

# DETERMINANTES DA ALOCAÇÃO DE TRABALHO AGRÍCOLA NAS REGIÕES BRASILEIRAS

## *DETERMININGS OF FARM WORK ALLOCATION IN BRAZILIAN REGIONS*

**Filipe de Moraes Cangussu Pessoa**

Instituto Federal de Educação, Ciência e Tecnologia de Brasília – DF – Brasil

**Daniel Arruda Coronel**

Universidade Federal de Santa Maria – RS – Brasil

**Airton Lopes Amorim**

Universidade Federal de Viçosa – MG – Brasil

**João Eustáquio de Lima**

Universidade Federal de Viçosa – MG – Brasil

---

**Resumo:** O objetivo deste trabalho foi analisar os determinantes da alocação de trabalho agrícola nas regiões brasileiras, com base nos microdados da Pesquisa Nacional por Amostra de Domicílios (PNAD, 2009). Para isso, o procedimento empírico consistiu na utilização de dois modelos: Análise Fatorial Confirmatória (AFC) e o Modelo *Logit*. O modelo de AFC apresentou bom ajustamento e definiu duas variáveis latentes qualificação e renda. No que tange ao Modelo *Logit* os resultados indicaram que o fato de um indivíduo ser do sexo masculino e residir no meio rural aumenta a probabilidade de ele estar alocado na atividade agrícola, enquanto ser da cor branca reduz essa probabilidade. Ademais, as variáveis renda e qualificação se relacionam de forma negativa com a alocação de trabalho agrícola, sendo a Região Centro-Oeste a que mais contribuiu para um indivíduo estar alocado em um trabalho agrícola.

**Palavras-chave:** PNAD; Análise Fatorial Confirmatória; Modelo *Logit*

**Abstract:** The aim of this paper was to analyze the determinings of farm work allocation in Brazilian regions, based on micro data of National Household Sample Survey 2009. For that, the empirical procedure consisted in the use of two models: Confirmatory Factor Analysis and the Logit Model. The model Confirmatory Factor Analysis showed good fit and defined two latent variables: qualification and income. As to the Logit Model, the results showed that the fact of a man living in rural areas increases the likelihood of being allocated in farm work, however if he is white this probability decrease. Besides, the variables qualification and income have a negative relation with farm work allocation, being the Midwest region that contributed most for people being allocated in farm work.

**Keywords:** National Household Sample Survey, Confirmatory Factor Analysis, Logit Model

---

## INTRODUÇÃO

Ao longo do século XX, o Brasil passou por um intenso processo de urbanização da sua estrutura produtiva, o que vem gerando significativas mudanças no meio rural brasileiro, tais como aumento das ocupações não agrícolas, elevação da produtividade agrícola e aumento da qualificação da mão de

obra. Tais questões vêm sendo estudadas por vários pesquisadores, merecendo destaque os trabalhos de Graziano da Silva (1996), Graziano da Silva e Del Grossi (2001) e Kageyama (2004)<sup>1</sup>.

Não obstante esse processo de urbanização, o Brasil iniciou este século com uma população rural ainda expressiva. Em 2008, 30,8 milhões de pessoas declararam residir em zona rural. Entretanto, a consolidação de uma malha urbana com elevado número de pequenas cidades determina que parte dos residentes na zona rural trabalhe em áreas urbanas, ocorrendo também a situação inversa. Desse modo, a medida mais precisa da ocupação no campo corresponde à população ativa vinculada às atividades agrícolas. Segundo essa perspectiva, tais atividades envolviam 16 milhões de pessoas em 2008, representando 17,4% da população ocupada do país (BUAINAIN; DEDECCA, 2010.)

De acordo com Buainain e Dedecca (2010), apesar de a redução da participação relativa do pessoal ocupado na agricultura ter continuado ao longo da primeira década do século XXI, o estoque de população ocupada de forma direta nas atividades agrícolas manteve-se praticamente inalterado. Ainda de acordo com esses autores, esta constatação permite explicitar pelo menos dois fatores relevantes que condicionam o funcionamento e a evolução do mercado de trabalho agrícola no Brasil. O primeiro é a própria existência de um estoque ainda elevado de população ocupada em atividades agrícolas, havendo sinais de que o país conviverá com a preservação de um segmento de trabalho agrícola expressivo nas próximas décadas. O segundo refere-se à elevada participação das ocupações não remuneradas no total do pessoal ocupado na agricultura.

Nessa perspectiva, o objetivo deste trabalho foi analisar os determinantes da alocação de trabalho agrícola nas regiões brasileiras, com base nos microdados da Pesquisa Nacional por Amostra de Domicílios (PNAD, 2009), realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Para isso, utilizou-se o Modelo de Análise Fatorial Confirmatória (AFC) e o Modelo Econométrico *Logit*.

Dessa forma, este trabalho tenta contribuir com a literatura relacionada à alocação de trabalho agrícola, gerando evidências com base em dados mais recentes sobre a população brasileira. Ademais, tais evidências são fornecidas por meio de um modelo pouco explorado em trabalhos econômicos, a Análise Fatorial Confirmatória, mas com potencial para produzir estimativas mais robustas sobre os determinantes da alocação de trabalho agrícola.

O presente trabalho está estruturado em três seções, além desta introdução. Na segunda seção, são apresentados os procedimentos metodológicos; na terceira, os resultados obtidos são analisados e discutidos e, na última seção, são apresentadas as principais conclusões.

---

<sup>1</sup>Apesar dessas transformações, o meio rural brasileiro ainda apresenta problemas sérios como: menor nível de renda *per capita*, piores condições de moradia, saneamento, higiene e de educação, dentre outros, conforme Campolina *et al.* (2009).

## METODOLOGIA

Este estudo investiga os determinantes das decisões de alocação de trabalho agrícola nas regiões brasileiras, por meio do exame de dados microeconômicos. Para isso, propôs-se uma estratégia empírica composta de duas etapas. A primeira consiste na estimação do modelo de Análise Fatorial Confirmatória (AFC) e a segunda etapa consiste de um modelo *Logit*.

A ideia de se utilizar a AFC consiste em obter escores fatoriais de variáveis latentes, as quais são formadas por variáveis observadas que possuem alto grau de associação entre si, que são importantes para a determinação da alocação de trabalho no setor agrícola, juntamente com outras variáveis, que serão utilizadas no modelo *Logit* para mensurar a contribuição de cada uma e explicar a alocação de trabalho no setor supracitado. O modelo *Logit* se faz necessário, neste contexto, pela natureza da variável que representa a alocação de trabalho, a qual é binária.

