente, ajuda na compreensão da interpretação das estimativas (HOFFMANN, 2004).

[37] É calculada facilmente tomando o exponencial do beta estimado. É a mesma coisa que a razão de chances (*odd-ratio*) do modelo logit binário, mas no multinomial é denominado de razão relativa de risco.

$$RRR = \frac{\frac{\Pr ob(Y = j \mid x+1)}{\Pr ob(Y = k \mid x+1)}}{\frac{\Pr ob(Y = j \mid x)}{\Pr ob(Y = k \mid x)}} \tag{44}$$

A variável dependente do modelo logit multinomial é policotômica, assumindo o valor 0 se a família for agrícola; 1, se for não agrícola; 2, quando for pluriativa; e 3, para os não-ocupados. As variáveis explicativas do modelo são: idade média da PEA restrita[38]; idade média da PEA restrita ao quadrado; média de anos de estudo; número de componentes da família; duas *dummies* para posição na ocupação, definida como 1 para conta-própria e 0, caso-contrário, além de 1 para empregados e 0, caso-contrário; e variável *dummy* para local de moradia, assumindo valor de 1 para o rural mais distante (agropecuário) e 0, caso-contrário. A razão de dependência é formada dividindo o número de membros da família considerados dependentes pela idade da PEA ativa, enquanto para a *proxy* para os insumos fixos da propriedade, são utilizadas duas variáveis relacionadas à existência de água proveniente de poço na propriedade e o tipo de iluminação que a propriedade possui. A água de poço é uma grande "riqueza" dentro de uma propriedade no rural do Nordeste, valorizando-a bastante, junto com a energia elétrica; uma variável *dummy* indicando se a família possui ou não renda do não-trabalho, como aposentadorias, pensões ou transferência de renda do governo, por exemplo; a renda *per capita*; e, finalmente, variáveis *dummy,* uma para cada estado nordestino.

Sobre os sinais, o esperado, com relação à idade média da PEA restrita, é um comportamento em forma de "U" invertido, tanto para as famílias não-agrícolas como para as pluriativas. Esta forma quadrática indica que o aumento da idade eleva a chance de as famílias possuírem algum membro ocupado em atividade não-agrícola até certo ponto, posteriormente esta chance começa a se reduzir. Com relação aos anos de estudo, o sinal esperado positivo indica que quanto maior o número de anos de estudo, maior a chance de algum membro estar ocupado em atividade não-agrícola.

[38] PEA restrita se refere às pessoas com 10 anos ou mais que trabalharam mais de 15 horas na semana de referência , excluídas as que se dedicam apenas ao auto-consumo.

O sinal esperado para o número de componentes da família é positivo, ou seja, uma família mais numerosa teria mais condições de liberar alguém para se ocupar em atividade não-agrícola. O local de moradia possui relação inversa com a chance de ocupação em atividade não-agrícola, pela maior distância que a pessoa está do mercado de trabalho não-agrícola. O sinal da razão de dependência realmente é ambíguo, como também considera Andrade (2003). Uma razão maior de dependência, ou seja, um maior número de dependentes em relação à quantidade de pessoas em idade ativa, pode tanto levar a família a optar por uma ocupação não-agrícola para elevar a renda e auxiliar na sua manutenção quanto a permanecer exclusivamente agrícola devido à "falta de braços" na propriedade).

Espera-se que a renda do não-trabalho reduza a chance de a família possuir atividade não-agrícola, ou seja, sinal negativo. Para a renda *per capita*, o sinal esperado é positivo, ou seja, espera-se que as famílias com renda mais elevada possuam ocupação não-agrícola. Finalmente, sobre as variáveis relacionadas aos insumos fixos da propriedade, o sinal esperado, no caso da região Nordeste, é negativo. Espera-se que possua relação inversa com a chance de algum membro estar ocupado em atividade não-agrícola. Esta seria a situação em que o lucro da atividade agrícola é baixo, estimulando a busca por diversificação das fontes de renda.

3.2 Efeitos da renda não-agrícola sobre a pobreza

Para analisar o efeito da renda não-agrícola sobre a pobreza, o primeiro passo é dado na estimação do modelo logit multinomial. O segundo passo é estimar a equação da renda média. Então, pode-se comparar a distribuição observada da renda total das famílias com outra, obtida por simulação. Diversos tipos de simulações podem ser feitas. Para as famílias exclusivamente agrícolas, existe a possibilidade de simular qual seria sua renda caso se tornassem pluriativas ou não-agrícolas. Nas famílias pluriativas, a simulação possibilita entender qual seria a renda caso se dedicassem exclusivamente às atividades agrícolas ou às não-agrícolas. Finalmente, para as famílias não-agrícolas, quais as rendas esperadas se passassem para exclusivamente

agrícola ou pluriativa. Não faz sentido analisar qual a renda da família pluriativa, caso estivesse não-ocupada, por exemplo. Disto depreende-se que todas as informações sobre as famílias de não-ocupados são importantes e utilizadas tanto no modelo que explica os determinantes de ocupação quanto no que estima as rendas médias. Porém, não são feitas simulações com esse tipo de família.

A estimação do modelo explicativo das rendas médias das famílias agrícolas, por exemplo, pode conter viés de seleção amostral, já que a amostra é composta das rendas de quatro tipos de famílias. Da mesma forma, isto pode ocorrer no caso da estimação das rendas médias dos outros tipos de famílias.

Toda literatura a que se teve acesso para confecção deste trabalho restringe os tipos de ocupação em duas: possui e não possui emprego não-agrícola. Então, como a variável "tipos de ocupação" é dicotômica, a estimação do modelo é feita com base no procedimento de Heckman (1979) de dois passos. No caso deste trabalho, esta restrição não é feita. A variável "tipos de ocupação" é policotômica, já que existem quatro tipos de famílias (agrícola, não-agrícola, pluriativa, não-ocupada). Como não se pode utilizar o procedimento de Heckman, já que a variável dependente da equação de seleção possui quatro categorias, o modelo utilizado é o de seleção amostral com logit multinomial. Lee (1983) sugere uma generalização do procedimento definido por Heckman (1979), estendendo-o para o caso em que o modelo de seleção amostral é o logit multinomial. Segundo Lee (1983), o segundo passo do procedimento é estimar

$$E(W_i|Y_i = j) = \gamma_i' Z_i + \sigma\rho_j \frac{\phi(f_{1i}(\beta_j' X_i))}{\Phi_s(\beta_j' X_i)} + v_i \tag{45}$$

em que W_i é a renda da família i e tipo j; γ_i é o conjunto de parâmetros a serem estimados; Z_i é o conjunto de variáveis explicativas da renda e que pode ser um subconjunto de X_i; σ é o erro padrão do erro aleatório; e ρ_j é a correlação entre o erro aleatório v_i e o $\varepsilon_i|(Y_i = j)$, que não precisa estar restrita a -1 e +1. Se esta correlação existe, o viés de seleção precisa ser corrigido. Os valores de $\beta_j' X_i$ são obtidos no primeiro passo, ao ser estimado o logit múltiplo; ϕ é a função de densidade de

probabilidade normal padrão (f) avaliada no ponto $(J_{1i}(\beta_j' X_i))$; J_{1i} é a inversa da função de distribuição de probabilidade normal padronizada acumulada (F^{-1}) avaliada no ponto $\beta_j' X_i$, necessário já que os resíduos deste modelo seguem a distribuição Weibull; e $\Phi_s(\beta_j' X_i)$ é a função de distribuição de probabilidade normal padronizada acumulada avaliada no ponto $\beta_j' X_i$.

Pelo menos três equações de renda podem ser estimadas (famílias agrícolas, não-agrícolas e pluriativas) e a partir dos resíduos, ser feitas as simulações citadas no início deste tópico. Inicialmente, para exemplificar a forma de simular as rendas, considera-se W_{0i} a renda de cada família *i* que participa apenas de atividades agrícolas[39]. Sabe-se que:

$$W_{0i} = E(W_{0i}|Y_{0i} = f) + v_{0i} = \gamma_i' Z_i + \sigma\rho_j\varphi + v_{0i} \qquad (46)$$

em que φ é igual a $\frac{\phi(J_{1i}(\beta_j' x_i))}{\Phi_s(\beta_j' x_i)}$, que se assemelha à razão inversa de Mills no Procedimento de Heckman com modelo binário; $E(W_{Ci}|Y_{Ci} = f)$ é o valor esperado da renda condicionado às características observadas pelas famílias e estas serem exclusivamente agrícolas; e v_{0i} se refere às características que afetam a renda mas que não são observadas. O interesse é prever a renda de cada família *i* exclusivamente agrícola. Para as famílias agrícolas esta é a sua renda observada. Para as famílias pluriativas e não-agrícolas, esta é a renda prevista que perceberiam caso se dedicassem apenas às atividades agrícolas. Esta previsão requer que $E(v_{0i}|Y_i) = 0$ e var $(v_{0i}|Y_i) = \sigma_0^2$.

Utilizando os parâmetros estimados, pode-se prever a renda W_{0i} para todas as *i* famílias (incluindo agrícolas, não-agrícolas e pluriativos). Para prever o valor esperado condicionado da renda ($\hat{E}W_{0i} = \hat{\gamma}_i' Z_i + \hat{\sigma}\hat{\rho}_j'\hat{\varphi}$) para os grupos de famílias exclusivamente não-agrícolas e pluriativas, é necessário gerar termos não-observados $\hat{v}_{0i}$.

[39] A notação segue W_{0i} para as famílias exclusivamente agrícolas, W_{1i} para aquelas pluriativas e W_{2i} para as que são exclusivamente não-agrícolas.

Segundo De Janvry *et al.* (2005) e Zhu e Luo (2006), para isto se constrói uma variável aleatória

$$v_{0i}^{*} = \delta_0 \Phi^{-1}(r) \tag{47}$$

em que δ_0 é o erro padrão estimado de v_{0i} (famílias agrícolas); r se refere a uma seqüência de números aleatórios entre 0 e 1; e Φ^{-1} é o inverso da função de distribuição normal padronizada acumulada. Ao final, têm-se as previsões da renda caso se dediquem exclusivamente às atividades agrícolas para todas as *i* famílias:

$$\hat{W}_{0i} = \begin{cases} W_i = \gamma_i' Z_i + \delta \hat{\rho}_j \hat{\varphi} + v_{0i} & quando\ Y_i = 0 \\ \gamma_i' Z_i + \delta \hat{\rho}_j \hat{\varphi} + v_{0i}^{*} & quando\ Y_i = 1 \\ \gamma_i' Z_i + \delta \hat{\rho}_j \hat{\varphi} + v_{0i}^{*} & quando\ Y_i = 2 \end{cases} \tag{48}$$

Este procedimento deve ser repetido para obter $\hat{W}_{1i}$ e $\hat{W}_{2i}$, de forma que se consiga comparar a renda de todos os tipos de famílias caso fossem exclusivamente agrícolas, exclusivamente não-agrícolas e pluriativas. Da comparação da renda com e sem a presença de atividades não-agrícolas, dentre classes decompostas pelo índice de Foster-Greer-Thorbecke (FGT) (FOSTER *et al.*, 1984), estuda-se o impacto do não-agrícola sobre a pobreza. Espera-se que os índices de pobreza se reduzam quando a família possui renda não-agrícola e aumentem se as famílias forem exclusivamente agrícolas.

Fazem parte da classe de índices FGT a Proporção de pobres (P_0), o Hiato da Pobreza (P_1) e a Severidade da Pobreza (P_2). O índice proporção de pobres mede a proporção de famílias que possuem renda *per capita* familiar não superior à linha de pobreza. Este indicador é importante, mas possui limitações[40] e por isto deve ser utilizado em conjunto com os outros dois indicadores. O índice hiato da pobreza mede a intensidade da pobreza e pode ser interpretado como um indicador do déficit de

[40] O indicador permanece o mesmo quando a renda se eleva sem alcançar a linha da pobreza ou mesmo se reduz entre os pobres (HOFFMANN, 1998). A proporção também é insensível à distribuição de renda entre os pobres, ou seja, permanece inalterada se houver transferência de renda entre um mais e outro menos pobre (MARIANO e NEDER, 2004).

pobreza. O índice hiato da pobreza ao quadrado dá um maior peso para as pessoas mais pobres e leva em conta a desigualdade de renda entre os pobres, sendo geralmente descrito como um indicador de severidade da pobreza. Os dois últimos índices enfatizam as famílias que estão muito abaixo da linha de pobreza predeterminada (*z*), ou seja, os mais pobres entre os pobres (MARIANO e NEDER, 2004). Como demonstrado em Hoffmann (1998), os índices FGT são calculados segundo a seguinte expressão:

$$\varphi(\alpha) = \frac{1}{nz^{\alpha}} \sum_{i=1}^{p} (z - x_i)^{\alpha} , \qquad \text{com } \alpha \geq 0 \qquad (49)$$

em que,

p é o número de pobres, ou seja, famílias com renda *per capita* não superior à linha de pobreza predeterminada;

n é o tamanho da população;

z é a linha da pobreza predeterminada; e

$\mathbf{x_i}$ é um vetor de renda *per capita* familiar da *i-ésima* família em ordem crescente.

Essa medida é igual à proporção de pobres (P_0) quando $\alpha = 0$; igual ao hiato da pobreza (P_1), quando $\alpha = 1$; e igual ao hiato da pobreza ao quadrado ou severidade da pobreza (P_2) quando $\alpha = 2$. Pode-se observar que os índices FGT variam de 0 a 1, com $\varphi(\alpha) = 0$ quando todas as famílias possuem renda maior do que z e $\varphi(\alpha) = 1$ quando todas as rendas são iguais a zero (HOFFMANN, 1998).

Após calcular os índices de pobreza, uma análise interessante é a estimação das elasticidades crescimento-renda da pobreza, ou seja, o que ocorre com os índices de pobreza se houver uma variação no crescimento da renda. A metodologia se baseia na Curva de Lorenz para determinar os índices de pobreza, sendo proposta por Datt (1998) e adaptada aos microdados da Pnad por Neder (2003b), seguindo a metodologia de Kakwani (1990) para a derivação das fórmulas das elasticidades[41].

[41] Hoffmann (2005) também calcula elasticidades da pobreza em relação ao rendimento médio e ao índice de Gini. Porém, segue outro método, considerando que a distribuição da renda é log-normal. Como isto é bastante controverso, opta-se, nesta pesquisa, por seguir o método adaptado de Neder (2003b).

O processo inicia-se, então, com a escolha de uma especificação para a Curva de Lorenz, a qual será estimada por regressão linear com o método de mínimos quadrados ordinários. Para simplicidade dos cálculos, normalmente a forma quadrática é a escolhida,

$$L(1-L) = a(p^2 - L) + bL(p-1) + c(p-L) \tag{50}$$

ou

$$L(p) = -\frac{1}{2}[bp + e + (mp^2 + np + e^2)^{1/2} \tag{51}$$

em que L é a participação dos *p* por cento inferiores da população na renda familiar; e *a*, *b* e *c* são parâmetros estimáveis da Curva de Lorenz. Os parâmetros *e, m e n* são assim definidos por Datt (1998):

$$e = -(a+b+c+1) \tag{52}$$
$$m = b^2 - 4a \tag{53}$$
$$n = 2be - 4c \tag{54}$$

As medidas de pobreza se restringem aos índices FGT e podem ser obtidos a partir da supracitada especificação da Curva de Lorenz (50) e (51) pelas expressões[42]

$$P_0 = -\frac{1}{2m}\left[n + r\left(b + \frac{2z}{\mu}\right)\left\{\left(b + \frac{2z}{\mu}\right)^2 - m\right\}^{-1/2}\right] \tag{55}$$

$$P_1 = P_0 - \left(\frac{\mu}{z}\right)L(P_0) \tag{56}$$

$$P_2 = 2(P_1) - P_0 - \left(\frac{z}{z}\right)^2\left[aP_0 + bL(P_0) - \left(\frac{r}{16}\right)ln\left(\frac{1-P_0/z_1}{1-P_0/z_2}\right)\right] \tag{57}$$

[42] As relações entre as expressões e suas derivações podem ser encontradas detalhadamente em Guimarães (2007).

em que P_0, P_1 e P_2 já foram definidos e os demais parâmetros são

$$r = (n^2 - 4me^2)^{\frac{1}{2}} \tag{58}$$

$$s_1 = \frac{r-n}{2m} \tag{59}$$

$$s_2 = -\frac{r+n}{2m} \tag{60}$$

Segundo Neder (2003b) e Guimarães (2007), os valores das elasticidades crescimento renda da pobreza para os índices FGT são calculados a partir das seguintes fórmulas,

$$\varepsilon_{P_0} = -\frac{z}{\mu P_0 L''(P_0)} \tag{61}$$

$$\varepsilon_{P_1} = 1 - \frac{P_0}{P_1} \tag{62}$$

$$\varepsilon_{P_2} = 2\left(1 - \frac{P_1}{P_2}\right) \tag{63}$$

sendo $L''(P_0)$ o valor da segunda derivada da Curva de Lorenz avaliada em P_0, definida por

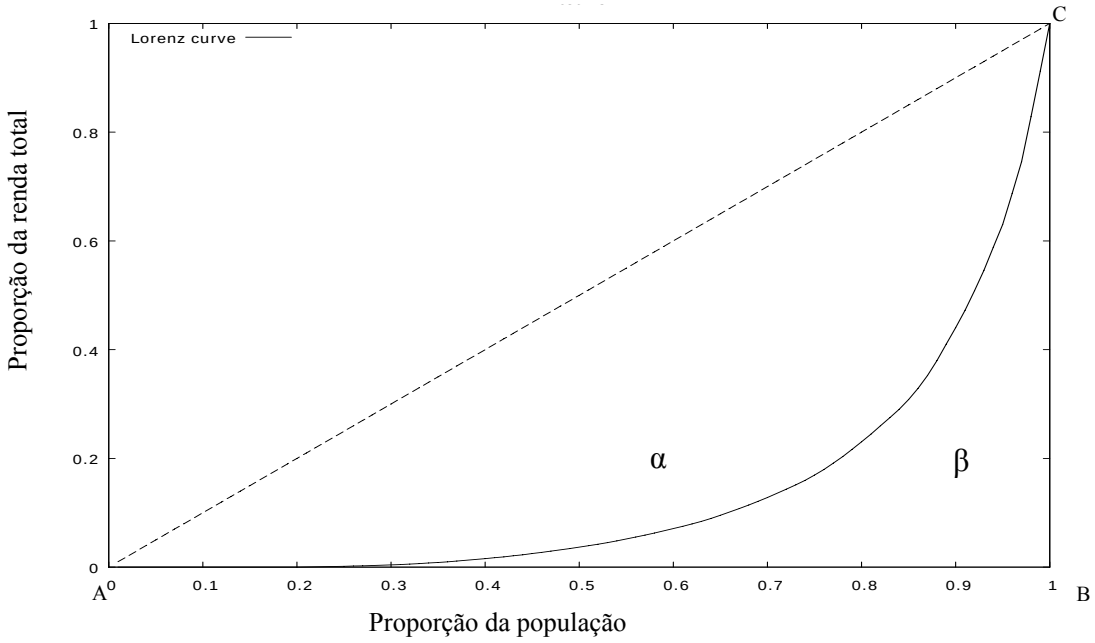
$$L''(P_0) = \frac{r^3(mP_0^2 + nP_0 + e^2)^{-3/2}}{8} \tag{64}$$

3.3 Efeitos da renda não-agrícola sobre a concentração

Se, por um lado, pode-se considerar que as rendas não-agrícolas contribuem para reduzir a pobreza, por outro, estas rendas podem ainda contribuir na desconcentração de renda. Como cita Hoffmann (2001), o aumento da renda de uma pessoa pobre, ou com baixas rendas, reduz a desigualdade. Existe uma relação negativa entre a renda dos mais pobres e a desigualdade. Isto é derivado da condição de Pigou-Dalton, a qual estabelece que as medidas de desigualdade devem ter seus valores aumentados quando há transferências regressivas de renda, ou seja, quando

ocorre transferência de renda do mais pobre para o mais rico ou uma transferência que cria desigualdade entre duas famílias que possuíam rendas iguais (HOFFMANN, 1998).

A partir da renda observada e da simulada obtida da equação de regressão (48), podem ser calculados dois índices de Gini. Este índice mede a desigualdade relativa da distribuição de renda pela razão entre a *área da desigualdade* (α) e a área de uma distribuição de perfeita igualdade da Curva de Lorenz[43]. A Curva de Lorenz mostra como a proporção da renda total (Φ) aumenta em função da proporção da população (*p)*. Se todas as famílias tivessem a mesma renda (igual a média das rendas), *p* seria sempre igual a Φ, como demonstra a linha pontilhada na Figura 3.



Fonte: Hoffmann (1998).

Figura 3: Curva hipotética de Lorenz.

Nesta figura, a Curva de Lorenz seria o segmento de renda AC, denominado de *linha da perfeita igualdade*. Caso toda a renda fosse apropriada por apenas uma família e a renda das demais fosse zero, no limite, a Curva de Lorenz seria formada pelos pontos ABC, denominada de *linha da perfeita desigualdade*. Quanto mais

[43] $G = \frac{\alpha}{0,5} = 2\alpha$, como $0 \leq \alpha \leq 0,5$, tem-se que o índice de Gini varia entre 0 e 1 (HOFFMANN, 1998).

afastada a Curva de Lorenz estiver da linha de perfeita igualdade, maior o grau de desigualdade da distribuição, sendo esta área denominada de *área de desigualdade*, indicada pela letra grega α. A letra β representa a área entre a Curva de Lorenz e o eixo das abscissas (HOFFMANN, 1998).

A associação direta com a Curva de Lorenz é uma das vantagens do índice de Gini. Dentre as diversas fórmulas existentes, a utilizada nesta pesquisa está demonstrada em Hoffmann (1998),

$$G = \frac{2}{n^2 \mu} \sum_{i=1}^{n} i x_i - \left(1 + \frac{1}{n}\right) \tag{65}$$

em que μ é a renda média; n é o número de observações; e x_i são as rendas.

Exemplificando, no caso da comparação entre as rendas das famílias pluriativas e exclusivamente agrícolas, se o índice obtido dos valores observados, que incluem as rendas não-agrícolas, for menor do que o simulado (sem as rendas não-agrícolas), isto significa que a presença da renda não-agrícola desconcentra a renda. Os erros-padrão, viés e intervalo de confiança dos índices são obtidos pelo método de *Linearização de Taylor (Método Delta)* e o de reamostragem *Bootstrap*, os quais são definidos no Anexo A.

De forma semelhante ao feito para a análise da pobreza, nos estudos de concentração pode-se calcular o impacto (em percentual) na pobreza de uma variação percentual no índice de desigualdade (Gini), ou seja, a elasticidade Gini da pobreza. Como demonstrado em Datt (1998) e Guimarães (2007), a partir da equação (50), pode-se calcular o índice de Gini pelas fórmulas[44]

$$G = \frac{a}{2} - \frac{n(b+2)}{4m} + \frac{r^2}{3m\sqrt{-m}} \left[sen^{-1} \frac{(2m+n)}{r} - sen^{-1} \frac{n}{r} \right] \qquad \text{quando m < 0} \tag{66}$$

e

$$G = \frac{a}{2} - \frac{n(b+2)}{4m} - \frac{r^2}{8m\sqrt{m}} ln \left[abs \left(\frac{2m+n+2\sqrt{m}\,(a+c-1)}{n-2a\sqrt{m}} \right) \right] \qquad \text{quando m > 0} \tag{67}$$

[44] Segundo Datt (1998), para a especificação quadrática da curva de Lorenz, as fórmulas para o índice de Gini só são válidas se a+c ≥ 1.

Como demonstrado em Kakwani (1990), Datt (1998), Neder (2003b) e Guimarães (2007), os valores da elasticidade Gini da pobreza podem ser obtidos pelas seguintes equações,

$$\xi_{P_0} = \left(1 - \frac{z}{\mu}\right) / (P_0 L''(P_0)) \tag{68}$$

$$\xi_{P_1} = 1 + \left(\frac{\mu}{z-1}\right) \frac{P_0}{P_1} \tag{69}$$

$$\xi_{P_2} = 2\left[1 + \left(\frac{\mu}{z-1}\right) \frac{P_1}{P_2}\right] \tag{70}$$

Tanto na análise de pobreza quanto na de concentração é utilizado, para auxiliar a identificar se a renda não-agrícola concentra ou não a renda e se reduz ou não a pobreza, um procedimento não-paramétrico de estimação de função de densidade denominado método de Kernel.

No método paramétrico, busca-se inferir sobre a população com base nos dados da amostra. Supõe-se conhecer a função de distribuição da qual os dados foram gerados, mas não o valor dos parâmetros desta função. Nesse processo, transforma-se a falta de conhecimento sobre uma característica da população em desconhecimento do parâmetro da função de distribuição (média e variância na distribuição normal, por exemplo). Nos métodos não paramétricos, não se faz necessário especificar a forma funcional da distribuição.

Toda variável aleatória possui uma distribuição, ou seja, aparece num intervalo de classe com "certa" freqüência e pode ser sintetizada por meio de métodos gráficos, como um histograma, por exemplo. Para construir um histograma, basta representar a variável de interesse no eixo horizontal e a freqüência no eixo vertical. A freqüência de cada classe é mostrada como o desenho de um retângulo (denominado de barra ou *bin*), cuja base é o intervalo da classe no eixo horizontal e a altura é a correspondente freqüência. O processo inicia-se com a escolha do ponto de partida e o desenho da

primeira barra (x_0+h). A segunda barra pode se iniciar onde termina a primeira e, assim, sucessivamente. A altura da barra é uma estimativa da densidade para um valor particular de x, definida por:

$$f(x) = \frac{1}{Nh}, \text{sendo N = número de obs. no intervalo (x, x + h)} \tag{71}$$

Ao fazer isso, tem-se idéia da densidade sem especificar, em nenhum momento, uma forma funcional para a distribuição. Diante disso, Johnston e DiNardo (1997) destacam que o histograma *"é a forma mais simples de estimação não paramétrica da densidade"*. Duas características importantes do histograma são: a) a estimativa da densidade depende da escolha da origem (x_0), pois diferentes pontos de partida originam histogramas diferentes, com abas mais ou menos longas; e b) o histograma não é contínuo, e sua forma depende da amplitude (h) das barras (quanto menos estreita as barras, menos o histograma será "suave", ou seja, parece ser mais recortado e com bicos nos dados em vários pontos).

Para solucionar os problemas relacionados com a dependência do valor de origem, não continuidade e falta de suavização do histograma, pode-se utilizar uma função de densidade de probabilidade da variável aleatória X, para um dado valor de h, definida como

$$\hat{f}(x) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x - X_i}{h}\right) \tag{72}$$

em que K é uma função de ponderação[45]; x é o centro do intervalo amostral (como se fosse o centro do intervalo de cada barra do histograma); e X_i representa cada observação do intervalo. A função $\hat{f}(x)$ passa a ser denominada de estimador Kernel da densidade de *Rosenblatt-Parzen*.

[45] Sendo Kernel uma função de pesos que satisfaz a seguinte propriedade $\int_{-\infty}^{\infty} K(x)dx = 1$ e determina a forma dos "ressaltos" (*bumps*). Existem diversos Kernels disponíveis nos *softwares*, com a Epanechnikov normalmente sendo a padrão.

Assim como o histograma, o estimador Kernel fornece uma estimação não-paramétrica da função de densidade. Contudo, não possui a propriedade indesejável de sua forma depender da escolha da origem e substitui os bicos por ressaltos suavizados (*smoothers*). O estimador é não paramétrico, mas a função estimada pode ser ponderada por diferentes núcleos (K), que são funções paramétricas conhecidas.

Quando K assume a função uniforme, tem-se o estimador simples (*naive*) de densidade. O problema desse estimador é que possui cantos nas barras e derivada zero no restante (menos suave). Por outro lado, as demais funções K admitem derivada.

Entretanto, o ponto chave para determinar o montante de suavidade não é o Kernel que vai ser escolhido, mas o valor da amplitude (h)[46]. Nesta pesquisa, opta-se pela escolha ótima feita pelo *software* Stata 10.1. Quanto maior a amplitude, mais suave será a estimativa e menor a variância, porém o enviesamento será maior (JOHNSTON e DINARDO, 1997).

Diante de certas hipóteses, o estimador Kernel é consistente, assintoticamente possui distribuição normal, mas é viesado em amostras pequenas. O viés está relacionado com a amplitude (h), porém, para uma escolha de h apropriada, o viés desaparece assintoticamente. A noção intuitiva, já citada anteriormente, é de que, quanto maior a amplitude, embora exista menor variância, o viés será maior (GREENE, 2003). O estimador Kernel possui, também, alguns obstáculos quando a densidade em questão possui caldas longas. Isso é causado pelo fato de a amplitude ser fixa para todas as observações, não considerando, assim, características locais dos dados.

3.4 Fontes dos Dados

Todas as informações estatísticas a serem empregadas neste trabalho foram obtidas da Pesquisa Nacional por Amostra de Domicílios (Pnad) do IBGE, para os anos de 2003 e 2005. A escolha por trabalhar com os dois anos se justifica por serem diferentes com relação à regularidade das chuvas, podendo viesar os resultados e as

[46] Técnicas de suavização são usadas para permitir a construção de um estimador da função da média condicional, sem fazer fortes hipóteses sobre o comportamento da função entre pontos (GREENE, 2003).

conclusões da pesquisa, se apenas um ano for objeto de estudo. A opção por trabalhar com dados agregados decorre basicamente de dois fatores. O primeiro, por uma questão de número de observações. Trabalhar com apenas um único estado, exclusivamente, é importante devido à heterogeneidade[47] existente na região, mas restringe o tamanho da amostra e, possivelmente, a qualidade das estimativas econométricas. Se a amostra é pequena, pode-se ter o problema da micronumerosidade.

O segundo motivo é que existe um importante fator comum (estrutural) em praticamente todas as cidades da região Nordeste, com menos de 50 mil habitantes[48], que são a pobreza[49] rural e urbana existente. Estes municípios se caracterizam "pela presença de pequenos centros urbanos cujas funções urbanas são precariamente exercidas, a maioria da população concentrada em atividades agrícolas ou em serviços de baixa qualificação, tendo baixíssima escolaridade e possuindo, por conseqüência, baixos índices de renda familiar" (IPEA/IBGE/NESUR, 2001, *apud* NASCIMENTO, 2005, p. 54).

3.4.1 Características básicas do plano amostral da Pnad

A Pnad é uma pesquisa anual realizada pela Fundação Instituto Brasileiro de Geografia e Estatística (IBGE), com exceção dos anos de Censo Demográfico. Utilizando de amostragem probabilística, tem o objetivo de produzir informações básicas para o estudo do desenvolvimento socioeconômico do país. A Pnad possui múltiplos propósitos e, por isto, analisa diferentes características socioeconômicas, como, por exemplo: educação, trabalho, rendimento, habitação, migração e fecundidade. A partir do ano de 2004, a Pnad inclui a área rural de Rondônia, Acre, Amazonas, Roraima, Pará e Amapá e passa a cobrir todo o território nacional.

[47] Os resultados obtidos considerando a região como um todo não são válidos para os estados de forma isolada ou qualquer outro tipo de desagregação.
[48] As cidades com menos de 50 mil habitantes correspondem a aproximadamente 92% do total das cidades da região, onde residem 84% das famílias rurais nordestinas, segundo o Censo Demográfico de 2000.
[49] O conceito de pobreza é analisado em maiores detalhes na segunda parte do trabalho.

A Tabela 7 trata da semana de referência, número de pessoas e de domicílios amostrados nos anos de 2003 (ano sem chuvas regulares) e 2005 (ano chuvoso), que são os anos de estudo deste trabalho. Os dados são sempre coletados no final do mês de setembro. Em 2003, foram entrevistadas, na forma face a face, 384.834 pessoas em 133.255 domicílios. Em 2005, foram entrevistadas 408.148 pessoas residentes em 142.471 domicílios. Maiores informações podem ser obtidas nas "notas metodológicas" que estão no CD com os micro-dados e junto à Coordenação de Trabalho e Rendimento da Diretoria de Pesquisas do IBGE.

Tabela 7. Informações Básicas da Pnad nos anos de 2003 e 2005

Ano	Semana de Referência	Número de Pessoas	Número de Domicílios
2003	21 a 27 de setembro	384.834	133.255
2005	18 a 24 de setembro	408.148	142.471

Fonte: Microdados da Pnad de 2003 e 2005.

