Universidade de Brasília

Instituto de Ciências Exatas

Departamento de Estatística

Dissertação de Mestrado

Análise de regressão em dois estágios com modelos DEA na presença de variáveis contextuais endógenas via distribuição Beta-Inflacionada

por

Bruno Soares de Castro

Brasília

2019

Análise de regressão em dois estágios com modelos DEA na presença de variáveis contextuais endógenas via distribuição Beta-Inflacionada

por

Bruno Soares de Castro

Dissertação apresentada ao Departamento de Estatística da Universidade de Brasília, como requisito parcial para obtenção do título de Mestre em Estatística.

Orientador: Prof. Dr. Geraldo da Silva e Souza
Coorientador: Prof. Dr. Bernardo Borba de Andrade

Brasília
2019

Dissertação submetida ao Programa de Pós-Graduação em Estatística do Departamento de Estatística da Universidade de Brasília como parte dos requisitos para a obtenção do grau de Mestre em Estatística.

Texto aprovado por:

Prof. Geraldo da Silva e Souza
Orientador, Embrapa/UnB

Prof. Bernardo Borba de Andrade
Coorientador, EST/UnB

Dr.ª Eliane Gonçalves Gomes
Embrapa

Prof. André Luiz Fernandes Cançado
EST/UnB

Algo só é impossível até que alguém duvide e resolva provar o contrário.

(Albert Einstein)

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Resumo

O presente trabalho estuda os modelos de análise envoltória de dados em dois-estágios. O objetivo geral é estudar a viabilidade do modelo de regressão Beta inflacionada em um para o estudo da influência de variáveis contextuais endógenas na eficiência medida por análise envoltória de dados, via método dos momentos generalizado, com uso de variáveis instrumentais e via método de máxima verossimilhança em dois estágios, com correção de Murphy e Topel (1985). O trabalho também discute o método bootstrap para comparação das estimativas dos métodos clássicos. Para ilustrar as metodologias, uma aplicação aos dados municipais do censo agropecuário brasileiro de 2006 é realizada. Os resultados da regressão Beta inflacionada em um pelo método de máxima verossimilhança ou do método dos momentos generalizado apresentou resultados similares. Na presença de variáveis contextuais endógenas, sugere-se uma abordagem por método de máxima verossimilhança com correção de Murphy e Topel (1985). Ademais, os métodos supracitados fornecem erros padrão confiáveis, sem a necessidade de métodos bootstrap.

Palavras-chave: Análise envoltória de dados; Distribuição Beta inflacionada em um; Endogeneidade.

Abstract

This work aims to study the data envelopment analysis models in two stages. The general objective is to study the viability of the one-inflated Beta regression model for the study of the endogenous contextual variables influence on the efficiency measured by data envelopment analysis, via generalized method of moments, with the use of instrumental variables and two-stage maximum likelihood method, with correction by Murphy and Topel (1985). The research also discusses the bootstrap method for comparison of classical method estimates. To illustrate the methodologies, an application to municipal data from the 2006 Brazilian agricultural census is performed. The results of one-inflated Beta regression by the maximum likelihood method or the generalized method of moments presented similar results. In the presence of endogenous contextual variables, a maximum likelihood approach with Murphy and Topel correction (1985) is suggested. In addition, the above methods provide reliable standard errors without the need for bootstrap methods.

Keywords: Data envelopment analysis; One-inflated Beta distributions; Endogeneity.

Lista de Tabelas

Lista de Figuras

Sumário

Capítulo 1

Introdução

Nos mais diversos setores, sejam privados ou governamentais, as unidades tomadoras de decisão buscam ser cada vez mais eficientes. Farrell (1957) definiu dois conceitos de eficiência, técnica e alocativa, em que o primeiro busca uma quantidade ótima de insumos para uma capacidade máxima de produção e o segundo busca uma quantidade ótima de insumos em relação ao preço de cada um deles.

Charnes, Cooper e Rhodes (1978) desenvolveram a técnica de análise envoltória de dados que é uma extensão da medida de eficiência de Debreu (1951) e Farrell (1957). A análise envoltória de dados consiste em uma abordagem de programação linear para o desenvolvimento de indicadores de eficiência técnica e fronteiras da produção, por meio de múltiplos insumos e produtos. Destaca-se que esta é uma técnica não paramétrica, o que elimina a necessidade de se fazer qualquer suposição a priori sobre a forma funcional entre os insumos e os produtos, diferente dos modelos de fronteira estocástica de produção.

Os modelos mais comuns para a análise envoltória de dados são o de retornos constantes à escala e o de retornos variáveis à escala. O modelo de retornos constantes à escala considera que qualquer variação nos insumos leva a uma variação proporcional nos produtos. Enquanto o modelo de retornos variáveis à escala considera retornos crescentes ou decrescentes de escala na curva de produção. Tais modelos foram desenvolvidos por Charnes, Cooper e Rhodes (1978) e Banker, Charnes e Cooper (1984), respectivamente. Trabalhos como Emrouznejad e Yang (2018) e Liu et al. (2013) são artigos de revisão de aplicações.

Além de estudar quais são as unidades tomadoras de decisão eficientes, é fundamental saber quais variáveis contextuais podem influenciar nessas eficiências. Na literatura encontram-se

estudos que abordam esta circunstância, denominada por análise envoltória de dados em dois-estágios. O primeiro estágio consiste no cálculo da eficiência via análise envoltória de dados e o segundo é um modelo de regressão, como os modelos tobit de dois-limites (Simar e Wilson, 2007), fracionado (Hoff, 2007; McDonald, 2009; Papke e Wooldridge, 1996) e fracionado em duas-partes (Ramalho, Ramalho e Henriques, 2010).

Contudo, o ideal é buscar uma distribuição que se adeque à variável em estudo: a medida de eficiência. Ospina e Ferrari (2010) propõem uma abordagem para quando uma variável está no intervalo de zero e um, inclusive contendo um dos extremos. A proposta é generalizar uma distribuição, de modo a incluir o ponto de massa no zero ou no um. Desta forma, ter-se-á uma mistura de distribuições, ou seja, para o ponto de massa atribui-se uma distribuição degenerada e para a parte contínua atribui-se a distribuição Beta. Tal distribuição designa-se Beta inflacionada, a qual é uma alternativa razoável para modelar a eficiência.

O modelo de regressão apropriado para estudar quais variáveis contextuais podem influenciar a eficiência é definido pela média da distribuição Beta inflacionada. As estimativas dos parâmetros para o modelo são obtidas pelo método de máxima verossimilhança, o qual baseia-se em resultados assintóticos. Porém, na prática, a teoria assintótica pode obter resultados que são inadequados, caso algumas das condições do modelo não sejam satisfeitas, problema este que pode estar associado à endogeneidade.

É possível que ocorra endogeneidade por diversos fatores, tais como, causalidade simultânea, omissão de variáveis, erros nas variáveis e função mal especificada. Há vários métodos de estimação para esta situação, como estimador de variáveis instrumentais, mínimos quadrados de dois estágios e método dos momentos generalizado via variáveis instrumentais.

Leiderman (1980) também sugeriu um procedimento de estimativa para o método de máxima verossimilhança com presença de endogeneidade, denominado de máxima verossimilhança de informação completa. Posteriormente, Murphy e Topel (1985) propuseram um estimador, descrito por Greene (2012, p. 536), chamado de máxima verossimilhança de dois-estágios. Informações adicionais podem ser consultadas em Davidson e MacKinnon (1995, cap. 7, 17 e 18).

1.1 Objetivos

O objetivo geral desta dissertação é estudar a viabilidade do modelo de regressão Beta inflacionada em um, para o estudo da influência de variáveis contextuais exógenas e endógenas na eficiência medida por análise envoltória de dados, estabelecendo um modelo que se adeque às medidas de eficiência.

Para alcançar o objetivo geral, é necessário alcançar os seguintes objetivos específicos:

(i) Modelar a eficiência medida por análise envoltória de dados pela regressão Beta inflacionada em um, via máxima verossimilhança e método dos momentos generalizado;

(ii) Tratar possíveis variáveis contextuais endógenas via:

  (a) Método dos momentos generalizado, com o uso de variáveis instrumentais;

  (b) Método de máxima verossimilhança em dois estágios, com correção de Murphy e Topel (1985);

(iii) Usar o método bootstrap para comparação dos resultados aos métodos clássicos.

Capítulo 2

Regressão Beta Inflacionada para modelagem da eficiência obtida por Análise Envoltória de Dados

2.1 Introdução

A análise de envoltória de dados (DEA, em inglês) é um procedimento não estatístico baseado em programação linear no qual dados referentes a insumos e produtos são utilizados para obtenção de um escore de eficiência técnica. Os dados compreendem uma matriz $\mathbf{U}_{s\times n}$ insumos e $\mathbf{W}_{m\times n}$ produtos, aos quais são compostos por s insumos e m produtos para cada unidade produtiva na amostra, denominada de unidade tomadora de decisão (DMU, em inglês). Denotaremos por n o número de DMU's disponível na análise.

A análise de envoltória de dados se baseia em diversos conceitos e definições formais como eficiência técnica, conjunto (tecnológico) de possibilidades, fronteira de produção, disponibilidade livre (de insumos e produtos), orientação (insumo ou produto) e etc. Ademais, existem diferentes modelos DEA. Na ilustração do Cap. 5 utilizaremos o modelo retornos variáveis à escala (VRS, em inglês) com orientação ao produto. Neste modelo os escores de eficiência são maiores ou iguais a 1. Por exemplo, uma DMU que receba escore de 1.2 é tal que poderia produzir 20% a mais mantendo seus insumos constantes. Uma DMU com escore 1 é eficiente e não teria como produzir mais sem alterar seus insumos. As DMUs com escore unitário formam a fronteira de

produção e são ditas eficientes no sentido de Farrel.

Os escores permitem ordenar as unidades em termos de suas eficiências, mas são relativos, podendo se alterar conforme as DMU's presentes na análise. A metodologia DEA é amplamente utilizada em pesquisa operacional, engenharia industrial e economia. A literatura associada é vasta, os vários modelos existentes possuem diversas propriedades e fornecem diferentes medidas. Não caberia neste trabalho uma revisão dessa metodologia e sugerimos as referências clássicas Coelli et al. (2005) e Cooper (2007) como ponto de partida. Para os fins desta dissertação as medidas de eficiência serão dadas e nosso objetivo será modelar estatisticamente tais medidas como funções de variáveis explicativas. Para tanto propomos utilizar a regressão beta inflacionada.

A regressão beta (Ferrari e Cribari-Neto, 2004) e versões inflacionadas (Ospina e Ferrari, 2010, 2012) foram propostas para modelagem de variáveis resposta na forma de taxas ou proporções, $y \in [0, 1]$. Por exemplo, pode-se modelar a fração de petróleo convertido em gasolina, após destilação e fracionamento, como função da temperatura na qual toda a gasolina foi vaporizada e em função de outros fatores experimentais (Ferrari e Cribari-Neto, 2004). Neste trabalho utilizaremos, de modo inédito, o modelo de regressão beta inflacionado para modelar as eficiências medidas por DEA com função de variáveis contextuais (não discretionárias, ambientais) que não tenham tido papel na análise de envoltória.

Consideraremos que as eficiências em questão, $\{\theta_t; t = 1, \ldots, n\}$, estejam no intervalo $(0, 1]$, não importanto a direção e os retornos de escala adotados na etapa DEA. No exemplo acima, a DMU com escore 1.2 teria escore $\theta_t = 1/1.2 = 0.83$ e as unidades eficientes continuariam com escore 1. Observa-se que em casos de grandes amostras é raro não se obter um número elevado de DMUs eficientes. Em termos do modelo estatístico, esse fenômeno é denominado de inflação de uns.

2.2 Modelagem dos escores da eficiência

Nesta seção será proposto um modelo de regressão para a eficiência θ_t medida por DEA dado por

$$\mu_t = \mathrm{E}(\theta_t | \mathbf{x_t}, \boldsymbol{\beta}), \tag{2.1}$$

em que $\mathbf{X}$ $(n \times p)$ tem t-ésima linha igual a $\mathbf{x}_t = (\mathbf{x}_{t1}, \ldots, \mathbf{x}_{tp})$, com observações das p variáveis

contextuais e $\boldsymbol{\beta} \in \Theta \subseteq \mathbb{R}^p$ é um parâmetro p-dimensional desconhecido.

Além do modelo clássico de regressão, a literatura contém diferentes modelos propostos para lidar com resposta no intervalo $(0, 1]$, incluindo a regressão tobit, Pseudo-Máxima verossimilhança, versões não inflacionada com a distribuição beta e métodos de dois estágios (Hoff, 2007; McDonald, 2009; Papke e Wooldridge, 1996; Ramalho, Ramalho e Henriques, 2010; Simar e Wilson, 2007). Ospina e Ferrari (2010) definiram a distribuição Beta inflacionada em zero ou um, a qual se ajusta tanto ao valor um, quanto à parte fracionada dos escores da eficiência medidas por DEA. A média desta distribuição é uma alternativa para estudar quais variáveis contextuais podem influenciar a eficiência.

Uma questão levantada na literatura é correlação entre as observações das eficiências medidas por DEA. A correlação entre as observações, θ_t, surge de forma semelhante aos resíduos da regressão por mínimos quadrados ordinários. Em amostras finitas, a estimativa da eficiência da DMU_k será influenciada pelas medidas das demais DMU's (Simar e Wilson, 2007). Desta forma, em conformidade com as condições de regularidade em Greene (2012, p. 514) e Daraio e Simar (2007, p. 47), em que o estimador $\hat{\beta}_n$ é consistente para β, se $\hat{\beta}_n \xrightarrow{p} \beta$ quando $n \to \infty$, espera-se que com grande amostra a correlação gerada pela eficiência medida por DEA seja irrelevante, com a garantia de consistência, situação semelhante ao que ocorre com os resíduos da regressão.

Uma outra questão presente em muitos modelos econométricos é a presença de varáveis endógenas, principalmente quando as variáveis contextuais são correlacionadas parcialmente ou totalmente com os resíduos do modelo. Na presença de endogeneidade necessita-se de um método que possa corrigir a influência da endogeneidade nas estimativas dos parâmetros do modelo. Há na literatura alguns métodos regularmente utilizados, tais como, método dos momentos generalizado (Davidson e MacKinnon, 1995, cap. 7) e máxima verossimilhança em dois-estágios (Murphy e Topel, 1985).

2.3 Distribuição Beta

Ferrari e Cribari-Neto (2004) apresentaram uma nova reparametrização em relação aos parâmetros de forma da distribuição Beta. Na nova reparametrização, a função densidade de Y é

$$f(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma(\phi(1-\mu))} y^{\mu\phi-1}(1-y)^{\phi(1-\mu)-1}\mathbb{I}_{(0,1)}(y), \tag{2.2}$$

com $\mu \in (0,1)$ e $\phi > 0$. O momento de ordem r é

$$\mu_r = \mathrm{E}(Y^r) = \frac{\Gamma(\phi)\Gamma(\mu\phi + r)}{\Gamma(\phi + r)\Gamma(\mu\phi)} = \frac{(\mu\phi)_r}{\phi_r}, \tag{2.3}$$

no qual $(a)_{(r)} = a(a+1)(a+2)...(a+r-1)$. Por conseguinte, $Y \sim Beta(\mu, \phi)$, com média e variância, respectivamente, dadas por

$$\begin{aligned} \mathrm{E}[Y] &= \mu, \\ \mathrm{Var}[Y] &= \frac{V(\mu)}{1+\phi}, \end{aligned} \tag{2.4}$$

em que $V(\mu) = \mu(1-\mu)$, é a função de variância, definição semelhante aos modelos lineares generalizados. Note que ϕ pode ser definido como um parâmetro de dispersão, desde que a média seja fixa. Quanto maior for ϕ menor será a variância.

Na Figura 2.1 temos vários exemplos de densidade da distribuição Beta. Para valores de $\mu \neq 0.5$ a densidade tem comportamento assimétrico e quando $Y \sim Beta(\mu = 0.5, \phi = 2)$, a distribuição reduz-se a $Uniforme(0,1)$.

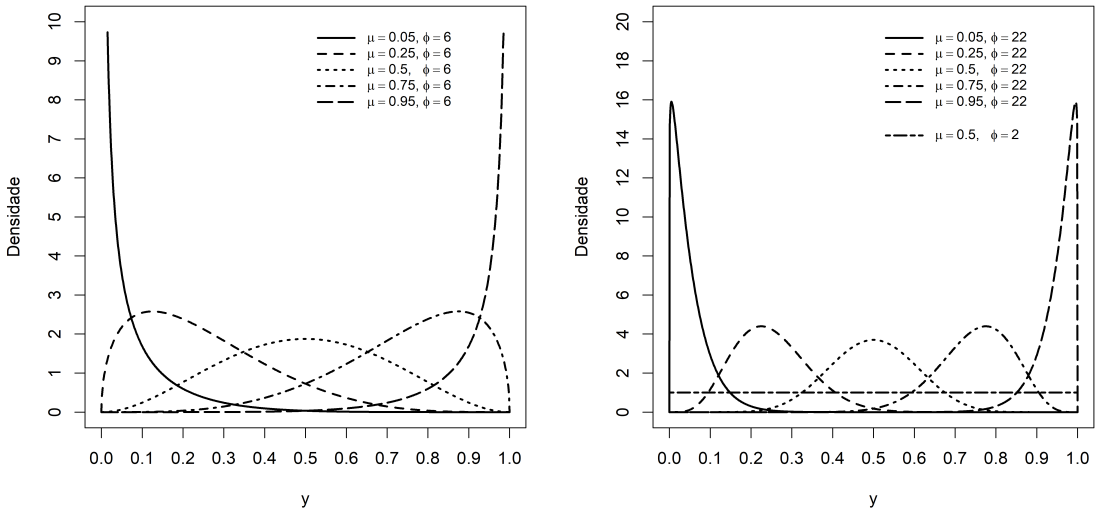

Figura 2.1: Gráficos de densidade da distribuição beta para diferentes valores de μ e ϕ

Além da flexibilidade de forma da distribuição Beta, um outro ponto importante a ser destacado é que ela também é um caso especial da família exponencial, em que a definição da preposição dada por este trabalho é expressa por:

Proposição 2.3.1. A distribuição Beta dada por (2.2) pertence à família exponencial de dimensão dois de posto completo.

Demonstração: Seja $\eta = (\eta_1, \eta_2)$, tal que $\eta_1 = \mu\phi$ e $\eta_2 = \phi$, $B(\eta_1, \eta_2) = \log(\Gamma(\eta_1)) + \log(\Gamma(\eta_2 - \eta_1)) - \log(\Gamma(\eta_2))$ é uma função com valores reais em η e o vetor de estatística $t(y) = (t_1(y), t_2(y))$, sendo que

$$\begin{aligned} t_1(y) &= \begin{cases} \log\left(\frac{y}{1-y}\right), & \text{se } y \in (0,1) \\ 0, & \text{se c. c.} \end{cases} \\ t_2(y) &= \begin{cases} \log(1-y), & \text{se } y \in (0,1) \\ 0, & \text{se c. c.} \end{cases} \end{aligned} \tag{2.5}$$

Reescrevendo a Equação (2.2), tem-se

$$f(y; \mu, \phi) = \exp\{\eta^\top T(y) - B(\eta)\}h(y), \tag{2.6}$$

em que $h(y)$ é uma função positiva definida como

$$h(y) = \begin{cases} \frac{1}{y(1-y)}, & \text{se } y \in (0,1) \\ 0, & \text{se c. c.} \end{cases}$$

A parametrização η constitui uma transformação bijetora que leva $\mathfrak{X} = \{(\mu, \phi) \in \mathbb{R}^2 : (0,1) \times \mathbb{R}^+\}$ a $\mathfrak{D} = \{\eta = (\eta_1, \eta_2) \in \mathbb{R}^2 : \mathbb{R}^+ \times \mathbb{R}^+\}$, um subconjunto aberto em $\mathbb{R}^2$. Os $t's$ e os $\eta's$ são linearmente independentes e o espaço paramétrico contêm retângulos bidimensionais.

A distribuição Beta apresentada em (2.6) pertence à família exponencial de posto completo, com as condições usuais de regularidade satisfeitas (Casella e Berger, 2002; McCullagh e Nelder, 1983). Consequentemente, o vetor de estatísticas $t(y) = (t_1(y), t_2(y)) = (y^{**}, y^{\dagger\dagger})$ apresentado em (2.5) possui estatísticas suficientes e completas, com momentos dados por

$$
\begin{aligned}
\mu^* &= \mathrm{E}(y^{**}) = \psi(\eta_1) - \psi(\eta_2 - \eta_1) = \psi(\mu\phi) - \psi(\phi(1-\mu)),\\
\mu^\dagger &= \mathrm{E}(y^{\dagger\dagger}) = \psi(\eta_2 - \eta_1) - \psi(\eta_2) = \psi(\phi(1-\mu)) - \psi(\phi),\\
v^* &= \mathrm{Var}(y^{**}) = \psi'(\eta_1) + \psi'(\eta_2 - \eta_1) = \psi'(\mu\phi) + \psi'(\phi(1-\mu)),\\
v^\dagger &= \mathrm{Var}(y^{\dagger\dagger}) = \psi'(\eta_2 - \eta_1) - \psi'(\eta_2) = \psi'(\phi(1-\mu)) - \psi'(\phi),\\
c^{*\dagger} &= \mathrm{Cov}(y^{**}, y^{\dagger\dagger}) = \mathrm{Cov}(y^{\dagger\dagger}, y^{**}) = -\psi'(\eta_2 - \eta_1) = -\psi'(\phi(1-\mu)),
\end{aligned}
\tag{2.7}
$$

em que $\psi(x) = \mathrm{dlog}(\Gamma(x))/dx = \Gamma'(x)/\Gamma(x)$, i.e., ψ é a função digama e ψ' é a derivada da função digama.

2.4 Distribuição Beta inflacionada

Quando há um número excessivo de observações nos extremos da variável Y, no zero ou um, é de interesse adicionar essa informação na distribuição. A distribuição Beta inflacionada é dividida em duas partes: uma é a distribuição Beta para as observações fracionada e a outra é uma distribuição degenerada em um ponto de massa (Ospina e Ferrari, 2010).

A função de distribuição acumulada da mistura é dada por

$$
BI_c(y; \alpha, \mu, \phi) = \alpha \mathbb{I}_{(c)}(y) + (1-\alpha)F(y; \mu, \phi), \tag{2.8}
$$

na qual $F(y; \mu, \phi)$ é a função de distribuição acumulada Beta, $\mathbb{I}_c(y)$ é a função indicadora do ponto de massa em c, a qual recebe valor 1 quando $y \in c$ e 0 quando $y \notin c$ e α é a probabilidade no ponto de massa, $P(y = c) = \alpha$.