### O modelo de Análise Fatorial Confirmatória

Para reduzir a massa de dados de variáveis observadas a um conjunto de fatores que representam um novo conjunto de variáveis latentes, de dimensão inferior ao das variáveis originais, utiliza-se a análise fatorial. Isso permite que variáveis de naturezas distintas possam ser analisadas conjuntamente, por meio de escores que podem ser hierarquizados ou utilizados em outros métodos estatísticos (MINGOTI, 2007).

A análise fatorial pode ser categorizada em duas classes: exploratória e confirmatória<sup>2</sup>. Na Análise Fatorial Exploratória (AFE), não existe um modelo específico relacionando variáveis latentes e observadas; o número de fatores (variáveis latentes) não é definido com antecedência; e as variáveis latentes influenciam todas as variáveis observadas. Assim, a técnica definirá quais variáveis estão relacionadas a determinado fator e não existe qualquer restrição imposta ao relacionamento entre os fatores e as variáveis. Ao contrário, na Análise Fatorial Confirmatória (AFC), o modelo é construído com antecedência, e o número de fatores e seu relacionamento com cada variável observada é determinado pelo conhecimento prévio do pesquisador, embasado em sua experiência sobre o problema de pesquisa, ou em um marco teórico.

O modelo de AFC pode ser representado, em termos matriciais, da seguinte forma:

$$X = \Gamma Y + \Lambda \quad (1)$$

em que:

$X$  é um vetor  $p \times 1$  das variáveis observadas;

---

<sup>2</sup>Para um exame detalhado ver Thompson (2004).

$\Gamma$  é uma matriz  $p \times m$  de cargas fatoriais estimadas para as variáveis observadas;

$Y$  é um vetor  $m \times 1$  representando as variáveis latentes; e

$\Lambda$  é um vetor de erros associados com cada variável observada para levar em conta erros de medida e variância não associados com a variável latente. Assume-se que a média de  $\Lambda$  é igual a zero, que sua variância é homocedástica, e não autocorrelacionada entre observações. Além disso, assume-se que  $\Lambda$  e  $Y$  não são correlacionados. (LEE, 2007)

A hipótese nula na AFC é que  $\Sigma = \Sigma(\theta)$ , ou seja, a matriz de covariância populacional das variáveis observadas  $\Sigma$  é igual à matriz de covariância reproduzida (estimada) como uma função dos parâmetros do modelo,  $\Sigma(\theta)$  (também denominada matriz de covariância implícita). Por sua vez,  $\Sigma(\theta)$  pode ser reescrito em termos da Equação (1) como:

$$\begin{aligned}\Sigma(\theta) &= E(XX') \\ &= E[(\Gamma Y + \Lambda)(Y' \Gamma' + \Lambda')] \\ &= \Gamma E(Y Y') \Gamma' + E(\Lambda \Lambda') \\ &= \Gamma \Phi \Gamma' + \Theta\end{aligned}\tag{2}$$

em que:

$\Phi$  e  $\Theta$  são as matrizes de covariância das variáveis latentes e de erros de medida, respectivamente. As variáveis observadas estão correlacionadas umas com as outras porque elas estão associadas com pelo menos uma variável latente. Uma vez que essa informação é extraída dos dados, visando transformá-los em um conjunto pré-determinado de variáveis latentes, permanece um erro de medida, também chamado de unicidade, o qual não está correlacionado com nenhuma das variáveis latentes.

O procedimento de estimação procura encontrar a matriz de covariância implícita  $\hat{\Sigma}(\theta)$  o mais próximo possível de  $S$ , a matriz de covariância amostral das variáveis observadas. Os métodos mais utilizados para estimação são Máxima Verossimilhança (MV) e Mínimos Quadrados Ponderados (MQP), também denominado *Asymptotically Distribution Free* (ADF).

O procedimento de MV leva a estimativas para os parâmetros  $\theta$  que maximizam a função de verossimilhança  $L$  de que a matriz de covariância amostral  $S$  foi extraída de uma população para a qual a matriz de covariância implícita  $\Sigma(\theta)$  é válida. A função de log-verossimilhança  $L$  a ser maximizada de acordo com Bollen (1989) é:

$$\log L = -\frac{1}{2}(N-1)\{\log|\Sigma(\theta)| + tr(S\Sigma(\theta)^{-1})\} + c\tag{3}$$

em que:

$\log$  é o logaritmo natural;

$L$  é a função de verossimilhança;

$N$  é o tamanho da amostra;

$\theta$  é o vetor de parâmetros;

$S$  é a matriz de covariância amostral;

$\Sigma(\theta)$  é a matriz de covariância implícita e

$|\Sigma(\theta)|$  seu determinante,  $tr$  é o traço da matriz e  $c$  é uma constante que contém termos da distribuição Wishart que não mudam, uma vez que a amostra é dada.

A estimação por MV tem as vantagens de oferecer estimativas dos parâmetros e erros-padrão que são consistentes e eficientes, e outra vantagem é que as estimativas são, em geral, invariantes com relação à escala e livres de escala. Como consequência, o resultado do modelo ajustado não depende se a matriz de covariância ou correlação e também se os dados originais ou transformados foram utilizados (BOLLEN, 1989).

As desvantagens da MV repousam em três pressupostos: i) as amostras devem ser grandes; ii) as variáveis observadas devem ser contínuas; e iii) as variáveis observadas devem ter distribuição normal multivariada (BROWN, 2006).

Se algumas variáveis observadas são de natureza ordinal, ou dicotômica, e outras de natureza contínua, e se as variáveis observadas contínuas se desviam, de maneira considerável, de uma distribuição normal multivariada, o método mais indicado para estimação do modelo de AFC é o MQP (BROWN, 2006).

O método de MQP busca minimizar a seguinte função de ajuste,

$$F_{MQP} = [s - \sigma(\theta)]' W^{-1} [s - \sigma(\theta)] \quad (4)$$

em que  $s$  é um vetor de elementos não redundantes da matriz de covariância empírica,  $\sigma(\theta)$  é um vetor de elementos não redundantes da matriz de covariância implícita do modelo,  $\theta$  é o vetor de parâmetros e  $W$  é uma matriz positiva definida de pesos com dimensão  $p \times p$  onde  $p = k(k+1)/2$  e  $p$  é o número de variáveis observadas.

