Modelos de regressão beta inflacionados truncados

Gustavo Henrique de Araujo Pereira

Tese apresentada
ao
Instituto de Matemática e Estatística
da
Universidade de São Paulo
para
obtenção do título
de
Doutor em Ciências

Programa: Estatística

Orientadora: Profa. Dra. Denise Aparecida Botter

Coorientadora: Profa. Dra. Mônica Carneiro Sandoval

Durante o desenvolvimento deste trabalho o autor recebeu
auxílio financeiro da CAPES e do CNPq.

São Paulo, julho de 2012

Modelos de regressão beta inflacionados truncados

Esta versão da tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 24/05/2012. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Profa. Dra. Denise Aparecida Botter (orientadora) - IME-USP
- Profa. Dra. Silvia Lopes de Paula Ferrari - IME-USP
- Prof. Dr. Klaus Leite Pinto Vasconcellos - UFPE
- Prof. Dr. Raydonal Ospina Martínez - UFPE
- Prof. Dr. Rinaldo Artes - Insper

"O sofrimento é passageiro,
desistir é para sempre".
(Lance Armstrong)

À minha esposa Patricia e aos meus
filhos Breno Gustavo e Ana Júlia.

Agradecimentos

Agradeço a Deus por ter me dado coragem para deixar minhas atividades profissionais para fazer o doutorado e também às seguintes pessoas:

Às minhas orientadoras Denise Botter e Mônica Sandoval que durante todo o meu doutorado estiveram sempre disponíveis para ajudar, para incentivar, para ler e reler cada parte do trabalho e para sugerir correções, modificações e complementações para a tese, sem nunca restringir a minha liberdade criativa.

À minha esposa Patricia pela paciência e amor que teve durante todo o desenvolvimento deste trabalho, me apoiando mesmo nos finais de semana em que tive que passar menos momentos com a família para finalizar a tese.

Aos meus filhos Breno Gustavo e Ana Júlia, que com sorrisos me incentivaram a desenvolver este trabalho da melhor forma que eu podia.

Aos meus pais Adevaldes e Clélia que desde cedo me ensinaram o valor da educação. À minha mãe também por tantas vezes ter cuidado do Breno e da Ana para que eu pudesse me dedicar mais ao desenvolvimento deste trabalho.

Às minhas irmãs Priscila e Tarsila que sempre apoiaram meus estudos. À Tarsila também pela revisão deste trabalho.

Aos meus sogros Romão e Odila pelos cuidados com meus filhos que sempre me deixaram mais tranquilo para o desenvolvimento desta tese.

Aos professores Marcos Magalhães e Rinaldo Artes que, ao me orientarem na iniciação científica e no mestrado, respectivamente, foram fundamentais na minha decisão de fazer carreira acadêmica.

Aos professores Silvia Ferrari e Raydonal Ospina pelos trabalhos com os modelos de regressão beta inflacionados que facilitaram o desenvolvimento desta tese.

Aos amigos Marco Riquelme e Cibele Russo que desde o início do doutorado me ajudaram, me incentivaram, me deram sugestões e contribuíram para que os anos de desenvolvimento deste trabalho fossem tão interessantes.

À amiga Marcela Bussadori pela revisão efetuada neste trabalho.

Aos amigos Mariana Melo, Tiago Magalhães, Alejandra Tapia, Francyelle Silva, Luz Marina Goméz, Nubia Duarte, Karin Tamura e Frederico Poleto pelas diversas conversas que incentivaram, distraíram e tornaram melhor o período de desenvolvimento deste trabalho.

À todos aqueles que foram alunos de pós-graduação do IME-USP comigo e com quem tive uma ótima convivência durante o doutorado: Eliane, Jalmar, Lizandra, Marcelo, Cristian, Germán, Michel, Amanda, Rafael, Renato, Tatiana, Hugo, Liliane, Arthur, Alexandre, Gleyce, João, José, Victor, Mirian, Guillermo, Rodrigo, Ivan, Valdivino, Tatiane e Fernando.

Ao grupo de professores de estatística do Insper: Kelly, Adriana, Sérgio, Tatiana, Rinaldo e Priscila, que me apoiaram neste trabalho e me ajudaram e continuam me ajudando a me desenvolver como professor.

Ao profissional e à instituição financeira que disponibilizaram os dados de cartão de crédito utilizados neste trabalho.

À Gabrielle Souza e Márcio Borges do Ministério do Trabalho e Emprego que disponibilizaram os dados de seguro desemprego.

À todos os professores do IME-USP que contribuíram com minha formação acadêmica desde a graduação ao doutorado.

Aos funcionários do IME-USP que tornaram esses anos na faculdade um pouco mais fáceis.

À Capes e ao CNPq pelo apoio financeiro.

À banca examinadora pelas sugestões que ajudaram a melhorar este trabalho.

Resumo

Os modelos de regressão beta e beta inflacionados conseguem ajustar adequadamente grande parte das variáveis do tipo proporção. No entanto, esses modelos não são úteis quando a variável resposta não pode assumir valores no intervalo $(0, c)$ e assume o valor c com probabilidade positiva. Variáveis relacionadas a algum tipo de pagamento limitado entre dois valores, quando estudadas em relação ao seu valor máximo, possuem essas características. Para ajustar essas variáveis, introduzimos a distribuição beta inflacionada truncada (BIZUT), que é uma mistura de uma distribuição beta com suporte no intervalo $(c, 1)$ e uma distribuição trinomial que assume os valores zero, um e c. Propomos ainda um modelo de regressão para as situações em que a variável resposta tem distribuição BIZUT. Admitimos que todos os parâmetros da distribuição podem variar em função de variáveis preditoras. Além disso, o modelo permite que o parâmetro conhecido c varie entre as unidades populacionais. Para esse modelo são desenvolvidos diversos aspectos inferenciais, são obtidos resultados para as situações em que c é variável e são conduzidos estudos de simulação de Monte Carlo. Além disso, discutimos análise de resíduos, desenvolvemos análise de influência local e realizamos uma aplicação a dados reais de cartão de crédito.

Palavras-chave: cartão de crédito, estimador de máxima verossimilhança, modelos inflacionados, modelos de regressão beta inflacionados truncados, proporções, regressão beta.

Abstract

The beta regression model or the inflated beta regression model may be a reasonable choice to fit a proportion in most situations. However, they do not fit well variables that do not assume values in the open interval $(0, c)$, $0 < c < 1$ and assume the c value with positive probability. Variables related to a kind of double bounded payment amount when studied as a proportion of the maximum payment amount have this feature. For these variables, we introduce the truncated inflated beta distribution (TBEINF). This proposed distribution is a mixture of the beta distribution bounded in the open interval $(c, 1)$ and a trinomial distribution that assumes the values zero, one and c. This work also proposes a regression model where the response variable is TBEINF distributed. The model allows all the unknown parameters of the conditional distribution of the response variable to be modeled as functions of explanatory variables. Moreover, the model allows nonconstant known parameter c across population units. For this model, some inferential aspects are developed, some results when c is not constant are obtained and Monte Carlo simulation studies are performed. In addition, residual and local influence analysis are discussed and an application to credit card data is presented.

Keywords: beta regression, credit card, inflated models, maximum likelihood estimator, proportions, truncated inflated beta regression.

Sumário

1 Introdução **1**
1.1 Objetivo e organização da tese 6

2 Distribuição beta inflacionada truncada **8**
2.1 Distribuição beta 8
2.2 Distribuição beta inflacionada 9
2.3 Distribuição beta inflacionada truncada 9
2.4 Variáveis inflacionadas em mais pontos 15
2.5 Estudos de simulação e comparação das distribuições BIZU e BIZUT 15
2.6 Aplicação 18
2.7 Conclusão 21

3 Modelos de regressão beta inflacionados truncados **23**
3.1 Definição e estimação pontual 23
3.2 Intervalos de confiança e testes de hipóteses 32
3.3 Modelo RBIZUT com c variável 36
3.3.1 Resultados para o modelo RBIZUT com c variável 36
3.3.2 Endogeneidade 38
3.4 Estudos de simulação 39
3.5 Seleção de modelos 46
3.6 Conclusão 48

4 Análise de diagnóstico **49**
4.1 Análise de resíduos e teste de qualidade de ajuste 49
4.2 Análise de influência local 55
4.2.1 Influência local no modelo RBIZUT 57
4.3 Conclusão 62

5 Aplicação **63**
5.1 Aplicação a dados reais de cartão de crédito 63
5.2 Conclusão 74

6 Conclusão **76**
6.1 Trabalhos futuros 77

A Provas dos resultados do Capítulo 2 **78**
A.1 Prova dos resultados dos momentos da distribuição beta 78
A.2 Prova dos resultados dos momentos da distribuição BIZUT 79
A.3 Matriz de informação de Fisher da distribuição BIZUT 79
A.4 Distribuição assintótica do estimador de máxima verossimilhança de $\mathrm{E}(Y)$ 80

B Provas dos resultados do Capítulo 3 **82**
B.1 Momentos de y_i^* e $y_i^\bullet$ 82
B.2 Obtenção da função escore para o modelo RBIZUT 82
B.3 Obtenção da matriz de informação de Fisher 85
B.4 Obtenção da partição do vetor escore 91
B.5 Prova dos resultados para o modelo RBIZUT com c variável 92
B.5.1 Prova da Proposição 3.3.1 92
B.5.2 Prova da Proposição 3.3.2 92
B.5.3 Prova da Proposição 3.3.3 93

C Provas dos resultados do Capítulo 4 **94**
C.1 Obtenção de Δ para a ponderação de casos 94
C.2 Obtenção de Δ para a perturbação individual de variáveis preditoras 95

Referências Bibliográficas **98**

Capítulo 1

Introdução

O estudo de taxas, razões e proporções é comum em diversas áreas do conhecimento. Muitas vezes essas variáveis variam no intervalo $(0, 1)$. Frequentemente o objetivo do trabalho é o estudo dessas variáveis em função de algumas outras através de um modelo de regressão. Diversas abordagens podem ser utilizadas nessas situações. A utilização de regressão linear normal não é adequada, já que essas variáveis, por serem limitadas, não seguem essa distribuição. Além disso, valores ajustados para essas variáveis através de um modelo de regressão linear não necessariamente pertencem ao intervalo $(0, 1)$. A forma mais simples de contornar o problema é efetuar uma transformação na variável resposta para que ela assuma valores na reta real e ajustar um modelo de regressão linear para a variável transformada. Segundo Kieschnick e McCullough (2003) a transformação mais utilizada para variáveis que assumem valores no intervalo $(0, 1)$ é a logito, definida como $g(y) = \log[y/(1-y)]$. Ospina (2008, página 4) cita diversas outras e Atkinson (1985, Capítulo 7) discute famílias de transformações que envolvem um parâmetro desconhecido. A utilização de transformações tem pelo menos dois inconvenientes. O primeiro é que elas alteram a natureza dos dados, dificultando assim a interpretação dos parâmetros do modelo. Além disso, o uso de uma transformação não garante que a nova variável tenha distribuição normal e seja homocedástica, que são suposições do modelo de regressão linear normal.

Uma alternativa à transformação da variável resposta é o ajuste de um modelo de regressão não linear que assume que a variável dependente é igual a uma função não linear das variáveis preditoras e de seus respectivos parâmetros mais um erro suposto normal homocedástico (Hermalin e Wallace, 1994; Kieschnick e McCullough, 2003). No entanto, a suposição de que o erro tem distribuição normal não é razoável, já que nesse caso ele necessariamente está contido no intervalo $(-1, 1)$. Uma abordagem mais interessante é a utilização de modelos de quase-verossimilhança introduzidos por Wedderburn (1974). Para o ajuste desses modelos não é necessária a suposição de que a variável resposta tenha alguma distribuição conhecida. O modelo requer apenas que se especifique uma relação entre a variância e a média da variável dependente. Cox (1996) e Papke e Wooldridge (1996) propuseram modelos de quase-verossimilhança para o ajuste de variáveis do tipo proporção. Diferentemente dos modelos já citados, não há importantes restrições ao uso desses modelos de quase-verossimilhança para o ajuste dessas variáveis. No entanto, segundo Kieschnick e McCullough (2003), para variáveis respostas do tipo proporção e amostras não muito grandes, a

performance desses modelos parece ser inferior à da regressão beta (que será discutida a seguir).

Uma forma mais natural de ajustar um modelo de regressão para variáveis que variam no intervalo $(0, 1)$ é supor que a variável resposta tem alguma distribuição conhecida com suporte no intervalo $(0, 1)$. Já foram desenvolvidos modelos de regressão para variáveis com distribuição beta (Ferrari e Cribari-Neto, 2004; Paolino, 2001; Kieschnick e McCullough, 2003) e simplex (Jorgensen, 1997*a*) e proposta uma nova parametrização da distribuição de Kumaraswamy para o desenvolvimento de modelos de regressão (Mitnik e Baek, 2012). As três distribuições de probabilidade citadas são descritas respectivamente em Johnson et al. (1995, Capítulo 25), Barndorff-Nielsen e Jorgensen (1991) e Kumaraswamy (1980). O modelo de regressão simplex tem a vantagem, em relação à regressão beta, de pertencer à classe de modelos de dispersão (Jorgensen, 1997*b*), que pode ser vista como uma extensão dos modelos lineares generalizados (McCullagh e Nelder, 1989). Assim a regressão simplex pode utilizar os resultados já desenvolvidos para essa classe de modelos. No entanto, em sua versão mais geral, a distribuição simplex tem o inconveniente de possuir quatro parâmetros enquanto a distribuição beta tem apenas dois. Além disso, em sua versão com dois parâmetros, o modelo de regressão simplex parece não ser flexível o suficiente para produzir, em situações reais, um ajuste de qualidade equivalente ao obtido com a regressão beta (Kieschnick e McCullough, 2003; Miyashiro, 2008). Aprofundaremos a discussão a respeito da regressão beta porque esta tese está diretamente relacionada com esse modelo.

Ferrari e Cribari-Neto (2004) desenvolveram a especificação mais estruturada e melhor formalizada da regressão beta. Para facilitar a interpretação dos parâmetros do modelo, os autores propuseram uma reparametrização da distribuição beta de modo que um de seus parâmetros fosse a média e o outro pudesse ser interpretado como um parâmetro de precisão. Assim, o modelo supõe que a distribuição da variável resposta é beta e que a resposta média está relacionada com um preditor linear a partir de uma função de ligação. Com essa especificação, em vários aspectos, a regressão beta é semelhante aos modelos lineares generalizados (McCullagh e Nelder, 1989). A distribuição beta é bastante flexível, podendo assumir diversos formatos simétricos e assiméticos, incluindo os formatos de 'J', 'U' e 'J' invertido. Dessa forma, em geral, pode ser bem razoável admitir que uma variável que assume valores no intervalo $(0, 1)$ tem distribuição beta.

Juntamente com o trabalho de Ferrari e Cribari-Neto (2004), dois outros artigos introduziram a regressão beta. Paolino (2001) comparou a regressão beta com modelos de regressão linear com e sem transformação logito da variável resposta. Utilizando dados simulados e reais, ele concluiu que a regressão beta apresenta vantagens significativas em relação à regressão linear para o estudo de variáveis resposta que assumem valores no intervalo $(0, 1)$. Kieschnick e McCullough (2003) compararam a performance de um modelo de regressão beta com ligação logito com as de diversos outros modelos usualmente utilizados na prática a partir de uma variação do critério de informação de Akaike proposta por McQuarrie e Tsai (1998). Em ambos os conjuntos de dados referentes a estudos de proporções, o modelo de regressão beta produziu um ajuste superior aos dos demais, segundo esse critério.

Após o artigo de Ferrari e Cribari-Neto (2004), vários trabalhos relacionados ao modelo de re-

gressão beta foram desenvolvidos e um pacote no software R foi criado para o ajuste do modelo (Cribari-Neto e Zeileis, 2010). Ospina et al. (2006) obtiveram estimadores de máxima verossimilhança dos parâmetros do modelo de regressão beta com correção do viés (ver também Ospina et al. (2011)). Já Espinheira et al. (2008*a*,*b*) introduziram novos resíduos e medidas de influência local. Para a mesma classe de modelos, Ferrari e Pinheiro (2011) desenvolveram correções para pequenas amostras do teste da razão de verossimilhanças. Andrade (2007) estudou o efeito da especificação incorreta da função de ligação em modelos de regressão beta. Miyashiro (2008) comparou os modelos de regressão beta e simplex em dois conjuntos de dados, observando que em ambos a regressão beta possibilitou um melhor ajuste. O modelo de regressão beta foi estendido por Smithson e Verkuilen (2006) para permitir que o parâmetro de precisão da distribuição beta também varie em função de variáveis preditoras e por Simas et al. (2010), que introduziram uma versão não linear do modelo. Cribari-Neto e Souza (2011) desenvolveram testes de hipóteses para os parâmetros que ajustam a média da variável resposta sem a necessidade de ajustar o parâmetro de precisão. Já Venezuela (2008) estendeu o modelo para situações em que há dependência entre as observações, em um contexto de equações de estimação generalizadas (Hardin e Hilbe, 2003), enquanto Souza (2011) e Melo et al. (2009) propuseram modelos de regressão beta multivariados. O modelo de regressão beta foi ainda desenvolvido do ponto de vista bayesiano por Branscum et al. (2007).

Em algumas situações, a variável de interesse pode assumir não apenas valores no intervalo $(0, 1)$ como também os valores zero ou um ou ambos. Nessas situações, o modelo de regressão beta não é adequado, já que a variável resposta é discreta nos pontos zero e/ou um e contínua no intervalo $(0, 1)$. A regressão linear com transformação logito ou outras transformações comumente utilizadas para dados de proporções não podem ser utilizadas porque não estão definidas para o valor zero. Há transformações que estão definidas para o valor zero e um (Piepho, 2003), mas a regressão linear com essas transformações para o ajuste de variáveis que assumem valores no intervalo $[0, 1]$ apresentam os mesmos problemas já discutidos para variáveis com suporte em $(0, 1)$. A regressão não linear da forma discutida anteriormente também não é recomendada porque a existência de valores discretos tende a acarretar que a distribuição dos erros seja ainda mais distante da distribuição normal. Os modelos de quase-verossimilhança podem ser utilizados. No entanto, nas suas versões propostas por Cox (1996) e Papke e Wooldridge (1996) eles assumem que variáveis que estão fortemente relacionadas com a média dos valores contínuos também estão fortemente relacionadas com a chance de ocorrência dos valores discretos. Em muitas situações, essa suposição não é razoável (Cook et al., 2008). O modelo de Papke e Wooldridge (1996) foi estendido por Ramalho e Silva (2009) para que essa suposição não seja necessária. No entanto, no modelo deles, apenas a chance de ocorrência do valor zero e a média dos valores contínuos não estão necessariamente relacionadas com as mesmas variáveis preditoras. O modelo ainda supõe que variáveis que estão fortemente relacionadas com a média dos valores contínuos também estão fortemente relacionadas com a chance de ocorrência do valor um.

Um modelo popular entre os econometristas para o estudo de variáveis respostas que estão limitadas entre dois valores é o modelo tobito duplamente limitado (Maddala, 1986, página 161).

Esse modelo assume que entre zero e um a variável resposta tem distribuição normal truncada. Devido a essa suposição, a variância da variável resposta não é uma função da média. No entanto, para variáveis limitadas entre dois pontos é mais razoável admitir que a variância decresce à medida que a média se aproxima dos pontos zero ou um (Cook et al., 2008). O modelo tobito tem ainda a mesma restrição já discutida para os modelos de quase-verossimilhança. Além disso, o modelo foi originalmente desenvolvido para dados censurados entre zero e um, o que não é o caso quando estamos estudando variáveis resposta que são uma proporção.

Uma classe de modelos de regressão para variáveis que assumem valores no intervalo $[0, 1]$ foi proposta por Lesaffre et al. (2007). No entanto, a classe de modelos proposta por eles pode ser utilizada apenas quando a variável resposta assume exatamente k valores discretos no intervalo $[0, 1]$, ou quando a variável dependente é resultado da razão entre uma variável de contagem e uma constante conhecida, que pode variar entre as unidades populacionais. Em muitos estudos, a variável resposta assume apenas valores no intervalo $[0, 1]$, mas não pode ser enquadrada em nenhuma dessas duas situações. Assim, é conveniente a obtenção de um modelo mais geral.

Uma possível abordagem é utilizar um modelo inflacionado. O termo inflacionado indica que a massa de probabilidade de alguns pontos é maior do que a suposta no modelo proposto. A maioria dos trabalhos na área de modelos inflacionados supõe que a distribuição da variável resposta é uma mistura entre uma distribuição degenerada no zero e uma outra distribuição conhecida. Aitchison (1955) foi o primeiro a sugerir esse tipo de mistura e discutiu a obtenção de estimadores eficientes quando o interesse é apenas o estudo da distribuição de uma variável desse tipo e não estudá-la em função de outras variáveis como em um modelo de regressão. O primeiro modelo de regressão inflacionado foi proposto por Feuerverger (1979) para o estudo de precipitações pluviométricas e utilizou a distribuição gama. Desde então foram desenvolvidos modelos inflacionados para diversas distribuições como, por exemplo, Poisson (Lambert, 1992; Ridout et al., 1998), binomial (Vieira et al., 2000; Hall, 2000), binomial negativa (Greene, 1994) e gaussiana inversa (Heller et al., 2006).

Ospina (2008) introduziu o modelo de regressão beta inflacionado em zero e um. Trata-se de uma extensão do modelo de regressão beta para situações em que a variável resposta varia no intervalo $[0, 1]$. O modelo assume que a distribuição da variável resposta é beta inflacionada (Ospina e Ferrari, 2010), sendo esta uma mistura de uma distribuição beta e uma distribuição de Bernoulli. Nesse modelo, tanto a média dos valores contínuos como a chance de assumir os valores zero e um são ajustadas em função de variáveis preditoras. No mesmo período, Cook et al. (2008) introduziram o modelo de regressão beta inflacionado em zero para situações em que a variável resposta assume valores no intervalo $[0, 1)$ e Hoff (2007) propôs um modelo de regressão beta inflacionado em um para problemas em que o suporte da variável dependente é o intervalo $(0, 1]$. Em ambos os modelos assume-se que a variável resposta segue uma distribuição de mistura entre uma distribuição beta e uma distribuição degenerada em um ponto: zero para o primeiro modelo e um para o último. No trabalho de Cook et al. (2008) foi ainda comparada, utilizando um conjunto de dados reais, a regressão beta inflacionada no zero, o modelo de quase-verossimilhança de Papke e Wooldridge (1996) e o modelo tobito duplamente limitado. Concluiu-se, conforme esperado diante do que já

foi discutido, que a regressão beta inflacionada produziu um ajuste bem superior aos dos demais modelos.

Alguns outros trabalhos relacionados ao modelo de regressão beta inflacionado já foram desenvolvidos. Um modelo de regressão não linear para variáveis resposta com distribuição beta inflacionada em zero ou um foi introduzido por Ospina e Ferrari (2012). Testes de erros de especificação e correções para pequenas amostras do teste da razão de verossimilhanças foram desenvolvidos por Pereira (2010).

Os modelos de regressão beta e beta inflacionado conseguem ajustar adequadamente grande parte das variáveis do tipo proporção. No entanto, algumas dessas variáveis não podem ser bem ajustadas por esses modelos. Isso ocorre, por exemplo, quando a variável resposta não pode assumir valores no intervalo $(0, c)$. Variáveis relacionadas a algum tipo de pagamento limitado entre dois valores, quando estudadas em relação ao seu valor máximo, possuem essa característica. Um exemplo dessas variáveis que tem grande interesse econômico na área bancária é o valor pago da fatura de cartão de crédito como proporção do valor total da fatura (valor relativo do pagamento). A administradora de cartão de crédito estabelece um valor de pagamento mínimo e portanto essa variável não pode assumir valores entre $(0, c)$, sendo c um valor conhecido. O valor relativo do pagamento pode assim assumir os valores discretos zero (em caso de não pagamento), um (em caso de pagamento total) e c (em caso de pagamento mínimo) ou um valor contínuo no intervalo $(c, 1)$. Outros exemplos podem ser encontrados na área de previdência social. O valor da contribuição mensal de um indivíduo está limitado entre a contribuição associada a um salário mínimo e um teto reajustado anualmente pelo Instituto Nacional de Seguridade Social (INSS). Dessa forma, o valor de contribuição do indivíduo ao INSS como proporção do teto de contribuição também pode assumir os valores discretos zero, um e c ou valores contínuos no intervalo $(c, 1)$. Os valores dos benefícios pagos pelo INSS aos idosos que trabalharam na iniciativa privada como proporção do benefício máximo também têm as mesmas características, já que o rendimento de aposentadorias é igual ou superior ao salário mínimo.

Não se tem conhecimento de nenhum trabalho em que são ajustados modelos de regressão para variáveis que assumem valores discretos em zero, um e c ou contínuos no intervalo $(c, 1)$. Estudos relacionados ao pagamento do cartão de crédito geralmente consideram a variável resposta como binária, com níveis pagou ou não o valor total da fatura (Hamilton e Khan, 2001; Abdul-Muhmin e Umar, 2007). A probabilidade do indivíduo não pagar a fatura total é assim ajustada através de um modelo de regressão logística. Outros trabalhos estimam conjuntamente dois parâmetros em função de variáveis preditoras (Kim e DeVaney, 2001; Robb e Sharpe, 2009). O primeiro parâmetro é a probabilidade do indivíduo não pagar a fatura total. O segundo é a esperança do valor financiado (valor da fatura que o indivíduo deixou de pagar) quando o cliente não paga a fatura total. As estimativas são obtidas a partir do modelo de dois estágios de Heckman (Maddala, 1983; Dow e Norton, 2003; Jonsson, 2012). No primeiro estágio é ajustada a probabilidade do cliente não pagar a fatura total através de um modelo probito. No segundo estágio estima-se a esperança do valor financiado (não pago) quando o cliente não paga a fatura total. Utiliza-se neste estágio apenas a

subamostra de indivíduos que não pagaram a fatura total. As estimativas são obtidas a partir do estimador de mínimos quadrados ordinários, tendo como variável preditora adicional uma função do preditor linear do modelo probito ajustado no primeiro estágio. O modelo de Heckman pode ser visto ainda como uma alternativa aos modelos inflacionados para variáveis resposta que são uma mistura entre uma variável degenerada em um ponto e uma outra com alguma distribuição com suporte em $\mathbb{R}^+$. Uma abordagem diferente foi utilizada por Régis e Artes (2008) que consideraram que o cartão de crédito tem 5 estados e ajustaram um modelo multi-estado de Markov (Kay, 1986), que estima a probabilidade de transição entre eles. Os 5 estados considerados foram: pagou o valor total da fatura (VRP = 1), financiou uma parte da fatura ($c \leq$ VRP < 1), não efetuou nenhum pagamento VRP = 0), *default* (fatura em atraso há pelo menos 3 meses) e cartão cancelado.

Na área de previdência há um detalhado trabalho que estima a renda de aposentados norte americanos a partir de diversas variáveis e estimativas e simulações de valores futuros para o estado civil, a idade de morte e a idade de aposentadoria (Toder et al., 2002). O trabalho foi aperfeiçoado diversas vezes ao longo dos anos, mas nenhuma versão leva em consideração o fato de que o valor de aposentadoria pública está limitado entre dois valores. Moura e Istake (2007) ajustou o valor de aposentadoria total (pública mais privada) de brasileiros e paranaenses utilizando dados da Pesquisa Nacional de Amostra de Domicílios (PNAD) do ano de 2004 do Instituto Brasileiro de Geografia e Estatística (IBGE). Foi utilizado um modelo de regressão linear múltipla com transformação logarítmica da variável resposta. Como a aposentadoria inclui o benefício privado, que não tem valores mínimos e máximos, não foi necessária a utilização de um modelo que levasse em consideração que a variável resposta era limitada.

1.1 Objetivo e organização da tese

O objetivo da tese é desenvolver aspectos de inferência e diagnóstico em modelos de regressão para variáveis que assumem valores em $[0, 1]$, têm probabilidade zero de assumir valores no intervalo $(0, c)$ e positiva em zero, um e c, com a suposição que em $(c, 1)$ sua distribuição é beta. Denominaremos a distribuição de probabilidade dessas variáveis beta inflacionada truncada. Assim, em outras palavras, o objetivo da tese é desenvolver aspectos de inferência para ajustar um modelo de regressão e realizar o seu diagnóstico para variáveis resposta que têm distribuição beta inflacionada truncada. O trabalho é uma extensão do estudo de Ospina (2008) e utiliza sua estrutura e muitos de seus resultados. O modelo a ser desenvolvido será denominado modelo de regressão beta inflacionado truncado.

A tese está organizada da seguinte forma. O Capítulo 2 introduz a distribuição beta inflacionada truncada, obtém seu vetor escore e matriz de informação de Fisher e discute estimação pontual e intervalar de seus parâmetros. As propriedades dos estimadores sugeridos também são estudadas utilizando-se dados simulados e uma aplicação a dados reais é realizada. O Capítulo 3 apresenta o modelo de regressão beta inflacionado truncado. O vetor escore e a informação de Fisher são obtidos e discute-se a estimação dos parâmetros por máxima verossimilhança. São apresentados ainda expressões para intervalos de confiança e testes de hipóteses para os parâmetros envolvidos. O modelo admite que o valor c pode variar entre as unidades populacionais. Assim, também são

apresentados três resultados para um particular modelo de regressão beta inflacionado truncado quando c é variável e também é discutida a possibilidade de c ser uma variável endógena (Stock e Watson, 2003, Seção 10.1), caso seja incluída como variável preditora do modelo. Em seguida, são feitos estudos de simulação de Monte Carlo para avaliar a performance dos estimadores apresentados para o modelo de regressão beta inflacionado truncado. Para finalizar o capítulo, discute-se ainda seleção de modelos. O Capítulo 4 discute análise de resíduos e de influência local específicas para o modelo do Capítulo 3. O Capítulo 5 traz uma aplicação do modelo desenvolvido neste trabalho para o pagamento do cartão de crédito utilizando dados reais obtidos de uma instituição financeira. O Capítulo 6 apresenta as conclusões da tese e sugestões para estudos posteriores.