A mistura tem a função densidade da seguinte maneira

$$
bi_c(y; \alpha, \mu, \phi) = \begin{cases} \alpha, & \text{se } y = c \\ (1-\alpha)f(y; \mu, \phi), & \text{se } y \in (0,1) \end{cases} \tag{2.9}
$$

ou, equivalentemente

$$
bi_c(y; \alpha, \mu, \phi) = \{\alpha^{\mathbb{I}_{(c)}(y)}(1-\alpha)^{1-\mathbb{I}_{(c)}(y)}\}\{f(y; \mu, \phi)^{1-\mathbb{I}_{(c)}(y)}\}, \tag{2.10}
$$

na qual $f(y; \mu, \phi)$ é a distribuição Beta definida em (2.2).

Ospina e Ferrari (2010) usam a seguinte termologia: Seja y uma variável aleatória dada por $Y \sim BI_c(\alpha, \mu, \phi)$, então

1. Se $c = 0$, ou seja, $P(y = 0) = \alpha$, é chamada de distribuição Beta inflacionada em zero - BIZ, onde $Y \sim BIZ(\alpha, \mu, \phi)$.

2. Se $c = 1$, ou seja, $P(y = 1) = \alpha$, é chamada de distribuição Beta inflacionada em um - BIU, onde $Y \sim BIU(\alpha, \mu, \phi)$.

O momento de ordem r é definido por $\mathrm{E}(Y^r) = \alpha c + (1 - \alpha)\mu_r$ com $r = 1, 2, ...$ e μ_r definido em (2.3). A média e a variância da variável aleatória y são dadas pela seguinte expressão

$$\begin{aligned} \mathrm{E}[Y] &= \alpha c + (1 - \alpha)\mu, \\ \mathrm{Var}[Y] &= (1 - \alpha)\frac{\mu(1 - \mu)}{1 + \phi} + \alpha(1 - \alpha)(c - \mu)^2. \end{aligned} \tag{2.11}$$

Atente-se que α exerce uma ponderação entre a parte degenerada e a parte da distribuição Beta, na esperança da distribuição Beta inflacionada.

Na Figura 2.2 temos a densidade da distribuição BIU. Um ponto importante a se destacar é que a distribuição sempre será assimétrica, havendo bimodalidade, desde que $\alpha > 0$.

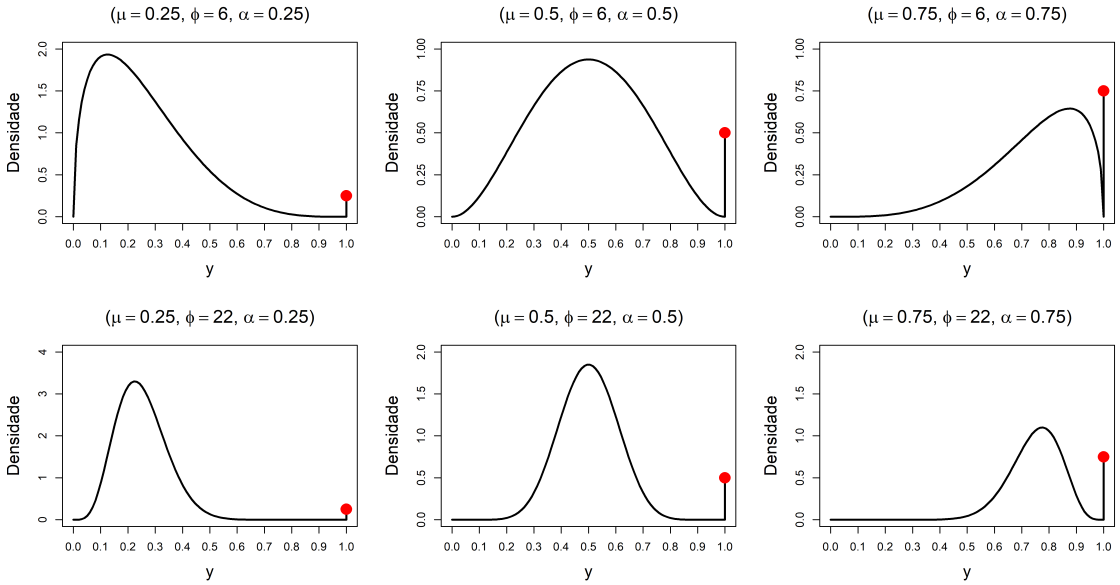

Figura 2.2: Gráficos da densidade da BIU para diferentes valores de μ, ϕ e α.

Quando $y \in (0, 1)$ a forma da densidade da distribuição não se altera para $y = 1$ ou $y = 0$, mantendo-se as propriedades da distribuição Beta. Ademais, a distribuição Beta inflacionada

faz parte da família exponencial. A proposição a seguir está em Ospina e Ferrari (2010), com modificações.

Proposição 2.4.1. A distribuição Beta inflacionada, a qual possui três parâmetros, pertence à família exponencial de posto completo.

Demonstração: Seja $\eta = (\eta_1, \eta_2, \eta_3)$, tal que $\eta_1 = [\log(\alpha/(1-\alpha)) + B(\eta_2, \eta_3)]$, $\eta_2 = \mu\phi$ e $\eta_3 = \phi$, onde $B(\eta_2, \eta_3) = \log(\Gamma(\eta_2)\Gamma(\eta_3)/\Gamma(\eta_3 - \eta_2))$ e os vetores de estatística $T(y) = (t_1(y), t_2(y), t_3(y))$, sendo que

$$\begin{aligned} t_1(y) &= \begin{cases} 1, & \text{se } y = c \\ 0, & \text{se } y \in (0,1) \end{cases} \\ t_2(y) &= \begin{cases} \log\left(\frac{y}{1-y}\right), & \text{se } y \in (0,1) \\ 0, & \text{se } y = c \end{cases} \\ t_3(y) &= \begin{cases} \log(1-y), & \text{se } y \in (0,1) \\ 0, & \text{se } y = c \end{cases} \end{aligned} \tag{2.12}$$

podendo reescrever a Equação (2.9) na seguinte expressão

$$bi_c(y; \alpha, \mu, \phi) = exp\{\eta^\top T(y) - B^*(\eta)\}h(y), \tag{2.13}$$

na qual $B^* = \log\{1 + exp[\eta_1 - B(\eta_2, \eta_3)]\} + B(\eta_2, \eta_3)$ é uma função com valores reais em η e

$$h(y) = \begin{cases} \dfrac{1}{y(1-y)}, & \text{se } y \in (0,1) \\ 0, & \text{se } y = c \end{cases}$$

$h(y)$ é uma função positiva definida sobre o conjunto $(0,1) \cup \{c\}$. A parametrização η estabelece uma transformação bijetora em $\mathfrak{X} = \{(\alpha, \mu, \phi) \in \mathbb{R}^3 : (0,1) \times (0,1) \times \mathbb{R}^+\}$ a $\mathfrak{D} = \{\eta = (\eta_1, \eta_2, \eta_3) \in \mathbb{R}^3 : \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+\}$, um conjunto aberto em $\mathbb{R}^3$. Todavia, os $t's$ e os $\eta's$ não satisfazem contrastes lineares, ou seja, são linearmente independentes e o espaço paramétrico contém retângulos tridimensionais.

Desta forma $bi_c(y; \alpha, \mu, \phi)$ é da família exponencial de posto completo, com as condições usuais de regularidade satisfeitas (Cox e Hinkley, 1979, pg. 107). As estatísticas $t(y) =$

$(t_1(y), t_2(y), t_3(y)) = (\mathbb{I}_c(y), y^*, y^\dagger)$ são suficientes e completas (Casella e Berger, 2002; McCullagh e Nelder, 1983), assim os momentos para $t(y)$ podem ser dados como

$$\begin{aligned}
\mathrm{E}(\mathbb{I}_c(y)) &= \frac{exp[\eta_1 - B(\eta_2, \eta_3)]}{1 + exp[\eta_1 - B(\eta_2, \eta_3)])} = \alpha, \\
\mathrm{Var}(\mathbb{I}_c(y)) &= \frac{exp[\eta_1 - B(\eta_2, \eta_3)]}{1 + exp[\eta_1 - B(\eta_2, \eta_3)])} \frac{1}{1 + exp[\eta_1 - B(\eta_2, \eta_3)])} = \alpha(1-\alpha),
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(y^*|y \in (0,1)) &= \mathrm{E}(y^{**}) = \mu^*, \\
\mathrm{Var}(y^*|y \in (0,1)) &= \mathrm{Var}(y^{**}) = v^*,
\end{aligned} \quad . \tag{2.14}$$

$$\begin{aligned}
\mathrm{E}(y^\dagger|y \in (0,1)) &= \mathrm{E}(y^{\dagger\dagger}) = \mu^\dagger, \\
\mathrm{Var}(y^\dagger|y \in (0,1)) &= \mathrm{Var}(y^{\dagger\dagger}) = v^\dagger,
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}(y^*, y^\dagger|y \in (0,1)) = \mathrm{Cov}(y^\dagger, y^*|y \in (0,1)) = \mathrm{Cov}(y^{**}, y^{\dagger\dagger}) = \mathrm{Cov}(y^{\dagger\dagger}, y^{**}) = c^{*\dagger}, \\
\mathrm{Cov}(\mathbb{I}_c(y), y^*) = \mathrm{Cov}(\mathbb{I}_c(y), y^\dagger) = \mathrm{Cov}(y^*, \mathbb{I}_c(y)) = \mathrm{Cov}(y^\dagger, \mathbb{I}_c(y)) = 0.
\end{aligned}$$

Note que a estatística da parte degenerada é independente da parte contínua e os momentos da parte contínua apresentada aqui são os mesmos dados em (2.7).

2.5 Verossimilhança do modelo Beta inflacionado

Sejam $y_1, y_2, ..., y_n$ uma amostra de variáveis aleatórias independentes e identicamente distribuídas $(i.i.d)$ com $y_t \sim bi_c(\alpha, \mu, \phi)$ para todo $t = 1, \ldots, n$. A função de verossimilhança da expressão (2.13) é dada por

$$L(\alpha, \mu, \phi; \mathbf{y}) = \prod_{t=1}^{n} bi_c(y_t; \alpha, \mu, \phi) = L_1(\alpha) L_2(\mu, \phi). \tag{2.15}$$

A função de verossimilhança é fatorada em dois termos independentes, onde $L_1(\alpha)$ e $L_2(\mu, \phi)$ é como

$$\begin{aligned}
L_1(\alpha) &= \prod_{t=1}^{n} \alpha^{\mathbb{I}_{(c)}(y_t)} (1-\alpha)^{1-\mathbb{I}_{(c)}(y_t)}, \\
L_2(\mu, \phi) &= \prod_{t=1}^{n} f(y_t; \mu, \phi)^{1-\mathbb{I}_{(c)}(y_t)} = \prod_{t: y_t \in (0,1)} f(y_t; \mu, \phi),
\end{aligned}$$

em que $L_1(\alpha)$ e $L_2(\mu, \phi)$ dependem apenas dos parâmetros α e (μ, ϕ), respectivamente.

A log-verossimilhança é dada como

$$l(\alpha, \mu, \phi; \mathbf{y}) = log(L(\alpha, \mu, \phi; \mathbf{y})) = \sum_{t=1}^{n} l_{1t}(\alpha) + \sum_{t:y_t \in (0,1)} l_{2t}(\mu, \phi),$$

em que

$$l_{1t}(\alpha) = \mathbb{I}_{(c)}(y_t) log(\alpha) + (1 - \mathbb{I}_{(c)}(y_t)) log(1 - \alpha),$$
$$l_{2t}(\mu, \phi) = log\Big(\frac{\Gamma(\phi)}{\Gamma(\phi\mu)\Gamma(\phi(1-\mu))}\Big) + (\mu\phi - 1)y_t^* + (\phi - 2)y_t^\dagger.$$

Derivando a log-verossimilhança em relação a (α, μ, ϕ), temos a seguinte função escore

$$U_\alpha(\alpha) = \frac{\partial l_1(\alpha)}{\partial \alpha} = \sum_{t=1}^{n} \Big(\frac{\mathbb{I}_{(c)}(y_t)}{\alpha} - \frac{1 - \mathbb{I}_{(c)}(y_t)}{1 - \alpha} \Big),$$
$$U_\mu(\mu, \phi) = \frac{\partial l_2(\mu, \phi)}{\partial \mu} = \sum_{t:y_t \in (0,1)} -\phi\{\psi(\mu\phi) - \psi(\phi(1-\mu)) - y_t^*\},$$
$$U_\phi(\mu, \phi) = \frac{\partial l_2(\mu, \phi)}{\partial \phi} = \sum_{t:y_t \in (0,1)} \{\psi(\phi) - \mu\psi(\mu\phi) - (1-\mu)\psi(\phi(1-\mu)) + \mu y_t^* + y_t^\dagger\}.$$

Dadas os gradientes acima e a separabilidade na equação (2.15) é possível ver que a solução para $U_\alpha(\alpha) = 0$ é $\hat{\alpha} = \sum_{t=1}^{n} \mathbb{I}_{(c)}(y_t)/n$. Desta forma, o estimador de máxima verossimilhança (EMV) para α coincide com o de momentos dado em (2.14). Ademais, o estimador $\hat{\alpha}$ é enviesado e é função da estatística suficiente e completa (Casella e Berger, 2002). Contudo, a solução para $U_\mu(\mu, \phi) = 0$ e $U_\phi(\mu, \phi) = 0$ não são triviais, havendo a necessidade de usar algum método de iteração para uma solução numérica que maximize a log-verossimilhança $l_2(\mu, \phi)$. Um estimador para (μ, ϕ) também pode ser obtido através dos momentos condicionais de $E(y|y \in (0,1)) = \mu$ e $Var(y|y \in (0,1)) = \mu(1-\mu)/(\phi+1)$. Para mais detalhes consulte Ospina e Ferrari (2010).

A hessiana da log-verossimilhança em relação a (α, μ, ϕ) no ponto de massa c é dada por

$$k_{\alpha\alpha} = \frac{\partial^2 l_1(\alpha)}{\partial \alpha^2} = -n\left(\frac{\hat{\alpha}}{\alpha^2} + \frac{1-\hat{\alpha}}{(1-\alpha)^2}\right),$$

$$k_{\mu\mu} = \frac{\partial^2 l_2(\mu,\phi)}{\partial \mu^2} = -n(1-\hat{\alpha})\phi^2\{\psi'(\mu\phi) + \psi'(\phi(1-\mu))\},$$

$$k_{\phi\phi} = \frac{\partial^2 l_2(\mu,\phi)}{\partial \phi^2} = -n(1-\hat{\alpha})\{\mu^2\psi'(\mu\phi) + (1-\mu)^2\psi'(\phi(1-\mu)) - \psi'(\phi)\},$$

$$k_{\phi\mu} = k_{\mu\phi}^{\top} = \frac{\partial^2 l_2(\mu,\phi)}{\partial \phi \partial \mu} = \frac{\partial^2 l_2(\mu,\phi)}{\partial \mu \partial \phi} =$$

$$\sum_{t:y_t\in(0,1)} \{\psi(\phi(1-\mu)) - \psi(\phi\mu) + y_t^* - \phi\mu\psi'(\mu\phi) + \phi(1-\mu)\psi'(\phi(1-\mu))\},$$

$$k_{\alpha\mu} = k_{\alpha\phi} = k_{\mu\alpha}^{\top} = k_{\phi\alpha}^{\top} = 0.$$

O parâmetro α é ortogonal aos parâmetros (μ, ϕ), deste modo a matriz hessiana é dada por

$$K(\alpha,\mu,\phi) = \begin{pmatrix} k_{\alpha\alpha} & 0 & 0 \\ 0 & k_{\mu\mu} & k_{\mu\phi} \\ 0 & k_{\phi\mu} & k_{\phi\phi} \end{pmatrix},$$

e desta forma, para se obter a informação de Fisher, sob as condições de regularidade dadas em 2.4.1 , basta aplicar o negativo da esperança matemática, $-\mathbb{E}_\theta(.)$. Note que os únicos termos da hessiana a depender da esperança são $\hat{\alpha}$ e y_t^*. Logo, $\mathbb{E}_\alpha(\hat{\alpha}) = \alpha$ e $E(y^*|y \in (0,1)) = E(y^{**}) = \mu^*$ em que μ^* está definido em (2.7).

Consequentemente, a matriz de informação de Fisher de $\theta = (\alpha, \mu, \phi)$ é

$$-\mathbb{E}_\theta(K(\alpha,\mu,\phi)) = H(\alpha,\mu,\phi) = \begin{pmatrix} h_{\alpha\alpha} & 0 & 0 \\ 0 & h_{\mu\mu} & h_{\mu\phi} \\ 0 & h_{\phi\mu} & h_{\phi\phi} \end{pmatrix},$$

em que

$$h_{\alpha\alpha} = -\mathbb{E}_\theta\Big(\frac{\partial^2 l_1(\alpha)}{\partial\alpha^2}\Big) = n\frac{1}{\alpha(1-\alpha)},$$

$$h_{\mu\mu} = -\mathbb{E}_\theta\Big(\frac{\partial^2 l_2(\mu,\phi)}{\partial\mu^2}\Big) = n(1-\alpha)\phi^2\{\psi'(\mu\phi)+\psi'(\phi(1-\mu))\},$$

$$h_{\phi\phi} = -\mathbb{E}_\theta\Big(\frac{\partial^2 l_2(\mu,\phi)}{\partial\phi^2}\Big) = n(1-\alpha)\{\mu^2\psi'(\mu\phi)+(1-\mu)^2\psi'(\phi(1-\mu))-\psi'(\phi)\},$$

$$\begin{aligned} h_{\phi\mu} = h_{\mu\phi}^\top = -\mathbb{E}_\theta\Big(\frac{\partial^2 l_2(\mu,\phi)}{\partial\phi\partial\mu}\Big) &= -\mathbb{E}_\theta\Big(\frac{\partial^2 l_2(\mu,\phi)}{\partial\mu\partial\phi}\Big) \\ &= n(1-\alpha)\phi\{\mu\psi'(\mu\phi)-(1-\mu)\psi'(\phi(1-\mu))\}, \end{aligned}$$

$$h_{\alpha\mu} = h_{\alpha\phi} = h_{\mu\alpha}^\top = h_{\phi\alpha}^\top = 0.$$

Um ponto importante a se destacar é que a matriz de informação não depende de c, e a ortogonalidade dada nos vetores escores permite dizer que α e (μ,ϕ) são assintoticamente independentes. Além disso, a proposição dada em 2.4.1 possibilita obter a normalidade assintótica dos estimadores de máxima verossimilhança, sendo um facilitador para usuais inferências.

2.6 Modelo de regressão Beta inflacionada

Sejam $y_1, y_2, ..., y_n$ variáveis aleatórias independentes, com $y_t \sim bi_c(\alpha_t, \mu_t, \phi_t)$. Portanto, pode-se definir um preditor para α_t, μ_t e ϕ_t como

$$\begin{aligned} \eta_{1t} &= g_1(\alpha_t) = \sum_{i=1}^{p} x_{ti}\beta_i, \\ \eta_{2t} &= g_2(\mu_t) = \sum_{j=1}^{m} z_{tj}\gamma_j, \\ \eta_{3t} &= g_3(\phi_t) = \sum_{l=1}^{k} v_{tl}\delta_l, \end{aligned} \tag{2.16}$$

em que $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_p)^\top$, $\boldsymbol{\gamma} = (\gamma_1,\ldots,\gamma_m)^\top$ e $\boldsymbol{\delta} = (\delta_1,\ldots,\delta_k)^\top$ são vetores de parâmetros desconhecidos, com $\boldsymbol{\beta}\in\mathbb{R}^p$, $\boldsymbol{\gamma}\in\mathbb{R}^m$ e $\boldsymbol{\delta}\in\mathbb{R}^k$. As matrizes $\mathbf{X} = (\mathbf{x}_1^\top,\ldots,\mathbf{x}_p^\top)^\top$, $\mathbf{Z} = (\mathbf{z}_1^\top,\ldots,\mathbf{z}_m^\top)^\top$ e $\mathbf{V} = (\mathbf{v}_1^\top,\ldots,\mathbf{v}_k^\top)^\top$ são de covariáveis exógenas conhecidas, podendo compartilhar as mesmas covariáveis. Para garantir que o modelo seja identificado, isto é, que haja apenas um único conjunto de solução, temos que $p+m+k<n$. Ao qual $\eta_1 = (\eta_{11},\ldots,\eta_{1n})^\top$, $\eta_2 = (\eta_{21},\ldots,\eta_{2n})^\top$ e $\eta_3 = (\eta_{31},\ldots,\eta_{3n})^\top$ são preditores lineares, embora possam ser funções não-lineares (Ospina e

Ferrari, 2012).

As funções de ligação $\boldsymbol{g_1}: (0,1) \to \mathbb{R}$, $\boldsymbol{g_2}: (0,1) \to \mathbb{R}$ e $\boldsymbol{g_3}: (0,+\infty) \to \mathbb{R}$ são estritamente monótonas e têm derivada de segunda ordem. As formas mais usuais para $g_1(.)$ e $g_2(.)$ são: $g(\theta) = log\{\theta/(1-\theta)\}$ (função logito), $g(\theta) = \Phi^{-1}(\theta)$ (função probito) em que $\Phi(.)$ é a função de distribuição acumulada normal padrão, $g(\theta) = log\{-log(1-\theta)\}$ (função de ligação log-log complementar), $g(\theta) = log\{-log(\theta)\}$ (função de ligação log-log), entre outras (Collett, 2002; McCullagh e Nelder, 1983). Para $g_3(.)$ tem-se usualmente $g(\theta) = log(\theta)$ (função log) e $g(\theta) = \sqrt{\theta}$ (função raiz-quadrada).