Uma das principais vantagens do método de MQP é que ele não assume normalidade multivariada dos dados, nem qualquer outra forma de distribuição, sendo, nesse aspecto, muito menos restritivo que o método de MV. A desvantagem do método é que ele requer amostras muito maiores do que aquelas requeridas pelo método de MV para produzir estimativas consistentes e eficientes dos parâmetros (SCHERMELLEN-ENGEL; MOOSBRUGGER; MÜLLER, 2003).

Afora as questões de estimação, o modelo de AFC ainda requer o cumprimento de critérios de identificação. Se um modelo tem  $p$  variáveis observadas, então os dados contêm  $k(k+1)/2$  momentos de segunda ordem e, conseqüentemente,  $p$ , o número de parâmetros baseados nos momentos de segunda ordem que podem ser estimados, não pode exceder  $k(k+1)/2$  (FEDERICI; MAZZITELLI, 2008).

Cumprido os critérios de identificação e estimação do modelo de AFC, passa-se a analisar seu ajustamento. Existem diversas estatísticas e índices para tal finalidade. No presente trabalho, optou-se por utilizar a estatística de teste  $\chi^2$ , a raiz quadrada do erro médio de aproximação (RMSEA<sup>3</sup>), a raiz quadrada do

<sup>3</sup> Do inglês *Root Mean Square Error of Approximation*.

resíduo médio padronizado (SRMR<sup>4</sup>), o índice Tucker-Lewis (TLI), o índice comparativo de ajuste (CFI<sup>5</sup>) e o coeficiente de determinação (CD).

Como critérios de aceitação de um bom ajuste do modelo, a literatura (HAIR JR. *et al.* (1998), BROWN (2006), BOLLEN (1989)) sugere para a estatística de  $\chi^2$  p-valores acima de 0,05, RMSEA abaixo de 0,08, SRMR abaixo de 0,10, TLI acima de 0,9, CFI acima de 0,8 e CD acima de 0,8.<sup>6</sup>

### O modelo *Logit*

Um modelo de regressão do tipo  $Y_i = X_i\beta + \varepsilon_i$  em que

$$Y = \begin{cases} 1 \text{ com probabilidade } P \\ 0 \text{ com probabilidade } 1 - P \end{cases}$$

é dito um modelo de regressão de escolha binária e explica a probabilidade de  $Y$  assumir o valor 1 ( $P(Y=1)$ ), podendo ser escrito da seguinte forma:

$$P_i = P(Y_i = 1|X) = F(X_i\beta) \quad (5)$$

em que  $F(\cdot)$  é uma função de  $X_i\beta$ , com  $X_i$  representando a matriz de variáveis  $n \times k$  e  $\beta$  o vetor de parâmetros  $k \times 1$ , utiliza-se uma função de distribuição de probabilidade acumulada para garantir que  $0 \leq P_i \leq 1$  (GREENE (2008), CAMERON ; TRIVEDI (2005)).

Os modelos mais utilizados para modelar situações em que a variável dependente é binária são o *Probit* e o *Logit*. Uma questão comumente enfrentada pelo pesquisador é a de qual modelo considerar na análise. Conforme Cameron e Trivedi (2005), em termos empíricos, não há, em princípio, nenhuma razão para utilizar um em detrimento do outro. Existe pouca diferença entre as probabilidades estimadas a partir do *Logit* ou *Probit*. A diferença só se acentua nas caudas da função de densidade de probabilidade acumulada (FDA) usada em cada modelo<sup>7</sup>. Contudo, se o foco da análise for o cálculo de efeitos marginais em pontos médios da amostra, ao invés do efeito marginal para observações individuais, pouca diferença qualitativa e quantitativa há na utilização de um ou outro.

Tendo em vista o exposto, optou-se por utilizar o modelo *Logit*, que possui um atrativo adicional de poder apresentar os resultados em termos de *odds-ratio* (razão de chances a favor).

<sup>4</sup> Do inglês *Standardized Root Mean Square Residual*.

<sup>5</sup> Do inglês *Comparative Fit Index*.

<sup>6</sup> Para maiores informações sobre cada um destes indicadores e de outros ver Schermellen-Engel; Moosbrugger; Müller (2003).

<sup>7</sup>No caso do modelo *Logit*, a função utilizada é a logística, e no caso do modelo *Probit*, a função normal padronizada. A função logística possui caudas mais pesadas e topo menos inclinado que a normal padronizada.

O modelo *Logit* especifica a função  $F(.)$  como a função logística que pode ser representada como

$$P_i = \Lambda(X_i\beta) = \frac{1}{1 + e^{-X_i\beta}} = \frac{e^{X_i\beta}}{1 + e^{X_i\beta}} \quad (6)$$

em que  $\Lambda(X_i\beta)$  representa a função de distribuição logística acumulada.

O efeito marginal de uma variável explicativa é dado por

$$EM_{X_j} = \Lambda(X_i\beta)[1 - \Lambda(X_i\beta)]\beta_j \quad (7)$$

A *odds-ratio* é dada por

$$\frac{P_i}{1 - P_i} = \frac{F(X_i\beta)}{1 - F(X_i\beta)} = \frac{\Lambda(X_i\beta)}{1 - \Lambda(X_i\beta)} = \frac{e^{X_i\beta}}{1 + e^{X_i\beta}} = e^{X_i\beta} \quad (8)$$

A estimação de modelos de escolha binária, e em particular o *Logit*, é feita pelo método de máxima verossimilhança. No caso do modelo *Logit*, busca-se maximizar a seguinte função com relação ao vetor de parâmetros  $\beta$ .

$$\ln L = \sum_{i=1}^n Y_i \ln[\Lambda(X_i\beta)] + \sum_{i=1}^n (1 - Y_i) \ln[1 - \Lambda(X_i\beta)] \quad (9)$$

em que:

$\ln L$  representa o logaritmo natural da função de verossimilhança<sup>8</sup>.

Para fins de estimação (linearidade dos parâmetros), trabalha-se com o modelo em termos de logaritmo de chances a favor, assim a partir de (8) tem-se:

$$\frac{P_i}{1 - P_i} = e^{X_i\beta} \Rightarrow \ln\left(\frac{P_i}{1 - P_i}\right) = X_i\beta \quad (10)$$

A avaliação do modelo será feita pela estatística da razão de verossimilhança, que mede a significância global dos parâmetros, significância estatística individual dos parâmetros e pseudo- $R^2$ .<sup>9</sup>

## Modelo analítico

O modelo de AFC pode ser visto em termos de diagrama na Figura 1. Nesta figura, os círculos representam as variáveis latentes, enquanto os quadrados representam as variáveis observadas. Setas unidirecionais representam o sentido de

<sup>8</sup> Para um exame detalhado da derivação desta função, bem como das condições de primeira e segunda ordem, ver Greene (2008).