Capítulo 2

Distribuição beta inflacionada truncada

Neste capítulo são apresentadas as distribuições beta (Seção 2.1), beta inflacionada (Seção 2.2) e beta inflacionada truncada (Seção 2.3), com ênfase nesta última, que é introduzida neste trabalho e para a qual é sugerida uma extensão na Seção 2.4. Na Seção 2.5 são feitos estudos de simulação e na Seção 2.6 é discutida uma aplicação a dados reais para a distribuição beta inflacionada truncada.

2.1 Distribuição beta

A distribuição beta tem suporte no intervalo aberto (a, b), a e b números reais conhecidos e é geralmente parametrizada utilizando-se dois parâmetros de forma. Nesta parametrização, sua função densidade de probabilidade é dada por

$$f(y; p, q, a, b) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \frac{(y-a)^{p-1}(b-y)^{q-1}}{(b-a)^{p+q-1}}, \quad y \in (a, b), \tag{2.1}$$

em que $\Gamma(.)$ é a função gama definida como $\Gamma(k) = \int_0^\infty t^{k-1}e^{-t}\,dt$ e $p > 0$ e $q > 0$ são os parâmetros de forma. Se $a = 0$ e $b = 1$, a distribuição assume sua forma padrão, que é a mais utilizada em aplicações. Definindo-se $\mu = a + p(b-a)/(p+q)$ e $\phi = p + q$, ou seja, $p = [(\mu - a)/(b-a)]\phi$ e $q = [(b-\mu)/(b-a)]\phi$, obtém-se uma parametrização alternativa da distribuição beta (Ferrari e Cribari-Neto, 2004) com função densidade de probabilidade dada por

$$f(y; \mu, \phi, a, b) = \frac{\Gamma(\phi)(y-a)^{\left(\frac{\mu-a}{b-a}\right)\phi-1}(b-y)^{\left(\frac{b-\mu}{b-a}\right)\phi-1}}{\Gamma\left[\left(\frac{\mu-a}{b-a}\right)\phi\right]\Gamma\left[\left(\frac{b-\mu}{b-a}\right)\phi\right](b-a)^{\phi-1}}, \quad y \in (a, b). \tag{2.2}$$

A parametrização (2.2) é mais conveniente no estudo de modelos de regressão pois, neste caso, o parâmetro $\mu \in (a, b)$ é a média de uma variável com distribuição beta e $\phi > 0$ é um parâmetro de precisão.

A distribuição beta é bastante conveniente no estudo de variáveis limitadas, pois sua função densidade de probabilidade é bastante flexível. Dependendo dos valores de μ e ϕ, (2.2) pode assumir diversas formas simétricas e assimétricas, como, por exemplo, 'U', 'J' e 'J' invertido. Johnson et al. (1995) e Gupta e Nadarajah (2004) apresentam diversas características, propriedades e aplicações da distribuição beta padrão com parametrização (2.1) e Ospina (2008) apresenta seus momentos na parametrização (2.2). Para quaisquer valores a e b em (2.2), se Y é uma variável aleatória com

distribuição beta, então seu momento (em torno do zero) de ordem r ($r = 1, 2, \ldots$) pode ser obtido recursivamente através de

$$\mathrm{E}(Y^r) = \mu_r = \mathrm{E}[(Y-a)^r] + \sum_{i=1}^{r} (-1)^{i+1} \binom{r}{i} a^i \mu_{r-i}, \tag{2.3}$$

em que $\mathrm{E}[(Y-a)^r] = (b-a)^r \prod_{i=1}^{r} \left(\frac{p+r-i}{\phi+r-i} \right)$ e $p = \frac{(\mu-a)\phi}{(b-a)}$, com $\mu_0 = 1$ e $\mu_1 = E(Y) = \mu$.

Utilizando-se (2.3), obtém-se que

$$\mathrm{Var}(Y) = \frac{(\mu-a)(b-\mu)}{1+\phi} = \frac{V(\mu)}{1+\phi}, \tag{2.4}$$

em que $V(\mu) = (\mu-a)(b-\mu)$ é a função de variância da distribuição beta. As provas de (2.3) e (2.4) estão no Anexo A.1.

2.2 Distribuição beta inflacionada

Em algumas áreas, há variáveis de interesse que são proporções que podem assumir os valores zero e um com probabilidade positiva. A proporção de óbitos por causas mal definidas nos municípios brasileiros em determinado ano e a proporção de grãos de milho de uma espiga que estão contaminados com um determinado gênero de fungo são exemplos de variáveis com essas características. Essas variáveis não são absolutamente contínuas e portanto não podem ser ajustadas pela distribuição beta. Para estudar variáveis que assumem valores no intervalo $[0, 1]$ e têm probabilidade positiva de assumir os valores zero e um, Ospina e Ferrari (2010) propuseram a distribuição beta inflacionada em zero e um (BIZU). Trata-se de uma mistura entre a distribuição beta e a distribuição de Bernoulli. Os autores obtiveram os momentos da distribuição BIZU, sua função escore, sua matriz de informação de Fisher e provaram que ela pertence à família exponencial de posto completo. Discutiram ainda estimação por momentos condicionais e máxima verossimilhança, desenvolveram estudos de simulação para estudar a performance dos estimadores obtidos e apresentaram aplicações. A distribuição de probabilidade que será introduzida neste trabalho na Seção 2.3 é uma extensão da distribuição BIZU. Os resultados obtidos por Ospina e Ferrari (2010) podem ser obtidos a partir dos resultados que serão apresentados na próxima seção. Por esse motivo, não serão apresentados nesta seção os resultados da distribuição beta inflacionada obtidos pelos autores do artigo.

2.3 Distribuição beta inflacionada truncada

Nas situações em que a variável de interesse é uma proporção que não pode assumir valores em um determinado intervalo, as distribuições beta, BIZU, BIZ (beta inflacionada no ponto zero) e BIU (beta inflacionada no ponto um) não são adequadas. Conforme discutido no Capítulo 1, esse é o caso, por exemplo, de variáveis relacionadas a um pagamento limitado entre dois valores, quando estudadas em função de seu valor máximo. Além dos exemplos já citados no capítulo anterior, o

valor que um desempregado brasileiro recebe de seguro desemprego como proporção do valor máximo e o valor mensal que um bancário paga ao sindicato como proporção do valor máximo também não podem assumir valores em um intervalo $(0, c)$, sendo $c \in (0, 1)$ um valor conhecido. Todas essas variáveis podem assumir os valores zero, um e c com probabilidade positiva e qualquer número real no intervalo $(c, 1)$. Assim, para ajustar essas variáveis é razoável introduzir uma distribuição que é uma mistura de uma distribuição beta e uma distribuição discreta. A distribuição proposta é uma mistura de uma distribuição beta com suporte no intervalo $(c, 1)$ e uma distribuição trinomial que assume os valores zero, um e c. Essa distribuição será denominada distribuição beta inflacionada truncada (BIZUT) e sua função densidade de probabilidade é dada por

$$f(y; \alpha, \gamma_0, \gamma_1, \mu, \phi, c) = \begin{cases} \alpha\gamma_0 & \text{se } y = 0 \\ \alpha\gamma_1 & \text{se } y = 1 \\ \alpha(1 - \gamma_0 - \gamma_1) & \text{se } y = c \\ (1 - \alpha) f(y; \mu, \phi, c, 1) & \text{se y} \in (c, 1), \end{cases} \tag{2.5}$$

em que α é o parâmetro de mistura $(P(Y \in \{0, 1, c\}))$; γ_0 e γ_1 são os parâmetros da trinomial $(P(Y = 0|Y \in \{0, 1, c\})$ e $P(Y = 1|Y \in \{0, 1, c\})$, respectivamente), Y é uma variável aleatória com distribuição BIZUT e $f(y; \mu, \phi, c, 1)$ é a função densidade de probabilidade da distribuição beta com vetor de parâmetros $(\mu, \phi, c, 1)$. Neste caso, o espaço paramétrico de α e γ_0 são os intervalos abertos $(0, 1)$, $c < \mu < 1$, $\phi > 0$ e $0 < \gamma_1 < 1 - \gamma_0$.

Na parametrização (2.5) da distribuição BIZUT é fácil perceber que a mesma é uma mistura entre uma distribuição beta e uma distribuição trinomial. No entanto, nessa parametrização, a interpretação dos parâmetros γ_0 e γ_1 não é tão simples. Assim, neste trabalho será utilizada uma parametrização mais conveniente que é obtida definindo-se $\delta_0 = \alpha\gamma_0$, $\delta_1 = \alpha\gamma_1$ e $\delta_c = \alpha(1 - \gamma_0 - \gamma_1)$ e, portanto, $\alpha = \delta_0 + \delta_1 + \delta_c$, $\gamma_0 = \delta_0/(\delta_0 + \delta_1 + \delta_c)$ e $\gamma_1 = \delta_1/(\delta_0 + \delta_1 + \delta_c)$. Dessa forma, propomos a definição a seguir.

Definição 2.3.1. *Seja Y uma variável aleatória com suporte em $\{0\} \cup [c, 1]$. Dizemos que Y tem distribuição beta inflacionada truncada com vetor de parâmetros $(\delta_0, \delta_1, \delta_c, \mu, \phi, c)$, se sua função densidade de probabilidade é dada por*

$$f(y; \delta_0, \delta_1, \delta_c, \mu, \phi, c) = \begin{cases} \delta_0 & \textit{se } y = 0 \\ \delta_1 & \textit{se } y = 1 \\ \delta_c & \textit{se } y = c \\ (1 - \delta_0 - \delta_1 - \delta_c) f(y; \mu, \phi, c, 1) & \textit{se } y \in (c, 1), \end{cases} \tag{2.6}$$

em que δ_0, δ_1 e δ_c são, respectivamente, $P(Y = 0)$, $P(Y = 1)$ e $P(Y = c)$, $\mu = E(Y|Y \in (c, 1))$, ϕ é um parâmetro de precisão, c é um parâmetro conhecido com espaço paramétrico no intervalo $(0, 1)$ e $f(y; \mu, \phi, c, 1)$ é definida como em (2.2).

Nesta parametrização, $\delta_0 \in (0,1)$, $\mu \in (c,1)$, $\phi > 0$, $0 < \delta_1 < 1-\delta_0$, $0 < \delta_c < 1-\delta_0-\delta_1$ e a notação utilizada para a distribuição será BIZUT$(\delta_0, \delta_1, \delta_c, \mu, \phi, c)$. Pode-se notar que tanto na parametrização (2.5) como na (2.6) há restrições no espaço paramétrico relacionadas aos parâmetros associados aos pontos em que a variável é discreta. Em (2.6), por exemplo, $0 < \delta_0+\delta_1+\delta_c < 1$. No entanto, conforme será visto posteriormente, isso não acarreta uma maior dificuldade na estimação dos parâmetros porque esses parâmetros têm estimadores de máxima verossimilhança de forma fechada que sempre satisfazem as restrições. Quando $c \to 0$ e $\delta_c \to 0$, a distribuição BIZUT tende à distribuição BIZU introduzida por Ospina e Ferrari (2010). A partir de (2.6) pode-se perceber que a função distribuição BIZUT é dada por

$$F(y;\delta_0,\delta_1,\delta_c,\mu,\phi,c) = \delta_0 I_{[0,\infty[}(y) + \delta_c I_{[c,\infty[}(y) + \delta_1 I_{[1,\infty[}(y) + (1-\alpha)F(y;\mu,\phi,c,1), \tag{2.7}$$

em que $\alpha = \delta_0+\delta_1+\delta_c$ e $F(y;\mu,\phi,c,1)$ é a função distribuição de uma beta com vetor de parâmetros $(\mu,\phi,c,1)$.

Os momentos da distribuição BIZUT podem ser obtidos a partir da variável indicadora $I_{\{0,1,c\}}(Y)$. Temos que $Y|I_{\{0,1,c\}}(Y) = 1$ tem distribuição trinomial e $Y|I_{\{0,1,c\}}(Y) = 0$ tem distribuição beta. Portanto,

$$\mathrm{E}(Y^r|I_{\{0,1,c\}}(Y)) = \begin{cases} (c^r\delta_c + \delta_1)/\alpha & \text{com probabilidade } \alpha \\ \mu_r & \text{com probabilidade } 1-\alpha \end{cases} \tag{2.8}$$

e

$$\mathrm{Var}(Y|I_{\{0,1,c\}}(Y)) = \begin{cases} V_1 & \text{com probabilidade } \alpha \\ V_2 & \text{com probabilidade } 1-\alpha, \end{cases} \tag{2.9}$$

nas quais μ_r é o momento de ordem r (em torno do 0) da distribuição beta com vetor de parâmetros $(\mu,\phi,c,1)$ apresentado em (2.3), $V_1 = \mathrm{Var}(Y|I_{\{0,c,1\}}(Y)=1) = (c^2\delta_c(\alpha-\delta_c)+\delta_1(\alpha-\delta_1)-2c\delta_c\delta_1)/\alpha^2$ e $V_2 = \mathrm{Var}(Y|I_{\{0,c,1\}}(Y)=0) = [(\mu-c)(1-\mu)]/(\phi+1)$. Utilizando-se (2.8) e (2.9), após alguns cálculos algébricos (Apêndice A.2), obtém-se

$$\mathrm{E}(Y^r) = c^r\delta_c + \delta_1 + (1-\alpha)\mu_r, \quad r = 1,2,\ldots, \tag{2.10}$$

$$\mathrm{Var}(Y) = \alpha V_1 + (1-\alpha)V_2 + (1-\alpha)\left[\frac{c\delta_c+\delta_1}{\sqrt{\alpha}} - \sqrt{\alpha}\mu\right]^2. \tag{2.11}$$

Pode-se observar a partir de (2.10) que $\mathrm{E}(Y) = c\delta_c + \delta_1 + (1-\alpha)\mu$.

A função densidade de probabilidade (2.6) pode ser reescrita como

$$f(y;\eta) = \exp[\eta^\top T(y) - A(\eta)]h(y), \tag{2.12}$$

em que $\eta = (\eta_1,\eta_2,\eta_3,\eta_4,\eta_5)^\top$, $T(y) = (t_1(y),t_2(y),t_3(y),t_4(y),t_5(y))^\top$, $\theta = (\delta_0,\delta_1,\delta_c,\mu,\phi)$, $\eta_1 = \log(\delta_0/C(\theta))$, $\eta_2 = \log(\delta_1/C(\theta))$, $\eta_3 = \log(\delta_c/C(\theta))$, $\eta_4 = [(\mu-c)/(1-c)]\phi$, $\eta_5 = [(1-\mu)/(1-c)]\phi$, $C(\theta) = [(1-\delta_0-\delta_1-\delta_c)\Gamma(\phi)]/\{\Gamma\{[(\mu-c)/(1-c)]\phi\}\Gamma\{[(1-\mu)/(1-c)]\phi\}(1-c)^{\phi-1}\}$, $t_1(y) = I_{\{0\}}(y)$, $t_2(y) = I_{\{1\}}(y)$, $t_3(y) = I_{\{c\}}(y)$, $t_4(y) = \log(y-c)$ se $y \in (c,1)$ e 0 caso contrário, $t_5(y) = \log(1-y)$

se $y \in (c,1)$ e 0 caso contrário, $A(\eta) = -\log\{\Gamma(\eta_4+\eta_5)/[\Gamma(\eta_4)\Gamma(\eta_5)(1-c)^{\eta_4+\eta_5-1} + (e^{\eta_1}+e^{\eta_2}+e^{\eta_3})\Gamma(\eta_4+\eta_5)]\}$, $h(y) = 1/[(y-c)(1-y)]$ se $y \in (c,1)$ e 1 caso contrário. Além disso, tanto as estatísticas t quanto os parâmetros η são linearmente independentes e o espaço paramétrico de η contém um retângulo aberto 5-dimensional. Assim, está provada (Lehmann e Casella, 1998, Seção 1.5) a seguinte proposição:

Proposição 2.3.1. *A distribuição BIZUT pertence à família exponencial de posto completo.*

Sejam $y_1, y_2, \ldots, y_n$ uma amostra aleatória de uma distribuição BIZUT$(\delta_0, \delta_1, \delta_c, \mu, \phi, c)$. Então, como consequência da proposição (2.3.1), $(T_1, T_2, T_3, T_4, T_5) = \sum_{i=1}^n (t_1(y_i), t_2(y_i), t_3(y_i), t_4(y_i), t_5(y_i))$ é uma estatística suficiente e completa, em que

$$\begin{aligned}
T_1 &= \sum_{i=1}^n I_{\{0\}}(y_i), \\
T_2 &= \sum_{i=1}^n I_{\{1\}}(y_i), \\
T_3 &= \sum_{i=1}^n I_{\{c\}}(y_i), \\
T_4 &= \sum_{i:y_i\in(c,1)} \log(y_i - c), \\
T_5 &= \sum_{i:y_i\in(c,1)} \log(1 - y_i).
\end{aligned} \tag{2.13}$$

Utilizando-se (2.6), obtém-se que o logaritmo da função de verossimilhança de Y associado a essa amostra é dado por

$$l(\theta) = l_1(\theta) + l_2(\theta), \tag{2.14}$$

em que

$$\begin{aligned}
l_1(\theta) &= T_1\log(\delta_0) + T_2\log(\delta_1) + T_3\log(\delta_c) + (n - T_1 - T_2 - T_3)\log(1-\delta_0-\delta_1-\delta_c), \\
l_2(\theta) &= (n - T_1 - T_2 - T_3)\log\left\{\frac{\Gamma(\phi)}{\Gamma\left[\left(\frac{\mu-c}{1-c}\right)\phi\right]\Gamma\left[\left(\frac{1-\mu}{1-c}\right)\phi\right](1-c)^{\phi-1}}\right\} + \left[\left(\frac{\mu-c}{1-c}\right)\phi - 1\right]T_4 + \\
&\quad \left[\left(\frac{1-\mu}{1-c}\right)\phi - 1\right]T_5.
\end{aligned}$$

Pode-se notar que $l_1(\theta)$ é função somente dos parâmetros δ_0, δ_1 e δ_c e $l_2(\theta)$ é função somente de μ e ϕ. Portanto, $(\delta_0, \delta_1, \delta_c)$ e (μ, ϕ) são parâmetros separáveis e a estimação de $(\delta_0, \delta_1, \delta_c)$ pode ser feita separadamente da estimação de (μ, ϕ) e vice-versa (Pace e Salvan, 1997, página 128).

A função escore associada a uma amostra de Y é obtida derivando-se (2.14) em relação a cada um dos componentes de θ. Lembrando que a função digama é definida como $\psi(j) = \partial \log\Gamma(j)/\partial j =$

$\Gamma'(j)/\Gamma(j)$ obtém-se a função escore $(U_{\delta_0}, U_{\delta_1}, U_{\delta_c}, U_\mu, U_\phi)$ em que

$$\begin{aligned}
U_{\delta_0} &= T_1/\delta_0 - (n - T_1 - T_2 - T_3)/(1-\alpha), \\
U_{\delta_1} &= T_2/\delta_1 - (n - T_1 - T_2 - T_3)/(1-\alpha), \\
U_{\delta_c} &= T_3/\delta_c - (n - T_1 - T_2 - T_3)/(1-\alpha), \\
U_\mu &= [\phi/(1-c)]\{(n - T_1 - T_2 - T_3)\{\psi[\phi(1-\mu)/(1-c)] - \psi[\phi(\mu-c)/(1-c)]\} + T_4 - T_5\}, \\
U_\phi &= (n - T_1 - T_2 - T_3)\{\psi(\phi) - [(\mu-c)/(1-c)]\psi[\phi(\mu-c)/(1-c)] - \log(1-c) - \\
&\quad [(1-\mu)/(1-c)]\psi[\phi(1-\mu)/(1-c)]\} + [(\mu-c)/(1-c)]T_4 + [(1-\mu)/(1-c)]T_5.
\end{aligned} \tag{2.15}$$

Resolvendo-se o sistema de equações $(U_{\delta_0}, U_{\delta_1}, U_{\delta c}) = (0,0,0)$, obtém-se que os estimadores de máxima verossimilhança de δ_0, δ_1 e δ_c são respectivamente dados por

$$\hat{\delta}_0 = \frac{T_1}{n}, \quad \hat{\delta}_1 = \frac{T_2}{n}, \hat{\delta}_c = \frac{T_3}{n}. \tag{2.16}$$

Eles são não viciados e funções da estatística suficiente e completa e, portanto, são estimadores não viciados de variância uniformemente mínima (ENVVUM) de δ_0, δ_1 e δ_c, respectivamente. A partir de (2.13), pode-se notar que T_1, T_2 e T_3 têm distribuição binomial com parâmetros (n, δ_0), (n, δ_1) e (n, δ_c), respectivamente. Logo, a variância de $\hat{\delta}_k, k = 0, 1$ ou c é $\delta_k(1-\delta_k)/n$. O sistema de equações $(U_\mu, U_\phi) = (0,0)$ não tem solução algébrica. Assim, os estimadores de máxima verossimilhança de μ $(\hat{\mu})$ e de ϕ $(\hat{\phi})$ devem ser obtidos utilizando-se métodos numéricos. Nas Seções 2.5 e 2.6, será utilizado o método quase-Newton BFGS com derivadas analíticas (Nocedal e Wright, 2006, Seção 6.1).

Estimadores de forma fechada de μ e ϕ podem ser obtidos, de forma semelhante aos obtidos por Ospina e Ferrari (2010) para a distribuição beta inflacionada, a partir dos momentos condicionais de Y dado que $Y \in (c, 1)$. A distribuição de $Y|Y \in (c,1)$ é beta com vetor de parâmetros $(\mu, \phi, c, 1)$, c conhecido e, portanto, $\mathrm{E}[Y|Y \in (c,1)] = \mu$ e $\mathrm{Var}[Y|Y \in (c,1)] = (\mu - c)(1-\mu)/(1+\phi)$. Resolvendo-se o sistema de equações

$$\begin{cases} \mu & = \bar{y}^* \\ \frac{(\mu-c)(1-\mu)}{(1+\phi)} & = s^{2*}, \end{cases}$$

em que $\bar{y}^* = \sum_{i:y_i \in (c,1)}[y_i/(n - T_1 - T_2 - T_3)]$ e $s^{2*} = \sum_{i:y_i \in (c,1)}[(y_i - \bar{y}^*)^2/(n - T_1 - T_2 - T_3)]$, obtêm-se os estimadores

$$\tilde{\mu} = \bar{y}^* \quad e \quad \tilde{\phi} = \frac{(\bar{y}^* - c)(1 - \bar{y}^*)}{s^{2*}} - 1. \tag{2.17}$$

Estimadores de forma fechada de $\mathrm{E}(Y)$ e de Var(Y) podem ser obtidos substituindo-se μ e ϕ por $\tilde{\mu}$ e $\tilde{\phi}$ em (2.10) e (2.11).

A matriz de Informação de Fisher, $K(\theta)$, pode ser obtida utilizando-se que, na família exponencial de posto completo, $\mathrm{E}[t_k(y_i)] = \partial A(\eta)/\partial \eta_k$ (Lehmann e Casella, 1998, página 27). Após alguns

cálculos algébricos (Apêndice A.3), obtém-se que

$$K(\theta) = \begin{pmatrix} k_{\delta_0\delta_0} & k_{\delta_0\delta_1} & k_{\delta_0\delta_c} & 0 & 0 \\ k_{\delta_0\delta_1} & k_{\delta_1\delta_1} & k_{\delta_1\delta_c} & 0 & 0 \\ k_{\delta_0\delta_c} & k_{\delta_1\delta_c} & k_{\delta_c\delta_c} & 0 & 0 \\ 0 & 0 & 0 & k_{\mu\mu} & k_{\mu\phi} \\ 0 & 0 & 0 & k_{\mu\phi} & k_{\phi\phi} \end{pmatrix}, \tag{2.18}$$

em que

$$\begin{aligned}
k_{\delta_0\delta_0} &= (1-\delta_1-\delta_c)/[\delta_0(1-\alpha)],\ k_{\delta_1\delta_1} = (1-\delta_0-\delta_c)/[\delta_1(1-\alpha)], \\
k_{\delta_c\delta_c} &= (1-\delta_0-\delta_1)/[\delta_c(1-\alpha)],\ k_{\delta_0\delta_1} = k_{\delta_0\delta_c} = k_{\delta_1\delta_c} = (1-\alpha)^{-1}, \\
k_{\mu\mu} &= [\phi/(1-c)]^2(1-\alpha)\{\psi'[\phi(\mu-c)/(1-c)] + \psi'[\phi(1-\mu)/(1-c)]\}, \\
k_{\phi\phi} &= (1-\alpha)\{[(\mu-c)/(1-c)]^2\psi'[\phi(\mu-c)/(1-c)] + [(1-\mu)/(1-c)]^2\psi'[\phi(1-\mu)/(1-c)] - \\
&\quad \psi'(\phi)\}, \\
k_{\mu\phi} &= [\phi(1-\alpha)/(1-c)^2]\{(\mu-c)\psi'[\phi(\mu-c)/(1-c)] - (1-\mu)\psi'[\phi(1-\mu)/(1-c)]\}
\end{aligned}$$

e $\psi'(.)$ é a função trigama definida como $\psi'(j) = \partial^2 \log\Gamma(j)/\partial j^2 = \partial\psi(j)/\partial j$. Pode-se notar em (2.18) que $(\delta_0, \delta_1, \delta_c)$ e (μ, ϕ) são parâmetros ortogonais e, portanto, os dois vetores de estimadores de máxima verossimilhança destes parâmetros são assintoticamente independentes (Cox e Reid, 1987).

Pelas propriedades assintóticas do estimador de máxima verossimilhança, $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N_5(0, K(\theta)^{-1})$ (condições de regularidade estão satisfeitas porque a distribuição BIZUT pertence à família exponencial de posto completo). Dessa forma, para amostras grandes, intervalos de confiança aproximados para cada um dos parâmetros com coeficiente de confiança de $(100 \times \gamma)\%$ são dados por

$$\mathrm{IC}(\theta_k; \gamma) = \hat{\theta}_k \pm z_{(1+\gamma)/2}(\hat{k}^{\theta_k\theta_k}/n)^{1/2}, \tag{2.19}$$

em que $z_{(1+\gamma)/2}$ é o quantil $(1+\gamma)/2$ da distribuição $N(0,1)$ e $\hat{k}^{\theta_k\theta_k}$ é o termo de posição k da diagonal de $K(\theta)^{-1}$ avaliada em $\hat{\theta}$. Além disso, uma região de confiança assintótica conjunta para os cinco parâmetros da distribuição BIZUT com coeficiente de confiança de $(100 \times \gamma)\%$ pode ser escrita (Johnson e Wichern, 2007, página 221) como

$$n(\hat{\theta} - \theta)^\top \widehat{K(\theta)}(\hat{\theta} - \theta) \leq [5(n-1)/(n-5)]F_{5,n-5;\gamma}, \tag{2.20}$$

em que $F_{5,n-5;\gamma}$ é o quantil γ da distribuição F de Fisher-Snedecor com 5 e $n-5$ graus de liberdade e $\widehat{K(\theta)}$ é $K(\theta)$ avaliada em $\hat{\theta}$. O intervalo de confiança e a região de confiança podem ser utilizados para testar hipóteses relacionadas com os parâmetros da distribuição BIZUT. A distribuição assintótica do estimador de máxima verossimilhança de qualquer função diferenciável de θ pode ser obtida utilizando-se o método delta. No Anexo A.4, como exemplo, é obtida a distribuição assintótica de $\widehat{E(Y)}$.

2.4 Variáveis inflacionadas em mais pontos

Os resultados apresentados na Seção 2.3 podem ser estendidos facilmente para as situações em que há um maior número de pontos em que a variável é inflacionada. Suponha que uma variável X assuma os valores zero, um, c e outros v (inteiro maior que zero) valores adicionais conhecidos no intervalo $(0, c)$ com probabilidade positiva e qualquer outro número real no intervalo $(c, 1)$. Se admitirmos que a distribuição de X é uma mistura de uma distribuição beta com suporte no intervalo aberto $(c, 1)$ e uma distribuição multinomial, então sua função densidade de probabilidade em uma parametrização similar a (2.6) pode ser escrita como

$$f(x; \delta_0, \delta_1, \delta_c, \lambda_1, \ldots, \lambda_v, \mu, \phi, c) = \begin{cases} \delta_0 & \text{se } x = 0 \\ \delta_1 & \text{se } x = 1 \\ \delta_c & \text{se } x = c \\ \lambda_i & \text{se } x = s_i, i = 1, 2, \ldots, v \\ (1-\tau) f(x; \mu, \phi, c, 1) & \text{se } x \in (c, 1), \end{cases} \tag{2.21}$$

em que s_i são valores conhecidos no intervalo $(0, c)$ para os quais $P(X = s_i) > 0$, $\lambda_i = P(X = s_i)$ e $\tau = \delta_0 + \delta_1 + \delta_c + \sum_{i=1}^{v} \lambda_i$.

Os momentos de X, a estimação de seus parâmetros e a informação de Fisher podem ser obtidos de forma semelhante à apresentada na Seção 2.3. Os resultados obtidos, em geral, têm mais termos mas são bastante similares aos apresentados na seção anterior. O momento de ordem r (em torno do 0) de X, por exemplo, é dado por $\mathrm{E}(X^r) = c^r \delta_c + \delta_1 + \sum_{i=1}^{v} {s_i}^r \lambda_i + (1-\tau)\mu_r$. Os estimadores de máxima verossimilhança de λ_i são $\hat{\lambda}_i = \sum_{j=1}^{n} I_{\{s_i\}}(x_j)/n$ e os estimadores de δ_0, δ_1, δ_c, μ e ϕ não sofrem alteração. Os elementos de matriz de informação de Fisher associados à λ_i podem ser escritos como $k_{\lambda_i\lambda_i} = (1 - \tau + \lambda_i)/[\lambda_i(1-\tau)]$, $k_{\delta_0\lambda_i} = k_{\delta_1\lambda_i} = k_{\delta_c\lambda_i} = k_{\lambda_i\lambda_j, j\neq i} = (1-\tau)^{-1}$ e $k_{\lambda_i\mu} = k_{\lambda_i\phi} = 0$. Os resultados também são semelhantes aos obtidos anteriormente mesmo que os valores s_i não estejam restritos ao intervalo $(0, c)$.