Uma questão importante com relação à Pnad é que ela possui um desenho de amostragem complexa, englobando: estratificação, conglomeração, probabilidades desiguais de seleção e ajustes de pesos amostrais para calibração com os totais populacionais. Tratar a Pnad como IID significa ignorar o efeito do seu plano amostral e, com isto, as estimativas de significância ou construção de intervalos de confiança são calculados incorretamente. Para utilizar as informações do desenho amostral para realizar inferências é necessário conhecer em qual estrato e em qual unidade primária amostral (PSU) está localizado o domicílio da amostra e utilizar os pesos corretamente.

A estratificação da amostra básica da Pnad é feita em duas etapas. Inicialmente, há uma estratificação geográfica que divide o país em 36 estratos. Destes, 18 Unidades da Federação formam estratos independentes. As 9 unidades restantes geram outros 18 estratos, pois em cada um desses são definidos 2 estratos naturais: um com todos os municípios da região metropolitana sediada na capital e outro com os demais municípios, conforme se pode observar no Quadro 1.

Quadro 1 – Estratos geográficos da Pnad de 2003

Estratos Geográficos	
01 Distrito Federal	19 Pará
02 Rondônia	20 Região metropolitana de Belém
03 Acre	21 Ceará
04 Amazonas	22 Região metropolitana de Fortaleza
05 Roraima	23 Pernambuco
06 Amapá	24 Região metropolitana de Recife
07 Tocantins	25 Bahia
08 Sergipe	26 Região metropolitana de Salvador
09 Mato Grosso do Sul	27 Minas Gerais
10 Mato Grosso	28 Região metropolitana de Belo Horizonte
11 Goiás	29 Rio de Janeiro
12 Piauí	30 Região metropolitana do Rio de Janeiro
13 Rio Grande do Norte	31 São Paulo
14 Paraíba	32 Região metropolitana de São Paulo
15 Alagoas	33 Paraná
16 Espírito Santo	34 Região metropolitana de Curitiba
17 Santa Catarina	35 Rio Grande do Sul
18 Maranhão	36 Região Metropolitana de Porto Alegre

Fonte: Faria (2006).

Tanto para o ano de 2003 quanto para o de 2005, conforme consta nas notas metodológicas da Pnad, o Plano de Amostragem considera três estágios de seleção: a) Unidades Primárias (PSU), os municípios; b) Unidades Secundárias (SSU), os setores censitários; e c) Unidades Terciárias, os domicílios particulares e unidades de habitação em domicílios coletivos. Na seleção de PSU e SSU (municípios e setores censitários) da Pnad, foram adotadas a divisão territorial e a malha setorial vigentes em 1º de agosto de 2000 e utilizadas para a realização do Censo Demográfico de 2000.

No primeiro estágio, os municípios são classificados em auto-representativos e não auto-representativos. Os municípios não auto-representativos são estratificados e, em cada estrato, é selecionada, com reposição e com probabilidade proporcional à população residente, obtida no Censo Demográfico de 2000. No segundo estágio, são selecionadas as unidades (setores censitários), em cada município da amostra, também com probabilidade proporcional ao tamanho e com reposição. Neste caso, é utilizado, como medida de tamanho, o número de unidades domiciliares existentes por ocasião do Censo de 2000. No último estágio são selecionados, com igual probabilidade, em cada setor censitário da amostra, os domicílios particulares e as unidades de habitação em domicílios coletivos. A amostra é complementada com unidades domiciliares do cadastro de projetos de novas construções, ou seja, de projetos habitacionais com mais de 30 unidades domiciliares que surgiram após o Censo de 2000.

A Tabela 8 mostra as frações amostrais da Pnad para os anos de 2003 e 2005, por estrato geográfico. Existem diferenças entre os dois anos, porque algumas partes do território nacional não eram amostradas em 2003. Podem-se perceber frações amostrais para a região metropolitana da UF e o restante da UF, como é o caso do Pará, Ceará, Pernambuco e Bahia, na região Nordeste. O inverso das frações amostrais, demonstradas na Tabela 8, multiplicadas pelo fator de ajuste disponibilizado no arquivo de domicílios da Pnad constituem os pesos amostrais, necessários para o processo de expansão da amostra (projeção para a população).

Cabe ressaltar que o desenho amostral da Pnad visa a possibilitar a expansão dos seus resultados para Brasil, Grandes Regiões, Unidades da Federação e nove Regiões Metropolitanas (Belém, Fortaleza, Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo, Curitiba e Porto Alegre). Então, não é garantida a representatividade da amostra para níveis geográficos menores (município, distrito e setor) e demais regiões metropolitanas (GUIMARÃES, 2007).

Tabela 8. Frações Amostrais da Pnad nos anos de 2003 e 2005, por estrato geográfico

Estratos geográficos	Fração Amostral 2003	Fração Amostral 2005
Roraima (rural)	-	1/50
Acre (rural), Amapá (rural)	-	1/80
Rondônia (rural), Acre (urbana), Roraima (urbana), região metropolitana de Belém	1/150*	1/150
Rondônia (urbana), Amazonas (rural), Amapá (urbana), Distrito Federal, Tocantins, região metropolitana de Fortaleza, de Recife, de Salvador e de Porto Alegre	1/200**	1/200
Amazonas (urbana), Pará (rural)	1/250***	1/250
Sergipe, Mato Grosso do Sul, Mato Grosso, Goiás	1/300	1/300
Pará (urbana), região metropolitana de Belo Horizonte e de Curitiba	1/350	1/350
Rio de Janeiro	1/400	1/400
Rio Grande do Norte, Paraíba, Alagoas, Bahia e Espírito Santo	1/450	1/450
Piauí, Ceará, Pernambuco	1/500	1/500
Minas Gerais, região Metropolitana do Rio de Janeiro, Paraná, Santa Catarina e Rio Grande do Sul	1/550	1/550
Maranhão	1/750	1/750
São Paulo, região metropolitana de São Paulo	1/800	1/800

Fonte: Microdados da Pnad de 2003 e 2005.
* excluído Rondônia (rural); ** excluído Amazonas (rural); *** excluído Pará (rural).

Existem algumas estatísticas para avaliar o impacto da incorporação do plano amostral, denominado de EPA (**Efeito do Plano Amostral**). O primeiro a ser desenvolvido foi o **DEFF** (*design-effect*) por Kish em 1965, definido como a razão entre a estimativa da variância do parâmetro obtida considerando o plano amostral e a estimativa da variância do parâmetro, considerando que os dados fossem obtidos como amostra aleatória simples (GUIMARÃES, 2007). Posteriormente, é criado um conceito ampliado de DEFF, bem mais simples de calcular, denominado **MEFF** (*Misspecification Effect*). O MEFF compara a estimativa da variância do parâmetro obtida considerando o plano amostral com outra estimativa do mesmo modelo, só que desconsiderando peso, conglomerado e estratificação.

$$MEFF = \frac{\hat{V}(\hat{\theta})}{\hat{V}_{DPA}(\hat{\theta}_{DPA})} \quad (73)$$

Quanto mais o valor calculado do MEFF estiver distante da unidade, mais incorreta será a estimação da variância se não for considerado o plano amostral. Quando o valor do MEFF é superior à unidade, desconsiderar o plano amostral subestima a variância verdadeira (denominador menor do que o numerador). Se o MEFF é menor do que um, não considerar o desenho amostral superestima (inflaciona) a variância verdadeira (denominador maior do que o numerado). Os valores calculados para esta estatística são demonstrados nos resultados após as diversas estimações usando os dados da Pnad.

3.4.2 Outros conceitos e procedimentos utilizados

O IBGE considera urbana a população residente dentro dos perímetros urbanos, definidos pela legislação de cada município brasileiro. A definição de rural na Pnad é obtida de forma residual. São rurais os habitantes residentes fora deste perímetro urbano (DEL GROSSI e GRAZIANO DA SILVA, 2002).

Por outro lado, existe a possibilidade de desdobrar as áreas rurais a partir de critérios como número de domicílios e existência de serviços: a) *áreas rurais-extensão urbanas* são aquelas com distância inferior a 1 km do perímetro urbano; b) *áreas rurais-povoados* são aglomerações não vinculadas a um único proprietário, com área continuamente construída e reconhecível; c) *áreas rurais-núcleo* são aqueles cujo solo pertence a um único proprietário, com mais de 10 e menos de 51 domicílios, possuindo ou não serviços para atender aos moradores; d) *áreas rurais-outros* são os aglomerados que não possuem serviços para atender os moradores e não pertencem a um único proprietário; e e) *áreas rurais-exclusive* são as áreas rurais propriamente ditas (DEL GROSSI e GRAZIANO DA SILVA, 2002). Este fato contribui para a análise, já que se espera que quanto mais próximo o local de moradia for de uma área urbana, maior a possibilidade de ocorrência da pluriatividade. A questão da delimitação é controversa, porém este trabalho se concentra apenas no rural

propriamente dito (áreas rurais, exclusive), ou *rural agropecuário*, segundo a denominação dos trabalhos do projeto Rurbano.

Ao trabalhar com o conceito de pluriatividade, analisam-se todos os integrantes da família. Disto, a unidade relevante de análise passa a ser a família e não a pessoa, principalmente quando se busca captar as diversas combinações possíveis de atividades agrícolas e não-agrícolas. Existe também a possibilidade de trabalhar com domicílios. Na Pnad, tanto o conceito de família quanto o de domicílio possuem limitações severas. Na Pnad, para a definição de família considera-se apenas a unidade nuclear. Porém, em muitas situações existem mais de um casal vivendo sob o mesmo teto e que possuem dependência econômica entre si. Domicílio é o local de moradia estruturalmente separado e independente, constituído por um ou mais cômodos. Em muitos casos, juntam-se famílias independentes simplesmente por não possuírem uma entrada privativa em seus aposentos. Como os dois conceitos possuem estas limitações, é adotado no âmbito do projeto Rurbano e seguido nesta pesquisa, uma unidade de análise denominada de *família extensa*, a qual inclui além da família nuclear, os parentes que vivem no mesmo domicílio, mesmo que formem outro casal, e os agregados (DEL GROSSI e GRAZIANO DA SILVA, 2002).

O conceito pode ainda ser desdobrado em famílias do tipo empregadores, conta-própria, empregados e não-ocupados. Nas tabulações do Projeto Rurbano, como citado na nota de rodapé 7, as famílias foram classificadas da seguinte forma: empregador, conta-própria, empregados e "não-ocupados". Esta abertura contribui para identificar, dentro de cada tipo de família, a probabilidade de ser agrícola, não-agrícola ou pluriativa.

Outra ponderação relevante, já ressaltada na segunda parte desta pesquisa, é a análise das rendas agropecuárias. Como em todas as pesquisas do tipo declaratória, na Pnad tem-se a limitação da subestimação dos valores. Além disso, a forma com que o valor da renda variável agrícola é coletado contribui para viesar o valor. Segundo Del Grossi e Graziano da Silva (2002), a Pnad solicita que o declarante informe o valor médio mensal do rendimento bruto (se empregado) ou da retirada (conta-própria ou

empregador), do ano anterior. Para a resposta, é necessária a lembrança do tamanho da produção, preço de venda, custos, além da operação aritmética para obter o resultado.

A linha de pobreza utilizada neste trabalho é definida em ½ salário mínimo do período de referência, ou seja, R$ 120,00 para 2003 e R$ 150,00 em 2005, da mesma forma que Nascimento (2005) e Guimarães (2007).

O *software* utilizado para realizar a parte empírica da pesquisa é o Stata[50] 10.1. Para estimar o modelo de seleção amostral com logit multinomial para amostra de dados complexa, é necessário o código "*svyselmlog*" (BOURGUIGNON *et al.*, 2001). Os índices de pobreza são decompostos utilizando o código "*sepov"*. Para o cálculo da desigualdade, os códigos utilizados são "*ineqerr*" e "*svylorenz*". Todos os algoritmos Stata utilizados estão no Anexo B.

[50] Produzido e comercializado pela *StataCorp,College Station, Texas.* Website: www.stata.com

4. RESULTADOS E DISCUSSÕES

Nesta parte do trabalho são apresentados os resultados obtidos na estimação dos determinantes da escolha entre alternativas de ocupação, nos anos de 2003 e 2005. Posteriormente, são calculados os efeitos das atividades e rendas não-agrícolas sobre a pobreza e a concentração, mediante análise comparativa das rendas observadas e simuladas entre o estimador Kernel, os índices de pobreza (FGT), o índice de Gini e as elasticidades da pobreza.

4.1 Estatísticas descritivas das variáveis

Dentre outras variáveis, a Tabela 9 apresenta a média, erro padrão e o intervalo de confiança para as rendas das famílias agrícolas, não-agrícolas, pluriativas e dos não-ocupados, nos anos de 2003 e 2005. Especificamente para 2003 (ano em que na maior parte da região não ocorreram chuvas regulares), o número de observações (3410) ponderado pelo peso, representa, aproximadamente, 1,7 milhão de famílias exclusivamente agrícolas no meio rural do Nordeste. Com relação às não-agrícolas, a expansão para a população indica estar-se trabalhando com perto de 394 mil famílias não-agrícolas, ou seja, cerca de 24% do total das famílias agrícolas. O total de famílias pluriativas no meio rural nordestino é de aproximadamente 880 mil, representando em torno de 52% do total das agrícolas. Se for considerado o número de famílias que

possuem pelo menos um membro ocupado em atividades não-agrícolas, ou seja, somando o total de pluriativas com a de não-agrícolas, para cada quatro famílias exclusivamente agrícolas, haverá três com pelo menos um membro ocupado em não-agrícola. Com relação aos não-ocupados, o montante é de aproximadamente 151 mil famílias.

Tabela 9 – Estatística descritiva (média e erro padrão) das variáveis, considerando o plano amostral, Nordeste, 2003 e 2005

Variável	Número de Observações	Média	Erro Padrão Linearizado	Intervalo de Confiança (95%)	
Ano de 2003					
Renda agrícola	3410	348,7616	10,0321	328,8842	368,6389
Renda não-agríc.	848	517,3042	45,2654	427,5987	607,0097
Renda pluriativ.	1735	519,1723	15,3845	488,6742	549,6705
Renda não-ocup.	317	282,3149	16,2519	249,9904	314,6393
Idade PEA	6310	30,3306	0,1331	30,0670	30,5941
Anos de estudo	6310	2,9165	0,0981	2,7223	3,1108
N° Compon. da Família	6310	4,1647	0,0286	4,1080	4,2214
Razão dependência	6310	0,5705	0,0080	0,5546	0,5865
Ren. *per capita*	6310	414,7284	10,9222	393,0993	436,3574
Ano de 2005					
Renda agrícola	3437	423,1324	10,8712	401,6063	444,6586
Renda não-agríc.	951	669,4647	69,9406	530,8726	808,0567
Renda pluriativ.	1941	676,1959	21,4158	633,7752	718,6165
Renda não-ocup.	335	343,5012	17,5110	308,7374	378,2650
Idade PEA	6664	30,4205	0,1568	30,1099	30,7310
Anos de estudo	6664	3,2871	0,0985	3,0920	3,4822
N° Compon. da Família	6664	4,0539	0,0351	3,9843	4,1235
Razão dependência	6664	0,5302	0,0088	0,5127	0,5477
Renda *per capita*	6664	528,3328	15,8797	496,9049	559,7606

Fonte: Dados reprocessados da Pnad (2003) e Pnad (2005), pelo autor.

Por outro lado, no ano de 2003, conforme se observa na Tabela 9, a renda média das famílias pluriativas (R$ 519,17) é superior à renda média das famílias agrícolas (R$ 348,76) e estatisticamente igual à das não-agrícolas (R$ 517,30). A renda das famílias agrícolas compreende, neste ano, aproximadamente, 67% da renda das pluriativas/não-agrícolas. Chama ainda atenção que, na média, a quantidade média de anos de estudo das famílias é de aproximadamente 2,92 anos e a idade média da PEA é

30,33 anos. A média do número de componentes da família é aproximadamente de quatro pessoas.

Para o ano de 2005, de um total de 1,69 milhões de famílias agrícolas, existem cerca de 435 mil exclusivamente não-agrícolas e 980 mil pluriativas, no meio rural do Nordeste. Em termos percentuais, isto representa, respectivamente, 26% e 58% do total de agrícolas. Sendo o ano de 2005 considerado de chuvas regulares na maior parte da região, vale ressaltar este crescimento na proporção de famílias com pelo menos um membro ocupado em atividades não-agrícolas, comparado com as famílias exclusivamente agrícolas entre 2003 e 2005. Isto pode ser um indicativo de que a questão climática não seja determinante na decisão de a família possuir ou não um membro ocupado em atividades não-agrícolas.

A renda média das famílias pluriativas (R$ 676,20) é superior à renda média das famílias agrícolas (R$ 423,13) e estatisticamente igual à renda das não-agrícolas (R$ 669,46). Isto significa que a renda das famílias agrícolas compreende, em 2005, aproximadamente 63% da renda das pluriativas. Na média, a quantidade de anos de estudo média das famílias é de 3,29 anos e a idade média da PEA restrita é de 30,42 anos. A média do número de componentes da família é de aproximadamente quatro pessoas. A razão de dependência se reduz para 0,5302, quando se compara com a razão de 0,5705 do ano de 2003.

A Tabela 10 também está relacionada à estatística descritiva das variáveis, considerando o plano amostral, indicando a proporção dentre as categorias de cada variável. No ano de 2003, a maior parte das famílias recebe algum tipo de renda do não-trabalho (62,88%), como aposentadorias, pensões, transferência de renda do governo ou de algum parente que migrou, por exemplo. Aproximadamente 81% das famílias ou são do tipo conta-própria (47,93%) ou empregada (33,29%).

A grande maioria reside no rural denominado nos estudos do projeto Rurbano de “agropecuário” (84,21%), ou seja, o rural mais distante possível do urbano. Com relação à iluminação do domicílio, mais de 72% deles possuem luz elétrica. Por outro lado, um número considerável de 26,20% das famílias ainda possuem iluminação a

óleo, querosene ou gás de cozinha. Quanto à fonte de água, 29,15% possuem um poço, 32,28% não possuem e 38,56% desconhecem a informação (Tabela 10).

Tabela 10 – Estatística descritiva (Proporção e erro padrão) das variáveis, considerando o plano amostral, Nordeste, 2003

Variável	Proporção %	Erro Padrão Linearizado	Intervalo de Confiança (95%) %	
Ano de 2003				
Renda não-trabalho				
não-possui	37,1164	0,0126	34,6182	39,6145
possui	62,8835	0,0126	60,3854	65,3817
Conta-própria	47,9345	0,0146	45,0322	50,8369
Empregado	33,2912	0,0106	31,1902	35,3922
Local de moradia				
Rural Agropecuário	84,2132	0,0206	80,1210	88,3053
Iluminação domiciliar				
Elétrica	72,8291	0,0206	68,7377	76,9204
Óleo, querosene ou gás	26,2094	0,0203	22,1794	30,2395
Outra forma	0,9614	0,0026	0,4444	1,4783
Água de poço				
sim	29,1531	0,0192	25,3317	32,9745
não	32,2844	0,0206	28,1979	36,3708
ignorado	38,5624	0,0226	34,0692	43,0555

Fonte: Dados reprocessados da Pnad (2003), pelo autor.

A Tabela 11 reporta informações sobre as mesmas variáveis da Tabela 10, para o ano de 2005. A maior parte das famílias recebe algum tipo de renda do não-trabalho, 68,39%. Esta proporção é maior do que em 2003, indicando que mais famílias estão tendo acesso a esta fonte de renda não-agrícola[51]. Em torno de 80% das famílias ou são do tipo conta-própria (44,21%) ou empregada (35,54%).

A maior parte das famílias reside no "rural agropecuário", 83,50%. Este percentual é bem próximo do encontrado para o ano de 2003, sendo outro indicativo de que se deve relativizar o papel da questão climática no entendimento dos determinantes das decisões das famílias. Com relação à iluminação do domicílio, mais de 78% deles possuem luz elétrica. Por outro lado, houve redução, comparado com 2003 para 20,82%, no percentual de famílias que ainda possuem iluminação por outras fontes que não sejam a elétrica. Certamente que este fato se deve ao Programa

[51] Como citado na descrição das variáveis, fica claro que a fonte de renda é não-agrícola, mas não é proveniente de ocupação não-agrícola.

Nacional de Universalização do Acesso e Uso da Energia Elétrica para Todos – Programa Luz para Todos, instituído no final de 2003 pelo Governo Federal. A meta do programa é prover energia elétrica para toda a população rural do país até 2008. Com relação à água, 29,06% possuem um poço, 28,42% não possuem e 42,52% desconhecem a informação.

Tabela 11 – Estatística descritiva (Proporção e erro padrão) das variáveis, considerando o plano amostral, Nordeste, 2005

Variável	Proporção %	Erro Padrão Linearizado	Intervalo de Confiança (95%) %	
Ano de 2005				
Renda não-trabalho				
não-possui	31,6046	0,0118	29,2553	33,9538
possui	68,3954	0,0118	66,0461	70,7446
Conta-própria	44,2143	0,0140	41,4377	58,5622
Empregado	35,5389	0,0100	33,5588	37,5191
Local de moradia				
Rural Agropecuário	83,5033	0,0228	78,9796	88,0270
Iluminação domiciliar				
Elétrica	78,0650	0,0199	74,1200	82,0101
Óleo, querosene ou gás	20,8192	0,0204	16,7723	24,8662
Outra forma	1,1156	0,0021	0,6938	1,5374
Água de poço				
sim	29,0584	0,0191	25,2708	32,8460
não	28,4210	0,0198	24,4963	32,3457
ignorado	42,5205	0,0223	38,0970	46,9439

Fonte: Dados reprocessados da Pnad (2005), pelo autor.

4.2 Determinantes da escolha de alternativas de ocupação nos anos de 2003 e 2005

Como já demonstrado, a família, na busca da melhor estratégia de sobrevivência, escolhe de forma simultânea apenas uma das opções (são mutuamente excludentes) de ocupação de seus membros: atividades agrícolas, não-agrícolas, ser pluriativas ou estar não-ocupadas. A cada uma das alternativas/categorias se associa uma utilidade, que a família obtém da decisão tomada. A família *i* escolhe a alternativa *j* em relação à *k*, se $U_j > U_{k,}$.

O teste de Wald é importante para analisar se as variáveis contribuem na explicação das alternativas de ocupação. Os resultados do teste, calculado inicialmente considerando o ano de 2003, para cada variável, estão na Tabela 12. Para o teste F com 3 e 118 graus de liberdade, todas as variáveis são significativas a 1%, indicando que são importantes para explicar as alternativas de ocupação do modelo logit multinomial: agrícola, não-agrícola, pluriativa ou não-ocupado.

Tabela 12 – Teste de Wald para nulidade dos efeitos de cada variável explicativa sobre a dependente. Modelo logit multinomial - ano de 2003

Variáveis	Valor Calculado F (3, 118)	Prob > F
Idade PEA	42,7400	0,0000
Idade PEA quadrática	36,8700	0,0000
Anos de estudo	51,3100	0,0000
N° Compon. da Família	50,6500	0,0000
Local de moradia	16,0500	0,0000
Razão dependência	38,7100	0,0000
Renda não-trabalho	14,2300	0,0000
Conta-própria	22929,6900	0,0000
Empregados	40,7600	0,0000
Renda per capta	20,3100	0,0000
Iluminação domiciliar	4,7900	0,0035
Água de poço	4,4300	0,0055

Fonte: Dados da Pesquisa. Pnad (2003).

Os coeficientes dos determinantes da escolha entre as alternativas de ocupação, estimados pelo modelo logit multinomial[52], para o ano de 2003, estão apresentados na Tabela 13. A base de comparação são as famílias agrícolas. Esta tabela demonstra ainda os efeitos parciais[53] no ponto médio e a razão relativa de risco de cada variável.

Com relação à significância, a maior parte das variáveis é significativa a 1% de probabilidade. Em algumas situações, uma variável é significativa em uma alternativa

[52] Os testes não detectaram colinearidade entre as covariáveis. O FIV, em todas as variáveis, é menor do que 2.

[53] Calculados com o auxílio do programa *margeff* no *software* Stata. Efeitos parciais são mudanças parciais discretas nas quantidades de interesse quando a variável em estudo aumenta em uma unidade de mensuração. No caso de *dummies* e variáveis contáveis, a unidade de mensuração é 1. Normalmente os valores são muito próximos aos obtidos via mudanças calculadas via efeito marginal, mas com uma redução de custo computacional considerável.

e não é em outra. Este resultado é comum em modelos deste tipo (HOFFMANN, 2004). Para a alternativa "não-agrícola", as variáveis *razão de dependência* e *conta-própria* não são significativas a 10%. No caso da alternativa "pluriativa", *renda não-trabalho*, *iluminação domiciliar* e *água de poço* não são estatisticamente significativas a 10%. Com relação à alternativa "não-ocupado", as variáveis não significativas são *anos de estudo*, *iluminação domiciliar* e *água de poço*.

Tabela 13 – Resultados do modelo Logit Multinomial para os determinantes da escolha de alternativas de ocupação, Nordeste, 2003

Variáveis	Não-agrícola			Pluriativa			Não-ocupado		
	Coef.	Efeito Parcial	RRR	Coef.	Efeito Parcial	RRR	Coef.	Efeito Parcial	RRR
Idade PEA	0,1596***	0,0010	1,1731	0,1781***	0,0066	1,1950	-0,1541***	0,0000	0,8572
Idade PEA quadrática	-0,0022***	-	0,9978	-0,0024***	-	0,9976	0,0020***	-	1,0020
Anos de estudo	0,2954***	0,0199	1,3436	0,1598***	0,0255	1,1732	0,0668	0,0000	1,0690
N° compon. da família	-0,0629**	-0,0096	0,9391	0,1664***	0,0364	1,1811	-0,6832***	0,0000	0,5050
Local de moradia	-1,0733***	-0,1075	0,3419	-0,2336*	-0,0088	0,7917	-1,3081***	0,0000	0,2703
Razão dependência	-0,1291	0,0043	0,8789	-0,5582***	-0,1129	0,5722	0,7416***	0,0000	2,0993
Renda não-trabalho	-0,3472***	-0,0312	0,7067	0,0608	0,0222	1,0627	1,5698***	0,0000	4,8055
Conta-Própria	-0,1097	-0,0190	0,8961	0,4117***	0,0911	1,5093	-33,4339***	-0,0086	0,0000
Empregados	0,7449***	0,0426	2,1062	0,7173***	0,1313	2,0489	-5,6581***	0,0000	0,0035
Renda per capta	0,0009***	0,0000	1,0009	0,0012***	0,0002	1,0012	-0,0031***	0,0000	0,9970
Iluminação domiciliar	-0,4174***	-0,0345	0,6587	-0,0037	0,0105	0,9963	-0,1040	0,0000	0,9012
Água de poço	0,0861***	0,0068	1,0899	0,0103	-0,0001	1,0103	0,0114	0,0000	1,0115
Piauí	-0,7502	-0,0613	0,4723	0,8079**	0,2046	2,2431	-0,6423	0,0000	0,5261
Ceará	-0,7636	-0,0443	0,4660	-0,3515	-0,0547	0,7036	0,7780	0,0000	2,1771
Rio Grande do Norte	-0,2170	-0,0043	0,8049	-0,5995**	-0,1064	0,5491	1,7990**	0,0000	6,0439
Paraíba	-0,0507	-0,0015	0,9505	-0,1016	-0,0195	0,9034	0,9638	0,0000	2,6216
Pernambuco	-1,0020**	-0,0538	0,3671	-0,5831**	-0,0937	0,5582	0,3640	0,0000	1,4390
Alagoas	-0,9020*	-0,0444	0,4058	-0,8019**	-0,1290	0,4485	1,3541**	0,0000	3,8732
Sergipe	-1,1811**	-0,0588	0,3069	-0,3385	-0,0482	0,7129	-1,1356	0,0000	0,3212
Bahia	-1,1367**	-0,0610	0,3209	-0,9698***	-0,1625	0,3792	1,0905**	0,0000	2,9757
Constante	-3,2782***		-	-4,7010***		-	3,0550***		-

Fonte: Dados da Pesquisa. Pnad (2003).

Categoria Base: agrícola.

Para o cálculo dos efeitos parciais foram considerados os valores médios das variáveis. Para as variáveis *dummy*, o efeito parcial representa a passagem de probabilidade quando x=0 para x=1.

***, **, *, representam, respectivamente, significativo a 1%, 5% e 10%.

Sobre os sinais, a maior parte está de acordo com o esperado, conforme se observa na Tabela 13. No caso da alternativa "não-agrícola", pode ser observada a forma quadrática em que existe um ponto de máximo. Até este ponto, a idade aumenta o *log* das chances em favor de as famílias serem não-agrícolas, em relação às famílias agrícolas. Depois deste, o aumento em uma unidade na média da *idade da PEA* familiar reduz o *log* das chances, em relação às famílias agrícolas, tudo mais constante. O ponto de máximo é alcançado em 36,27 anos[54]. No caso de *água de poço*, o sinal do coeficiente é positivo, mas esta variável assume valor 2 se possui poço ou nascente e 4, caso contrário. Isto significa que o sinal positivo indica uma probabilidade de ser não-agrícola mais elevada para as famílias que não possuem poço ou nascente.

Sobre a alternativa "pluriativas", grande parte dos sinais também está como o esperado. O ponto de máximo na forma quadrática de *idade da PEA* é 37,10 anos. Para as famílias de não-ocupados, a idade tem um comportamento em forma de U, com valor de mínimo de 38,53 anos. Isto indica que até esta idade média limite, tem-se redução no *log* das chances em favor de as famílias serem não-ocupadas. Após este ponto, começa a crescer a probabilidade de a família ser não-ocupada, frente às famílias agrícolas, tudo mais constante.

Com relação à razão relativa de risco, o aumento em uma unidade na média de *anos de estudo* da família eleva em 34,36% a chance de a família ser não-agrícola e em 17,32% a chance de a família ser pluriativa, em relação às famílias agrícolas, tudo mais constante. O acréscimo em uma unidade do *número de componentes da família* reduz em 6,09% a chance de a família ser exclusivamente não-agrícola, eleva em 18,11% a chance de a família ser pluriativa e representa um decréscimo de 49,50% na chance de a família ser não-ocupada (dadas as suas características), frente às famílias agrícolas e tudo mais não alterado.

Para a variável *local de moradia*, no caso de a família residir no "rural agropecuário", a chance de a família ser exclusivamente não-agrícola se reduz em

[54] Segundo a fórmula $x^* = \left|\frac{\beta_{idpearer}}{2\beta_{idpearer2}}\right| = \left|\frac{0,1866}{2(-0,00512)}\right| = 36,27$, sugerida em Wooldridge (2006).

65,81% e a chance de ser pluriativa, decresce em 20,83%. A chance de a família ser não-ocupada tem decréscimo de 72,97% em relação às famílias agrícolas, *ceteris paribus*. De forma significativa, o aumento unitário na *razão de dependência*, em relação às famílias agrícolas e tudo mais constante, reduz em 42,78% a chance de ser pluriativa e aumenta em 109,93% a chance de ser não-ocupada, dadas as suas características.

A variável *renda não-trabalho* é uma *dummy* indicando se a família possui ou não renda do não-trabalho. A chance de as famílias que possuem renda do não-trabalho serem exclusivamente não-agrícolas é 29,33% menor do que a chance daquelas que não possuem, relativamente às famílias exclusivamente agrícolas. A chance dos que possuem renda do não-trabalho ser pluriativas, relativamente a serem agrícolas, é 1,06 vezes maior do que a chance daquelas que não possuem. Contudo, neste caso, a variável não é significativa a 10% de probabilidade.

A chance de as famílias que possuem renda do não-trabalho serem não-ocupadas é 380,55% maior do que a chance daquelas que não possuem, relativamente às famílias exclusivamente agrícolas. Isto certamente demonstra que as rendas de aposentadorias e/ou transferências são importantes para manutenção das famílias no rural do Nordeste, como também considera Maia Gomes (2001). O efeito da elevação em uma unidade na *renda per capita* praticamente não acresce ou reduz a probabilidade de a família ser do tipo não-agrícola, agrícola ou não-ocupada.