A função de verossimilhança para o modelo com os parâmetros $\theta = (\beta^\top, \gamma^\top, \delta^\top)^\top$ pode ser definida por

$$\mathrm{L}(\theta) = \prod_{t=1}^{n} bi_c(y_t; \alpha_t, \mu_t, \phi_t) = \mathrm{L}_1(\boldsymbol{\beta})\, \mathrm{L}_2(\boldsymbol{\gamma}, \boldsymbol{\delta}), \tag{2.17}$$

em que

$$\mathrm{L}_1(\boldsymbol{\beta}) = \prod_{t=1}^{n} \alpha_t^{\mathbb{I}_{(c)}(y_t)} (1-\alpha_t)^{1-\mathbb{I}_{(c)}(y_t)},$$

$$\mathrm{L}_2(\boldsymbol{\gamma}, \boldsymbol{\delta}) = \prod_{t=1}^{n} f(y_t; \mu_t, \phi_t)^{1-\mathbb{I}_{(c)}(y_t)} = \prod_{t: y_t \in (0,1)} f(y_t; \mu_t, \phi_t),$$

nas quais $\alpha_t = g_1^{-1}(\eta_{1t})$, $\mu_t = g_2^{-1}(\eta_{2t})$ e $\phi_t = g_3^{-1}(\eta_{3t})$, como definido em (2.16), são funções de β, γ e δ, respectivamente. O termo $\mathrm{L}(\theta)$ é fatorado em dois, sendo um dependente somente de $\beta^\top$ e o outro apenas de $(\gamma^\top, \delta^\top)^\top$. O processo de inferência para os parâmetros θ se dá independente entre $\beta^\top$ e $(\gamma^\top, \delta^\top)^\top$ por causa da separabilidade da verossimilhança apresentada acima.

O logaritmo da função de verossimilhança para o modelo com $\theta = (\beta^\top, \gamma^\top, \delta^\top)^\top$ é dado pela forma

$$l(\theta; \mathbf{y}) = l_1(\boldsymbol{\beta}) + l_2(\boldsymbol{\gamma}, \boldsymbol{\delta}) = \sum_{t=1}^{n} l_{1t}(\alpha_t) + \sum_{t: y_t \in (0,1)} l_{2t}(\mu_t, \phi_t),$$

em que

$$l_{1t}(\alpha_t) = \mathbb{I}_{(c)}(y_t) log(\alpha_t) + (1 - \mathbb{I}_{(c)}(y_t)) log(1-\alpha_t),$$

$$l_{2t}(\mu_t, \phi_t) = log\left(\frac{\Gamma(\phi_t)}{\Gamma(\phi_t\mu_t)\Gamma(\phi_t(1-\mu_t))}\right) + (\mu_t\phi_t - 1)y_t^* + (\phi_t - 2)y_t^\dagger.$$

Desta forma, observa-se que $l_{1t}(\alpha_t)$ é a verossimilhança da distribuição Bernoulli. Portanto, $\mathbb{I}_{(c)}(y_t) \sim Bern(\alpha_t)$, com os momentos dados em (2.14).

A função escore com relação a $(\beta^\top, \gamma^\top, \delta^\top)^\top$ é dada por

$$U(\theta) = (U_\beta(\boldsymbol{\beta})^\top, U_\gamma(\boldsymbol{\gamma}, \boldsymbol{\delta})^\top, U_\delta(\boldsymbol{\gamma}, \boldsymbol{\delta})^\top)^\top,$$

em que

$$\begin{aligned}
U_{\beta_i}(\boldsymbol{\beta}) &= \frac{\partial l_1(\boldsymbol{\beta})}{\partial \beta_i} = \sum_{t=1}^{n} \frac{\partial l_{1t}(\alpha_t)}{\partial \alpha_t} \frac{d\alpha_t}{d\eta_{1t}} \frac{\partial \eta_{1t}}{\partial \beta_i}, \\
U_{\gamma_j}(\boldsymbol{\gamma}, \boldsymbol{\delta}) &= \frac{\partial l_2(\boldsymbol{\gamma}, \boldsymbol{\delta})}{\partial \gamma_j} = \sum_{t:y_t \in (0,1)} \frac{\partial l_{2t}(\mu_t, \phi_t)}{\partial \mu_t} \frac{d\mu_t}{d\eta_{2t}} \frac{\partial \eta_{2t}}{\partial \gamma_j}, \\
U_{\delta_l}(\boldsymbol{\gamma}, \boldsymbol{\delta}) &= \frac{\partial l_2(\boldsymbol{\gamma}, \boldsymbol{\delta})}{\partial \delta_l} = \sum_{t:y_t \in (0,1)} \frac{\partial l_{2t}(\mu_t, \phi_t)}{\partial \phi_t} \frac{d\phi_t}{d\eta_{3t}} \frac{\partial \eta_{3t}}{\partial \delta_l},
\end{aligned} \tag{2.18}$$

sendo

$$\begin{aligned}
\frac{d\alpha_t}{d\eta_{1t}} &= \frac{dg_1^{-1}(\eta_{1t})}{d\eta_{1t}} = \frac{1}{g_1'(\alpha_t)} & & \frac{\partial \eta_{1t}}{\partial \beta_i} = x_{ti} \text{ ou } \frac{\partial \eta_1}{\partial \boldsymbol{\beta}} = \mathbf{X} \\
\frac{d\mu_t}{d\eta_{2t}} &= \frac{dg_2^{-1}(\eta_{2t})}{d\eta_{2t}} = \frac{1}{g_2'(\mu_t)} & \text{e} & \frac{\partial \eta_{2t}}{\partial \gamma_j} = z_{tj} \text{ ou } \frac{\partial \eta_2}{\partial \boldsymbol{\gamma}} = \mathbf{Z} \\
\frac{d\phi_t}{d\eta_{3t}} &= \frac{dg_3^{-1}(\eta_{3t})}{d\eta_{3t}} = \frac{1}{g_3'(\phi_t)} & & \frac{\partial \eta_{3t}}{\partial \delta_l} = v_{tl} \text{ ou } \frac{\partial \eta_3}{\partial \boldsymbol{\delta}} = \mathbf{V}
\end{aligned}$$

$\mathbf{X}$ $(n \times p)$, $\mathbf{Z}$ $(n \times m)$ e $\mathbf{V}$ $(n \times k)$ são matrizes em que a t-ésima linha e (i, j, l)-ésima coluna são x_{ti}, z_{tj} e v_{tl}, respectivamente, i $(i') = 1, \ldots, p$, j $(j') = 1, \ldots, m$ e l $(l') = 1, \ldots, k$ com respeito a $\boldsymbol{\beta} \in \mathbb{R}^p$, $\boldsymbol{\gamma} \in \mathbb{R}^m$ e $\boldsymbol{\delta} \in \mathbb{R}^k$. Isso se dá porque se assume que o preditor é linear.

Portanto, a função escore dada em (2.18) é exemplificada como

$$\begin{aligned}
U_{\beta_i}(\boldsymbol{\beta}) &= \frac{\partial l_1(\boldsymbol{\beta})}{\partial \beta_i} = \sum_{t=1}^{n} \frac{\mathbb{I}_{(c)}(y_t) - \alpha_t}{\alpha_t(1-\alpha_t)} \frac{1}{g_1'(\alpha_t)} x_{ti}, \\
U_{\gamma_j}(\boldsymbol{\gamma}, \boldsymbol{\delta}) &= \frac{\partial l_2(\boldsymbol{\gamma}, \boldsymbol{\delta})}{\partial \gamma_j} = \sum_{t:y_t \in (0,1)} \phi_t(y_t^* - \mu_t^*) \frac{1}{g_2'(\mu_t)} z_{tj}, \\
U_{\delta_l}(\boldsymbol{\gamma}, \boldsymbol{\delta}) &= \frac{\partial l_2(\boldsymbol{\gamma}, \boldsymbol{\delta})}{\partial \delta_l} = \sum_{t:y_t \in (0,1)} \left[\mu_t(y_t^* - \mu_t^*) + (y_t^\dagger - \mu_t^\dagger)\right] \frac{1}{g_3'(\phi_t)} v_{tl}.
\end{aligned} \tag{2.19}$$

As estatísticas $(y_t^*, y_t^\dagger)$ são suficientes e completas com média $(\mu_t^*, \mu_t^\dagger)$, respectivamente, denotadas em (2.14). Ospina e Ferrari (2012) apresentaram a seguinte forma matricial para a função escore, definida como

$$U_\beta(\boldsymbol{\beta}) = \mathbf{X}^\top \mathcal{A} D \mathcal{A}^*(y^c - \alpha),$$

$$U_\gamma(\boldsymbol{\gamma}, \boldsymbol{\delta}) = \mathbf{Z}^\top (I_n - Y^c) T \Phi (y^* - \mu^*),$$

$$U_\delta(\boldsymbol{\gamma}, \boldsymbol{\delta}) = \mathbf{V}^\top (I_n - Y^c) H \left[\mathcal{M}(y^* - \mu^*) + (y^\dagger - \mu^\dagger) \right],$$

a qual $y^c = (\mathbb{I}_{(c)}(y_1), \ldots, \mathbb{I}_{(c)}(y_n))^\top$, $\alpha = (\alpha_1, \ldots, \alpha_n)^\top$, $y^* = (y_1^*, \ldots, y_n^*)^\top$, $\mu^* = (\mu_1^*, \ldots, \mu_n^*)^\top$, $y^\dagger = (y_1^\dagger, \ldots, y_n^\dagger)^\top$ e $\mu^\dagger = (\mu_1^\dagger, \ldots, \mu_n^\dagger)^\top$ são vetores de dimensão $(n \times 1)$ e $\mathcal{A} = \text{diag}(1/\alpha_1, \ldots, 1/\alpha_n)$, $\mathcal{A}^* = \text{diag}(1/(1-\alpha_1), \ldots, 1/(1-\alpha_n))$, $D = \text{diag}(1/g_1'(\alpha_1), \ldots, 1/g_1'(\alpha_n))$, $T = \text{diag}(1/g_2'(\mu_1), \ldots, 1/g_2'(\mu_n))$, $H = \text{diag}(1/g_3'(\phi_1), \ldots, 1/g_3'(\phi_n))$, $\Phi = \text{diag}(\phi_1, \ldots, \phi_n)$, $\mathcal{M} = \text{diag}(\mu_1, \ldots, \mu_n)$, $Y^c = \text{diag}(\mathbb{I}_{(c)}(y_1), \ldots, \mathbb{I}_{(c)}(y_n))$ e $I_n = \text{diag}(1, \ldots, 1)$ são matrizes de dimensão $(n \times n)$.

Ao calcular a segunda derivada da log-verossimilhança $l(\theta; \mathbf{y})$, pode-se obter a matriz de informação observada e, posteriormente, a matriz de informação de Fisher. Portanto, a matriz de informação observada é dada como

$$\begin{aligned}
J_{ii'} = &- \sum_{t=1}^{n} \left\{ \left(\frac{-\mathbb{I}_{(c)}(y_t)}{\alpha_t^2} - \frac{1 - \mathbb{I}_{(c)}(y_t)}{(1-\alpha_t)^2} \right) \left[\frac{1}{g_1'(\alpha_t)} \right] \right. \\
&+ \left. \left(\frac{\mathbb{I}_{(c)}(y_t) - \alpha_t}{\alpha_t(1-\alpha_t)} \right) \left[\frac{-g_1''(\alpha_t)}{(g_1'(\alpha_t))^2} \right] \right\} \left[\frac{1}{g_1'(\alpha_t)} \right] \frac{\partial \eta_{1t}}{\partial \beta_i} \frac{\partial \eta_{1t}}{\partial \beta_{i'}} \\
&- \sum_{t=1}^{n} \left\{ \left(\frac{\mathbb{I}_{(c)}(y_t) - \alpha_t}{\alpha_t(1-\alpha_t)} \right) \left[\frac{1}{g_1'(\alpha_t)} \right] \frac{\partial^2 \eta_{1t}}{\partial \beta_i \partial \beta_{i'}} \right\},
\end{aligned}$$

$$\begin{aligned}
J_{jj'} = &- \sum_{t:y_t \in (0,1)} \left\{ -\phi_t^2 v_t^* \left[\frac{1}{g_2'(\mu_t)} \right] + \phi_t (y_t^* - \mu_t^*) \left[\frac{-g_2''(\mu_t)}{(g_2'(\mu_t))^2} \right] \right\} \left[\frac{1}{g_2'(\mu_t)} \right] \frac{\partial \eta_{2t}}{\partial \gamma_j} \frac{\partial \eta_{2t}}{\partial \gamma_{j'}} \\
&- \sum_{t:y_t \in (0,1)} \left\{ \phi_t (y^* - \mu^*) \left[\frac{1}{g_2'(\mu_t)} \right] \frac{\partial^2 \eta_{2t}}{\partial \gamma_j \gamma_{j'}} \right\},
\end{aligned}$$

$$\begin{aligned}
J_{ll'} = &- \sum_{t:y_t \in (0,1)} \left\{ \left(-\mu_t^2 v_t^* - 2\mu_t c_t^{*\dagger} - v_t^\dagger \right) \left[\frac{1}{g_3'(\phi_t)} \right] \right. \\
&+ \left. \left[\mu_t (y_t^* - \mu_t^*) + (y_t^\dagger - \mu_t^\dagger) \right] \left[\frac{-g_3''(\phi_t)}{(g_3'(\phi_t))^2} \right] \right\} \left[\frac{1}{g_3'(\phi_t)} \right] \frac{\partial \eta_{3t}}{\partial \delta_l} \frac{\partial \eta_{3t}}{\partial \delta_{l'}}
\end{aligned}$$

$$- \sum_{t:y_t\in(0,1)} \left\{ [\mu_t(y_t^* - \mu_t^*) + (y_t^\dagger - \mu_t^\dagger)] \left[\frac{1}{g_3'(\phi_t)}\right] \frac{\partial^2 \eta_{3t}}{\partial\delta_l \partial\delta_{l'}} \right\},$$

$$J_{jl} = J_{lj}^\top = - \sum_{t:y_t\in(0,1)} [(y_t^* - \mu_t^*) - \phi_t(\mu_t v_t^* + c_t^{*\dagger})] \left[\frac{1}{g_2'(\mu_t)}\right] \frac{\partial \eta_{2t}}{\partial \gamma_j} \left[\frac{1}{g_3'(\phi_t)}\right] \frac{\partial \eta_{3t}}{\partial \delta_l},$$

$$J_{ij} = J_{il} = J_{ji}^\top = J_{li}^\top = 0,$$

para a forma matricial da matriz observada, consulte Ospina e Ferrari (2012).

A matriz de informação de Fisher é simplesmente $I_F(\theta) = \mathrm{E}(J(\theta))$. Por conseguinte, pode ser escrita como

$$I_F(\theta) = \begin{pmatrix} I_{\beta\beta} & 0 & 0 \\ 0 & I_{\gamma\gamma} & I_{\gamma\delta} \\ 0 & I_{\delta\gamma} & I_{\delta\delta} \end{pmatrix}, \tag{2.20}$$

em que

$$I_{\beta\beta} = \sum_{t=1}^{n} \left(\frac{1}{\alpha_t(1-\alpha_t)}\right) \left[\frac{1}{g_1'(\alpha_t)}\right]^2 \frac{\partial \eta_{1t}}{\partial \beta_i} \frac{\partial \eta_{1t}}{\partial \beta_{i'}},$$

$$I_{\gamma\gamma} = \sum_{t:y_t\in(0,1)} \phi_t^2 v_t^* \left[\frac{1}{g_2'(\mu_t)}\right]^2 \frac{\partial \eta_{2t}}{\partial \gamma_j} \frac{\partial \eta_{2t}}{\partial \gamma_{j'}},$$

$$I_{\delta\delta} = \sum_{t:y_t\in(0,1)} (\mu_t^2 v_t^* + 2\mu_t c_t^{*\dagger} + v_t^\dagger) \left[\frac{1}{g_3'(\phi_t)}\right]^2 \frac{\partial \eta_{3t}}{\partial \delta_l} \frac{\partial \eta_{3t}}{\partial \delta_{l'}},$$

$$I_{\gamma\delta} = I_{\delta\gamma}^\top = \sum_{t:y_t\in(0,1)} [\phi_t(\mu_t v_t^* + c_t^{*\dagger})] \left[\frac{1}{g_2'(\mu_t)}\right] \frac{\partial \eta_{2t}}{\partial \gamma_j} \left[\frac{1}{g_3'(\phi_t)}\right] \frac{\partial \eta_{3t}}{\partial \delta_l},$$

$$I_{\beta\gamma} = I_{\beta\delta} = I_{\gamma\beta}^\top = I_{\delta\beta}^\top = 0.$$

Com notação matricial, tem-se

$$I_{\beta\beta} = \mathbf{X}^\top \mathbf{W_1} \mathbf{X},$$

$$I_{\gamma\gamma} = \mathbf{Z}^\top \mathbf{W_2} \mathbf{Z},$$

$$I_{\delta\delta} = \mathbf{V}^\top \mathbf{W_3} \mathbf{V},$$

$$I_{\gamma\delta} = I_{\delta\gamma}^\top = \mathbf{Z}^\top \mathbf{W_4} \mathbf{V},$$

$$I_{\beta\gamma} = I_{\beta\delta} = I_{\gamma\beta}^{\top} = I_{\delta\beta}^{\top} = 0,$$

em que $\mathbf{W_1} = (\mathcal{A}^* + \mathcal{A})D^2$, $\mathbf{W_2} = \Phi T\{\mathcal{V}^*\mathcal{A}^{*-1}\}T\Phi$, $\mathbf{W_3} = H\{(\mathcal{M}^2\mathcal{V}^* + 2\mathcal{M}\mathcal{C} + \mathcal{V}^{\dagger})\mathcal{A}^{*-1}\}H$ e $\mathbf{W_4} = T\{\Phi(\mathcal{M}\mathcal{V}^* + \mathcal{C})\mathcal{A}^{*-1}\}H$ onde $\mathcal{V}^* = \text{diag}(v_1^*, \ldots, v_n^*)$ e $\mathcal{C} = \text{diag}(c_1^{*\dagger}, \ldots, c_n^{*\dagger})$.

Como de hábito, a estimativa por máxima verossimilhança dá-se por métodos numéricos. Para mais detalhes veja Ospina e Ferrari (2012).

2.7 Regressão Beta inflacionada em um para os escores da eficiência

Dado um vetor de eficiências, medidas por DEA, θ, assuma que $\theta_t \sim BIU(\alpha_t, \mu_t, \phi_t)$. O modelo de regressão para a eficiência será dado da seguinte forma

$$\text{E}(\theta_t) = \alpha_t + (1 - \alpha_t)\mu_t, \tag{2.21}$$

e a variância para θ_t é

$$\text{Var}(\theta_t) = (1 - \alpha_t)\frac{\mu_t(1 - \mu_t)}{1 - \phi_t} + \alpha_t(1 - \alpha_t)(1 - \mu_t)^2, \tag{2.22}$$

na qual

$$\alpha_t = \Phi(\eta_{1t}),$$

$$\mu_t = \frac{1}{1+\exp(-\eta_{2t})},$$

$$\phi_t = \exp(\eta_{3t}),$$

em que η_{1t}, η_{2t} e η_{3t} são preditores lineares, os mesmos dados em (2.16). Note que (η_{1t}, α_t), (η_{2t}, μ_t) e (η_{3t}, ϕ_t) estão ligados pela função probito, logística e log, respectivamente; $\Phi(.)$ é a função de distribuição normal padrão. O processo de inferência para a estimativa dos parâmetros do modelo (2.21) é o mesmo dado na Seção 2.6.

As covariáveis que estão no modelo de regressão Beta inflacionada são exógenas em relação aos resíduos. No capitulo seguinte, discute-se a hipótese de endogeneidade no contexto de uma aplicação com dados do censo agropecuário.

Capítulo 3

Endogeneidade

Estudos das mais diversas áreas estão interessados em uma relação de causa e efeito. A aplicação dos modelos estatísticos é primordial para descrever tais fenômenos. Entretanto, ao estimar o modelo estatístico, são necessárias algumas condições. Uma das exigências é que as covariáveis do modelo em estudo sejam exógenas. A suposição de exogeneidade dada em Greene (2012, pg. 223) é que $\mathrm{E}(\epsilon_t|x_{tk}) = 0,\ \forall\ t = 1, \ldots, n$ e $k = 1, \ldots, p$. Portanto,

$$\begin{aligned}
\mathrm{E}(\epsilon_t x_{tk}) &= \mathrm{E}(\mathrm{E}(\epsilon_t x_{tk}|x_{tk})),\\
\mathrm{E}(\epsilon_t x_{tk}) &= \mathrm{E}(x_{tk}\,\mathrm{E}(\epsilon_t|x_{tk})),\\
\mathrm{E}(\epsilon_t x_{tk}) &= 0.
\end{aligned}$$

Note que a condição 'estritamente exógena' ocorre quando x_{tk} é ortogonal a ϵ_t sendo este o erro experimental, seja para o modelo linear ou não-linear. Logo, a definição de endogeneidade pode ser escrita como

$$\mathrm{E}(\epsilon_t|x_{tk}) \neq 0,$$

em que o conceito de endogeneidade é que as variáveis preditoras podem estar correlacionadas com o erro do modelo.

3.1 Método dos momentos generalizado

Atribui-se o método dos momentos generalizado (GMM, em inglês) a Hansen (1982), embora outros autores tenham se dedicado a essa temática. Nesta abordagem, a estimação dos parâmetros é baseada na generalização dos momentos populacionais.

Seja $\rho_0 = (\rho_1, \ldots, \rho_s)^\top \in \Theta \subseteq \mathbb{R}^s$ o verdadeiro vetor de parâmetros a serem estimados, $\mathbf{X}$ é a matriz de covariáveis com a t-ésima linha $\mathbf{x}_t = (x_{t1}, \ldots, x_{tp}) \in \mathbb{R}^p$ e $f(.) = (f_1(.), \ldots, f_q(.))^\top \in \mathbb{R}^q$, um vetor de funções de momentos. Os momentos populacionais são tomados da seguinte maneira

$$\mathrm{E}(f(\mathbf{x}_t, \rho_0)) = 0, \tag{3.1}$$

em que uma condição de ortogonalidade indica que um conjunto de momentos da população é igual a zero. As funções de momentos $f(.)$ serão condições de momentos do modelo linear ou não-linear.

Se o número de condições de momentos q for igual ao número de parâmetros s, ou seja, $q = s$, diz-se que o modelo é identificado e o estimador para ρ é

$$n^{-1} \sum_{t=1}^{n} f(\mathbf{x}_t, \hat{\rho}) = 0. \tag{3.2}$$

Tal estimador é chamado de método dos momentos (MM), ao qual é equivalente a minimizar

$$\mathbf{Q}_n(\rho) = \left(n^{-1} \sum_{t=1}^{n} f(\mathbf{x}_t, \rho)^\top \right) \left(n^{-1} \sum_{t=1}^{n} f(\mathbf{x}_t, \rho) \right) \tag{3.3}$$

tendo-se uma única solução trivial para ρ.