2.5 Estudos de simulação e comparação das distribuições BIZU e BIZUT

Os estimadores de momentos condicionais (MC) e os estimadores de máxima verossimilhança (MV) de μ, ϕ, $\mathrm{E}(Y)$ e $\mathrm{Var}(Y)$ na distribuição BIZUT foram comparados através de dois estudos de simulação de Monte Carlo. Os estimadores de máxima verossimilhança de δ_0, δ_1 e δ_c não foram estudados porque eles são ENVVUM. Para cada caso, foram realizadas 10000 replicações e obtidas estimativas do viés e da raiz do erro quadrático médio (REQM) dos estimadores. Os dois estudos de simulação discutidos abaixo e a aplicação da Seção 2.6 foram desenvolvidos utilizando-se a linguagem de programação Ox (Doornik, 2007). Uma boa introdução ao uso do Ox em Estatística pode ser vista em Cribari-Neto e Zarkos (2003).

O primeiro estudo de simulação foi feito para a distribuição BIZUT com parâmetros $\delta_0 = 0{,}3$, $\delta_1 = 0{,}1$, $\delta_c = 0{,}2$, $\mu = 0{,}4$, $\phi = 2$, $c = 0{,}2$ e alguns valores de n. A Tabela 2.1 apresenta os resultados. Os estimadores $\hat{\mu}$, $\widehat{\mathrm{E}(Y)}$, $\widehat{\mathrm{Var}(Y)}$ e $\widetilde{\mathrm{Var}(Y)}$ são ligeiramente viciados em pequenas

amostras, enquanto os estimadores $\tilde{\mu}$ e $\widetilde{\mathrm{E}(Y)}$ parecem ser não viciados. No entanto, analisando-se a REQM, a performance dos estimadores MV parece ser ligeiramente melhor do que a dos estimadores MC. Em geral, mesmo para amostras pequenas, os estimadores MC e MV de μ, $\mathrm{E}(Y)$ e $\mathrm{Var}(Y)$ têm uma boa performance. Por outro lado, os estimadores $\hat{\phi}$ e $\tilde{\phi}$ superestimam consideravelmente ϕ para amostras pequenas e $\hat{\phi}$ apresenta performance superior a $\tilde{\phi}$. Conforme esperado, o viés e a REQM dos estimadores de todos os parâmetros estudados diminui à medida que se aumenta o tamanho da amostra. É interessante notar que, para o vetor de parâmetros estudado e amostras pequenas, o viés de $\hat{\phi}$ é consideravelmente grande, mas $\widehat{\mathrm{Var}(Y)}$, que é função de $\hat{\phi}$, é apenas ligeiramente não viciado.

Tabela 2.1: Resultados da simulação da distribuição BIZUT para alguns valores de n: $\mu = 0{,}4$, $\phi = 2$, $\delta_0 = 0{,}3$, $\delta_1 = 0{,}1$, $\delta_c = 0{,}2$, $c = 0{,}2$, $\mathrm{E}(Y) = 0{,}3$, $\mathrm{Var}(Y) = 0{,}098$.

n	μ				ϕ			
	Viés		REQM		Viés		REQM	
	$\tilde{\mu}$	$\hat{\mu}$	$\tilde{\mu}$	$\hat{\mu}$	$\tilde{\phi}$	$\hat{\phi}$	$\tilde{\phi}$	$\hat{\phi}$
20	-0,0003	-0,0035	0,0742	0,0727	2,5870	2,2729	45,8107	45,7167
50	0,0000	-0,0015	0,0457	0,0441	0,4090	0,3840	1,1602	0,9719
100	0,0002	-0,0007	0,0318	0,0307	0,1735	0,1665	0,6261	0,5299
500	0,0000	-0,0002	0,0143	0,0137	0,0350	0,0342	0,2327	0,1964
1000	0,0000	-0,0001	0,0100	0,0096	0,0146	0,0141	0,1612	0,1361
10000	0,0000	0,0000	0,0032	0,0030	0,0022	0,0022	0,0501	0,0423
n	$\mathrm{E}(Y)$				$\mathrm{Var}(Y)$			
	Viés		REQM		Viés		REQM	
	$\mathrm{E}(\tilde{Y})$	$\mathrm{E}(\hat{Y})$	$\mathrm{E}(\tilde{Y})$	$\mathrm{E}(\hat{Y})$	$\widetilde{\mathrm{Var}}(Y)$	$\widehat{\mathrm{Var}}(Y)$	$\widetilde{\mathrm{Var}}(Y)$	$\widehat{\mathrm{Var}}(Y)$
20	0,0000	-0,0012	0,0710	0,0707	-0,0050	-0,0058	0,0320	0,0319
50	-0,0001	-0,0007	0,0443	0,0440	-0,0018	-0,0022	0,0204	0,0203
100	0,0000	-0,0003	0,0313	0,0311	-0,0008	-0,0010	0,0144	0,0143
500	0,0000	0,0000	0,0141	0,0139	-0,0002	-0,0002	0,0066	0,0065
1000	-0,0001	-0,0001	0,0099	0,0098	-0,0002	-0,0002	0,0046	0,0046
10000	0,0000	0,0000	0,0032	0,0031	0,0000	0,0000	0,0014	0,0014

No segundo estudo de simulação, fixou-se $c = 0{,}2$ e $n = 100$ e executou-se a simulação para alguns diferentes valores dos parâmetros da distribuição BIZUT (Tabela 2.2). O espaço paramétrico de ϕ não tem um limite superior e, na tabela, é comparada a performance dos estimadores para valores muito diferentes de ϕ. Por essa razão, para uma melhor comparação dos estimadores de ϕ, para esse parâmetro, a tabela apresenta o viés relativo (viés de ϕ/ϕ) e a REQM relativa (REQM de ϕ/ϕ). Para todos os valores dos parâmetros estudados, os estimadores de μ, $\mathrm{E}(Y)$ e $\mathrm{Var}(Y)$ parecem ser não viciados ou apenas ligeiramente viciados e $\hat{\phi}$ e $\tilde{\phi}$ superestimam ϕ. Além disso, a performance de todos os estimadores é melhor quando os valores dos parâmetros δ são pequenos do que quando eles são grandes. Esse resultado era esperado, já que, na prática (veja Seção 2.3), os estimadores de μ e ϕ utilizam apenas as observações para as quais $y_i \in (c, 1)$. À medida que os parâmetros δ crescem, o

número de observações utilizadas para estimar μ e ϕ decresce e, em consequência, a performance dos estimadores se reduz. As REQMs de $\hat{\mu}$ e $\tilde{\mu}$ são consideravelmente menores quando $\phi = 100$ do que quando $\phi = 2$. Isso é razoável, já que à medida que ϕ cresce, a variância de Y decresce. Além disso, as performances de $\hat{\phi}$ e $\tilde{\phi}$ são bem semelhantes quando $\phi = 100$, mas o primeiro é superior quando $\phi = 2$. Por outro lado, em relação à performance dos estimadores, o valor de μ parece ser menos importante que os valores de ϕ e dos parâmetros δ. Em resumo, os estimadores estudados de μ, $\mathrm{E}(Y)$ e $\mathrm{Var}(Y)$ apresentam boa performance, $\hat{\phi}$ e $\tilde{\phi}$ são bons estimadores de ϕ apenas para amostras grandes e, em geral, conforme esperado, os estimadores MV parecem ser ligeiramente superiores aos estimadores MC.

Tabela 2.2: Resultados da simulação da distribuição BIZUT para alguns valores de μ, ϕ, δ_0, δ_1, δ_c: $n = 100$, $c = 0{,}2$ (δ_s: $\delta_0 = 0{,}075$, $\delta_1 = 0{,}025$, $\delta_c = 0{,}050$, δ_l: $\delta_0 = 0{,}3$, $\delta_1 = 0{,}1$, $\delta_c = 0{,}2$).

Valores dos parâmetros	μ				ϕ			
	Viés		REQM		Viés relat. (%)		REQM relat. (%)	
	$\tilde{\mu}$	$\hat{\mu}$	$\tilde{\mu}$	$\hat{\mu}$	$\tilde{\phi}$	$\hat{\phi}$	$\tilde{\phi}$	$\hat{\phi}$
δ_s, $\mu = 0{,}4$, $\phi = 2$	0,0000	-0,0003	0,0219	0,0210	3,4907	3,6775	18,9642	16,3079
δ_s, $\mu = 0{,}4$, $\phi = 100$	0,0000	0,0000	0,0037	0,0037	3,7220	3,6756	16,8567	16,7500
δ_s, $\mu = 0{,}6$, $\phi = 2$	0,0001	0,0001	0,0249	0,0239	2,5823	2,9657	15,6590	14,0930
δ_s, $\mu = 0{,}6$, $\phi = 100$	0,0000	0,0000	0,0044	0,0044	3,7919	3,8267	16,7634	16,7709
δ_l, $\mu = 0{,}4$, $\phi = 2$	0,0002	-0,0007	0,0318	0,0307	8,6767	8,3245	31,3073	26,4972
δ_l, $\mu = 0{,}4$, $\phi = 100$	-0,0001	-0,0001	0,0055	0,0055	8,2501	8,2402	27,7933	27,6028
δ_l, $\mu = 0{,}6$, $\phi = 2$	-0,0006	-0,0007	0,0367	0,0354	6,5076	7,4273	25,3738	23,6132
δ_l, $\mu = 0{,}6$, $\phi = 100$	0,0000	0,0000	0,0063	0,0063	8,4845	8,5580	27,5290	27,5513
Valores dos parâmetros	$\mathrm{E}(Y)$				$\mathrm{Var}(Y)$			
	Viés		REQM		Viés		REQM	
	$\mathrm{E}(\tilde{Y})$	$\mathrm{E}(\hat{Y})$	$\mathrm{E}(\tilde{Y})$	$\mathrm{E}(\hat{Y})$	$\mathrm{Var}(\tilde{Y})$	$\mathrm{Var}(\hat{Y})$	$\mathrm{Var}(\tilde{Y})$	$\mathrm{Var}(\hat{Y})$
δ_s, $\mu = 0{,}4$, $\phi = 2$	0,0000	-0,0003	0,0238	0,0232	-0,0005	-0,0007	0,0085	0,0083
δ_s, $\mu = 0{,}4$, $\phi = 100$	0,0003	0,0003	0,0153	0,0153	-0,0002	-0,0002	0,0069	0,0069
δ_s, $\mu = 0{,}6$, $\phi = 2$	-0,0001	-0,0001	0,0285	0,0278	-0,0007	-0,0009	0,0084	0,0081
δ_s, $\mu = 0{,}6$, $\phi = 100$	0,0003	0,0003	0,0192	0,0192	-0,0005	-0,0005	0,0084	0,0084
δ_l, $\mu = 0{,}4$, $\phi = 2$	0,0000	-0,0003	0,0313	0,0311	-0,0008	-0,0010	0,0144	0,0143
δ_l, $\mu = 0{,}4$, $\phi = 100$	-0,0003	-0,0003	0,0288	0,0288	-0,0009	-0,0009	0,0140	0,0140
δ_l, $\mu = 0{,}6$, $\phi = 2$	-0,0005	-0,0006	0,0362	0,0360	-0,0016	-0,0018	0,0120	0,0117
δ_l, $\mu = 0{,}6$, $\phi = 100$	0,0001	0,0001	0,0331	0,0331	-0,0011	-0,0011	0,0103	0,0103

A Figura 2.1 compara a função distribuição acumulada (FDA) da distribuição BIZUT com parâmetros $\delta_0 = 0{,}3$, $\delta_1 = 0{,}1$, $\mu = 0{,}4$, $\phi = 2$ e alguns valores de c e δ_c e a FDA da distribuição BIZU com mesma média, mesma variância e mesma probabilidade de assumir os valores zero e um. Pode-se notar que para $c = 0{,}01$ e $\delta_c = 0{,}01$, ambas as distribuições têm aproximadamente a mesma FDA. No entanto, à medida que c e δ_c crescem, a diferença entre as FDAs das duas distribuições aumenta rapidamente. O valor de c parece acarretar maior diferença nas FDAs do que o valor de δ_c. A diferença entre as duas FDAs, por exemplo, é maior quando $c = 0{,}3$ e $\delta_c = 0{,}01$ do que quando

$c = 0{,}01$ e $\delta_c = 0{,}3$. Os gráficos sugerem que se um indivíduo não notar que $P(Y \in (0, c)) = 0$ e $P(Y = c) > 0$ e ajustar a distribuição BIZU, ele cometerá um erro considerável. O ajuste obtido será razoável apenas se tanto c quanto δ_c forem bem pequenos.

2.6 Aplicação

O valor do seguro desemprego no Brasil é uma função dos três últimos salários do desempregado. Há alguns requisitos para se ter direito ao benefício, sendo que um deles é ter recebido salários consecutivos nos últimos 6 meses anteriores à data de demissão. Cada desempregado elegível ao benefício tem direito a receber entre 3 e 5 parcelas (varia de acordo com o número de meses trabalhados nos 3 anos anteriores à data de demissão) de mesmo valor do seguro desemprego. Se o desempregado, enquanto recebe o benefício, obtém um novo emprego, ele perde o direito de receber as parcelas restantes. O valor do benefício varia (em reais no ano de 2009) no intervalo fechado $[465{,}00; 870{,}01]$. Muitos desempregados recebem o valor mínimo do benefício e muitos outros recebem o valor máximo. Além disso, as pessoas que não são elegíveis ao benefício não recebem nenhum valor. Portanto, quando deseja-se estudar o valor do seguro desemprego como proporção do valor máximo do benefício, a distribuição BIZUT pode ser uma escolha razoável para ajustar a variável.

Obteve-se uma amostra de 2322 desempregados do estado da Bahia que receberam a primeira parcela do seguro desemprego em Maio/2009. Os dados foram fornecidos pelo *Ministério do Trabalho e Emprego do Brasil*. Estudou-se a variável definida como a razão entre o valor da terceira parcela de desemprego e o valor máximo do benefício (valor proporcional do seguro desemprego na Bahia, VPSDBA). Assim, observações de valores zero são referentes a pessoas que perderam direito ao benefício, por exemplo, porque obtiveram um novo emprego. Na amostra, os valores zero, um e $c = 0{,}5345$ foram assumidos por, respectivamente, 99, 198 e 1210 pessoas. A média e o desvio padrão amostral para as observações no intervalo $(c, 1)$ foram, respectivamente, 0,7196 e 0,1258. Os valores no intervalo contínuo da variável variaram entre 0,5346 e 0,9983 e os quartis obtidos foram respectivamente, 0,6050, 0,7078 e 0,8092. O histograma da variável no intervalo $(c, 1)$ (Figura 2.2) sugere que a variável VPSDBA tem distribuição assimétrica positiva no intervalo $(0{,}5345, 1)$. Utilizando-se os resultados dados na Seção 2.3, obtiveram-se estimativas de máxima verossimilhança (MV) e intervalos de confiança para os parâmetros (Tabela 2.3). A partir da equação (2.20) obteve-se ainda uma região de confiança para os 5 parâmetros da distribuição BIZUT, que foi omitida por conter muitos termos.

Tabela 2.3: Estimativas pontual e intervalar (95% de confiança) dos parâmetros da distribuição BIZUT para a variável valor proporcional do seguro desemprego na Bahia.

	δ_0	δ_1	δ_c	μ	ϕ
Estimativas MV	0,0426	0,0853	0,5211	0,7226	2,2250
Limite inferior	0,0344	0,0739	0,5008	0,7142	2,0393
Limite superior	0,0509	0,0966	0,5414	0,7310	2,4107

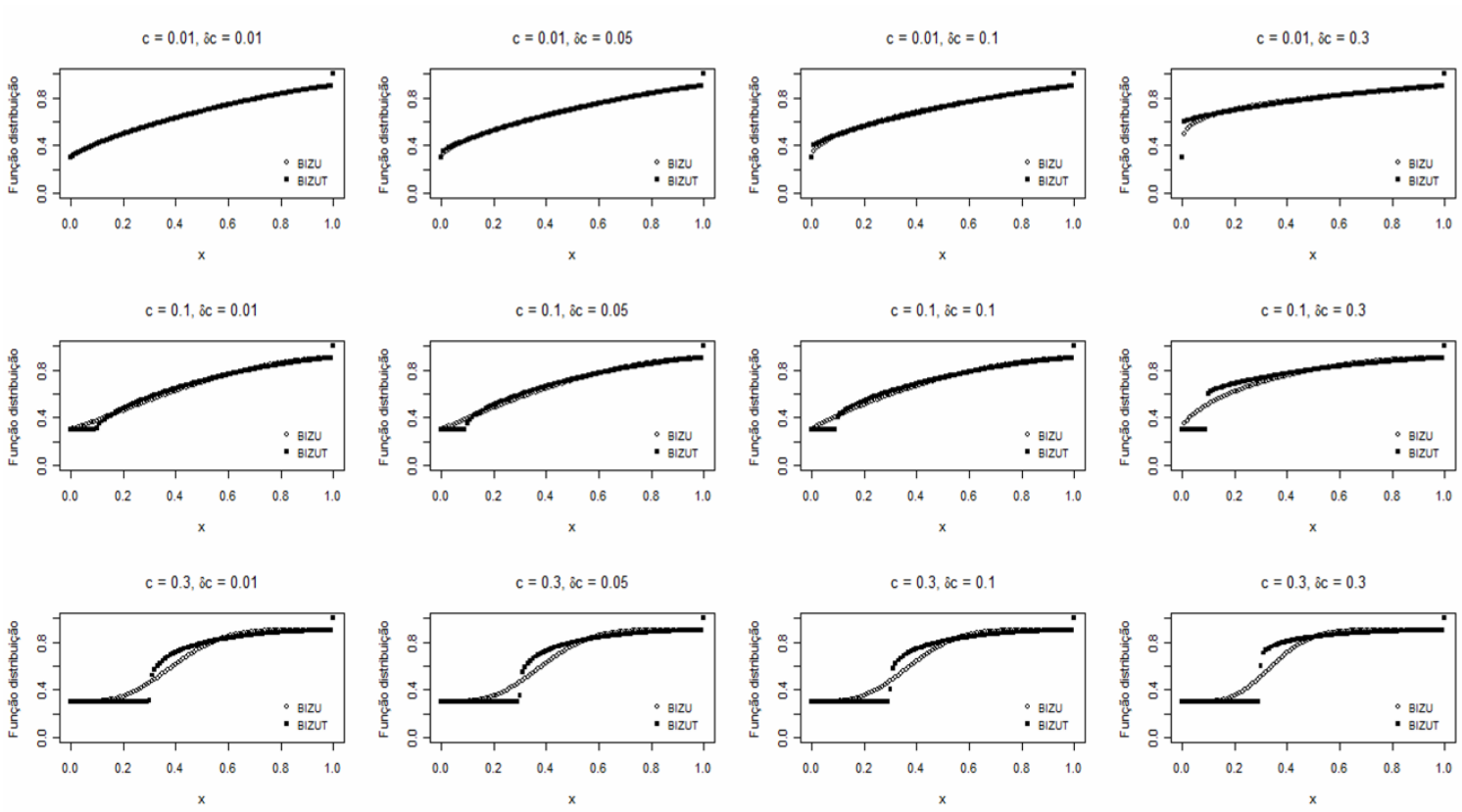

Figura 2.1: Função distribuição BIZUT para alguns valores de c e δ_c ($\mu = 0{,}4$, $\phi = 2$, $\delta_0 = 0{,}3$, $\delta_1 = 0{,}1$) e função distribuição BIZU com a mesma média, a mesma variância e os mesmos valores de δ_0 e δ_1.

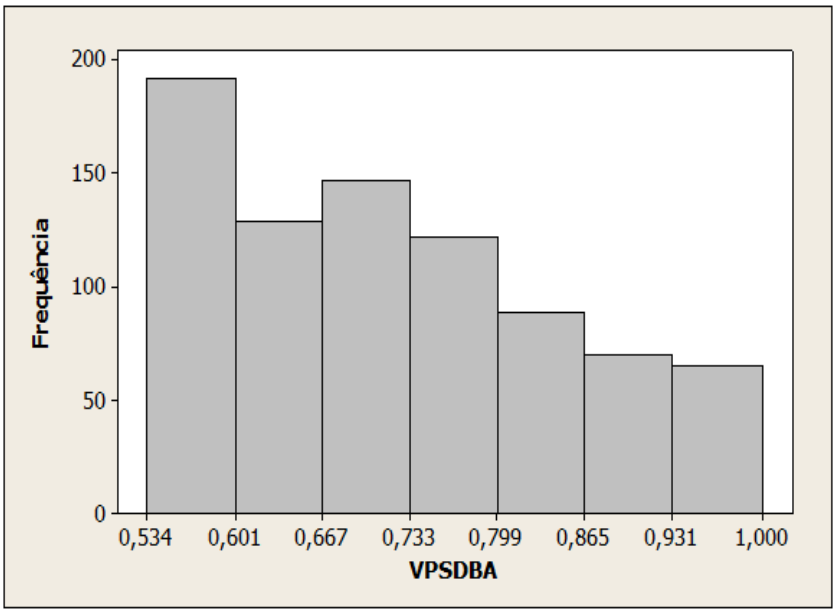

Figura 2.2: Histograma para a variável VPSDBA no estado da Bahia.

O intervalo de confiança e a região de confiança obtidos são baseados nas propriedades assintóticas do estimador de máxima verossimilhança. Para estudar a performance deles em uma amostra de tamanho 2322, conduziu-se um estudo de simulação de Monte Carlo com 10000 réplicas. O estudo foi feito para uma variável com distribuição BIZUT e parâmetros iguais aos estimados para a variável VPSDBA e apresentados na Tabela 2.3. A Tabela 2.4 apresenta estimativas da probabilidade de cobertura dos intervalos e da região de confiança. Pode-se notar que todas as estimativas estão próximas dos coeficientes de confiança. A tabela sugere que os intervalos e região de confiança apresentados em (2.19) e (2.20) têm boa performance se a amostra tem tamanho 2322. A distribuição BIZUT foi proposta neste trabalho para ajustar algumas variáveis na área de econometria. Nesta área, geralmente não há dificuldade em se obter amostras ainda maiores do que a que utilizamos neste capítulo. Por esta razão, na prática, as expressões (2.19) e (2.20) apresentam boa performance para estimar os parâmetros e testar hipóteses.

Tabela 2.4: Estimativas da probabilidade de cobertura dos intervalos e região de confiança para os parâmetros da distribuição BIZUT (n = 2322 e os valores dos parâmetros são iguais às estimativas apresentadas na Tabela 2.3).

Coeficiente de confiança	δ_0	δ_1	δ_c	μ	ϕ	Região
0,90	0,8940	0,8962	0,9026	0,8971	0,9062	0,8983
0,95	0,9449	0,9472	0,9501	0,9533	0,9507	0,9503
0,99	0,9899	0,9880	0,9919	0,9899	0,9893	0,9906

A Figura 2.3, no gráfico à esquerda, traz a função distribuição empírica da variável VPSDBA e a FDA da distribuição BIZUT com parâmetros iguais às estimativas de máxima verossimilhança

apresentadas na Tabela 2.3. Pode-se notar que ambas as funções distribuição estão bem próximas entre si. No entanto, como a distribuição BIZUT tem um parâmetro para cada um dos valores discretos que a variável pode assumir ($\delta_0 = P(Y = 0)$, por exemplo), ambas as curvas são exatamente iguais no intervalo $[0, c]$. Assim, é mais conveniente estudar a função distribuição condicional (P(VPSDBA $\leq x | c <$ VPSDBA < 1)). O gráfico à direita da Figura 2.3 apresenta as funções distribuição condicionais empírica e da distribuição BIZUT. Ao longo de todo o intervalo $(0{,}5345, 1{,}0000)$, ambas as curvas não estão distantes entre si. Para nenhum valor x, a diferença entre as duas funções distribuição em x é, em módulo, superior a 0,04. Como a variável VPSDBA obviamente tem probabilidade positiva de assumir os valores zero, um e c e pode assumir qualquer valor no intervalo $(c, 1)$, ela terá distribuição BIZUT se ela tiver distribuição beta no intervalo $(c, 1)$. Para testar a hipótese de que a variável VPSDBA tem distribuição beta no intervalo $(c, 1)$, utilizou-se o teste de aderência de Anderson-Darling com método da meia-amostra (Stephens, 1986, página 169). A hipótese de que a variável VPSDBA tem distribuição beta no intervalo $(c, 1)$ não foi rejeitada ($p > 0,10$). A Figura 2.3 e o resultado do teste de aderência sugerem que é razoável ajustar a distribuição BIZUT para a variável VPSDBA.

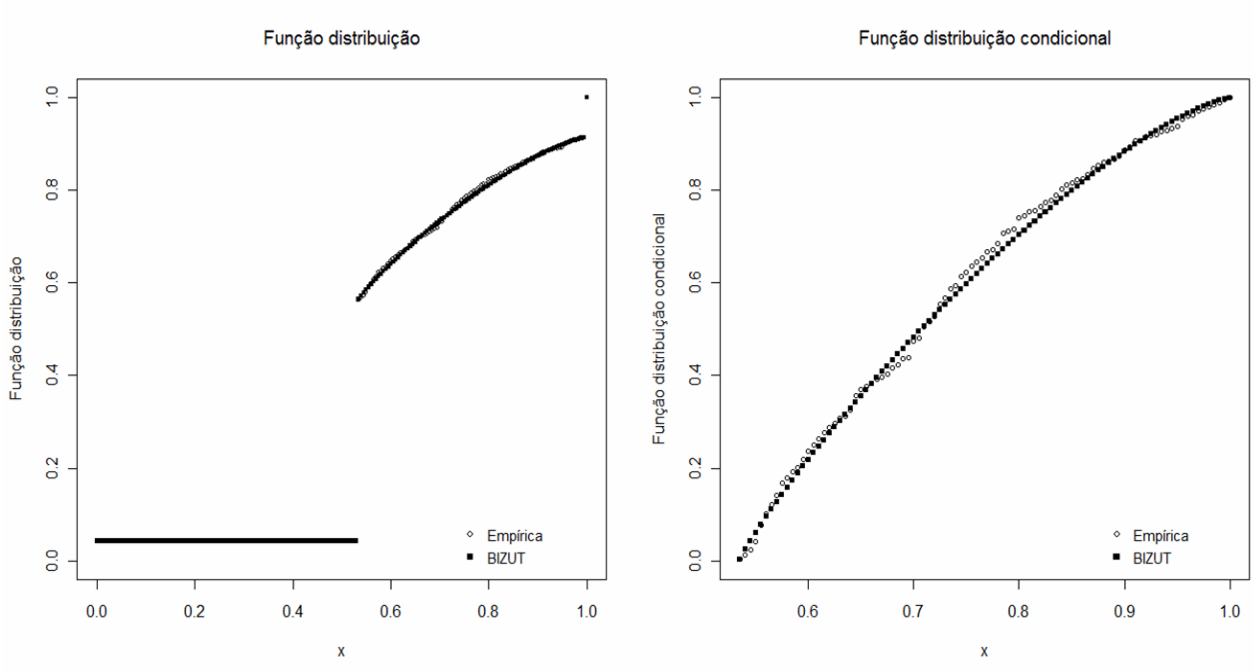

Figura 2.3: Função distribuição para a variável VPSDBA.

2.7 Conclusão

Neste capítulo introduzimos a distribuição BIZUT, que é discreta nos pontos zero, um e c e contínua no intervalo aberto $(c, 1)$. Foram apresentadas as propriedades da distribuição, discutidos métodos de estimação de seus parâmetros, executados estudos de simulação e conduzida uma aplicação a dados de seguro desemprego. Os resultados sugerem que a distribuição BIZUT pode ser

útil no estudo de variáveis do tipo proporção que não podem assumir valores no intervalo $(0, c)$.

Capítulo 3

Modelos de regressão beta inflacionados truncados

Neste capítulo são introduzidos os modelos de regressão beta inflacionados truncados (RBIZUT). Na Seção 3.1 são definidos os modelos RBIZUT, obtêm-se o vetor escore e a matriz de informação de Fisher e discute-se a estimação dos parâmetros por máxima verossimilhança. Intervalos de confiança e testes de hipóteses para os modelos RBIZUT são desenvolvidos na Seção 3.2. Algumas questões relativas ao modelo RBIZUT com c variável são apresentadas e discutidas na Seção 3.3. Em seguida (Seção 3.4) são desenvolvidos estudos de simulação de Monte Carlo que avaliam a performance dos estimadores de máxima verossimilhança dos parâmetros dos modelos RBIZUT. Para finalizar o capítulo, discute-se seleção de modelos de regressão na Seção 3.5 e as considerações finais são feitas na Seção 3.6.