Com grande significância, as chances de a família *conta-própria* ser pluriativa são 50,93% maiores do que ser agrícola. Da mesma forma, bastante significativa, as chances de a família *conta-própria* ser não-ocupada são 100% menores do que ser agrícola. As chances de as famílias *empregados* serem não-agrícolas, em relação à mesma razão de chances para a categoria agrícolas, são 110,62% maiores, tudo mais constante. A chance de as famílias *empregados* serem pluriativas é 104,89% mais elevada, comparando com as agrícolas. As famílias de *empregados* possuem quase 100% a mais de chance de estarem não-ocupadas em relação às de *empregados* exclusivamente agrícolas.

Sobre as duas variáveis *proxy* escolhidas para representar os insumos fixos da propriedade, as famílias sem acesso à energia elétrica possuem 34,13% menos chances de serem não-agrícolas frente a serem exclusivamente agrícolas, do que as que possuem energia elétrica. Para as demais categorias, o efeito não é significativo. Sobre a existência de poço ou nascente de água, o efeito é estatisticamente significativo apenas para a alternativa não-agrícola. Neste caso, as chances de as famílias que não possuem, serem não-agrícolas, são 8,99% maiores do que as chances das famílias com este insumo fixo, em relação às famílias agrícolas e sob a condição *ceteris paribus*.

Com relação às variáveis para "controlar" o efeito do estado em que a família reside, visando a reduzir o problema da heterogeneidade das informações agregadas, no caso da alternativa "não-agrícola" todos os coeficientes estimados são negativos, resultando em razões relativas de risco inferiores à unidade. Isto significa que, em relação ao estado do Maranhão, residir em qualquer outro estado da região Nordeste reduz a chance em favor da família ser exclusivamente não-agrícola.

Dentre os valores significativos, as chances de ser não-agrícola comparada a ser exclusivamente agrícola reduz em 63,29% quando se passa do estado do Maranhão para o estado de Pernambuco. Neste estado, as atividades não-agrícolas estão bastante relacionadas com a agroindústria canavieira. Nos anos 90 houve grande redução na produção e na área plantada de cana-de-açúcar e a crise neste setor além de reduzir o emprego agrícola, puxa para baixo as ocupações não-agrícolas (BASALDI *et al.*, 2000). Este caso é o mesmo para o estado de Alagoas, onde a atividade agrícola predominante é a cana-de-açúcar (GOMES DA SILVA e VERAS, 2001). Residir neste estado reduz a chance de ser não-agrícola em 59,42%. Em Sergipe, a chance decresce em 69,31% e na Bahia, em 67,91%.

No caso das famílias pluriativas, residir no estado do Piauí eleva a chance de a família ser pluriativa em 124,31%. Este estado possui um setor têxtil em expansão, que tem gerado uma grande quantidade de empregos (MAIA GOMES, 2001). A maior parte destes empregos se encontra na capital Teresina (pólo de confecção de Teresina), que é a única capital dentre os estados da região que não se localiza no litoral e que

certamente atrai, pela possibilidade de encontrar ocupação, um grande número de trabalhadoras que residem no meio rural de seu entorno.

Outro fator que contribui na explicação do que ocorre no Piauí é o deslocamento da fronteira agrícola para a parte do cerrado piauiense, que passa a elevar a produção de soja, atrai grande quantidade de trabalhadores e incrementa uma demanda por uma série de ocupações não-agrícolas derivadas. Soma-se a isso, toda a cadeia produtiva relacionada à apicultura. Esta atividade vem se tornando a principal fonte de renda dos produtores que antes priorizavam cultivar feijão e milho (VILELA, 2000). O estado já é o maior produtor do Nordeste e a tendência é de crescimento, pois possui as condições de clima e vegetação que favorecem a produção, o mel produzido é orgânico, podendo ser certificado e ser exportado para os mercados americano e europeu.

Por outro lado, nos demais estados a chance é menor. Se a família mora no Rio Grande do Norte, a chance de ser pluriativa se reduz em 45,09% em relação às famílias agrícolas, tudo mais constante. Este valor é próximo ao encontrado para Pernambuco, em que a chance se reduz em 44,18%. As menores chances são para as famílias da Bahia, 62,08% a menos. Os três estados com maiores chances de as famílias serem não-ocupadas são o Rio Grande do Norte, Alagoas e a Bahia.

A Tabela 13 também reporta os efeitos parciais médios, ou seja, a variação nas probabilidades dada a mudança unitária em uma ou mais variáveis dependentes. Para *idade da PEA*, a variação em uma unidade na idade média da PEA familiar eleva a probabilidade de a família ser não-agrícola em 0,0010 e a probabilidade de ser pluriativa em 0,0066 (0,66%). Para as famílias de não-ocupados, em quase todos os casos o efeito encontrado é muito pequeno, apesar de muitos serem estatisticamente significativos. Para *anos de estudo*, tem-se que um ano a mais na média de anos de estudo familiar eleva a probabilidade de a família ser não-agrícola em 0,0199 e a probabilidade de ser pluriativa em 0,0255. Para as famílias de não-ocupados, o efeito parcial não é significativo.

Com relação à variável *número de componentes da família*, um componente a mais reduz a probabilidade de a família ser não-agrícola em 0,0096. Para as famílias

pluriativas, ocorre aumento de 0,0364. Com relação ao *local de moradia*, residir no rural agropecuário reduz a probabilidade de a família ser exclusivamente não-agrícola (-0,1075). O efeito não é significativo para as famílias pluriativas.

Para a *razão de dependência*, o efeito não é significativo na alternativa não-agrícola. O aumento na *razão de dependência* reduz a probabilidade de a família ser pluriativa em 0,1129. Já a variável *renda não-trabalho*, uma variável *dummy*, representa o efeito da família que não possuía passar a ter renda do não-trabalho. Isto significa uma redução na probabilidade de ser exclusivamente não-agrícola de 0,0312. Quanto à *renda per capita*, os valores são muito baixos em todos os casos.

Com relação às famílias *conta-própria*, a probabilidade de ser não-agrícola se reduz em 0,0190 e 0,0086, para serem não-ocupadas. Contudo, se eleva em 0,0911 para ser pluriativo. Na variável *empregados*, todas as alternativas também são significativos, elevando a probabilidade em 0,0426 no caso dos não-agrícolas e 0,1313, nos pluriativos. Para *iluminação do domicílio* e *água de poço*, os efeitos parciais são significativos apenas para as famílias não-agrícolas. As famílias sem energia elétrica possuem probabilidade 0,0345 menor do que as que possuem de serem não-agrícolas. Por outro lado, aquelas que não possuem água de poço ou nascente possuem probabilidade 0,0068 maior de ser não-agrícola do que aquelas que possuem.

Analisando o estado em que a família reside, com relação às famílias não-agrícolas, as reduções na probabilidade são maiores no Piauí (-0,0613) e Bahia (-0,0610), todos os valores negativos. Para as famílias pluriativas, no estado do Piauí é encontrado um efeito significativo de elevação na probabilidade em 0,2046. Os estados de Alagoas (-0,1290) e Bahia (-0,1625) apresentam maiores reduções na probabilidade.

A Tabela 14 mostra os resultados da estatística MEFF (*Misspecification Effect*) para o logit multinomial, considerando a Pnad de 2003. O MEFF compara a estimativa da variância do parâmetro obtida considerando o plano amostral com outra estimativa, do mesmo modelo, só que desconsiderando peso, conglomerado e estratificação. Isto é importante para avaliar o Efeito do Plano Amostral (EPA), conforme descrito no

Anexo A. Na maior parte dos casos analisados, o MEFF é superior à unidade, ou seja, a não consideração do plano amostral subestima a variância verdadeira.

Tabela 14 – Resultados do MEFF para as estimações do modelo Logit Multinomial para os determinantes da escolha de alternativas de ocupação, Nordeste, 2003

Variáveis	Não-agrícola	Pluriativa	Não-ocupado
Idade PEA	1,0785	1,0784	1,4744
Idade PEA quadrática	1,1566	1,0975	1,4249
Anos de estudo	1,7581	1,4342	1,5959
N° compon. da família	0,9249	1,2627	1,4945
Local de moradia	2,9721	1,7771	1,5519
Razão dependência	1,2434	0,8767	1,6742
Renda não-trabalho	1,3812	1,2470	1,4346
Conta-Própria	1,3367	1,6757	0,0000
Empregados	1,2043	1,4931	1,0245
Renda per capta	2,9594	3,7626	3,3263
Iluminação domiciliar	2,3980	1,7948	1,2866
Água de poço	2,1314	2,1211	1,2024
Piauí	4,2534	5,9518	1,4870
Ceará	6,0511	3,9870	1,8468
Rio Grande do Norte	10,9943	2,5758	2,6159
Paraíba	6,7848	5,2863	1,8092
Pernambuco	6,6693	4,1003	2,0744
Alagoas	5,2941	4,7297	1,9020
Sergipe	3,2292	2,0252	1,1780
Bahia	7,0402	5,2071	1,8402
Constante	1,7511	0,9989	2,1023

Fonte: Dados da Pesquisa. Pnad (2003).

A Tabela 15 trata dos valores calculados do teste de Wald, utilizado para analisar se as variáveis contribuem na explicação das alternativas de ocupação, considerando as informações do ano de 2005. Para o teste F, com 3 e 125 graus de liberdade, todas as variáveis são significativas a 1%, indicando que são importantes para explicar as alternativas de ocupação do modelo logit multinomial.

Tabela 15 – Teste de Wald para nulidade dos efeitos de cada variável explicativa sobre a dependente do modelo logit multinomial para o ano de 2005

Variáveis	Valor Calculado F (3, 125)	Prob > F
Idade PEA	16,0800	0,0000
Idade PEA quadrática	15,8200	0,0000
Anos de estudo	42,2300	0,0000
N° Compon. da Família	85,8800	0,0000
Local de moradia	16,3400	0,0000
Razão dependência	43,6600	0,0000
Renda não-trabalho	23,5600	0,0000
Conta-própria	30903,8800	0,0000
Empregados	26,5200	0,0000
Renda per capta	27,0400	0,0000
Iluminação domiciliar	5,4300	0,0015
Água de poço	6,9700	0,0002

Fonte: Dados da Pesquisa. Pnad (2003).

A análise dos valores estimados[55] para o ano de 2005, apresentados na Tabela 16, em muito se assemelha ao caso de 2003, sendo novo indicativo de que a questão climática não tem o efeito imaginado na tomada de decisão da família sobre tipo de ocupação. Disto, são ressaltadas mais as diferenças. Com relação à significância, para a alternativa "não-ocupada", a variável não significativa é *anos de estudo*. As variáveis *iluminação domiciliar* e *água de poço* apresentam efeitos significativos sobre esta alternativa de escolha. Para as *dummies* que buscam controlar o estado de residência das famílias, na alternativa "não-agrícola" apenas para Sergipe o coeficiente estimado mostrou-se significativo. Para as famílias de não-ocupados, nenhuma *dummy* por estado é significativa. Sobre os sinais das estimativas, a maior parte também está de acordo com o esperado. Para a alternativa "pluriativas", no ano de 2005, o sinal de *iluminação domiciliar* é positivo, mas não significativo estatisticamente.

Com relação à razão relativa de risco, o aumento em uma unidade na *média de anos de estudo da família* eleva em 26,60% a chance de a família ser não-agrícola e,

[55] Os testes para o ano de 2005, também não detectaram colinearidade entre as covariáveis, o FIV (Fator de Inflação da Variância) não ultrapassou 2 em nenhum caso.

em 16,24%, a chance de a família ser pluriativa, em relação às famílias exclusivamente agrícolas, *ceteris paribus*. O acréscimo em uma unidade do *número de componentes da família* reduz em 12,37% a chance de a família ser exclusivamente não-agrícola, eleva em 22,12% as chances de a família ser pluriativa e representa uma redução de 52,29% nas chances em favor da família ser não-ocupada, frente às famílias agrícolas.

Tabela 16– Resultados do modelo Logit Multinomial para os determinantes da escolha de alternativas de ocupação, Nordeste, 2005

Variáveis	Não-agrícola			Pluriativa			Não-ocupado		
	Coef.	Efeito Parcial	RRR	Coef.	Efeito Parcial	RRR	Coef.	Efeito Parcial	RRR
Idade PEA	0,0904***	0,0000	1,0946	0,1319***	0,0050	1,1410	-0,0990***	0,0000	0,9058
Idade PEA quadrática	-0,0013***	-	0,9987	-0,0018***	-	0,9982	0,0013**	-	1,0013
Anos de estudo	0,2358***	0,0159	1,2660	0,1505***	0,0255	1,1624	0,0547	0,0000	1,0562
N° compon, da família	-0,1320***	-0,0178	0,8763	0,1999***	0,0478	1,2212	-0,7401***	0,0000	0,4771
Local de moradia	-1,1781***	-0,1282	0,3079	-0,2316*	0,0019	0,7933	-1,2022***	0,0000	0,3005
Razão dependência	-0,0975	0,0089	0,9071	-0,5581***	-0,1190	0,5723	0,7951***	0,0000	2,2147
Renda não-trabalho	-0,4275***	-0,0396	0,6521	-0,0102	0,0121	0,9899	1,5240***	0,0000	4,5904
Conta-Própria	-0,1820	-0,0229	0,8336	0,2531**	0,0631	1,2880	-40,5067***	-0,0063	0,0000
Empregados	0,7235***	0,0506	2,0617	0,4809***	0,0812	1,6175	-5,8847***	0,0000	0,0028
Renda per capta	0,0011***	0,0001	1,0011	0,0013***	0,0002	1,0013	-0,0028***	0,0000	0,9972
Iluminação domiciliar	-0,4487***	-0,0418	0,6385	0,0620	0,0277	1,0639	-0,3728**	0,0000	0,6888
Água de poço	0,0982***	0,0080	1,1032	0,0176	0,0008	1,0177	0,1083***	0,0000	1,1144
Piauí	-0,5221	-0,0562	0,5933	0,7672*	0,1970	2,1537	-0,3577	0,0000	0,6993
Ceará	-0,5614	-0,0373	0,5704	-0,1996	-0,0283	0,8191	-0,5782	0,0000	0,5609
Rio Grande do Norte	0,1898	0,0413	1,2090	-0,7704**	-0,1499	0,4628	0,4114	0,0000	1,5090
Paraíba	-0,1609	-0,0013	0,8514	-0,4655	-0,0896	0,6278	0,5699	0,0000	1,7681
Pernambuco	-0,6206	-0,0317	0,5376	-0,6910**	-0,1228	0,5011	-0,2602	0,0000	0,7709
Alagoas	-0,7193	-0,0322	0,4871	-1,1057***	-0,1835	0,3310	0,1601	0,0000	1,1737
Sergipe	-1,3318*	-0,0632	0,2640	-0,9320**	-0,1504	0,3938	-0,9708	0,0000	0,3788
Bahia	-1,0225	-0,0533	0,3597	-1,0643***	-0,1888	0,3450	-0,0223	0,0000	0,9779
Constante	-2,0104**		-	-4,0283***		-	3,0278***		

Fonte: Dados da Pesquisa. Pnad (2005)

Categoria Base: agrícola.

Para o cálculo dos efeitos parciais foram considerados os valores médios das variáveis. Para as variáveis *dummy*, o efeito parcial representa a passagem de probabilidade quando x=0 para x=1.

***, **, *, representam, respectivamente, significativo a 1%, 5% e 10%.

Para "*local de moradia*", se a família residir no "rural agropecuário", a chance em favor de a família ser exclusivamente não-agrícola se reduz em 69,21%, a de ser pluriativa, diminui em 20,67% e, em favor de ser não-ocupada, tem decréscimo de 69,95%, sempre em relação às famílias agrícolas e tudo mais constante. De forma significativa, o aumento unitário na *razão de dependência* reduz em 42,77% a chance de ser pluriativa e aumenta em 121,47% a chance de ser não-ocupada.

A chance em favor de as famílias que possuem *renda do não-trabalho* serem exclusivamente não-agrícolas é 34,79% menor do que a chance daquelas que não possuem, relativamente às famílias exclusivamente agrícolas. A chance de as famílias que possuem *renda do não-trabalho* serem não-ocupadas é 4,59 vezes maior do que a chance daquelas que não possuem. Isto certamente demonstra que, em 2005, as rendas de aposentadorias e/ou transferências são ainda mais importantes para manutenção das famílias no rural do Nordeste do que em 2003. Para as variáveis relacionadas à posição na ocupação, as chances de as famílias *conta-própria* serem pluriativas são 28,80% maiores do que serem exclusivamente agrícolas. Para *empregados*, as chances de serem não-agrícolas são 106,17% maiores e de ser pluriativas são 61,75% mais elevadas, tudo mais constante.

Sobre as variáveis de controle por estado de residência, a família rural residir em Sergipe reduz as chances de ser exclusivamente não-agrícola em 73,60%. Para as famílias pluriativas, residir no Piauí eleva a chance de a família ser pluriativa em 115,37%. Contudo, se a família mora no Rio Grande do Norte, a chance de ser pluriativa se reduz em 53,72% e em Pernambuco se reduz em cerca de 50%. Os estados de Alagoas, Sergipe e Bahia possuem menores chances a favor de as famílias serem pluriativas.

Com relação aos efeitos parciais, pode ser observado pela Tabela 16 que, para a variável *idade PEA*, a variação em uma unidade na idade média da PEA familiar eleva a probabilidade de a família ser pluriativa em 0,0050. Para *anos de estudo*, um ano a mais na média de anos de estudo familiar eleva a probabilidade de a família ser não-agrícola em 0,0159 e a probabilidade de ser pluriativa em 0,0255. Para as famílias de não-ocupados, o efeito parcial não é significativo. Sobre a variável *número de*

componentes da família, um componente a mais reduz a probabilidade de a família ser não-agrícola em 0,0178. Para as famílias pluriativas, ocorre aumento de 0,0478 na probabilidade.

Para a variável *local de moradia*, residir no rural mais distante do urbano significa menor probabilidade de a família ser exclusivamente não-agrícola (-0,1282). Para a alternativa "pluriativas", o efeito não é estatisticamente significativo. O aumento na *razão de dependência* reduz a probabilidade de a família ser pluriativa em 0,1190. O efeito da família que não possuía passar a ter *renda do não-trabalho*, significa uma redução na probabilidade de ser exclusivamente não-agrícola de 0,0396.

Ainda analisando a Tabela 16, para a família de *conta-própria*, a probabilidade é 0,0229 menor de ser não-agrícola, 0,0631 maior de ser pluriativa e 0,0063 menor de ser não-ocupada. Para as *empregados*, a probabilidade é 0,0506 maior de ser não-agrícola e 0,0812 mais elevada de ser pluriativa. As famílias sem energia elétrica possuem probabilidade 0,0418 menor do que as que possuem de serem não-agrícolas e 0,0277 maior de serem pluriativas. As famílias que não possuem *água de poço* ou nascente possuem probabilidade 0,0080 maior de ser não-agrícola do que aquelas que possuem.

Se a família reside no estado do Piauí, possui probabilidade 0,0562 menor de ser não-agrícola e 0,1970 maior de ser pluriativa. No caso do Rio Grande do Norte, a probabilidade de ser pluriativa é 0,1499 menor do que ser agrícola. Nos demais estados com valores significativos, a probabilidade de ser pluriativa ou não-agrícola é sempre menor do que ser exclusivamente agrícola.

De forma semelhante ao que ocorre no ano de 2003, na maior parte dos casos o valor de MEFF é superior à unidade, ou seja, a não consideração do plano amostral subestima a variância verdadeira, conforme se observa na Tabela 17. Porém, foram também encontrados valores menores do que a unidade. Isto significa que a não consideração do plano amostral está superestimando a variância calculada para *idade PEA* e *idade PEA quadrática* na alternativa "pluriativa" e "não-ocupados", assim como para *renda do não-trabalho* e *conta-própria* na categoria "não-ocupado" e *razão dependência* em "não-agrícola".

Tabela 17 – Resultados do MEFF para as estimações do modelo Logit Multinomial para os determinantes da escolha de alternativas de ocupação, Nordeste, 2005

Variáveis	Não-agrícola	Pluriativa	Não-ocupado
Idade PEA	1,0609	0,8524	0,9082
Idade PEA quadrática	1,0211	0,8357	0,9158
Anos de estudo	2,0697	1,2905	1,0995
N° compon. da família	1,0443	1,7309	1,2288
Local de moradia	4,0030	2,1860	1,2047
Razão dependência	0,9547	1,1642	1,1346
Renda não-trabalho	1,7944	2,3000	0,9120
Conta-Própria	1,7805	2,3065	0,0000
Empregados	1,2355	2,1951	1,0257
Renda per capta	3,9279	4,8600	2,4344
Iluminação domiciliar	2,2927	1,7635	1,4594
Água de poço	2,7414	1,4353	1,3718
Piauí	7,0119	9,1155	1,9846
Ceará	11,3445	6,2213	1,8974
Rio Grande do Norte	15,3818	3,3935	2,1336
Paraíba	10,3958	4,3424	1,4428
Pernambuco	12,0234	5,4921	1,7372
Alagoas	9,2271	4,3019	1,7347
Sergipe	7,9816	5,2463	1,1760
Bahia	13,5271	7,3521	1,7110
Constante	2,3611	1,3934	1,0854

Fonte: Dados da Pesquisa. Pnad (2005).

Em resumo, este tópico tem a finalidade de ajudar a compreender melhor os determinantes das escolhas de alternativas de ocupação das famílias rurais do Nordeste. Considera-se estar esclarecida, tanto no ano de 2003 quanto no ano de 2005, a existência de um grande número de famílias com pelo menos um membro ocupado em Orna (Ocupação Rural Não-Agrícola). Nos dois anos supracitados, a maior parte das famílias é do tipo conta-própria ou empregada, residindo no rural agropecuário e recebendo algum tipo de renda do não-trabalho.

Chama a atenção também que cerca de 10% das famílias estão não-ocupadas, residindo no meio rural e sobrevivendo da renda do não-trabalho. As famílias não-agrícolas e pluriativas são menos dependentes deste tipo de renda, comparado com as de não-ocupados, como demonstra a variação na probabilidade na presença da *renda do não trabalho*. Apresentam ainda impacto importante na probabilidade de a família

possuir pelo menos um membro ocupado em Orna, a idade média da família, o nível educacional, o número de componentes da família, a proximidade do centro urbano/mercado de trabalho mais desenvolvido e, principalmente, residir no estado do Piauí.

4.3 Efeitos da renda não-agrícola sobre a pobreza e a concentração

Com o modelo de seleção amostral via logit multinomial, após analisar os determinantes das escolhas de ocupação, podem ser estimadas as rendas médias para cada tipo de família, corrigindo o viés de seleção. O que se pretende é simular, por exemplo, quais as rendas das famílias agrícolas se elas fossem pluriativas, ou ainda, quais as rendas das famílias pluriativas se elas fossem não-agrícolas.

Não foram feitas simulações com as rendas das famílias de não-ocupados, conforme já discutido na parte metodológica. Então, tenta-se esclarecer se o não-agrícola contribui de alguma forma para redução da pobreza no rural do Nordeste, comparando as rendas *per capita* com e sem a presença de atividades não-agrícolas, via índices de Foster-Greer-Thorbecke (FGT): Proporção de Pobres (P_0), Hiato da Pobreza (P_1) e Severidade da Pobreza (P_2).

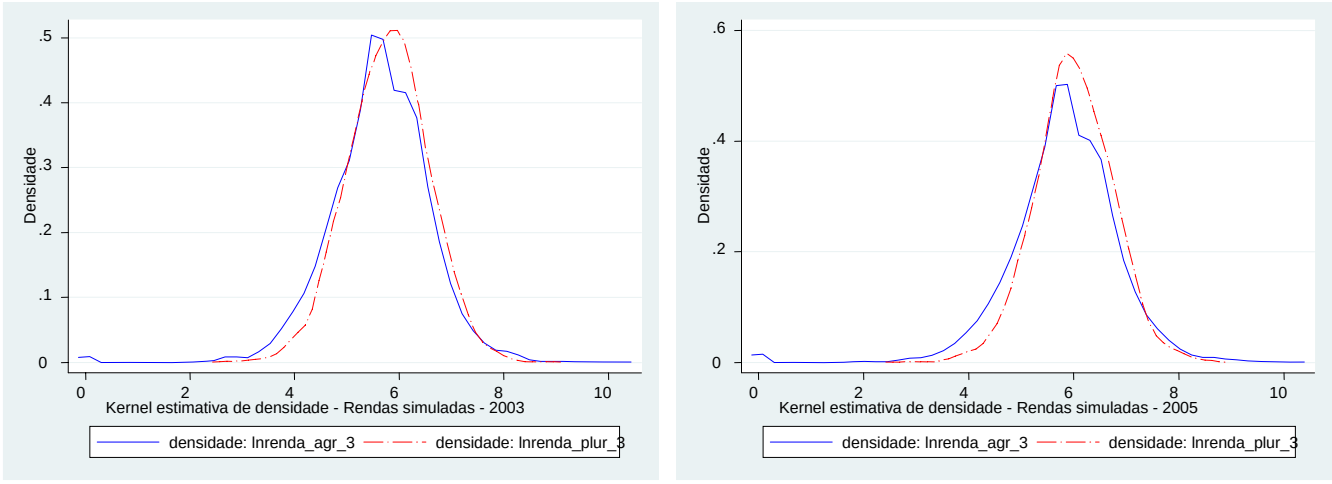
4.3.1 Simulação considerando todas as famílias agrícolas ou pluriativas

A primeira situação analisada considera as rendas observadas das famílias agrícolas e as simuladas, tanto das famílias pluriativas quanto das famílias não-agrícolas, se elas fossem agrícolas. Este é o caso em que todas as famílias, exceto as de não-ocupados, são exclusivamente agrícolas. A comparação é feita com todas as famílias sendo pluriativas, ou seja, considerando as rendas observadas das famílias pluriativas e as simuladas das famílias agrícolas e não-agrícolas, se fossem pluriativas.

A Figura 4 demonstra as funções Kernel para os logaritmos das rendas[56] dos dois tipos de famílias, nos anos de 2003 e 2005. A visualização das distribuições é

[56] Considerando a renda log-normal, a visualização fica mais clara do comportamento da distribuição. Se não considerar log-normal, a forma da distribuição geralmente é bastante assimétrica positiva.

interessante como abordagem inicial da compreensão da pobreza e da concentração. No caso simulado de todas as famílias serem agrícolas, observam-se, tanto no ano seco quanto no ano chuvoso, "caudas" mais pesadas na parte inferior ao centro da distribuição da renda agrícola, principalmente com concentração próximo do valor zero. Isto significa que o número de famílias com renda inferior à renda média é maior do que o número de famílias com renda superior à média, indicando o grau pobreza caso todos fossem agrícolas. Indica também que, possivelmente, a intensidade da pobreza é superior nas famílias agrícolas, comparado com as pluriativas.



Fonte: Dados da Pesquisa.

Figura 4 – Funções Kernel para os logaritmos das rendas simuladas com todas as famílias agrícolas e todas pluriativas, Nordeste, anos de 2003 e 2005.

Com relação à forma da distribuição, nos dois tipos de famílias a curva é leptocúrtica, ou seja, existe uma maior concentração de famílias com renda em torno da média. No ano de 2003, o centro da Kernel para as famílias agrícolas está mais à esquerda, em comparação com a distribuição para as famílias pluriativas. Isto indica que a renda média para as pluriativas é mais elevada. Nos dois anos, o "pico" da Kernel é maior para as famílias pluriativas, indicando maior concentração de famílias com renda próxima do valor médio. Com base na função Kernel, não se pode definir que a pluriatividade reduza a concentração. Porém, pode-se afirmar que devido às caudas mais pesadas na parte inferior da distribuição e à semelhança no formato

leptocúrtico, é possível que exista uma maior desigualdade no caso de todas as famílias serem do tipo agrícola.

A Tabela 18 trata dos índices de pobreza (FGT) para as rendas simuladas, com todas as famílias sendo exclusivamente agrícolas ou pluriativas. O índice P_0 mede a proporção de famílias que não possuem renda *per capita* superior à linha de pobreza (R$ 120,00 e R$ 150,00, para 2003 e 2005, respectivamente). Dado que este indicador possui limitações, são calculados o P_1 e o P_2, dando mais pesos aos mais pobres entre os pobres. Para o ano de 2003, os valores calculados dos FGT a partir da simulação com todas as famílias agrícolas correspondem a uma proporção de pobres (P_0) em torno de 16,28%, ao hiato da pobreza (P_1) em 6,22% e à severidade da pobreza (P_2) em, aproximadamente, 3,36%.

Tabela 18 - Índices de pobreza (FGT) para as rendas simuladas com todas as famílias agrícolas e todas pluriativas, Nordeste, anos de 2003 e 2005

	Estimativa FGT	Erro Padrão	Intervalo de Confiança (95%)		Valor-P do teste de igualdade entre os índices
Ano de 2003					
Renda Simulada todas Agrícolas					
P_0	0,1628	0,0077	0,1477	0,1780	-
P_1	0,0622	0,0040	0,0542	0,0702	-
P_2	0,0336	0,0028	0,0281	0,0391	-
Renda Simulada todas Pluriativas					
P_0	0,1030	0,0059	0,0914	0,1146	0,0000
P_1	0,0298	0,0023	0,0253	0,0342	0,0000
P_2	0,0127	0,0012	0,0103	0,0152	0,0000
Ano de 2005					
Renda Simulada todas Agrícolas					
P_0	0,1683	0,0086	0,1512	0,1853	-
P_1	0,0691	0,0043	0,0605	0,0776	-
P_2	0,0398	0,0030	0,0339	0,0457	-
Renda Simulada todas Pluriativas					
P_0	0,0848	0,0048	0,0752	0,0944	0,0000
P_1	0,0246	0,0015	0,0217	0,0276	0,0000
P_2	0,0107	0,0008	0,0091	0,0122	0,0000

Fonte: Dados da Pesquisa.

P_0, P_1 e P_2 indicam, respectivamente, a proporção de pobres, o hiato da pobreza e a severidade da pobreza.

Pela observação do valor de probabilidade do teste de igualdade entre os índices, podem ser considerados estatisticamente diferentes dos relativos às famílias pluriativas, que são 10,30%, 2,98% e 1,27%, respectivamente. Pode-se considerar a importância da pluriatividade, pois os valores simulados são menores, com redução de cerca de 50% no hiato da pobreza e de dois terços na severidade da pobreza.

No ano de 2005, as simulações indicam que se a totalidade das famílias fossem agrícolas, todos os indicadores de pobreza seriam mais elevados do que em 2003 e, no caso de serem pluriativas, seriam ainda menores. Os valores para o caso das agrícolas são 16,83%, 6,91% e 3,98%. Para as pluriativas, os índices diminuem para 8,48%, 2,46% e 1,07%, respectivamente, para P_0, P_1 e P_2, sendo todos estatisticamente diferentes.

A Tabela 19 mostra os valores do índice de Gini para as rendas simuladas das famílias, se fossem agrícolas ou pluriativas, nos anos de 2003 e 2005. Os valores de erro padrão são calculados via *bootstrap* e linearização de Taylor. Em todos os casos, os valores calculados por linearização de Taylor são ligeiramente inferiores. Para o ano de 2003, o valor para as famílias agrícolas é cerca de 0,4826 e, para as famílias pluriativas de, aproximadamente, 0,4155. A diferença entre os valores é bem mais elevada no ano de 2005, de 0,5009 e 0,4016, respectivamente, para agrícolas e pluriativas.

Tabela 19 - Índice de concentração de Gini para as rendas simuladas com todas as famílias agrícolas e todas pluriativas. Erros padrão obtidos por *bootstrap (b)* com 200 replicações e linearização de Taylor (*lt*), Nordeste, anos de 2003 e 2005

Índice	Estimativa FGT	Viés	Erro Padrão	Intervalo de Confiança com correção de viés (95%)	
Ano de 2003					
Renda Simulada todas Agrícolas					
Gini (*b*)	0,4826	0,0031	0,0146	0,4543	0,5076
Gini (*lt*)	0,4851	-	0,0119	0,4617	0,5084
Renda Simulada todas Pluriativas					
Gini *(b)*	0,4155	0,0010	0,0057	0,4036	0,4257
Gini *(lt)*	0,4185	-	0,0044	0,4098	0,4271
Ano de 2005					
Renda Simulada todas Agrícolas					
Gini *(b)*	0,5009	0,0045	0,0112	0,4809	0,5182

Gini *(lt)*	0,5032	-	0,0109	0,4817	0,5246
Renda Simulada todas Pluriativas					
Gini *(b)*	0,4016	0,0002	0,0060	0,3899	0,4120
Gini *(lt)*	0,4004		0,0045	0,3917	0,4091

Fonte: Dados da Pesquisa.