O estimador $\hat{\rho}_{MM}$ para ρ_0 é consistente e assintoticamente normal desde que as seguintes condições sejam satisfeitas (Hamilton, 1995, pg. 414):

Teorema 1. Sob as seguintes condições para os momentos em (3.2):

(i) $n^{-1} \sum_{t=1}^{n} f(\mathbf{x}_t, \rho_0) \xrightarrow{p} \mathrm{E}(f(\mathbf{x}_t, \rho_0)) = 0,$

(ii) $n^{-1} \sum_{t=1}^{n} \frac{\partial f(\mathbf{x}_t, \rho_0)}{\partial \rho_0} \xrightarrow{p} A_0,$

(iii) $n^{-1/2} \sum_{t=1}^{n} f(\mathbf{x}_t, \rho_0) \xrightarrow{d} N(0, S(\rho_0)),$

tem-se:

1. Sob a condição (i), $\hat{\rho}_{MM}$ é um estimador consistente, ou seja: $\hat{\rho}_{MM} \xrightarrow{p} \rho_0$,

2. Sob as condições (ii) e (iv), $\hat{\rho}_{MM}$ é um estimador assintoticamente normal, isso é:
   $n^{1/2}(\hat{\rho}_{MM} - \rho_0) \xrightarrow{d} N(0, A_0^{-1} S(\rho_0)(A_0^{\top})^{-1})$,

na qual A_0 é uma matriz finita de ordem $q \times s$ com $\dim(A_0) = s$ e $S(\rho_0)_{q\times q}$ é uma matriz de covariância assintótica dos momentos.

Ao considerar que a quantidade de momentos é igual a de parâmetros, já evidencia que o modelo em estudo é identificado e tem apenas um único conjunto-solução. Contudo, há casos em $q > s$. Nesta situação temos uma superidentificação e o estimador (3.2) não tem uma solução trivial. Desta forma, pode-se ter uma generalização do estimador MM, passando a ser chamado de método dos momentos generalizado, o qual considera uma forma quadrática para aproximar $n^{-1} \sum_{t=1}^{n} f(\mathbf{x}_t, \hat{\rho})$ a zero.

O estimador GMM, do mesmo modo que o estimador MM, é baseado nas condições de momentos populacionais, as quais minimizam o verdadeiro valor do vetor de parâmetros ρ_0. Portanto, a função objetivo que minimiza a estimativa $\hat{\rho}_{GMM}$ é dada por

$$\mathbf{Q}_n(\rho) = \Big(n^{-1} \sum_{t=1}^{n} f(\mathbf{x}_{t1}, \ldots, \mathbf{x}_{tp}, \rho)^{\top}\Big) \mathbf{W}_n \Big(n^{-1} \sum_{t=1}^{n} f(\mathbf{x}_{t1}, \ldots, \mathbf{x}_{tp}, \rho)\Big), \tag{3.4}$$

em que $\mathbf{W}_n$ é uma matriz de dimensão $q \times q$ não-negativa definida, que depende somente das covariáveis e converge para uma matriz finita e de posto completo quando $n \to \infty$. $\hat{\rho}_{GMM}$ é um estimador consistente e assintoticamente normal. Como dado por Hamilton (1995, pg. 414), tais suposições para $\hat{\rho}_{GMM}$ são válidas a exemplo dos seguintes pressupostos:

Teorema 2. Sob as seguintes condições para os momentos em (3.4):

(i) $n^{-1} \sum_{t=1}^{n} f(\mathbf{x}_t, \rho_0) \xrightarrow{p} \mathrm{E}(f(\mathbf{x}_t, \rho_0)) = 0$,

(ii) $\mathbf{Q}_n(\rho_0) \xrightarrow{p} 0$,

(iii) $n^{-1} \sum_{t=1}^{n} \frac{\partial f(\mathbf{x}_t, \rho_0)}{\partial \rho_0} \xrightarrow{p} A_0$,

(iv) $n^{-1/2} \sum_{t=1}^{n} f(\mathbf{x}_t, \rho_0) \xrightarrow{d} N(0, S(\rho_0))$,

(v) $\mathbf{W}_n \xrightarrow{p} \mathbf{W}_0$,

tem-se:

1. Sob as condições (i) e (ii), $\hat{\rho}$ é um estimador consistente, ou seja: $\hat{\rho} \xrightarrow{p} \rho_0$,

2. Sob as condições (iii), (iv) e (v), $\hat{\rho}$ é um estimador assintoticamente normal, isso é:
$n^{1/2}(\hat{\rho} - \rho_0) \xrightarrow{d} N(0, (A_0^\top W_0 A_0)^{-1}(A_0^\top W_0 S(\rho_0) W_0 A_0)(A_0^\top W_0 A_0)^{-1})$,

as matrizes A_0 e $S(\rho_0)$ são as mesmas dadas para o estimador MM.

A matriz W_n é uma matriz de pesos para a estimativa de $\hat{\rho}_{GMM}$. Então, conseguir uma matriz ótima para W_n é essencial para se obter um estimador mais eficiente. Porém, quando $q > s$ uma matriz de peso ótima para W_n é a $S(\rho)^{-1}$, sendo esse um estimador ótimo GMM. Logo, a distribuição assintótica do estimador passa a ser da seguinte maneira

$$n^{1/2}(\hat{\rho}_{GMM} - \rho_0) \xrightarrow{d} N(0, (A_0^\top S(\rho_0)^{-1} A_0)^{-1}), \tag{3.5}$$

resultado esse dado por Hansen (1982). O estimador ótimo GMM é dado por

$$\mathbf{Q}_n(\rho) = \left(n^{-1}\sum_{t=1}^{n} f(\mathbf{x}_{t1}, \ldots, \mathbf{x}_{tp}, \rho)^\top\right) S(\rho)^{-1} \left(n^{-1}\sum_{t=1}^{n} f(\mathbf{x}_{t1}, \ldots, \mathbf{x}_{tp}, \rho)\right). \tag{3.6}$$

Hansen, Heaton e Yaron (1996) apresentaram três estimadores distintos para estimar o estimador ótimo GMM para (3.6). Tais estimadores são chamados de estimador de dois-estágios, iterativo e atualização continuada, em que as suposições dadas no Teorema (2) e a distribuição assintótica do estimador (3.5) são válidas para todos.

O estimador de dois-estágios GMM dar-se-á partir da substituição de $S(\rho)^{-1}$ por I_n, a matriz identidade, pois $S(\rho)^{-1}$ só dependerá das covariáveis, sendo esta a matriz inicial para a estimativa no primeiro estágio. Em um segundo estágio, $S(\hat{\rho})^{-1}$, estimado neste primeiro estágio, passará a ser a matriz de peso para estimar ρ_{GMM}. O estimador iterativo é uma consequência do estimador de dois-estágios. Porém, $S(\hat{\rho})_j^{-1}$ passa a ser estimado com base em cada novo vetor de parâmetros estimados no processo iterativo j. O processo termina quando $\hat{\rho}_j$ e $\hat{\rho}_{j+1}$ passarem a não ser mais diferentes baseando-se em uma tolerância definida. Para o estimador de atualização continuada, considere desde o inicio que $S(\rho)^{-1}$ depende dos parâmetros ρ. Portanto, ambos passam a ser estimados simultaneamente.

Embora se discuta um estimador ótimo GMM, não foi determinado qual seria o melhor estimador para $S(\rho)$. Neste tocante, se os momentos são independentes, um estimador consistente para $S(\rho)$ é

$$S(\rho) = n^{-1} \sum_{t=1}^{n} f(\mathbf{x}_t, \rho) f(\mathbf{x}_t, \rho)^\top, \tag{3.7}$$

sendo este um estimador natural e robusto para a heterocedasticidade, por construção.

Em concordância às suposições de um estimador consistente e assintoticamente normal, pode-se ter a performance do teste de superidentificação, conhecido como teste J. O teste vem dado por

$$J = n\Big(n^{-1} \sum_{t=1}^{n} f(\mathbf{x}_{t1}, \ldots, \mathbf{x}_{tp}, \hat{\rho}_{GMM})^\top\Big) S(\hat{\rho})^{-1} \Big(n^{-1} \sum_{t=1}^{n} f(\mathbf{x}_{t1}, \ldots, \mathbf{x}_{tp}, \hat{\rho}_{GMM})\Big).$$

O teste J é calculado sob $H_0 = \mathrm{E}(f(\mathbf{x}_t, \rho)) = 0$. A estatística J tem distribuição qui-quadrado com $p = q - s$ graus de liberdade. Desta forma, o teste é apenas para o caso de superidentificação, pois quando é identificado a estatística do teste é igual a zero. Salienta-se que $\hat{\rho}_{GMM}$ e $\hat{S}(\rho)$ são as estimativas baseadas na escolha do processo de estimação de dois-estágios, iterativo ou atualização continuada.

3.1.1 Variáveis instrumentais

O problema em ter endogeneidade é que os estimadores dos parâmetros desconhecidos produzem estimativas não consistentes para quaisquer método de inferência. Na prática é necessário um processo que possa dar estimativas consistentes. A inconsistência dos estimadores pode ser corrigida pela abordagem GMM baseando-se em variáveis instrumentais (VI) (Wooldridge, 2010). Neste trabalho essa abordagem passará a ser chamado de estimador VI.

Considere que existe um conjunto de variáveis instrumentais $\mathbf{D}_{(n\times q)}$ que satisfaça a condição

$$E(\epsilon_t d_{tj}) = 0, \tag{3.8}$$

no qual $j = 1, \ldots, q$. A condição dada em (3.8) é a mesma de (3.1), em que a condição de ortogonalidade dos momentos populacionais é igual a zero.

Portanto, o erro experimental é expresso da seguinte maneira

$$\epsilon_t = y_t - \mathrm{E}(y_t|\mathbf{x_t}, \boldsymbol{\rho}), \tag{3.9}$$

no qual assume-se que o erro é aditivo, independente da forma de $\mathrm{E}(y_t|\mathbf{x_t}, \boldsymbol{\rho})$, sendo $\mathbf{x_t}$ composto por variáveis exógenas e endógenas. Note que $\mathbf{D}_{(n\times q)}$ é um conjunto de variáveis exógenas em $\mathbf{x_t}$ e instrumentos novos.

Neste caso, as condições de momentos estão vinculadas à dimensão da matriz de variáveis instrumentais, sendo $\dim(\mathbf{D}) = q$. Ademais, não obstante ao que foi descrito, note que quando $q = s$, sendo s a ordem do vetor de $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_s)$, o modelo é identificado e seu estimador VI é tido minimizando a seguinte expressão

$$\mathbf{Q}_n(\rho) = \Big(n^{-1}\boldsymbol{\epsilon}^\top \mathbf{D}\Big)\Big(n^{-1}\mathbf{D}^\top \boldsymbol{\epsilon}\Big). \tag{3.10}$$

Quando $q > s$, ou seja, havendo uma superidentificação, o estimador VI é obtido minimizando

$$\mathbf{Q}_n = \Big(n^{-1}\boldsymbol{\epsilon}^\top \mathbf{D}\Big)\mathbf{W}_n\Big(n^{-1}\mathbf{D}^\top \boldsymbol{\epsilon}\Big), \tag{3.11}$$

valendo-se das mesmas propriedades da seção 3.1. Para mais consulte Davidson e MacKinnon (1995, cap. 17), Cameron e Trivedi (2005, cap. 6) e Greene (2012, cap. 13).

3.2 Máxima verossimilhança de dois-estágios

Seja $y_t \sim f_1(y_t|\mu_{1t})$, em que $f_1(y_t|\mu_{1t})$ é uma distribuição de probabilidade qualquer, definida em um espaço de probabilidade, para a qual podem-se definir as seguintes médias

$$\begin{aligned} \mu_{\mathbf{1}} &= E(\mathbf{y}|\mathbf{X}_{(\mathbf{1})}, \rho_1, \mu_{\mathbf{0}}), \\ \mu_{\mathbf{0}} &= E(\mathbf{x}_{(\mathbf{2})}|\mathbf{D}, \rho_0), \end{aligned} \tag{3.12}$$

em que $\mathbf{x}_{(\mathbf{2})} \sim f_0(\mathbf{x}_{(\mathbf{2})}|\mu_{\mathbf{0}})$ é uma variável endógena; $\mathbf{X}_{(\mathbf{1})}$ é a matriz de variáveis exógenas; $\mathbf{D} = \begin{bmatrix} \mathbf{X}_{(\mathbf{1})} & \mathbf{K} \end{bmatrix}$ é uma matriz de variáveis instrumentais, em que a matriz $\mathbf{K}$ é a matrix de instrumentos novos; ρ_0 e ρ_1 são vetores de parâmetros do modelo e $\mathbf{y}$ é o vetor de variável resposta (Hardin, 2002). O modelo dado em (3.12) pode ser estimado em dois estágios por máxima verossimilhança. Portanto, as funções de log-verossimilhança para $f_0(.)$ e $f_1(.)$ são dadas por $l_0(\rho_0) = \sum_{t=1}^n l_{0t}(\rho_0) = \sum_{t=1}^n \log f_0(\mathbf{x}_{(\mathbf{2})\mathbf{t}}|\mu_{\mathbf{0t}})$ e $l_1(\rho_1, \hat{\rho_0}) = \sum_{t=1}^n l_{1t}(\rho_1, \hat{\rho_0}) = \sum_{t=1}^n \log f_1(y_t|\mu_{1t})$, respectivamente.

Murphy e Topel (1985) enunciaram que, sob as condições de regularidade, identificabilidade e que ρ_0^* e ρ_1^* sejam os verdadeiros parâmetros, a máxima verossimilhança do primeiro estágio para ρ_0 é um estimador consistente para μ_0. Consequentemente, maximizar $n^{-1}\sum_{t=1}^{n} l_{1t}(\rho_1, \hat{\rho_0})$ com respeito a ρ_1 é assintoticamente o mesmo que maximizar $n^{-1}\sum_{t=1}^{n} l_{1t}(\rho_1, \rho_0^*)$. Neste caso, o estimador de máxima verossimilhança para o segundo estágio de ρ_1 é também um estimador consistente.

Ao aplicar o teorema do limite central, tem-se que

$$\left.\begin{array}{l} \dfrac{1}{\sqrt{n}}\sum_{t=1}^{n} \dfrac{\partial l_0(\rho_0^*)}{\partial \rho_0} \\ \dfrac{1}{\sqrt{n}}\sum_{t=1}^{n} \dfrac{\partial l_{1t}(\rho_1^*, \rho_0^*)}{\partial \rho_1} \end{array}\right\} \xrightarrow{d} N(0, \Omega), \tag{3.13}$$

na qual

$$\Omega = \begin{pmatrix} \mathcal{R}_1(\rho_0) & \mathcal{R}_4(\rho) \\ \mathcal{R}_4(\rho)^\top & \mathcal{R}_2(\rho_1) \end{pmatrix}.$$

Ao expandir (3.13) sobre $\rho^* = (\rho_0^*, \rho_1^*)$ e aplicando a lei dos grandes números, tem-se

$$\begin{aligned} \sqrt{n}(\hat{\rho_0} - \rho_0^*) &\xrightarrow{d} -\mathcal{R}_1(\rho_0)^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_0(\rho_0^*)}{\partial \rho_0}, \\ \sqrt{n}(\hat{\rho_1} - \rho_1^*) &\xrightarrow{d} -\mathcal{R}_2(\rho_1)^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_{1t}(\rho_1^*, \rho_0^*)}{\partial \rho_1} \\ &\quad + \mathcal{R}_2(\rho_1)^{-1}\mathcal{R}_3(\rho)^\top\mathcal{R}_1(\rho_0)^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_0(\rho_0^*)}{\partial \rho_0}. \end{aligned} \tag{3.14}$$

Note-se que a primeira equação foi substituída na segunda em (3.14) para a aplicação da lei dos grandes números. Para mais consulte Murphy e Topel (1985) e Greene (2012, pg. 536).

Sabida a veracidade em (3.13) e (3.14), a distribuição assintótica para ρ_1 é dada por

$$\sqrt{n}(\hat{\rho_1} - \rho_1) \overset{a}{\sim} N(0, \Sigma),$$

sendo

$$\begin{aligned} \Sigma = \mathcal{R}_2(\rho_1) + \mathcal{R}_2(\rho_1)[&\mathcal{R}_3(\rho)^\top\mathcal{R}_1(\rho_0)\mathcal{R}_3(\rho) \\ &- \mathcal{R}_4(\rho)^\top\mathcal{R}_1(\rho_0)\mathcal{R}_3(\rho) - \mathcal{R}_3(\rho)^\top\mathcal{R}_1(\rho_0)\mathcal{R}_4(\rho)]\mathcal{R}_2(\rho_1), \end{aligned} \tag{3.15}$$

em que as matrizes podem ser estimadas pela definição em Greene (2012, pg. 536). Assim

$$\begin{aligned}
\hat{\mathcal{R}}_1(\hat{\rho}_0) &= \left[\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial ln f_{t0}}{\partial \hat{\rho}_0}\right)\left(\frac{\partial ln f_{t0}}{\partial \hat{\rho}_0^\top}\right)\right]^{-1}, \\
\hat{\mathcal{R}}_2(\hat{\rho}_1) &= \left[\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial ln f_{t1}}{\partial \hat{\rho}_1}\right)\left(\frac{\partial ln f_{t1}}{\partial \hat{\rho}_1^\top}\right)\right]^{-1}, \\
\hat{\mathcal{R}}_3(\hat{\rho}) &= \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial ln f_{t1}}{\partial \hat{\rho}_1}\right)\left(\frac{\partial ln f_{t1}}{\partial \hat{\rho}_0^\top}\right), \\
\hat{\mathcal{R}}_4(\hat{\rho}) &= \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial ln f_{t1}}{\partial \hat{\rho}_1}\right)\left(\frac{\partial ln f_{t0}}{\partial \hat{\rho}_0^\top}\right).
\end{aligned} \tag{3.16}$$

O termo $\hat{\Sigma}$ é uma matriz de variância-covariância para ρ_1 corrigida, a qual garante que o estimador é consistente e assintoticamente normal.

Capítulo 4

Bootstrap

Os métodos bootstrap foram introduzidos por Efron e Tibshirani (1986) e têm como objetivo reamostrar as observações para estimar a distribuição da estatística de interesse. No anexo B são discutidos o processo de estimação, o cálculo do viés e o intervalo de confiança para métodos bootstrap gerais. Na seção abaixo será discutida a metodologia que será usada para obter as estimativas dos estimadores VI e de máxima verossimilhança de dois-estágios via bootstrap.

4.1 Bootstrap para a modelagem dos escores da eficiência

Simar e Wilson (2007) apresentaram metodologias para a estimativa da eficiência medida por DEA sobre algumas variáveis contextuais no segundo estágio através do método bootstrap, no qual os parâmetros estimados são consistentes para o processo gerador de dados apresentado pelos mesmos. Similarmente a Simar e Wilson (2007), a ideia deste trabalho é fornecer uma metodologia para a estimação em dois-estágios via bootstrap, na presença de endogeneidade.

O objetivo em utilizar a metodologia bootstrap para o caso do estimador VI e da máxima verossimilhança de dois-estágios é comparar os erros padrão estimados na seção 3 com os obtidos por bootstrap. O processo que será apresentado é bem semelhante ao bootstrap geral retratado na seção B.1.

Recorde que, as matrizes $\mathbf{U}^\top$ e $\mathbf{W}^\top$ são as matrizes de insumos e produtos, respectivamente. A matrizes $\mathbf{X}$, $\mathbf{Z}$ e $\mathbf{V}$ são covariáveis dos preditores lineares de $\boldsymbol{\alpha}$, $\boldsymbol{\mu}$ e $\boldsymbol{\phi}$, respetivamente. Ademais, somente as variáveis que estejam nas componentes $\boldsymbol{\alpha}$ e $\boldsymbol{\mu}$ são tratadas como variáveis endógenas

e exógenas. A matriz $\mathbf{D}$ é tratada como matriz de variáveis instrumentais, a qual compartilha as mesmas varáveis exógenas de $\mathbf{X}$ e de $\mathbf{Z}$ e contém os instrumentos novos.

Portanto, ao considerar a estimação dos parâmetros para a Expressão 2.21 pelo estimador VI, tem-se que:

(i) Seja a matriz de variáveis $\mathbf{Q} = \begin{bmatrix}\mathbf{U}^\top & \mathbf{W}^\top & \mathbf{X} & \mathbf{Z} & \mathbf{D}\end{bmatrix}$ é a matriz de dados baseada nas n DMU's.

(ii) Para cada réplica bootstrap, b, indexado em $b = 1, \ldots, 2000$, faça:

   (a) Gere uma amostra pseudo-aleatória com repetição, por linha da matriz $\mathbf{Q}^{(b)}$, a partir da função de distribuição $\hat{F}(x)$.

   (b) Calcule $\theta^{(b)}$ escores de eficiência por DEA a partir da pseudo-amostra gerada em (a).

   (c) Dada a réplica $\theta^{(b)}$, calcule a b-ésima réplica das estimativas dos parâmetros $(\boldsymbol{\beta}^{(b)}, \boldsymbol{\gamma}^{(b)})$ do modelo em questão.

(iii) Por último, calcule a média e o desvio-padrão dos estimadores por bootstrap de $(\hat{\boldsymbol{\beta}}^{(b)}, \hat{\boldsymbol{\gamma}}^{(b)})$.

Atente-se que, para o caso do GMM via VI, o modelo em estudo é tratado apenas como um modelo não-linear, sem a necessidade de qualquer especificação da distribuição da variável resposta.

O processo para estimar o modelo de dois-estágios via máxima verossimilhança de dois-estágios por bootstrap é dado da seguinte maneira

(i) Seja a matriz de variáveis $\mathbf{Q} = \begin{bmatrix}\mathbf{U}^\top & \mathbf{W}^\top & \mathbf{X} & \mathbf{Z} & \mathbf{V} & \mathbf{D}\end{bmatrix}$, em que $\mathbf{Q}$ é a matriz de dados baseada nas n DMU's.

(ii) Para cada réplica bootstrap, b, indexado em $b = 1, \ldots, 2000$, faça:

   (a) Gere uma amostra pseudo-aleatória com repetição, por linha da matriz $\mathbf{Q}^{(b)}$, a partir da função de distribuição $\hat{F}(x)$.

   (b) Calcule $\theta^{(b)}$ escores de eficiência por DEA a partir da pseudo-amostra, gerada em (a).

(c) Estime o vetor de parâmetros $\boldsymbol{\rho}^{(b)}$ do modelo da variável endógena, do primeiro estágio, baseado nas variáveis exógenas e instrumentais da pseudo-amostra $\mathbf{Q}^{(b)}$. Lembrando que o vetor de parâmetros $\boldsymbol{\rho}^{(b)}$ pode representar os parâmetros para mais de um modelo, caso haja mais de uma variável endógena.

(d) Dada a réplica $\theta^{(b)}$ e $\boldsymbol{\rho}^{(b)}$, calcule a b-ésima réplica das estimativas dos parâmetros $(\boldsymbol{\beta}^{(b)}, \boldsymbol{\gamma}^{(b)}, \boldsymbol{\delta}^{(b)})$ do modelo em questão.