3.1 Definição e estimação pontual

Nos Capítulos 1 e 2 foram mencionadas algumas variáveis que têm probabilidade positiva de assumir os valores zero, um e c e podem assumir ainda qualquer valor no intervalo $(c, 1)$. Dessa forma, é razoável esperar que algumas dessas variáveis, em algumas populações, possam ser bem ajustadas pela distribuição beta inflacionada truncada. No entanto, é natural imaginar que, em muitas situações, mesmo que a utilização da distribuição beta seja factível, nem todas as unidades populacionais sejam identicamente distribuídas. Para os dados utilizados na aplicação do Capítulo 2, por exemplo, indivíduos com curso superior certamente têm maior probabilidade de receber o valor máximo do seguro desemprego do que aqueles que sequer completaram o primeiro grau. Assim, para o estudo dessas variáveis, é necessária a introdução de modelos de regressão beta inflacionados truncados. Eles assumem que, dadas as variáveis preditoras consideradas, a variável resposta do modelo tem distribuição BIZUT. No modelo proposto neste trabalho todos os cinco parâmetros desconhecidos da distribuição BIZUT são ajustados em função de variáveis preditoras. Além disso, admitimos que o valor c da distribuição BIZUT, mesmo sendo conhecido, possa variar entre as unidades populacionais. No exemplo do cartão de crédito mencionado no Capítulo 1, por exemplo, o parâmetro conhecido c para a distribuição do valor relativo do pagamento varia em função da unidade populacional, já que o valor mínimo de pagamento como proporção do valor da fatura não é constante para todos os clientes. Já no exemplo da aplicação do Capítulo 2, c é constante, já que nenhum brasileiro pode receber parcela de seguro desemprego inferior a um salário minimo. Poderíamos propor um modelo considerando c fixo e uma extensão para a qual fosse permitido que

c variasse entre as unidades populacionais. No entanto, observou-se que as propriedades dos dois modelos são bem semelhantes e o fato de c ser fixo pouco simplifica o modelo mais geral. Dessa forma propomos a seguinte definição para os modelos de regressão beta inflacionados truncados.

Definição 3.1.1. *Sejam $y_1, y_2, \ldots, y_n$ variáveis aleatórias independentes, cada uma delas com distribuição BIZUT$(\delta_{i0}, \delta_{i1}, \delta_{ic}, \mu_i, \phi_i, c_i)$ conforme definido em (2.6), ou seja, com função densidade de probabilidade dada por*

$$f(y_i; \delta_{i0}, \delta_{i1}, \delta_{ic}, \mu_i, \phi_i, c_i) = \begin{cases} \delta_{i0} & se\ y_i = 0 \\ \delta_{i1} & se\ y_i = 1 \\ \delta_{ic} & se\ y_i = c_i \\ (1 - \delta_{i0} - \delta_{i1} - \delta_{ic}) f(y_i; \mu_i, \phi_i, c_i, 1) & se\ y_i \in (c_i, 1), \end{cases} \tag{3.1}$$

em que $f(y_i; \mu_i, \phi_i, c_i, 1)$ é a função densidade de probabilidade de uma variável com distribuição beta e vetor de parâmetros $(\mu_i, \phi_i, c_i, 1)$, $c \in (0, 1)$ conhecido. Os modelos de regressão beta inflacionados truncados (RBIZUT) são definidos por (3.1) e pelos seguintes componentes sistemáticos:

$$\begin{cases} g_1(\mu_i) & = \eta_{i1} \\ g_2(\phi_i) & = \eta_{i2} \\ H(\delta_{i0}, \delta_{i1}, \delta_{ic}) & = (\zeta_{i0}, \zeta_{i1}, \zeta_{ic}), \end{cases} \tag{3.2}$$

em que $\eta_{i1} = x_{i1}^\top \beta_1$, $\eta_{i2} = x_{i2}^\top \beta_2$, $\zeta_{i0} = z_{i0}^\top \gamma_0$, $\zeta_{i1} = z_{i1}^\top \gamma_1$ e $\zeta_{ic} = z_{ic}^\top \gamma_c$ são os preditores lineares, $\beta_1 = (\beta_{11}, \beta_{21}, \ldots, \beta_{p_\mu 1})^\top$, $\beta_2 = (\beta_{12}, \beta_{22}, \ldots, \beta_{p_\phi 2})^\top$, $\gamma_0 = (\gamma_{10}, \gamma_{20}, \ldots, \gamma_{p_0 0})^\top$, $\gamma_1 = (\gamma_{11}, \gamma_{21}, \ldots, \gamma_{p_1 1})^\top$ e $\gamma_c = (\gamma_{1c}, \gamma_{2c}, \ldots, \gamma_{p_c c})^\top$ são os vetores de parâmetros desconhecidos, $x_{i1} = (x_{i11}, x_{i21}, \ldots, x_{ip_\mu 1})^\top$, $x_{i2} = (x_{i12}, x_{i22}, \ldots, x_{ip_\phi 2})^\top$, $z_{i0} = (z_{i10}, z_{i20}, \ldots, z_{ip_0 0})^\top$, $z_{i1} = (z_{i11}, z_{i21}, \ldots, z_{ip_1 1})^\top$ e $z_{ic} = (z_{i1c}, z_{i2c}, \ldots, z_{ip_c c})^\top$ são constantes que representam os valores das variáveis preditoras para cada um dos parâmetros da distribuição BIZUT, g_1 e g_2 são funções de ligação estritamente monótonas e duplamente diferenciáveis de $(c_i, 1)$ em $\mathbb{R}$ e de $\mathbb{R}^+$ em $\mathbb{R}$, respectivamente, e H é uma função de ligação bijetora duplamente diferenciável de C em $\mathbb{R}^3$, em que C é um subespaço de $\mathbb{R}^3$ definido como $C = \{(\delta_{i0}, \delta_{i1}, \delta_{ic}) : 0 < \delta_{i0} < 1, 0 < \delta_{i1} < 1 - \delta_{i0}, 0 < \delta_{ic} < 1 - \delta_{i0} - \delta_{ii}\}$.

Observe que há uma única função de ligação abrangendo simultaneamente os parâmetros δ_{i0}, δ_{i1} e δ_{ic}. É necessário definir o modelo desta forma para garantir que a restrição linear $0 < \delta_{i0} + \delta_{i1} + \delta_{ic} < 1$ seja atendida. Como δ_{i0}, δ_{i1} e δ_{ic} são probabilidades da variável resposta assumir determinados valores, é natural utilizar a mesma função H usada na regressão logística multinomial (Hosmer e Lemeshow, 2000, página 261), ou seja, $H(\delta_{i0}, \delta_{i1}, \delta_{ic}) = (h_0(\delta_{i0}, \delta_{i1}, \delta_{ic}), h_1(\delta_{i0}, \delta_{i1}, \delta_{ic}), h_c(\delta_{i0}, \delta_{i1}, \delta_{ic})) = (\log[\delta_{i0}/(1-\delta_{i0}-\delta_{i1}-\delta_{ic})], \log[\delta_{i1}/(1-\delta_{i0}-\delta_{i1}-\delta_{ic})], \log[\delta_{ic}/(1-\delta_{i0}-\delta_{i1}-\delta_{ic})])$. Neste caso, exponenciais dos parâmetros γ podem ser interpretados como acréscimos relativos na razão de chances. Assim, por exemplo, $\exp\{\gamma_{20}\}$ é o valor pelo qual é multiplicada a chance de ocorrência do valor 0 na variável resposta (em relação a ocorrência de valor no intervalo contínuo

da variável resposta) quando aumenta-se z_{20} em uma unidade e mantém-se as demais variáveis preditoras constantes. Embora a função H mencionada seja a mais simples, conforme observou Ospina (2008) em seu modelo inflacionado nos pontos 0 e 1, a função H pode ser definida até mesmo de forma genérica, por exemplo, utilizando

$$H(\delta_{i0}, \delta_{i1}, \delta_{ic}) = \left(h\left(\frac{\delta_{i0}}{1-\alpha_i}\right), h\left(\frac{\delta_{i1}}{1-\alpha_i}\right), h\left(\frac{\delta_{ic}}{1-\alpha_i}\right)\right), \tag{3.3}$$

em que $\alpha_i = \delta_{i0} + \delta_{i1} + \delta_{ic}$ e h é uma função estritamente monótona e duplamente diferenciável de $\mathbb{R}^+$ em $\mathbb{R}$. Observe que a função H utilizada na regressão logística multinomial corresponde a (3.3) com h sendo a função logarítmica.

No estudo da média de variáveis do tipo proporção, a razão $\mu/(1-\mu)$ também pode ser vista como uma chance (Ferrari e Cribari-Neto, 2004). Assim, uma extensão natural da chance para variáveis de média μ limitadas no intervalo (a, b), é a razão $(\mu-a)/(b-\mu)$. Logo, uma escolha interessante para a função de ligação $g_1(.)$ para o modelo RBIZUT é a extensão natural da ligação logito para parâmetros com espaço paramétrico no intervalo $(c_i, 1)$ dada por $g_1(\mu_i) = \log[(\mu_i - c_i)/(1 - \mu_i)]$. Neste caso, $\exp\{\beta_{j1}\}$ é a razão entre a chance da variável resposta quando $x_{ij1} = k + 1$ e quando $x_{ij1} = k$ dado que a variável dependente assumiu valores no intervalo $(c_i, 1)$ e mantidas as demais variáveis preditoras constantes. Como $g_2(.)$ é uma função de $\mathbb{R}^+$ em $\mathbb{R}$, pode-se utilizar, por exemplo, para o parâmetro ϕ_i, $g_2(\phi_i) = \log(\phi_i)$. De forma similar ao denominado por Ospina (2008), utilizaremos o termo regressão logística beta inflacionada truncada (RLBIZUT) para designar o modelo RBIZUT da Definição 3.1.1 quando $g_1(\mu_i) = \log[(\mu_i - c_i)/(1 - \mu_i)]$, $g_2(\phi_i) = \log(\phi_i)$ e $H(\delta_{i0}, \delta_{i1}, \delta_{ic}) = (\log[\delta_{i0}/(1-\delta_{i0}-\delta_{i1}-\delta_{ic})], \log[\delta_{i1}/(1-\delta_{i0}-\delta_{i1}-\delta_{ic})], \log[\delta_{ic}/(1-\delta_{i0}-\delta_{i1}-\delta_{ic})])$. Todos os resultados apresentados nesta seção e na próxima valem para qualquer modelo RBIZUT. Nas situações em que há uma simplificação considerável dos resultados, apresentamos também a versão válida apenas para os modelos RLBIZUT.

Os modelos RBIZUT podem ser vistos como uma extensão dos modelos de regressão beta inflacionados em zero e um (RBIZU) introduzidos por Ospina (2008). Os modelos RBIZUT estendem os modelos RBIZU em relação a três aspectos. O primeiro é que o modelo permite a existência de um terceiro ponto de inflação c e que a variável resposta não assuma valores no intervalo $(0, c)$. O segundo aspecto é que o modelo admite que o suporte da variável resposta varie entre as unidades populacionais. O último é que o modelo admite que o parâmetro de precisão ϕ varie em função de outras variáveis. Os modelos RBIZUT pertencem ainda à classe dos modelos aditivos generalizados para locação, escala e forma (GAMLSS) introduzidos por Rigby e Stasinopoulos (2005). No entanto, a biblioteca GAMLSS do software R desenvolvida pelos autores do artigo para ajustar modelos desta classe não pode ser utilizada para o ajuste de modelos RBIZUT. O motivo é que os modelos RBIZUT possuem 5 vetores de parâmetros e a biblioteca admite a existência de no máximo 4.

A partir da Definição 3.1.1, obtém-se que o logaritmo da função de verossimilhança dos modelos

RBIZUT baseada em uma amostra aleatória contendo n observações independentes é dado por

$$l(\theta) = l_1(\gamma_0, \gamma_1, \gamma_c) + l_2(\beta_1, \beta_2), \tag{3.4}$$

em que

$$\begin{aligned} l_1(\gamma_0, \gamma_1, \gamma_c) &= \sum_{i=1}^{n} l_{i1}(\delta_{i0}, \delta_{i1}, \delta_{ic}), \\ l_2(\beta_1, \beta_2) &= \sum_{i:y_i \in (c_i,1)} l_{i2}(\mu_i, \phi_i), \end{aligned} \tag{3.5}$$

$$\begin{aligned} l_{i1}(\delta_{i0}, \delta_{i1}, \delta_{ic}) &= I_{\{0\}}(y_i)\log\delta_{i0} + I_{\{1\}}(y_i)\log\delta_{i1} + I_{\{c_i\}}(y_i)\log\delta_{ic} \\ &\quad +(1 - I_{\{0\}}(y_i) - I_{\{1\}}(y_i) - I_{\{c_i\}}(y_i))\log(1 - \delta_{i0} - \delta_{i1} - \delta_{ic}), \\ l_{i2}(\mu_i, \phi_i) &= \log\left\{\frac{\Gamma(\phi_i)}{\Gamma\left[\left(\frac{\mu_i - c_i}{1-c_i}\right)\phi_i\right]\Gamma\left[\left(\frac{1-\mu_i}{1-c_i}\right)\phi_i\right](1-c_i)^{\phi_i - 1}}\right\} + \\ &\quad \left[\left(\frac{\mu_i - c_i}{1-c_i}\right)\phi_i - 1\right]\log(y_i - c_i) + \left[\left(\frac{1-\mu_i}{1-c_i}\right)\phi_i - 1\right]\log(1 - y_i) \end{aligned}$$

e δ_{i0}, δ_{i1}, δ_{ic}, μ_i e ϕ_i são definidos implicitamente por (3.2). De forma coerente com o que foi observado no Capítulo 2, os vetores de parâmetros $(\gamma_0^\top, \gamma_1^\top, \gamma_c^\top)^\top$ e $(\beta_1^\top, \beta_2^\top)^\top$ são separáveis e a estimação de $(\gamma_0^\top, \gamma_1^\top, \gamma_c^\top)^\top$ pode ser feita separadamente da estimação de $(\beta_1^\top, \beta_2^\top)^\top$ e vice-versa.

A função escore é obtida derivando-se o logaritmo da função de verossimilhança (3.4) em relação a cada um dos parâmetros do modelo RBIZUT. Após alguns cálculos algébricos (ver Apêndice B.2), obtém-se que a função escore do modelo RBIZUT é dada por

$$U(\theta) = (U_D(\gamma_0, \gamma_1, \gamma_c)^\top, U_C(\beta_1, \beta_2)^\top)^\top, \tag{3.6}$$

em que

$$\begin{aligned} U_D(\gamma_0, \gamma_1, \gamma_c) &= (U_{\gamma_0}(\gamma_0, \gamma_1, \gamma_c)^\top, U_{\gamma_1}(\gamma_0, \gamma_1, \gamma_c)^\top, U_{\gamma_c}(\gamma_0, \gamma_1, \gamma_c)^\top)^\top \text{ e} \\ U_C(\beta_1, \beta_2) &= (U_{\beta_1}(\beta_1, \beta_2)^\top, U_{\beta_2}(\beta_1, \beta_2)^\top)^\top. \end{aligned} \tag{3.7}$$

Assim, como consequência do fato dos parâmetros associados à parte discreta e à parte contínua da distribuição BIZUT serem separáveis, os componentes de $U(\theta)$ correspondentes à γ_0, γ_1 e γ_c são função apenas dos parâmetros γ_0, γ_1 e γ_c, enquanto os componentes correspondentes à β_1 e β_2 são função apenas de β_1 e β_2.

Para apresentar cada um dos vetores que compõem (3.7), é necessário definir algumas transformações de y_i e alguns vetores e matrizes. Seja

$$y_i^* = \begin{cases} \log\left(\frac{y_i - c_i}{1 - y_i}\right) & \text{se } y_i \in (c_i, 1) \\ 0 & \text{caso contrário} \end{cases} \tag{3.8}$$

e

$$y_i^\bullet = \begin{cases} \log\left(\frac{1-y_i}{(y_i-c_i)^{c_i}}\right) & \text{se } y_i \in (c_i, 1) \\ 0 & \text{caso contrário.} \end{cases} \tag{3.9}$$

Sob condições de regularidade, é provado no Apêndice B.1 que os primeiros momentos condicionais (dado que $y_i \in (c_i, 1)$) centrais de y_i^* e $y_i^\bullet$ são dados por

$$\begin{aligned}
\mu_i^* &= \mathrm{E}(y_i^*|I_{(c_i,1)}(y_i)=1) = \psi\left[\left(\frac{\mu_i - c_i}{1-c_i}\right)\phi_i\right] - \psi\left[\left(\frac{1-\mu_i}{1-c_i}\right)\phi_i\right], \\
\mu_i^\bullet &= \mathrm{E}(y_i^\bullet|I_{(c_i,1)}(y_i)=1) = \psi\left[\left(\frac{1-\mu_i}{1-c_i}\right)\phi_i\right] - c_i\psi\left[\left(\frac{\mu_i - c_i}{1-c_i}\right)\phi_i\right] + (1-c_i)[\log(1-c_i) - \\
&\quad \psi(\phi_i)], \\
\upsilon_i^* &= \mathrm{Var}(y_i^*|I_{(c_i,1)}(y_i)=1) = \psi'\left[\left(\frac{\mu_i - c_i}{1-c_i}\right)\phi_i\right] + \psi'\left[\left(\frac{1-\mu_i}{1-c_i}\right)\phi_i\right], \\
\upsilon_i^\bullet &= \mathrm{Var}(y_i^\bullet|I_{(c_i,1)}(y_i)=1) = \psi'\left[\left(\frac{1-\mu_i}{1-c_i}\right)\phi_i\right] + c_i^2\psi'\left[\left(\frac{\mu_i - c_i}{1-c_i}\right)\phi_i\right] + (1-c_i)^2\psi'(\phi_i), \\
\xi^{*\bullet} &= \mathrm{Cov}(y_i^*, y_i^\bullet|I_{(c_i,1)}(y_i)=1) = -\psi'\left[\left(\frac{1-\mu_i}{1-c_i}\right)\phi_i\right] - c_i\psi'\left[\left(\frac{\mu_i - c_i}{1-c_i}\right)\phi_i\right],
\end{aligned} \tag{3.10}$$

em que $\psi(.)$ é a função digama.

Definimos ainda os vetores $y_{\{0\}} = (I_{\{0\}}(y_1), \ldots, I_{\{0\}}(y_n))^\top$, $y_{\{1\}} = (I_{\{1\}}(y_1), \ldots, I_{\{1\}}(y_n))^\top$, $y_{\{c\}} = (I_{\{c_1\}}(y_1), \ldots, I_{\{c_n\}}(y_n))^\top$, $y_{(c,1)} = (I_{(c_1,1)}(y_1), \ldots, I_{(c_n,1)}(y_n))^\top$, $y^* = (y_1^*, \ldots, y_n^*)^\top$, $y^\bullet = (y_1^\bullet, \ldots, y_n^\bullet)^\top$ $\mu = (\mu_1, \ldots, \mu_n)^\top$, $\mu^* = (\mu_1^*, \ldots, \mu_n^*)^\top$ e $\mu^\bullet = (\mu_1^\bullet, \ldots, \mu_n^\bullet)^\top$ e as matrizes diagonais $T_0 = \mathrm{diag}\{\partial\delta_{10}/\partial\zeta_{10}, \ldots, \partial\delta_{n0}/\partial\zeta_{n0}\}$, $T_1 = \mathrm{diag}\{\partial\delta_{11}/\partial\zeta_{11}, \ldots, \partial\delta_{n1}/\partial\zeta_{n1}\}$, $T_c = \mathrm{diag}\{\partial\delta_{1c}/\partial\zeta_{1c}, \ldots, \partial\delta_{nc}/\partial\zeta_{nc}\}$, $T_\mu = \mathrm{diag}\{\partial\mu_1/\partial\eta_{11}, \ldots, \partial\mu_n/\partial\eta_{n1}\}$, $T_\phi = \mathrm{diag}\{\partial\phi_1/\partial\eta_{12}, \ldots, \partial\phi_n/\partial\eta_{n2}\}$, $T_{01} = \mathrm{diag}\{\partial\delta_{10}/\partial\zeta_{11}, \ldots, \partial\delta_{n0}/\partial\zeta_{n1}\}$, $T_{10} = \mathrm{diag}\{\partial\delta_{11}/\partial\zeta_{10}, \ldots, \partial\delta_{n1}/\partial\zeta_{n0}\}$, $T_{0c} = \mathrm{diag}\{\partial\delta_{10}/\partial\zeta_{1c}, \ldots, \partial\delta_{n0}/\partial\zeta_{nc}\}$, $T_{c0} = \mathrm{diag}\{\partial\delta_{1c}/\partial\zeta_{10}, \ldots, \partial\delta_{nc}/\partial\zeta_{n0}\}$, $T_{1c} = \mathrm{diag}\{\partial\delta_{11}/\partial\zeta_{1c}, \ldots, \partial\delta_{n1}/\partial\zeta_{nc}\}$, $T_{c1} = \mathrm{diag}\{\partial\delta_{1c}/\partial\zeta_{11}, \ldots, \partial\delta_{nc}/\partial\zeta_{n1}\}$, $\Delta_0 = \mathrm{diag}\{1/\delta_{10}, \ldots, 1/\delta_{n0}\}$, $\Delta_1 = \mathrm{diag}\{1/\delta_{11}, \ldots, 1/\delta_{n1}\}$, $\Delta_c = \mathrm{diag}\{1/\delta_{1c}, \ldots, 1/\delta_{nc}\}$, $\Delta_{(c,1)} = \mathrm{diag}\{1/(1-\alpha_1), \ldots, 1/(1-\alpha_n)\}$, $\phi^* = \mathrm{diag}\{\phi_1/(1-c_1), \ldots, \phi_n/(1-c_n)\}$ e $c^* = \mathrm{diag}\{1/(1-c_1), \ldots, 1/(1-c_n)\}$. Assim, os vetores que compõem a função escore (3.6) são dados por

$$\begin{aligned}
U_{\gamma_0}(\gamma_0, \gamma_1, \gamma_c) &= Z_0^\top T_0(\Delta_0 y_{\{0\}} - \Delta_{(c,1)} y_{(c,1)}) + Z_0^\top T_{10}(\Delta_1 y_{\{1\}} - \Delta_{(c,1)} y_{(c,1)}) \\
&\quad + Z_0^\top T_{c0}(\Delta_c y_{\{c\}} - \Delta_{(c,1)} y_{(c,1)}), \\
U_{\gamma_1}(\gamma_0, \gamma_1, \gamma_c) &= Z_1^\top T_{01}(\Delta_0 y_{\{0\}} - \Delta_{(c,1)} y_{(c,1)}) + Z_1^\top T_1(\Delta_1 y_{\{1\}} - \Delta_{(c,1)} y_{(c,1)}) \\
&\quad + Z_1^\top T_{c1}(\Delta_c y_{\{c\}} - \Delta_{(c,1)} y_{(c,1)}), \\
U_{\gamma_c}(\gamma_0, \gamma_1, \gamma_c) &= Z_c^\top T_{0c}(\Delta_0 y_{\{0\}} - \Delta_{(c,1)} y_{(c,1)}) + Z_c^\top T_{1c}(\Delta_1 y_{\{1\}} - \Delta_{(c,1)} y_{(c,1)}) \\
&\quad + Z_c^\top T_c(\Delta_c y_{\{c\}} - \Delta_{(c,1)} y_{(c,1)}), \\
U_{\beta_1}(\beta_1, \beta_2) &= X_1^\top \phi^* \mathrm{mdiag}(y_{(c,1)}) T_\mu (y^* - \mu^*), \\
U_{\beta_2}(\beta_1, \beta_2) &= X_2^\top c^* \mathrm{mdiag}(y_{(c,1)}) T_\phi [\mathrm{mdiag}(\mu)(y^* - \mu^*) + (y^\bullet - \mu^\bullet)],
\end{aligned} \tag{3.11}$$

em que $Z_0 = (z_{10}, \ldots, z_{n0})^\top$, $Z_1 = (z_{11}, \ldots, z_{n1})^\top$, $Z_c = (z_{1c}, \ldots, z_{nc})^\top$, $X_1 = (x_{11}, \ldots, x_{n1})^\top$, $X_2 = (x_{12}, \ldots, x_{n2})^\top$ e mdiag(.) é o operador que transforma um vetor em uma matriz diagonal.

O componente $U_D(\gamma_0, \gamma_1, \gamma_c)$ do vetor escore definido em (3.7) pode ser escrito de forma mais compacta utilizando-se matrizes aumentadas. Sejam

$$\tilde{T} = \begin{pmatrix} T_0 & T_{10} & T_{c0} \\ T_{01} & T_1 & T_{c1} \\ T_{0c} & T_{1c} & T_c \end{pmatrix} \text{ e } \tilde{Z} = \begin{pmatrix} Z_0 & 0 & 0 \\ 0 & Z_1 & 0 \\ 0 & 0 & Z_c \end{pmatrix}, \tag{3.12}$$

matrizes de dimensão $(3n \times 3n)$ e $(3n \times (p_0 + p_1 + p_c))$, respectivamente, e seja $y_\Delta = ((\Delta_0 y_{\{0\}} - \Delta_{(c,1)} y_{(c,1)})^\top, (\Delta_1 y_{\{1\}} - \Delta_{(c,1)} y_{(c,1)})^\top, (\Delta_c y_{\{c\}} - \Delta_{(c,1)} y_{(c,1)})^\top)^\top$ vetor de dimensão $3n$. Então é imediato que

$$U_D(\gamma_0, \gamma_1, \gamma_c) = \tilde{Z}^\top \tilde{T} y_\Delta. \tag{3.13}$$

Os componentes do vetor escore $U_D(\gamma_0, \gamma_1, \gamma_c)$ simplificam-se consideravelmente se a função de ligação H do modelo (3.2) for a mesma do modelo RLBIZUT. Nesse caso é provado no Apêndice B.2 que

$$\begin{aligned} U_{\gamma_0}(\gamma_0, \gamma_1, \gamma_c) &= Z_0^\top (y_{\{0\}} - \delta_0), \\ U_{\gamma_1}(\gamma_0, \gamma_1, \gamma_c) &= Z_1^\top (y_{\{1\}} - \delta_1), \\ U_{\gamma_c}(\gamma_0, \gamma_1, \gamma_c) &= Z_c^\top (y_{\{c\}} - \delta_c), \end{aligned} \tag{3.14}$$

em que $\delta_0 = (\delta_{10}, \ldots, \delta_{n0})^\top$, $\delta_1 = (\delta_{11}, \ldots, \delta_{n1})^\top$ e $\delta_c = (\delta_{1c}, \ldots, \delta_{nc})^\top$.

O sistema de equações $U(\theta) = 0$ não tem solução algébrica. Por esse motivo, os estimadores de máxima verossimilhança dos parâmetros do modelo RBIZUT devem ser obtidos por métodos numéricos. Em todos os estudos de simulação de Monte Carlo e na aplicação do Capítulo 5 foi utilizado o método quase-Newton BFGS com derivadas analíticas (Nocedal e Wright, 2006, Seção 6.1).

Para a utilização do método BFGS e da maior parte dos métodos numéricos, é necessário fornecer valores iniciais para cada um dos parâmetros. De (3.5) pode-se notar que o termo do logaritmo da função de verossimilhança correspondente aos vetores de parâmetros β_1 e β_2 soma apenas as contribuições das observações i para as quais $y_i \in (c_i, 1)$. Por esse motivo, a forma mais simples de obter valores iniciais para $\hat{\beta}_1$ e $\hat{\beta}_2$ é criar novos vetores e matrizes que consideram apenas as observações para as quais $y_i \in (c_i, 1)$. Seja m o número de observações da amostra para as quais $y_i \in (c_i, 1)$. Definimos o vetor de observações $y_C = (y_{1C}, \ldots, y_{mC})^\top$ que considera do vetor y apenas as observações para as quais $y_i \in (c_i, 1)$ e as matrizes X_{C1} e X_{C2} construídas a partir de X_1 e X_2 excluindo-se as linhas correspondentes às observações para as quais $y_i \notin (c_i, 1)$. Assim, se por exemplo, $n = 5$, $c_i = 0{,}2$ para todo i e $y = (0{,}4, 0{,}2, 0, 0{,}7, 1)^\top$, temos que $m = 2$, $y_C = (0{,}4, 0{,}7)^\top$, X_{C1} é uma matriz de dimensão $(2 \times \mathrm{p}_\mu)$ com primeira linha igual à primeira linha de X_1 e segunda linha igual à quarta linha de X_1 e X_{C2} é uma matriz de dimensão $(2 \times \mathrm{p}_\phi)$ com primeira linha igual à primeira linha de X_2 e segunda linha igual à quarta linha de X_2.

Para o vetor de parâmetros β_1 sugerimos a utilização do mesmo método proposto por Ferrari e Cribari-Neto (2004). Trata-se de utilizar como valor inicial para $\hat{\beta}_1$ os estimadores de mínimos quadrados ordinários do modelo de regressão linear tendo como variável resposta $g_1(y_C)$ e como variáveis preditoras aquelas contidas em X_{C1}. No modelo de Ferrari e Cribari-Neto (2004), ϕ é constante. Assim, para o vetor de parâmetros β_2 sugerimos uma extensão do que eles propuseram como valor inicial para ϕ para o caso em que o parâmetro de precisão varia em função das variáveis preditoras. Seja $\check{\mu}_i = g_1^{-1}(\check{y}_{iC})$ em que $\check{y}_{iC}$ é o valor ajustado para a observação i na regressão linear para a variável transformada $g_1(y_C)$ e $\check{r}_i$ o resíduo ordinário para a observação i neste mesmo modelo. Da equação (2.9), tem-se que $\mathrm{Var}(y_{iC}) = [(\mu_i - c_i)(1-\mu_i)]/(\phi_i + 1)$ o que implica que

$$\phi_i = \frac{(\mu_i - c_i)(1-\mu_i)}{\mathrm{Var}(y_{iC})} - 1. \tag{3.15}$$

Fazendo-se a expansão de Taylor de $g_1(y_{iC})$ em torno de μ_i, tem-se que

$$\mathrm{Var}[g_1(y_{iC})] \approx \mathrm{Var}[g_1(\mu_i) + (y_{iC} - \mu_i)g_1'(\mu_i)] = \mathrm{Var}(y_{iC})[g_1'(\mu_i)]^2 \tag{3.16}$$

e portanto

$$\mathrm{Var}(y_{iC}) \approx \mathrm{Var}[g_1(y_{iC})][g_1'(\mu_i)]^{-2}. \tag{3.17}$$

Em um modelo de regressão linear normal homocedástico, um estimador não viciado da variância da variável resposta é a razão entre a soma de quadrados de resíduos e os graus de liberdade dos resíduos. Como no modelo RBIZUT o parâmetro ϕ_i da distribuição de y_{iC} varia em função das variáveis preditoras, um estimador razoável da $\mathrm{Var}(y_{iC})$ a partir de (3.17) é dado por $\check{\sigma}_i^2 = m\check{r}_i^2/\{(m-p_\mu)[g_1'(\check{\mu}_i)]^2\}$ e assim de (3.15) temos que um estimador inicial de ϕ_i é dado por

$$\check{\phi}_i = \frac{(\check{\mu}_i - c_i)(1-\check{\mu}_i)}{\check{\sigma}_i^2} - 1. \tag{3.18}$$

Assim, valores iniciais para $\hat{\beta}_2$ podem ser obtidos a partir dos estimadores de mínimos quadrados ordinários de um modelo de regressão linear tendo como variável resposta $(g_2(\check{\phi}_1), \ldots, g_2(\check{\phi}_m))^\top$ e as variáveis que compõem X_{C2} como preditoras. Se $\check{\phi}_i$ da equação (3.18) for negativo para algum i (ou mesmo para mais de um), basta substituir o valor negativo por algum valor positivo próximo de zero antes de ajustar o modelo de regressão linear.