<sup>9</sup> Maiores informações sobre os aspectos formais destas medidas podem ser encontradas em Greene (2008) e Cameron e Trivedi (2005).

causalidade entre as variáveis latentes e observadas. Nesse sentido, pode-se dizer que, no modelo, há duas variáveis latentes exógenas e cinco variáveis observadas endógenas. Setas bidirecionais indicam covariância entre variáveis, sejam elas latentes ou observadas. Por fim, a cada variável observada, têm-se um termo de erro associado.

Para atender a necessidade de identificação e padronizar as estimativas dos parâmetros do modelo, restringiu-se a variância dos erros e das variáveis latentes como um. Isso pode ser visto na Figura 1 pela presença do número um que acompanha cada um destes termos.

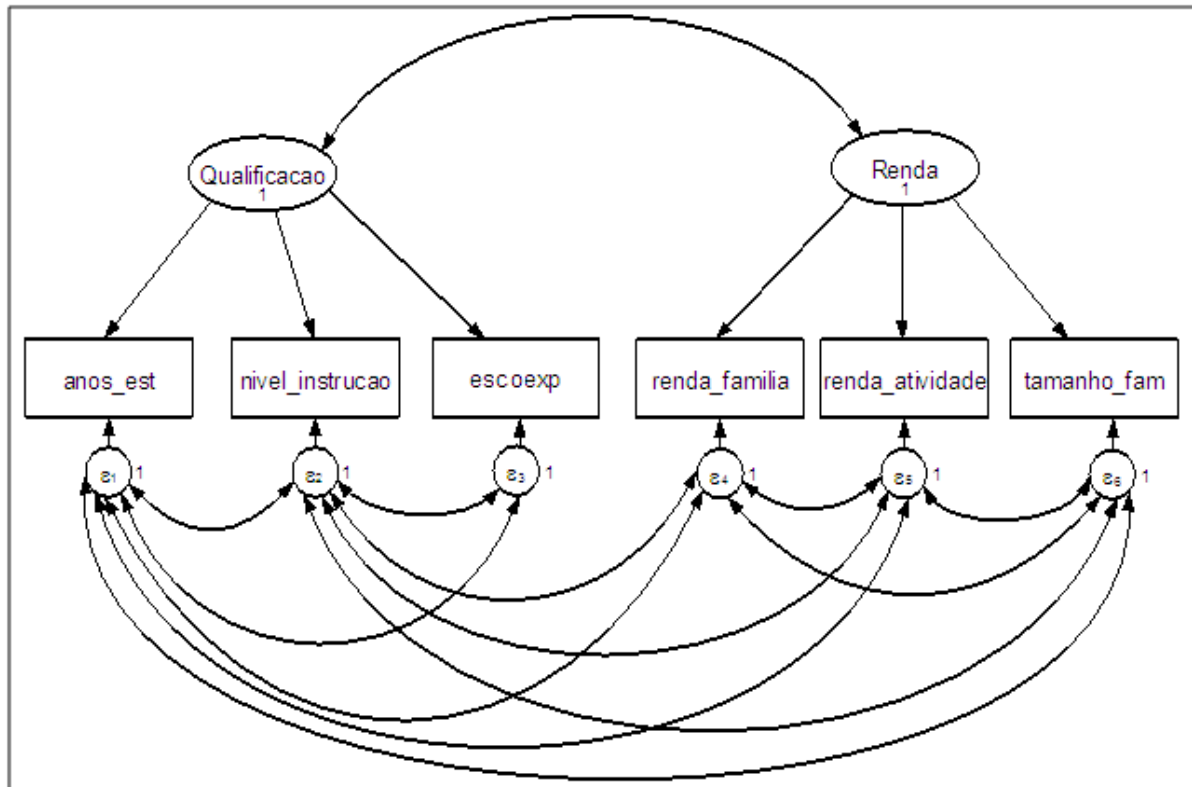


Figura 1 – Modelo de Análise Fatorial Confirmatória

Fonte: Elaboração dos autores.

O modelo estabelece a existência de duas variáveis latentes, uma denominada qualificação, que é manifestada pelas variáveis observadas anos de estudo, nível de instrução e pela multiplicação de escolaridade por experiência<sup>10</sup>, e a outra renda, manifestada pelas variáveis renda da família, renda da atividade e tamanho da família. Ademais, se estabelece a existência de covariância entre renda e qualificação e também entre as variáveis observadas. As covariâncias foram estabelecidas com base em índices de modificação<sup>11</sup>.

<sup>10</sup> A variável escoexp representa a multiplicação dos anos de estudo pela experiência (idade da pessoa), como em Figueiredo, Marta e Guimarães (2010).

<sup>11</sup>Para uma abordagem didática do papel de índices de modificação, bem como suas limitações, ver Brown (2006).



Com já mencionado, o modelo de AFC irá fornecer os escores fatoriais para as variáveis latentes renda e qualificação que serão utilizadas no modelo *Logit*.

O modelo *Logit* a ser estimado foi especificado como segue:

$$L_i = \ln\left(\frac{P_i}{1-P_i}\right) = \beta_1 + \beta_2 \text{genero} + \beta_3 \text{idade} + \beta_4 \text{idade}^2 + \beta_5 \text{etnia} + \beta_6 \text{estado\_civil} + \beta_7 \text{local\_nascimento} + \beta_8 \text{local\_residencia} + \beta_9 \text{renda} + \beta_{10} \text{qualificacao} + \beta_{11} \text{sul} + \beta_{12} \text{nordeste} + \beta_{13} \text{centro\_oeste} + \beta_{14} \text{norte} \quad (11).$$

A próxima seção detalha cada uma das variáveis utilizadas em (10) e (11), bem como a base de dados utilizada.

### Fonte de dados

A análise empírica realizada neste trabalho foi baseada nos microdados da Pesquisa Nacional por Amostra de Domicílio (PNAD) para o ano de 2009, ou seja, a pesquisa mais recente disponível. Os microdados da PNAD consistem em dados individuais das principais características socioeconômicas dos indivíduos e famílias, umas de caráter permanente, como as características gerais da população quanto a educação, trabalho, rendimento e habitação, e outras com periodicidade variável, como as características sobre migração, fecundidade, nupcialidade, saúde, nutrição e outros temas que são incluídos no sistema de acordo com as necessidades de informação para o País (IBGE, 2007).