Com relação à concentração, nos dois anos o valor do índice de Gini calculado é mais elevado para as famílias agrícolas, ou seja, neste nível de agregação a pluriatividade contribui para redução da desigualdade no meio rural. O valor menor para o índice de concentração, se todos fossem agrícolas, no ano de 2003, possivelmente possui relação com a questão climática. A renda obtida da atividade agrícola no "ano seco" é muito baixa, independentemente do tamanho da propriedade, do tipo de produção ou da cultura produzida. No ano de seca, sofrem redução de produção mesmo os perímetros irrigados, devido aos baixos volumes das barragens/reservatórios de água que passa a ser direcionada para consumo humano e animal. Então, para este tipo de simulação, a renda agrícola se "nivela por baixo", levando ao menor valor para o índice de Gini.

4.3.2 Análise das simulações por tipos de famílias

Os resultados do tópico anterior são importantes para ter uma idéia geral do problema em questão. Contudo, a restrição de todas as famílias serem de um tipo ou de outro pode não ser útil para perceber se as características das famílias ou do contexto em que elas estão inseridas contribuem para alterar os resultados dos índices de pobreza e concentração. Por este motivo, neste item são feitas novas análises, baseadas nas diversas combinações possíveis: família agrícola com renda observada e simulada, caso pluriativa ou não-agrícola; família pluriativa com renda observada e simulada, caso seja exclusivamente agrícola ou não-agrícola; e, por último, o caso da família não-agrícola com renda observada e simulada, caso agrícola ou pluriativa.

Para a obtenção das rendas estimadas para cada tipo de família, considerando a possibilidade de viés de seleção, é utilizado o modelo de seleção amostral com logit multinomial. Utilizando os parâmetros estimados e os resíduos, como demonstrado na

metodologia, é possível encontrar a renda observada das famílias e as rendas simuladas. A Tabela 20 apresenta os valores de coeficientes estimados para a renda das famílias agrícolas nos anos de 2003 e 2005. Os valores do erro padrão foram obtidos pelo procedimento de reamostragem *bootstrap* com 1000 replicações. Nos dois anos praticamente todas as variáveis foram significativas em nível de 5% de probabilidade, com os sinais iguais aos esperados. O valor significativo de *lambda* indica que é correto considerar o viés de seleção neste modelo.

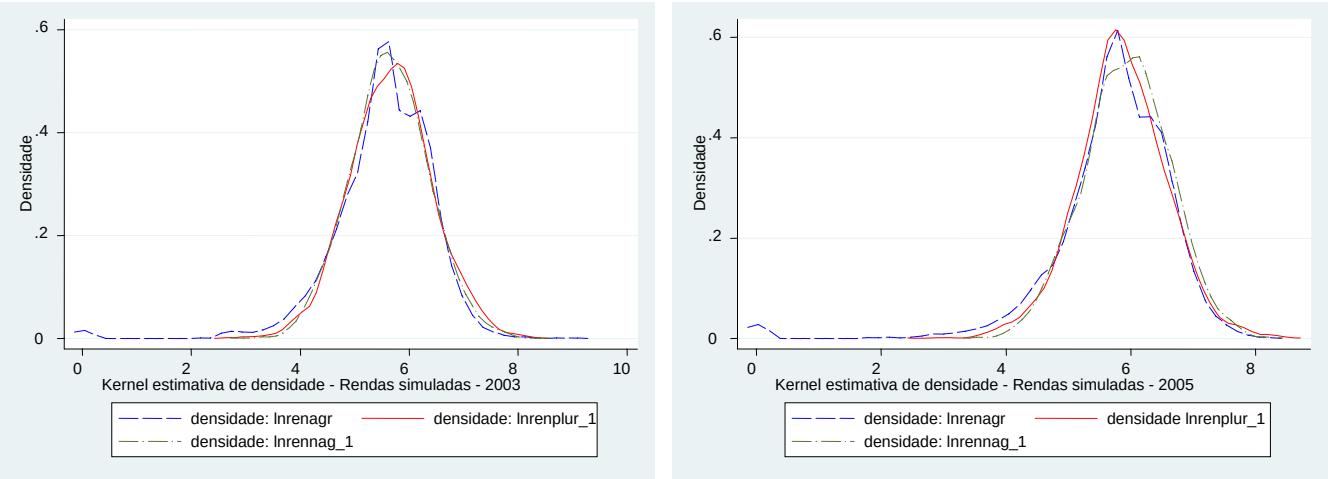
Tabela 20 – Coeficientes estimados para a equação do *log* da renda média das famílias agrícolas com correção de seleção, baseado no logit multinomial, Nordeste, anos de 2003 e 2005

	Ano de 2003			Ano de 2005		
Variáveis	Coefic.	Erro Padrão*	Valor-P	Coef.	Erro Padrão*	Valor-P
Idade da PEA	0,0048	0,0025	0,0530	0,0057	0,0031	0,0690
Anos de Estudo	-0,0714	0,0168	0,0000	-0,1035	0,0188	0,0000
N° Compon. da Família	0,0372	0,0113	0,0010	0,0440	0,0142	0,0020
Local de Moradia	0,3549	0,0798	0,0000	0,4969	0,0994	0,0000
Razao Dependência	0,0882	0,0335	0,0080	0,2167	0,0402	0,0000
Piauí	-0,2262	0,1179	0,0550	-0,5073	0,1378	0,0000
Ceará	0,2786	0,1012	0,0060	0,0908	0,1115	0,4150
Rio Grande do Norte	0,1251	0,1169	0,2840	0,1926	0,1227	0,1170
Paraíba	0,1482	0,1111	0,1820	0,3623	0,1197	0,0020
Pernambuco	0,3557	0,1004	0,0000	0,3256	0,1105	0,0030
Alagoas	0,4748	0,1073	0,0000	0,2971	0,1284	0,0210
Sergipe	0,4534	0,1294	0,0000	0,8053	0,1375	0,0000
Bahia	0,6912	0,0966	0,0000	0,7060	0,1146	0,0000
Lambda	1,3451	0,1761	0,0000	1,6847	0,2182	0,0000
Constante	3,8011	0,2055	0,0000	3,5952	0,2586	0,0000

Fonte: Dados da Pesquisa.
* Erro-padrão obtido por *bootstrap* com 1000 replicações

A Figura 5 mostra as funções Kernel para o *log* da renda das famílias agrícolas, observada e simulada, caso fossem pluriativas ou exclusivamente não-agrícolas. Nos dois anos, verifica-se a existência de caudas mais longas e observações concentradas próximas ao valor de zero, para o *log* das rendas agrícolas observadas. As funções Kernel para as simulações do caso da pluriatividade e do não-agrícola possuem a parte

inferior da distribuição mais distante do valor zero. Isto sugere a possibilidade de as atividades e rendas não-agrícola, isoladas ou combinadas com as agrícolas, contribuirem para a redução da pobreza rural no Nordeste. Em 2003 e 2005, as distribuições apresentam formato leptocúrtico semelhantes, com "picos" ligeiramente mais elevados para o *log* da renda agrícola. Não é possível observar, então, diferenças de assimetria entre as funções Kernel.



Fonte: Dados da Pesquisa

Figura 5 – Funções Kernel para os logaritmos das rendas observadas e simuladas das famílias agrícolas, caso fossem pluriativas ou não-agrícolas, Nordeste, anos de 2003 e 2005.

A Tabela 21 mostra os valores calculados para os índices FGT e MEFF (*Misspecification Effect*), considerando as rendas das famílias agrícolas observadas e simuladas, caso fossem pluriativas ou não-agrícolas, tanto no ano de 2003 quanto 2005. Como se pode observar para 2003, com relação à renda observada, a proporção de pobres está em torno de 15,53%, o hiato da pobreza em 6,40% e a severidade da pobreza em 3,74%.

Após a simulação, ou seja, considerando a renda *per capita* das famílias agrícolas, caso fossem pluriativas, o valor de P_0 se reduz para 12,26%, estatisticamente diferente ao nível de 1% de probabilidade. O valor de P_1 se reduz para 3,47% e o valor de P_2 para 1,47%. Em ambos os casos, os valores são estatisticamente diferentes. De forma semelhante, as rendas simuladas das famílias agrícolas, caso fossem

exclusivamente não-agrícolas, indicam que os três índices decrescem para 11,96%, 3,21% e 1,25%, sendo estatisticamente diferentes das rendas observadas.

Continuando a análise da Tabela 21, no ano de 2005, para a renda observada das famílias agrícolas, os valores de P_0, P_1 e P_2 são, respectivamente, 14,62%, 6,77% e 4,29%. Após as simulações, no caso se fossem pluriativas, o impacto é significativo em todos os índices de pobreza, resultando em menores valores, 10,45%, 3,13% e 1,39%, para P_0, P_1 e P_2. Fica claro, também, que ser não-agrícola contribui para redução dos três índices de pobreza. Considerando os percentuais na simulação, o valor em P_1 passa a ser 9,79%, P_2 se reduz para 2,48% e P_3 decresce para 0,92%, praticamente um quinto do índice da severidade da pobreza (P_2) obtido com a renda agrícola observada.

Tabela 21 - Índices de pobreza (FGT) para as rendas das famílias agrícolas: observadas e simuladas, caso fossem pluriativas e não-agrícolas e MEFF, Nordeste, anos de 2003 e 2005

	Estimativa FGT	Erro Padrão	Intervalo de Confiança (95%)		Valor-P do teste de igualdade entre os índices	MEFF
Ano de 2003						
Renda Observada						
P_0	0,1553	0,0125	0,1306	0,1800	-	4,1789
P_1	0,0640	0,0066	0,0510	0,0771	-	4,7325
P_2	0,0374	0,0046	0,0282	0,0466	-	4,2069
Renda Simulada se Pluriativas						
P_0	0,1226	0,0064	0,1099	0,1353	0,0030	1,3449
P_1	0,0347	0,0024	0,0299	0,0395	0,0000	1,5160
P_2	0,0147	0,0013	0,0121	0,0174	0,0000	1,4351
Renda Simulada se Não-agrícolas						
P_0	0,1196	0,0063	0,1071	0,1321	0,0040	1,2897
P_1	0,0321	0,0020	0,0281	0,0361	0,0000	1,1943
P_2	0,0125	0,0009	0,0106	0,0143	0,0000	1,0497
Ano de 2005						
Renda Observada						
P_0	0,1462	0,0140	0,1184	0,1739	-	5,5114
P_1	0,0677	0,0072	0,0534	0,0820	-	4,9126
P_2	0,0429	0,0050	0,0329	0,0528	-	3,8078
Renda Simulada se Pluriativas						
P_0	0,1045	0,0062	0,0922	0,1167	0,0020	1,3939
P_1	0,0313	0,0020	0,0273	0,0354	0,0000	1,0859

P_2	0,0139	0,0011	0,0116	0,0161	0,0000	1,0175
Renda Simulada se Não-agrícolas						
P_0	0,0979	0,0051	0,0878	0,1081	0,0020	0,9983
P_1	0,0248	0,0017	0,0213	0,0282	0,0000	1,1602
P_2	0,0092	0,0009	0,0075	0,0109	0,0000	1,2637

Fonte: Dados da Pesquisa.

P_0, P_1 e P_2 indicam, respectivamente, a proporção de pobres, o hiato da pobreza e a severidade da pobreza.

Dado que o ano de 2003 não é considerado de chuvas regulares para toda a região, isto reduz a renda de todas as famílias, resultando em índices de pobreza mais elevados na comparação com 2005. Contudo, fica demonstrado que a pluriatividade isolada reduz a pobreza rural nos dois anos e que esta diminuição poderia ser ainda maior se as famílias se tornassem exclusivamente não-agrícolas.

Deste fato, os valores aqui encontrados não corroboram os resultados de Nascimento (2005), de que a pluriatividade não está fazendo com que as famílias passem para a parte de cima da linha da pobreza. As simulações nas rendas das famílias agrícolas demonstram que as atividades e rendas não-agrícolas são importantes para redução não só da proporção de pobres, objeto de estudo de Nascimento (2005), mas também do hiato e da severidade da pobreza.

Sobre o MEFF, em grande parte dos casos a não consideração do desenho amostral subestima a variância verdadeira, pois a maior parte dos valores para as rendas nos dois anos possuem MEFF > 1, principalmente no caso das rendas observadas em que os valores são os mais elevados.

Continuando a análise por tipo de família, a Tabela 22 mostra os valores dos coeficientes estimados para as rendas das famílias pluriativas nos anos de 2003 e 2005, considerando a possibilidade de viés de seleção. No ano de 2003, as variáveis *idade da PEA* e *anos de estudo* não são estatisticamente significativas a 10%. Em 2005, apenas *anos de estudo* e a *dummy* relativa a residir no estado do *Ceará* não são significativas ao nível de 10%. Nos dois anos, os valores significativos de *lambda* indicam a necessidade de considerar o viés de seleção nestes modelos.

Tabela 22 – Coeficientes estimados para a equação do *log* da renda média das famílias pluriativas com correção de seleção, baseado no logit multinomial, Nordeste, anos de 2003 e 2005

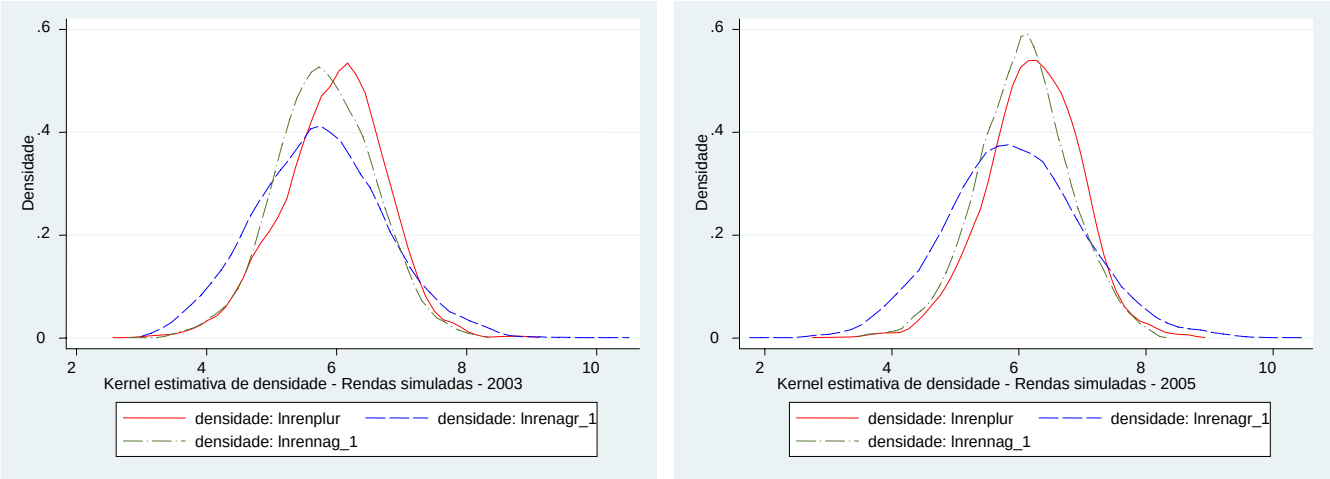
	Ano de 2003			Ano de 2005		
Variáveis	Coefic.	Erro Padrão*	Valor-P	Coef.	Erro Padrão*	Valor-P
Idade da PEA	0,0041	0,0036	0,2480	0,0069	0,0040	0,0820
Anos de Estudo	0,0211	0,0150	0,1600	-0,0141	0,0142	0,3180
N° Compon. da Família	-0,0398	0,0193	0,0400	-0,1240	0,0225	0,0000
Local de Moradia	-0,1440	0,0814	0,0770	-0,2969	0,0927	0,0010
Razao Dependência	0,1935	0,0555	0,0000	0,3939	0,0635	0,0000
Piauí	-0,6236	0,1255	0,0000	-0,7649	0,1137	0,0000
Ceará	0,1915	0,1014	0,0590	0,1350	0,1046	0,1970
Rio Grande do Norte	0,6893	0,1486	0,0000	0,9789	0,1646	0,0000
Paraíba	0,2463	0,1119	0,0280	0,6769	0,1327	0,0000
Pernambuco	0,3017	0,1077	0,0050	0,5052	0,1260	0,0000
Alagoas	0,5177	0,1683	0,0020	0,8686	0,1797	0,0000
Sergipe	0,2965	0,1269	0,0190	0,4985	0,1773	0,0050
Bahia	0,6623	0,1073	0,0000	0,8888	0,1131	0,0000
Lambda	-1,5528	0,1780	0,0000	-2,1348	0,1635	0,0000
Constante	7,3738	0,3723	0,0000	8,5110	0,3447	0,0000

Fonte: Dados da Pesquisa.
* Erro padrão obtido por *bootstrap* com 1000 replicações

As funções Kernel para o *log* das rendas observadas e simuladas das famílias pluriativas podem ser analisadas na Figura 6. Nos dois anos, considerando a renda observada das famílias pluriativas, não se observam caudas longas na parte inferior da distribuição, indicando que a pobreza é menor para este tipo de família. No ano de 2003, a Kernel para a simulação do *log* da renda das famílias pluriativas, se fossem agrícolas, apresenta valores concentrados na parte superior da cauda. Possivelmente isto vai estar relacionado a uma maior concentração de renda neste caso.

Nos dois anos, é observável que existem diferenças na curtose entre a renda observada das pluriativas e a simulada, caso fossem agrícolas. A distribuição do *log* da renda agrícola é mais platicúrtica, ou seja, com valores mais "espalhados" da média e "pico" mais baixo, consequentemente. Isto pode ser um sinal de maior desigualdade,

comparado com os outros tipos de renda. Com relação à assimetria, nos dois anos pode-se observar que as médias das três distribuições são diferentes, sendo maior para a renda advinda da pluriatividade e menor no caso exclusivamente agrícola.



Fonte: Dados da Pesquisa.

Figura 6 – Funções Kernel para os logaritmos das rendas observadas e simuladas das famílias pluriativas, caso fossem agrícolas ou exclusivamente não-agrícolas, Nordeste, anos de 2003 e 2005.

A Tabela 23 apresenta os valores calculados para os índices FGT e MEFF, considerando as rendas das famílias pluriativas observadas e simuladas, caso fossem agrícolas ou não-agrícolas, para os anos de 2003 e 2005. Os valores estimados dos índices FGT para as rendas observadas correspondem, respectivamente, a 7,87%, 2,48% e 1,11%, em 2003. Os três índices obtidos das rendas simuladas das famílias pluriativas, se fossem agrícolas, são estatisticamente diferentes daqueles com rendas observadas. Os valores se elevam bastante, para 17,34%, 6,03% e 2,90%, ou seja, praticamente mais do que dobram a proporção de pobres, o hiato da pobreza e a severidade da pobreza. Considerando as rendas simuladas, caso as famílias se tornassem exclusivamente não-agrícolas, pode-se afirmar que estatisticamente não existem diferenças entre os índices de pobreza comparativamente à renda observada das famílias pluriativas.

Para o ano de 2005, a combinação de elevação da renda agrícola, devido ao não problema climático, somado com a atividade não-agrícola, faz elevar as rendas observadas das famílias pluriativas de forma a reduzir a proporção de pobres, o hiato e

a severidade da pobreza, na comparação com 2003. Os valores calculados com a renda observada representam 5,60%, 1,69% e 0,75%, para P_0, P_1 e P_2, respectivamente. Estes valores são os mais baixos dentre as famílias, no ano com chuvas regulares. As simulações, de forma semelhante ao encontrado em 2003, indicam que se estas famílias fossem agrícolas, ocorreriam grandes elevações em todos os índices de pobreza.

Tabela 23 - Índices de pobreza (FGT) e MEFF para as rendas das famílias pluriativas: observada e simulada, caso fossem agrícolas e não-agrícolas, Nordeste, anos de 2003 e 2005

	Estimativa FGT	Erro Padrão	Intervalo de Confiança (95%)		Valor-P do teste de igualdade entre os índices	MEFF
Ano de 2003						
Renda Observada						
P_0	0,0787	0,0115	0,0559	0,1015	-	3,3863
P_1	0,0248	0,0044	0,0161	0,0336	-	3,5326
P_2	0,0111	0,0024	0,0063	0,0159	-	3,1490
Renda Simulada se Agrícolas						
P_0	0,1734	0,0100	0,1535	0,1932	0,0000	1,2342
P_1	0,0603	0,0045	0,0515	0,0691	0,0000	1,3879
P_2	0,0290	0,0024	0,0242	0,0338	0,0000	1,1984
Renda Simulada se Não-Agrícolas						
P_0	0,0839	0,0069	0,0703	0,0976	0,6790	1,1027
P_1	0,0224	0,0023	0,0178	0,0269	0,5810	1,0020
P_2	0,0095	0,0013	0,0070	0,0120	0,5200	1,0353
Ano de 2005						
Renda Observada						
P_0	0,0560	0,0072	0,0416	0,0703	-	1,9653
P_1	0,0169	0,0026	0,0117	0,0220	-	1,8518
P_2	0,0075	0,0014	0,0047	0,0104	-	1,6246
Renda Simulada se Agrícolas						
P_0	0,1957	0,0105	0,1750	0,2164	0,0000	1,3679
P_1	0,0723	0,0042	0,0639	0,0806	0,0000	1,0667
P_2	0,0376	0,0027	0,0323	0,0428	0,0000	1,0220
Renda Simulada se Não-Agrícolas						
P_0	0,0734	0,0065	0,0606	0,0862	0,0480	1,1571
P_1	0,0192	0,0021	0,0150	0,0235	0,4150	1,1363
P_2	0,0080	0,0011	0,0058	0,0101	0,7970	1,0621

Fonte: Dados da Pesquisa.

P_0, P_1 e P_2 indicam, respectivamente, a proporção de pobres, o hiato da pobreza e a severidade da pobreza.

Isto é importante para desconsiderar a pluriatividade como algo transitório, relacionado apenas com a crise da atividade agrícola. Dadas as características da atividade agropecuária no Nordeste, é muito improvável que uma família pluriativa passe a ser exclusivamente agrícola. Para as famílias pluriativas, mesmo no ano chuvoso, a diversificação das atividades é a melhor estratégia de sobrevivência. Além disso, independentemente do ano ser ou não com chuvas regulares, a pluriatividade é importante para reduzir a pobreza.

Para confirmar esta idéia, a Tabela 24 reporta as rendas observadas e simuladas das famílias agrícolas e pluriativas, para os anos de 2003 e 2005. Seguindo o raciocínio desenvolvido por De Janvry *et al.* (2005), as comparações dos resultados da Tabela 24 permitem concluir que se as famílias pluriativas fossem exclusivamente agrícolas, teriam rendas médias (R$ 531,59 e R$ 681,75 para 2003 e 2005, respectivamente) superiores às rendas das famílias exclusivamente agrícolas (R$ 348,76 e R$ 423,13). Isto significa que as atuais famílias exclusivamente agrícolas do rural nos estados do Nordeste não são as mais eficientes economicamente.

Tabela 24 - Renda média observada e simulada das famílias agrícolas e pluriativas, Nordeste, anos de 2003 e 2005

Tipos de Famílias	Fonte 0 – Agrícolas (W_0)	Fonte 1 – Pluriatividade (W_1)	Valor-P da diferença de médias
Ano de 2003			
Famílias Agrícolas (P=0)	$W_0 \mid_{P=0}$ = R$ 348,76	$E\hat{W}_1 \mid_{P=0}$ = R$ 380,71	0,0000
Famílias Pluriativas (P=1)	$E\hat{W}_0 \mid_{P=1}$ = R$ 531,59	$W_1 \mid_{P=1}$ = R$ 519,17	0,5990
Ano de 2005			
Famílias Agrícolas (P=0)	$W_0 \mid_{P=0}$ = R$ 423,13	$E\hat{W}_1 \mid_{P=0}$ = R$ 453,12	0,0000
Famílias Pluriativas (P=1)	$E\hat{W}_0 \mid_{P=1}$ =R$ 681,75	$W_1 \mid_{P=1}$ = R$ 676,20	0,1340

Fonte: Dados da Pesquisa.

Consequentemente, pode-se ainda esperar uma redução do número de famílias exclusivamente agrícolas e uma elevação do número de pluriativas ou até mesmo de exclusivamente não-agrícolas. Com relação às famílias pluriativas, pode-se considerar que são os mais eficientes, pois suas rendas médias observadas (R$ 519,17 e R$ 676,20, para 2003 e 2005, respectivamente) são superiores às rendas das famílias agrícolas, se fossem pluriativas (R$ 380,71 e R$ 453,12). Em regiões mais desenvolvidas, como a Sul do Brasil, por exemplo, o número de famílias agrícolas também decresce. Contudo, elas passam a ser exclusivamente não-agrícolas (NASCIMENTO, 2005). Dada a característica do meio rural do Nordeste, principalmente com relação ao mercado de trabalho não-agrícola, que não tem condições de absorver toda a mão de obra agrícola, a pluriatividade deve ser a estratégia das famílias com melhores resultados.

Além disso, os valores da renda média no caso das famílias agrícolas, observados e simulados, são maiores e estatisticamente diferentes nos dois anos. Por outro lado, os valores para as rendas das famílias pluriativas são iguais estatisticamente, como demonstrado pelo valor de probabilidade do teste de diferença entre as médias. Então, pode-se concluir também que as rendas não-agrícolas criam efeitos positivos[57] para a atividade agropecuária, podendo ajudar as famílias a se manter no estabelecimento e até mesmo na aquisição do necessário para a preparação do solo, plantio ou colheita agrícola, por exemplo.

Finalizando a análise da obtenção das rendas estimadas, a Tabela 25 mostra os valores estimados dos coeficientes para as rendas das famílias não-agrícolas em 2003 e 2005, considerando o viés de seleção. No ano de 2003, só foi possível obter valores de erro padrão via *bootstrap* com no máximo 70 replicações. Em todas as tentativas com número maior de replicações, o algoritmo apresenta erro. Mesmo assim, a análise foi validada, pois as estimativas pontuais estão calculadas corretamente e as variáveis significativas para 2003 são praticamente as mesmas de 2005. As *dummies* para todos

[57] Delgado e Cardoso (2000) defendem a tese de que, em muitos casos, uma parte da renda das aposentadorias percebidas pelos mais velhos é utilizada como recurso para investimento produtivo no estabelecimento rural. A dificuldade do acesso ao crédito, o risco da sua tomada e/ou taxas de juros incompatíveis com a capacidade de pagamento dos produtores podem dificultar a atividade agropecuária. Por outro lado, a renda não-agrícola pode relaxar esta restrição.

os estados não se apresentaram significativas nos dois anos, assim como o *local de moradia*. O valor de *lambda* também não é estatisticamente significativo, indicando não existir neste caso viés de seleção.

Tabela 25 – Coeficientes estimados para a equação da renda média das famílias não-agrícolas com correção de seleção, baseado no logit multinomial, Nordeste, anos 2003 e 2005

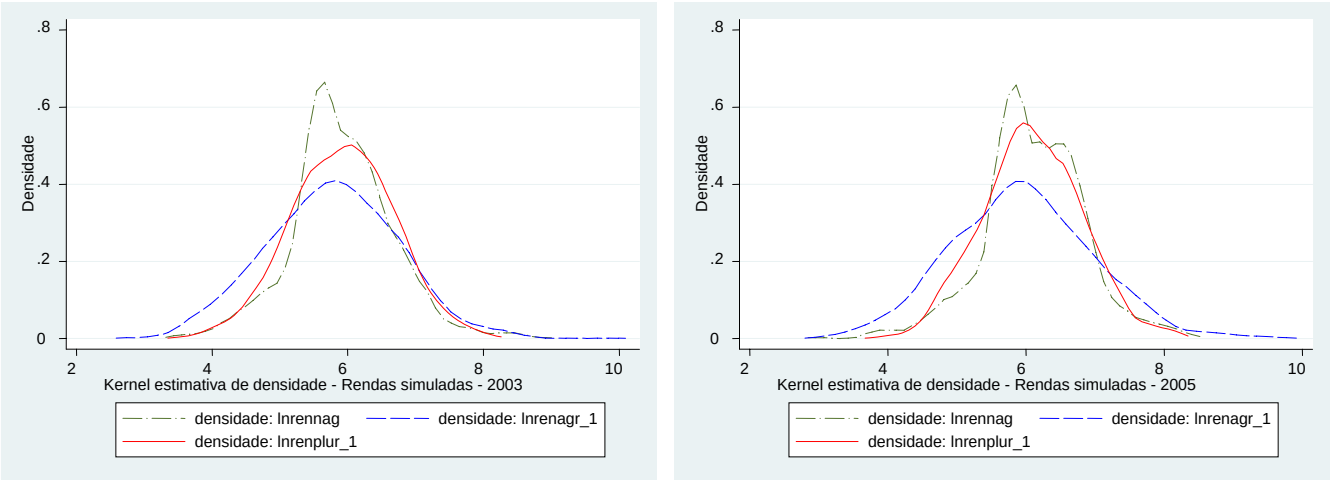
	Ano de 2003			Ano de 2005		
Variáveis	Coefic.	Erro Padrão*	Valor-P	Coef.	Erro Padrão**	Valor-P
Idade da PEA	0,0260	0,0031	0,0000	0,0213	0,0035	0,0000
Anos de Estudo	0,1038	0,0187	0,0000	0,0820	0,0193	0,0000
N° Compon. da Família	0,1234	0,0211	0,0000	0,1109	0,0300	0,0000
Local de Moradia	0,0612	0,0960	0,5240	0,0675	0,1074	0,5300
Razao Dependência	-0,1246	0,0511	0,0150	-0,0011	0,0559	0,9840
Piauí	-0,1149	0,1872	0,5390	-0,2370	0,1801	0,1880
Ceará	0,1403	0,1128	0,2130	-0,1022	0,1389	0,4620
Rio Grande do Norte	0,0418	0,0927	0,6520	-0,1267	0,1400	0,3650
Paraíba	-0,0252	0,1070	0,8130	-0,1327	0,1211	0,2730
Pernambuco	-0,0278	0,1097	0,8000	-0,1516	0,1150	0,1870
Alagoas	-0,0046	0,1809	0,9800	-0,0273	0,1259	0,8290
Sergipe	-0,1639	0,1464	0,2630	-0,2037	0,1937	0,2930
Bahia	0,0109	0,1009	0,9140	-0,1162	0,1102	0,2920
Lambda	-0,1019	0,1733	0,5570	-0,1422	0,1975	0,4710
Constante	4,3143	0,2261	0,0000	4,9498	0,2692	0,0000

Fonte: Dados da Pesquisa.

* Erro-padrão obtido por *bootstrap* com 70 replicações.

** Erro-padrão obtido por *bootstrap* com 1000 replicações.

A Figura 7 reporta as funções Kernel para o *log* das rendas observadas e simuladas das famílias exclusivamente não-agrícolas, caso fossem agrícolas ou pluriativas. Nos dois anos não se observa a presença de caudas mais pesadas na parte inferior da média. Especificamente em 2005, não é claro que se tenha deslocamento nas distribuições, ou seja, as médias aparentam ser bastante próximas. As curtoses das distribuições, por outro lado, são claramente distintas, indicando que a forma mais leptocúrtica é encontrada na renda observada das famílias não-agrícolas e a mais platicúrtica no caso da simulação, se as famílias fossem agrícolas.



Fonte: Dados da Pesquisa.

Figura 7 – Funções Kernel para os logaritmos das rendas observadas e simuladas das famílias não-agrícolas, caso fossem agrícolas ou pluriativas, Nordeste, anos de 2003 e 2005.

Para concluir a análise dos índices de pobreza para todos os tipos de famílias, na Tabela 26 são apresentados os valores considerando as rendas observadas das famílias não-agrícolas e as respectivas rendas simuladas, caso fossem agrícolas ou pluriativas, em 2003 e 2005. No ano de 2003, para a renda observada, o valor de P_0 equivale a 6,38%, P_1 a 2,06% e P_2 a 0,86%. Dentre todos os tipos de famílias, é a que apresenta a menor proporção de pobres, hiato da pobreza e severidade da pobreza, no ano de seca. Caso se tornassem exclusivamente agrícolas, isto acarretaria em grande aumento nos três índices, que passariam para 17,19%, 5,87% e 2,78%. No caso da pluriatividade, em todos os índices a diferença não é estatisticamente diferente a 10%.