Na situação em que as observações são independentes e identicamente distribuídas, vimos no Capítulo 2 que $\hat{\delta}_0 = \sum_{i=1}^n I_{\{0\}}(y_i)/n$, $\hat{\delta}_1 = \sum_{i=1}^n I_{\{1\}}(y_i)/n$ e $\hat{\delta}_c = \sum_{i=1}^n I_{\{c_i\}}(y_i)/n$ são ENVVUM de δ_0, δ_1 e δ_c Assim, pode-se utilizar como valores iniciais de $\hat{\gamma}_0$, $\hat{\gamma}_1$ e $\hat{\gamma}_c$, respectivamente, $\check{\gamma}_{10} = h_0(\sum_{i=1}^n I_{\{0\}}(y_i)/n, \sum_{i=1}^n I_{\{1\}}(y_i)/n, \sum_{i=1}^n I_{\{c_i\}}(y_i)/n), \check{\gamma}_{20} = 0, \ldots, \check{\gamma}_{p_0 0} = 0, \check{\gamma}_{11} = h_1(\sum_{i=1}^n I_{\{0\}}(y_i)/n, \sum_{i=1}^n I_{\{1\}}(y_i)/n, \sum_{i=1}^n I_{\{c_i\}}(y_i)/n), \check{\gamma}_{21} = 0, \ldots, \check{\gamma}_{p_1 0} = 0, \check{\gamma}_{1c} = h_c(\sum_{i=1}^n I_{\{0\}}(y_i)/n, \sum_{i=1}^n I_{\{1\}}(y_i)/n, \sum_{i=1}^n I_{\{c_i\}}(y_i)/n), \check{\gamma}_{2c} = 0, \ldots, \check{\gamma}_{p_c 0} = 0$.

Podem ainda ser obtidos valores iniciais de $\hat{\gamma}_0$, $\hat{\gamma}_1$ e $\hat{\gamma}_c$ mais próximos dos valores finais se a função de ligação H puder ser escrita como em (3.3). Para isso, criamos y_{D_0} a partir de y da seguinte forma:

para cada observação de y, se $y_i = 0$ inclui-se o valor 1 como uma observação de y_{D_0}, se $y_i \in (c_i, 1)$ inclui-se o valor 0 como uma observação de y_{D_0} e em caso contrário não se inclui nenhuma observação adicional em y_{D_0}. Cria-se ainda Z_{D_0} em que são mantidas de Z_0 apenas as linhas correspondentes às observações para as quais $y_i = 0$ ou $y_i \in (c_i, 1)$ e define-se m_0 como o número de linhas de Z_{D_0}. Assim, se por exemplo, $n = 5$, $c_i = 0{,}2$ para todo i e $y = (0{,}4, 0{,}2, 0, 0{,}7, 1)^\top$, temos que $m_0 = 3$, $y_{D_0} = (0, 1, 0)^\top$ e Z_{D_0} é uma matriz de dimensão $(3 \times p_0)$ com primeira linha igual à primeira linha de Z_0, segunda linha igual à terceira linha de Z_0 e terceira linha igual à quarta linha de Z_0. Sendo $\check{y}_{iD_0}$ o valor ajustado para a observação i na regressão linear com variável resposta y_{D_0} e variáveis preditoras Z_{D_0}, então $\check{\delta}_{i0} = \check{y}_{iD_0}$ é um estimador de $\delta_{i0}/(\delta_{i0} + (1 - \alpha_i))$. Em consequência, $(1 - \check{\delta}_{i0})$ é um estimador de $(1 - (\delta_{i0}/(\delta_{i0} + (1 - \alpha_i)))) = (1 - \alpha_i)/(\delta_{i0} + (1 - \alpha_i))$ e $\check{\delta}_{i0}/(1 - \check{\delta}_{i0})$ é um estimador de $\delta_{i0}/(1 - \alpha_i)$. Assim, o estimador de mínimos quadrados ordinários da regressão linear que tem como variável resposta $(h(\check{\delta}_{10}/(1 - \check{\delta}_{10})), \ldots, h(\check{\delta}_{m_0 0}/(1 - \check{\delta}_{m_0 0})))^\top$ e variáveis preditoras contidas em Z_{D_0} é um possível valor inicial para $\hat{\gamma}_0$. Se $\check{y}_{iD_0}$ não estiver contido no intervalo $(0, 1)$ para algum i (ou mesmo para mais de um), basta substituir o valor obtido por algum valor real próximo do valor obtido, mas que esteja no intervalo $(0, 1)$, antes de ajustar o segundo modelo de regressão linear. De forma análoga, obtêm-se y_{D_1}, Z_{D_1}, y_{D_c} e Z_{D_c} e valores iniciais para $\hat{\gamma}_1$ e $\hat{\gamma}_c$. Nos estudo de simulação da Seção 3.4 foi utilizado o método descrito acima para a obtenção de valores iniciais para $\hat{\gamma}_0$, $\hat{\gamma}_1$ e $\hat{\gamma}_c$ e os valores obtidos foram, em geral, muito próximos dos valores encontrados para $\hat{\gamma}_0$, $\hat{\gamma}_1$ e $\hat{\gamma}_c$ após a convergência do método BFGS.

A matriz de informação de Fisher do modelo RBIZUT é obtida a partir da função escore (3.6) (ver Apêndice B.3) e é dada por

$$K(\theta) = \begin{pmatrix} K_D(\gamma_0, \gamma_1, \gamma_c) & 0 \\ 0 & K_C(\gamma_0, \gamma_1, \gamma_c, \beta_1, \beta_2) \end{pmatrix}, \tag{3.19}$$

em que

$$K_D(\gamma_0, \gamma_1, \gamma_c) = \begin{pmatrix} K_{\gamma_0\gamma_0} & K_{\gamma_0\gamma_1} & K_{\gamma_0\gamma_c} \\ K_{\gamma_1\gamma_0} & K_{\gamma_1\gamma_1} & K_{\gamma_1\gamma_c} \\ K_{\gamma_c\gamma_0} & K_{\gamma_c\gamma_1} & K_{\gamma_c\gamma_c} \end{pmatrix}, \; K_C(\gamma_0, \gamma_1, \gamma_c, \beta_1, \beta_2) = \begin{pmatrix} K_{\beta_1\beta_1} & K_{\beta_1\beta_2} \\ K_{\beta_2\beta_1} & K_{\beta_2\beta_2} \end{pmatrix}. \tag{3.20}$$

Pode-se notar por (3.19) e (3.20) que os vetores de parâmetros γ_0, γ_1 e γ_c são ortogonais a β_1 e β_2 e, portanto, os correspondentes estimadores de máxima verossimilhança destes vetores de parâmetros são assintoticamente independentes. Definindo as matrizes diagonais, $W = \text{diag}\{w_1, \ldots, w_n\}$, $A =$

$\mathrm{diag}\{a_1,\ldots,a_n\}$ e $B=\mathrm{diag}\{b_1,\ldots,b_n\}$, em que

$$\begin{aligned}
w_i &= \left(\frac{\phi_i}{1-c_i}\right)^2\left\{\psi'\left[\left(\frac{1-\mu_i}{1-c_i}\right)\phi_i\right]+\psi'\left[\left(\frac{\mu_i-c_i}{1-c_i}\right)\phi_i\right]\right\},\\
a_i &= \left(\frac{\phi_i}{1-c_i}\right)\left\{\left(\frac{\mu_i-c_i}{1-c_i}\right)\psi'\left[\left(\frac{\mu_i-c_i}{1-c_i}\right)\phi_i\right]-\left(\frac{1-\mu_i}{1-c_i}\right)\psi'\left[\left(\frac{1-\mu_i}{1-c_i}\right)\phi_i\right]\right\},\\
b_i &= \left(\frac{\mu_i-c_i}{1-c_i}\right)^2\psi'\left[\left(\frac{\mu_i-c_i}{1-c_i}\right)\phi_i\right]+\left(\frac{1-\mu_i}{1-c_i}\right)^2\psi'\left[\left(\frac{1-\mu_i}{1-c_i}\right)\phi_i\right]-\psi'(\phi_i),
\end{aligned} \tag{3.21}$$

então as matrizes apresentadas em (3.20) que compõem a matriz de informação de Fisher são dadas por

$$\begin{aligned}
K_{\gamma_0\gamma_0} &= Z_0^\top[\Delta_0T_0^2+\Delta_1T_{10}^2+\Delta_cT_{c0}^2+\Delta_{(c,1)}(T_0+T_{10}+T_{c0})^2]Z_0,\\
K_{\gamma_1\gamma_1} &= Z_1^\top[\Delta_0T_{01}^2+\Delta_1T_1^2+\Delta_cT_{c1}^2+\Delta_{(c,1)}(T_{01}+T_1+T_{c1})^2]Z_1,\\
K_{\gamma_c\gamma_c} &= Z_c^\top[\Delta_0T_{0c}^2+\Delta_1T_{1c}^2+\Delta_cT_c^2+\Delta_{(c,1)}(T_{0c}+T_{1c}+T_c)^2]Z_c,\\
K_{\gamma_0\gamma_1} &= K_{\gamma_1\gamma_0}^\top = Z_0^\top[\Delta_0T_0T_{01}+\Delta_1T_{10}T_1+\Delta_cT_{c0}T_{c1}\\
&\quad+\Delta_{(c,1)}(T_0+T_{10}+T_{c0})(T_{01}+T_1+T_{c1})]Z_1,\\
K_{\gamma_0\gamma_c} &= K_{\gamma_c\gamma_0}^\top = Z_0^\top[\Delta_0T_0T_{0c}+\Delta_1T_{10}T_{1c}+\Delta_cT_{c0}T_c\\
&\quad+\Delta_{(c,1)}(T_0+T_{10}+T_{c0})(T_{0c}+T_{1c}+T_c)]Z_c,\\
K_{\gamma_1\gamma_c} &= K_{\gamma_c\gamma_1}^\top = Z_1^\top[\Delta_0T_{01}T_{0c}+\Delta_1T_1T_{1c}+\Delta_cT_{c1}T_c\\
&\quad+\Delta_{(c,1)}(T_{01}+T_1+T_{c1})(T_{0c}+T_{1c}+T_c)]Z_c,\\
K_{\beta_1\beta_1} &= X_1^\top[\Delta_{(c,1)}^{-1}WT_\mu^2]X_1,\\
K_{\beta_2\beta_2} &= X_2^\top[\Delta_{(c,1)}^{-1}BT_\phi^2]X_2,\\
K_{\beta_1\beta_2} &= K_{\beta_2\beta_1}^\top = X_1^\top[\Delta_{(c,1)}^{-1}AT_\mu T_\phi]X_2.
\end{aligned} \tag{3.22}$$

Os componentes (3.20) da matriz de informação de Fisher podem ser escritos de forma mais simples utilizando-se matrizes aumentadas. Sejam

$$\tilde{Q}=\begin{pmatrix} Q_0 & Q_{01} & Q_{0c}\\ Q_{01}^\top & Q_1 & Q_{1c}\\ Q_{0c}^\top & Q_{1c}^\top & Q_c\end{pmatrix},\ \tilde{C}=\begin{pmatrix}\Delta_{(c,1)}^{-1}WT_\mu^2 & \Delta_{(c,1)}^{-1}AT_\mu T_\phi\\ \Delta_{(c,1)}^{-1}AT_\mu T_\phi & \Delta_{(c,1)}^{-1}BT_\phi^2\end{pmatrix},\ \tilde{X}=\begin{pmatrix}X_1 & 0\\ 0 & X_2\end{pmatrix}, \tag{3.23}$$

matrizes de dimensão $(3n \times 3n)$, $(2n \times 2n)$ e $(2n \times (p_\mu + p_\phi))$, em que

$$\begin{aligned}
Q_0 &= \Delta_0 T_0^2 + \Delta_1 T_{10}^2 + \Delta_c T_{c0}^2 + \Delta_{(c,1)}(T_0 + T_{10} + T_{c0})^2,\\
Q_1 &= \Delta_0 T_{01}^2 + \Delta_1 T_1^2 + \Delta_c T_{c1}^2 + \Delta_{(c,1)}(T_{01} + T_1 + T_{c1})^2,\\
Q_c &= \Delta_0 T_{0c}^2 + \Delta_1 T_{1c}^2 + \Delta_c T_c^2 + \Delta_{(c,1)}(T_{0c} + T_{1c} + T_c)^2,\\
Q_{01} &= Q_{01}^\top = \Delta_0 T_0 T_{01} + \Delta_1 T_{10} T_1 + \Delta_c T_{c0} T_{c1} + \Delta_{(c,1)}(T_0 + T_{10} + T_{c0})(T_{01} + T_1 + T_{c1}),\\
Q_{0c} &= Q_{0c}^\top = \Delta_0 T_0 T_{0c} + \Delta_1 T_{10} T_{1c} + \Delta_c T_{c0} T_c + \Delta_{(c,1)}(T_0 + T_{10} + T_{c0})(T_{0c} + T_{1c} + T_c),\\
Q_{1c} &= Q_{1c}^\top = \Delta_0 T_{01} T_{0c} + \Delta_1 T_1 T_{1c} + \Delta_c T_{c1} T_c + \Delta_{(c,1)}(T_{01} + T_1 + T_{c1})(T_{0c} + T_{1c} + T_c),
\end{aligned} \tag{3.24}$$

e seja Z como definido em (3.12). Então

$$\begin{aligned}
K_D(\gamma_0, \gamma_1, \gamma_c) &= \tilde{Z}^\top \tilde{Q} \tilde{Z},\\
K_C(\gamma_0, \gamma_1, \gamma_c, \beta_1, \beta_2) &= \tilde{X}^\top \tilde{C} \tilde{X}.
\end{aligned} \tag{3.25}$$

Caso a função de ligação H seja a mesma utilizada no modelo RLBIZUT, no Apêndice B.3 prova-se que os componentes de $K_D(\gamma_0, \gamma_1, \gamma_c)$ apresentados em (3.22) simplificam-se para

$$\begin{aligned}
K_{\gamma_0\gamma_0} &= Z_0^\top \operatorname{diag}\{\delta_{10}(1-\delta_{10}), \ldots, \delta_{n0}(1-\delta_{n0})\} Z_0,\\
K_{\gamma_1\gamma_1} &= Z_1^\top \operatorname{diag}\{\delta_{11}(1-\delta_{11}), \ldots, \delta_{n1}(1-\delta_{n1})\} Z_1,\\
K_{\gamma_c\gamma_c} &= Z_c^\top \operatorname{diag}\{\delta_{1c}(1-\delta_{1c}), \ldots, \delta_{nc}(1-\delta_{nc})\} Z_c,\\
K_{\gamma_0\gamma_1} &= K_{\gamma_1\gamma_0}^\top = Z_0^\top \operatorname{diag}\{-\delta_{10}\delta_{11}, \ldots, -\delta_{n0}\delta_{n1}\} Z_1,\\
K_{\gamma_0\gamma_c} &= K_{\gamma_c\gamma_0}^\top = Z_0^\top \operatorname{diag}\{-\delta_{10}\delta_{1c}, \ldots, -\delta_{n0}\delta_{nc}\} Z_c,\\
K_{\gamma_1\gamma_c} &= K_{\gamma_c\gamma_1}^\top = Z_1^\top \operatorname{diag}\{-\delta_{11}\delta_{1c}, \ldots, -\delta_{n1}\delta_{nc}\} Z_c.
\end{aligned} \tag{3.26}$$

3.2 Intervalos de confiança e testes de hipóteses

Intervalos de confiança e testes de hipóteses sobre os parâmetros do modelo RBIZUT podem ser obtidos utilizando-se as propriedades assintóticas do estimador de máxima verossimilhança. Temos que, sob condições de regularidade,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N_p(0, J(\theta)^{-1}) \tag{3.27}$$

em que $J(\theta) = \lim_{n\to\infty} K(\theta)/n$, $K(\theta)$ conforme definido em (3.19) e $p = p_o + p_1 + p_c + p_\mu + p_\phi$. Dessa forma, para amostras grandes, intervalos de confiança aproximados para cada um dos parâmetros com coeficiente de confiança de $(100 \times \gamma)\%$ são dados por

$$\mathrm{IC}(\theta_k; \gamma) = \hat{\theta}_k \pm z_{(1+\gamma)/2} (\hat{k}^{\theta_k\theta_k})^{1/2}, \tag{3.28}$$

em que $z_{(1+\gamma)/2}$ é o quantil $(1+\gamma)/2$ da distribuição $N(0,1)$ e $\hat{k}^{\theta_k\theta_k}$ é o termo de posição k da diagonal de $K(\theta)^{-1}$ avaliada em $\hat{\theta}$. Observe que, ao contrário do intervalo de confiança para a distribuição

BIZUT apresentado em (2.19), a expressão do intervalo (3.28) não contém explicitamente o tamanho da amostra n. Isso ocorre porque para a distribuição BIZUT obtivemos a informação de Fisher para uma única observação, enquanto que para o modelo RBIZUT a obtivemos para uma amostra de tamanho n. Assim, no intervalo de confiança (3.28) o tamanho da amostra é considerado através de $\hat{k}^{\theta_k\theta_k}$.

Uma região de confiança assintótica para todos os parâmetros do modelo RBIZUT com coeficiente de confiança de $(100 \times \gamma)\%$ pode ser escrita (Johnson e Wichern, 2007, página 221) como

$$(\hat{\theta} - \theta)^\top \widehat{K(\theta)}(\hat{\theta} - \theta) \leq [p(n-1)/(n-p)]F_{p,n-p;\gamma}, \tag{3.29}$$

em que $F_{p,n-p;\gamma}$ é o quantil γ da distribuição F de Fisher-Snedecor com p e $n-p$ graus de liberdade e $\widehat{K(\theta)}$ é $K(\theta)$ avaliada em $\hat{\theta}$. A distribuição assintótica do estimador de máxima verossimilhança de qualquer função diferenciável de θ pode ser obtida utilizando-se o método delta.

Para testar hipóteses a respeito dos parâmetros do modelo RBIZUT é necessário definir algumas partições de vetores e matrizes. Suponha que temos interesse em testar hipóteses relacionadas a um subconjunto dos vetores de parâmetros γ_0, γ_1, γ_c, β_1 e β_2. Assim, esses vetores podem ser escritos como $\gamma_0 = (\gamma_{A0}^\top, \gamma_{B0}^\top)^\top$, $\gamma_1 = (\gamma_{A1}^\top, \gamma_{B1}^\top)^\top$, $\gamma_c = (\gamma_{Ac}^\top, \gamma_{Bc}^\top)^\top$, $\beta_1 = (\beta_{A1}^\top, \beta_{B1}^\top)^\top$ e $\beta_2 = (\beta_{A2}^\top, \beta_{B2}^\top)^\top$, em que γ_{A0}, γ_{A1}, γ_{Ac}, β_{A1}, β_{A2} são vetores que contêm os parâmetros de interesse de dimensões, respectivamente, q_0, q_1, q_c, q_μ e q_ϕ e $0 \leq q_0 \leq p_0$, $0 \leq q_1 \leq p_1$, $0 \leq q_c \leq p_c$, $0 \leq q_\mu \leq p_\mu$ e $0 \leq q_\phi \leq p_\phi$. Dessa forma, as hipóteses do teste são:

$$\begin{aligned} &\mathrm{H}_0\colon \gamma_{A0} = \gamma_{A0}^{(0)}, \gamma_{A1} = \gamma_{A1}^{(0)}, \gamma_{Ac} = \gamma_{Ac}^{(0)}, \beta_{A1} = \beta_{A1}^{(0)}, \beta_{A2} = \beta_{A2}^{(0)}; \\ &\mathrm{H}_1\colon \text{violação de pelo menos uma das desigualdades de } \mathrm{H}_0, \end{aligned} \tag{3.30}$$

em que $\gamma_{A0}^{(0)}$, $\gamma_{A1}^{(0)}$, $\gamma_{Ac}^{(0)}$, $\beta_{A1}^{(0)}$ e $\beta_{A2}^{(0)}$ são vetores contendo valores específicos para os parâmetros de interesse.

A hipótese nula gera as seguintes partições do vetor escore, das matrizes de variáveis preditoras e da matriz de informação de Fisher:

$$U(\theta) = (U_{\gamma_{A0}}^\top, U_{\gamma_{B0}}^\top, U_{\gamma_{A1}}^\top, U_{\gamma_{B1}}^\top, U_{\gamma_{Ac}}^\top, U_{\gamma_{Bc}}^\top, U_{\beta_{A1}}^\top, U_{\beta_{B1}}^\top, U_{\beta_{A2}}^\top, U_{\beta_{B2}}^\top)^\top,$$

$$Z_0 = (Z_{A0}, Z_{B0}), Z_1 = (Z_{A1}, Z_{B1}), Z_c = (Z_{Ac}, Z_{Bc}), X_1 = (X_{A1}, X_{B1}), X_2 = (X_{A2}, X_{B2}) \text{ e}$$

$$K(\theta) = \begin{pmatrix} K_{\gamma_{A0}\gamma_{A0}} & K_{\gamma_{A0}\gamma_{B0}} & K_{\gamma_{A0}\gamma_{A1}} & K_{\gamma_{A0}\gamma_{B1}} & K_{\gamma_{A0}\gamma_{Ac}} & K_{\gamma_{A0}\gamma_{Bc}} & 0 & 0 & 0 & 0 \\ K_{\gamma_{B0}\gamma_{A0}} & K_{\gamma_{B0}\gamma_{B0}} & K_{\gamma_{B0}\gamma_{A1}} & K_{\gamma_{B0}\gamma_{B1}} & K_{\gamma_{B0}\gamma_{Ac}} & K_{\gamma_{B0}\gamma_{Bc}} & 0 & 0 & 0 & 0 \\ K_{\gamma_{A1}\gamma_{A0}} & K_{\gamma_{A1}\gamma_{B0}} & K_{\gamma_{A1}\gamma_{A1}} & K_{\gamma_{A1}\gamma_{B1}} & K_{\gamma_{A1}\gamma_{Ac}} & K_{\gamma_{A1}\gamma_{Bc}} & 0 & 0 & 0 & 0 \\ K_{\gamma_{B1}\gamma_{A0}} & K_{\gamma_{B1}\gamma_{B0}} & K_{\gamma_{B1}\gamma_{A1}} & K_{\gamma_{B1}\gamma_{B1}} & K_{\gamma_{B1}\gamma_{Ac}} & K_{\gamma_{B1}\gamma_{Bc}} & 0 & 0 & 0 & 0 \\ K_{\gamma_{Ac}\gamma_{A0}} & K_{\gamma_{Ac}\gamma_{B0}} & K_{\gamma_{Ac}\gamma_{A1}} & K_{\gamma_{Ac}\gamma_{B1}} & K_{\gamma_{Ac}\gamma_{Ac}} & K_{\gamma_{Ac}\gamma_{Bc}} & 0 & 0 & 0 & 0 \\ K_{\gamma_{Bc}\gamma_{A0}} & K_{\gamma_{Bc}\gamma_{B0}} & K_{\gamma_{Bc}\gamma_{A1}} & K_{\gamma_{Bc}\gamma_{B1}} & K_{\gamma_{Bc}\gamma_{Ac}} & K_{\gamma_{Bc}\gamma_{Bc}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & K_{\beta_{A1}\beta_{A1}} & K_{\beta_{A1}\beta_{B1}} & K_{\beta_{A1}\beta_{A2}} & K_{\beta_{A1}\beta_{B2}} \\ 0 & 0 & 0 & 0 & 0 & 0 & K_{\beta_{B1}\beta_{A1}} & K_{\beta_{B1}\beta_{B1}} & K_{\beta_{B1}\beta_{A2}} & K_{\beta_{B1}\beta_{B2}} \\ 0 & 0 & 0 & 0 & 0 & 0 & K_{\beta_{A2}\beta_{A1}} & K_{\beta_{A2}\beta_{B1}} & K_{\beta_{A2}\beta_{A2}} & K_{\beta_{A2}\beta_{B2}} \\ 0 & 0 & 0 & 0 & 0 & 0 & K_{\beta_{B2}\beta_{A1}} & K_{\beta_{B2}\beta_{B1}} & K_{\beta_{B2}\beta_{A2}} & K_{\beta_{B2}\beta_{B2}} \end{pmatrix}.$$

Sejam ainda $U_A(\theta) = (U_{\gamma_{A0}}^\top, U_{\gamma_{A1}}^\top, U_{\gamma_{Ac}}^\top, U_{\beta_{A1}}^\top, U_{\beta_{A2}}^\top)^\top$,

$$[K(\theta)]^{-1} = \begin{pmatrix} K_{AA}^{\gamma_0\gamma_0} & K_{AB}^{\gamma_0\gamma_0} & K_{AA}^{\gamma_0\gamma_1} & K_{AB}^{\gamma_0\gamma_1} & K_{AA}^{\gamma_0\gamma_c} & K_{AB}^{\gamma_0\gamma_c} & 0 & 0 & 0 & 0 \\ K_{BA}^{\gamma_0\gamma_0} & K_{BB}^{\gamma_0\gamma_0} & K_{BA}^{\gamma_0\gamma_1} & K_{BB}^{\gamma_0\gamma_1} & K_{BA}^{\gamma_0\gamma_c} & K_{BB}^{\gamma_0\gamma_c} & 0 & 0 & 0 & 0 \\ K_{AA}^{\gamma_1\gamma_0} & K_{AB}^{\gamma_1\gamma_0} & K_{AA}^{\gamma_1\gamma_1} & K_{AB}^{\gamma_1\gamma_1} & K_{AA}^{\gamma_1\gamma_c} & K_{AB}^{\gamma_1\gamma_c} & 0 & 0 & 0 & 0 \\ K_{BA}^{\gamma_1\gamma_0} & K_{BB}^{\gamma_1\gamma_0} & K_{BA}^{\gamma_1\gamma_1} & K_{BB}^{\gamma_1\gamma_1} & K_{BA}^{\gamma_1\gamma_c} & K_{BB}^{\gamma_1\gamma_c} & 0 & 0 & 0 & 0 \\ K_{AA}^{\gamma_c\gamma_0} & K_{AB}^{\gamma_c\gamma_0} & K_{AA}^{\gamma_c\gamma_1} & K_{AB}^{\gamma_c\gamma_1} & K_{AA}^{\gamma_c\gamma_c} & K_{AB}^{\gamma_c\gamma_c} & 0 & 0 & 0 & 0 \\ K_{BA}^{\gamma_c\gamma_0} & K_{BB}^{\gamma_c\gamma_0} & K_{BA}^{\gamma_c\gamma_1} & K_{BB}^{\gamma_c\gamma_1} & K_{BA}^{\gamma_c\gamma_c} & K_{BB}^{\gamma_c\gamma_c} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & K_{AA}^{\beta_1\beta_1} & K_{AB}^{\beta_1\beta_1} & K_{AA}^{\beta_1\beta_2} & K_{AB}^{\beta_1\beta_2} \\ 0 & 0 & 0 & 0 & 0 & 0 & K_{BA}^{\beta_1\beta_1} & K_{BB}^{\beta_1\beta_1} & K_{BA}^{\beta_1\beta_2} & K_{BB}^{\beta_1\beta_2} \\ 0 & 0 & 0 & 0 & 0 & 0 & K_{AA}^{\beta_2\beta_1} & K_{AB}^{\beta_2\beta_1} & K_{AA}^{\beta_2\beta_2} & K_{AB}^{\beta_2\beta_2} \\ 0 & 0 & 0 & 0 & 0 & 0 & K_{BA}^{\beta_2\beta_1} & K_{BB}^{\beta_2\beta_1} & K_{BA}^{\beta_2\beta_2} & K_{BB}^{\beta_2\beta_2} \end{pmatrix},$$

$$K^{\theta_A}(\theta) = \begin{pmatrix} K_{AA}^{\gamma_0\gamma_0} & K_{AA}^{\gamma_0\gamma_1} & K_{AA}^{\gamma_0\gamma_c} & 0 & 0 \\ K_{AA}^{\gamma_1\gamma_0} & K_{AA}^{\gamma_1\gamma_1} & K_{AA}^{\gamma_1\gamma_c} & 0 & 0 \\ K_{AA}^{\gamma_c\gamma_0} & K_{AA}^{\gamma_c\gamma_1} & K_{AA}^{\gamma_c\gamma_c} & 0 & 0 \\ 0 & 0 & 0 & K_{AA}^{\beta_1\beta_1} & K_{AA}^{\beta_1\beta_2} \\ 0 & 0 & 0 & K_{AA}^{\beta_2\beta_1} & K_{AA}^{\beta_2\beta_2} \end{pmatrix} \text{ e}$$

$$[K^{\theta_A}(\theta)]^{-1} = \begin{pmatrix} K^{\gamma_{A0}\gamma_{A0}} & K^{\gamma_{A0}\gamma_{A1}} & K^{\gamma_{A0}\gamma_{Ac}} & 0 & 0 \\ K^{\gamma_{A1}\gamma_{A0}} & K^{\gamma_{A1}\gamma_{A1}} & K^{\gamma_{A1}\gamma_{Ac}} & 0 & 0 \\ K^{\gamma_{Ac}\gamma_{A0}} & K^{\gamma_{Ac}\gamma_{A1}} & K^{\gamma_{Ac}\gamma_{Ac}} & 0 & 0 \\ 0 & 0 & 0 & K^{\beta_{A1}\beta_{A1}} & K^{\beta_{A1}\beta_{A2}} \\ 0 & 0 & 0 & K^{\beta_{A2}\beta_{A1}} & K^{\beta_{A2}\beta_{A2}} \end{pmatrix}.$$

O vetor $U_A(\theta)$ pode ser obtido substituindo em $U(\theta)$, cujos componentes foram apresentados em (3.11), Z_0, Z_1, Z_c, X_1 e X_2 por Z_{A0}, Z_{A1}, Z_{Ac}, X_{A1} e X_{A2}, respectivamente. A prova desse resultado pode ser vista no Apêndice B.4.