(iii) Por último, calcule a média e o desvio-padrão dos estimadores por bootstrap de $(\hat{\boldsymbol{\rho}}^{(b)}, \hat{\boldsymbol{\beta}}^{(b)}, \hat{\boldsymbol{\gamma}}^{(b)}, \hat{\boldsymbol{\delta}}^{(b)})$.

Sendo estes os processos via bootstrap para a presença de variáveis endógenas.

Capítulo 5

Implementação com Dados do Censo Agropecuário

5.1 Dados

Os dados compreendem as unidades produtoras, estabelecimentos rurais, agregadas em nível municipal, que serão as DMU's sob análise, são de 4965 municípios, representando quase 90% do total de municípios no Brasil. Os dados do censo agropecuário IBGE de 2006 foram utilizados para extrair informações a partir do sistema produtivo de cada unidade produtora. Portanto, para o processo de manufaturação temos como insumos os gastos totais com terra, com trabalho e com insumos tecnológicos (fatores de produção), contendo em cada insumo os fluxos de gastos das unidades; para o produto temos a renda bruta rural total dos estabelecimentos.

As variáveis contextuais observadas relevantes ao estudo foram acesso a crédito, índice de concentração de renda municipal (Gini), proporção de agricultores que receberam assistência técnica (Assist.-Técnica), índice municipal de desenvolvimento social (Social), índice de desenvolvimento demográfico (Demográfico), índice de desenvolvimento ambiental (Ambiental) e as regiões geográficas brasileira. As variáveis contextuais foram extraídas do censo demográfico brasileiro de 2010, do Ministério da Saúde (2011) e bases de dados do Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), data base 2009.

As variáveis sociais têm como finalidade quantificar o bem estar de cada indivíduo baseando-se em aspectos como disponibilidade de água, energia e esgoto sanitário, além de indicadores

do nível educacional e de saúde nas unidades rurais. As variáveis demográficos compreendem as circunstâncias relacionadas à performance do desenvolvimento rural, com maior importância as idades potencialmente ativa entre 15 e 59 anos. As variáveis ambientais são aquelas variáveis relacionadas às melhores práticas de conservação, juntamente com a produção agrícola para o estabelecimento.

As medidas social, ambiental e demográfica consideradas neste trabalho foram aquelas orientadas pelo instituto Confederação Nacional da Agricultura - CNA, conforme descrita em Souza, Gomes e Alves (2018). As variáveis foram ordenadas e normalizadas pelo máximo, quando necessário. Todas as análises foram realizadas no programa R (R Core Team, 2018).

5.2 Resultados DEA

A eficiência foi calculada através do modelo VRS com orientação a produto, tendo como insumos gastos totais com terra, trabalho e insumos tecnológicos e como produto a renda bruta total do estabelecimento, todos agregados por município. Na Tabela 5.1 têm-se as estatísticas dos escores da eficiência medida por DEA para cada região.

Tabela 5.1: Medidas dos escores da eficiência por VRS.

Região	Média	Desvio Padrão	Coeficiente de variação	Frequência	Quantidade de DMU's eficientes
Norte	0.4643	0.2085	0.4492	405	2
Nordeste	0.2945	0.2371	0.8049	1666	19
Sudeste	0.6448	0.2586	0.4011	1531	1
Sul	0.6406	0.1758	0.2744	1143	1
Centro-Oeste	0.7744	0.1963	0.2534	220	0
Geral	0.5173	0.2831	0.5473	4965	23

O Centro-Oeste é a região com maior eficiência média, seguida por Sudeste e Sul. O Nordeste tem a pior eficiência média, logo após aparece a região Norte. Ademais, note-se que a região Nordeste tem o maior coeficiente de variação, ou seja, uma maior variabilidade em relação à média. É possível observar uma concordância do coeficiente de variação com o Figura 5.1.

Os coeficientes de variação da região Norte e Sudeste estão bem próximos, embora ao observar o gráfico do box-plot é visível que para a região Norte a maior variabilidade está abaixo da mediana, enquanto para a região Sudeste, a maior variabilidade está acima da mediana. O

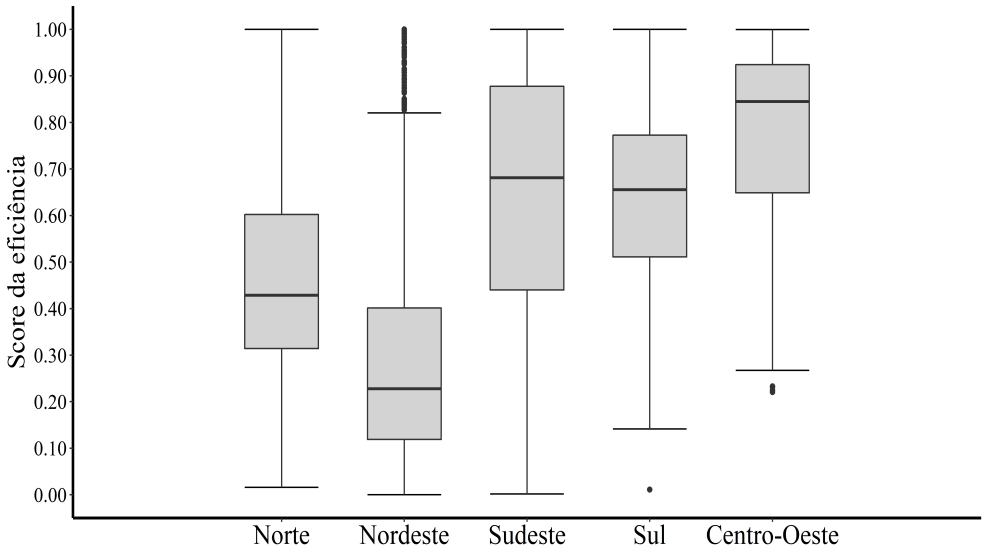

Figura 5.1: Box-Plot dos scores da eficiência medida pelo modelo VRS

Nordeste é a região com mais outliers, no qual esses outliers estão sendo causados pelo maior número de unidades totalmente eficientes. Embora a região Nordeste seja a de menor eficiência média, é a região com mais unidades eficientes, ou seja, $\hat{\theta} = 1$, enquanto que a região de maior eficiência média, Centro-Oeste, não há unidade eficiente.

5.3 Resultados do modelo de regressão Beta inflacionada em um

Na Tabela 5.2 tem-se o resultado do ajuste do modelo de regressão Beta inflacionada em um. As especificações em relação à função de ligação de cada componente do modelo foi dada na Seção 2.7. Optou-se em deixar apenas as variáveis significativas em cada componente. As variáveis contextuais foram todas significativas na componente contínua, μ, enquanto na parte da inflação, α, foram significativas apenas crédito, índice de desenvolvimento social e índice de concentração de renda municipal.

Ademais, o modelo considerou o fator região para modelar a heterogeneidade. A região Norte foi tomada como base para o preditor linear dos parâmetros de dispersão. Note que o parâmetro de dispersão das unidades produtoras da região Sul é maior do que o da região Norte, causando assim uma menor variância dos escores de eficiência estimados da região Sul. Enquanto os escores da eficiência estimada da região Centro-Oeste têm um parâmetro de dispersão semelhante ao da

região Norte. Sendo assim, o efeito do parâmetro de dispersão das duas regiões têm resultados semelhantes na variância dos escores da eficiência estimada.

Tabela 5.2: Estimativa da regressão Beta inflacionada em um.

Componente	Coeficiente	Estimado	Erro Padrão	Intervalo de confiança		P-valor
				Inferior	Superior	
$\hat{\mu}$	Intercepto	-6.3270	0.1049	-6.5327	-6.1214	0.0000
	Crédito	1.4886	0.0461	1.3981	1.5790	0.0000
	Social	1.1805	0.0753	1.0329	1.3280	0.0000
	Demográfico	1.3980	0.1002	1.2016	1.5945	0.0000
	Ambiental	-0.8889	0.1341	-1.1518	-0.6260	0.0000
	Assist.-Técnica	1.2188	0.0563	1.1085	1.3290	0.0000
	Gini	5.3152	0.1075	5.1044	5.5259	0.0000
$\hat{\phi}$	Intercepto	1.6586	0.0634	1.5343	1.7829	0.0000
	Nordeste	0.2081	0.0725	0.0660	0.3501	0.0041
	Sudeste	0.6454	0.0736	0.5011	0.7897	0.0000
	Sul	1.1630	0.0759	1.0143	1.3117	0.0000
	Centro-Oeste	0.1877	0.1107	-0.0293	0.4047	0.0901
$\hat{\alpha}$	Intercepto	-4.1999	0.7868	-5.7420	-2.6578	0.0000
	Crédito	-1.3992	0.5362	-2.4501	-0.3483	0.0091
	Social	-2.5454	0.7394	-3.9945	-1.0963	0.0006
	Gini	3.4139	0.9891	1.4752	5.3525	0.0006

Na Figura 5.2 temos os valores estimados dos componentes μ e α em relação às variáveis contextuais. Para a visualização das figuras, considerou-se que apenas a variável em estudo esteja variando, enquanto as demais são mantidas constantes em sua média. Por exemplo, para a variável Crédito, o modelo foi estimado usando a média das demais variáveis e somente crédito entrou no modelo em sua t-ésima observação.

Para este conjunto de dados, espera-se que os sinais dos parâmetros estimados das covariáveis que estejam em ambas componentes μ e α sejam iguais. Na componente μ (Tabela 5.2) os parâmetros estimados das variáveis Crédito, Social, Demográfico, Assist.-Técnica e Gini foram positivos, indicando que quanto maiores os valores destas variáveis maior a eficiência no sistema produtivo. Observa-se que a estimativa da variável Ambiental foi negativa, sinalizando que as unidades produtoras que têm maior preocupação com o meio ambiente apresentam decréscimo na eficiência. Pode-se encontrar este comportamento no Figura 5.2.

Ao contrário do que se esperava, no componente α, o sinal para as estimativas dos parâme-

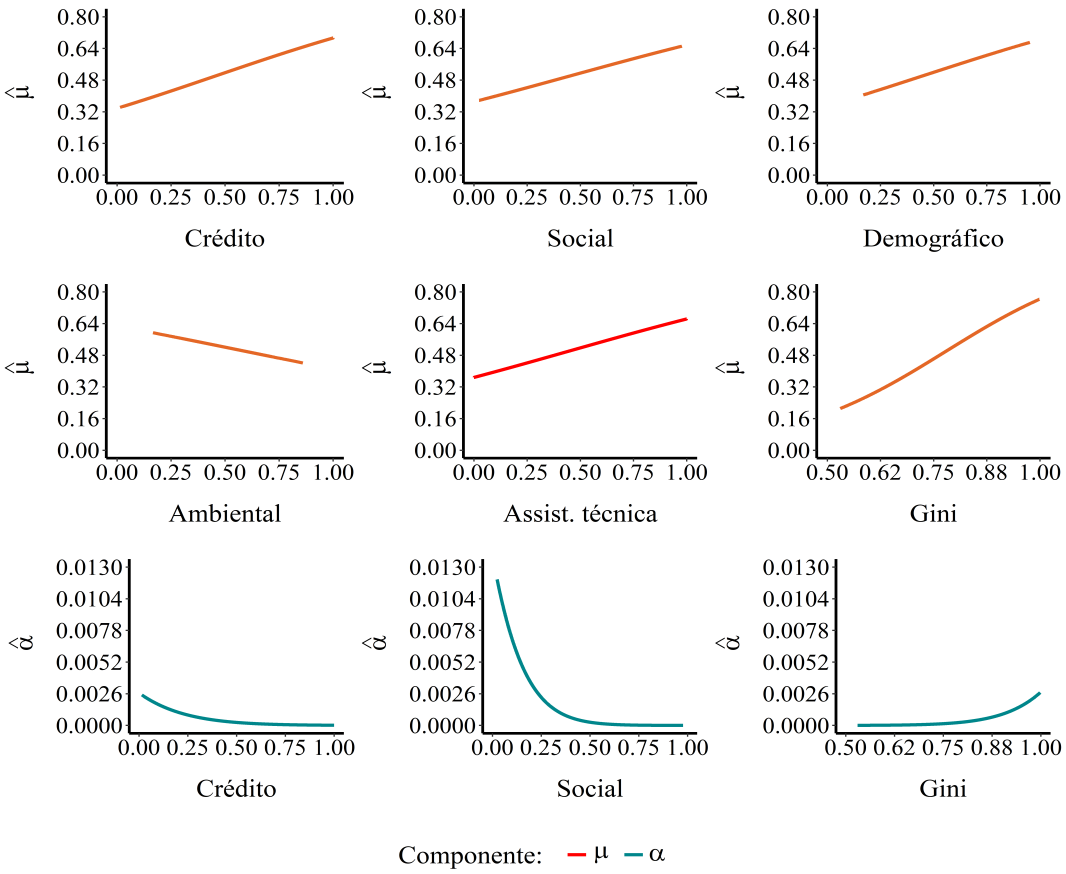

Figura 5.2: Probabilidade da parte contínua $\hat{\mu}$

tros Crédito e Social foram negativos, evidenciando que quanto maior os valores destas variáveis menor a probabilidade de se ter um sistema produtivo eficiente. Tal resultado pode estar relacionado às unidades eficientes situarem-se majoritariamente no Nordeste. Desta forma, presume-se que as eficiências destas unidades não serão afetadas, independente do quanto em crédito esteja disponível à elas ou a disponibilidade de serviços de água, energia elétrica, sistema de esgoto e outros, por parte de fatores externos.

Sintetizando, essas unidades produtoras buscam por si próprias meios para serem produtivas independente de fatores externos, por mais que a eficiência média da região Nordeste seja a menor. Acredita-se que tal situação gerou as estimativas dos parâmetros negativas paras as variáveis Crédito e Social. A variável Gini tem uma associação forte positiva com os escores da eficiência, indicando que as unidades produtoras eficientes estão relacionadas com a concentração

de renda municipal.

A correlação de Pearson entre os valores preditos do modelo e os observados foi de 0.8511, evidenciando uma correlação forte. Embora as covariáveis expliquem grande parte da variação da variável resposta, suspeita-se que há problema de endogeneidade.

Na seguinte seção, o modelo de regressão Beta inflacionada em um será estimado baseado nos gradientes, apresentado na expressão (2.19), com momentos igualados a zeros, sem considerar endogeneidade. Para o caso de endogeneidade, as variáveis Crédito e Assis.-Técnica serão tratadas como endógenas. Desse modo, nas próximas seções serão discutidos os processos de estimação pelo estimador VI e por Murphy Topel, na presença de endogeneidade. As duas formas foram tratadas na Seção 3.1.1 e 3.2, respectivamente.

5.4 Resultados do modelo de regressão Beta inflacionada em um via GMM com e sem variáveis contextuais endógenas

A suposição para estimar qualquer modelo via GMM é que os momentos sejam igualados a zero. O modelo de regressão Beta inflacionada em um foi estimado baseado no estimador ótimo GMM (3.6), por meio da abordagem de dois-estágios usando os momentos (2.19) e os resíduos (3.9) iguais a zero.

Para a realização do teste J, inseriu-se o vetor de resíduos para que o modelo seja superidentificado, embora a estimativa para o caso identificado também dê resultados bem similares. A Tabela 5.3 apresenta as estimativas dos parâmetros do modelo.

Tabela 5.3: Estimativa da regressão Beta inflacionada em um via GMM

Componente	Coeficiente	Estimado	Erro Padrão	Intervalo de confiança		Pr(>|t|)
				Inferior	Superior	
$\hat{\mu}$	Intercepto	-6.3292	0.1205	-6.5653	-6.0931	0.0000
	Crédito	1.4856	0.0681	1.3520	1.6191	0.0000
	Social	1.1822	0.0860	1.0136	1.3508	0.0000
	Demográfico	1.4004	0.1226	1.1600	1.6408	0.0000
	Ambiental	-0.8910	0.1530	-1.1909	-0.5912	0.0000
	Assist.-Técnica	1.2179	0.0709	1.0790	1.3568	0.0000
	Gini	5.3204	0.1199	5.0855	5.5553	0.0000
$\hat{\phi}$	Intercepto	1.6717	0.1071	1.4618	1.8816	0.0000
	Nordeste	0.1890	0.1108	-0.0282	0.4062	0.0880
	Sudeste	0.6343	0.1211	0.3969	0.8718	0.0000
	Sul	1.1450	0.1137	0.9221	1.3679	0.0000
	Centro-Oeste	0.1805	0.1620	-0.1370	0.4979	0.2652
$\hat{\alpha}$	Intercepto	-4.2638	1.0652	-6.3515	-2.1761	0.0001
	Crédito	-1.3959	0.5122	-2.3998	-0.3920	0.0064
	Social	-2.5476	0.9860	-4.4800	-0.6151	0.0098
	Gini	3.4918	1.2100	1.1203	5.8634	0.0039

O teste J apresentou p-valor de 0.74, aceitando-se a hipótese nula, ou seja, a performance do teste indica que os momentos são iguais a zero para a solução encontrada. A estimativa dos parâmetros via GMM foi similar aos resultados encontrados por EMV, embora os erros padrão por GMM tiveram um tênue aumento, salvo a variável Crédito da parte discreta. A correlação entre os valores preditos e os reais foi idêntica à apresentada por EMV, evidenciando que a estimativa via GMM é uma excelente forma de estimar os parâmetros com base nos gradientes e nos resíduos, como momentos.

Para contornar o problema de endogeneidade, estimou-se o modelo (2.21) satisfazendo a suposição (3.8) a partir do estimador GMM (3.11). Para o processo de estimativa do estimador ótimo GMM usou-se o estimador iterativo. Os resultados do modelo encontram-se na Tabela 5.4. As tentativas de estimar os parâmetros com preditores diferentes para μ e α foram todas inadequadas, seja por não convergir o algoritmo ou pela rejeição da hipótese nula do teste J. Logo, assumiu-se que as componentes μ e α tivessem o mesmo preditor linear, pois como já mencionado anteriormente, as componentes μ e α devem apresentar os coeficientes estimados com os mesmos sinais.

As variáveis instrumentais usadas foram gastos com terra, trabalho, insumos tecnológicos, Social, Demográfico, Ambiental e Gini. Além de considerar certas interações entre Social & Ambiental, Ambiental & terra, Gini & Ambiental e Gini & Social.

O teste J para o modelo dado na Tabela 5.4 obteve p-valor igual a 0.12. Desta forma, a solução encontrada para os parâmetros estimados satisfaz a condição de momentos iguais a zero.

Tabela 5.4: Estimativa da regressão Beta inflacionada em um por estimador VI com $(\hat{\mu}, \hat{\alpha})$ tendo o mesmo preditor linear

Componente	Coeficiente	Estimado	Erro Padrão	Intervalo de confiança		Pr(>|t|)
				Inferior	Superior	
$(\hat{\mu}, \hat{\alpha})$	Intercepto	-7.4587	0.9799	-9.3794	-5.5380	0.0000
	Crédito	0.1561	0.5126	-0.8486	1.1607	0.7607
	Social	-1.9101	0.6015	-3.0889	-0.7312	0.0015
	Demográfico	0.8169	0.1714	0.4808	1.1529	0.0000
	Ambiental	-1.4552	0.4751	-2.3863	-0.5240	0.0022
	Assist.-Técnica	5.5609	1.5041	2.6129	8.5090	0.0002
	Gini	6.5957	1.2262	4.1924	8.9990	0.0000

A variável Crédito, por mais que tenha um coeficiente positivo, não foi estatisticamente significativa. Ademais, nota-se que o coeficiente da variável Social ainda continua negativo, evidenciando o mesmo problema dado para o modelo presente na Tabela 5.2. Para as demais, os coeficientes das variáveis como Demográfico, Gini e Ambiental demonstraram comportamento similar ao modelo de regressão Beta inflacionada em um.

Na Tabela (5.5) temos a estimativa dos parâmetros pelo estimador VI por bootstrap. A decisão por estimar via bootstrap deveu-se ao propósito de diminuir a influência da correlação dos escores de eficiência medida por DEA e prover erros padrão mais confiáveis às estimativas.

Tabela 5.5: Estimativa da regressão Beta inflacionada em um pelo estimador VI por bootstrap com $(\hat{\mu}, \hat{\alpha})$ tendo o mesmo preditor linear

Componente	Coeficiente	Estimado	Erro Padrão	Intervalo de confiança		P-valor
				Inferior	Superior	
$(\hat{\mu}, \hat{\alpha})$	Intercepto	-7.4587	1.2107	-9.0607	-4.8239	0.0000
	Crédito	0.1561	0.5858	-0.7707	1.5180	0.7899
	Social	-1.9101	0.7098	-2.9908	-0.3274	0.0071
	Demográfico	0.8168	0.1860	0.4982	1.2341	0.0000
	Ambiental	-1.4552	0.5530	-2.2514	-0.2117	0.0085
	Assist.-Técnica	5.5609	1.7893	1.6320	8.2516	0.0019
	Gini	6.5957	1.5240	3.2933	8.5564	0.0000

Os erros padrão e a significância dos parâmetros estimados tiveram um pequeno aumento em comparação aos da Tabela 5.4. Os intervalos de confiança para o modelo (5.5) foram calculados com base no percentil centrado.

No Figura 5.3 temos o comportamento de cada variável em relação às componentes contínua e discreta. As componentes estão representadas separadamente, pois mesmo que os preditores sejam iguais, as funções de ligação são diferentes. As variáveis Assist.-Técnica e Gini possuem comportamento semelhante entre as componentes, em que os maiores valores destas variáveis apresentam influência relativamente acentuada para a componente discreta.

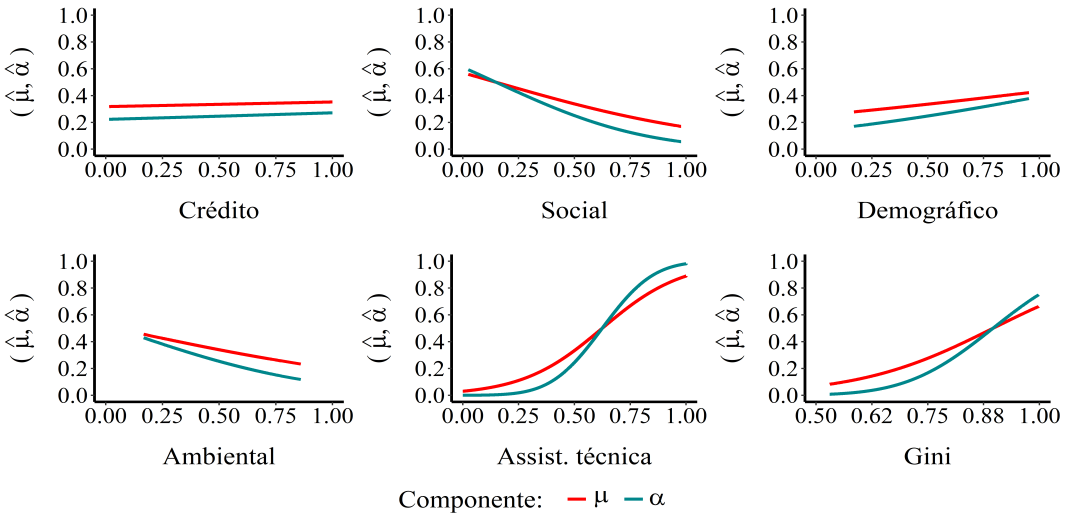

Figura 5.3: Probabilidade da parte contínua e discreta em que $(\hat{\alpha}, \hat{\mu})$ tem o mesmo preditor.