A pesquisa é realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE) e incorpora todos os aspectos que definem um “plano amostral complexo”, a saber: estratificação das unidades de amostragem, conglomeração, probabilidades desiguais de seleção em um ou mais estágios e ajustes dos pesos amostrais para calibração com os totais populacionais conhecidos.

Tendo em vista que a PNAD é uma amostra complexa, a análise desse tipo de dados deve ser realizada utilizando-se todas as informações do desenho amostral da pesquisa e mensurando os efeitos dessa estrutura nas estimativas. O desenho amostral da PNAD é definido através de três informações complementares: os pesos de expansão da amostra, o estrato e a unidade primária de amostragem (UPA ou *PSU*). Esse conjunto de informações permite mensurar o efeito do desenho amostral nas estimativas e deve ser selecionado junto com as variáveis de interesse na PNAD (FIGUEIREDO; MARTA ; GUIMARÃES, 2010.)

As variáveis selecionadas para estimação do modelo estão descritas na Tabela 1, que define as variáveis da AFC e do modelo *logit*.

Tabela 1- Descrição das variáveis utilizadas no estudo

| VARIÁVEL                                  | DESCRIÇÃO  | TIPO DA VARIÁVEL |
|---|--|------------------|
|   | <b>Variável dependente</b>   |                  |
| Participação                              | Variável <i>dummy</i> que assume valor 1 se o indivíduo está ocupado na atividade agrícola e zero caso contrário | Categórica       |
|   | <b>Variáveis explicativas</b>  |                  |
| Gênero <sup>a</sup>                       | Variável <i>dummy</i> que assume valor 1 se o indivíduo é do sexo masculino                                      | Categórica       |
| Idade <sup>a</sup>                        | Idade do morador na data de referência   | Quantitativa     |
| Etnia <sup>a</sup>                        | Variável <i>dummy</i> que assume valor 1 se o indivíduo é branco   | Categórica       |
| Estado Civil <sup>a</sup>                 | Variável <i>dummy</i> que assume valor 1 se o indivíduo é casado   | Categórica       |
| Anos de estudo <sup>a</sup>               | Anos de estudo   | Quantitativa     |
| Local de nascimento <sup>a</sup>          | Variável <i>dummy</i> que assume valor 1 se o indivíduo nasceu na Unidade da Federação de residência (Imigração) | Categórica       |
| Tamanho da família <sup>b</sup>           | Número de componentes da família   | Quantitativa     |
| Nível de instrução familiar <sup>b</sup>  | Nível de instrução mais elevado alcançado (todas as pessoas)   | Categórica       |
| Local da residência familiar <sup>b</sup> | Variável <i>dummy</i> que assume valor 1 se o indivíduo reside na zona rural                                     | Categórica       |
| Renda da família <sup>b</sup>             | Rendimento mensal familiar <i>per capita</i>   | Quantitativa     |
| Sul <sup>c</sup>                          | Variável <i>dummy</i> que assume valor 1 se o indivíduo reside na região Sul                                     | Categórica       |
| Nordeste <sup>c</sup>                     | Variável <i>dummy</i> que assume valor 1 se o indivíduo reside na região Nordeste                                | Categórica       |
| Centro-oeste <sup>c</sup>                 | Variável <i>dummy</i> que assume valor 1 se o indivíduo reside na região Centro-Oeste                            | Categórica       |
| Norte <sup>c</sup>                        | Variável <i>dummy</i> que assume valor 1 se o indivíduo reside na região Norte                                   | Categórica       |

Fonte: Elaboração dos autores.

Nota: (a) variáveis que representam características individuais; (b) variáveis que representam características da família; e (c) *dummies* regionais, referência sudeste.

## ANÁLISE E DISCUSSÕES DOS RESULTADOS

### Evidências da população ocupada a partir da seleção da amostra

Com base na PNAD 2009, a amostra selecionada (indivíduos ocupados no período de referência de 365 dias) ficou representada por 208.580 pessoas (34.407 da área rural e 174.173 da área urbana) sendo 55,94% do sexo masculino e 44,06% do sexo feminino. Na Tabela 2, é possível observar que o percentual de trabalhadores urbanos é superior ao de trabalhadores rurais em todas as regiões brasileiras.

Tabela 2- Percentual de trabalhadores ocupados, por área e região brasileira

| <b>Região</b> | <b>Rural</b> | <b>Urbano</b> |
|---------------|--------------|---------------|
| Norte         | 22,96        | 77,04         |
| Centro-Oeste  | 12,27        | 87,73         |
| Nordeste      | 23,08        | 76,92         |
| Sudeste       | 8,88         | 91,12         |
| Sul           | 15,93        | 84,07         |
| <b>Total</b>  | <b>16,50</b> | <b>83,50</b>  |

Fonte: Elaboração dos autores a partir dos dados da PNAD/2009.

A Tabela 3, por sua vez, contém informações mais detalhadas a respeito da amostra selecionada, enfatizando a questão do gênero, idade, escolaridade e setor de atividade.

Tabela 3- Percentual de trabalhadores ocupados, por setor de atividade e região brasileira

| <b>Variável</b>                    | <b>Norte</b> | <b>Centro-Oeste</b> | <b>Nordeste</b> | <b>Sudeste</b> | <b>Sul</b> |
|------------------------------------|--------------|---------------------|-----------------|----------------|------------|
| <b><i>Gênero</i></b>               |              |                     |                 |                |            |
| Masculino                          | 57,69        | 55,66               | 56,81           | 55,25          | 54,40      |
| Feminino                           | 42,31        | 44,34               | 43,19           | 44,75          | 45,60      |
| <b><i>Faixa Etária</i></b>         |              |                     |                 |                |            |
| De 10 a 29 anos                    | 38,43        | 35,53               | 36,96           | 32,93          | 33,58      |
| De 30 a 49 anos                    | 44,45        | 45,91               | 44,24           | 46,08          | 44,66      |
| De 50 a 69 anos                    | 15,96        | 17,10               | 16,98           | 19,41          | 20,06      |
| De 70 anos ou mais                 | 1,16         | 1,46                | 1,82            | 1,58           | 1,70       |
| <b><i>Grau de Escolaridade</i></b> |              |                     |                 |                |            |
| Sem instrução                      | 8,82         | 5,98                | 13,08           | 4,67           | 3,68       |
| 1º Grau incompleto                 | 35,31        | 32,10               | 35,85           | 29,46          | 32,37      |
| 1º Grau completo                   | 8,97         | 8,81                | 7,78            | 10,45          | 11,64      |
| 2º Grau incompleto                 | 7,82         | 7,91                | 7,04            | 6,84           | 7,15       |

## Continuação

| Variável                  | Norte | Centro-Oeste | Nordeste | Sudeste | Sul   |
|---------------------------|-------|--------------|----------|---------|-------|
| 2º Grau completo          | 26,53 | 26,39        | 25,20    | 29,94   | 26,44 |
| Superior                  | 12,55 | 18,81        | 11,05    | 18,64   | 18,73 |
| <b>Setor de Atividade</b> |       |              |          |         |       |
| Agrícola                  | 19,26 | 13,22        | 23,13    | 9,47    | 13,84 |
| Industrial                | 10,63 | 11,17        | 10,20    | 16,72   | 19,20 |
| Comércio                  | 19,10 | 18,80        | 18,14    | 17,90   | 18,51 |
| Serviços                  | 51,02 | 56,81        | 48,52    | 55,91   | 48,45 |

Fonte: Elaboração dos autores a partir dos dados da PNAD/2009.