A partir dos vetores e matrizes definidos, podem-se obter as seguintes estatísticas para o teste de hipóteses (3.30): da razão de verossimilhanças (Q_{RV}), escore (Q_E), de Wald (Q_W) e gradiente

(Q_G, Terrell (2002)). Sob H_0, todas elas tem distribuição assintótica qui-quadrado com q graus de liberdade, em que $q = q_o + q_1 + q_c + q_\mu + q_\phi$.

A estatística da razão de verossimilhanças é dada por

$$Q_{RV} = 2[l(\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_c, \hat{\beta}_1, \hat{\beta}_2) - l(\tilde{\gamma}_0, \tilde{\gamma}_1, \tilde{\gamma}_c, \tilde{\beta}_1, \tilde{\beta}_2)], \tag{3.31}$$

em que $l(\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_c, \hat{\beta}_1, \hat{\beta}_2)$ é o logaritmo da função de verossimilhança (3.4) aplicado no estimador de máxima verossimilhança de θ e $l(\tilde{\gamma}_0, \tilde{\gamma}_1, \tilde{\gamma}_c, \tilde{\beta}_1, \tilde{\beta}_2)$ é o logaritmo da função de verossimilhança (3.4) aplicado no estimador de máxima verossimilhança restrito de θ (sob H_0).

Nas estatísticas seguintes, o símbolo "$\hat{}$" indica que os correspondentes vetores e matrizes devem ser aplicados no estimador de máxima verossimilhança e "$\tilde{}$" indica que os mesmos devem ser aplicados no estimador de máxima verossimilhança sob H_0. A estatística do teste escore é dada por

$$\begin{aligned}
Q_E &= [\tilde{U}_A(\theta)]^\top \tilde{K}^{\theta_A}(\theta)\tilde{U}_A(\theta) \\
&= \tilde{U}_{\gamma_{A0}}^\top \tilde{K}_{AA}^{\gamma_0\gamma_0}\tilde{U}_{\gamma_{A0}} + \tilde{U}_{\gamma_{A1}}^\top \tilde{K}_{AA}^{\gamma_1\gamma_0}\tilde{U}_{\gamma_{A0}} + \tilde{U}_{\gamma_{Ac}}^\top \tilde{K}_{AA}^{\gamma_c\gamma_0}\tilde{U}_{\gamma_{A0}} + \tilde{U}_{\gamma_{A0}}^\top \tilde{K}_{AA}^{\gamma_0\gamma_1}\tilde{U}_{\gamma_{A1}} \\
&\quad + \tilde{U}_{\gamma_{A1}}^\top \tilde{K}_{AA}^{\gamma_1\gamma_1}\tilde{U}_{\gamma_{A1}} + \tilde{U}_{\gamma_{Ac}}^\top \tilde{K}_{AA}^{\gamma_c\gamma_1}\tilde{U}_{\gamma_{A1}} + \tilde{U}_{\gamma_{A0}}^\top \tilde{K}_{AA}^{\gamma_0\gamma_c}\tilde{U}_{\gamma_{Ac}} + \tilde{U}_{\gamma_{A1}}^\top \tilde{K}_{AA}^{\gamma_1\gamma_c}\tilde{U}_{\gamma_{Ac}} \\
&\quad + \tilde{U}_{\gamma_{Ac}}^\top \tilde{K}_{AA}^{\gamma_c\gamma_c}\tilde{U}_{\gamma_{Ac}} + \tilde{U}_{\beta_{A1}}^\top \tilde{K}_{AA}^{\beta_1\beta_1}\tilde{U}_{\beta_{A1}} + \tilde{U}_{\beta_{A2}}^\top \tilde{K}_{AA}^{\beta_2\beta_1}\tilde{U}_{\beta_{A1}} + \tilde{U}_{\beta_{A1}}^\top \tilde{K}_{AA}^{\beta_1\beta_2}\tilde{U}_{\beta_{A2}} \\
&\quad + \tilde{U}_{\beta_{A2}}^\top \tilde{K}_{AA}^{\beta_2\beta_2}\tilde{U}_{\beta_{A2}},
\end{aligned} \tag{3.32}$$

enquanto a estatística de Wald pode ser escrita como

$$\begin{aligned}
Q_W &= [(\hat{\gamma}_{A0} - \gamma_{A0}^{(0)})^\top, (\hat{\gamma}_{A1} - \gamma_{A1}^{(0)})^\top, (\hat{\gamma}_{Ac} - \gamma_{Ac}^{(0)})^\top, (\hat{\beta}_{A1} - \beta_{A1}^{(0)})^\top, (\hat{\beta}_{A2} - \beta_{A2}^{(0)})^\top][\hat{K}^{\theta_A}(\theta)]^{-1} \\
&\quad [(\hat{\gamma}_{A0} - \gamma_{A0}^{(0)})^\top, (\hat{\gamma}_{A1} - \gamma_{A1}^{(0)})^\top, (\hat{\gamma}_{Ac} - \gamma_{Ac}^{(0)})^\top, (\hat{\beta}_{A1} - \beta_{A1}^{(0)})^\top, (\hat{\beta}_{A2} - \beta_{A2}^{(0)})^\top]^\top \\
&= (\hat{\gamma}_{A0} - \gamma_{A0}^{(0)})^\top \hat{K}^{\gamma_{A0}\gamma_{A0}}(\hat{\gamma}_{A0} - \gamma_{A0}^{(0)}) + (\hat{\gamma}_{A1} - \gamma_{A1}^{(0)})^\top \hat{K}^{\gamma_{A1}\gamma_{A0}}(\hat{\gamma}_{A0} - \gamma_{A0}^{(0)}) \\
&\quad + (\hat{\gamma}_{Ac} - \gamma_{Ac}^{(0)})^\top \hat{K}^{\gamma_{Ac}\gamma_{A0}}(\hat{\gamma}_{A0} - \gamma_{A0}^{(0)}) + (\hat{\gamma}_{A0} - \gamma_{A0}^{(0)})^\top \hat{K}^{\gamma_{A0}\gamma_{A1}}(\hat{\gamma}_{A1} - \gamma_{A1}^{(0)}) \\
&\quad + (\hat{\gamma}_{A1} - \gamma_{A1}^{(0)})^\top \hat{K}^{\gamma_{A1}\gamma_{A1}}(\hat{\gamma}_{A1} - \gamma_{A1}^{(0)}) + (\hat{\gamma}_{Ac} - \gamma_{Ac}^{(0)})^\top \hat{K}^{\gamma_{Ac}\gamma_{A1}}(\hat{\gamma}_{A1} - \gamma_{A1}^{(0)}) \\
&\quad + (\hat{\gamma}_{A0} - \gamma_{A0}^{(0)})^\top \hat{K}^{\gamma_{A0}\gamma_{Ac}}(\hat{\gamma}_{Ac} - \gamma_{Ac}^{(0)}) + (\hat{\gamma}_{A1} - \gamma_{A1}^{(0)})^\top \hat{K}^{\gamma_{A1}\gamma_{Ac}}(\hat{\gamma}_{Ac} - \gamma_{Ac}^{(0)}) \\
&\quad + (\hat{\gamma}_{Ac} - \gamma_{Ac}^{(0)})^\top \hat{K}^{\gamma_{Ac}\gamma_{Ac}}(\hat{\gamma}_{Ac} - \gamma_{Ac}^{(0)}) + (\hat{\beta}_{A1} - \beta_{A1}^{(0)})^\top \hat{K}^{\beta_{A1}\beta_{A1}}(\hat{\beta}_{A1} - \beta_{A1}^{(0)}) \\
&\quad + (\hat{\beta}_{A2} - \beta_{A2}^{(0)})^\top \hat{K}^{\beta_{A2}\beta_{A1}}(\hat{\beta}_{A1} - \beta_{A1}^{(0)}) + (\hat{\beta}_{A1} - \beta_{A1}^{(0)})^\top \hat{K}^{\beta_{A1}\beta_{A2}}(\hat{\beta}_{A2} - \beta_{A2}^{(0)}) \\
&\quad + (\hat{\beta}_{A2} - \beta_{A2}^{(0)})^\top \hat{K}^{\beta_{A2}\beta_{A2}}(\hat{\beta}_{A2} - \beta_{A2}^{(0)})
\end{aligned} \tag{3.33}$$

e a estatística gradiente é dada por

$$\begin{aligned}
Q_G &= [\hat{U}_A(\theta)]^\top [(\hat{\gamma}_{A0} - \gamma_{A0}^{(0)})^\top, (\hat{\gamma}_{A1} - \gamma_{A1}^{(0)})^\top, (\hat{\gamma}_{Ac} - \gamma_{Ac}^{(0)})^\top, (\hat{\beta}_{A1} - \beta_{A1}^{(0)})^\top, (\hat{\beta}_{A2} - \beta_{A2}^{(0)})^\top]^\top \\
&= \hat{U}_{\gamma_{A0}}^\top (\hat{\gamma}_{A0} - \gamma_{A0}^{(0)}) + \hat{U}_{\gamma_{A1}}^\top (\hat{\gamma}_{A1} - \gamma_{A1}^{(0)}) + \hat{U}_{\gamma_{Ac}}^\top (\hat{\gamma}_{Ac} - \gamma_{Ac}^{(0)}) + \hat{U}_{\beta_{A1}}^\top (\hat{\beta}_{A1} - \beta_{A1}^{(0)}) \\
&\quad + \hat{U}_{\beta_{A2}}^\top (\hat{\beta}_{A2} - \beta_{A2}^{(0)}).
\end{aligned} \tag{3.34}$$

3.3 Modelo RBIZUT com c variável

Um dos aspectos interessantes do modelo RBIZUT é permitir que o parâmetro conhecido c da distribuição BIZUT da variável resposta varie entre as unidades populacionais. Na prática, isso significa que o modelo permite que o suporte da distribuição da variável resposta varie entre as unidades populacionais, já que y_i pode assumir valores em $\{0\} \cup [c_i, 1]$. Conforme mencionado na Seção 3.1, o fato de c variar entre as unidades populacionais não torna o modelo consideravelmente mais complexo, em relação à situação em que c é fixo. No entanto, alguns aspectos práticos adicionais devem ser discutidos quando c é variável. Na Seção 3.3.1 apresentamos alguns resultados para o modelo RBIZUT com c variável e na Seção 3.3.2 discutimos que incluir c entre as variáveis preditoras do modelo pode trazer endogeneidade ao modelo (Stock e Watson, 2003, Capítulo 10).

3.3.1 Resultados para o modelo RBIZUT com c variável

Em modelos RBIZUT com c variável, à medida que c cresce, a amplitude do intervalo $(c, 1)$ diminui. Portanto, é razoável imaginar que, mesmo que duas unidades populacionais possuam os mesmos valores para todas as variáveis preditoras, elas tenham valores diferentes de μ caso tenham valores de c diferentes. Assim, pode ser interessante incluir c_i ou até mesmo uma função de c_i entre as variáveis preditoras para μ_i. Como o espaço paramétrico do parâmetro μ_i da variável resposta é igual a $(c_i, 1)$, a função de ligação mais simples para μ_i é a utilizada no modelo RLBIZUT, ou seja, $g_1(\mu_i) = \log[(\mu_i - c_i)/(1 - \mu_i)]$. Para essa função de ligação, apresentaremos e provaremos três proposições relacionadas ao modelo RBIZUT com c variável. Antes de apresentarmos as proposições, é necessário definir uma determinada função de μ_i. Sendo y_i uma variável aleatória com distribuição BIZUT$(\delta_{i0}, \delta_{i1}, \delta_{ic}, \mu_i, \phi_i, c_i)$, definimos $\mu_i^R = (\mu_i - c_i)/(1 - c_i)$. Observe que μ_i^R é a posição relativa de μ_i no intervalo $(c_i, 1)$. Assim, se $\mu_i^R = 0,4$, $\mu_i - c_i$ é igual a 40% da amplitude do intervalo $(c_i, 1)$. Apresentamos a seguir a primeira proposição, que é provada no Apêndice B.5.

Proposição 3.3.1. *Sejam $y_1, y_2, \ldots, y_n$, variáveis aleatórias independentes, em que y_i tem distribuição BIZUT($\delta_{i0}, \delta_{i1}, \delta_{ic}, \mu_i, \phi_i, c_i$), c_i conhecido. Admita que essas variáveis são ajustadas por um modelo RBIZUT com função de ligação para μ_i dada por $g_1(\mu_i) = \log[(\mu_i - c_i)/(1 - \mu_i)]$ e função de ligação para ϕ_i dada por $g_2(\phi_i) = \log(\phi_i)$. Se nem c_i e nem nenhuma função de c_i estão entre as variáveis preditoras para μ e ϕ, então a variância assintótica dos estimadores de máxima verossimilhança dos vetores de parâmetros β_1 e β_2 não é função de c_i.*

Assim, para as funções de ligação mais simples para μ e ϕ, ao contrário do esperado, a variância assintótica dos estimadores de máxima verossimilhança dos vetores de parâmetros β_1 e β_2 não cresce à medida que aumenta a variabilidade de c_i entre as unidades populacionais e não é menor quando c é fixo do que quando ele é variável.

As demais proposições estão relacionadas à inclusão de uma função de c_i entre as variáveis preditoras para μ_i e também são provadas no Apêndice B.5.

Proposição 3.3.2. *Sejam $y_1, y_2, \ldots, y_n$, variáveis aleatórias independentes, em que y_i tem distribuição BIZUT($\delta_{i0}, \delta_{i1}, \delta_{ic}, \mu_i, \phi_i, c_i$), c_i conhecido. Admita que essas variáveis são ajustadas por*

um modelo RBIZUT com função de ligação para μ_i *dada por* $g_1(\mu_i) = \log[(\mu_i - c_i)/(1 - \mu_i)]$. *Seja* $f_c(c_i)$ *uma função crescente, derivável e com domínio que contém o intervalo* $(0, 1)$. *Se incluirmos* $f_c(c_i)$ *entre as variáveis preditoras para o parâmetro* μ_i, *então a equação do modelo RBIZUT para o parâmetro* μ_i *pode ser escrita como* $\log[(\mu_i - c_i)/(1 - \mu_i)] = \eta_{i1} + \tau f_c(c_i)$, *em que* $\tau \in \mathbb{R}$ *é um parâmetro. Nesse caso, variando-se* c_i *(e consequentemente* $f_c(c_i)$*) e mantendo-se as demais variáveis preditoras constantes, tem-se que* μ_i^R, *em relação à* c_i, *é:*
I crescente se $\tau > 0$;
II constante se $\tau = 0$;
III decrescente se $\tau < 0$.

Note que, pela Proposição 3.3.2, se $\tau = 0$ ou se, equivalentemente, $f_c(c_i)$ não for incluída entre as variáveis preditoras para μ_i, então, variando-se c_i e mantendo-se as demais variáveis preditoras constantes, μ_i varia mas a posição relativa de μ_i (μ_i^R) mantém-se constante. Na prática, se $f_c(c_i)$ for significante e/ou se houver interesse em acrescentá-la no modelo, o termo $\tau f_c(c_i)$ é incluído em η_{i1}. Dessa forma, o modelo RBIZUT continua podendo ser representado como apresentado em (3.2).

A terceira proposição está relacionada à variação de μ_i quando se varia c_i. Na Proposição 3.3.3 e daqui em diante, exp(.), exp[.] e exp{.} são representações para a função exponencial.

Proposição 3.3.3. *Sejam* $g_c(c_i) = \mu_i = \{c_i + \exp[\eta_{i1} + \tau f_c(c_i)]\}/\{1 + \exp[\eta_{i1} + \tau f_c(c_i)]\}$ *e* $g_j(j) = g_c(c_i + j) - g_c(c_i)$. *Nas condições da Proposição 3.3.2 e fixado* $0 < j < 1 - c_i$, *se* $f_c(c_i)$ *é positiva no intervalo* $(0, 1)$, *então* $\exists\ \eta_{i1}$ *tal que* $\forall\, \tau$, $g_j(j) > 0$.

O que a Proposição 3.3.3 garante é que se $f_c(c_i)$ for incluída entre as variáveis preditoras para μ_i e $f_c(c_i)$ for positiva no intervalo $(0, 1)$, então aumentando-se c_i e mantendo-se as demais variáveis preditoras constantes, μ_i necessariamente cresce para determinados conjuntos de valores das demais variáveis preditoras. Assim, a Proposição 3.3.3 apresenta uma restrição do modelo RBIZUT (com função de ligação para μ_i do modelo RLBIZUT) quando é incluída uma função $f_c(c_i)$ positiva no intervalo $(0, 1)$. Na prática, porém, a Proposição 3.3.3 não deve trazer problema ao ajuste de um modelo RBIZUT. Em geral, mantendo-se as demais variáveis preditoras constantes, parece ser razoável que μ_i cresça quando c_i cresce. No exemplo do cartão de crédito mencionado no Capítulo 1, por exemplo, essa hipótese é bem razoável. Se dois clientes de um cartão de crédito têm valores de pagamento mínimo (em relação ao valor da fatura, c_i) diferentes, mas são idênticos em relação a todas as demais variáveis preditoras para μ_i, é natural imaginar que o cliente com maior c_i apresente um maior μ_i. É importante lembrar ainda que o modelo não restringe o decrescimento de μ_i^R quando c_i cresce (de acordo com a Proposição 3.3.2).

Caso, em alguma situação real, a restrição apresentada na Proposição 3.3.3 dificulte a obtenção de um bom ajuste para uma variável de interesse, ela pode ser contornada. Para isso, basta incluir entre as variáveis preditoras para μ_i uma função negativa no intervalo $(0, 1)$. Se $f_c(c_i) = \log(c_i)$, por exemplo, for incluída entre as variáveis preditoras para μ_i, o modelo RBIZUT não possui a restrição apresentada na Proposição 3.3.3. No entanto, em geral, não é conveniente utilizar uma função $f_c(c_i)$ negativa no intervalo $(0, 1)$. O motivo é que, nesse caso, μ_i torna-se decrescente em

relação a τ, ao contrário do que ocorre geralmente em modelos de regressão, dificultando assim a interpretação dos resultados.

3.3.2 Endogeneidade

Conforme discutido na Seção 3.3.1, em modelos RBIZUT com c variável pode ser interessante incluir c como variável preditora do modelo. Embora tenhamos discutido na Seção 3.3.1 a inclusão de c como variável preditora para μ, pode ser importante a inclusão de c em todas as 5 equações do modelo RBIZUT. No exemplo do cartão de crédito mencionado no Capítulo 1, suponha que dois clientes são semelhantes em relação a todas as variáveis preditoras disponíveis, mas para um deles o valor de c é pequeno e para o outro o valor de c é grande. O primeiro tende a ter maior δ_c e menor δ_0 que o segundo, já que c grande é naturalmente um desincentivo ao pagamento mínimo e tende a aumentar a probabilidade do cliente não pagar nada, já que ele precisa ter mais recursos para efetuar pelo menos o pagamento mínimo.

Embora possa ser importante a inclusão de c como variável preditora em um modelo RBIZUT, é conveniente avaliar se sua inclusão pode trazer problemas de endogeneidade. Uma variável é endógena quando é correlacionada com o termo de erro do modelo (Stock e Watson, 2003, Seção 10.1). Este é definido como a diferença entre o valor da variável resposta e o valor que seria ajustado pelo modelo se todos os parâmetros fossem conhecidos. Entre os motivos que podem tornar uma variável endógena estão, por exemplo, o fato dela ser medida com erro e causalidade simultânea. Esta última ocorre quando não somente a distribuição da variável resposta varia em função da variável preditora, mas também o valor da variável resposta pode afetar o valor da variável preditora. No exemplo do cartão de crédito, dependendo do procedimento adotado pela instituição financeira para definir o valor de c, a mesma pode ser ou não uma variável endógena devido a causalidade simultânea. Suponha, por exemplo, que a instituição reduza o valor de c para clientes que nos últimos 12 meses sempre pagaram o valor total da fatura. Nesse caso c será uma variável endógena, já que o mesmo tende a ser pequeno para observações com valor da variável resposta igual a 1.

Para uma variável endógena, os estimadores de máxima verossimilhança dos parâmetros associados a essa variável não são consistentes. Dessa forma, se c for endógena e for incluída em todas as equações do modelo RBIZUT, utilizaremos 5 estimadores não consistentes no ajuste do modelo. Assim, nesse caso, para o exemplo do cartão de crédito, não conseguiremos obter boas estimativas de, por exemplo, como varia a probabilidade do cliente não pagar a fatura, quando aumentamos o valor de c em 0,1 e mantemos o valor das demais variáveis preditoras constantes.

Uma forma de obtenção de estimadores consistentes na presença de variáveis endógenas é trabalhar com variáveis instrumentais (Stock e Watson, 2003, Capítulo 10), que são variáveis que são correlacionadas com a variável endógena, mas que não são correlacionadas com o termo de erro do modelo. Assim, um possível tema de pesquisa futura pode ser o desenvolvimento de estimadores consistentes dos parâmetros do modelo RBIZUT com variáveis preditoras endógenas utilizando variáveis instrumentais.

Como neste trabalho não desenvolvemos estimadores consistentes nessas situações, caso c seja uma variável endógena, duas estratégias podem ser adotadas. A primeira é não incluir c entre as

variáveis preditoras do modelo. Nesse caso, se a associação entre c e a variável resposta for forte, o ajuste do modelo RBIZUT será satisfatório apenas se c for definida a partir de outras variáveis já presentes no modelo. Outra estratégia é incluir c mesmo sabendo que as estimativas dos parâmetros associados à variável serão obtidas a partir de estimadores que não são consistentes. Nesse caso, as estimativas obtidas dos parâmetros associados a c não terão interpretação prática.

Dentre as variáveis citadas ao longo deste trabalho que possivelmente têm um bom ajuste usando o modelo RBIZUT, apenas o VRP do cartão de crédito apresenta c variável. Além disso, é possível que muitas instituições não usem critérios para definir o valor de c que tornem a variável endógena. Mesmo nos casos em que c é endógena, é possível que a não inclusão de c entre as variáveis preditoras do modelo não afete consideravelmente a qualidade do ajuste. O motivo disso é que geralmente é bem grande o percentual de clientes que apresentam valores pequenos de c e próximos entre si. Assim, como é razoável imaginar que uma pequena mudança no valor de c pouco afete os parâmetros da distribuição da variável resposta, c tende a não ser uma variável importante do modelo. Dessa forma, se c for uma variável endógena e se um percentual grande da população de interesse apresentar valores próximos de c, recomendamos a não inclusão de c como variável preditora, já que isso não deve afetar consideravelmente a qualidade do ajuste do modelo. Por outro lado, caso isso não ocorra e haja indícios de que c seja uma variável importante do modelo, sugerimos o desenvolvimento teórico de estimadores consistentes para os parâmetros do modelo RBIZUT na presença de variáveis endógenas.

3.4 Estudos de simulação

Nesta seção são apresentados estudos de simulação de Monte Carlo para avaliar a performance dos estimadores de máxima verossimilhança dos parâmetros do modelo RBIZUT em diferentes situações. Foram conduzidos cinco estudos de simulação de Monte Carlo. Todos eles foram desenvolvidos utilizando-se a linguagem de programação Ox. Para cada caso, foram realizadas 5000 replicações e obtidas estimativas do viés e da raiz do erro quadrático médio (REQM) dos estimadores. Nos cinco estudos, foram considerados casos particulares do modelo RLBIZUT. No entanto, é razoável esperar que se tivessem sido escolhidas outras funções de ligação, a maior parte das conclusões seriam semelhantes. As tabelas contendo os resultados dos estudos de simulação desta seção estão no Apêndice **??**.

O modelo RBIZUT contém cinco vetores de parâmetros e, assim, possui no mínimo 5 parâmetros. Na prática, o número de parâmetros do modelo RBIZUT é bem maior. Na maior parte das vezes, o ajuste preliminar de um modelo RBIZUT conterá pelo menos um intercepto e uma variável preditora para modelar cada um dos 5 parâmetros da variável resposta, resultando em um mínimo de 10 parâmetros. Além disso, conforme mencionado na Seção 2.6, a distribuição BIZUT foi proposta para ajustar algumas variáveis na área de econometria. Nesta área, geralmente não há dificuldade em se obter amostras com milhares de observações. Para todas as variáveis citadas nos Capítulos 1 e 2, que possivelmente têm um bom ajuste utilizando um modelo RBIZUT, podem-se obter facilmente amostras grandes. Assim, em situações reais, tanto o número de parâmetros como o tamanho da amostra geralmente são grandes. Por esse motivo, os estudos de simulação desenvolvidos nesta seção

não consideram amostras muito pequenas. Nesta seção, quando for conveniente e a omissão não gerar nenhuma dúvida, nos referimos ao estimador de máxima verossimilhança simplesmente como estimador.

Nos 4 primeiros estudos de simulação, o modelo RLBIZUT possui a seguinte estrutura:

$$\begin{cases} \log\left(\dfrac{\mu_i - c_i}{1-\mu_i}\right) & = \beta_{11} + \beta_{21}x_{i21} + \beta_{31}x_{i31} \\ \log(\phi_i) & = \beta_{12} + \beta_{22}x_{i22} + \beta_{32}x_{i32} \\ \left(h\left(\dfrac{\delta_{i0}}{1-\alpha_i}\right), h\left(\dfrac{\delta_{i1}}{1-\alpha_i}\right), h\left(\dfrac{\delta_{ic}}{1-\alpha_i}\right)\right) & = (\gamma_{10} + \gamma_{20}z_{i20}, \gamma_{11} + \gamma_{21}z_{i21}, \gamma_{1c} + \gamma_{2c}z_{i2c}), \end{cases} \tag{3.35}$$

em que h é a função logarítmica. O modelo (3.35) é um caso particular de (3.2) com $x_{i11} = x_{i12} = z_{i10} = z_{i11} = z_{i1c} = 1$ para todo i e, portanto, β_{11}, β_{12}, γ_{10}, γ_{11} e γ_{1c} são interceptos. No primeiro, terceiro e quarto estudos de simulação foram adotados os seguintes valores para os verdadeiros valores dos parâmetros: β_{11} = -1,0, β_{21} = 2,5, β_{31} = -1,7, β_{12} = 2,5, β_{22} = 0,9, β_{32} = -0,9, γ_{10} = -0,6, γ_{20} = -1,2, γ_{11} = -1,5, γ_{21} = 1,1, γ_{1c} = -0,6 e γ_{2c} = -1,1.

Em muitos trabalhos nos quais os modelos de regressão desenvolvidos ajustam mais de um parâmetro da variável resposta em função de outras variáveis, os estudos de simulação consideram variáveis preditoras diferentes para modelar cada um dos parâmetros da variável resposta. No entanto, na prática, muitas variáveis preditoras são utilizadas para modelar todos os parâmetros da variável resposta. Por esse motivo, nos estudos de simulação desenvolvidos, $x_{i21} = x_{i22} = z_{i20} = z_{i21} = z_{i2c}$. As variáveis preditoras foram geradas independentemente a partir de variáveis aleatórias com distribuição uniforme no intervalo $(0, 1)$ e mantidas contantes em todas as replicações. Nos estudos em que n é mantido constante, as variáveis preditoras também foram mantidas contantes mesmo quando variamos os valores dos parâmetros ou a amplitude do intervalo de variação de c. Como $x_{i21} = x_{i22} = z_{i20} = z_{i21} = z_{i2c}$, foi necessário gerar apenas x_{i21}, x_{i31} e x_{i32}.

No primeiro estudo de simulação, c foi mantido constante e igual a 0,2 e n variou entre 50 e 1000. Com os verdadeiros valores dos parâmetros mencionados anteriormente, tem-se que o intervalo de variação dos parâmetros associados à distribuição BIZUT são: $\delta_0 \in (0{,}082, 0{,}236)$, $\delta_1 \in (0{,}096, 0{,}332)$, $\delta_c \in (0{,}091, 0{,}236)$, $\mu \in (0{,}250, 0{,}854)$ e $\phi \in (4{,}953, 29{,}964)$. A Tabela 3.1 apresenta os resultados do primeiro estudo de simulação. Pode-se notar que os estimadores de máxima verossimilhança dos parâmetros do vetor β_1 parecem apresentar vieses pequenos mesmo para pequenas amostras. Porém, os estimadores dos parâmetros dos vetores β_2, γ_0, γ_1 e γ_c apresentam vieses moderados para pequenas amostras. É interessante ressaltar que, em relação ao valor verdadeiro, os vieses dos estimadores de máxima verossimilhança dos parâmetros do vetor β_2 no modelo RLBIZUT parecem ser menores do que o viés do estimador de máxima verossimilhança de ϕ na distribuição BIZUT (estudado na Seção 2.5). Pode-se perceber também que a REQM parece ser consideravelmente menor para os estimadores dos parâmetros do vetor β_1 do que para os estimadores de β_2, γ_0, γ_1 e γ_c. Pode-se notar ainda que, independente do tamanho da amostra, a REQM dos estimadores de máxima verossimilhança dos parâmetros dos vetores β_1 e β_2 no modelo RLBIZUT parecem ser, respectivamente, maiores do que a REQM dos estimadores de máxima verossimilhança de μ e ϕ na

distribuição BIZUT. Conforme esperado, em geral, o viés e a REQM dos estimadores de todos os parâmetros considerados diminui à medida que se aumenta o tamanho da amostra.