O modelo de regressão beta-inflacionado em um pelo estimador VI apresentou resultados

bastante satisfatórios, com uma correlação entre os valores preditos e os reais de 0.6974. As estimativas negativas dos coeficientes para a componente α das variáveis Crédito e Social ainda permaneceram no modelo estimado por GMM. As estimativas dos erros padrão dos parâmetros estimados do modelo pelo estimador VI via abordagens tradicionais e bootstrap também tiveram resultados semelhantes.

Na próxima seção, será apresentada a estimativa por EMV com correção de Murphy e Topel, para o caso com endogeneidade.

5.5 Resultados do modelo de regressão Beta inflacionada em um por EMV com correção de Murphy e Topel

Para tratar a presença de endogeneidade por meio da metodologia de Murphy e Topel (1985) exige dois estágios. As variáveis endógenas são as mesmas consideradas na estimação pelo estimador VI. Decidiu-se, então, por manter as mesmas variáveis instrumentais para maior facilidade de comparação entre as metodologias. Para o primeiro estágio, usou-se o modelo de regressão linear e a regressão fracionada para as variáveis Crédito e Assist.-Técnica, respectivamente. Os resultados de tais modelos podem ser analisados nas Tabelas A.1 e A.2 no anexo A.

Na Tabela 5.6 tem-se o modelo de regressão Beta inflacionada em um com α e μ tendo o mesmo preditor linear. Ademais, os erros padrão já estão corrigidos via expressão (3.15) desenvolvida por Murphy e Topel (1985).

Na componente ϕ, a estimativa dos parâmetros das regiões Sudeste, Sul e Centro-oeste tiveram sinais negativos quando comparada à região Norte. Portanto, somente a região Nordeste tem influência positiva sobre ϕ, causando uma menor variabilidade aos valores da eficiência quando comparado com a região Norte.

Tabela 5.6: Estimativa da regressão Beta inflacionada em um por EMV via correção de Murphy e Topel com $(\hat{\mu}, \hat{\alpha})$ tendo o mesmo preditor linear.

Componente	Coeficiente	Estimado	Erro Padrão	Intervalo de confiança		P-valor
				Inferior	Superior	
$(\hat{\mu}, \hat{\alpha})$	Intercepto	-5.3064	0.1861	-5.6712	-4.9416	0.0000
	Crédito	0.5591	0.2390	0.0907	1.0275	0.0193
	Social	-1.1706	0.3202	-1.7981	-0.5432	0.0003
	Demográfico	0.5382	0.1737	0.1977	0.8786	0.0019
	Ambiental	-0.3495	0.2644	-0.8678	0.1687	0.1862
	Assist.-Técnica	3.3468	0.6306	2.1107	4.5828	0.0000
	Gini	4.3729	0.3157	3.7542	4.9916	0.0000
$\hat{\phi}$	Intercepto	1.7538	0.0550	1.6460	1.8616	0.0000
	Nordeste	0.1272	0.0665	-0.0032	0.2575	0.0559
	Sudeste	-0.4119	0.0697	-0.5485	-0.2753	0.0000
	Sul	-0.0425	0.0800	-0.1993	0.1144	0.5957
	Centro-Oeste	-0.6280	0.1361	-0.8947	-0.3613	0.0000

Ao comparar os resultados do modelo na Tabela 5.6 com os resultado da Tabela 5.4, os sinais dos parâmetros mantiveram-se os mesmos. Observa-se que a estimativa do parâmetro da variável Crédito foi significativa e a estimativa do parâmetro da variável ambiental não foi significativa na Tabela 5.6, em comparação ao da Tabela 5.4. Para a variável Social, a estimativa negativa do parâmetro ainda persiste.

Na Tabela 5.7 encontra-se o modelo de regressão Beta inflacionada em um com α e μ tendo o mesmo preditor linear estimado pelo processo de dois-estágios por máxima verossimilhança via bootstrap. Contrastando-se os erros padrão corrigidos por Murphy e Topel (1985) e por bootstrap, é notório que a diferença é mínima, não havendo discrepância na estimativa dos erros padrão entre as duas técnicas.

Tabela 5.7: Estimativa da regressão Beta inflacionada em um com $(\hat{\mu}, \hat{\alpha})$ tendo o mesmo preditor linear via bootstrap

Componente	Coeficiente	Estimado	Erro Padrão	Intervalo de confiança		P-valor
				Inferior	Superior	
$(\hat{\mu}, \hat{\alpha})$	Intercepto	-5.3064	0.1747	-5.8204	-5.1285	0.0000
	Crédito	0.5591	0.1830	0.1986	0.9191	0.0023
	Social	-1.1706	0.2090	-1.6386	-0.8191	0.0000
	Demográfico	0.5382	0.1312	0.2954	0.8053	0.0000
	Ambiental	-0.3495	0.1984	-0.7085	0.0614	0.0782
	Assist.-Técnica	3.3468	0.4323	2.6587	4.3356	0.0000
	Gini	4.3729	0.2081	4.0674	4.8677	0.0000
$\hat{\phi}$	Intercepto	1.7538	0.1188	1.5455	2.0214	0.0000
	Nordeste	0.1272	0.1292	-0.1360	0.3730	0.3251
	Sudeste	-0.4119	0.1233	-0.6412	-0.1482	0.0008
	Sul	-0.0425	0.1269	-0.2637	0.2389	0.7379
	Centro-Oeste	-0.6280	0.1420	-0.8912	-0.3360	0.0000

Os valores estimados e os reais apresentaram correlação de 0.8657. Na Figura 5.4 verifica-se o comportamento das variáveis em relação a α e μ. O gráfico salienta inclusive o quanto as variáveis contextuais no modelo pelo estimador VI ou correção de Murphy e Topel (1985) tem comportamento similar, salvo a variável Crédito que aparece levemente acentuada positivamente, como já era esperado, pois o parâmetro foi significativo e positivo.

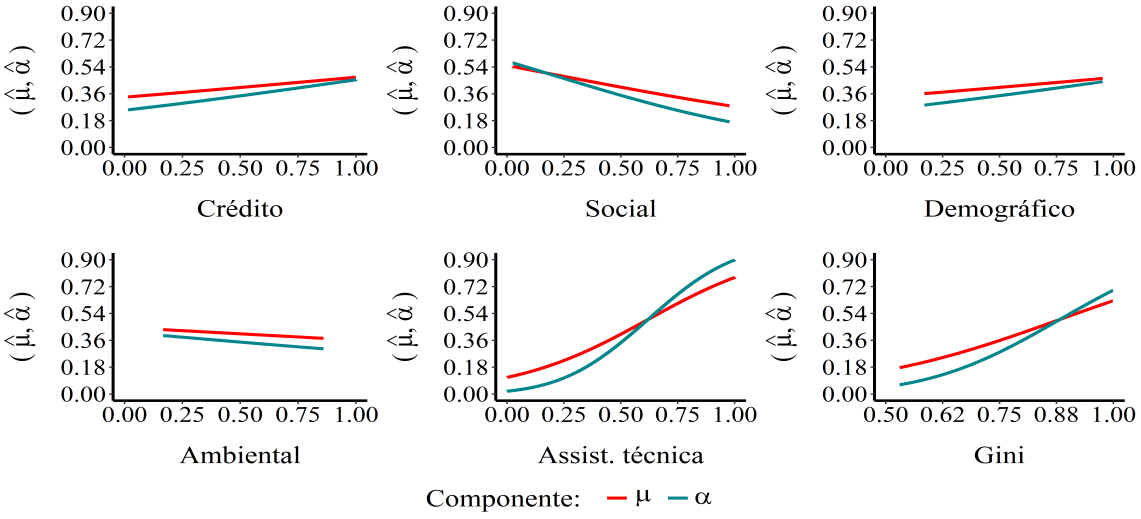

Figura 5.4: Probabilidade da parte contínua e inflacionada em que $(\hat{\alpha}, \hat{\mu})$ tem o mesmo preditor.

Embora haja uma similaridade no comportamento das variáveis contextuais em relação a

(μ, α), o modelo por EMV via correção de Murphy e Topel (1985) teve um desempenho melhor quando comparado ao valor da correlação entre os valores preditos e os observados do modelo pelo estimador VI. Isso pode ter sido ocasionado pela falta de informação do parâmetro ϕ quando estimado por estimador VI. No caso em que a estimativa é obtida por EMV, essa informação não é descartada.

5.6 Conclusão

Neste capítulo foi abordada a aplicação das teorias apresentadas. Nos resultados do modelo DEA foi possível observar que a região Centro-Oeste tem a maior eficiência média, embora não exista nenhuma DMU que seja totalmente eficiente. Contudo, mesmo a região Nordeste sendo aquela com menor eficiência média, é a região que apresenta mais unidades totalmente eficientes.

Ao estimar o modelo Beta inflacionada em um, as variáveis contextuais foram todas significativas na componente μ, enquanto que em α apenas a variável Crédito, Social e Gini foram significativas. Além disso, as variáveis Crédito e Social na componente α, apresentaram valores negativos para seus respectivos parâmetros estimados, tanto por EMV ou GMM, problema esse ocasionado possivelmente pelas unidades de produção eficientes que estão em sua maioria na região Nordeste, sendo essa a região de menor eficiência média, como já mencionado.

As variáveis Crédito e Assis.-Técnica foram consideradas como endógenas. Ao tratar a endogeneidade, pelo estimador VI ou por EMV via correção de Murphy e Topel (1985), julgou-se necessário tomar α e μ com o mesmo preditor. Este procedimento ocorreu essencialmente devido ao estimador VI não haver convergido, ao se considerar ambas componentes com preditores diferentes.

Para a variável Social, o valor do parâmetro estimado negativo ainda persiste, no caso em que a estimativa foi dada pelo estimador IV ou por EMV via correção de Murphy e Topel (1985). Não houve discrepância entre os erros padrão estimados pelo estimador IV ou por EMV via correção de Murphy e Topel (1985) em comparação ao bootstrap, ou seja, ambos os métodos capturam a variabilidade da estimativa dos parâmetros de forma satisfatória.

Capítulo 6

Considerações finais

Apesar de diversas relevantes publicações em relação à regressão de dois-estágios, este trabalho tem o propósito de contribuir para uma distribuição que possa adequar-se aos escores da eficiência calculados por DEA, neste caso a distribuição Beta inflacionada em um. As estimativas dos parâmetros deste modelo podem ser obtidas através de EMV ou GMM, os quais produzem resultados similares.

Na presença de variáveis contextuais endógenas, sugere-se uma abordagem por EMV com correção de Murphy e Topel, principalmente ao considerar o poder de predição do modelo e a informação em relação parâmetro de escala da distribuição beta inflacionada, pois o parâmetro de escala surge a partir do segundo momento. No entanto, ainda há um caminho a ser trilhado para melhoria do processo de estimação pelo estimador VI ao se considerar o modelo de regressão beta-inflacionada em um, principalmente uma metologia em que possa lidar com a endogeneidade e estimar o parâmetro de escala conjuntamente.

Ademais, os métodos supracitados fornecem erros padrão confiáveis tanto quanto os estimados por bootstrap na aplicação dos dados agropecuário. Em um trabalho futuro, subsistem perspectivas por questões sobre a análise dos resíduos e outliers para o caso de dois-estágios, quer a estimativa seja obtida por EMV ou GMM.

Bibliografia

Banker, Rajiv D, Charnes, Abraham e Cooper, William Wager (1984). "Some models for estimating technical and scale inefficiencies in data envelopment analysis". *Management science* 30.9, pp. 1078–1092.

Cameron, A Colin e Trivedi, Pravin K (2005). *Microeconometrics: methods and applications*. Cambridge university press.

Casella, George e Berger, Roger L (2002). *Statistical inference*. Vol. 2. Duxbury Pacific Grove, CA.

Charnes, Abraham, Cooper, William W e Rhodes, Edwardo (1978). "Measuring the efficiency of decision making units". *European journal of operational research* 2.6, pp. 429–444.

Coelli, Timothy J et al. (2005). *An introduction to efficiency and productivity analysis*. Springer Science & Business Media.

Collett, David (2002). *Modelling binary data*. CRC press.

Cooper, WW (2007). *Seiford. LM and Tone, K.(2000) Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*.

Cox, David Roxbee e Hinkley, David Victor (1979). *Theoretical statistics*. Chapman e Hall/CRC.

Daraio, Cinzia e Simar, Leopold (2007). *Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications*. Springer Science & Business Media.

Davidson, Russell e MacKinnon, James G (1995). "Estimation and inference in econometrics". *Econometric Theory* 11.3, pp. 631–635.

Debreu, Gerard (1951). "The coefficient of resource utilization". *Econometrica: Journal of the Econometric Society*, pp. 273–292.

Efron, Bradley e Tibshirani, Robert (1986). "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy". *Statistical science*, pp. 54–75.

Emrouznejad, Ali e Yang, Guo-liang (2018). “A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016”. *Socio-Economic Planning Sciences* 61, pp. 4–8.

Farrell, Michael James (1957). “The measurement of productive efficiency”. *Journal of the Royal Statistical Society: Series A (General)* 120.3, pp. 253–281.

Ferrari, Silvia e Cribari-Neto, Francisco (2004). “Beta regression for modelling rates and proportions”. *Journal of applied statistics* 31.7, pp. 799–815.

Greene, William H (2012). *Econometric analysis, 71e*. New York University: Prentice Hall.

Hamilton, James D (1995). “Time series analysis”. *Economic Theory. II, Princeton University Press, USA*, pp. 625–630.

Hansen, Lars Peter (1982). “Large sample properties of generalized method of moments estimators”. *Econometrica: Journal of the Econometric Society*, pp. 1029–1054.

Hansen, Lars Peter, Heaton, John e Yaron, Amir (1996). “Finite-sample properties of some alternative GMM estimators”. *Journal of Business & Economic Statistics* 14.3, pp. 262–280.

Hardin, James W (2002). “The robust variance estimator for two-stage models”. *The Stata Journal* 2.3, pp. 253–266.

Hoff, Ayoe (2007). “Second stage DEA: Comparison of approaches for modelling the DEA score”. *European Journal of Operational Research* 181.1, pp. 425–435.

Leiderman, Leonardo (1980). “Macroeconometric testing of the rational expectations and structural neutrality hypotheses for the United States”. *Journal of Monetary Economics* 6.1, pp. 69–82.

Liu, John S et al. (2013). “A survey of DEA applications”. *Omega* 41.5, pp. 893–902.

McCullagh, Peter e Nelder, John A (1983). “1989”. *Generalized linear models* 37.

McDonald, John (2009). “Using least squares and tobit in second stage DEA efficiency analyses”. *European journal of operational research* 197.2, pp. 792–798.

Murphy, Kevin M e Topel, Robert H (1985). “Estimation and inference in two-step econometric models”. *Journal of Business & Economic Statistics* 3.4, pp. 88–97.

Ospina, Raydonal e Ferrari, Silvia LP (2010). “Inflated beta distributions”. *Statistical Papers* 51.1, p. 111.

— (2012). “A general class of zero-or-one inflated beta regression models”. *Computational Statistics & Data Analysis* 56.6, pp. 1609–1623.

Papke, Leslie E e Wooldridge, Jeffrey M (1996). "Econometric methods for fractional response variables with an application to 401 (k) plan participation rates". *Journal of applied econometrics* 11.6, pp. 619–632.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Ramalho, Esmeralda A, Ramalho, Joaquim JS e Henriques, Pedro D (2010). "Fractional regression models for second stage DEA efficiency analyses". *Journal of Productivity Analysis* 34.3, pp. 239–255.

Rizzo, Maria L (2007). *Statistical computing with R*. Chapman e Hall/CRC.

Simar, Leopold e Wilson, Paul W (2007). "Estimation and inference in two-stage, semi-parametric models of production processes". *Journal of econometrics* 136.1, pp. 31–64.

Souza, G da S, Gomes, Eliane Gonçalves e Alves, ER de A (2018). "Imperfeições de mercado e concentração de renda na produção agrícola." *Revista de Política Agrícola, Brasília, DF, v. 27, n. 2, p. 31-38.*

SOUZA, G de S (1998). *Introdução aos modelos de regressão linear e não-linear*. EMBRAPA-SPI Brasília.

Wooldridge, Jeffrey M (2010). *Econometric analysis of cross section and panel data*. MIT press.

Apêndice A

Regressão do primeiro estágio para a correção Murphy e Topel

Aqui estão os resultados da regressão linear e fracionada para a variável Crédito e Assist.-Técnica, respectivamente. A variável Crédito é assimétrica e com alguns valores zero, portanto para a regressão linear para a mesma usou-se o valor um somado a Crédito e depois o logaritmo natural, ou seja, $\log(\text{Crédito} + 1)$. A regressão linear para a varável $\log(\text{Crédito} + 1)$ é dada na tabela A.1 abaixo.

Tabela A.1: Estimativa da regressão linear.

	Estimativa	Erro Padrão	valor t	Pr(>|t|)
Intercepto	2.0383	1.1518	1.7697	0.0768
Terra	-0.1557	0.1129	-1.3795	0.1678
Trabalho	-0.1580	0.0298	-5.3080	0.0000
Insumo-Técn.	0.9303	0.0418	22.2672	0.0000
Social	4.6741	1.1923	3.9204	0.0001
Demográfico	-2.5526	0.2578	-9.9031	0.0000
Ambiental	-6.3462	2.6314	-2.4117	0.0159
Gini	-0.2992	1.3340	-0.2243	0.8225
Ambiental & Gini	1.5633	3.0747	0.5085	0.6112
Ambiental & Social	-1.8116	1.3848	-1.3082	0.1909
Social & Gini	-2.5325	1.1541	-2.1944	0.0283
Ambiental & Terra	0.5080	0.2137	2.3765	0.0175

A Regressão fracionada para a variável Assist.-Técnica é dada na tabela A.2 abaixo.

Tabela A.2: Estimativa da regressão fracionada.

	Estimativa	Erro Padrão	valor t	Pr(>|t|)
Intercepto	-5.0492	0.6779	-7.4488	0.0000
Terra	-0.1985	0.0672	-2.9532	0.0032
Trabalho	0.0164	0.0172	0.9574	0.3384
Resto	0.3923	0.0249	15.7704	0.0000
Social	5.6700	0.7126	7.9573	0.0000
Demográfico	-0.4772	0.1470	-3.2458	0.0012
Ambiental	0.9917	1.5293	0.6485	0.5167
Gini	1.6376	0.7869	2.0810	0.0375
Ambiental & Gini	-4.5358	1.7817	-2.5458	0.0109
Ambiental & Social	-0.6178	0.8092	-0.7635	0.4452
Social & Gini	-3.3201	0.6977	-4.7589	0.0000
Ambiental & Terra	0.3890	0.1268	3.0667	0.0022

Apêndice B

Bootstrap

B.1 Método bootstrap

Métodos de bootstrap são uma classe de métodos de Monte Carlo não-paramétricos que estimam a distribuição da população por reamostragem. São frequentemente usados quando a distribuição da população de interesse não é especificada. SOUZA (1998, cap. 8) introduziu a seguinte definição:

Definição B.1.1. Seja $\mathbf{x} = (x_1, \ldots, x_n)^\top$ uma amostra aleatória de tamanho n de uma população com função de distribuição $F(x) \in \mathbb{R}^k$. Seja $\theta = \theta(F)$ um parâmetro populacional de interesse. Suponha que esteja disponível, com base em $\mathbf{x}$, um estimador $\hat{F}(x)$ de $F(x)$. O estimador $\hat{\theta} = T(\hat{F}(x))$ é um estimador bootstrap resultante da substituição de F por $\hat{F}$ no funcional que define θ.

A definição B.1.1 tem por finalidade a substituição da $F(x)$ desconhecida por $\hat{F}(x)$ estimada, ou seja, a $\hat{F}(x)$ assumirá o papel da $F(x)$. A estimativa por bootstrap pode ser paramétrica ou não-paramétrica, mais conhecida por "bootstrap paramétrico" ou "bootstrap não-paramétrico". A diferença entre os dois processos basicamente está na escolha da $\hat{F}(x)$. Para entender melhor, assuma o seguinte conceito:

(i) Bootstrap paramétrico: Se $F(x)$ for conhecido, $F(x) = F_\theta(x)$, então $\hat{F}(x) = F_{\hat{\theta}}(x)$, em que $\hat{\theta}$ é uma estimativa de θ, caso semelhante dos EMV. Desta forma, a metodologia de reamostragem para $\mathbf{x}^{*(b)}$ será dada pela amostra pseudo-aleatória obtida a partir da

própria distribuição $F_{\hat{\theta}}(x)$ e não mais da amostra original.

(ii) Bootstrap não-paramétrico: Se $F(x)$ for desconhecido, $F(x) = \hat{F}(x)$, então tem-se uma função de distribuição empírica, a qual definimos que $\hat{F}(x) = n^{-1} \sum_{j=1}^{n} \mathbf{I}_{\{\mathbf{x}_j \leq x\}}(x)$, em que quando há valores repetidos na amostra, $\hat{F}(x)$ atribui probabilidade proporcional à frequência amostral, a cada valor específico de x. Portanto, amostrar $\mathbf{x}^{*(b)}$ a partir $\hat{F}(x)$ é o mesmo que tomar uma amostra com reposição dos dados originais de tamanho n, com probabilidade n^{-1}.

O estimador bootstrap para θ deste processo será a média das réplicas $\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(B)}$, dado pelo processo gerador de $\hat{F}(x)$, onde $\hat{F}(x)$ é conhecido como distribuição bootstrap. Rizzo (2007, pg. 184) ilustra o seguinte processo de estimativa por bootstrap para θ:

(i) Para cada réplica bootstrap, indexado em $b = 1, \ldots, B$:

(a) Gere uma amostra pseudo-aleatória com repetição $\mathbf{x}^{*(b)} = x_1^*, \ldots, x_n^*$ a partir da função de distribuição $\hat{F}(x)$, na qual $\hat{F}(x)$ pode ser gerada via bootstrap paramétrico ou não-paramétrico.