Do total de trabalhadores ocupados, observou-se que a maioria era do gênero masculino, em todas as regiões brasileiras. Ainda com respeito ao gênero do trabalhador, embora o número relativo de trabalhadores masculinos seja superior ao feminino, percebe-se que esta diferença é menor nas regiões Sudeste e Sul do país.

Quanto à análise por faixa etária, destaca-se que a participação dos ocupados nas faixas etárias mais baixas (10 a 29 anos e 30 a 49 anos) é muito superior às demais faixas etárias.

Em relação ao grau de escolaridade, percebe-se a redução no percentual de ocupados das classes com menor grau de escolaridade e o aumento da participação das classes com maior grau de escolaridade, independente da região brasileira.

As últimas variáveis de análise da Tabela 3 referem-se ao setor de atividade, no qual fica evidente a predominância do setor terciário em todas as regiões brasileiras. Entretanto, vale destacar o elevado percentual de pessoas ocupadas na atividade agrícola na região Nordeste do Brasil, bem como o elevado percentual de pessoas ocupadas na indústria nas regiões Sudeste e Sul do país.

## Determinantes da alocação de trabalho agrícola

Antes de se proceder à estimação do modelo de AFC, apresentado na Figura 1, realizou-se o teste de normalidade multivariada. Dada a rejeição de normalidade multivariada dos dados, a presença de variável ordinal (nível de instrução) e a utilização de uma amostra grande (87.646)<sup>12</sup>, optou-se pela estimação por MQP.

A Tabela 4 apresenta o modelo de AFC estimado e as medidas de avaliação de ajuste do mesmo. Como as estimativas estão padronizadas, os coeficientes das

<sup>12</sup> O valor de 87.646 é inferior ao valor de 208.580 reportado na subseção 3.1 porque só foi considerado na estimação do modelo observações disponíveis para todas as variáveis analisadas.

cargas fatoriais podem ser interpretados como medidas de correlação entre as variáveis observadas e as variáveis latentes<sup>13</sup>.

Tabela 4 – Modelo de Análise Fatorial Confirmatória e Medidas de Avaliação

| Variáveis                 | Qualificação |             |         | Renda |              |         | Avaliação do Modelo |        |
|---------------------------|--------------|-------------|---------|-------|--------------|---------|---------------------|--------|
|                           | Coef.        | Z           | P-Valor | Coef. | Z            | P-valor | Graus de Liberdade  |        |
| anos de estudo            | 0,97         | 11000,00    | 0,00    | -     | -            | -       | 8                   |        |
| nível de instrução        | 0,84         | 1587,54     | 0,00    | -     | -            | -       | $\chi^2$            | 751,59 |
| experiência               | 0,99         | 15000000,00 | 0,00    | -     | -            | -       | P-valor( $\chi^2$ ) | 0,00   |
| renda da família          | -            | -           | -       | 0,99  | 290000000,00 | 0,00    | RMSEA               | 0,03   |
| renda da atividade        | -            | -           | -       | 0,99  | 440000000,00 | 0,00    | CFI                 | 0,99   |
| tamanho da família        | -            | -           | -       | -0,71 | -274,65      | 0,00    | TLI                 | 0,98   |
| Corr(Qualificação,Renda): | 0,27         |             |         |       |              |         | SRMR                | 0,07   |
|                           |              |             |         |       |              |         | CD                  | 0,90   |

Fonte: Resultados da pesquisa.

Nota: (a) a idade da pessoa multiplicada pelos anos de estudo foi utilizada como *proxy* para experiência.

Dessa forma, percebe-se que a variável latente “Qualificação” tem correlações elevadas com todas as variáveis observadas, com destaque para anos de estudo e experiência, que apresentam correlações positivas e superiores a 0,95. Para a variável latente “Renda”, também se tem correlações positivas e elevadas, fato válido para as variáveis renda da família e renda da atividade. Por outro lado, percebe-se que, quanto maior a família, menor é a renda, constatação feita pela correlação negativa de -0,71.

Ao observar as medidas de avaliação do modelo, vê-se que, no geral, ele apresentou um bom ajustamento, com todas as cargas fatoriais significativas a 1% de significância. Da mesma forma, a estatística de  $\chi^2$  foi significativa a 1% de significância. Aqui cabem algumas ressalvas feitas em Schermellen-Engel; Moosbrugger; Müller (2003) à estatística de  $\chi^2$ :

- A estatística de  $\chi^2$  se baseia na suposição de que as variáveis observadas possuem distribuição normal multivariada.
- Com o aumento do tamanho da amostra e um número constante de graus de liberdade, o valor da estatística de  $\chi^2$  aumenta. Isso leva ao problema de que modelos plausíveis possam ser rejeitados baseados na significância da estatística de  $\chi^2$ , embora a diferença entre a matriz de covariância amostral e a matriz de covariância implícita do modelo não seja relevante.

<sup>13</sup> Optou-se por não reportar as covariâncias entre as variáveis observadas conforme Figura 1 por não ser este o foco da análise. O interesse primeiro aqui é o de obter os escores fatoriais das variáveis latentes a partir de um modelo bem ajustado.

Feitas essas observações, decidiu-se por avaliar o modelo, atribuindo-se maior peso às demais medidas de ajuste. Assim, com o RMSEA de 0,03 (abaixo do valor de 0,08), CFI de 0,99 (acima de 0,8), TLI de 0,98 (acima de 0,9), SRMR de 0,07 (abaixo de 0,10) e CD de 0,9 (acima de 0,8), conclui-se que o modelo apresentou um bom ajuste.