Tabela 3.1: Estimativas do viés e da raiz do erro quadrático médio do modelo RBIZUT com c fixo (0,2) obtidas a partir de estudo de simulação de Monte Carlo com 5000 réplicas ($\delta_0 \in (0{,}082, 0{,}236), \delta_1 \in (0{,}096, 0{,}332), \delta_c \in (0{,}091, 0{,}236), \mu \in (0{,}250, 0{,}854), \phi \in (4{,}953, 29{,}964)$).

n	γ_{10} = -0,6		γ_{20} = -1,2		γ_{11} = -1,5		γ_{21} = 1,1		γ_{1c} = -0,6		γ_{2c} = -1,1	
	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM
50	-0,071	0,984	-0,160	2,318	-0,198	1,220	0,197	1,814	-0,040	0,871	-0,178	1,941
70	-0,025	0,735	-0,111	1,524	-0,112	0,844	0,115	1,378	-0,009	0,707	-0,149	1,497
100	-0,013	0,611	-0,092	1,151	-0,079	0,706	0,078	1,028	-0,023	0,603	-0,076	1,106
200	-0,008	0,417	-0,035	0,798	-0,043	0,474	0,043	0,736	-0,012	0,418	-0,031	0,790
500	-0,005	0,245	-0,014	0,470	-0,008	0,272	0,005	0,423	0,000	0,242	-0,019	0,459
1000	0,000	0,182	-0,009	0,347	-0,006	0,200	0,008	0,313	0,000	0,178	-0,006	0,339
n	β_{11} = -1,0		β_{21} = 2,5		β_{31} = -1,7		β_{12} = 2,5		β_{22} = 0,9		β_{32} = -0,9	
	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM
50	-0,011	0,411	0,016	0,570	-0,011	0,491	0,336	1,219	0,084	1,574	-0,143	1,304
70	-0,001	0,318	0,012	0,430	-0,021	0,399	0,199	0,857	0,093	1,167	-0,075	1,115
100	-0,006	0,260	0,014	0,342	-0,012	0,313	0,173	0,675	-0,027	0,788	-0,044	0,766
200	-0,004	0,171	0,006	0,233	-0,001	0,209	0,064	0,401	0,014	0,534	-0,022	0,513
500	0,001	0,107	0,002	0,142	-0,004	0,135	0,024	0,246	0,010	0,315	-0,009	0,313
1000	-0,001	0,078	0,002	0,103	-0,001	0,093	0,012	0,170	0,003	0,224	-0,002	0,215

No segundo estudo de simulação c é constante e igual a 0,2, $n = 500$ e variou-se os valores dos parâmetros do modelo RLBIZUT. Os verdadeiros valores considerados para os parâmetros do modelo, bem como o intervalo de variação dos parâmetros da variável resposta em cada situação estão na Tabela 3.2. O primeiro vetor de valores verdadeiros para os parâmetros foi o utilizado no primeiro estudo de simulação. A partir dele foram obtidos 5 outros vetores de valores verdadeiros para θ. No segundo vetor, alteramos apenas os valores dos parâmetros de β_1 de forma que o intervalo de variação de μ reduzisse, mas continuasse contendo o valor 0,6 (ponto médio entre $c = 0{,}2$ e 1,0). No terceiro e quarto vetores consideramos, respectivamente, μ próximo de c e próximo de 1 e mantivemos os valores dos vetores de parâmetros β_2, γ_0, γ_1 e γ_c constantes. No vetor de parâmetros seguinte, aumentamos o intervalo de variação de ϕ de $(4{,}95, 29{,}96)$ para $(20{,}09, 121{,}51)$. No último vetor de parâmetros, reduzimos o intervalo de variação de δ_0, δ_1 e δ_c. Os resultados obtidos são coerentes com aqueles observados no segundo estudo de simulação da Seção 2.5 e são apresentados na Tabela 3.3. O valores de μ e de ϕ não parecem ter relação com a performance dos estimadores de γ_0, γ_1 e γ_c. Isso é razoável, já que $K_D(\gamma_0, \gamma_1, \gamma_c)$ apresentado em (3.20) não é função de β_1 e β_2. O valor de μ parece afetar o desempenho do estimador de β_2, mas a ordenação da performance não tem um padrão. Os estimadores dos parâmetros β_{12} e β_{22} têm REQM menor quando o intervalo de variação de μ inclui o ponto 0,6 do que quando ele está próximo de c ou de 1. No entanto, para o estimador de β_{32} a REQM é maior quando o intervalo de variação de μ inclui o ponto 0,6 (ponto médio entre c e 1) do que quando ele está próximo de c ou de 1. A REQM dos estimadores

de β_1 parece decrescer à medida que ϕ cresce, já que a REQM dos estimadores dos 3 parâmetros que compõem β_1 decresceram quando o intervalo de variação de ϕ passou de $(4{,}95, 29{,}96)$ para $(20{,}09, 121{,}51)$. Isso era esperado, já que quando ϕ cresce a variância da variável resposta diminui. Ao compararmos os resultados para os vetores de parâmetros 1 e 6, pode-se observar ainda que a redução do valor de δ_0, δ_1 e δ_c parece acarretar uma redução da REQM dos estimadores de β_1 e β_2. Isso é coerente com o resultado da função escore para o modelo RBIZUT cujos componentes são dados em (3.11). Observando-se $U_{\beta_1}(\beta_1, \beta_2)$ e $U_{\beta_2}(\beta_1, \beta_2)$, pode-se notar que apenas as observações tais que $y_i \in (c_i, 1)$ são utilizadas na estimação de β_1 e β_2. Assim, à medida que os parâmetros da variável resposta δ_0, δ_1 e δ_c crescem, o número de observações utilizadas para estimar μ e ϕ decresce e, em consequência, a performance dos estimadores se reduz.

Tabela 3.2: Intervalo de variação dos parâmetros da variável resposta e valores dos parâmetros do modelo de regressão utilizados nos estudos de simulação de Monte Carlo com c fixo.

Vetores de parâmetros	Variação δ_0	Vetor γ_0	Variação δ_1	Vetor γ_1	Variação δ_c	Vetor γ_c
1	(0,082, 0,236)	(-0,6, -1,2)	(0,096, 0,332)	(-1,5, 1,1)	(0,091, 0,236)	(-0,6, -1,1)
2	(0,082, 0,236)	(-0,6, -1,2)	(0,096, 0,332)	(-1,5, 1,1)	(0,091, 0,236)	(-0,6, -1,1)
3	(0,082, 0,236)	(-0,6, -1,2)	(0,096, 0,332)	(-1,5, 1,1)	(0,091, 0,236)	(-0,6, -1,1)
4	(0,082, 0,236)	(-0,6, -1,2)	(0,096, 0,332)	(-1,5, 1,1)	(0,091, 0,236)	(-0,6, -1,1)
5	(0,082, 0,236)	(-0,6, -1,2)	(0,096, 0,332)	(-1,5, 1,1)	(0,091, 0,236)	(-0,6, -1,1)
6	(0,044, 0,148)	(-1,5, -1,3)	(0,054, 0,195)	(-2,5, 1,2)	(0,044, 0,134)	(-1,6, -1,2)
Vetores de parâmetros	Variação μ	Vetor β_1	Variação ϕ	Vetor β_2		
1	(0,250, 0,854)	(-1,0, 2,5, -1,7)	(4,95, 29,96)	(2,5, 0,9, -0,9)		
2	(0,502, 0,698)	(-0,2, 0,7, -0,3)	(4,95, 29,96)	(2,5, 0,9, -0,9)		
3	(0,211, 0,273)	(-3,6, 1,3, -0,7)	(4,95, 29,96)	(2,5, 0,9, -0,9)		
4	(0,974, 0,995)	(4,1, 0,9, -0,7)	(4,95, 29,96)	(2,5, 0,9, -0,9)		
5	(0,250, 0,854)	(-1,0, 2,5, -1,7)	(20,09, 121,51)	(3,9, 0,9, -0,9)		
6	(0,250, 0,854)	(-1,0, 2,5, -1,7)	(4,95, 29,96)	(2,5, 0,9, -0,9)		

O terceiro estudo de simulação é semelhante ao primeiro e os resultados são apresentados na Tabela 3.4. A única alteração é que no terceiro estudo c varia no intervalo $(0{,}15, 0{,}25)$. O valor de c para cada observação foi gerado de forma independente a partir de uma variável aleatória com distribuição uniforme no intervalo $(0{,}15, 0{,}25)$ e mantido constante em todas as replicações. Dessa forma, em média, o valor de c no terceiro e no primeiro estudo são iguais. Comparando-se as Tabelas 3.1 e 3.4, pode-se notar que o viés e a REQM não parecem se alterar quando passamos de uma situação em que c é fixo para uma em que c é variável. Assim, a performance dos estimadores dos parâmetros do modelo RLBIZUT parece ser semelhante quando c é fixo e quando c é variável.

No quarto estudo de simulação, foram mantidos os valores dos parâmetros do primeiro estudo e, com $n = 500$, variou-se a amplitude do intervalo de variação (AIV) de c. Em todas as situações consideradas, gerou-se o valor de c a partir de variáveis aleatórias uniformes em um intervalo cujo

Tabela 3.3: Estimativas do viés e da raiz do erro quadrático médio do modelo RBIZUT com c fixo (0,2) obtidas a partir de estudo de simulação de Monte Carlo com 5000 réplicas ($n = 500$ e vários diferentes vetores de parâmetros)

Vetores de	γ_{10}		γ_{20}		γ_{11}		γ_{21}		γ_{1c}		γ_{2c}	
parâmetros	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM
1	-0,005	0,245	-0,014	0,470	-0,008	0,272	0,005	0,423	0,000	0,242	-0,019	0,459
2	0,002	0,247	-0,020	0,471	-0,012	0,276	0,009	0,429	-0,003	0,239	-0,010	0,459
3	0,003	0,246	-0,023	0,480	-0,009	0,277	0,004	0,426	0,003	0,243	-0,028	0,461
4	-0,007	0,248	-0,011	0,478	-0,026	0,280	0,031	0,434	-0,004	0,240	-0,008	0,455
5	0,002	0,246	-0,035	0,474	-0,004	0,275	-0,006	0,431	0,005	0,239	-0,028	0,455
6	0,001	0,291	-0,042	0,590	-0,033	0,344	0,028	0,520	-0,012	0,300	-0,019	0,594
Vetores de	β_{11}		β_{21}		β_{31}		β_{12}		β_{22}		β_{32}	
parâmetros	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM
1	0,001	0,107	0,002	0,142	-0,004	0,135	0,024	0,246	0,010	0,315	-0,009	0,313
2	0,002	0,097	-0,002	0,121	0,000	0,124	0,023	0,244	0,015	0,307	-0,011	0,317
3	-0,020	0,251	0,012	0,325	0,006	0,206	0,043	0,305	-0,017	0,406	-0,003	0,263
4	0,008	0,312	-0,007	0,490	0,034	0,220	0,034	0,345	-0,004	0,538	0,052	0,238
5	0,000	0,055	0,000	0,073	-0,001	0,070	0,026	0,249	0,014	0,322	-0,018	0,321
6	-0,002	0,087	0,004	0,113	-0,003	0,108	0,021	0,197	0,000	0,253	-0,009	0,249

Tabela 3.4: Estimativas do viés e da raiz do erro quadrático médio do modelo RBIZUT com c variável (0,15, 0,25) obtidas a partir de estudo de simulação de Monte Carlo com 5000 réplicas ($\delta_0 \in (0{,}082, 0{,}236), \delta_1 \in (0{,}096, 0{,}332), \delta_c \in (0{,}091, 0{,}236), \mu \in (0{,}204, 0{,}863), \phi \in (4{,}953, 29{,}964)$).

n	γ_{10} = -0,6		γ_{20} = -1,2		γ_{11} = -1,5		γ_{21} = 1,1		γ_{1c} = -0,6		γ_{2c} = -1,1	
	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM
50	-0,051	1,067	-0,166	2,088	-0,192	1,220	0,206	1,713	-0,075	1,122	-0,103	1,937
70	-0,028	0,693	-0,125	1,546	-0,125	0,837	0,130	1,420	-0,032	0,702	-0,097	1,504
100	-0,022	0,609	-0,070	1,167	-0,093	0,715	0,096	1,084	-0,024	0,618	-0,049	1,157
200	-0,003	0,385	-0,045	0,747	-0,051	0,463	0,058	0,703	-0,017	0,389	-0,009	0,722
500	-0,004	0,250	-0,009	0,483	-0,015	0,281	0,014	0,442	-0,005	0,245	-0,008	0,476
1000	-0,005	0,167	-0,002	0,326	-0,012	0,190	0,015	0,300	-0,005	0,166	-0,001	0,322
n	β_{11} = -1,0		β_{21} = 2,5		β_{31} = -1,7		β_{12} = 2,5		β_{22} = 0,9		β_{32} = -0,9	
	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM
50	-0,006	0,400	0,017	0,530	-0,018	0,450	0,257	1,209	0,109	1,448	-0,029	1,575
70	-0,002	0,284	0,021	0,478	-0,025	0,394	0,151	0,851	0,118	1,192	0,003	1,082
100	-0,002	0,250	0,008	0,343	-0,007	0,282	0,128	0,603	0,072	0,834	-0,055	0,820
200	-0,005	0,184	0,012	0,230	-0,006	0,203	0,070	0,402	0,015	0,499	-0,028	0,528
500	-0,001	0,112	0,004	0,148	-0,003	0,129	0,023	0,249	0,009	0,319	-0,008	0,310
1000	-0,001	0,077	0,003	0,098	-0,002	0,093	0,009	0,167	0,006	0,214	-0,004	0,219

ponto médio fosse igual a 0,2, mantendo-se assim a média constante. Para uma amplitude do intervalo de variação de c de 0,1, por exemplo, c varia de forma uniforme entre 0,15 e 0,25. Quando a amplitude do intervalo de variação de c é igual a 0, c é fixo. Os resultados são apresentados na Tabela 3.5. As conclusões são semelhantes às obtidas no terceiro estudo. Pode-se notar que o viés e a REQM de todos os parâmetros parecem se manter constantes independente da amplitude do intervalo de variação de c. Assim, mesmo que a amplitude do intervalo de variação de c seja grande, a performance dos estimadores do modelo RLBIZUT parece ser semelhante quando c é fixo e quando c é variável.

Tabela 3.5: Estimativas do viés e da raiz do erro quadrático médio do modelo RBIZUT com c variável (com média 0,2) obtidas a partir de estudo de simulação de Monte Carlo com 5000 réplicas ($n = 500, \delta_0 \in (0{,}082, 0{,}236), \delta_1 \in (0{,}096, 0{,}332), \delta_c \in (0{,}091, 0{,}236), \mu \in (0{,}063, 0{,}891), \phi \in (4{,}953, 29{,}964)$).

AIV	γ_{10} = -0,6		γ_{20} = -1,2		γ_{11} = -1,5		γ_{21} = 1,1		γ_{1c} = -0,6		γ_{2c} = -1,1	
de c	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM
0,00	0,001	0,250	-0,020	0,483	-0,005	0,280	0,004	0,445	0,001	0,248	-0,013	0,483
0,01	-0,001	0,248	-0,021	0,485	-0,019	0,279	0,020	0,441	-0,006	0,245	-0,005	0,466
0,05	-0,005	0,248	-0,006	0,483	-0,010	0,274	0,006	0,437	-0,006	0,249	-0,007	0,476
0,10	-0,004	0,250	-0,009	0,483	-0,015	0,281	0,014	0,442	-0,005	0,245	-0,008	0,476
0,20	-0,003	0,252	-0,012	0,483	-0,013	0,279	0,014	0,438	0,001	0,246	-0,017	0,474
0,40	-0,004	0,247	-0,011	0,483	-0,020	0,281	0,025	0,444	-0,008	0,250	-0,005	0,479
AIV	β_{11} = -1,0		β_{21} = 2,5		β_{31} = -1,7		β_{12} = 2,5		β_{22} = 0,9		β_{32} = -0,9	
de c	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM
0,00	0,002	0,111	-0,001	0,148	-0,001	0,130	0,033	0,248	-0,001	0,319	-0,017	0,307
0,01	0,003	0,111	0,000	0,146	-0,005	0,130	0,026	0,245	0,010	0,318	-0,016	0,303
0,05	0,001	0,114	0,002	0,148	-0,005	0,132	0,020	0,242	0,015	0,317	-0,003	0,300
0,10	-0,001	0,112	0,004	0,148	-0,003	0,129	0,023	0,249	0,009	0,319	-0,008	0,310
0,20	-0,001	0,108	0,001	0,143	-0,002	0,131	0,028	0,246	0,004	0,321	-0,008	0,307
0,40	0,002	0,109	-0,001	0,146	-0,004	0,133	0,023	0,246	0,010	0,319	-0,006	0,309

No quinto estudo de simulação foi feita uma alteração na estrutura do modelo (3.35). Incluiu-se $f_c(c_i) = c_i$ entre as variáveis preditoras para μ_i, excluiu-se uma variável preditora para ϕ_i alterando-se os valores dos parâmetros remanescentes e manteve-se inalterada a estrutura relacionada aos parâmetros δ_{i0}, δ_{i1} e δ_{ic} da variável resposta. Assim, a estrutura do modelo RLBIZUT no quinto estudo de simulação é dada por

$$\begin{cases} \log\left(\dfrac{\mu_i - c_i}{1-\mu_i}\right) & = \beta_{11} + \beta_{21}x_{i21} + \beta_{31}x_{i31} + \beta_{41}x_{i41} \\ \log(\phi_i) & = \beta_{12} + \beta_{22}x_{i22} \\ \left(h\left(\dfrac{\delta_{i0}}{1-\alpha_i}\right), h\left(\dfrac{\delta_{i1}}{1-\alpha_i}\right), h\left(\dfrac{\delta_{ic}}{1-\alpha_i}\right)\right) & = (\gamma_{10} + \gamma_{20}z_{i20}, \gamma_{11} + \gamma_{21}z_{i21}, \gamma_{1c} + \gamma_{2c}z_{i2c}), \end{cases} \tag{3.36}$$

em que h é a função logarítmica, $x_{i41} = c_i$, β_{41} = -5,0, β_{12} = 1,6, β_{22} = 1,8 e os valores dos demais parâmetros são iguais aos do primeiro estudo de simulação. Com a alteração dos valores de β_{12} e

β_{22}, o intervalo de variação de ϕ manteve-se igual ao do primeiro estudo de simulação. Neste último estudo de simulação, fixou-se $n = 500$ e variou-se a amplitude do intervalo de variação de c entre 0,01 e 0,4 com os mesmos critérios utilizados no quarto estudo. Os resultados são apresentados na Tabela 3.6.

Tabela 3.6: Estimativas do viés e da raiz do erro quadrático médio do modelo RBIZUT com c variável (com média 0,2) obtidas a partir de estudo de simulação de Monte Carlo com 5000 réplicas ($n = 500, \delta_0 \in (0{,}082, 0{,}236), \delta_1 \in (0{,}096, 0{,}332), \delta_c \in (0{,}091, 0{,}236), \mu \in (0{,}063, 0{,}891), \phi \in (4{,}953, 29{,}964)$).

AIV	γ_{10} = -0,6		γ_{20} = -1,2		γ_{11} = -1,5		γ_{21} = 1,1		γ_{1c} = -0,6		γ_{2c} = -1,1	
de c	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM
0,01	-0,004	0,247	-0,009	0,492	-0,009	0,277	0,005	0,443	0,001	0,240	-0,024	0,480
0,05	-0,003	0,247	-0,009	0,492	-0,012	0,272	0,010	0,439	0,000	0,242	-0,022	0,469
0,10	-0,005	0,240	-0,006	0,477	-0,019	0,278	0,023	0,444	-0,004	0,240	-0,009	0,471
0,20	-0,004	0,241	-0,014	0,471	-0,016	0,282	0,015	0,445	-0,004	0,242	-0,011	0,475
0,40	-0,002	0,241	-0,011	0,474	-0,007	0,274	-0,003	0,438	0,001	0,240	-0,020	0,475
AIV	β_{11} = -1,0		β_{21} = 2,5		β_{31} = -1,7		β_{41} = -5,0		β_{12} = 1,6		β_{22} = 1,8	
de c	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM	Viés	REQM
0,01	-0,052	2,766	0,007	0,182	-0,004	0,141	0,236	13,849	0,017	0,195	0,019	0,329
0,05	0,001	0,561	0,003	0,181	-0,001	0,144	-0,011	2,713	0,013	0,197	0,022	0,334
0,10	0,000	0,293	0,006	0,184	-0,005	0,141	-0,010	1,324	0,012	0,196	0,029	0,330
0,20	0,003	0,199	0,005	0,179	-0,004	0,140	-0,032	0,711	0,012	0,192	0,033	0,330
0,40	-0,006	0,156	0,008	0,178	-0,001	0,142	-0,008	0,381	0,024	0,194	0,010	0,332

Pode-se notar que, mesmo com a inclusão de $f_c(c_i) = c_i$ como variável preditora para μ_i, a performance dos estimadores dos parâmetros γ_0, γ_1, γ_c e β_2 não parece se alterar em função da amplitude do intervalo de variação de c. A performance dos estimadores dos parâmetros de β_1 associados às duas variáveis preditoras que não são $f_c(c_i) = c_i$ também não parecem ser função da amplitude do intervalo de variação de c. No entanto, as performances do estimador do intercepto associado a μ_i e do estimador do parâmetro associado à $f_c(c_i) = c_i$ parecem variar bastante em função da amplitude do intervalo de variação de c. Em ambos os casos, a REQM parece decrescer à medida que a amplitude do intervalo de variação de c cresce. Em princípio isso pode parecer estranho, já que quanto maior a amplitude do intervalo de variação de c, menor é a semelhança do modelo RLBIZUT com os modelos de regressão usuais. Assim, seria razoável esperar que a REQM aumentasse à medida que a amplitude do intervalo de variação de c crescesse. No entanto, vimos nos estudos de simulação 3 e 4 que os estimadores dos parâmetros do modelo RLBIZUT têm performance semelhante nas situações em que c é fixo e naquelas em que c é variável. Desta forma, o decrescimento da REQM dos estimadores do intercepto associado a μ_i e do parâmetro associado à $f_c(c_i) = c_i$ parece ser consequência do aumento da dispersão da variável preditora c_i à medida que a amplitude do intervalo de variação de c cresce. Na regressão linear simples normal com variável resposta y_i com variância σ^2, variável preditora x_i e tamanho de amostra n, por exemplo, mantidos constantes n, σ^2 e $\bar{x} = \sum_{i=1}^{n} x_i/n$, a variância do estimador de máxima verossimilhança tanto

do intercepto quando do parâmetro associado a x_i decresce à medida que $\sum_{i=1}^{n}(x_i - \bar{x})^2$ cresce. Pelos resultados observados na Tabela 3.6, assim como na regressão linear simples, o aumento da dispersão de uma variável preditora para μ_i parece melhorar a performance tanto do estimador do parâmetro associado a essa variável quanto do estimador do intercepto associado a μ_i. Pode-se notar ainda que, para a amplitude do intervalo de variação de c igual a 0,01, os vieses dos mesmos estimadores são bem maiores do que nas demais situações. Possivelmente, isso se deve ao fato de c_i neste caso ficar próximo de ser uma constante, o que faz com que os vetores $(x_{111}, x_{211}, \ldots, x_{n11})^\top$ e $(x_{141}, x_{241}, \ldots, x_{n41})^\top$ fiquem próximos de ser linearmente dependentes.

3.5 Seleção de modelos

Conforme mencionado na Seção 3.1, o modelo RBIZUT pertence à classe dos modelos aditivos generalizados para locação, escala e forma (GAMLSS) introduzidos por Rigby e Stasinopoulos (2005). Essa classe de modelos é bastante ampla contendo inclusive modelos de regressão para variáveis resposta que têm distribuições que não pertencem à família exponencial. Além disso, a classe admite modelos em que todos os parâmetros da distribuição da variável resposta são ajustados em função de variáveis preditoras e os modelos dessa classe podem ter termos fixos, aleatórios e de alisamento. O modelo RBIZUT pertence à subclasse dos modelos paramétricos GAMLSS que contém apenas termos fixos. Para a comparação de modelos encaixados nessa subclasse, Rigby e Stasinopoulos (2005) sugerem a utilização do teste da razão de verossimilhanças discutido na Seção 3.2. Podemos fazer ainda essa comparação utilizando o teste escore, de Wald e gradiente também discutidos na Seção 3.2. Para a comparação de modelos não encaixados, Rigby e Stasinopoulos (2005) sugerem o uso do critério de informação de Akaike generalizado, GAIC (Akaike, 1983), que para o modelo RBIZUT, utilizando a notação introduzida na Seção 3.1, é dado por

$$\text{GAIC} = -2l(\hat{\theta}) + h_n(n)(p_0 + p_1 + p_c + p_\mu + p_\phi), \tag{3.37}$$

em que $h_n(n)$ é uma função não decrescente e positiva para $n > 1$ e o melhor modelo é aquele com menor GAIC. Algumas sugestões de $h_n(n)$ encontradas na literatura são $h_n(n) = 2$ (AIC, Akaike (1974)) e $h_n(n) = \log(n)$ (BIC, Schwarz (1978)).

Além do GAIC, existem diversas outras sugestões de critérios para a seleção de modelos. Para o modelo de regressão beta com dispersão fixa, Giampaoli et al. (2009) compararam, a partir de estudos de simulação de Monte Carlo, o AIC e o BIC com 4 outros critérios. Giampaoli et al. (2009) concluíram que o BIC é o melhor critério no cenário em que os valores dos parâmetros são grandes e que não é possível escolher o melhor critério no cenário em que os valores dos parâmetros são pequenos. Para o modelo de regressão beta com dispersão variável, Bayer (2011) comparou a partir de estudos de simulação de Monte Carlo, alguns critérios existentes com alguns propostos por ele em diferentes cenários (valores dos parâmetros) e abordagens (estratégias de seleção). Bayer (2011) observou que não é possível eleger um critério que seja melhor em todos os cenários. Uma das abordagens consideradas por Bayer (2011) é primeiro selecionar o submodelo da média supondo a dispersão constante e depois, assumindo que o modelo selecionado na etapa 1 seja correto, selecionar

o submodelo da dispersão. Ele concluiu que essa abordagem tem melhor performance do que a abordagem de selecionar simultaneamente os dois submodelos e requer a estimação de um menor número de modelos.

Os trabalhos que comparam critérios de seleção de modelos geralmente consideram que todos os possíveis modelos serão ajustados e dentre eles será selecionado o melhor modelo segundo o critério escolhido. No entanto, na prática, utilizar esse procedimento é inviável para o modelo RBIZUT se o número de variáveis preditoras disponíveis não for bem pequeno. Como o modelo RBIZUT ajusta os 5 parâmetros desconhecidos da distribuição da variável resposta em função de variáveis preditoras, se utilizarmos esse procedimento, precisaremos estimar 2^{5m} modelos, em que m é o número de variáveis preditoras disponíveis. Se $m = 5$, por exemplo, já seria necessário estimar mais de 33 milhões de modelos. Mesmo utilizando a abordagem proposta por Bayer (2011) e citada no parágrafo anterior, o número de modelos a ser estimado é muito grande.

Para contornar esse problema, existem algoritmos que procuram identificar o melhor sem ajustar todos os possíveis modelos. O mais simples desses algoritmos é o *stepwise* e suas variações (Kutner et al., 2004, Seção 9.4), que adiciona e retira variáveis do modelo de forma sequencial até que a adição e a retirada de variáveis do modelo não melhorem sua performance. Tradicionalmente, o *stepwise* considera o resultado de algum teste de hipótese para a adição ou retirada de variáveis do modelo, mas ele pode ser usado de forma semelhante utilizando-se qualquer critério de seleção de modelos. Diversos outros algoritmos são discutidos em Miller (2002, Capítulo 3).

No caso de modelos em que mais de um parâmetro da distribuição da variável resposta é ajustado em função de variáveis preditoras, como é o caso do modelo RBIZUT, o uso de um algoritmo de seleção de modelos envolve uma escolha adicional. Deve-se decidir se a seleção de variáveis em cada um dos submodelos deve ser feita de forma simultânea ou se deve selecionar as variáveis de um submodelo de cada vez. Caso seja utilizada essa última abordagem, deve-se decidir ainda qual a ordem de seleção de variáveis entre os submodelos. Para o modelo de regressão beta com dispersão variável pode-se adotar, por exemplo, a sugestão de Bayer (2011) já discutida de selecionar primeiro as variáveis do submodelo da média e depois aquelas do submodelo da dispersão. Além do trabalho dele, não temos conhecimento de nenhum outro que discute essa importante questão na classe de modelos GAMLSS.