Em 2005, os valores calculados para os índices FGT equivalem a 5,78%, 1,93% e 0,90%, com relação à renda observada. Caso passassem a ser exclusivamente agrícolas, os valores seriam 19,26%, 6,75% e 3,35%, ou seja, bastante elevados comparativamente e diferentes estatisticamente. Se fossem pluriativas, a proporção de pobres (P_0) é acrescida para 7,32% e P_1 para 1,60%, mas haveria redução para 0,53% na severidade da pobreza. Os valores de P_0 e P_1 são estatisticamente iguais aos obtidos com a renda observada, mas os de P_2 são diferentes a 10% de significância, conforme demonstrado no valor de probabilidade do teste de igualdade entre os índices, mostrado na Tabela 26.

Tabela 26 - Índices de pobreza (FGT) e MEFF para as rendas das famílias não-agrícolas: observada e simulada, caso fossem agrícolas e pluriativas, Nordeste, anos de 2003 e 2005

	Estimativa FGT	Erro Padrão	Intervalo de Confiança (95%)		Valor-P do teste de igualdade entre os índices	MEFF
Ano de 2003						
Renda Observada						
P_0	0,0638	0,0079	0,0482	0,0794	-	0,9794
P_1	0,0206	0,0031	0,0146	0,0267	-	1,0536
P_2	0,0086	0,0017	0,0054	0,0119	-	1,0978
Renda Simulada se Agrícolas						
P_0	0,1719	0,0138	0,1446	0,1992	0,0000	1,1628
P_1	0,0587	0,0062	0,0465	0,0709	0,0000	1,3820
P_2	0,0278	0,0036	0,0207	0,0349	0,0000	1,3794
Renda Simulada se Pluriativas						
P_0	0,0725	0,0108	0,0511	0,0938	0,5160	1,5950
P_1	0,0193	0,0036	0,0123	0,0264	0,7800	1,6525
P_2	0,0076	0,0016	0,0044	0,0108	0,6610	1,3606
Ano de 2005						
Renda Observada						
P_0	0,0578	0,0089	0,0402	0,0754	-	1,2890
P_1	0,0193	0,0037	0,0119	0,0266	-	1,3743
P_2	0,0090	0,0022	0,0046	0,0134	-	1,3621
Renda Simulada se Agrícolas						
P_0	0,1926	0,0154	0,1620	0,2231	0,0000	1,4540
P_1	0,0675	0,0070	0,0536	0,0814	0,0000	1,6716
P_2	0,0335	0,0042	0,0250	0,0419	0,0000	1,6053
Renda Simulada se Pluriativas						
P_0	0,0732	0,0086	0,0561	0,0903	0,1780	1,0226
P_1	0,0160	0,0022	0,0117	0,0203	0,3760	0,8924
P_2	0,0053	0,0010	0,0033	0,0073	0,0910	0,9818

Fonte: Dados da Pesquisa.

P_0, P_1 e P_2 indicam, respectivamente, a proporção de pobres, o hiato da pobreza e a severidade da pobreza.

Analisando a questão da desigualdade, a Tabela 27 trata do índice de Gini para as rendas das famílias agrícolas, observadas e simuladas, nos anos de 2003 e 2005. No ano de 2003, a renda média das famílias passando a ser pluriativas, se eleva de R$ 348,76 para R$ 380,71. No caso de passar a ser não-agrícola, este valor aumenta para

R$ 364,39. Contudo, apesar do valor mais elevado e do efeito positivo sobre a pobreza, a comparação entre os valores do índice de Gini mostra valores praticamente iguais para a renda observada e as simuladas (0,4027, 0,4095 e 0,3930).

Tabela 27 - Índice de concentração de Gini para as rendas das famílias agrícolas: observadas e simuladas, se fossem pluriativas e não-agrícolas. Erros padrão obtidos por *bootstrap (b)* com 200 replicações e linearização de Taylor (*lt*), Nordeste, anos de 2003 e 2005

Índice	Estimativa	Viés	Erro Padrão	Intervalo de Confiança com correção de viés (95%)	
Ano de 2003					
Renda Observada					
Gini (*b*)	0,4027	0,0012	0,0116	0,3813	0,4237
Gini (*lt*)	0,4056	-	0,0104	0,3853	0,4260
Renda Simulada se Pluriativas					
Gini *(b)*	0,4095	-0,0008	0,0072	0,3979	0,4264
Gini *(lt)*	0,4091	-	0,0053	0,3987	0,4194
Renda Simulada se Não-agrícolas					
Gini *(b)*	0,3930	-0,0012	0,0076	0,3791	0,4074
Gini *(lt)*	0,3914	-	0,0052	0,3813	0,4015
Ano de 2005					
Renda Observada					
Gini *(b)*	0,3886	0,0034	0,0094	0,3650	0,4008
Gini *(lt)*	0,3902	-	0,0077	0,3751	0,4053
Renda Simulada se Pluriativas					
Gini *(b)*	0,3915	-0,0015	0,0081	0,3778	0,4079
Gini *(lt)*	0,3895	-	0,0065	0,3769	0,4022
Renda Simulada se Não-agrícolas					
Gini *(b)*	0,3730	-0,0015	0,0066	0,3614	0,3889
Gini *(lt)*	0,3713	-	0,0044	0,3628	0,3799

Fonte: Dados da Pesquisa.

Para o ano de 2005, também não há mudanças elevadas no nível de concentração da renda das famílias agrícolas (0,3886), caso passassem para pluriativas (0,3915) ou não-agrícolas (0,3730). Do ponto de vista das famílias agrícolas, comparando com os valores obtidos na primeira simulação (considerando ou todos agrícolas ou todos pluriativos), esta abordagem mais desagrupada não confirma a hipótese de que as atividades/rendas não-agrícolas contribuem para reduzir a concentração, independentemente de o ano ser ou não com chuvas regulares e do ponto de vista das famílias agrícolas. Apenas caso passassem para exclusivamente

não-agrícolas é que haveria reduções nos índices nos dois anos. Por outro lado, confirma os resultados encontrados por Mariano e Neder (2004) e Ney e Hoffman (2007) de que as rendas não-agrícolas elevam a desigualdade no meio rural.

A Tabela 28 trata do índice de Gini considerando as rendas das famílias pluriativas observadas e simuladas, em 2003 e 2005. Como demonstrado na Tabela 24, para o ano de 2003, os valores calculados para as médias de cada renda são R$ 519,17, R$ 531,59 e R$ 443,46, respectivamente, para a renda observada e simulada se agrícola e não-agrícola. Já para o ano de 2005, os valores são R$ 676,20 R$ 681,75 e R$ 561,46. Para este tipo de família, observa-se a importância da renda não-agrícola para redução da concentração de renda. Os valores calculados com a renda observada são 0,4083 e 0,3908. Se passassem a ser exclusivamente agrícolas, o valor do Gini se elevaria para 0,5471 no ano de 2003 e para 0,5744 em 2005. Na comparação com as não-agrícolas, os valores são praticamente os mesmos, como indica a Tabela 28.

Tabela 28 - Índice de concentração de Gini para as rendas das famílias pluriativas: observada e simulada, se fossem agrícolas e não-agrícolas. Erros padrão obtidos por *bootstrap (b)* com 200 replicações e linearização de Taylor (*lt*), Nordeste, Anos de 2003 e 2005

Índice	Estimativa	Viés	Erro Padrão	Intervalo de Confiança com correção de viés (95%)	
Ano de 2003					
Renda Observada					
Gini (*b*)	0,4083	0,0039	0,0115	0,3682	0,4260
Gini (*lt*)	0,4140	-	0,0098	0,3948	0,4332
Renda Simulada se Agrícolas					
Gini *(b)*	0,5471	0,0034	0,0228	0,5025	0,5912
Gini *(lt)*	0,5492	-	0,0196	0,5108	0,5877
Renda Simulada se Não-Agrícolas					
Gini *(b)*	0,4042	-0,0013	0,0088	0,3912	0,4252
Gini *(lt)*	0,4042	-	0,0078	0,3889	0,4195
Ano de 2005					
Renda Observada					
Gini *(b)*	0,3908	0,0004	0,0106	0,3733	0,4168
Gini *(lt)*	0,3907	-	0,0080	0,3751	0,4063
Renda Simulada se Agrícolas					
Gini *(b)*	0,5744	-0,0024	0,0195	0,5369	0,6156
Gini *(lt)*	0,5728	-	0,0138	0,5456	0,5999
Renda Simulada se Não-Agrícolas					
Gini *(b)*	0,3865	-0,0012	0,0074	0,3726	0,3995
Gini *(lt)*	0,3852	-	0,0063	0,3729	0,3975

Fonte: Dados da Pesquisa.

Os valores calculados do índice de Gini para as rendas (observadas e simuladas) das famílias não-agrícolas são apresentados na Tabela 29. Com relação aos valores da renda observada, em 2003 é 0,4220, reduzindo-se para 0,3885 em 2005. Caso se tornassem exclusivamente agrícolas, as desigualdades se elevariam fortemente, com os índices se ampliando para 0,5258 e 0,5584. Se fossem pluriativas, o efeito sobre a concentração não seria significativo.

Tabela 29 - Índice de concentração de Gini para as rendas das famílias não-agrícolas: observadas e simuladas, se fossem agrícolas e pluriativas. Erros padrão obtidos por *bootstrap (b)* com 200 replicações e linearização de Taylor (*lt*), Nordeste, anos de 2003 e 2005

Índice	Estimativa	Viés	Erro Padrão	Intervalo de Confiança com correção de viés (95%)	
Ano de 2003					
Renda Observada					
Gini (*b*)	0,4220	0,0102	0,0317	0,3535	0,4819
Gini (*lt*)	0,4384	-	0,0325	0,3746	0,5021
Renda Simulada se Agrícolas					
Gini *(b)*	0,5258	-0,0014	0,0323	0,4760	0,6148
Gini *(lt)*	0,5302	-	0,0242	0,4827	0,5777
Renda Simulada se Pluriativas					
Gini *(b)*	0,4032	0,0069	0,0154	0,3752	0,4250
Gini *(lt)*	0,4100	-	0,0102	0,3900	0,4301
Ano de 2005					
Renda Observada					
Gini *(b)*	0,3885	0,0109	0,0299	0,3307	0,4292
Gini *(lt)*	0,4004	-	0,0360	0,3299	0,4708
Renda Simulada se Agrícolas					
Gini *(b)*	0,5584	0,0079	0,0245	0,4985	0,5952
Gini *(lt)*	0,5689	-	0,0204	0,5289	0,6088
Renda Simulada se Pluriativas					
Gini *(b)*	0,3908	-0,0006	0,0135	0,3664	0,4191
Gini *(lt)*	0,3913		0,0098	0,3720	0,4105

Fonte: Dados da Pesquisa.

Em resumo, os resultados das simulações, considerando o conjunto da região Nordeste mas desagregando por tipo de família demonstram, do ponto de vista das atuais famílias agrícolas, que a pluriatividade é importante para redução de todos os

índices de pobreza considerados, independentemente das condições climáticas. Entretanto, não tem influência na mesma proporção sobre a concentração de renda. Para as famílias pluriativas, o abandono das atividades não-agrícolas tem grande efeito na elevação de todos os indicadores de pobreza e de concentração. As famílias pluriativas e não-agrícolas são as que possuem a menor incidência de pobreza, comparado com as agrícolas.

O caso em questão é que para as famílias pluriativas, o abandono das atividades não-agrícolas gera efeitos negativos tanto nos indicadores de pobreza quanto no índice de concentração. Então, para estas famílias, a melhor estratégia de sobrevivência é realmente diversificar suas atividades. No caso das famílias exclusivamente agrícolas, a pluriatividade consegue reduzir a pobreza mas não a concentração. Dado que a pluriatividade é importante para melhorar a condição das famílias que estão abaixo da linha da pobreza, seu estímulo deve ser combinado com outros tipos de incentivos visando a reduzir a concentração de renda existente no meio rural da região.

4.3.3 Simulação considerando apenas as famílias nos estados do Maranhão, Piauí e Rio Grande do Norte

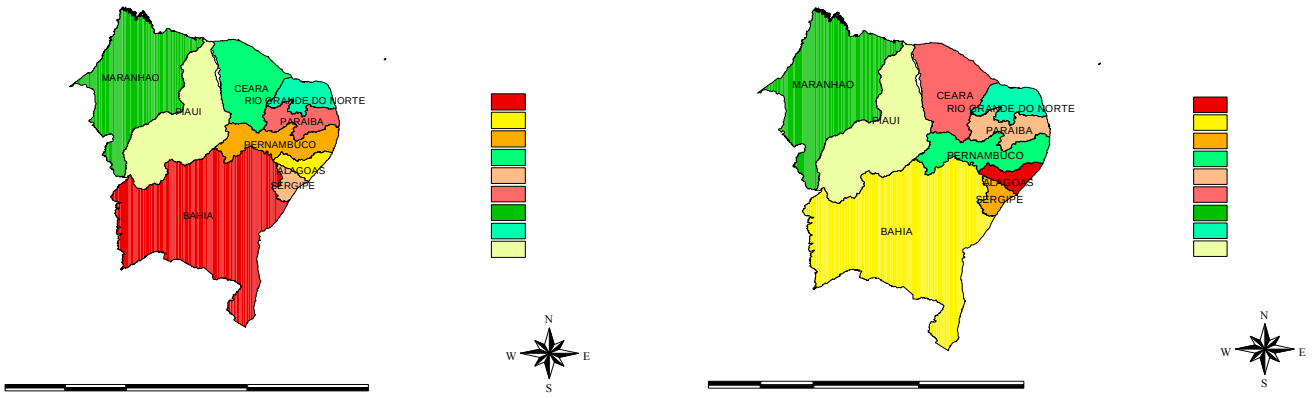
Tão importante quanto analisar as características intrínsecas das famílias é entender o que De Janvry *et al.* (2005) definiram como "*village fixed effects"*, ou seja, um conjunto de variáveis que captam o impacto do desenvolvimento econômico local sobre a renda e as atividades não-agrícolas das famílias. Na base de dados da Pnad não se conseguem variáveis que captem o grau de desenvolvimento da localidade de residência das famílias, as relações sociais ou a influência que a proximidade de uma cidade "pólo", como Caruaru-PE, Campina Grande-PB ou Mossoró-RN, irradiam nas famílias de cidades circunvizinhas.

Este item, então, refaz as análises, considerando agora apenas os estados que podem ser considerados com "maior propensão" (maior percentual de PEA rural pluriativa e não-agrícola em relação à agrícola) para as atividades não-agrícolas e tenta observar se o contexto em que as famílias estão inseridas influencia na relação entre atividades/rendas não-agrícolas, pobreza e concentração. É outra possibilidade para

tentar entender melhor o impacto sobre a desigualdade da renda, pois a abordagem mais desagrupada do item anterior não consegue confirmar a hipótese de que as atividades/rendas não-agrícolas contribuem para reduzir a concentração, indicada na primeira simulação.

Considerando apenas os estados com maior percentual de PEA rural pluriativa e não-agrícola em relação à agrícola, imagina-se que os efeitos locais sejam mais fortes e que a conseqüência sobre a pobreza e a concentração seja mais evidenciada. Este tópico centra a análise apenas nas famílias agrícolas e pluriativas, pois parece mais esclarecido o impacto na pobreza e na concentração das famílias exclusivamente não-agrícolas.

O Mapa 1 reporta as proporções de famílias do tipo não-agrícola e pluriativa sobre o total, residentes no meio rural dos nove estados da região Nordeste. Observa-se que os três estados com as maiores proporções, tanto em 2003 quanto em 2005, são: Piauí, Maranhão e Rio Grande do Norte. Em 2005, inclusive, os três estados possuem mais de 50% das famílias com pelo menos um membro ocupado em atividade não-agrícola, sendo que, desde o início dos anos 90, a PEA rural ocupada em atividades não-agrícolas supera o número de pessoas ocupadas na agricultura no Rio Grande do Norte (GOMES DA SILVA, 2000).



Fonte: Dados da Pesquisa. Pnad (2003) e (2005).

Mapa 1- Proporções de famílias do tipo não-agrícola e pluriativa sobre o total de famílias no meio rural, Nordeste, anos de 2003 e 2005.

O Maranhão (MA), desde o início dos anos 90, possui taxas bastante elevadas de crescimento das atividades não-agrícolas em quase todos os ramos de atividades. Isto é devido à existência de fronteiras agrícolas, principalmente no caso da soja, que contribui para geração de Orna (GOMES DA SILVA e VERAS, 2001).

O estado do Piauí (PI) também apresenta na porção Sul/Sudoeste uma parte da fronteira agrícola com produção de soja e arroz. A produção de mel e castanha de caju também são importantes atividades agrícolas que geram atividades não-agrícolas derivadas. O setor têxtil alavancado pelo pólo de confecção de Teresina também é importante fonte de geração de emprego para a mão-de-obra rural residente nos municípios próximos. Embora nos anos 90 ainda não transparecesse nos dados o peso do não-agrícola, suas taxas de crescimento se elevavam, principalmente quando combinadas com atividades agrícolas. Como observa Vilela (2000, p. 66), no "Piauí, o que vem assumindo grande importância é a combinação da relação entre atividades agrícolas e não-agrícolas, formando uma 'cadeia', uma articulação a partir do qual é potencializada a geração de índices significativos de postos de trabalho".

Historicamente, a população do Rio Grande do Norte sempre conviveu com alternativas não-agrícolas à ocupação agrícola. Entre essas ocupações, destacam-se a exploração do sal (Oeste do estado) e a mineração (região do Seridó). Dado o caráter não-empresarial da maior parte das atividades agrícolas no estado (exceto o pólo Assú-Mossoró), nos períodos de seca ou sazonalidade a população não migrava para centros urbanos maiores, pois podia compatibilizar a ocupação em atividades não-agrícolas e se manter no rural potiguar. Soma-se a isso uma importante interiorização de serviços ligados a instituições como Sesi/Senai/Sebrae e tem-se a explicação da grande parcela da PEA ocupada em Orna (GOMES DA SILVA, 2000; GOMES DA SILVA e VERAS, 2001).

As estatísticas descritivas das variáveis, considerando apenas estes três estados, possuem diferenças dos valores para a região como um todo, conforme pode ser observado nas Tabelas 30 e 31. Em 2003, com relação à educação, a média dos anos de estudo das famílias é de 2,99, com o limite superior do intervalo de confiança chegando a 3,6 anos. Para o ano de 2005, o valor médio sobe para 3,37 anos. Na

comparação com os valores regionais, nos anos de 2003 e 2005, a média de idade é menor (30,18 e 29,62), mas o número de componentes da família é maior (4,23 e 4,20), como indicado na Tabela 30.

Tabela 30 – Estatística descritiva (média e erro-padrão) das variáveis, considerando o plano amostral, Maranhão, Piauí e Rio Grande do Norte, anos 2003 e 2005

Variável	Média	Erro-Padrão Linearizado	Intervalo de Confiança (95%)	
Ano de 2003				
Renda agrícola	295,1867	22,0894	249,5964	340,7771
Renda Pluriativ.	466,6827	27,1890	410,4379	522,9276
Idade PEA	30,1768	0,2218	29,7199	30,6338
Anos de estudo	2,9859	0,3139	2,3392	3,6323
N° Compon. da Família	4,2334	0,0617	4,1063	4,3605
Ano de 2005				
Renda agrícola	377,4730	24,6424	326,9110	428,0350
Renda Pluriativ.	588,9365	36,5491	513,6623	664,2107
Idade PEA	29,6197	0,36072	28,8955	30,3734
Anos de estudo	3,3656	0,3072	2,7364	3,9949
N° Compon. da Família	4,1982	0,08557	4,0229	4,3734

Fonte: Dados reprocessados da Pnad (2003) e Pnad (2005), pelo autor.

A Tabela 31 se refere às proporções das variáveis dos estados em análise. A proporção de pessoas que não possui renda do não trabalho (43,40% em 2003 e 34,23% em 2005), conjuntamente nos estados do PI, MA e RN, é mais elevada do que a proporção regional (37,12% e 31,6% em 2003 e 2005, respectivamente). Nestes estados, existe uma maior quantidade de famílias do tipo conta-própria (55,11% e 54,43%), relativamente aos dados da região.

A maior parte das famílias reside no meio rural mais distante do urbano (83%). Estas famílias têm menos acesso à energia elétrica, dado que apenas 65,7% declararam ter a eletricidade como fonte de iluminação do domicílio, em 2005. Por outro lado, é bem maior a proporção de pessoas com acesso à água de poço dentro da propriedade, respectivamente, 39,3% e 36,6% em 2003 e 2005. Em nível regional, estes valores são 29,2% e 29,1%.

Tabela 31 – Estatística descritiva (Proporção e erro padrão) das variáveis, considerando o plano amostral, Maranhão, Piauí e Rio Grande do Norte, 2005

Variável	Proporção %	Erro Padrão Linearizado	Intervalo de Confiança (95%) %	
Ano de 2003				
Renda não-trabalho				
não-possui	43,3997	0,0252	38,2019	48,5975
possui	56,6600		51,4024	61,7980
Posição na ocupação				
conta-própria	55,1108	0,0364	47,6089	62,6129
empregado	30,2476	0,0286	24,3468	36,1484
Local de moradia				
Rural Agropecuário	83,0470	0,0480	73,1466	92,9473
Iluminação domiciliar				
Elétrica	61,9798	0,0619	49,2109	74,7487
Óleo, querosene ou gás	36,8890	0,0596	24,6128	49,1668
Água de poço				
sim	39,3050	0,0536	28,2563	50,3546
não	25,6900	0,0406	17,3119	34,0682
Ano de 2005				
Renda não-trabalho				
não-possui	34,2273	2,4714	29,1649	39,2897
possui	65,7727		60,7103	70,8351
Posição na ocupação				
conta-própria	54,4287	03,0091	48,2648	60,5926
empregado	32,9582	02,1473	28,5596	37,3567
Local de moradia				
Rural Agropecuário	83,9958	04,6047	74,5634	93,4282
Iluminação domiciliar				
Elétrica	65,6800	06,1635	53,0547	78,3054
Óleo, querosene ou gás	33,6087	06,3379	20,6261	46,5914
Água de poço				
sim	36,6345	05,5426	25,2811	47,9880
não	25,5766	04,5851	16,1845	34,9688

Fonte: Dados reprocessados da Pnad (2005), pelo autor.

A Tabela 32 reporta os valores estimados dos índices de pobreza FGT, tanto para a renda observada das famílias agrícolas, quanto para a renda simulada, nos anos de 2003 e 2005. Como se pode observar, tanto para a proporção de pobres, quanto para o hiato da pobreza e a severidade da pobreza, os valores se reduzem bastante no caso das famílias agrícolas, se fossem pluriativas. Para a renda observada em 2003, os valores calculados são 24,11%, 10,09% e 6,08%. Com as rendas simuladas, estes

valores se reduzem para 16,89%, 4,05% e 1,55%, sendo considerados diferentes estatisticamente dos primeiros a 5% de probabilidade.

Tabela 32- Índices de pobreza (FGT) e MEFF para rendas das famílias agrícolas: observadas e simuladas, caso fossem pluriativas, Maranhão, Piauí e Rio Grande do Norte, anos de 2003 e 2005

	Estimativa FGT	Erro Padrão	Intervalo de Confiança (95%)		Valor-P teste de igualdade entre os índices	MEFF
Ano de 2003						
Renda Observada						
P_0	0,2411	0,0348	0,1692	0,3130	-	3,9711
P_1	0,1009	0,0178	0,0641	0,1377	-	3,9366
P_2	0,0608	0,0139	0,0320	0,0896	-	3,9603
Renda Simulada se Pluriativa						
P_0	0,1689	0,0136	0,1407	0,1970	0,0290	0,8014
P_1	0,0405	0,0053	0,0296	0,0514	0,0000	1,2116
P_2	0,0155	0,0031	0,0092	0,0219	0,0010	1,2450
Ano de 2005						
Renda Observada						
P_0	0,1939	0,0444	0,1028	0,2850	-	7,2736
P_1	0,0927	0,0215	0,0486	0,1369	-	5,4323
P_2	0,0596	0,0139	0,0310	0,0882	-	3,8068
Renda Simulada se Pluriativa						
P_0	0,1037	0,0134	0,0761	0,1312	0,0460	1,0954
P_1	0,0284	0,0039	0,0205	0,0363	0,0040	0,7374
P_2	0,0118	0,0020	0,0077	0,0158	0,0010	0,6609

Fonte: Dados da Pesquisa.

P_0, P_1 e P_2 indicam, respectivamente, a proporção de pobres, o hiato da pobreza e a severidade da pobreza.

Para a proporção de pobres, isto representa uma redução de aproximadamente 30%, para o hiato da pobreza, cerca de 60% e para a pobreza severa, uma diminuição de 75%. Para o ano de 2005, os valores dos índices são menores, algo esperado por se tratar de um ano sem seca. No ano de chuvas regulares, a combinação de atividades agrícolas com não agrícolas reduz P_0, P_1 e P_2 para 10,37%, 2,84% e 1,18%, nesta ordem.

Esta desagregação, então, é importante para demonstrar que nos lugares em que as famílias têm mais acesso às atividades não-agrícolas e que o nível educacional médio é maior, a pluriatividade contribui fortemente para redução da pobreza,

independentemente do índice FGT analisado. Isto reforça o resultado encontrado na análise regional, cujo peso para os mais pobres entre os pobres de uma renda não-agrícola (que por suas características é também mais estável ao longo do ano, além de mais elevada do que a agrícola) é fundamental para redução da situação de miséria em que muitas famílias do rural nordestino se encontram.

A Tabela 33 se refere aos índices de concentração de Gini para a renda observada das famílias agrícolas e a simulada, se fossem pluriativas, com os erros padrão obtidos por *bootstrap* e linearização de Taylor. A comparação entre estes índices mostra valores maiores, ou seja, maior concentração, no caso da renda observada, tanto no ano seco (0,4117 contra 0,4038 da renda simulada) como no chuvoso (0,4045 contra 0,3805). Este resultado difere do encontrado com os dados agrupados para todos os estados da região, dado que está indicando que nos estados com "maior propensão" para as atividades não-agrícolas, elas também podem contribuir para reduzir as desigualdades de renda.

Tabela 33 - Índice de concentração de Gini para as rendas das famílias agrícolas: observadas e simuladas, se fossem pluriativas. Erros padrão obtidos por *bootstrap (b)* com 200 replicações e linearização de Taylor (*lt*). Maranhão, Piauí e Rio Grande do Norte, anos de 2003 e 2005

Índice	Estimativa	Erro Padrão	Intervalo de Confiança com correção de viés (95%)	
Ano de 2003				
Renda Observada				
Gini (*b*)	0,4117	0,0221	0,3726	0,4516
Gini (*lt*)	0,4162	0,0209	0,3752	0,4571
Renda Simulada se Pluriativa				
Gini *(b)*	0,4038	0,0196	0,3774	0,4652
Gini *(lt)*	0,3980	0,0135	0,3716	0,4244
Ano de 2005				
Renda Observada				
Gini *(b)*	0,4045	0,0230	0,3657	0,4535
Gini *(lt)*	0,4026	0,0192	0,3649	0,4403
Renda Simulada se Pluriativa				
Gini *(b)*	0,3805	0,0226	0,3463	0,4434
Gini *(lt)*	0,3713	0,0134	0,3451	0,3975

Fonte: Dados da Pesquisa.

Numa região em que a atividade agropecuária não incorporou elevado nível tecnológico, ou seja, onde a maior parte dos produtores usa para arar a terra tração

animal ou mesmo apenas enxadas, as diferenças de rendimento estão bastante correlacionadas com a disponibilidade de terras. Esta disponibilidade, por sua vez, sabe-se que é extremamente concentrada no Nordeste como um todo. Esta relação entre concentração de terras e concentração de renda não possui simetria quando a renda em questão provém de atividades não-agrícolas. Consequentemente, o acesso as rendas não-agrícolas pode sim ter efeito positivo na redução da desigualdade da renda.

A Tabela 34 se refere aos índices FGT para as rendas das famílias pluriativas. Os três índices, quando considerada a renda observada, apresentam valores menores comparativamente aos obtidos com a renda simulada. Isto significa que, se as famílias pluriativas não tivessem a renda não-agrícola, haveria um significativo aumento na proporção de pobres (P_0), a qual passaria de 12,13% para cerca de 23,19% em 2003 e de 8,36% para 22,70% em 2005.

Tabela 34- Índices de pobreza (FGT) e MEFF para rendas das famílias pluriativas: observadas e simuladas, caso fossem agrícolas. Maranhão, Piauí e Rio Grande do Norte, Anos de 2003 e 2005

	Estimativa FGT	Erro Padrão	Intervalo de Confiança (95%)		Valor-P do teste de igualdade entre os índices	MEFF
Ano de 2003						
Renda Observada						
P_0	0,1213	0,0249	0,0698	0,1728	-	2,9299
P_1	0,0423	0,0104	0,0207	0,0639	-	2,9649
P_2	0,0200	0,0060	0,0075	0,0324	-	2,7331
Renda Simulada se Agrícola						
P_0	0,2319	0,0194	0,1918	0,2719	0,0070	1,0445
P_1	0,0806	0,0089	0,0623	0,0990	0,0130	1,1838
P_2	0,0393	0,0047	0,0297	0,0490	0,0160	0,9245
Ano de 2005						
Renda Observada						
P_0	0,0836	0,0151	0,0524	0,1147	-	1,7127
P_1	0,0246	0,0051	0,0141	0,0351	-	1,6504
P_2	0,0104	0,0027	0,0050	0,0159	-	1,5087
Renda Simulada se Agrícola						
P_0	0,2270	0,0219	0,1818	0,2722	0,0000	1,6172
P_1	0,0863	0,0084	0,0690	0,1035	0,0000	1,0644
P_2	0,0454	0,0051	0,0349	0,0559	0,0000	0,9702

Fonte: Dados da Pesquisa.

P_0, P_1 e P_2 indicam, respectivamente, a proporção de pobres, o hiato da pobreza e a severidade da pobreza.

A Tabela 35 trata do índice de Gini considerando as rendas em questão. Do ponto de vista das famílias pluriativas, observa-se a importância da renda não-agrícola para redução da concentração de renda. Isto devido aos valores calculados para o Gini serem maiores para a renda simulada comparando com a renda observada (com renda não-agrícola). De forma semelhante ao que ocorre nos itens anteriores, os erros padrão calculados por linearização de Taylor são ligeiramente menores do que aqueles calculados por *bootstrap*.

Tabela 35 - Índice de concentração de Gini para as rendas das famílias pluriativas: observadas e simuladas, caso fossem agrícolas. Erros padrão obtidos por *bootstrap* com 200 replicações e Linearização de Taylos. Maranhão, Piauí e Rio Grande do Norte, anos de 2003 e 2005

Índice	Estimativa	Erro Padrão	Intervalo de Confiança com correção de viés (95%)	
Ano de 2003				
Renda Observada				
Gini *(b)*	0,4308	0,0198	0,3926	0,4676
Gini *(lt)*	0,4355	0,0134	0,4093	0,4617
Renda Simulada se Agrícola				
Gini *(b)*	0,5321	0,0359	0,4792	0,6111
Gini *(lt)*	0,5413	0,0319	0,4788	0,6039
Ano de 2005				
Renda Observada				
Gini *(b)*	0,4155	0,0196	0,3890	0,4765
Gini *(lt)*	0,4087	0,0161	0,3771	0,4403
Renda Simulada se Agrícola				
Gini *(b)*	0,5944	0,0391	0,5527	0,7138
Gini *(lt)*	0,5845	0,0277	0,5303	0,6388

Fonte: Dados da Pesquisa.

Este tópico aprofunda o entendimento do efeito do não-agrícola sobre a pobreza e a concentração. A nova desagregação, considerando apenas os estados do Maranhão, Piauí e Rio Grande do Norte, ou seja, aqueles com maior proporção de famílias do tipo não-agrícola e pluriativa sobre o total de famílias, confirma que o não-agrícola é importante para reduzir a pobreza rural e demonstra que pode contribuir também para redução da desigualdade.