A Tabela 5 apresenta o modelo *Logit* estimado com as medidas de avaliação de razão de verossimilhança, pseudo-R<sup>2</sup> e significância individual dos parâmetros.

**Tabela 5 – Modelo *Logit* de alocação de trabalho agrícola**

| Variável            | Trabalho Agrícola |        |         |
|---------------------|-------------------|--------|---------|
|                     | Coeficiente       | Z      | P-valor |
| Gênero              | 1,29              | 39,91  | 0,00    |
| Idade               | 0,00              | 1,08   | 0,28    |
| Idade <sup>2</sup>  | -0,00             | -0,77  | 0,44    |
| Etnia               | -0,05             | -1,88  | 0,06    |
| Estado Civil        | 0,09              | 3,02   | 0,00    |
| Local de Nascimento | 0,02              | 0,96   | 0,33    |
| Local de Residência | 2,70              | 93,31  | 0,00    |
| Renda               | -0,13             | -7,58  | 0,00    |
| Qualificação        | -0,60             | -42,35 | 0,00    |
| Sul                 | 0,12              | 2,54   | 0,01    |
| Nordeste            | 0,12              | 2,96   | 0,00    |
| Centro-Oeste        | 0,30              | 6,52   | 0,00    |
| Norte               | 0,12              | 2,79   | 0,00    |
| Constante           | -4,55             | -40,75 | 0,00    |

Número de Observações: 87.646

LR Chi<sup>2</sup>(13) = 18.402,92 Prob > chi2 = 0,0000

Pseudo R2 = 0,3327

**Fonte:** Resultados da pesquisa.

Dado que os coeficientes apresentados na Tabela 5 representam o impacto sobre o logaritmo da razão de chances de um indivíduo estar alocado no trabalho agrícola, *a priori*, não se tem uma interpretação econômica para tais coeficientes.

Nota-se que as variáveis idade, idade<sup>2</sup> e local de nascimento não são estatisticamente significativas a 10% de significância. Todas as outras variáveis são estatisticamente significativas a 1% de significância (exceto etnia, que foi significativa a 10%). A estatística de razão de verossimilhança indica que, em conjunto, todas as variáveis são importantes para explicar a alocação de trabalho agrícola. O valor do pseudo-R<sup>2</sup> foi de 0,33, relativamente baixo, mas que tem suporte no teste de razão de verossimilhança.

A Tabela 6 apresenta as estimativas dos efeitos marginais para a média da amostra e a *odds-ratio* absoluta e percentual.

Tabela 6 – Estimativas do efeito marginal para o ponto médio da amostra, *odds-ratio* absoluta e percentual

| Variável            | Efeito Marginal |        |         | Chances a favor |        |         | Chances a favor % |        |         |
|---------------------|-----------------|--------|---------|-----------------|--------|---------|-------------------|--------|---------|
|                     | Coefficiente    | Z      | P-valor | Coefficiente    | Z      | P-valor | Coefficiente      | Z      | P-valor |
| Gênero              | 0.0493          | 39.48  | 0.00    | 3.65            | 39.91  | 0.00    | 264.95            | 39.91  | 0.00    |
| Idade               | 0.0018          | 1.70   | 0.08    | 1.00            | 1.70   | 0.08    | 0.19              | 1.70   | 0.08    |
| Etnia               | -0.0022         | -1.88  | 0.06    | 0.94            | -1.88  | 0.06    | -5.66             | -1.88  | 0.06    |
| Estado Civil        | 0.0036          | 3.03   | 0.00    | 1.09            | 3.02   | 0.00    | 9.69              | 3.02   | 0.00    |
| Local de Nascimento | 0.0011          | 0.97   | 0.33    | 1.03            | 0.96   | 0.33    | 3.03              | 0.96   | 0.33    |
| Local de Residência | 0.2756          | 55.81  | 0.00    | 14.93           | 93.31  | 0.00    | 1393.74           | 93.31  | 0.00    |
| Renda               | -0.0051         | -7.58  | 0.00    | 0.87            | -7.58  | 0.00    | -12.44            | -7.58  | 0.00    |
| Qualificação        | -0.0235         | -41.36 | 0.00    | 0.54            | -42.35 | 0.00    | -45.29            | -42.35 | 0.00    |
| Sul                 | 0.0048          | 2.45   | 0.01    | 1.12            | 2.54   | 0.01    | 12.85             | 2.54   | 0.01    |
| Nordeste            | 0.0050          | 2.87   | 0.00    | 1.13            | 2.96   | 0.00    | 13.34             | 2.96   | 0.00    |
| Centro-Oeste        | 0.0132          | 5.92   | 0.00    | 1.35            | 6.52   | 0.00    | 35.90             | 6.52   | 0.00    |
| Norte               | 0.0052          | 2.68   | 0.00    | 1.13            | 2.79   | 0.00    | 13.72             | 2.79   | 0.00    |

Fonte: Resultados da pesquisa.

Tendo em vista o exposto na Tabela 6, vê-se que, para o ponto médio da amostra, o fato de um indivíduo ser do sexo masculino aumenta a probabilidade de este estar alocado em um trabalho agrícola em 4,93 pontos percentuais. Ou em termos de razão de chances a favor, pode-se dizer que um indivíduo do sexo masculino tem 3,65 mais chances de estar alocado em um trabalho agrícola que indivíduos do sexo feminino (em termos percentuais 264,95% mais chances). Esses resultados vão ao encontro do trabalho de Figueiredo; Marta; Guimarães (2010), que analisaram os determinantes da ocupação do núcleo familiar rural no Brasil.

O fato de um indivíduo ser da cor branca reduz sua probabilidade de estar alocado em trabalho agrícola, mas este efeito é modesto, sendo de 0,22%, resultado que pode ser corroborado pelo trabalho de Cavalieri; Fernandes (1998), que investigou os diferenciais salariais nas regiões metropolitanas brasileiras. Indivíduos casados têm mais chance de estarem alocados em trabalho agrícola que indivíduos solteiros, mas tal efeito também é pequeno, 0,36%.

Obviamente, indivíduos que residem no meio rural têm maiores chances de estarem alocados em trabalho agrícola e isso representou o fator mais importante na determinação de tal alocação.

As variáveis latentes "Renda" e "Qualificação" se relacionam de forma negativa com a alocação de trabalho agrícola, indicando que indivíduos com maior renda e qualificação têm menores chances de estarem alocados em trabalho agrícola, sendo que o impacto maior é atribuído à variável "Qualificação".