(b) Calcule a b-ésima réplica $\hat{\theta}^{(b)}$ a partir da pseudo-amostra gerada em (a), $\mathbf{x}^{*(b)}$.

(ii) A média e o desvio-padrão do estimador bootstrap de $\hat{\theta}^{(b)}$ são dados, respectivamente, por

$$\hat{\theta} = \frac{\sum_{b=1}^{B} \hat{\theta}^{(b)}}{B} \qquad \text{e} \qquad se(\hat{\theta}) = \sqrt{\frac{\sum_{b=1}^{B} (\hat{\theta}^{(b)} - \hat{\theta})^2}{B-1}}.$$

O viés de um estimador é definido como $\text{viés}(\hat{\theta}) = E[\hat{\theta}] - \theta$. Ao tomar $\underline{\hat{\theta}}$ como um estimador para θ baseado na amostra original tem-se que o viés para o estimador $\hat{\theta}$ obtido por bootstrap é dado por $\text{viés}(\hat{\theta}) = \hat{\theta} - \underline{\hat{\theta}}$. Os resultados para estimativas pontuais por bootstrap aqui apresentados, também podem ser entendidos para intervalos de confiança, sendo este apresentado na próxima seção.

B.2 Intervalo de confiança bootstrap

A estimativa intervalar de um parâmetro é de suma importância, principalmente por conter mais informação do que uma estimativa pontual. SOUZA (1998, pag. 451) descreve as propriedades dos principais intervalos de confiança bootstrap (ICB), ademais Rizzo (2007, pg. 197) descreve as vantagens e desvantagens ao utilizar alguns ICB. O ICB mais usual, principalmente por causa da teoria assintótica, é o normal.

Seja $\hat{\theta}$ um estimador bootstrap não enviesado para θ e n proveniente de uma amostra grande, então pode-se construir um intervalo de confiança z-normal para $\hat{\theta}$, da forma

$$\left(\hat{\theta} - z_{\frac{\alpha}{2}} se(\hat{\theta}) \ ; \ \hat{\theta} + z_{\frac{\alpha}{2}} se(\hat{\theta})\right),$$

em que $se(\hat{\theta})$ é o desvio padrão bootstrap, o mesmo dado anteriormente; $z_{\frac{\alpha}{2}} = \Phi^{-1}(\alpha/2)$, o qual $\Phi(.)$ é a distribuição acumulada normal padrão. A suposição ao usar o ICB z-normal é que a distribuição de $\hat{\theta}$ é normal. Entretanto, tal suposição pode não ser satisfeita. Neste caso, outra possibilidade é ICB t-student, dada como

$$\left(\hat{\theta} - t^*_{1-\frac{\alpha}{2}} se(\hat{\theta}) \ ; \ \hat{\theta} + t^*_{\frac{\alpha}{2}} se(\hat{\theta})\right),$$

vale observar que $t^*_{1-\frac{\alpha}{2}}$ e $t^*_{\frac{\alpha}{2}}$ são os quantis empíricos de $t^{(b)} = \frac{\hat{\theta}^{(b)} - \hat{\theta}}{se(\hat{\theta}^{(b)})}$ ordenados, ou seja, $t^{(1)} \leq \cdots \leq t^{(B)}$; $se(\hat{\theta}^{(b)})$ é calculado a cada amostra pseudo-aleatória com repetição, $\mathbf{x}^{*(b)}$. A suposição mais forte que se tem para este intervalo é que a distribuição para $\hat{\theta}$ seja aproximadamente normal e com um viés baixo.

Os ICB's apresentados anteriormente são todos baseados nas distribuições tradicionais, embora exista outros ICB's, tais como o percentil, percentil centrado e percentil viés corrigido acelerado. O ICB percentil leva apenas em consideração a distribuição amostral da estimativa dos $\hat{\theta}^{(b)}$, dado da seguinte forma

$$\left(H^{-1}_{\frac{\alpha}{2}} \ ; \ H^{-1}_{1-\frac{\alpha}{2}}\right) = \left(\hat{\theta}_{\frac{\alpha}{2}} \ ; \ \hat{\theta}_{1-\frac{\alpha}{2}}\right),$$

no qual H é a função de distribuição empírica de $\hat{\theta}$. Uma suposição importante é que quando $n \to \infty$ a distribuição H de $\hat{\theta}$ se aproxima da distribuição normal. Portanto, para este intervalo

de confiança uma amostra grande é essencial. O ICB percentil não nos garante a simetria do intervalo, uma saída para este caso é o ICB percentil centrado, ou mais conhecido como ICB básico, dado como

$$\Big(2\hat{\theta} - H^{-1}_{1-\frac{\alpha}{2}} \ ; \ 2\hat{\theta} - H^{-1}_{\frac{\alpha}{2}}\Big),$$

sendo que $H^{-1}_{1-\frac{\alpha}{2}}$ e $H^{-1}_{\frac{\alpha}{2}}$ é o mesmo dado para o ICB percentil.

Um excelente estimador intervalar bootstrap é o percentil viés corrigido acelerado, mas conhecido como BCa, dado por

$$\Big(H^{-1}_{\alpha_1} \ ; \ H^{-1}_{\alpha_2}\Big) = \Big(\hat{\theta}_{\alpha_1} \ ; \ \hat{\theta}_{\alpha_2}\Big),$$

em que

$$\alpha_1 = \Phi\Big(\hat{z}_0 + \frac{\hat{z}_0 + z_{\frac{\alpha}{2}}}{1-\hat{a}(\hat{z}_0 + z_{\frac{\alpha}{2}})}\Big),$$

$$\alpha_1 = \Phi\Big(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\frac{\alpha}{2}}}{1-\hat{a}(\hat{z}_0 + z_{1-\frac{\alpha}{2}})}\Big),$$

em que $\Phi(.)$ é a função de distribuição acumulada normal padrão; $z_{\frac{\alpha}{2}}$ e $z_{1-\frac{\alpha}{2}}$ são os quantis da distribuição normal padrão. O valor $\hat{z}_0$ representa o fator de correção de viés da mediana de $\hat{\theta}$ para $\underline{\hat{\theta}}$, dada por

$$\hat{z}_0 = \Phi^{-1}\left(\sum_{b=1}^{B} \frac{\mathbf{I}_{\{\hat{\theta} < \underline{\hat{\theta}}\}}(\hat{\theta})}{B}\right),$$

note que se $\hat{z}_0 = 0$, então $\underline{\hat{\theta}}$ será a mediana das réplicas bootstrap. O fator de aceleração $\hat{a}$ é estimado a partir das réplicas jackknife, dado por

$$\hat{a} = \frac{\sum_{i=1}^{n}\Big(\hat{\theta}_{(.)} - \hat{\theta}_{(i)}\Big)^3}{6\left[\sum_{i=1}^{n}\Big(\hat{\theta}_{(.)} - \hat{\theta}_{(i)}\Big)^2\right]^{\frac{2}{3}}},$$

onde $\frac{\sum_{i=1}^{n} \hat{\theta}_{(i)}}{n}$; $\hat{a}$ pode ser interpretado como uma medida de assimetria para $\hat{\theta}_{(i)}$. Lembrando que não discutiu-se o método de reamostragem Jackknife neste trabalho, para saber sobre o método Jackknife consulte Rizzo (2007, pg. 190).