Na comparação com os resultados para todos os estados da região, obtém-se a confirmação de que o efeito na pobreza é bastante relevante, reduzindo no caso da família agrícola passar a ser pluriativa e elevando, no caso da família pluriativa passar

a ser exclusivamente agrícola. A grande diferença se dá no resultado sobre a concentração, pois no caso das simulações com apenas os três estados supracitados, a conclusão a que se chega é de que a desigualdade pode ser reduzida, mesmo que ligeiramente, no caso da família agrícola passar a ser pluriativa. Ao considerar todos os estados, o índice de Gini aumentava na simulação da renda da família agrícola passando para pluriativa.

4.4 Elasticidades da pobreza em relação ao crescimento da renda familiar agrícola observada e simulada e ao índice de Gini

Neste item, analisam-se as estimações das elasticidades crescimento-renda da pobreza (Elcres) e elasticidades Gini da pobreza (Elgini), nos anos de 2003 e 2005, considerando os dados regionais. Uma análise desagregada neste item serviria apenas para identificar mudanças nos valores pontuais das elasticidades, sem alterar sua direção. Dado que o interesse maior desta pesquisa não é o tamanho e sim o sentido do "efeito", não há prejuízo em analisar regionalmente. A elasticidade crescimento-renda da pobreza se refere ao que ocorre com os índices FGT, se o crescimento da renda sofrer variação e a elasticidade Gini da pobreza, ao impacto na pobreza de uma variação percentual no índice de Gini.

A análise deste item está restrita às famílias agrícolas e pluriativas, mas com simulações caso fossem não-agrícolas, já que as famílias exclusivamente não-agrícolas são as que apresentam os menores valores tanto para os índices de pobreza quanto de concentração, não fazendo sentido ver o impacto que teria se passassem a ser de outro tipo. Os valores calculados estão mostrados na Tabela 36. Com relação à elasticidade crescimento-renda, nos dois anos e para todos os tipos de famílias, os valores calculados são maiores do que a unidade (em módulo), ou seja, uma variação percentual no crescimento da renda tem impacto mais do que proporcional na redução da pobreza. Observa-se que as maiores elasticidades estão em P_1 e P_2, ou seja, entre os mais pobres dentre os pobres. Os valores são ainda mais elevados para as famílias agrícolas, se fossem pluriativas ou não-agrícolas, para as famílias pluriativas e para estas, se exclusivamente não-agrícolas.

Se as famílias agrícolas fossem pluriativas, uma elevação em suas rendas de 10% reduziria em 38,51% a severidade da pobreza neste tipo de família. Com relação às rendas observadas, um aumento de 10% na renda das famílias agrícolas reduz em 14,36%, 16,29% e 18,42%, os valores dos índices FGT, P_0, P_1 e P_2, respectivamente. Para as famílias pluriativas, a elevação nos rendimentos em 10% reduz os índices em 17,13%, 22,57% e 28,18%.

Tabela 36 - Elasticidade da pobreza em relação ao crescimento da renda familiar e ao índice de Gini. Famílias agrícola e pluriativa. Renda observada e simulada, Nordeste, anos de 2003 e 2005

	2003		2005	
	Elcres	Elgini	Elcres	Elgini
FAMÍLIA AGRÍCOLA				
Renda Observada				
P_0	-1,4356	2,7366	-1,4300	2,5563
P_1	-1,6292	6,0118	-1,5881	5,6267
P_2	-1,8421	9,3239	-1,7687	8,7374
Renda Simulada se Pluriativa				
P_0	-2,0347	4,4202	-2,3578	4,7643
P_1	-2,9385	9,5562	-3,5027	10,0983
P_2	-3,8517	14,7125	-4,6669	15,4713
Renda Simulada se Não-agrícola				
P_0	-2,2089	4,4985	-2,3772	5,2320
P_1	-3,2770	9,7104	-3,5715	11,0616
P_2	-4,3558	14,9439	-4,7834	16,9299
FAMÍLIA PLURIATIVA				
Renda Observada				
P_0	-1,7134	5,6994	-2,5951	8,6042
P_1	-2,2568	11,8332	-4,0150	17,6273
P_2	-2,8179	18,0260	-5,4594	26,7152
Renda Simulada se Agrícola				
P_0	-1,4449	4,9557	-1,3357	4,7347
P_1	-2,0220	11,3650	-1,9037	11,2929
P_2	-2,5901	17,7054	-2,4410	17,7421
Renda Simulada se Não-agrícola				
P_0	-2,7184	7,3274	-2,8547	7,8307
P_1	-4,3728	15,4822	-4,5346	16,1818
P_2	-6,0317	23,6494	-6,2331	24,5840

Fonte: Dados da Pesquisa.

Elcres=Elasticidade crescimento-renda da pobreza; Elgini=Elasticidade Gini da pobreza.

P_0, P_1 e P_2 indicam, respectivamente, a proporção de pobres, o hiato da pobreza e a severidade da pobreza.

Sobre as elasticidades crescimento para o ano de 2005, igualmente se observa que as elasticidades mais elevadas estão em P_1 e P_2 e que são maiores para as famílias agrícolas, se fossem pluriativas ou não-agrícolas, para as famílias pluriativas e para estas, se exclusivamente não-agrícolas. Se as famílias agrícolas fossem pluriativas, um aumento de 10% em suas rendas reduziria o P_2 em 46,67% neste tipo de família. Com relação às rendas observadas, uma elevação de 10% na renda das famílias agrícolas reduz em 14,28%, 15,88% e 17,69% os valores dos índices FGT, P_0, P_1 e P_2, respectivamente. Para as famílias pluriativas, a elevação nos rendimentos em 10% reduz os índices em 25,95%, 40,15% e 54,59%, como indicado na Tabela 36.

Com relação à elasticidade Gini, um primeiro aspecto a ser salientado é que os valores calculados são mais elevados do que os da elasticidade crescimento-renda da pobreza, principalmente em P_1 e P_2. Como cita Neder (2003b, p. 33), "verifica-se que as medidas de pobreza rural são mais sensíveis a mudanças na desigualdade do que a mudanças nas rendas médias". Também vale ressaltar que os valores observados das elasticidades no ano de 2005, para as famílias pluriativas, são maiores do que os encontrados para 2003, indicando que a pobreza está se tornando mais facilmente aliviada devido ao crescimento econômico e pela redistribuição da renda neste tipo de família. Para as famílias agrícolas, isto não ocorre, indicando a dificuldade de redução da pobreza para elas.

Sobre os valores, com relação à renda observada para as famílias agrícolas, em 2003, uma redução de 1% na concentração de renda reduz os valores do FGT em aproximadamente 2,74% 6,01% e 9,32%. Estes valores poderiam ser mais elevados caso as famílias fossem pluriativas ou não-agrícolas, como demonstra a simulação na Tabela 36. Especificamente, caso se tornassem pluriativas, a redução seria de 4,42%, 9,56% e 14,71%. Em todos os casos, um aumento bastante significativo.

Para as famílias pluriativas, a redução na concentração faz decrescer em 5,70%, 11,83% e 18,03% os índices FGT. Então, deste ponto de vista, a pluriatividade também é importante para redução da pobreza. No ano de 2005, estes valores são ainda maiores, 8,60%, 17,63% e 26,71%, respectivamente, de redução para P_0, P_1 e P_2, se a concentração se reduzisse em 1%. Se estas famílias fossem agrícolas, a

desconcentração também contribuiria para diminuir a pobreza, principalmente no hiato e na severidade da pobreza, mas em menores proporções, 11,29% e 17,74%, como apresentado na Tabela 36.

Com isto, pode-se agregar mais este conjunto de informações para confirmar que comparando os diversos tipos de famílias, naquelas com atividade não-agrícola a pobreza pode ser aliviada com maior intensidade se ocorrer uma variação positiva na renda, principalmente nas famílias que mais precisam. Além disso, reforça o quão fundamental é reduzir as desigualdades e como isto se relaciona de forma positiva com a diminuição da pobreza.

5. RESUMO E CONCLUSÕES

A região Nordeste concentra a maior parte dos agricultores familiares do país, sendo que a maioria deles não se modernizou, não estão integrados em complexos agroindustriais e convivem com as maiores adversidades climáticas, tornando a atividade agrícola ainda mais arriscada. Apesar da grande heterogeneidade regional, a pobreza tanto urbana quanto rural é uma característica presente em praticamente todas as cidades nordestinas de pequeno porte. Apesar de esforços serem feitos para redução da pobreza intra-regional, ao comparar os dados com os de outras regiões é possível verificar que no Nordeste se encontra a maioria dos pobres brasileiros.

Se o Brasil pretende se aproximar dos ditos "países desenvolvidos", necessariamente precisa equacionar o subdesenvolvimento do meio rural nordestino, criando políticas para "puxar" as famílias para a parte de cima da linha da pobreza. Estas políticas precisam ser orquestradas em várias frentes, pois isoladamente não se conseguem resultados satisfatórios. O Bolsa-Família é um exemplo disto.

Dentre as políticas públicas possíveis, o estímulo às atividades não-agrícolas para famílias que residem no meio rural pode contribuir neste processo, desde que esteja esclarecido se o acesso a estas políticas tem impacto positivo na redução da pobreza e/ou concentração de renda. Por este motivo, o objetivo do presente estudo foi demonstrar o efeito da renda e da atividade não-agrícola - enquanto estratégia de sobrevivência para as famílias - sobre a pobreza e a desigualdade rural da região

Nordeste, considerando um ano de seca (2003) e outro com chuvas regulares (2005) na maior parte da região. Especificamente, buscou-se entender melhor os determinantes da escolha familiar em ser exclusivamente agrícola, pluriativa, não-agrícola ou estar não-ocupada, simular as rendas dos diversos tipos de famílias, identificando o impacto na renda total do não-agrícola e, consequentemente, na pobreza e na concentração.

O referencial teórico utilizado neste trabalho está relacionado à decisão de oferta de mão-de-obra rural, focando a possibilidade de as famílias poderem se engajar apenas em atividades agrícolas e/ou não-agrícolas. Segundo esta abordagem, a família compara as opções de trabalho e aloca seu tempo total disponível de forma a maximizar sua função conjunta de utilidade.

A metodologia utilizada para alcançar os objetivos propostos foi baseada, principalmente, em estimação de modelos de seleção amostral com logit multinomial, seguida dos cálculos dos índices de pobreza FGT (Foster-Greer-Thorbecke), do índice de Gini para avaliar concentração e as elasticidades crescimento-renda e Gini da pobreza. Para atingir os objetivos, foram utilizados os dados da Pnad de 2003 e 2005, com vistas a captar as diferenças existentes nos anos com chuvas irregulares e regulares na região Nordeste. Todas as estimações foram feitas considerando a Pnad amostra complexa, ou seja, incorporando os pesos, estratos, múltiplos estágios e as probabilidades desiguais de seleção.

Inicialmente, procurou-se estimar os determinantes de alternativas de ocupação (agrícola, não-agrícola, pluriativa, não-ocupado). Estes determinantes demonstram que quanto maior a escolaridade maior a chance de posicionar no mercado de trabalho não-agrícola. Um maior número de componentes da família eleva a probabilidade possibilidade de ser pluriativa ou não estar sem ocupação. As famílias exclusivamente agrícolas e de não-ocupados são mais dependentes da renda do não-trabalho, como as aposentadorias, pensões, bolsas do governo, quando comparadas as famílias não-agrícolas. As famílias de conta própria possuem maior chance de ser pluriativas, enquanto as de empregados, maior possibilidade de estar não-ocupadas ou ser exclusivamente não-agrícolas. Residir no estado do Piauí eleva a probabilidade de a família ser não-agrícola. Por outro lado, na Bahia e Sergipe, as possibilidades são

menores. A análise feita considerando o ano de seca e chuvoso serviu para demonstrar que o clima não afeta a decisão da família. Não é possível observar mudanças nos parâmetros que alteram os resultados comparando os dois anos em análise (2003 e 2005).

Em seguida, investigou-se qual o impacto da ocupação/renda não-agrícola sobre a pobreza e a concentração. A primeira simulação feita considerou o caso de todas as famílias serem agrícolas e, de forma inversa, todas pluriativas. Nos dois anos, os índices FGT de proporção de pobres (P_0), hiato da pobreza (P_1) e severidade da pobreza (P_2) foram mais elevados para as famílias agrícolas e se reduzem consideravelmente, principalmente, em P_1 e P_2, ou seja, nos mais pobres entre os pobres. Com relação à concentração, o impacto da pluriatividade para redução do índice de Gini se apresenta bastante significativo nos dois anos analisados. De acordo com esta simulação, conclui-se que a pluriatividade reduz tanto a pobreza quanto a concentração de renda, independentemente da questão climática.

Posteriormente, foram feitas diversas simulações de forma mais desagregada por família. Com relação à pobreza, no caso das famílias agrícolas, os resultados indicam que a pluriatividade isolada é importante para redução da proporção de pobres, do hiato e da severidade da pobreza. Os três índices FGT também se reduzem no caso de as famílias agrícolas passarem a ser exclusivamente não-agrícolas.

Do ponto de vista das famílias pluriativas, é nítido que o abandono do não-agrícola, mantendo, obviamente, as condições atuais da agropecuária nordestina, eleva acentuadamente a pobreza rural. Isto vale para ratificar que a pluriatividade não é um fenômeno passageiro, sendo de grande valia para aumento no bem-estar deste tipo de famílias. Dentre todos os tipo de famílias, as exclusivamente não-agrícolas são as que possuem a menor proporção de pobres, hiato da pobreza e severidade da pobreza. Este resultado é esperado, pois é este tipo de família que percebe as rendas mais elevadas, além de estáveis durante todo o ano. É diferente da renda agrícola, em que o produtor a cada mês de safra tem despesas, só conseguindo receita no final da colheita, com a venda do produto.

Na análise da concentração, para as atuais famílias pluriativas, se fossem agrícolas, o valor do índice de Gini se elevaria, confirmando o resultado na primeira simulação realizada no estudo. Para as rendas das famílias agrícolas, se fossem pluriativas, o resultado seria diferente daquele obtido com a primeira simulação realizada, cuja passagem de agrícola para pluriativo reduz a concentração, já que o índice de Gini calculado aumentou. Contudo, isto confirma outros estudos (que também consideram os dados agregados) que indicam que a parcela da renda não-agrícola contribui para elevar a desigualdade.

Considerando os dados agregados para a região Nordeste, o efeito concentrador da parcela da renda não-agrícola é bastante influenciado pela renda das famílias residentes em estados em que a dinâmica de geração das atividades não-agrícolas não é bastante desenvolvida. Esta conclusão foi obtida após novas análises, considerando apenas os estados do Maranhão, Piauí e Rio Grande do Norte, os quais possuem maior percentual de PEA pluriativa e não-agrícola sobre a PEA total. As simulações, então, confirmam que o não-agrícola também pode contribuir para reduzir a desigualdade. A redução, mesmo pequena, é importante e deve ser considerada em conjunto com instrumentos mais específicos para combater a desigualdade de renda no rural do Nordeste, como projetos de reforma agrária e de apoio ao pequeno produtor. Mais concretamente, isto sinaliza, para os formuladores de políticas públicas que visam reduzir a concentração na região, que os instrumentos utilizados não podem ser os mesmos para toda a região. As características locais devem ser observadas.

No estudo das elasticidades, a pluriatividade contribui com mais intensidade para a redução da pobreza, caso ocorra elevação na renda média da família. Por outro lado, o resultado que deve ser mais observado é que as medidas de pobreza rural são mais sensíveis a mudanças na desigualdade do que nas rendas médias, conforme indicado pelos valores mais elevados das elasticidade Gini da pobreza.

Com isto, considera-se esclarecido que a pluriatividade e as rendas não-agrícolas são importantes para reduzir a pobreza e a concentração no meio rural da região Nordeste, confirmando as hipóteses testadas na pesquisa. A redução da desigualdade da renda deve ser trabalhada por um conjunto de instrumentos

específicos, que também tratem da redução da concentração de terras. Novamente ressalta-se que o estímulo às atividades não-agrícolas deve fazer parte de um conjunto maior de medidas, que visem ao desenvolvimento rural da região e não como uma fonte isolada de solução dos problemas.

Com relação ao peso dado à seca como fonte explicativa da pobreza na região e diferenciadora das demais regiões brasileiras, o que se percebe é que os resultados para o ano de chuva se assemelham aos do ano de seca. A decisão da família sobre qual a melhor estratégia a ser adotada não está relacionada com a presença ou ausência de chuvas, e o impacto sobre a pobreza quando se considera a renda não-agrícola é positivo com ou sem chuvas. A perpetuação da condição de pobreza da população rural nordestina e o aumento da concentração da riqueza são menos uma "questão da natureza" e mais uma questão política. Esta perpetuação está, na verdade, relacionada à estrutura de poder da região, ao que se denomina "indústria da seca", uma elite que explora a problemática da ausência de chuvas para tirar proveito de verbas e obras do Estado brasileiro.

Como toda pesquisa científica, este trabalho também apresenta limitações que podem ser fonte de inspiração para trabalhos futuros. Uma destas limitações está relacionada ao uso de uma linha de pobreza única, baseada na renda da família. Pode-se trabalhar com várias linhas de pobreza, baseadas na quantidade de calorias a serem ingeridas de acordo com a cesta de consumo caracteristica da localidade em que a família está inserida. Outra limitação está relacionada com o método escolhido para tentar identificar o viés de seleção amostral. Pode-se tentar utilizar outros métodos diferentes disponíveis na literatura, mais intesivos na questão computacional, que podem proporcionar resultados com maior acurácia.

REFERÊNCIAS BIBLIOGRÁFICAS

ABRAMOVAY, Ricardo. **Paradigmas do capitalismo agrário em questão**. 2.ed. São Paulo: HUCITEC; Campinas: Editora da Unicamp, 1998. 275 p.

ADAMS, Richard H., HE, Jane J. Sources of income, inequality and poverty in rural Pakistan: a decomposition analysis. **International Food Policy Research Institute, Research Report,** n. 102, 1995. 39 p.

ADAMS, Richard H. Non-farm income, inequality and land in rural Egypt. **Policy Research Working Paper,** n. 2178, Word Bank, 1999. 39 p.

ADAMS, R. H. Non-farm income, inequality and poverty in rural Egypt and Jordan. **Policy Research Working Paper**, Word Bank, n. 2572, 2001. 48 p.

ALVAREZ, A. R. **Desenvolvimento teóricos sobre distribuição de renda, com ênfase em seus limites.** 76f. Dissertação (Mestrado em Economia) – Instituto de Pesquisas Econômicas da Faculdade de Economia, Administração de Contabilidade, Universidade de São Paulo-USP, São Paulo, 1996.

ALVES, M. O.. Pluriatividade no sertão nordestino: uma estratégia de sobrevivência. **Revista Raízes**, Campina Grande, v.21, n. 01, p.114-121, jan.-jun. 2002.

ANDRADE, A. O.. **Aplicação do modelo logístico multinomial no estudo da decisão do voto.** 141f. Dissertação (Mestrado em Estudos Populacionais e Pesquisas Sociais) – Escola Nacional de Ciências Estatísticas, Rio de Janeiro, 2006.

ANDRADE, V. D. de A. **O papel do estabelecimento agrícola e das características pessoais e familiares na alocação de trabalho no meio rural brasileiro.** 136f. Tese (Doutorado em Economia Aplicada) - Departamento de Economia Rural, Universidade Federal de Viçosa, Minas Gerais, 2003.

ARAÚJO, C.. "Non-agricultural employment growth and rural poverty reduction in Mexico during the 90s" **Cuadernos de Economía**, Vol. 41 (Diciembre), p.383-399, 2004.

ARRAES, R. A.; DINIZ, M. B.; DINIZ, M. T. Curva Ambiental de Kuznets e Desenvolvimento Econômico Sustentável. **Revista de Economia Rural**. Rio de Janeiro, v.44, n.3, p. 525-547, jul/set, 2006.

BALSADI, O. V., GRAZIANO DA SILVA, J., JULIO, J. E., JUNIOR, O. A.. Evolução das ocupações agrícolas e não-agrícolas no estado de Pernambuco: 1981-97. In: CAMPANHOLA, C. *et al.* **O novo rural brasileiro: uma análise estadual - Nordeste**. Jaguariúna, SP: Embrapa, 2000. p. 139-155.

BOURGUIGNON, F., FOURNIER, M., GURGAND, M. **Selection bias correction based on the multinomial logit model.** 2001, 12p. Disponível em: < http://www.crest.fr/pageperso/lmi/gurgand/selmlog.htm> Acesso em: ago. 2007.

CARNEIRO, M. J.. Significados da pluriatividade para a família rural. In: SEMINÁRIO NACIONAL DE DESENVOLVIMENTO RURAL SUSTENTÁVEL. Ministério do Desenvolvimento Agrário (MDA) – Conselho Nacional de Desenvolvimento Rural Sustentável (CONDRAF). Brasília, agosto de 2005. Texto para Discussão. 7p.

CHERNICK, M. R. **Bootstrap methods: a practitioner's guide.** New York: John Wiley & Sons, 1999. 264p.

COSTA, G.; ALVES, P.; BITTENCOURT, M; ARAÚJO, K.; DOYLE, H. Avaliação de uma *proxy* para a idade da firma utilizando amostragem complexa. Rio de Janeiro: IPEA, dez/2006. (**Texto para discussão IPEA**, n. 1240). 26 p.

DATT, G. **Computacionl tools for poverty measurement and analysis.** Washington: International Food Policy Research Institute, 1998. 21p. (FCND Discussion Paper, 50)

DE JANVRY, Alain e SADOULET. Income Strategies Among Rural Households in Mexico: The role of off-farm activities. **World Development.** v. 29, p. 467-480, 2001.

DE JANVRY, A.; SADOULET, E.; ZHU, N.. **The Role of Non-Farm Incomes in Reducing Rural Poverty and Inequality in China**. Department of Agricultural & Resource Economics, UCB. CUDARE Working Paper 1001, Mar. 2005. 29 p. Disponível em: <http://repositories.cdlib.org/are_ucb/1001> Acesso em: fev. 2007.

DELGADO, G. e CARDOSO JR., J. C. – Principais Resultados da Pesquisa Domiciliar sobre a Previdência Rural na Região Sul do Brasil (projeto avaliação socioeconômica da previdência social rural) . Brasília: IPEA, jun/2000. (**Texto para discussão IPEA**, n. 734). 63 p.

DEL GROSSI, M. e GRAZIANO DA SILVA, J. O uso das Pnads para as áreas rurais. Rio de Janeiro: IPEA, abr/2002. (**Texto para discussão IPEA**, n. 874). 33 p.

DOW, W. Flexible discrete choice demand models consistent with utility maximizations: An application to health care demand. **American Journal of Agricultural Economics.** v. 81, n. 3, p. 680-85, 1999.

ESCOBAL, Javier. The determinants of non-farm diversification in rural Peru. **Word Development.** V. 29, p. 481-496, 2001.

FARIA, A. L. C.. **Aplicação do teste de elegibilidade multidimensional na definição do público alvo beneficiário de políticas públicas.** 2006. 187f. Dissertação (Mestrado em Estudos Populacionais e Pesquisas Sociais) – Escola Nacional de Ciências Estatísticas, Rio de Janeiro, 2006.

FERREIRA, F.H.G., e LANJOUW, P.. Rural non-agricultural activities and poverty in the Brazilian Northeast. **World Development**. v. 29, p. 509–528, 2001.

FOSTER, J., J. GREER, THORBECKE, E.. A Class of Decomposable Poverty Measures. **Econometrica**, v. 52, n.3, p. 761-766, May 1984.

GARCIA JR, A.. **O Sul: caminho do roçado: estratégias de reprodução camponesa e transformação social**. São Paulo: Marco Zero; Brasília, DF: Editora da UnB, 1989. 285 p.

GOIS, A. Bolsa Família aplaca a fome, mas não acaba com a miséria. **Folha de São Paulo**. São Paulo, 28 de junho de 2008. Caderno Brasil. Disponível em <http://www.uol.com.br/fsp>. Acesso em: 28 jun. 2008.

GOMES DA SILVA, A. Novas configurações no espaço rural do Rio Grande do Norte. In: CAMPANHOLA, C. *et al.* **O novo rural brasileiro: uma análise estadual - Nordeste**. Jaguariúna, SP: Embrapa, 2000. p. 13-44.

GOMES DA SILVA, A., VERAS, E. C. A Heterogeneidade da dinâmica das ocupações no rural do Nordeste. In: CONGRESSO BRASILEIRO DE ECONOMIA E SOCIOLOGIA RURAL*,* 39, 2001, Recife. **Anais** (CD-ROM) Brasília, SOBER, 2001. 10 p.

GRAZIANO DA SILVA, J.. **O novo rural brasileiro**. 2ª ed. rev. Campinas, SP: Editora da Unicamp, 1999. (Coleção Pesquisas, 1). 151 p.

GRAZIANO DA SILVA, J. e DEL GROSSI, M. E.. A evolução das rendas e atividades rurais não-agrícolas no Brasil. In: SEMINÁRIO SOBRE O NOVO RURAL BRASILEIRO ("A DINÂMICA DAS ATIVIDADES AGRÍCOLAS E NÃO-

AGRÍCOLAS NO NOVO RURAL BRASILEIRO: FASE III DO PROJETO RURBANO"). 2, 2001, Campinas. Disponível em: < http://www.eco.unicamp.br/nea/rurbano/rurbanw.html>. Acesso em: 02 abr. 2002.

GRAZIANO DA SILVA, J. ; DEL GROSSI, M. E. ; CAMPANHOLA, C. . Novo Rural Brasileiro: uma atualização. In: CONGRESSO BRASILEIRO DE ECONOMIA E SOCIOLOGIA RURAL, 43, 2005, Ribeirão Preto. **Anais** (CD-ROM) Brasília, SOBER, 2005. 13p.

GREENE, W. **Econometric Analysis**. 5. ed. New York: Prentice Hall, 2003. 1026p.

GTDN. **Uma política de desenvolvimento para o Nordeste**. 3ª edição. Recife: Sudene, 1978. 92p.

GUIMARÃES, P. W. Variação de renda familiar, desigualdade e pobreza no Brasil no período 2001 a 2005. **Revista Economia e Desenvolvimento.** Santa Maria/RS, Edição n. 18, p. 73-99, 2006.

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.

HECKMAN, J.. Sample Selection Bias as a Specification Error. **Econometrica**, v. 47, n. 1, p. 153-161, Jan. 1979.

HOFFMANN, J. P. **Generalized linear models: an applied approach**. Boston: Pearson Education, 2004. 204p.

HOFFMAN, R. **Distribuição de renda: medidas de desigualdade e pobreza.** São Paulo: EDUSP, 1998. 275p.

HOFFMANN, R. Efeito do aumento de renda de uma pessoa sobre a desigualdade. In: SEMINÁRIO SOBRE O NOVO RURAL BRASILEIRO ("A DINÂMICA DAS ATIVIDADES AGRÍCOLAS E NÃO-AGRÍCOLAS NO NOVO RURAL BRASILEIRO: FASE III DO PROJETO RURBANO"). 2, 2001, Campinas. Disponível em: < http://www.eco.unicamp.br/nea/rurbano/rurbanw.html>. Acesso em: 02 abr. 2002.

HOFFMANN, R.. Elasticidade da pobreza em relação à renda média e à desigualdade no Brasil e nas Unidades da Federação. **Revista Economia**, Brasília (DF), v.6, n. 02, p.255-289, Jul.-Dez. 2005.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Microdados da Pesquisa Nacional por Amostra de Domicílios Pnad 2003, 2004, 2005.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Síntese de indicadores sociais – 2005. **Estudos e pesquisas: informação demográfica e sócio-econômica.** Rio de Janeiro, n.17, 2006. 330 p.

IPEADATA. **Base de dados.** Disponível em: < http://www.ipeadata.gov.br/>. Acesso em: 23 de set. 2008.

JOHNSTON. J. e DiNARDO. J.. **Econometric methods**. 4th Ed.. New York: McGraw-Hill, 1997. 531p.

KAGEYAMA, A. Pluriatividade e Ruralidade: aspectos metodológicos. **Economia Aplicada**. São Paulo: FIPE/FEA-USP, v. 2, n. 3, p.515-551, Jul.-Set. 1998.

KAKWANI, N. **Poverty and economic growth with applications to Côte D'Ivoire.** Washington D.C.: World Bank (LSMS), 1990. 68p. (Working Paper, 63)

KALTON, G. **Introduction to survey sampling**. Beverly Hills: California: Sage, 1983. 96p. (Series: Quantitative Applications in the Social Sciences, 35).

KRAAY, A.. Whein is growth pro-poor? Cross-country evidence. **Policy Research Working Paper**, Word Bank, n. 3225, 2004. 40 p.

KUZNETS, S. Economic growth and income inequality. **American Economic Review**, v.45, n.1, p. 1-28, mar., 1995.

LAMARCHE, H. (Coord.). **Agricultura familiar**: comparação internacional. Campinas, SP: Editora da Unicamp, 1993. (Volume 1). 336 p.

LEE. E. S.; FORTHOFER, R. N.; LORIMER, R. J. **Analyzing complex survey data**. Beverly Hills: California: Sage, 1989. 78p. (Series: Quantitative Applications in the Social Sciences, 71).

LEE, Lung-Fei. Generalized econometric models with selectivity. **Econometrica**, v. 51, n. 2, p. 507-512, Mar. 1983.

LEE, M. **Off-farm labor supply and various related aspects of resource allocation by agricultural households.** 125f. (Elektronische Dissertatonen) - Georg-August-Universität Göttingen, 1998. Disponível em: <http://webdoc.sub.gwdg.de/diss/1998/lee/>. Acesso em: <08/2007>.

LEONE, E. T. Renda familiar e trabalho feminino nas áreas rurais do Brasil nos anos 90. In: CAMPANHOLA, C. *et al.* **O novo rural brasileiro: novas atividades rurais**. Brasília, DF: Embrapa Informação Tecnológica, 2004. p. 219-242.

LIMA, J. R. F. de. A pluriatividade das famílias rurais: análise comparativa entre Brasil, região Nordeste, Paraíba e Rio Grande do Norte. In: CONGRESSO BRASILEIRO DE ECONOMIA E SOCIOLOGIA RURAL*,* 39, 2001, Recife. **Anais** (CD-ROM) Pernambuco, SOBER, 2001. 8 p.

LIMA, J. R. F. de. **A evolução das rendas e atividades não-agrícolas na Paraíba dos anos 90.** 86f. Dissertação (Mestrado em Economia Rural) – Departamento de Economia e Finanças, Universidade Federal da Paraíba, Campina Grande, 2002 . Disponível em <http://www.eco.unicamp.br/nea/rurbano/zipados/lima.pdf>.

LIMA, J. R. F. de. A evolução das atividades e rendas não-agrícolas na Paraíba nos anos 90. In: CONGRESSO BRASILEIRO DE ECONOMIA E SOCIOLOGIA RURAL*,* 44, 2006, Fortaleza. **Anais** (CD-ROM) Brasília, SOBER, 2006. 18 p.

LINS, M. P. E.; CALÔBA, G. M.. **Programação linear: com aplicações em teoria dos jogos e avaliação de desempenho (data envelopment anaysis).** Rio de Janeiro: Editora Interciência, 2006. 299 p.

LONG, J. S.; FREESE, J. **Regression models for categorical dependent variables using Stata**. 2nd ed.. Texas: Stata Press, 2006. 527 p.

LOPEZ, J. H. **Pro-poor growth: a review of what we know (and what we don't)**, 2004. 20p. Disponível em: <http://www.nadel.ethz.ch/lehre/ppg_review.pdf>. Acesso em: 20 de set. de 2008.

MADDALA, G. S. **Limited-dependent and qualitative variables in econometrics**. New York: Cambridge University Press, 1986. 401p.

MAIA GOMES, G. **Velhas secas em Novos sertões**. Brasília, DF: IPEA, 2001. 293 p.

MARIANO, J. L. e NEDER, H. D.. Renda e Pobreza entre Famílias no meio Rural do Nordeste. In: CONGRESSO BRASILEIRO DE ECONOMIA E SOCIOLOGIA RURAL, 42, 2004, Cuiabá. **Anais** (CD-ROM) Brasília, SOBER, 2004. 19 p.

MATTEI, Lauro F.. **Pluriatividade e Desenvolvimento Rural no Estado de Santa Catarina**. 211f. Tese (Doutorado em Economia Aplicada) - Instituto de Economia, Universidade Estadual de Campinas, Campinas, 1999.