Por fim, em termos geográficos, a região que mais contribui para um indivíduo estar alocado em um trabalho agrícola é a Centro-Oeste. Residir nessa

região aumenta a probabilidade de um indivíduo estar alocado em uma atividade agrícola em 1,32 pontos percentuais.

## CONCLUSÕES

As principais análises deste trabalho indicaram que ser do sexo masculino aumenta a probabilidade de estar alocado em um trabalho agrícola; o fato de um indivíduo ser da cor branca reduz sua probabilidade de estar alocado em trabalho agrícola, com os indivíduos residentes no meio rural apresentando maiores chances de estarem alocados em trabalho agrícola.

Renda e qualificação se relacionam de forma negativa com a alocação de trabalho agrícola, indicando que indivíduos com maior renda e qualificação têm menores chances de estarem alocados nesse tipo de atividade. Em termos geográficos, a região que mais contribui para um indivíduo estar alocado em um trabalho agrícola é a Centro-Oeste.

Este trabalho contribui com o estudo dos determinantes da alocação de trabalho agrícola nas regiões brasileiras. Contudo, torna-se necessário que pesquisas futuras considerem e aprofundem as discussões relacionadas a questões como os determinantes da alocação de trabalho não agrícola nos estados brasileiros, a incorporação de outros modelos e métodos de análise e uma maior comparação com outros trabalhos.

Outra abordagem que poderia enriquecer a análise seria a incorporação de uma dimensão temporal, dado que a única dimensão presente neste trabalho é a de seção cruzada. Isso poderia ser feito ao empilhar diversas PNADs mediante a utilização de um pseudopainel, tornando os resultados aqui encontrados mais robustos.

## REFERÊNCIAS

BOLLEN, K. A. **Structural equations with latent variables**. New York: Wiley, 1989.

BROWN T. A. **Confirmatory Factor Analysis for Applied Research-methodology in the Social Sciences Series**. New York: Guilford, 2006.

BUAINAIN, A. M.; DEDECCA, C. S. Mudanças e reiteração da heterogeneidade do mercado de trabalho agrícola. In: GASQUES, J. G.; VIEIRA FILHO, J. E. R.; NAVARRO, Z. (Org.) **A Agricultura Brasileira: desempenho, desafios e perspectivas**. Brasília: IPEA, 2010.

CAMERON, A. C.; TRIVEDI, P. K. **Microeconometrics: methods and applications**. New York: Cambridge University Press, 2005.



CAMPOLINA, B. *et al.* **O mercado de trabalho rural**: evolução recente, composição da renda e dimensão regional. Rio de Janeiro: IPEA, 2009. (Texto para discussão, 1398).

CAVALIERI, C.; FERNANDES, R. Diferenciais de salários por gênero e cor: uma comparação entre regiões metropolitanas brasileiras. **Revista de Economia Política**, v.18, n.1, p.158-175, 1998.

FEDERICI, A.; MAZZITELLI, A. **Dynamic factor analysis with STATA**. 2008. Disponível em: <http://www.stata.com/meeting/2italian/Federici.pdf>. Acesso em: 30 de março, 2012.

FIGUEIREDO, A. M. R.; MARTA, J. M. C.; GUIMARAES, P. W. Determinantes da ocupação do núcleo familiar composto na área rural do Brasil. **Revista Economia e Desenvolvimento**, v. 22, n.1, p. 44-61, 2010.

GEENE, W. H. **Econometric analysis**. 6 ed. New Jersey: Prentice Hall, 2008.

GRAZIANO DA SILVA, J. **A nova dinâmica da agricultura brasileira**. Campinas: UNICAMP, 1996.

GRAZIANO DA SILVA, J. DEL GROSSI, M. E. Rural nonfarm employment and incomes in Brazil: patterns and evolution. **World Development**, v.29, n.3, p.443-453, 2001.

HAIR JR., J. *et al.* **Multivariate data analysis**. 5. ed. New Jersey: Prentice Hall, 1998.

IBGE, Instituto Brasileiro de Geografia e Estatística. **PNAD**. 2007. Disponível em <<http://www.sidra.ibge.gov.br/bda/tabela/protabl.asp?z=p&o=16&i=P>>. Acesso em: 20 de março, 2012.

KAGEYAMA, A. Mudanças no trabalho rural no Brasil, 1992-2002. **Agricultura em São Paulo**. São Paulo, v.51, n.2, p.71-84, 2004.

LEE, S. Y. **Structural equation modelling**: a Bayesian approach. Wiley series in probability and statistics. Lee, Sik-Yum New York, NY, US: John Wiley & Sons Ltd, 2007.

MINGOTI, S. A. **Análise de dados através de método de estatística multivariada** - uma abordagem aplicada. Belo Horizonte: Editora UFMG, 2007.

SCHERMELLEH-ENGEL, K., MOOSBRUGGER, H; MÜLLER, H. Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures, **Methods of Psychological Research Online**, v. 8, n 2, p. 23-74, 2003.

THOMPSON, B. Exploratory and confirmatory factor analysis: Understanding concepts and applications. **American Psychological Association, Washington, DC, US, 2004.**

Submetido em 26/10/2012.

Aprovado em 20/03/2013.

**Sobre os autores:**

**Filipe de Moraes Cangussu Pessoa**

Doutorando e Mestre em Economia Aplicada pela Universidade Federal de Viçosa e Economista do Instituto Federal de Educação, Ciência e Tecnologia de Brasília.

E-mail: [filipe\\_morais\\_pessoa@yahoo.com.br](mailto:filipe_morais_pessoa@yahoo.com.br)

**Daniel Arruda Coronel**

Professor Adjunto do Programa de Pós-Graduação em Administração da Universidade Federal de Santa Maria (UFSM) e Doutor em Economia Aplicada pela Universidade Federal de Viçosa.

E-mail: [daniel.coronel@uol.com.br](mailto:daniel.coronel@uol.com.br)

**Airton Lopes Amorim**

Economista e Mestre em Economia pela Universidade Federal de Viçosa.

Endereço: Avenida Peter Henry Rolfs, s/n, Centro. CEP: 36570-000 - Vicososa, MG - Brasil

E-mail: [aimorim2007@yahoo.com.br](mailto:aimorim2007@yahoo.com.br)

**João Eustáquio de Lima**

Professor Titular do Programa de Pós-Graduação em Economia Aplicada da Universidade Federal de Viçosa e Bolsista de Produtividade do CNPq.

Endereço: Endereço: Avenida Peter Henry Rolfs, s/n, Centro. CEP: 36570-000 - Vicososa, MG - Brasil

E-mail: [jelima@ufv.br](mailto:jelima@ufv.br)