Apêndice C

Entrada de dados no R

```
###### Cáculo da eficiência por DEA

> rm(list = ls())
> ##### Pacote
>
> library(Benchmarking)
Carregando pacotes exigidos: lpSolveAPI
Carregando pacotes exigidos: ucminf
> library(dplyr)
> library(readxl)
> #library(rDEA)
> ## library(help = Benchmarking)
>
> ##### Diretório
>
> getwd()
[1] "C:/Users/Bruno Soares/Google Drive/Trabalho/1. Mestrado - Estatística - UnB/Disserta
    ção/0. Disertação"
> setwd("C:\\Users\\Bruno\\Google Drive\\Trabalho\\1. Mestrado - Estatística - UnB\\
    Dissertação\\0. Disertação")
>
> #### Data
> load("dados.RData")
>
> xy <- dados[, 3:6]
> names(xy)
[1] "ly"      "lxterra" "lxtrab"  "lxresto"
```

```
> v <- dim(xy)
>
> n <- v[1] ; p <- v[2]
> matriz <- matrix(ncol = p, nrow = n)
> colnames(matriz) <- c("lyR", "lxterraR", "lxtrabR", "lxrestoR")
> bra <- data.frame(matriz)
>
> for ( i in 1:p) {
+   bra[, i] <- rank(xy[,i])
+ }
>
> nome <- c("lxterraR", "lxtrabR", "lxrestoR")
>
> input  <- bra[ , nome]/4965
> output <- bra[,"lyR"]/4965
>
> dados <- cbind(dados, "lyR" = output, input)
>
> #save(dados, file = "dados.RData")
> #rm(list = ls())
> #load("dados.RData")
> #names(dados)
>
> ###### DEA - Retorno de escala
> ### Indice de eficiencia radial
> e_vrs_out <- dea(input, output, RTS="vrs", ORIENTATION= "out")
> str(e_vrs_out)
List of 12
$ eff         : num [1:4965] 2.07 2.32 2.18 2.45 1.89 ...
$ lambda      : num [1:4965, 1:4965] 0 0 0 0 0 0 0 0 0 0 ...
..- attr(*, "dimnames")=List of 2
.. ..$ : NULL
.. ..$ : chr [1:4965] "L1" "L2" "L3" "L4" ...
$ objval      : num [1:4965] 2.07 2.32 2.18 2.45 1.89 ...
$ RTS         : chr "vrs"
$ primal      : NULL
$ dual        : NULL
$ ux          : NULL
$ vy          : NULL
$ gamma       :function (x)
$ ORIENTATION: chr "out"
$ TRANSPOSE   : logi FALSE
$ param       : NULL
```

```
- attr(*, "class")= chr "Farrell"
> Efi_out = 1/e_vrs_out$eff
> efi_out <- data.frame(Efi_out)

###### Segundo Estágio para a eficência

> rm(list = ls())
> ###library(help = frm)
>
> library(frm)
> require(optimx)
> library(Matrix)
> library(xtable)
> library(gmm)
>
> ##### Diretório
> getwd()
> setwd("C:\\Users\\Bruno Soares\\Google Drive\\Trabalho\\1. Mestrado - Estatística - UnB
    \\Dissertação\\0. Disertação")
>
> ### load("dados.RData")
> function_beta <- function (b0, b1_fin, b2_soc, b3_dem, b4_amb, b5_ica, b6_gini,
+                            b00, b11_fin, b22_soc, b33_dem, b44_amb, b55_ica, b66_gini,
+                            x1, x2, x3, x4, x5, x6){
+
+   mu_cont    <- b0  + b1_fin  * x1 + b2_soc  * x2 + b3_dem   * x3 + b4_amb  * x4 + b5_
    ica  * x5 + b6_gini  * x6
+   alpha_disc <- b00 + b11_fin * x1 + b22_soc * x2 + b33_dem  * x3 + b44_amb * x4 + b55_
    ica * x5 + b66_gini * x6
+
+   p_alpha <- pnorm(alpha_disc)
+   p_mu <- (1/(1 + exp(-(mu_cont))))
+
+   mu_res_nls <- p_alpha + (1 - p_alpha)*p_mu
+   return(mu_res_nls)
+ }
> xx <-  dados
> y  <- xx$y
> x1 <- xx$financi
> x2 <- xx$social
> x3 <- xx$demo
> x4 <- xx$ambi
> x5 <- xx$ica460s
```

```
> x6 <- xx$ginitotal
>
> bin   <- c(1,0,0,0,0,0,0)
> frac  <- c(1,0,0,0,0,0,0)
> foo <- function(b){
+   sum((y- function_beta(b0=b[1],  b1_fin=b[2],  b2_soc=b[3],   b3_dem=b[4], b4_amb=b
    [5], b5_ica=b[6], b6_gini=b[7],
+                   b00=b[8], b11_fin=b[9], b22_soc=b[10], b33_dem=b[11], b44_amb=b[12],
    b55_ica=b[13], b66_gini=b[14],
+                         x1 = x1, x2 = x2, x3 = x3, x4 = x4, x5 = x5, x6 = x6))^2)
+ }
> est_beta <- optim(
+   par = c(bin, frac),
+   fn = foo,
+   gr = NULL,
+   hessian = TRUE,
+   method = "BFGS",
+   control = list(fnscale = 1, trace = TRUE)
+ )
initial  value 1359.072766
iter  10 value 111.217950
final  value 111.089353
converged
> est_beta$par
[1] -6.5812951  1.5389089  1.2157189  1.4509610 -0.6442823  1.0826989  5.4700144
    -2.9389718 -1.4687920
[10] -1.5742866 -1.7448923 -1.9267416 -1.4995665 -3.0271371


##### Regressão Beta inflacionada em um

> ## Valores iniciais
> beta <- as.numeric(est_beta$par)
> mu.start <- beta[c(1,2,3,4,5,6,7)]
> alpha.start <- c(beta[c(8,11,12,13)])
> phi.start <- c(1,0.5,0.5,0.5,0.5)
> result <- reg.betaflacionada(mu.formula = y ~ financi + social + demo + ambi + ica460s
    + ginitotal,
+                    phi.formula = ~ factor(regiao),
+                    alpha.formula = ~  financi + social  + ginitotal,
+                    mu.start = mu.start,
+                    phi.start = phi.start,
+                    alpha.start = alpha.start,
```

```
+                     data = dados)
initial  value -2362.750370
iter  10 value -3212.787808
iter  20 value -3320.055117
final  value -3320.066550
converged
> result$`Regressão mu`
Estimado DesvioPadrao L_inferior L_superior P.valor
mu.(Intercept) -6.326978     0.104927  -6.532631  -6.121325       0
mu.financi      1.488579     0.046141   1.398145   1.579013       0
mu.social       1.180431     0.075292   1.032861   1.328001       0
mu.demo         1.398041     0.100237   1.201580   1.594501       0
mu.ambi        -0.888887     0.134136  -1.151789  -0.625985       0
mu.ica460s      1.218746     0.056270   1.108459   1.329033       0
mu.ginitotal    5.315120     0.107535   5.104355   5.525885       0
> result$`Regressão phi`
Estimado DesvioPadrao L_inferior L_superior  P.valor
phi.(Intercept)     1.658612     0.063430   1.534291   1.782933 0.000000
phi.factor(regiao)2 0.208031     0.072488   0.065958   0.350104 0.004106
phi.factor(regiao)3 0.645366     0.073627   0.501061   0.789672 0.000000
phi.factor(regiao)4 1.163005     0.075863   1.014317   1.311694 0.000000
phi.factor(regiao)5 0.187661     0.110713  -0.029333   0.404656 0.090071
> result$`Regressão alpha`
Estimado DesvioPadrao L_inferior L_superior  P.valor
alpha.(Intercept) -4.199657     0.786795  -5.741746  -2.657568 0.000000
alpha.financi     -1.398941     0.536159  -2.449794  -0.348088 0.009076
alpha.social      -2.546100     0.739431  -3.995359  -1.096841 0.000575
alpha.ginitotal    3.413724     0.989124   1.475076   5.352372 0.000558

> ##### GMM sem endogeneidade
> mu.formula = y ~ financi + social + demo + ambi + ica460s + ginitotal
> phi.formula = ~ factor(regiao)
> alpha.formula = ~ financi + social + ginitotal
>
> m <- model.frame(mu.formula, dados_pred)
> X_mu <- model.matrix(mu.formula, m)
> y <- model.response(m, "numeric")
>
> position <- which(y %in% 1)
>
> mm <- model.frame(phi.formula, dados_pred)
> X_phi <- model.matrix(phi.formula, mm)
>
```

```
> mmm <- model.frame(alpha.formula, dados_pred)
> X_alpha <- model.matrix(alpha.formula, mmm)
> phi.start <- c(1,0.5,0.5,0.5,0.5)
>
> names(mu.start) <- colnames(X_mu)
> names(phi.start) <- colnames(X_phi)
> names(alpha.start) <- colnames(X_alpha)
>
> parms <- c(mu = mu.start, alpha = alpha.start, phi = phi.start)
>
> kk <- length(parms)
> k0 <- length(mu.start)
> k1 <- k0 + length(alpha.start)
> k2 <- k1 + length(phi.start)
>
> yy <- dados$yy
>
> dados0 <- cbind(y, yy, X_mu, X_alpha, X_phi)
>
> kd <- dim(dados)[2]
> kd1 <- 1; kd2 <- 2
> kd3 <- kd2 + dim(X_mu)[2]
> kd4 <- kd3 + dim(X_alpha)[2]
> kd5 <- kd4 + dim(X_phi)[2]
>
> {
+   iv.moments <- function(z, data){
+     betas <- z[1:k0]
+     b_alfa_eta  <- z[c(k0 + 1):k1]
+     b_phi_eta <- z[c(k1 + 1):k2]
+
+       y       <- data.matrix(data[ ,            kd1])
+       yy      <- data.matrix(data[ ,            kd2])
+       X_mu    <- data.matrix(data[ , c(1 + kd2):kd3])
+       X_alpha <- data.matrix(data[ , c(1 + kd3):kd4])
+       X_phi   <- data.matrix(data[ , c(1 + kd4):kd5])
+
+       mu_X <- X_mu[-position,]
+       phi_X <- X_phi[-position,]
+       y_mu <- y[-position]
+
+
+       ###### Parte discreta
```

```
+          alfa_est  <- as.vector(X_alpha   %*%   b_alfa_eta)
+          alpha <- pnorm(alfa_est)
+
+          lalpha <- (yy/alpha -  (1 - yy)/(1 - alpha)) * dnorm(alfa_est)
+          Balpha <- X_alpha*as.vector(lalpha)
+
+          ###### Parte contínua
+          mu_eta     <- as.vector(mu_X    %*%   betas)
+          phi_eta    <- as.vector(phi_X   %*%   b_phi_eta)
+
+          phi <- pmax(exp(phi_eta), 1e-150)
+          mu <- pmax(( 1 / (1 + exp(-mu_eta))), 1e-150)
+          q_mu <- pmax((1 - mu), 1e-150)
+
+          lmu   <- phi*(digamma(q_mu*phi) - digamma(mu*phi) + log(y_mu/(1 - y_mu)))*(mu*
    (1 - mu))
+          lphi  <- (digamma(phi) - (1 - mu)*digamma(q_mu*phi) - mu*digamma(mu*phi) + mu*
    log(y_mu/(1 - y_mu)) + log(1 - y_mu))*phi
+
+          Bmu <- mu_X*as.vector(lmu)
+          Bphi <- phi_X*as.vector(lphi)
+
+          m_mu <- ( 1 / (1 + exp(- X_mu %*% betas)))
+          m_phi <- exp(X_phi %*% b_phi_eta)
+
+          m_1 <- y - ( alpha + (1 -  alpha)*m_mu)
+
+          new_Bmu<- matrix(0, nrow= dim(X_alpha)[1], ncol= dim(mu_X)[2])
+          new_Bmu[-position,] <- Bmu
+
+          new_Bphi <- matrix(0, nrow= dim(X_alpha)[1], ncol= dim(phi_X)[2])
+          new_Bphi[-position,] <- Bphi
+
+          gra <- cbind(new_Bmu, Balpha, new_Bphi, m_1)
+          colnames(gra) <- c(colnames(Bmu), colnames(Balpha), colnames(Bphi), 'm_1')
+
+      return(gra)
+    }
+ }
> my_gmm <- gmm(iv.moments,
x = dados0,
t0 = parms,
vcov= "iid",
```

```
type = "twoStep",
wmatrix = "optimal",
optfct = 'optim',
method = "BFGS",
control = list(trace = 1,
#reltol = 1e-25,
maxit = 200000))
>
> summary(my_gmm)

Call:
gmm(g = iv.moments, x = dados0, t0 = parms, type = "twoStep",
wmatrix = "optimal", vcov = "iid", optfct = "optim", method = "BFGS",
control = list(trace = 1, maxit = 2e+05))


Method:  twoStep

Coefficients:
                    Estimate      Std. Error    t value       Pr(>|t|)
mu.(Intercept)     -6.3292e+00    1.2048e-01   -5.2534e+01    0.0000e+00
mu.financi          1.4856e+00    6.8118e-02    2.1809e+01   1.9199e-105
mu.social           1.1822e+00    8.6037e-02    1.3740e+01    5.8170e-43
mu.demo             1.4004e+00    1.2264e-01    1.1418e+01    3.3858e-30
mu.ambi            -8.9105e-01    1.5298e-01   -5.8244e+00    5.7305e-09
mu.ica460s          1.2179e+00    7.0870e-02    1.7185e+01    3.4467e-66
mu.ginitotal        5.3204e+00    1.1986e-01    4.4388e+01    0.0000e+00
alpha.(Intercept)  -4.2638e+00    1.0652e+00   -4.0029e+00    6.2577e-05
alpha.financi      -1.3959e+00    5.1220e-01   -2.7253e+00    6.4251e-03
alpha.social       -2.5476e+00    9.8598e-01   -2.5838e+00    9.7726e-03
alpha.ginitotal     3.4918e+00    1.2100e+00    2.8859e+00    3.9034e-03
phi.(Intercept)     1.6717e+00    1.0710e-01    1.5609e+01    6.3694e-55
phi.factor(regiao)2 1.8904e-01    1.1082e-01    1.7058e+00    8.8039e-02
phi.factor(regiao)3 6.3435e-01    1.2115e-01    5.2362e+00    1.6392e-07
phi.factor(regiao)4 1.1450e+00    1.1372e-01    1.0069e+01    7.6002e-24
phi.factor(regiao)5 1.8046e-01    1.6198e-01    1.1141e+00    2.6524e-01

J-Test: degrees of freedom is 1
J-test   P-value
Test E(g)=0:    0.11336  0.73636

Initial values of the coefficients
mu.(Intercept)          mu.financi           mu.social             mu.demo
```

```
-6.3294901            1.4885941            1.1822757            1.3950020
mu.ambi           mu.ica460s         mu.ginitotal    alpha.(Intercept)
-0.8880022            1.2187585            5.3182292           -0.1630973
alpha.financi         alpha.social     alpha.ginitotal     phi.(Intercept)
-1.3332372           -1.4706778           -1.9168587            1.6560335
phi.factor(regiao)2 phi.factor(regiao)3 phi.factor(regiao)4 phi.factor(regiao)5
0.2111623            0.6481544            1.1659604            0.1896103

##############
Information related to the numerical optimization
Convergence code =  0
Function eval. =  1515
Gradian eval. =  1436

> ##### GMM com endogeneidade
> ##### Variáveis endógenas
> End = ~ lxterra + lxtrab + lxresto +  social + demo + ambi + ginitotal +
+     I(ambi*ginitotal) + I(ambi*social) + I(social*ginitotal) +
+     I(ambi*lxterra)
>
> m <- model.frame(End, dados)
> Z <- model.matrix(End, m)
>
> formula_mu = y ~ financi + social + demo +  ambi + ica460s + ginitotal
> mm <- model.frame(formula_mu, dados)
> X <- model.matrix(formula_mu, mm)
> y <- model.response(mm, 'numeric')
>
> ######## Parametro
> coef_mu <-est_beta$par[c(1,2,3,4,5,6,7)]
>
> names(coef_mu) <- colnames(X)
>
> parms <- coef_mu
>
> kk <- length(parms)
>
> kd <- 1 + dim(X)[2]
> kd0 <- kd + dim(Z)[2]
>
> {
+     function_beta <- function (z, X0){
+         b_mu <- z[1:kk]
```

```
+
+         mu  <- X0 %*% b_mu
+
+         p_alpha <- pnorm(mu)
+         p_mu <- (1/(1 + exp(-mu)))
+
+         y_est <- p_alpha + (1 - p_alpha)*p_mu
+
+         return(y_est)
+     }
+
+     iv.moments = function(parms,data) {
+         y0 <- data[,  1]
+         X0 <- data.matrix(data[,2:kd])
+         Z0 <- data.matrix(data[,c(1 + kd):kd0])
+
+         y_est <- function_beta(parms, X0)
+
+         erro_y <- as.vector(y0 - y_est)
+
+         results_y <-  Z0*erro_y
+
+         return(cbind(results_y))
+     }
+ }
>
> dat <- data.matrix(cbind(y, X, Z))
> data <- dat
> gmm.fit  <- gmm(iv.moments,
+                 x = dat,
+                 t0 = parms,
+                 vcov= "iid",
+                 type = "iterative",
+                 wmatrix = "optimal",
+                 optfct = 'optim',
+                 method = "BFGS",
+                 control = list(trace = 1,
+                                #reltol = 1e-25,
+                                maxit = 200000))
> summary(gmm.fit)


Call:
```

```
gmm(g = iv.moments, x = dat, t0 = parms, type = "iterative",
wmatrix = "optimal", vcov = "iid", optfct = "optim", method = "BFGS",
control = list(trace = 1, maxit = 2e+05))


Method:  iterative

Coefficients:
Estimate      Std. Error    t value       Pr(>|t|)
(Intercept)  -7.4587e+00    9.7994e-01   -7.6113e+00    2.7125e-14
financi       1.5609e-01    5.1259e-01    3.0451e-01    7.6074e-01
social       -1.9101e+00    6.0147e-01   -3.1757e+00    1.4950e-03
demo          8.1685e-01    1.7144e-01    4.7647e+00    1.8915e-06
ambi         -1.4552e+00    4.7509e-01   -3.0629e+00    2.1918e-03
ica460s       5.5609e+00    1.5041e+00    3.6971e+00    2.1805e-04
ginitotal     6.5957e+00    1.2262e+00    5.3789e+00    7.4929e-08

J-Test: degrees of freedom is 5
J-test    P-value
Test E(g)=0:    8.66580  0.12316

Initial values of the coefficients
(Intercept)     financi      social        demo        ambi     ica460s    ginitotal
-6.6238097   0.7930263  -0.9358687   1.2250186  -0.8628511   3.2285699   5.3637066

##############
Information related to the numerical optimization
Convergence code =  0
Function eval. =  2
Gradian eval. =  1

##### GMM via VI por bootstrap

DEA <- dados[, c(15,16,17,18) ]

dat <- data.matrix(cbind(DEA, X, Z))

boot_betaflacionadaGMM <- function(data, indices){

dados <- data[indices,]

input <- dados[,2:4]
output <- dados[,1]
```

```
e_vrs_out <- dea(input, output, RTS="vrs", ORIENTATION= "out", SLACK = F, DUAL=TRUE)
yest = 1/e_vrs_out$eff

parte2 <- cbind(dados[,-c(1:4)])
dados0 <- cbind(yest, parte2)

my_gmm0 <- gmm(iv.moments,
x = dados0,
t0 = parms,
vcov= "iid",
#type = "cue",
type = "iterative",
#type = "twoStep",
wmatrix = "optimal",
optfct = 'optim',
method = "BFGS",
control = list( #trace = 1,
maxit = 200000))

print(summary( my_gmm0)[[6]])
return(my_gmm0[[2]])

}

library(boot)
system.time(duncan.boot <- boot(dat, boot_betaflacionadaGMM, R = 2000))

###### Estimativa da regressão beta-inflacionada de um com mu e alpha tendo o mesmo
    preditor linear via correção de Murphy e Topel.

End <-  ~ lxterra + lxtrab + lxresto +  social + demo + ambi + ginitotal +
I(ambi*ginitotal) + I(ambi*social) + I(social*ginitotal) +
I(ambi*lxterra)

mmm <- model.frame(End, dados)
X_end <- model.matrix(End, mmm)

kd <- dim(X_end)[2]

#### Financit
End = log(financit + 1) ~ lxterra + lxtrab + lxresto +  social + demo + ambi + ginitotal
    +
```

```
I(ambi*ginitotal) + I(ambi*social) + I(social*ginitotal) +
I(ambi*lxterra)

reg <- lm(End, data = dados)
fin_pred <- exp(predict(reg)) - 1

fin_predK <- rank(fin_pred)/4965

##### ica
End0 = ica460s ~ lxterra + lxtrab + lxresto +  social + demo + ambi + ginitotal +
I(ambi*ginitotal) + I(ambi*social) + I(social*ginitotal) +
I(ambi*lxterra)

reg_glm <- glm(End0, data = dados, family = quasibinomial(link = "logit"))

pred0 <- predict(reg_glm)
ica_pred <- 1/ (1 + exp(-pred0))

#### Regressão Beta inflacionada de um

result0 <- reg.betaflacionada(
+   mu.formula = y ~ fin_predK + social + demo + ambi + ica_pred + ginitotal,
+   phi.formula = ~ factor(regiao),
+   mu.start = mu.start,
+   phi.start = phi.start,
+   data = dadosA)
initial  value 1402.759518
iter  10 value 1013.996138
final  value 805.582761
converged
>
> result0
$`Regressão mu`
Estimado DesvioPadrao L_inferior L_superior  P.valor
mu.(Intercept) -5.306414     0.131807  -5.564752  -5.048076 0.000000
mu.fin_predK    0.559112     0.156420   0.252534   0.865689 0.000351
mu.social      -1.170647     0.189540  -1.542139  -0.799155 0.000000
mu.demo         0.538156     0.122730   0.297609   0.778703 0.000012
mu.ambi        -0.349520     0.173487  -0.689547  -0.009492 0.043939
mu.ica_pred     3.346761     0.380975   2.600065   4.093458 0.000000
mu.ginitotal    4.372922     0.185737   4.008885   4.736959 0.000000

$`Regressão phi`
```

```
Estimado DesvioPadrao L_inferior L_superior  P.valor
phi.(Intercept)      1.753769     0.065480   1.625432   1.882107 0.000000
phi.factor(regiao)2  0.127154     0.073006  -0.015935   0.270242 0.081562
phi.factor(regiao)3 -0.411916     0.077804  -0.564408  -0.259424 0.000000
phi.factor(regiao)4 -0.042460     0.083351  -0.205826   0.120905 0.610461
phi.factor(regiao)5 -0.628015     0.109744  -0.843109  -0.412921 0.000000

$ParÃ¢metro
mu.(Intercept)        mu.fin_predK           mu.social             mu.demo             mu
    .ambi         mu.ica_pred        mu.ginitotal
-5.30641417          0.55911185         -1.17064703          0.53815605
    -0.34951964          3.34676125          4.37292179
phi.(Intercept) phi.factor(regiao)2 phi.factor(regiao)3 phi.factor(regiao)4 phi.factor(
    regiao)5
1.75376941          0.12715359         -0.41191575         -0.04246048
    -0.62801504

$matriz_cov
mu.(Intercept)  mu.fin_predK     mu.social       mu.demo       mu.ambi    mu.ica_pred  mu.
    ginitotal phi.(Intercept)
mu.(Intercept)        0.0173731936  0.0101684115  0.0129183688  0.0020697622  1.558591e
    -04 -0.0277358545 -0.0203543108   -1.088364e-04
mu.fin_predK          0.0101684115  0.0244671922  0.0218798900  0.0074679048  1.493447e
    -02 -0.0554208192 -0.0213636575    2.492262e-04
mu.social             0.0129183688  0.0218798900  0.0359255788  0.0039492684  1.067519e
    -02 -0.0641609840 -0.0216349531   -8.596101e-04
mu.demo               0.0020697622  0.0074679048  0.0039492684  0.0150627482  4.305424e
    -03 -0.0190728954 -0.0099421970    2.545659e-04
mu.ambi               0.0001558591  0.0149344737  0.0106751935  0.0043054237  3.009766e
    -02 -0.0367027627 -0.0153983816    8.118501e-05
mu.ica_pred          -0.0277358545 -0.0554208192 -0.0641609840 -0.0190728954 -3.670276e
    -02  0.1451416002  0.0547478077    3.505455e-04
mu.ginitotal         -0.0203543108 -0.0213636575 -0.0216349531 -0.0099421970 -1.539838e
    -02  0.0547478077  0.0344980534    2.531352e-04
phi.(Intercept)      -0.0001088364  0.0002492262 -0.0008596101  0.0002545659  8.118501e
    -05  0.0003505455  0.0002531352    4.287574e-03
phi.factor(regiao)2  -0.0012753455 -0.0006545799 -0.0001862083 -0.0001387439 -2.715276e
    -04  0.0015103022  0.0013066288   -4.240939e-03
phi.factor(regiao)3  -0.0007798772  0.0004721471  0.0020568612  0.0002009107 -1.157107e
    -03 -0.0008408368  0.0007486109   -4.164853e-03
phi.factor(regiao)4  -0.0002062179  0.0008796323  0.0024942338 -0.0007847281  1.382981e
    -03 -0.0017610211 -0.0008349056   -4.171927e-03
```

```
phi.factor(regiao)5  -0.0018303434  0.0003826695 -0.0018992158  0.0017565491  4.619144e
    -04  0.0021953689  0.0007096974   -3.922775e-03
phi.factor(regiao)2 phi.factor(regiao)3 phi.factor(regiao)4 phi.factor(regiao)5
mu.(Intercept)            -0.0012753455       -0.0007798772       -0.0002062179
    -0.0018303434
mu.fin_predK              -0.0006545799        0.0004721471        0.0008796323
    0.0003826695
mu.social                 -0.0001862083        0.0020568612        0.0024942338
    -0.0018992158
mu.demo                   -0.0001387439        0.0002009107       -0.0007847281
    0.0017565491
mu.ambi                   -0.0002715276       -0.0011571066        0.0013829815
    0.0004619144
mu.ica_pred                0.0015103022       -0.0008408368       -0.0017610211
    0.0021953689
mu.ginitotal               0.0013066288        0.0007486109       -0.0008349056
    0.0007096974
phi.(Intercept)           -0.0042409391       -0.0041648526       -0.0041719268
    -0.0039227751
phi.factor(regiao)2        0.0053298458        0.0042932290        0.0041991242
    0.0042030227
phi.factor(regiao)3        0.0042932290        0.0060533879        0.0049711818
    0.0046011925
phi.factor(regiao)4        0.0041991242        0.0049711818        0.0069474319
    0.0045970709
phi.factor(regiao)5        0.0042030227        0.0046011925        0.0045970709
    0.0120437036

$Gradiente
[1]  0.0041456224  0.0048586994  0.0027477471  0.0027386925  0.0027797382  0.0036837428
    0.0035008209 -0.0006979026 -0.0011815650 -0.0017829132
[11] -0.0007263384 -0.0010571586


##### Correção de Murphy e Topel.

> mu.formula = y ~ fin_predK + social + demo + ambi + ica_pred + ginitotal
> phi.formula = ~ factor(regiao)
>
> m <- model.frame(mu.formula, dadosA)
> X_mu <- model.matrix(mu.formula, m)
> y <- model.response(m, "numeric")
>
```

```
> position <- which(y %in% 1)
> yy <- replace(y, y != 1, 0)
>
> mm <- model.frame(phi.formula, dadosA)
> X_phi <- model.matrix(phi.formula, mm)
>
> mu_X <- X_mu[-position,]
> phi_X <- X_phi[-position,]
> y_mu <- y[-position]
>
> y_ica <- dados$ica460s
> y_fin <-  log(dados$financit + 1)
>
>
> {
+   lG22 <- function(z) {
+     betas <- z[1:k0]
+     b_phi_eta <- z[(k0 + 1):kk]
+
+     y_eta_total  <- as.vector(X_mu    %*%   betas)
+     phi_eta   <- pmin(as.vector(phi_X   %*%   b_phi_eta), 700)
+
+     y_eta <- y_eta_total[-position]
+
+     ###### Parte discreta
+
+     alpha <- pnorm(y_eta_total)
+
+     lalpha <- (yy/alpha -  (1 - yy)/ (1 - alpha))*dnorm(y_eta_total)
+     Balpha <- X_mu*as.vector(lalpha)
+
+     ###### Parte contínua
+
+     phi <- pmax(exp(phi_eta), 1e-150)
+     mu <- pmax(( 1 / (1 + exp(-y_eta))), 1e-150)
+     q_mu <- pmax((1 - mu), 1e-150)
+
+     lmu <- phi*(digamma(q_mu*phi) - digamma(mu*phi) + log(y_mu/(1 - y_mu)))*(mu*q_mu)
+     lphi <- (digamma(phi) - q_mu*digamma(q_mu*phi) - mu*digamma(mu*phi) + mu*log(y_mu/
    (1 - y_mu)) + log(1 - y_mu))*phi
+
+     Bmu <- mu_X*as.vector(lmu)
+     Bphi <- phi_X*as.vector(lphi)
```

```
+
+     new_Bmu<- matrix(0, nrow= dim(X_mu)[1], ncol= dim(mu_X)[2])
+     new_Bmu[-position,] <- Bmu
+
+     new_Bphi <- matrix(0, nrow= dim(X_mu)[1], ncol= dim(phi_X)[2])
+     new_Bphi[-position,] <- Bphi
+
+     colnames(new_Bphi) <- colnames(Bphi)
+
+     Btotal <- new_Bmu + Balpha ### Vetor de parÃ¢metro
+
+     gra <- cbind(Btotal, new_Bphi)
+
+     return(-gra)
+   }
+
+   lG21 <- function(z) {
+     b_mu    <- z[1:k0]
+     b_phi   <- z[(k0 + 1):kk]
+
+     X_fin <- X_end
+     X_ica <- X_end
+
+     ###### Parte discreta
+     y_eta_total     <- as.vector(X_mu    %*%   b_mu)
+     phi_eta   <- as.vector(phi_X   %*%   b_phi)
+
+     mu_eta <- y_eta_total[-position]
+
+     phi <- pmax(exp(phi_eta), 1e-150)
+     mu <- pmax(( 1 / (1 + exp(-mu_eta))), 1e-150)
+     alpha <- pnorm(y_eta_total)
+     q_mu <- pmax((1 - mu), 1e-150)
+
+     ############ Verossimilhança
+     mu_a <- digamma(mu*phi) - digamma(q_mu*phi)
+     y_a  <- log(y_mu/(1 - y_mu))
+
+     lalpha <- (yy/alpha -  (1 - yy)/ (1 - alpha))*dnorm(y_eta_total)
+     lmu <- phi*(y_a - mu_a)*(mu*q_mu)
+
+     Lmu <- matrix(0, nrow= dim(X_mu)[1], ncol= 1)
+     Lmu[-position,] <- lmu
```

```
+
+     mu0 <- matrix(0, nrow= dim(X_mu)[1], ncol= 1)
+     mu0[-position,] <- mu
+
+     g_fin <- X_fin * as.vector(lalpha*b_mu[2] + Lmu*b_mu[2]) ### Regressão linear
+     g_ica <- X_ica * as.vector(mu0*(1 - mu0)*(lalpha*b_mu[6] +  Lmu*b_mu[6])) ###
    Regressão logistica
+
+     gra <- cbind(g_fin, g_ica)
+     colnames(gra) <- c(colnames(X_fin), colnames(X_ica))
+
+     return(gra)
+   }
+
+   G1_ica <- function(z){
+     betas <- z[1:kd]
+     betas <- reg_glm$coefficients
+
+     X_ica <- X_end
+
+     eta <- as.vector(X_ica %*% betas)
+     y_est<- (1/(1 + exp(-(eta))))
+
+     erro <- as.vector(y_ica - y_est)
+     g <- X_ica*erro
+     return(g)
+   }
+
+   G1_fin <- -2 * X_end * as.vector(y_fin - predict(reg))
+
+ }
>
> G22 <- lG22(result0$ParÃ¢metro)
>
> G21 <- lG21(result0$ParÃ¢metro)
> G_ica <- G1_ica(reg_glm$coefficients)
>
> V0 <- vcov(reg)
> V00 <- vcov(reg_glm)
>
> library(Matrix)
>
> V1 <- bdiag(V0,V00)
```

```
> V2 <- result0$matriz_cov
> G11 <- cbind(G1_fin, G_ica)
>
> #### Corrigido
>
> C <- t(G22) %*% G21
> R <- t(G22) %*% G11
>
> V33 <- ((C %*% V1 %*% t(C)) - (R %*% V1 %*% t(C)) - (C %*% V1 %*% t(R)))
>
> V2_corrigida <- V2 + V2 %*% V33 %*% V2
>  V2_corrigida
12 x 12 Matrix of class "dgeMatrix"
mu.(Intercept)  mu.fin_predK    mu.social       mu.demo       mu.ambi  mu.ica_pred  mu.
    ginitotal
mu.(Intercept)        0.0346439730  0.0262417202  0.039572669  0.0105809042  0.0104211459
     -0.078781802 -0.0495633429
mu.fin_predK          0.0262417202  0.0571098534  0.060780168  0.0181818529  0.0439856797
     -0.139546307 -0.0608834926
mu.social             0.0395726693  0.0607801685  0.102497129  0.0180801199  0.0446679739
     -0.190159653 -0.0756016215
mu.demo               0.0105809042  0.0181818529  0.018080120  0.0301765991  0.0140875395
     -0.051732096 -0.0326136165
mu.ambi               0.0104211459  0.0439856797  0.044667974  0.0140875395  0.0699219433
     -0.114085077 -0.0528043381
mu.ica_pred          -0.0787818015 -0.1395463074 -0.190159653 -0.0517320965 -0.1140850771
      0.397713928  0.1678037161
mu.ginitotal         -0.0495633429 -0.0608834926 -0.075601622 -0.0326136165 -0.0528043381
      0.167803716  0.0996403054
phi.(Intercept)       0.0001008043  0.0005319104 -0.003227338  0.0002317609 -0.0002336374
      0.003258712 -0.0003029426
phi.factor(regiao)2   0.0002961960  0.0027016896  0.005637645  0.0029227427  0.0029130001
     -0.009510677 -0.0037811809
phi.factor(regiao)3  -0.0021619794 -0.0006835914  0.001149651 -0.0010790218 -0.0023181443
      0.001904494  0.0036695241
phi.factor(regiao)4   0.0006217496  0.0019765330  0.004352944 -0.0017746083  0.0023992769
     -0.004744128 -0.0019191832
phi.factor(regiao)5   0.0004930353  0.0070611968  0.012413056  0.0005547376  0.0061063416
     -0.022189780 -0.0032250448
phi.(Intercept) phi.factor(regiao)2 phi.factor(regiao)3 phi.factor(regiao)4 phi.factor(
    regiao)5
mu.(Intercept)         0.0001008043         0.000296196       -0.0021619794
    0.0006217496        0.0004930353
```

```
mu.fin_predK            0.0005319104          0.002701690        -0.0006835914
    0.0019765330         0.0070611968
mu.social              -0.0032273380          0.005637645         0.0011496514
    0.0043529444         0.0124130562
mu.demo                 0.0002317609          0.002922743        -0.0010790218
    -0.0017746083         0.0005547376
mu.ambi                -0.0002336374          0.002913000        -0.0023181443
    0.0023992769         0.0061063416
mu.ica_pred             0.0032587121         -0.009510677         0.0019044944
    -0.0047441278        -0.0221897801
mu.ginitotal           -0.0003029426         -0.003781181         0.0036695241
    -0.0019191832        -0.0032250448
phi.(Intercept)         0.0030259100         -0.002652823        -0.0026752187
    -0.0028107603        -0.0030784840
phi.factor(regiao)2    -0.0026528233          0.004423824         0.0025589456
    0.0027849519         0.0026608614
phi.factor(regiao)3    -0.0026752187          0.002558946         0.0048563591
    0.0037290662         0.0035684115
phi.factor(regiao)4    -0.0028107603          0.002784952         0.0037290662
    0.0064031051         0.0042852614
phi.factor(regiao)5    -0.0030784840          0.002660861         0.0035684115
    0.0042852614         0.0185133627
> EP_corrigido <- sqrt(diag(V2_corrigida))
> EP_corrigido
[1] 0.18612891 0.23897668 0.32015173 0.17371413 0.26442758 0.63064564 0.31565853
    0.05500827 0.06651183
[10] 0.06968758 0.08001940 0.13606382

##### Correção de Murphy e Topel por Bootstrap

{
boot_betaflacionada <- function(data, indices){

dados <- data[indices,]

input <- dados[,16:18]
output <- dados[,15]

e_vrs_out <- dea(input, output, RTS="vrs", ORIENTATION= "out", SLACK = F, DUAL=TRUE)
yest = 1/e_vrs_out$eff

dados0 <- cbind(dados, yest)
```

```
reg <- lm(End, data = dados)
fin_pred <- exp(predict(reg)) - 1
fin_predK <- rank(fin_pred)/4965

reg_glm <- glm(End0, data = dados, family = quasibinomial(link = "logit"))
pred0 <- predict(reg_glm)
ica_pred <- 1/ (1 + exp(-pred0))

dados00 <- cbind(dados0, fin_predK, ica_pred)

result0 <- reg.betaflacionada(
mu.formula = yest ~ fin_predK + social + demo + ambi + ica_pred + ginitotal,
phi.formula = ~ factor(regiao),
mu.start = mu.start,
phi.start = phi.start,
data = dados00)

print(result0$ParÃ¢metro)

return(result0$ParÃ¢metro)

}
}

library(boot)
system.time(duncan.boot <- boot(dados, boot_betaflacionada, R = 2000))
```

Programa C.1: Códigos R