MOLION, L. C. B. Secas – o eterno retorno. In: FURTADO, C. *et al*. **Nordeste: o tempo perdido.** Recife: Editora ASA Pernambuco, 1986. p. 37-59. (Coleção Nordeste em evidência, 18)

NASCIMENTO, Carlos Alves do. Pluriatividade, Pobreza Rural e serviço doméstico remunerado. **Revista de Economia e Sociologia Rural**. Brasília-DF, v.42, n.2, p. 31-64, 2004.

NASCIMENTO, C. A. do. **Pluriatividade, pobreza rural e políticas públicas.** 226f. Tese (Doutorado em Economia Aplicada) - Instituto de Economia, Universidade Estadual de Campinas, Campinas, 2005.

NASCIMENTO, C. A., CARDOZO, S. A. Redes urbanas regionais e a pluriatividade das famílias rurais no Nordeste e no Sul do Brasil, 1992-1999 e 2001-2005. **Revista Econômica do Nordeste**, Fortaleza, v.38, n. 34, out-dez., p. 637-658, 2007.

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.

NAVARRO, Z. Desenvolvimento rural no Brasil: os limites do passado e os caminhos do futuro. **Revista Estudos Avançados**, SP: USP, v. 15, n. 43, p.83-100, 2001.

NEDER, H. D.. Os efeitos das atividades não agrícolas na distribuição de renda do meio rural do Brasil. **Revista de Economia e Sociologia Rural**, v. 41, n. 1, p. 265-290, 2003a.

NEDER, H. D.. Desenvolvimento de metodologias estatísticas aplicadas aos dados das Pnads. In: SEMINÁRIO SOBRE O NOVO RURAL BRASILEIRO ("A DINÂMICA DAS ATIVIDADES AGRÍCOLAS E NÃO-AGRÍCOLAS NO NOVO RURAL BRASILEIRO: FASE III DO PROJETO RURBANO"). 3, 2003b, Campinas. Disponível em: < http://www.eco.unicamp.br/nea/rurbano/rurbanw.html>. Acesso em: 12 abr. 2004.

NEDER, H. D. **Amostragem em Pesquisas Socioeconômicas**. São Paulo: Editora Alínea, 2008. 112p.

NEY, M. G., HOFFMANN, R**.** Atividades não–agrícolas e desigualdade de renda no meio rural brasileiro. In: CONGRESSO BRASILEIRO DE ECONOMIA E SOCIOLOGIA RURAL, 45, 2007, Londrina. **Anais** (CD-ROM) Brasília, SOBER, 2007. 21 p.

OFICINA INTERNACIONAL DEL TRABAJO (OIT). Por que la agricultura sigue siendo importante. **Empleo en el Mundo 2004-2005: empleo, productividad y reducción de la pobreza.** Genebra: OIT, p. 135-193, 2005.

PESSOA, D. G. C.; NASCIMENTO SILVA, P. L. **Análise de dados amostrais complexos**. São Paulo: Associação Brasileira de Estatística, 1998. 170p.

RAVALLION, M. Pro-poor Growth: a Primer. **Policy Research Working Paper**, Word Bank, n. 3242, 2004. 28 p.

ROCHA, S. **Pobreza no Brasil: afinal, do que se trata?** 3.ed. Rio de Janeiro: FGV, 2006. 244p.

RODRIGUEZ, Janete Lins (coord.). **Atlas escolar da Paraíba**. 3ª ed. João Pessoa: Editora Grafset, 2002. 112p.

ROMERO, A. R. **Meio Ambiente e dinâmica de inovações na agricultura**. São Paulo: Annablume/FAPESP, 1998. 278p.

SCHNEIDER, S; FIALHO, M. A. V. Pobreza rural, desequilíbrios regionais e desenvolvimento agrário no Rio Grande do Sul. **Teoria e Evidência Econômica**. Passo Fundo, v.8, n.15, p.117-149, 2000.

SCHNEIDER, S.. Desenvolvimento regional e articulações extra-regionais. In: I FÓRUM INTERNACIONAL TERRITÓRIO, DESENVOLVIMENTO RURAL E DEMOCRACIA. Fortaleza-CE, 2003. Anais (CD-ROM 20 p.).

SCHNEIDER, S.. O papel da pluriatividade numa estratégia de desenvolvimento rural. In: SEMINÁRIO NACIONAL DE DESENVOLVIMENTO RURAL SUSTENTÁVEL. Ministério do Desenvolvimento Agrário (MDA) – Conselho Nacional de Desenvolvimento Rural Sustentável (CONDRAF). Brasília, agosto de 2005. Texto para Discussão. 13 p.

SCHNEIDER, S. . Agricultura familiar e emprego no meio rural brasileiro: análise comparativa das Regiões Sul e Nordeste. **Parcerias estratégicas**, Brasília - DF, v. 1, p. 217-244, 2006.

SILVA JÚNIOR, Luiz Honorato da. Pobreza na população rural nordestina: uma análise de suas características durante os anos 90. **Revista do BNDES**, v.13, n.26, p. 275-290, 2006.

STEWART, F. **Distribuição de renda e desenvolvimento.** 2000**.** Disponível em: <http://www.nead.org.br/download.php?form=.pdf&id=49>. Acesso em: 20 set. 2008.

TEIXEIRA, V. L. Ocupação e renda de famílias rurais da região serrana do estado Rio de Janeiro. In: CAMPANHOLA, C. *et al.* **O novo rural brasileiro: rendas das famílias rurais**. Brasília, DF: Embrapa Informação Tecnológica, 2004. p. 159-204.

VALDÉS, A. & MISTIAEN, Johan. A. **Rural Poverty Alleviation in Brazil: Towards an Integrated Strategy.** Washington DC: World Bank, report no. 21790-BR, v.1, 2001. 62 p..

VEIGA, José Eli da *et al.* **O Brasil rural precisa de uma estratégia de desenvolvimento**. Brasília: Convênio FIPE-IICA (MDA/CNDRS/NEAD), 2001. Disponível em: <www.nead.gov.br>

VERAS, E. C. Perfil das famílias rurais de Arapiraca, AL: ocupação e renda. In: CAMPANHOLA, C. *et al.* **O novo rural brasileiro: rendas das famílias rurais**. Brasília, DF: Embrapa Informação Tecnológica, 2004. p. 39-70.

VILELA, Sérgio. O meio rural piauiense na virada do século: novos processos, novos atores e novas oportunidades. In: CAMPANHOLA, C. *et al.* **O novo rural brasileiro: uma análise estadual - Nordeste**. Jaguariúna, SP: Embrapa, 2000. p. 45-69.

WOOLDRIDGE, J. M. **Econometric Analysis of cross-section and panel data**. Cambridge: The MIT Press, 2002. 735p.

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

WORD BANK. **Pro-poor growth in the 1990s: lessons ans insights from 14 countries**. Washington: DC., 2005. 104p.

ZENG, T.. **Chinese Agricultural Household Farming Efficiency and Off-Farm Labor Supply**. 169f. (Ph.D Thesis). North Carolina State University, 2005. Disponível em: < http://www.lib.ncsu.edu/theses/available/etd-07262005-161114/unrestricted/etd.pdf>, Acesso em: ago. 2007>.

ZHU, N. & LUO, X. Nonfarm activity and rural income inequality: a case study of two provinces in China. **Policy Research Working Paper**, Word Bank, n. 3811, 2006. 26 p.

ANEXOS

ANEXO A

Conceitos básicos sobre amostragem e plano amostral complexo

Em pesquisas sócio-econômicas, a amostragem é muito importante, principalmente, mesmo com população finita, para poder reduzir os custos, sem perder a possibilidade de extrapolar os resultados para a *população* (N), ou seja, para o conjunto integral de indivíduos acerca dos quais a inferência será realizada. Um *elemento* da população é um indivíduo para o qual uma medida é tomada. Então, a população pode ser definida como o conjunto de todos os elementos. Uma *unidade amostral* (n) é a unidade que efetivamente constará na amostra. Unidade amostral e elemento são equivalentes em pesquisas com amostra aleatória simples, mas são diferentes em amostras de múltiplos estágios, onde existem diferentes níveis de unidades amostrais e o elemento amostral será a unidade de amostragem de nível mais baixo (Neder, 2008).

Uma forma de classificar os métodos de amostragem é com base no desenho (plano) amostral. Por plano amostral entende-se a forma como o processo de seleção das amostras (*s*) é executado, sendo que cada uma possui probabilidade de seleção *p(s)* conhecida ou calculável. Apenas uma amostra *s* é escolhida dentro do conjunto de todas as amostras *S* ($s \in S$) possíveis (GUIMARÃES, 2007).

Segundo o delineamento amostral, os métodos de amostragem podem ser classificados em amostra aleatória simples (com ou sem reposição), amostra sistemática, amostragem estratificada, amostragem por conglomerados e amostragem com probabilidade desigual. O que diferencia um plano amostral baseado em amostra aleatória simples de uma amostra complexa são a estratificação, a conglomeração e as probabilidades desiguais de seleção, podendo ainda existir ajustes dos pesos amostrais para calibração com totais populacionais conhecidos (PESSOA e NASCIMENTO SILVA, 1998). Neste tópico, estes tipos de amostragem são ligeiramente comentados,

sendo Neder (2008) recomendado para uma abordagem mais aprofundada, assim como os livros-referência de Kalton (1983) e Lee *et al* (1989).

Com relação à amostragem aleatória simples com reposição (AASCR), seleciona-se uma amostra de *n* unidades amostrais, de maneira que cada unidade tenha a mesma probabilidade de ser selecionada (NEDER, 2008). Pode ser considerada como a forma mais simplificada de amostragem. É de especial interesse porque o processo de reposição simplifica a inferência estatística, pois elimina qualquer covariância entre a seleção dos elementos (LEE *et al*, 1989). Cada seleção é independente das anteriores e os elementos da população possuem probabilidade de seleção (fração amostral) igual a $1-\left(1-\frac{1}{n}\right)^n$, podendo ser incluídos mais de uma vez na amostra (GUIMARÃES, 2007). Os pacotes econométricos básicos normalmente utilizados em modelagem consideram que os dados são obtidos através de amostra aleatória simples com reposição, ou seja, que são IID (NASCIMENTO SILVA *et al*, 2002). Qualquer processo de amostragem que não seja AASCR possui *design* complexo e requer considerações especiais na análise, para não se obter resultados viesados ou calculados erroneamente (LEE *et al,* 1989).

Amostra sistemática é importante por prover um meio de reduzir substancialmente o esforço necessário para a seleção da amostra. É bastante fácil de aplicar, sendo necessário apenas pegar cada *k-ésimo* elemento depois de um início aleatório (KALTON, 1983). Como um exemplo simples, supondo um bairro com 300 residências na qual 35 fazem parte da amostra, tem-se uma fração de 1 em 9 (35/300). Uma amostra sistemática do tamanho em questão é obtida tomando um número aleatório entre 1 e 9, para determinar a primeira residência da amostra e depois selecionar cada nona residência posterior.

Com amostragem sistemática, cada elemento da população possui a mesma probabilidade de ser selecionado para a amostra, o que assegura que a média amostral será um estimador não viesado da média populacional quando o número de elementos da população (N) é igual a *k* vezes o número de elementos na amostra (n) (LEE *et al*, 1989). Porém, difere da amostra aleatória simples com reposição, já que as probabilidades de conjuntos de elementos serem incluídos na amostra não são todas

iguais (KALTON, 1983). No exemplo acima, a probabilidade dos elementos 1 e 3 serem ambos selecionados é zero e a probabilidade de seleção dos elementos 1 e 10 é 1/9, complicando o cálculo da variância.

É geralmente utilizada no estágio final da amostragem em múltiplos estágios, quando o pesquisador de campo é instruído a selecionar uma proporção predeterminada de unidades em uma lista de moradias num bairro de uma cidade (LEE *et al*, 1989).

Normalmente, ao se fazer alguma pesquisa sócio-econômica, algum tipo de informação suplementar sobre os elementos da população a ser estudada é conhecida, como a parte urbana e a rural, tamanho da população, PIB, índice de desenvolvimento humano, entre outras. Isto pode ser utilizado para melhorar o plano amostral através da técnica da estratificação. Essencialmente, estratificação é a classificação da população em sub-populações, denominados estratos, baseado em informação suplementar. Então, dentro de cada estrato pode ser selecionada uma ou várias amostras. A vantagem derivada do fato do tamanho da amostra no estrato ser controlado pelo amostrista. Quando os tamanhos das amostras nos estratos são proporcionais ao tamanho da população no estrato, a estratificação é denominada de proporcional[58]. Caso contrário, é não-proporcional e denominado de método de repartição de Neyman. Normalmente se toma amostras aleatórias simples dentro de cada estrato, apesar de outros métodos serem possíveis de utilização (KALTON, 1983).

A estratificação é bastante utilizada, principalmente, por reduzir a variância amostral quando o estrato é homogêneo. Como demonstrado em Kalton (1983), considerando um grande N_h (tamanho da população no estrato h), pode-se decompor a variância e obter

$$S^2 = S_w^2 + \sum W_h (\bar{Y}_h - \bar{Y})^2 \qquad (74)$$

onde S^2 é a variância considerando amostra aleatória simples; S_w^2 é a variância média ponderada dentro do estrato, considerando que a estratificação é proporcional; W_h é a

[58] Com estratificação proporcional garante-se que a amostra seja auto-ponderada, ou seja, as proporções amostrais são estimativas não viesadas das proporções populacionais (NEDER, 2008).

proporção da população no estrato h; $\bar{Y}_h$ é a média da população no estrato h e $\bar{Y}$ é a média populacional global. Dado que $\sum W_h\ (\bar{Y}_h - \bar{Y})^2$ é uma quantidade não-negativa (soma de quadrado de dois termos), segue que $S^2 \geq S_w^2$. Isto significa que uma amostra estratificada proporcional não pode ser menos precisa do que uma amostra aleatória simples de mesmo tamanho. Quanto mais heterogêneos forem os estratos, maior será o ganho de precisão (KALTON, 1983).

Para a escolha dos estratos, duas condições são necessárias: a) proporção da população em cada estrato (W_h); b) deve ser possível retirar amostras separadas de cada estrato. Pelo menos duas amostras por estrato para ser possível estimar o erro-padrão da média populacional. A escolha de como formar os estratos depende dos objetivos da estratificação. Por exemplo, se for para aumentar os ganhos de precisão, devem-se formar estratos os mais homogêneos internamente possíveis (KALTON, 1983). Em outros casos, a estratificação serve para obter estimativas não apenas para o conjunto da população, mas também para os estratos previamente definidos (NEDER, 2008).

Uma outra forma de se selecionar a amostra é tratar as sub-populações como grupos (*clusters*). É sempre uma abordagem mais prática já que seleciona grupos e não elementos individuais, porém faz com que se perca a característica dos elementos amostrais serem selecionados independentemente (não é mais IID). Com isto, poupa recursos financeiros da pesquisa. Diferentemente da estratificação, apenas uma amostra é selecionada dentro do conglomerado. Caso todos os elementos dentro do conglomerado sejam incluídos na amostra, o método é conhecido como amostragem por conglomerados (*cluster*). Caso apenas alguns elementos sejam selecionados, tem-se a amostragem em dois estágios. Finalmente, tem-se a amostragem em múltiplos estágios quando a população é dividida em conglomerados hierarquizados por tamanho, depois são escolhidos dentro dos maiores grupos, de forma aleatória, alguns conglomerados menores e, dentre estes, selecionam-se de forma aleatória as unidades amostrais, pela justificativa de se tentar elevar os ganhos de precisão com redução de custos. Então, na abordagem em múltiplos estágios as unidades amostrais são grupos

de elementos exceto no último estágio de amostragem (KALTON, 1983; LEE *et al*, 1989; NEDER, 2008).

Apesar de estratos e conglomerados tratarem de grupos de elementos, servem para diferentes propósitos. Dado que os estratos estão todos representados na amostra, é vantajosa sua utilização se existe homogeneidade interna. Por outro lado, como apenas um conglomerado é selecionado, este deve ser representativo dos demais. Enquanto a estratificação proporcional aumenta o ganho de precisão, no geral, a conglomeração é menos precisa do que a amostragem aleatória simples[59]. Porém, dada a grande economia nos custos, é necessária uma análise custo-benefício para avaliar se é ou não apropriada. O primeiro estágio desta abordagem é denominado PSU[60] (Unidade Primária Amostral); o segundo, SSU (Unidade Secundária Amostral) e, assim, sucessivamente. Quando um PSU único é selecionado de um estrato, a variância dentro deste não pode ser estimada diretamente. Para se estimar a variância, pares de estratos similares são comumente combinados e tratados como se cada par constituísse um único estrato maior, denominado estrato colapsado (*collapsed strata*) (KALTON, 1983). Outra opção é incluir o estrato com PSU único em outro estrato, com maior número de observações[61].

Normalmente os conglomerados não possuem o mesmo tamanho. Setores censitários, bairros, municípios, seja o que for os *clusters*, quase sempre possuem tamanhos diferentes. É necessário algum controle nesta questão da variabilidade no tamanho da amostra. Uma maneira de reduzir a variabilidade é estratificar os conglomerados de acordo com o tamanho. Porém, esta não é a alternativa mais adequada, pois limita a vantagem que outros usos de estratificação possuem. A alternativa preferida é utilizar probabilidade de seleção proporcional ao tamanho[62]. Como demonstrado em Kalton (1983), no *design* com 3 estágios, como é o caso da Pnad, a probabilidade do elemento β no *cluster* α ser selecionado, com *a* PSUs

[59] Em exemplo meramente ilustrativo, Kalton (1983) demonstra que o tamanho da amostra com conglomerados precisa ser cerca de 3 vezes maior do que a aleatória simples para se obter o mesmo grau de precisão.
[60] Do inglês *primary sampling units*.
[61] Que pode ser feito facilmente no *software* Stata com a rotina IDONEPSU.
[62] Na amostragem com probabilidade proporcional ao tamanho a representatividade é mais "fina" em toda a escala da variável tamanho (NEDER, 2008).

selecionados, *b* unidades secundárias amostrais (SSU) selecionados dentro de cada PSU escolhido e *c* elementos amostrados dentro de cada SSU é dada por

$$P(\alpha\beta) = f = \frac{n}{N} = \left(\frac{aB_\alpha}{\sum B_\alpha}\right)\left(\frac{bB_{\alpha\beta}}{B_\alpha}\right)\left(\frac{c}{B_{\alpha\beta}}\right) \quad (75)$$

em que f é denominado de fração amostral; n = abc; N = $\sum B_\alpha$; $B_{\alpha\beta}$ é o tamanho do SSU β no PSU α.

Inicialmente deve-se tomar o tamanho de cada PSU, em seguida "acumular" os tamanhos e, posteriormente, selecionar de forma aleatória *x* números com 1 ≤ x ≤ maior B_α acumulado. A vantagem do método é que tende a garantir que a amostra seja representativa em relação ao tamanho de cada PSU.

Os pesos amostrais (w_i) são usados para refletir as diferentes probabilidades de seleção dos elementos amostrais. Sua omissão resulta em estimativas pontuais de parâmetros e erros-padrões, viesadas. Pode envolver a correção por diferencial de taxas de respostas (não-resposta) dentro das classes. Além disso, diferentes pesos são necessários para diferentes unidades de análise. No caso de pessoas e domicílios, por exemplo, os pesos a ser desenvolvidos são diferentes. Dois tipos de pesos são normalmente encontrados em pesquisas com amostra complexa: a) o peso (w_i) que possibilita a expansão dos dados para a população, que é o inverso da fração amostral ($1/f_i$); b) o peso relativo (rw_i), obtido pelo ajustamento do peso de expansão para refletir o tamanho da amostra, sendo definido pela razão entre w_i e a média de w_i, ou seja, $w_i/\overline{w_i}$ (LEE *et al*, 1989).

No caso de um plano amostral estratificado, a população N é dividida em L estratos, baseado em informação suplementar de alguma variável com N_1, N_2, ..., N_L elementos. São selecionados n_h (h= 1,2,..., L) elementos independentes de cada *h-ésimo* estrato. Quando a fração amostral é a mesma em cada estrato, o plano amostral é auto-ponderado. No caso da estratificação ser desproporcional, ou seja, $f_1 \neq f_2 \neq f_h$, os pesos de expansão e os relativos são diferentes nos estratos (LEE *et al*, 1989). Os pesos são tomados de forma proporcional ao inverso da probabilidade de seleção em

cada estrato (w_{hi}=N_h/n_h) para todos os elementos selecionados no estrato *h* (KALTON, 1983).

Um conceito utilizado em amostragem, relevante neste estudo, é o de modelo de superpopulação, que pode ser considerado um modelo intermediário entre a modelagem Clássica e a Probabilística. Na modelagem Clássica, considera-se y_1, ..., y_n, como realizações de variáveis aleatórias (IID) Y_i, ..., Y_n, com função de densidade *f* (y; θ) e $\theta \in \Theta$ sendo o parâmetro indexador de *f* e Θ, o espaço paramétrico. A partir das observações y_1, ..., y_n, são feitas inferências a respeito do parâmetro θ. A abordagem Probabilística é essencialmente não-paramétrica, já que não supõe uma distribuição paramétrica para as observações da amostra. A partir de uma população finita U= {1, ..., N} é selecionada uma amostra s={i_1, ..., i_n}, segundo um plano amostral caracterizado por *p(s)*, ou seja, pela probabilidade calculável de ser selecionada a amostra *s*. A partir dos valores da amostra y_1, ..., y_N, são feitas inferências sobre as funções de valores populacionais $g(y_1, ..., y_N)$ (PESSOA e NASCIMENTO SILVA, 1998).

No modelo de superpopulação, considera-se y_1, ..., y_N, como realizações conjuntas de vetores aleatórios (IID) $\mathbf{Y}_i$, ..., $\mathbf{Y}_N$, com função de densidade *f* (y; θ) e $\theta \in \Theta$. A partir de um plano amostral definido por *p(s)*, obtém-se os valores amostrais y_1, ..., y_n, que não são considerado IID e se busca fazer inferências sobre o parâmetro θ, considerando a estratificação, a conglomeração e as diferentes probabilidades de seleção dos elementos amostrais. Como pode ser observado, são utilizados tanto os métodos da abordagem Clássica quanto da Probabilística (PESSOA e NASCIMENTO SILVA, 1998).

Como demonstrado em Pessoa e Nascimento Silva (1998), sendo $\mathbf{y}_i=(y_{i1}, \dots, y_{iR})'$ o vetor Rx1 de valores das variáveis da pesquisa e $\mathbf{x}_i=(x_{i1}, \dots, x_{iQ})'$ o vetor Qx1 de variáveis auxiliares da i-ésima unidade da população, ou seja, das variáveis que contém a informação necessária para o desenho amostral (estrato, psu, pesos). Então, pode-se especificar a distribuição conjunta combinada das variáveis de pesquisa e das auxiliares. Considerando $f[(y_1, x_1), \dots, (y_N, X_n; \eta)$ a função de densidade de probabilidade de (Y_1, X_1), ..., (Y_N, X_N), onde η é um vetor de parâmetros,

pode-se definir um tipo de modelo de superpopulação, que não é degenerado, o qual supõe que os valores numa população finita são fixos mas desconhecidos descrito por,

$$P[(Y_i, X_i) = (y_i, x_i)] = 1 \, \forall_i \in U \tag{76}$$

Esta abordagem é denominada de aleatorização ou baseada no planejamento amostral, já que a única fonte de aleatoriedade provém do delineamento amostral. Com este modelo se pode fazer inferência analítica, ou seja, se pode explicar a relação entre as variáveis não apenas para a população finita em questão, mas também para outras que poderiam ter sido gerados pelo modelo de superpopulação (PESSOA e NASCIMENTO SILVA, 1998).

O modelo de superpopulação é uma fonte de aleatoriedade, mas existem outras como o processo de medição, o planejamento amostral e o mecanismo de resposta. Além do modelo de superpopulação, o planejamento amostral enquanto fonte de variação é outro ponto importante que deve ser enfatizado.

Os planos amostrais podem ser *informativos, não-informativos*. Um desenho amostral é informativo quando o mecanismo de seleção das unidades amostrais depende dos valores das variáveis de pesquisa, como nas pesquisas de caso-controle. Se a seleção não depende diretamente dos valores das variáveis da pesquisa, o delineamento amostral é denominado de *não-informativo*. Já os planos não-informativos podem também ser classificados como *ignoráveis* ou *não-ignoráveis*. Se o plano amostral é uma amostra aleatória simples com reposição, o plano é ignorável, pois o modelo adotado para a amostra é o mesmo da população. Um desenho amostral é não-ignorável quando é complexo. Neste caso, não considerar o plano amostral pode levar a estimativas pontuais viesadas e erros-padrões calculados erroneamente (GUIMARÃES, 2007).

Estimação de variância e parâmetros incorporando o plano amostral

A análise da variância é importante para avaliar a precisão das estimativas. Quando se está trabalhando com amostra complexa, para seu cálculo e, conseqüentemente, o de intervalos de confiança, faz-se uso de duas metodologias: a) Linearização; b) Replicação. Dentre os métodos de linearização, o método de **linearização de Taylor (ou método Delta**) é um dos mais usados. Na estatística, a aplicação do método de expansão de uma série de Taylor é para obter uma aproximação de um valor de uma função, a qual é difícil de calcular por ser não-linear e se basear na variância desta aproximação para obter a variância da função não-linear.

Conforme demonstrado em Pessoa e Nascimento Silva (1998), considere, por exemplo, a expansão de g($\hat{Y}$) em torno de Y, até o termo de primeira ordem, desprezando o resto, com g sendo não-linear e Y um vetor de totais de R variáveis de pesquisa, além de $\theta = g(Y)$,

$$\hat{\theta} \simeq \hat{\theta}_L = \frac{g(Y)}{0!} + \frac{\Delta g(Y)(\hat{Y}-Y)}{1!} \quad (77)$$

com $\Delta g(Y)$ sendo uma matriz Jacobiana (matriz com as derivadas parciais) KxR, com K=R cuja r-ésima coluna é $\partial g(Y)/\partial Y_r$, para r=1, ..., R.

Tomando a V($\hat{\theta}$), dado que g(Y) não possui variância de aleatorização (variância referente à distribuição de probabilidade p(s) induzida pelo plano amostral) é possível demonstrar que um estimador consistente de V($\hat{\theta}$) pode ser obtido por

$$\hat{V}(\hat{\theta}) = a\Sigma a \quad (78)$$

com a= $\Delta g(\hat{Y})$ e $\Sigma = \hat{V}(\hat{Y}) = \sum_{i\in s}\sum_{j\in s}\frac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}}\frac{y_i}{\pi_i}\frac{y_j}{\pi_j}$ é o estimador de variância de aleatorização não-viesado de $\hat{Y}$ (PESSOA e NASCIMENTO SILVA, 1998).

Com relação ao método de Replicação, dentre os métodos existentes[63], **Bootstrap** é bastante utilizado, apesar de ser computação intensivo. Esse método busca gerar uma distribuição de freqüências após a replicação de inúmeras amostras

[63] No *software* Stata, além de Bootstrap, é disponível *brr* (*balanced repeated replication*) e *jacknife*.

selecionadas com reposição da amostra original e com o mesmo tamanho "*n*". Trata a amostra como se fosse uma população e aplica um procedimento tipo Monte Carlo para gerar uma estimativa empírica da distribuição amostral da estatística envolvida. Segundo Chernick (1999) *bootstrap* é freqüentemente usado quando é muito difícil calcular a estimativa do erro padrão de um estimador. Sua vantagem frente ao método de Monte Carlo é que não é necessário conhecer o processo gerador dos dados para ser utilizado.

Como já discutido, uma questão de grande relevância é que os dados não podem ter estratos com **PSU (unidade primária amostral) único**. Com PSU único não se tem como calcular variância. É necessário ter variabilidade para se poder calcular variância. Em algumas situações, além de PSU único deve-se observar a existência também de PSU com observação única. Em algumas situações é possível calcular variância com PSU com observação única e outras vezes não é possível. Para solucionar o problema de PSU único, pode-se agregar as observações de estratos com PSU único em estrato na mesma UF com maior número de observações, por exemplo. Já com relação a questão do PSU com observação única, não se pode fazer esta agregação e, por enquanto, a alternativa é excluir esta observação única.

Com relação aos parâmetros, o ajuste de modelos considerando o desenho amostral é feito pelo método da Máxima Pseudo-Verossimilhança (MPV), que é uma adaptação do método da Máxima-Verossimilhança, utilizando os pesos amostrais (w_i). Como demonstrando em Pessoa e Nascimento Silva (1998), Nascimento Silva *et al*, (2002), Costa *et al* (2006), Andrade (2006) e Guimarães (2007) o método da Máxima Pseudo-Verossimilhança (MPV) pode ser resumidamente descrito da seguinte forma: seja $y_j = (y_{j1}, \cdots, y_{jR})'$ o vetor Rx1 das variáveis de pesquisa observadas para a unidade elementar *j*, gerado por um vetor aleatório Y_j, para $j \in U$, em que U={1,..., N}. Este vetor é considerado IID com função de densidade $f(y;\theta)$ e θ é um vetor Kx1 de parâmetros desconhecidos de interesse.

O vetor de parâmetros θ do modelo definido por $f(y;\theta)$ para a população finita pode ser estimado usando o estimador de MPV $\hat{\theta}_{MPV}$, o qual é um valor de θ que serve de solução das equações de

$$\hat{T}(\theta) - \sum_{j \in S} w_j u_j(\theta) - 0 \tag{79}$$

em que w_j são pesos amostrais adequados para a estimação de totais populacionais partindo da amostra *s*; $u_j(\theta) = \partial \log [f(y_j;\theta)]/\partial\theta$ é o vetor Kx1 dos escores da unidade elementar *j*, para $j \in U$; $\hat{T}(\theta)$ é a soma dos escores ponderados.

A partir da equação (78) é possível estimar a variância assintótica de $\hat{\theta}$, usando a matriz de primeira ordem da expansão da série de Taylor para o estimador de MPV. Com isto, o estimador da variância assintótica é dado por:

$$\hat{V}(\hat{\theta}_{MPV}) = [\hat{J}(\hat{\theta}_{MPV})]^{-1} \hat{V}[\sum_{j \in S} w_j u_j(\hat{\theta}_{MPV})][\hat{J}(\hat{\theta}_{MPV})]^{-1} \tag{80}$$

onde $[\hat{J}(\hat{\theta}_{MPV})]^{-1} = \frac{\partial \hat{T}(\theta)}{\partial\theta}\Big|_{\theta=\hat{\theta}_{MPV}} = \sum_{j \in s} w_j \frac{\partial u_j(\theta)}{\partial\theta}\Big|_{\theta=\hat{\theta}_{MPV}}$

e

$\hat{V}[\sum_{j \in S} w_j u_j(\hat{\theta}_{MPV})]$ é um estimador consistente para a matriz de variância do estimador do total populacional dos escores.

Dado que existe mais de uma forma de calcular os pesos w_j, os estimadores de MPV não serão únicos[64]. Dependendo do modelo que se pretende ajustar, é necessário apenas calcular os escores $u_j(\theta)$, os estimadores de total e variância, para se poderem calcular as estimativas pontuais $\hat{\theta}_{MPV}$ dos parâmetros θ e as estimativas da matriz de variâncias $\hat{V}(\hat{\theta}_{MPV})$. Estas podem ser utilizadas para calcular intervalos de confiança ou estatísticas de teste, sendo a distribuição assintótica de $\hat{\theta}$ normal multivariada.

[64] Porém, normalmente se utiliza o inverso da probabilidade de inclusão na amostra (GUIMARÃES, 2007).

ANEXO B

Algoritmos Stata utilizados para as estimativas do ano de 2003

1) Do File para Preparação dos Dados

```
*************************************************************
* PREPARACAO DE DADOS PNAD 2003
* 02/10/08
* Versao 3.0 da Tese
*************************************************************

cd "c:\tempecon\dados_temp\"

cap log close
log using algoritmo_2003, replace
set more off
set memory 256m

 * LEITURA DAS INFORMACOES DO DESENHO DA AMOSTRA NO ARQUIVO DE DOMICILIOS

 #delimit ;
 infix ano 1-4 uf 5-6 controle 5-12 serie 13-15
 tipo 16-17 espdom 22-22 aguapoco 66-66 ilumidom 73-73 probmun 101-112
 probsetor 116-127 intervalo 128-136 strat 169-175 psu 176-182 sitcen 88-88
 areacen 90-90 using dom2003.txt, clear;
 #delimit cr

 #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 & sitcen >= 4;
 #delimit cr

 save dom_ne_2003, replace

* 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 idade 27-29 conddom 30-30
 condfam 31-31 numfam 32-32 cor 33-33
 trabalhou 132-132 afastado 133-133 autocon 134-134
 constru 135-135 numtrab 136-136 codocu 137-140 codati 141-145
 ocupagr 146-147 ocupnag 298-298 rend_prin 323-334 rend_merc_prin 337-348
 horastrab 355-356 naoocup 369-369 codocusec 414-414
 maisumtrab 414-414 secmil 417-417 secfunpub 418-418 seccartass 419-419
 rend_sec 421-432 rend_merc_sec 435-446 horastrabsec 451-452 rendnt 517-517
 rend_apo 519-530 rend_pen 533-544 rend_oap 547-558 rend_ope 561-572
```

```
 rend_abo 575-586 rend_alu 589-600 rend_doa 603-614 rend_jur 617-628
 anosest 666-667 condativ 668-668 condocu 669-669 posocup 670-671
 horas 672-672 ativprin 673-673 ramos 674-675 grupocu 676-677
 rend_tr1 688-699 rend_tr2 700-711 rend_tod 712-723 rend_dom 724-735
 rend_fa1 736-747 tipofam 748-749 numfam1 750-751
 numfam2 752-753 rend_fa2 754-765 areacen 766-766
 sitcen 767-767 peso 768-772 pesofam 773-777
 using pes2003.txt, clear;
 #delimit cr

 #delimit;
 keep if uf > = 21 & uf < = 29;
 keep if sitcen >= 4;
 #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;
 keep if _merge == 3;
 drop _merge;
 #delimit cr

 save pes_ne_2003, replace

* COLOCACAO DOS NOVOS PESOS NO ARQUIVO DE PESSOAS

* LEITURA DO ARQUIVO DE NOVOS PESOS

 #delimit ;
 infix v0101 1-4 v0102 5-12 v0103 13-15
 v0301 16-17 pesopes2
 18-22 pesofam2 23-27
 using PesoPes2003.txt, clear;
 rename v0102 controle;
 rename v0103 serie;
 rename v0301 ordem;
 sort controle serie ordem, stable;
 #delimit cr

 save novospesos_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 novospesos_ne_2003;
 keep if _merge == 3;
 drop _merge;
```

```
 #delimit cr

 table conddom [pweight = pesopes2], contents(freq )

 save pes_ne_2003, replace

 * GERACAO DA VARIAVEL TIPO DE FAMILIA, POSICAO NA OCUPACAO E LOCAL DE
MORADIA

 use pes_ne_2003, clear

 #delimit;
 gen va=. ;
 replace va=1 if ocupagr != . & conddom < 6 & idade >= 10;

 gen vna=.;
 replace vna=1 if ocupnag != . & conddom < 6 & idade >= 10;

 gen vmaisumtrab = .;
 replace vmaisumtrab = 1 if maisumtrab != . & conddom < 6 & idade >= 10;

 gen vdes=.;
 replace vdes=1 if naoocup == 3 | ativprin == 3 & conddom < 6 & idade >=
10;
 #delimit cr

 save pes_ne_2003, replace

 * GERACAO DA VARIAVEL POSICAO NA OCUPACAO

 use pes_ne_2003, clear

 #delimit;
 gen vempregador=.;
 replace vempregador=1 if posocup == 10;

 gen vcontapro=.;
 replace vcontapro=1 if posocup == 9;

 gen vempregado=.;
 replace vempregado=1 if posocup < 9;

 gen vnaoocup=.;
 replace vnaoocup=1 if posocup >= 13;

 gen vautocon=.;
 replace vautocon=1 if posocup == 11 | posocup == 12;
 #delimit cr

 save pes_ne_2003.dta, replace

 * GERACAO DA VARIAVEL LOCAL DE MORADIA

 use pes_ne_2003, clear

 #delimit;
 gen vextensao=.;
 replace vextensao=1 if sitcen == 4;
```

```
gen vpovoado=.;
replace vpovoado=1 if sitcen == 5;

gen vnucleo=.;
replace vnucleo=1 if sitcen == 6;

gen voutros=.;
replace voutros=1 if sitcen == 7;

gen vruragrop=.;
replace vruragrop=1 if sitcen == 8;
#delimit cr

* GERACAO DA VARIAVEL RAZAO DE DEPENDENCIA

#delimit;
gene dependente = 1 if idade < 10 | idade > 60;
replace dependente = 0 if idade >= 10 & idade <= 60;
gene idadeativa = 1 if idade >= 10 & idade <= 60;
replace idadeativa = 0 if idade < 10 | idade > 60;
#delimit cr

* GERACAO DA VARIAVEL IDADE MEDIA DA PEA RESTRITA E IDADE MEDIA AO
QUADRADO

#delimit;
gene idpeares = idade if idade >= 10 & idade <= 60;
gene idpeares2 = idpeares^2;
#delimit cr

* GERACAO DA VARIAVEL NIVEL EDUCACIONAL MEDIO

#delimit;
replace anosest = . if anosest == 17;
replace anosest = . if idade < 15;
replace anosest = anosest - 1;
#delimit cr

* GERACAO DO NUMERO DE HORAS TRABALHADAS EM ATIVIDADES AGRICOLAS

gen numhoraagr = horas if ativprin == 1

* GERACAO DO NUMERO DE HORAS TRABALHADAS EM ATIVIDADES NAO-AGRICOLAS

gen numhoranagr = horas if ativprin == 2

* GERACAO DA VARIAVEL NUMERO DE COMPONENTES DA FAMILIA

gen numcompfam1 = numfam1

* GERACAO DA VARIAVEL RENDA FAMILIAR PER CAPITA

gen renfam = rend_fa1

* GERACAO DA VARIAVEL DUMMY SE POSSUI OU NAO RNT

#delimit;
gen vrendnt=0;
replace vrendnt=1 if rend_apo != . | rend_pen != .
| rend_oap != . | rend_ope != . | rend_abo != .
```

```
 | rend_alu != . | rend_doa != . | rend_jur != .;
 #delimit cr

 save pes_ne_2003, replace

 * Criação da variável família

 use pes_ne_2003, clear

 #delimit;
 collapse (sum) dependente idadeativa renfam vempregador
 vcontapro vempregado vnaoocup vautocon vextensao vpovoado
 vnucleo voutros vruragrop vrendnt (max) va vna vmaisumtrab
 vdes (mean) areacen idpeares idpeares2 anosest numcompfam1
 numhoraagr numhoranagr aguapoco ilumidom strat psu pesopes2,
 by(uf controle serie);
 #delimit cr

 save pes_ne_2003, replace

 * GERACAO DA VARIAVEL TIPO DE FAMILIA (AGRICOLA, NAO-AGRICOLA, PLURIATIVA,
NAO-ATIVO/DESOCUPADO)

 use pes_ne_2003, clear

 #delimit;
 gen famtipo=3;
 replace famtipo=0 if va == 1 & vna == . ;
 replace famtipo=1 if va == . & vna == 1;
 replace famtipo=2 if va == 1 & vna == 1 | vmaisumtrab != .;
 replace famtipo=3 if va == . & vna == . & vdes == 1;
 label values famtipo ;
 label define Labeltipo 0 "agricola";
 label define Labeltipo 1 "nao_agric", add;
 label define Labeltipo 2 "pluriativa", add;
 label define Labeltipo 3 "nao_at_des", add;
 label values famtipo Labeltipo;
 #delimit cr

 table famtipo [pweight = pesopes2], contents(freq )

* GERACAO DA VARIAVEL DUMMY SE POSSUI OU NAO RNT

 gen rendant=0
 replace rendant=1 if vrendnt != 0

* GERACAO DA VARIAVEL DUMMY POSICAO NA OCUPACAO (CONTA-PROPRIA, EMPREGADA)

 #delimit;
 gen contaprop=0;
 replace contaprop=1 if vcontapro == 1;
 cap label drop Labelposocup;
 label values contaprop ;
 label define Labelposocup 0 "caso contrario", add;
 label define Labelposocup 1 "conta-propria", add;
 label values contaprop Labelposocup;
 #delimit cr

 #delimit;
 gen empregados=0;
```

```
 replace empregados=1 if vempregado == 1;
 cap label drop Labelposocupa;
 label values empregados ;
 label define Labelposocupa 0 "caso contrario", add;
 label define Labelposocupa 1 "empregados", add;
 label values empregados Labelposocupa;
 #delimit cr

* GERACAO DA VARIAVEL LOCAL DE MORADIA (RURAL AGROPECUARIO)

 #delimit;
 gen localmora=0;
 replace localmora=1 if vruragrop !=0;
 cap label drop Labellocalmora;
 label values localmora ;
 label define Labellocalmora 0 "caso contrario", add;
 label define Labellocalmora 1 "rural agropecuario", add;
 label values localmora Labellocalmora;
 #delimit cr

* GERACAO DA VARIAVEL RAZAO DE DEPENDENCIA

 gen razaodep = dependente / idadeativa

* GERACAO DA VARIAVEL RENDA PER CAPITA

 gen renpercap = renfam / numcompfam1

 drop if renpercap > 5000000

* GERACAO DAS VARIAVEIS RENDA AGRICOLA, NAO-AGRICOLA, PLURIATIVA E NAO-
OCUPADOS

* Nao logaritmizadas

 #delimit;
 gen renagr = renpercap if famtipo == 0;
 replace renagr=0.00001 if renagr == 0;

 gen rennag = renpercap if famtipo == 1;
 replace rennag=0.00001 if rennag == 0;

 gen renplur = renpercap if famtipo == 2;
 replace renplur=0.00001 if renplur == 0;

 gen rennocu = renpercap if famtipo == 3;
 replace rennocu=0.00001 if rennocu == 0;
 #delimit cr

* Logaritmizadas

 #delimit;
 gen lnrenagr = ln(renagr);
 replace lnrenagr=0.00001 if lnrenagr <= 0;

 gen lnrennag = ln(rennag);
 replace lnrennag=0.00001 if lnrennag <= 0;

 gen lnrenplur = ln(renplur);
 replace lnrenplur=0.00001 if lnrenplur <= 0;
```

```
 gen lnrennocu = ln(rennocu);
 replace lnrennocu=0.00001 if lnrennocu <= 0;
 #delimit cr

* Sobre a variável aguapoco

 replace aguapoco=9 if aguapoco == .

 save pes_ne_2003, replace

* DECLARANDO/SETANDO O CONJUNTO DE DADOS COMO SENDO DE AMOSTRA COMPLEXA

 use pes_ne_2003, clear

 #delimit;
 local vars_drop "idpeares anosest numcompfam1 rendant contaprop
 empregados localmora razaodep aguapoco ilumidom";
 #delimit cr

 foreach i of local vars_drop {
            drop if `i' == .
 }

 svyset psu [pweight=pesopes2], strata(strat) vce(linearized)
singleunit(missing) || _n

 svydes, single

 save pes_ne_2003, replace

* 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=pesopes2], strata(strat) vce(linearized)
singleunit(missing) || _n
 svydes, single

 save pes_ne_2003, replace

* ROTINA DE ALOCACAO DE ESTRATOS COM UM UNICO PSU EM ESTRATOS COM MAIOR
NUMERO
* DE OBSERVACOES UTILIZANDO O DO.FILE idonepsu RENNAG - 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 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, replace
 append using transf
 save acum, replace
 }
 end

 capture program drop prog2
 program define prog2
 use pes_ne_2003, clear
 keep if uf == estado
 idonepsu rennag , 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 24 25 27 {
      scalar estado = `i'
      prog1
 }
```

```
 foreach i in 21 22 23 26 28 29 {
      scalar estado = `i'
      prog2
 }


 drop strat psu
 rename novo_str strat
 rename novo_psu psu

 svyset psu [pweight=pesopes2], strata(strat) vce(linearized)
singleunit(missing) || _n
 svydes rennag, single

 save pes_ne_2003, replace

* ROTINA DE ALOCACAO DE ESTRATOS COM UM UNICO PSU EM ESTRATOS COM MAIOR
NUMERO
* DE OBSERVACOES UTILIZANDO O DO.FILE idonepsu RENAGR - 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 prog1
 program define prog1
 use pes_ne_2003, clear
 keep if uf == estado
 gene novo_str = strat
 gene novo_psu = psu
 quietly {
 save transf, replace
 use acum, replace
 append using transf
 save acum, replace
 }
 end

 capture program drop prog2
 program define prog2
 use pes_ne_2003, clear
 keep if uf == estado
 idonepsu renagr , 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 27 28 29 {
```

```
      scalar estado = `i'
      prog1
 }

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

 drop strat psu
 rename novo_str strat
 rename novo_psu psu

 svyset psu [pweight=pesopes2], strata(strat) vce(linearized)
singleunit(missing) || _n
 svydes renagr, single

 save pes_ne_2003, replace

* ROTINA DE ALOCACAO DE ESTRATOS COM UM UNICO PSU EM ESTRATOS COM MAIOR
NUMERO
* DE OBSERVACOES UTILIZANDO O DO.FILE idonepsu RENPLUR - 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 prog1
 program define prog1
 use pes_ne_2003, clear
 keep if uf == estado
 gene novo_str = strat
 gene novo_psu = psu
 quietly {
 save transf, replace
 use acum, replace
 append using transf
 save acum, replace
 }
 end

 capture program drop prog2
 program define prog2
 use pes_ne_2003, clear
 keep if uf == estado
 idonepsu renplur , 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 22 23 25 26 27 28 29 {
```

```
      scalar estado = `i'
      prog1
 }

 foreach i in 21 24 {
      scalar estado = `i'
      prog2
 }

 drop strat psu
 rename novo_str strat
 rename novo_psu psu

 svyset psu [pweight=pesopes2], strata(strat) vce(linearized)
singleunit(missing) || _n
 svydes renplur, single

 save pes_ne_2003, replace

* ROTINA DE ALOCACAO DE ESTRATOS COM UM UNICO PSU EM ESTRATOS COM MAIOR
NUMERO
* DE OBSERVACOES UTILIZANDO O DO.FILE idonepsu RENNOCU - 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 rennocu , strata(strat) psu(psu) generate(novo_)
 quietly {
 save transf, replace
 use acum.dta, 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=pesopes2], strata(strat) vce(linearized)
singleunit(missing) || _n

 save pes_ne_2003, replace
```

```
* Eliminando as observações únicas

 use pes_ne_2003, clear

 #delimit;
 drop if strat == 240015 & psu == 62 | strat == 240015 & psu == 63
 | strat == 240015 & psu == 66 ;
 drop if strat == 290030 & psu == 394;
 #delimit cr

 drop if renplur < 1 | rennag < 1

 save pes_ne_2003, replace
 log close
 exit
```

2) Do File para o Logit Multinomial

```
*************************************************************************
**************
* MODELO LOGIT MULTINOMIAL para 2003 - entender os determinantes da escolha
do indivíduo
* em participar apenas de atividades agrícolas, não-agrícolas, ou ser
pluriativo.
*************************************************************************
**************

* 21/10/2008

cd "c:\tempecon\dados_temp\"

cap log close
log using mlog2003, replace
set more off
set mem 256m

use pes_ne_2003, clear

* Declarando os dados como amostra complexa

svyset psu [pweight=pesopes2], strata(strat) vce(linearized)
singleunit(missing)

svydes, single

* Estatísticas descritivas dos dados

set more off

* Média

#delimit;
local vars_med "renagr rennag renplur rennocu
idpeares anosest numcompfam1 razaodep renpercap";
#delimit cr

foreach i of local vars_med {
         svy: mean `i'
}
```

```
* Proporção

#delimit;
local vars_prop "rendant contaprop empregados localmora ilumidom aguapoco";
#delimit cr

foreach x of local vars_prop {
            svy: proportion `x'
}

* Modelo Logit Multinomial com a Familia do Tipo Agrícola como base

#delimit;
global mvars "famtipo idpeares idpeares2 anosest numcompfam1 localmora
razaodep rendant contaprop empregados renpercap ilumidom aguapoco";
#delimit cr

xi, prefix(dum): svy: mlogit $mvars i.uf, base(0)

* Teste de Colinearidade

#delimit;
local vars "idpeares anosest numcompfam1 localmora razaodep rendant
                  contaprop empregados renpercap ilumidom aguapoco";
#delimit cr
foreach y of local vars {
       local x : list vars - y
       svy: regress `y' `x'
         display "tolerance_`y' = " 1-e(r2) " VIF_`y' = " 1/(1-e(r2))
}
* Teste de Wald

qui xi, prefix(dum): svy: mlogit $mvars i.uf, base(0)
#delimit;
local mvars_b "idpeares idpeares2 anosest numcompfam1 localmora
razaodep rendant contaprop empregados renpercap ilumidom aguapoco";
#delimit cr

foreach v of local mvars_b {
           test `v', nosvyadjust
}

* Modelo Logit Multinomial com a Familia do Tipo Agricola como base - Razão
Relativa de Risco
xi, prefix(dum): svy: mlogit $mvars i.uf, base(0) rrr

* Efeitos Marginais

qui xi, prefix(dum): svy: mlogit $mvars i.uf, base(0)
margeff, at(mean) replace

* Estatistica MEFF
qui xi, prefix(dum): svy: mlogit $mvars i.uf, base(0)
estat effects, meff

* Margeff idpeares

#delimit;
qui xi, prefix(dum): svy: mlogit $mvars i.uf, base(0);
```

```
qui margeff , at(mean) replace;
qui su idpeares;
local m1 = r(mean);
local betas_logit "[agricola]idpeares [pluriativa]idpeares
[nao_agric]idpeares [nao_at_des]idpeares";
#delimit cr

foreach z of local betas_logit {
           nlcom `z'+ 2*`m1'*`z'2
}
```

3) Do File para o SELMLOG sem Bootstrap

```
*************************************************************
* MODELO SELMLOG 2003
* 08/10/08
* Autor: JOAO RICARDO LIMA
*************************************************************

cap log close selmlog2003
set more off
set matsize 7000
set seed 123456789
log using selmlog2003, replace

cd "c:\tempecon\dados_temp\"
use pes_ne_2003, clear
version 8: svyset [pweight = pesopes2], psu(psu) strata(strat)
global pline=120
gen cons = 1

*****************************************************************************
*****************************************
* Método escolhido para ser usado na Tese - LEE
*****************************************************************************
*****************************************

cap program drop selnoboot03
program define selnoboot03
            local vars "idpeares anosest numcompfam1 localmora razaodep"
            local vardepsel "renpercap idpeares2 ilumidom aguapoco rendant
contaprop empregados"
            cap drop singletonr
            di "modelo para as famílias `2'"
            qui xi, prefix(dum): svyselmlog `1' `vars' i.uf, sel (famtipo
`vars' `vardepsel' i.uf) force meth(lee) mloptions(base(0)) gen(lee`2')
showmlogit
            mkmat `vars' dumuf* lee`2'* cons, matrix(var_ind1)
            matrix B1=e(b)
            matrix y_hat1 = var_ind1*B1'
            svmat y_hat1, names(`1'_est)
            gen `1'_res = `1' - `1'_est1
            sum `1'_res
            gen va_res_`2' = r(mean)+r(sd)*invnorm(runiform())
            mkmat va_res_`2' , matrix(vares`2')
            matrix y_hat2 = var_ind1*B1'+vares`2'
            svmat y_hat2, names(`1'_)
            gen `3'=exp(`1'_1) if famtipo == `4'
            gen `5'=exp(`1'_1) if famtipo == `6'
```

```
end

selnoboot03 lnrenagr agr renagr_nag 1 renagr_plur 2
selnoboot03 lnrenplur plur renplur_agr 0 renplur_nag 1
selnoboot03 lnrennag nag rennag_agr 0 rennag_plur 2

**********************
* Cálculo dos índices
**********************

cap program drop indicesestprog
program indicesestprog
        ineqerr `1' `2' if famtipo == `3', reps(200) psu(psu)
psuwt(pesopes2)
            svylorenz `1' if famtipo == `3'
            svylorenz `2' if famtipo == `3'
            povdeco `1' if famtipo == `3', pline($pline) summarize
            povdeco `2' if famtipo == `3', pline($pline) summarize
            sepov `1' `2' if famtipo == `3' [w=pesopes2], p($pline)
strata(strat) psu(psu) ci meff
end

indicesestprog renagr_nag rennag 1
indicesestprog renagr_plur renplur 2
indicesestprog renplur_nag rennag 1
indicesestprog renplur_agr renagr 0
indicesestprog rennag_agr renagr 0
indicesestprog rennag_plur renplur 2

****************************
* Cálculo das significâncias
****************************

cap program drop signifprog
program signifprog
            di "Teste de igualdade para a proporção de pobres"
            gen `1' = `2' < $pline  if famtipo == `3'
            cap gen `4' = `5' < $pline if famtipo == `3'
            svy: mean `1' `4' if famtipo == `3'
            lincom `1' - `4'
            di "Teste de igualdade para o hiato da pobreza"
            gen `6' = `1'*($pline - `2' )/$pline if famtipo == `3'
            cap gen `7' = `4'*($pline - `5' )/$pline if famtipo == `3'
            svy: mean `6' `7' if famtipo == `3'
            lincom `6' - `7'
            di "Teste de igualdade para a severidade da pobreza"
            gen `8' = `1'*[($pline - `2' )/$pline]^2 if famtipo == `3'
            cap gen `9' = `4'*[($pline - `5' )/$pline]^2 if famtipo == `3'
            svy: mean `8' `9' if famtipo == `3'
            lincom `8' - `9'
end

signifprog prop_por_agrnag renagr_nag 1 prop_por_nag rennag ngap_agrnag
ngap1_nag sever_agrnag sever1_nag
signifprog prop_por_agrplur renagr_plur 2 prop_por_plur renplur
ngap_agrplur ngap1_plur sever_agrplur sever1_plur
signifprog prop_por_plurnag renplur_nag 1 prop_por_nag rennag ngap_plurnag
ngap1_nag sever_plurnag sever1_nag
signifprog prop_por_pluragr renplur_agr 0 prop_por_agr renagr ngap_pluragr
ngap1_agr sever_pluragr sever1_agr
```

```
signifprog prop_por_nagagr rennag_agr 0 prop_por_agr renagr ngap_nagagr
ngap1_agr sever_nagagr sever1_agr
signifprog prop_por_nagplur rennag_plur 2 pro_por_plur renplur ngap_nagplur
ngap1_plur sever_nagplur sever1_plur

**********************
* Situação Geral
**********************

* a) Todos fossem pluriativos
* b) Todos fossem agrícolas

* Situação A - Todos pluriativos

cap program drop todos3prog
program todos3prog
        gen `1' = .
        replace `1' = `2'_1 if famtipo == `3'
        replace `1' = `2'_1 if famtipo == `4'
        replace `1' = `2' if famtipo == `5'
        drop if `1' == .
            gen renda_`6'_3 = exp(`1')
end

todos3prog lnrenda_plur_3 lnrenplur 0 1 2 plur
todos3prog lnrenda_agr_3 lnrenagr 1 2 0 agr

* Concentracao e Pobreza

cap program drop indicesprog
program indicesprog
            ineqerr `1' , reps(200) psu(psu) psuwt(pesopes2)
            svylorenz `1'
            povdeco `1' , pline($pline) summarize
            sepov `1' [w=pesopes2], p($pline) strata(strat) psu(psu) ci
meff
end

indicesprog renda_plur_3
indicesprog renda_agr_3

* Teste de hipotese de igualdade dos indices

cap program drop signifprog3
program signifprog3
            di "Teste de Igualdade para a Proporção de Pobres"
            gen prop_por_`1'3 = renda_`1'_3 < $pline
            gen prop_por_`2'3 = renda_`2'_3 < $pline
            svy: mean prop_por_`1'3 prop_por_`2'3
            lincom prop_por_`1'3-prop_por_`2'3
            di "Teste de Igualdade para o Hiato da Pobreza"
            gen ngap_`1'3 = prop_por_`1'3*($pline - renda_`1'_3 )/$pline
            gen ngap1_`2'3 = prop_por_`2'3*($pline - renda_`2'_3 )/$pline
            svy: mean ngap_`1'3 ngap1_`2'3
            lincom ngap_`1'3- ngap1_`2'3
            di "Teste de igualdade para a severidade da pobreza"
            gen sever_`1'3 = prop_por_`1'3*[($pline - renda_`1'_3
)/$pline]^2
            gen sever1_`2'3 = prop_por_`2'3*[($pline - renda_`2'_3
)/$pline]^2
```

```
            svy: mean sever_`1'3 sever1_`2'3
            lincom sever_`1'3- sever1_`2'3
end

signifprog3 plur agr

*****************************************************
* Função Kernel para as rendas observadas e estimadas
*****************************************************

* Para o log da renda observada das familias agricolas

kdensity lnrenagr if famtipo == 0, generate(xagr fxagr) nograph
kdensity lnrenplur_1 if famtipo == 0, generate(xpluragr fxpluragr) nograph
kdensity lnrennag_1 if famtipo == 0, generate(xnagagr fxnagagr) nograph

label var xagr "ln_agr"
label var xpluragr "ln_plur_agr"
label var xnagagr "ln_nag_agr"

graph twoway (line fxagr xagr, sort) (line fxpluragr xpluragr, sort) (line
fxnagagr xnagagr, sort),    ///
ytitle("Densidade") xtitle("Kernel estimativa de densidade - Rendas
simuladas - 2003")

* Para a renda observada das familias pluriativos

kdensity lnrenplur if famtipo == 2, generate(xplur fxplur) nograph
kdensity lnrenagr_1 if famtipo == 2, generate(xagrplur fxagrplur) nograph
kdensity lnrennag_1 if famtipo == 2, generate(xnagplur fxnagplur) nograph

label var xplur "ln_plur"
label var xagrplur "ln_agr_plur"
label var xnagplur "ln_nag_plur"

graph twoway (line fxplur xplur, sort) (line fxagrplur xagrplur, sort)
(line fxnagplur xnagplur, sort),    ///
ytitle("Densidade") xtitle("Kernel estimativa de densidade - Rendas
simuladas - 2003")

* Para a renda observada das familias nao-agricolas

kdensity lnrennag if famtipo == 1, generate(xnag fxnag) nograph
kdensity lnrenagr_1 if famtipo == 1, generate(xagrnag fxagrnag) nograph
kdensity lnrenplur_1 if famtipo == 1, generate(xplurnag fxplurnag) nograph

label var xnag "ln_nag"
label var xagrnag "ln_agr_nag"
label var xplurnag "ln_plur_nag"

graph twoway (line fxnag xnag, sort) (line fxagrnag xagrnag, sort) (line
fxplurnag xplurnag, sort),    ///
ytitle("Densidade") xtitle("Kernel estimativa de densidade - Rendas
simuladas - 2003")

* Gráficos para os dois Kernel - todas agrícolas e todos não-agrícolas

kdensity lnrenda_agr_3, generate(xagr3 fxagr3) nograph
kdensity lnrenda_plur_3, generate(xplur3 fxplur3) nograph
```

```
label var xagr3 "lnrenda_agr_3"
label var xplur3 "lnrenda_plur_3"

graph twoway (line fxagr3 xagr3, sort) (line fxplur3 xplur3, sort) ,
ytitle("Densidade")       ///
xtitle("Kernel estimativa de densidade - Rendas simuladas - 2003")

save dadosok_mapirn03, replace

log close
exit
```

4) Do File para o SELMLOG com Bootstrap

```
*************************************************************
* MODELO SELMLOG 2003
* 02/10/08
*************************************************************

cap log close selboot2003
set more off
log using selboot2003, replace

cd "c:\tempecon\dados_temp\"
use pes_ne_2003, clear
version 8: svyset [pweight = pesopes2], psu(psu) strata(strat)

*************************************************************
* Método de LEE
*************************************************************

cap program drop selboot03
program define selboot03
    local vars "idpeares anosest numcompfam1 localmora razaodep"
    local vardepsel "renpercap idpeares2 ilumidom aguapoco rendant
contaprop empregados"
    cap drop singletonr
    xi, prefix(dum): svyselmlog `1' `vars' i.uf, sel (famtipo `vars'
`vardepsel' i.uf) force meth(lee) mloptions(base(0)) boot(1000)
end

foreach vardep of varlist lnrennag lnrenagr lnrenplur {
            selboot03 `vardep'
}

log close
exit
```

5) Do File para a Elasticidade da Pobreza.

```
/* PROGRAMA DE CÁLCULO DA ELASTICIDADE-POBREZA DO CRESCIMENTO  */
/* AUTOR: HENRIQUE DANTAS NEDER */
/* Modificado por JOAO RICARDO LIMA*/
/* 04/10/2008 */
```

```
cap log close elast_2003
clear
set mem 256m
set more off
log using elast_2003, replace
cd "c:\tempecon\dados_temp"
use dadosok_mapirn03.dta, clear

global i = 0
gen lp = 120
gen int pesopes = pesopes2

*****************************************************************
* rotina de calculo de elasticidades da pobreza

capture program drop elast
program define elast
preserve
keep if famtipo == `1'
keep `2' pesopes lp strat psu
summarize `2' [fw=pesopes]
scalar mu = r(mean)
summarize lp
scalar z = r(mean)
glcurve `2' [fw=pesopes], gl(L1) p(p) nograph
generate L = L1/mu

* Especificação da Curva de Lorenz: Quadrática Geral

generate y1 = L*(1-L)
generate x1 = p^2 - L
generate x2 = L*(p-1)
generate x3 = p - L
regress y x1 x2 x3

* verificar se o intercepto é não significativo

regress y x1 x2 x3, noconstant
matrix b1 = e(b)
scalar a = b1[1,1]
scalar b = b1[1,2]
scalar c = b1[1,3]
scalar e = -(a + b + c + 1)
scalar m = b^2 - 4*a
scalar n = 2*b*e - 4*c
scalar r1 = (n^2 - 4*m*(e^2))^(1/2)
scalar s1 = (r1 - n)/(2*m)
scalar s2 = -(r1 + n)/(2*m)
display a
display b
display c
display e
display m
display n
display r1
display s1
display s2
scalar H = -(1 / (2*m))*(n + r1*(b+2*z/mu)*((b+2*z/mu)^2-m)^(-1/2))
scalar LH = -(1/2)*(b*H + e + (m*H^2 + n*H + e^2)^(1/2))
scalar PG = H - (mu/z)*LH
```

```
scalar P2 = 2*PG - H - ((mu/z)^2)*(a*H + b*LH - (r1/16)*ln((1-H/s1)/(1-
H/s2)))
display "H  = " H
display "PG = " PG
display "P2 = " P2


* calculo da segunda derivada da curva de Lorenz

scalar L2H = (r1^2*(m*H^2 + n*H + e^2)^(-3/2))/8

* calculo do indice de Gini

if m < 0 {
scalar gini = e/2 - n*(b + 2)/(4*m) + r1^2 / (8*m*sqrt(-m))*(asin((2*m +
n)/r1) - asin(n/r1))
}
else {
scalar gini = e/2 - n*(b + 2)/(4*m) + r1^2 / (8*m*sqrt(m))*ln(abs((2*m + n
+ 2*sqrt(m)*(a + c - 1))/(n - 2* e * sqrt(m))))
}

display "Gini =  " gini

* calculo das elasticidades da pobreza

* com relação à média mu

scalar col1 = - z/(mu*H*L2H)
scalar col2 = 1 - H/PG
scalar col3 = 2*(1-PG / P2)

* com relação ao indice de Gini

scalar col4 = (1 - z / mu) / (H*L2H)
scalar col5 = 1 + (mu/z - 1)*H/PG
scalar col6 = 2*(1 + (mu/z - 1)*PG /P2)
display "Elasticidades da pobreza com relacao a renda media"
display "H   : " col1
display "PG  : " col2
display "SPG : " col3
display "Elasticidades da pobreza com relacao ao indice de Gini"
display "H   : " col4
display "PG  : " col5
display "SPG : " col6
global i = $i + 1
display $i
restore
end

foreach x of varlist renagr renplur_agr rennag_agr {
            elast 0 `x'
}

foreach x of varlist renplur renagr_plur rennag_plur {
            elast 2 `x'
}

log close
exit
```