Na aplicação do Capítulo 5 consideramos um modelo inicial com todas as variáveis preditoras disponíveis presentes em cada um dos 5 submodelos. Para a exclusão das variáveis, consideramos como critério o valor-p do teste de Wald para o parâmetro associado a cada uma das variáveis. A exclusão das variáveis foi feita até que nenhuma variável tivesse valor-p do teste de Wald para o parâmetro associado a ela superior a 0,05 e pudesse ser retirada sem ocasionar um aumento do AIC. Primeiro excluímos as variáveis dos submodelos que ajustam δ_0, δ_1 e δ_c. Em seguida, excluímos as variáveis que ajustam ϕ e por último as variáveis que ajustam μ. O motivo de termos iniciado pelas variáveis do componente discreto do modelo é que a variância assintótica (e assim possivelmente a variância em amostras finitas) dos estimadores de β_1 e β_2 são função de γ_0, γ_1 e γ_c, mas a variância assintótica dos estimadores de γ_0, γ_1 e γ_c não são função de β_1 e β_2. O motivo de termos excluído

as variáveis que ajustam ϕ antes daquelas que ajustam μ deve-se ao fato de que μ geralmente é um parâmetro de maior interesse que ϕ. É importante ressaltar que na aplicação do Capítulo 5 cada variável foi analisada separadamente em cada um dos submodelos. Assim, uma variável pode estar presente no modelo final em apenas alguns dos 5 submodelos.

Para o leitor interessado em seleção de modelos, uma boa introdução sobre o assunto pode ser encontrada em McQuarrie e Tsai (1998). Bayer (2011), Ngoc (2011) e diversos outros introduzem novos critérios de seleção de modelos e fazem revisão bibliográfica sobre o tema.

3.6 Conclusão

Neste capítulo introduzimos os modelos de regressão beta inflacionados truncados (RBIZUT). Esses modelos podem ser úteis para o ajuste de variáveis que são discretas nos pontos 0, 1 e em um terceiro ponto $0 < c < 1$, que pode variar entre as observações, e contínuas no intervalo $(c, 1)$. Obtivemos formas fechadas para a função escore e matriz de informação de Fisher, discutimos estimação dos parâmetros por máxima verossimilhança, apresentamos expressões para intervalos de confiança para os parâmetros e testes de hipóteses. Apresentamos ainda alguns resultados para os modelos RBIZUT com c variável, conduzimos estudos de simulação de Monte Carlo e discutimos seleção de modelos e endogeneidade. Dessa forma, com os resultados obtidos neste capítulo, é possível o ajuste de um modelo RBIZUT para as diversas variáveis da área de econometria que possivelmente podem ter distribuição BIZUT (dadas variáveis preditoras). No Capítulo 5 apresentamos uma aplicação do modelo RBIZUT para dados reais de cartão de crédito.

Capítulo 4

Análise de diagnóstico

Neste capítulo é discutida análise de diagnóstico para os modelos RBIZUT. Na Seção 4.1 discute-se análise de resíduos e um teste de qualidade de ajuste e na Seção 4.2.1 é desenvolvida análise de influência local.

4.1 Análise de resíduos e teste de qualidade de ajuste

Os resultados apresentados no Capítulo 3 permitem o ajuste de um modelo RBIZUT a um conjunto de dados reais e a realização de análise inferencial usando o modelo obtido. No entanto, as conclusões obtidas serão válidas apenas se o ajuste do modelo for satisfatório. Nesta seção, discutimos análise de resíduos e um teste de qualidade de ajuste que permitem avaliar se um modelo RBIZUT está bem ajustado.

Os vetores de parâmetros γ_0, γ_1 e γ_c são separáveis e ortogonais a β_1 e β_2. Além disso, enquanto os parâmetros dos primeiros vetores estão relacionados à componente discreta da variável resposta, os parâmetros dos demais vetores estão relacionados à componente contínua da variável resposta. Por esses motivos, é razoável realizar análise de resíduos das observações tais que $y_i \in (c_i, 1)$ separadamente daquelas tais que $y_i \notin (c_i, 1)$. A distribuição BIZUT é uma mistura entre a distribuição multinomial e a distribuição beta. Assim, podem ser utilizados os resíduos já desenvolvidos para os modelos de regressão beta e para os modelos de regressão multinomial.

Alguns trabalhos sobre resíduos para a regressão beta já foram realizados. No artigo que introduziu a regressão beta, Ferrari e Cribari-Neto (2004) sugeriram o uso do resíduo padronizado ou do resíduo deviance para o modelo de regressão beta com dispersão constante. Espinheira et al. (2008*b*) introduziram dois resíduos para essa mesma classe de modelos. Eles mostraram a partir de estudos de simulação que o segundo resíduo que eles propuseram, denominado resíduo ponderado padronizado 2, tem distribuição mais próxima da normal padrão e identifica melhor observações influentes do que o outro resíduo sugerido por eles e que os resíduos definidos por Ferrari e Cribari-Neto (2004). Esse resíduo foi estendido para o diagnóstico de modelos de regressão beta com dispersão variável por Ferrari et al. (2011). Correções para o resíduo padronizado e para o resíduo ponderado padronizado 1 de Espinheira et al. (2008*b*) foram desenvolvidas para o modelo de regressão beta com dispersão constante por Anholeto (2010). A partir de estudos de simulação, Anholeto (2010) concluiu que os resíduos corrigidos apresentam média mais próxima de zero e variância mais próxima de um do que as versões originais dos resíduos.

Para a análise de diagnóstico do componente contínuo do modelo RBIZUT sugerimos uma extensão natural do resíduo ponderado padronizado 2. Precisamos propor uma extensão do resíduo, porque a variável resposta no modelo considerado por Ferrari et al. (2011) assume valores no intervalo $(0, 1)$ e não no intervalo $(c_i, 1)$. Dessa forma, o resíduo proposto para análise de diagnóstico do componente contínuo do modelo é dado por

$$r_i^{\{c\}} = \frac{y_i^* - \mu_i^*}{\sqrt{\upsilon_i^*(1 - h_{ii})}}, \tag{4.1}$$

em que y_i^* foi definido em (3.8), μ_i^* e υ_i^* foram definidos em (3.10), h_{ii} é o elemento (i, i) da matriz $H = \hat{W}^{*1/2} X_1 (X_1^\top \hat{W}^* X_1)^{-1} X_1^\top \hat{W}^{*1/2}$ e W^* é uma matriz diagonal com elemento (i, i) dado por $w_i(\partial\mu_i/\partial\eta_{i1})^2$, w_i definido em (3.21). Apesar da expressão de $r_i^{\{c\}}$ ser semelhante à proposta por Ferrari et al. (2011), y_i^*, μ_i^*, υ_i^* e h_{ii} são diferentes para o modelo RBIZUT, em virtude do suporte da variável resposta ser diferente. Dessa forma, para as observações tais que $y_i \in (c_i, 1)$, propomos a utilização de 2 gráficos de diagnóstico utilizando $r_i^{\{c\}}$:

- diagrama de dispersão entre $r_i^{\{c\}}$ e $\hat{\mu}_i$. Caso c seja variável, pode ser mais interessante fazer o diagrama de dispersão entre $r_i^{\{c\}}$ e $\hat{\mu}_i^R = (\hat{\mu}_i - c_i)/(1 - c_i)$ para que o suporte da variável aleatória do eixo horizontal $(\hat{\mu}_i^R)$ seja o mesmo para todas as observações;

- gráfico de probabilidade meio-normal com envelope simulado para $r_i^{\{c\}}$ (Atkinson, 1985, Seção 4.2), já que o resíduo não tem distribuição exatamente normal.

Para a construção do envelope é necessário considerar que $r_i^{\{c\}}$ é definido apenas para as observações tais que $y_i \in (c_i, 1)$. Assim, o gráfico de probabilidade meio-normal com envelope simulado para $r_i^{\{c\}}$ no modelo RBIZUT é obtido a partir do seguinte procedimento (Ferrari e Cribari-Neto, 2004):

I Gere uma amostra da variável resposta considerando o modelo RBIZUT ajustado como o modelo verdadeiro para as observações tais que $y_i \in (c_i, 1)$ na amostra original e considerando o valor de y_i na amostra original para as observações tais que $y_i \notin (c_i, 1)$.

II Ajuste o modelo RBIZUT para a amostra simulada e obtenha o valor absoluto de cada um dos $r_i^{\{c\}}$ para as observações tais que $y_i \in (c_i, 1)$ e ordene-as em ordem crescente.

III Repita os passos I e II k vezes.

IV Considerando os $n^* = n - \sum_{i=1}^n [(I_{\{0\}}(y_i) + I_{\{1\}}(y_i) + I_{\{c_i\}}(y_i)]$ conjuntos de estatísticas de ordem contendo k observações cada, obtenha para cada conjunto sua média, seu mínimo e seu máximo.

V Represente nos gráficos esses valores e os valores absolutos dos resíduos $r_i^{\{c\}}$ ordenados no eixo vertical do gráfico e os escores meio normais dados por $\Phi^{-1}[(i + n^* - 1/8)/2n^* + 1/2)$ no eixo horizontal, em que Φ é a função distribuição normal padrão.

Na aplicação do Capítulo 5 utilizamos $k = 19$ como sugerido por Atkinson (1985).

Os trabalhos de análise de diagnóstico para modelos de regressão multinomiais foram desenvolvidos para a regressão logística multinomial. O primeiro trabalho nessa área foi desenvolvido por Lesaffre e Albert (1989), que estenderam a análise de diagnóstico para modelos de regressão

logística binária (Pregibon, 1981) para o caso multinomial. Os autores introduziram medidas para identificar observações discrepantes e influentes. Gupta et al. (2008) introduziram outros resíduos sugeridos principalmente quando os parâmetros do modelo são estimados utilizando o estimador de mínima ϕ-divergência (Gupta et al., 2006).

Para realizar a análise de diagnóstico do componente discreto do modelo RBIZUT sugerimos considerar que cada uma das observações pode assumir 4 diferentes valores: zero, um, c_i e um quarto valor, 2 por exemplo, para aquelas observações tais que $y_i \in (c_i, 1)$. Procedendo dessa forma, podemos avaliar simultaneamente não somente o ajuste do componente discreto do modelo como o ajuste do parâmetro de mistura α_i, que é a probabilidade de $y_i \in (c_i, 1)$. Assim para todas as observações da amostra sugerimos obter o resíduo estudentizado para a regressão logística multinomial (Lesaffre e Albert, 1989), que é um vetor com 3 posições para cada uma das unidades amostrais dado por

$$r_i^{\{d\}} = \left(r_i^{\{d0\}}, r_i^{\{d1\}}, r_i^{\{dc\}}\right)^\top = M_{ii}^{-1/2}\chi_i, \tag{4.2}$$

em que

$$\chi_i = \left(\frac{I_{\{0\}}(y_i) - \hat{\delta}_{i0}}{\hat{\delta}_{i0}}, \frac{I_{\{1\}}(y_i) - \hat{\delta}_{i1}}{\hat{\delta}_{i1}}, \frac{I_{\{c_i\}}(y_i) - \hat{\delta}_{ic}}{\hat{\delta}_{ic}}\right)^\top$$

e M_{ii} é o bloco de dimensão 3×3 de posição i da matriz quadrada $M = I - H$, ou seja, M_{ii} é uma matriz de dimensão 3×3 que contém as linhas e colunas $3i-2$, $3i-1$ e $3i$ da matriz M, em que $H = \hat{V}^{1/2}X(X^\top\hat{V}X)^{-1}X^\top\hat{V}^{1/2}$, $V = \text{diag}(V_1, V_2, \ldots, V_n)$, $X^\top = (X_{10}^\top, X_{20}^\top, \ldots, X_{n0}^\top)$,

$$V_i = \begin{pmatrix} \delta_{i0}(1-\delta_{i0}) & -\delta_{i0}\delta_{i1} & -\delta_{i0}\delta_{ic} \\ -\delta_{i0}\delta_{i1} & \delta_{i1}(1-\delta_{i1}) & -\delta_{i1}\delta_{ic} \\ -\delta_{i0}\delta_{ic} & -\delta_{i1}\delta_{ic} & \delta_{ic}(1-\delta_{ic}) \end{pmatrix}, X_{i0} = \begin{pmatrix} z_{i0}^\top & 0^\top & 0^\top \\ 0^\top & z_{i1}^\top & 0^\top \\ 0^\top & 0^\top & z_{ic}^\top \end{pmatrix}.$$

As três posições dos vetores se referem, respectivamente, aos valores zero, um e c_i que podem ser assumidos por y_i. O quarto valor não tem uma componente por ser tratado como valor de referência. Para que cada observação tenha apenas um resíduo, pode-se ainda obter (Lesaffre e Albert, 1989)

$$r_i^{\{dq\}} = \left(r_i^{\{d\}}\right)^\top r_i^{\{d\}}.$$

Os resíduos $r_i^{\{d\}}$ e $r_i^{\{dq\}}$ foram propostos para a regressão logística multinomial e assim, em princípio, só poderiam ser usados no modelo RBIZUT quando a função de ligação H fosse composta por funções de ligação logito. No entanto, $r_i^{\{d\}}$ e $r_i^{\{dq\}}$ dependem basicamente dos vetores de variáveis preditoras z_{i0}, z_{i1} e z_{ic} e das probabilidades ajustadas $\hat{\delta}_{i0}$, $\hat{\delta}_{i1}$ e $\hat{\delta}_{ic}$ e assim parece razoável utilizá-los para outras funções de ligação.

Para o componente discreto e o parâmetro de mistura sugerimos os seguintes gráficos de análise de diagnóstico:

- diagrama de dispersão entre $r_i^{\{d0\}}$ e $\hat{\delta}_{i0}$, entre $r_i^{\{d1\}}$ e $\hat{\delta}_{i1}$, entre $r_i^{\{dc\}}$ e $\hat{\delta}_{ic}$ e entre $r_i^{\{dq\}}$ e $\hat{\alpha}_i$;

- gráfico de probabilidade meio-normal com envelope simulado para $r_i^{\{d0\}}$, $r_i^{\{d1\}}$, $r_i^{\{dc\}}$ e $r_i^{\{dq\}}$.

O resíduo $r_i^{\{dq\}}$ e cada um dos componentes de $r_i^{\{d\}}$ têm distribuição assimétrica e bem diferente da distribuição normal. Além disso, para cada observação, a distribuição deles é discreta podendo assumir apenas 4 diferentes valores e a variância não é constante. Assim, no diagrama de dispersão entre $r_i^{\{d0\}}$ e $\hat{\delta}_{i0}$, por exemplo, não é esperado que os pontos se distribuam sem nenhum padrão como geralmente é observado com bons resíduos para modelos de regressão para variáveis contínuas quando o modelo está bem ajustado. Por todos esses motivos, assim como ocorre para resíduos de modelos para variáveis resposta que assumem um número pequeno de valores, a análise de resíduos do componente discreto do modelo e do parâmetro de mistura no modelo RBIZUT traz menos informação a respeito da qualidade do ajuste do que a análise de resíduos do componente contínuo. Dessa forma, para o componente discreto do modelo e o parâmetro de mistura é interessante a realização de um teste de qualidade de ajuste.

Alguns testes de qualidade de ajuste foram desenvolvidos para modelos de regressão com resposta multinomial. Fagerland et al. (2008) estudaram, utilizando resultados de simulação, as propriedades da extensão do teste de Hosmer e Lemeshow (1980), proposto originalmente para a regressão logística binária, para modelos multinomiais. Eles concluíram que a estatística do teste, sob a hipótese de que o modelo está bem ajustado, é bem aproximada pela distribuição qui-quadrado com $3(g-2)$ graus de liberdade, em que g é o número de grupos utilizados no teste conforme explicado a seguir. Eles concluíram ainda que o teste necessita de amostras grandes para ter um poder satisfatório. Outros testes de qualidade de ajuste para modelos multinomiais foram propostos por Osius e Rojek (1992), Pigeon e Heyse (1999) e Goeman e le Cessie (2006). Conforme já discutido, para os tipos de variáveis que podem ser ajustadas pelo modelo RBIZUT, não há dificuldade na obtenção de amostras grandes. Dessa forma, para testar a qualidade de ajuste do componente discreto do modelo e do parâmetro de mistura no modelo RBIZUT sugerimos a extensão do teste de Hosmer e Lemeshow (1980) para modelos multinomiais, já que é o mais simples dos testes propostos para modelos multinomiais e possui poder satisfatório para amostras grandes.

O teste sugerido para o modelo RBIZUT pode ser conduzido utilizando o seguinte procedimento:

- Divida as observações da amostra em g grupos com (aproximadamente) o mesmo número de observações, utilizando os quantis de $\hat{\alpha}_i$. Dessa forma, o primeiro grupo conterá as n/g observações com menores $\hat{\alpha}_i$, o segundo grupo conterá as n/g observações com menores $\hat{\alpha}_i$ e que não estão no primeiro grupo e assim por diante.

- Para cada grupo, obtenha O_{k1}, O_{k2}, O_{k3}, O_{k4}, o número de observações no grupo k tais que $y_i = 0$, $y_i = 1$, $y_i = c_i$ e $y_i \in (c_i, 1)$, respectivamente.

- Para cada grupo, calcule E_{k1}, E_{k2}, E_{k3}, E_{k4}, respectivamente, o valor esperado de observações se o modelo está bem ajustado no grupo k tais que $y_i = 0$, $y_i = 1$, $y_i = c_i$ e $y_i \in (c_i, 1)$, que

podem ser calculados a partir de

$$E_{k1} = \sum_{\Omega_k} \hat{\delta}_{i0},\ E_{k2} = \sum_{\Omega_k} \hat{\delta}_{i1},\ E_{k3} = \sum_{\Omega_k} \hat{\delta}_{ic},\ \text{e}\ E_{k4} = \sum_{\Omega_k} (1 - \hat{\alpha}_i),$$

em que Ω_k denota os índices das observações do grupo k.

- Obtenha a estatística do teste dada por

$$C_g = \sum_{k=1}^{g} \sum_{j=1}^{4} (O_{kj} - E_{kj})^2 / E_{kj} \tag{4.3}$$

- Calcule o nível descritivo do teste a partir da distribuição qui-quadrado com $3(g-2)$ graus de liberdade.

Fagerland et al. (2008) efetuou estudos de simulação para $g = 8, 10$ e 12 e nenhum deles apresentou melhor performance que os demais em todos os cenários estudados. Na aplicação do Capítulo 5 utilizamos $g = 10$.

Conforme já discutimos, é razoável realizar separadamente análise de resíduos dos componentes discretos e contínuos do modelo. No entanto, suponha que estamos utilizando $r_i^{\{c\}}$ definido em (4.1) para avaliar o ajuste da parte contínua e que a observação i apresente um alto resíduo. É razoável imaginar que, fixado o valor desse resíduo, quanto menor $\mathrm{P}(y_i \in (c_i, 1))$, mais discrepante será a observação i. Assim, para as observações tais que $y_i \in (c_i, 1)$ seria interessante introduzir uma modificação de $r_i^{\{c\}}$ que considere $\hat{\alpha}_i$, que é um estimador de $\mathrm{P}(y_i \in (c_i, 1))$. Para o modelo não linear de regressão beta inflacionado no 0, Ospina e Ferrari (2012) sugeriram um resíduo com essa característica, a partir do algoritmo scoring de Fisher para estimação dos parâmetros que modelam a média dos valores do componente contínuo da variável resposta. Para o modelo RBIZUT, o resíduo proposto por eles é dado por

$$r_i^{\{cd\}} = \frac{y_i^* - \mu_i^*}{\sqrt{\upsilon_i^*(1 - h_{ii})(1 - \hat{\alpha}_i)}}, \tag{4.4}$$

em que todas as quantidades são obtidas conforme definido em (4.1), com exceção dos elementos de W^*, que é utilizado no cálculo de h_{ii}, e em (4.4) é uma matriz diagonal com elementos dados $w_i(1 - \hat{\alpha}_i)(\partial\mu_i/\partial\eta_{i1})^2$.

O resíduo $r_i^{\{c\}}$ apresenta boas propriedades conforme estudado por Espinheira et al. (2008*b*) a partir de estudos de simulação. Observando-se a expressão de $r_i^{\{cd\}}$ em (4.4), podemos imaginar que, caso a variância de $r_i^{\{c\}}$ não varie em função de μ_i então provavelmente a variância de $r_i^{\{cd\}}$ variará. Assim, para estudar se $r_i^{\{cd\}}$ é heterocedástico, conduzimos um estudo de simulação utilizando-se a linguagem de programação Ox. Utilizou-se 5000 replicações, $n = 500$, c variável gerado a partir de uma distribuição uniforme no intervalo (0,1, 0,3) e a estrutura do modelo (3.35) com os seguintes valores para os parâmetros: β_{11} = -1,0, β_{21} = 2,5, β_{31} = -1,7, β_{12} = 2,5, β_{22} = 0,9, β_{32} = -0,9, γ_{10} = -3,0, γ_{20} = -0,5, γ_{11} = -2,0, γ_{21} = 4,0, γ_{1c} = -2,5 e γ_{2c} = -0,7. Com esses valores para os parâmetros, $\delta_0 \in (0{,}004, 0{,}039), \delta_1 \in (0{,}107, 0{,}873), \delta_c \in (0{,}005, 0{,}065), \mu \in (0{,}157, 0{,}872), \phi \in (4{,}953, 29{,}964)$.

Assim como nos estudos de simulação do Capítulo 3, $x_{i11} = x_{i12} = z_{i10} = z_{i11} = z_{i1c} = 1$, $x_{i21} = x_{i22} = z_{i20} = z_{i21} = z_{i2c}$ e as variáveis preditoras foram geradas independentemente a partir de variáveis aleatórias com distribuição uniforme no intervalo $(0, 1)$ e mantidas constantes em todas as replicações.

Os resultados do estudo de simulação são apresentados na Tabela 4.1. Para a obtenção dos resultados apresentados na tabela, em cada replicação dividimos as observações em 4 grupos em função do valor de $\hat{\mu}_i^R$ e calculamos a média e a variância de $r_i^{\{c\}}$ e $r_i^{\{cd\}}$. Os valores apresentados na tabela são as médias dos valores obtidos nas 5000 replicações. Conforme esperado, pode-se notar que a variância de $r_i^{\{cd\}}$ varia consideravelmente em função de $\hat{\mu}_i^R$. Dessa forma, em gráficos de dispersão entre $r_i^{\{cd\}}$ e $\hat{\mu}_i$ ou entre $r_i^{\{cd\}}$ e $\hat{\mu}_i^R$ é provável que os pontos se distribuam com um padrão que sugira a presença de heterocedasticidade mesmo que o componente contínuo do modelo esteja bem ajustado. Por outro lado, a variação da variância do resíduo $r_i^{\{c\}}$ em função de $\hat{\mu}_i^R$ parece ser bem menor do que a observada para $r_i^{\{cd\}}$. Diante do que foi discutido e também das boas propriedades de $r_i^{\{c\}}$ estudadas por Espinheira et al. (2008*b*), sugerimos o uso de $r_i^{\{c\}}$ e não de $r_i^{\{cd\}}$ para a análise de diagnóstico do componente contínuo do modelo RBIZUT, mesmo levando-se em consideração que $r_i^{\{c\}}$ não é função de um estimador de $\mathrm{P}(y_i \in (c_i, 1))$.

Tabela 4.1: Resultados do estudo de simulação para os resíduos $r_i^{\{c\}}$ e $r_i^{\{cd\}}$

Resíduo	Estimativa	$\hat{\mu}_i^R \in (0{,}0, 0{,}2]$	$\hat{\mu}_i^R \in (0{,}2, 0{,}4]$	$\hat{\mu}_i^R \in (0{,}4, 0{,}6]$	$\hat{\mu}_i^R \in (0{,}6, 1{,}0)$
$r_i^{\{c\}}$	Média	0,004	-0,003	0,002	0,002
	Variância	0,988	1,004	0,988	0,929
$r_i^{\{cd\}}$	Média	0,004	-0,004	0,004	0,004
	Variância	1,400	1,833	3,179	4,580

Como $r_i^{\{cd\}}$ é heterocedástico, seria bastante interessante a obtenção de uma outra modificação de $r_i^{\{c\}}$ que considere $\mathrm{P}(y_i \in (c_i, 1))$, mas seja homocedástica. No entanto, a obtenção de uma modificação de $r_i^{\{cd\}}$ com essa caraterística não é simples. Suponha que α_i fosse conhecido e propuséssemos, por exemplo, um resíduo da forma $r_i^{\{cd*\}}{=}h(\alpha_i)r_i^{\{c\}}$, sendo h uma função qualquer. Nesse caso, $\mathrm{VAR}(r_i^{\{cd*\}}) = [h(\alpha_i)]^2\mathrm{VAR}(r_i^{\{c\}})$ que seria heterocedástico se $(r_i^{\{c\}})$ for homocedástico. Na prática, α_i é desconhecido, mas esse exemplo simples ilustra a dificuldade de obtenção de uma modificação de $r_i^{\{c\}}$ que considere $\mathrm{P}(y_i \in (c_i, 1))$, mas seja homocedástico.

Também seria interessante a definição de um resíduo que pudesse ser utilizado na análise de diagnóstico tanto do componente contínuo do modelo quanto do componente discreto. Para o modelo RBIZU, Ospina (2008) sugere o uso de resíduos quantis aleatorizados (Dunn e Smyth, 1996). Esse resíduo é definido como

$$r_i^{\{g\}} = \Phi^{-1}(u_i), \tag{4.5}$$

em que Φ é a função distribuição normal padrão e, para o modelo RBIZUT, u_i é definido da seguinte

forma:

$$u_i = \begin{cases} U(0, \hat{\delta}_{i0}) & \text{se } y_i = 0 \\ U(\hat{\delta}_{i0}, \hat{\delta}_{i0} + \hat{\delta}_{ic}) & \text{se } y_i = c_i \\ F(y_i; \hat{\delta}_{i0}, \hat{\delta}_{i1}, \hat{\delta}_{ic}, \hat{\mu}_i, \hat{\phi}_i, c_i) & \text{se } y_i \in (c_i, 1), \\ U(1 - \hat{\delta}_{i1}, 1) & \text{se } y_i = 1, \end{cases} \tag{4.6}$$

em que $F(.)$ foi definida em (2.7) e $U(a, b)$ é uma variável aleatória com distribuição uniforme no intervalo (a, b). Assim, se $y_i = 0, 1$ ou c_i, $r_i^{\{g\}}$ é obtido gerando-se uma variável aleatória uniforme no intervalo especificado em (4.6) e portanto um único modelo pode produzir diferentes vetores de resíduos quantis aleatorizados. Se o modelo estiver bem ajustado, $r_i^{\{g\}}$ tem distribuição normal padrão.

Apesar da distribuição de $r_i^{\{g\}}$ ser conhecida se o modelo estiver bem ajustado e ele ser bastante interessante para o modelo RBIZU, $r_i^{\{g\}}$ não é útil para o modelo RBIZUT. Observando-se (4.5) e (4.6) pode-se notar que, se $y_i \in (c_i, 1)$, $r_i^{\{g\}}$ é necessariamente maior que $\Phi^{-1}(\hat{\delta}_{i0} + \hat{\delta}_{ic})$. Isso significa que se $y_i \in (c_i, 1)$ e $\hat{\delta}_{i0}$ não for pequeno, $r_i^{\{g\}}$ necessariamente não poderá ser ao mesmo tempo negativo e apresentar um valor grande, em módulo. Isso em geral não é razoável. Imaginemos, por exemplo, que estamos ajustando um modelo RBIZUT para uma variável definida como a razão entre o valor recebido de seguro desemprego em setembro de 2011 e o máximo valor do benefício naquela data. Suponha ainda que nossa população sejam os brasileiros acima de 18 anos que perderam o emprego em junho de 2011. Para um profissional jovem com curso superior, espera-se que $\hat{\delta}_{i0}$ não seja tão pequeno, porque o profissional, por ser jovem, tem uma probabilidade razoável de não ser elegível ao benefício por não ter trabalhado pelo menos 6 meses consecutivos antes de perder o emprego. Por outro lado, por ele ser bem qualificado, espera-se que $\hat{\mu}_i$ esteja próximo de 1, já que o valor do benefício está relacionado com os 3 últimos salários do desempregado. Assim, um profissional com essas características que recebesse um valor de seguro desemprego próximo ao mínimo deveria ser considerado uma observação discrepante. No entanto, como $\hat{\delta}_{i0}$ não é pequeno, o valor de $r_i^{\{g\}}$ nesse caso não nos levaria a classificar essa observação como discrepante.

Conforme foi discutido, não foi possível obter um resíduo que pudesse ser utilizado na análise de diagnóstico tanto do componente contínuo do modelo quanto do componente discreto e nem mesmo um bom resíduo para o componente contínuo do modelo que leve em consideração um estimador de $\mathrm{P}(y_i \in (c_i, 1))$. Dessa forma, sugerimos que as principais ferramentas de diagnóstico para o modelo RBIZUT sejam $r_i^{\{c\}}$ definido em (4.1) para o componente contínuo do modelo e o teste de bondade do ajuste com estatística C_g definida em (4.3) e estudado por Fagerland et al. (2008) para o componente discreto do modelo.

4.2 Análise de influência local

A análise de influência local é uma método de diagnóstico proposto por Cook (1986). Seu objetivo é verificar se existem observações que causam efeitos desproporcionais nos resultados da análise inferencial, quando pequenas perturbações são incorporadas ao modelo e/ou aos dados do estudo. Caso haja observações com essas características, teremos indícios de que o modelo está mal

ajustado, ou alguma suposição do modelo não é razoável para o conjunto de dados em análise, ou pode haver observações que foram obtidas de forma incorreta (por erro de digitação por exemplo). Nesse caso, a identificação dessas observações pode tanto auxiliar na escolha de um modelo mais adequado, como também fornecer informações sobre a sensibilidade do modelo a essas perturbações.

Um dos motivos da popularidade desta técnica é que ela pode ser aplicada a qualquer modelo desde que se tenha a função de verossimilhança. Assim, inúmeros trabalhos desenvolveram análise de influência local para diferentes modelos baseados na proposta de Cook (1986) ou em extensões da mesma, sendo que em algumas extensões não é necessária nem mesmo a existência da função de verossimilhança. Citamos alguns trabalhos sobre influência local a seguir. Lesaffre e Verbeke (1998) apresentaram análise de influência local em modelos lineares mistos normais e Ouwens et al. (2001) em modelos lineares mistos generalizados. Galea et al. (1997) desenvolveram análise de influência local em modelos elípticos lineares. Já Zhao et al. (1994) propuseram análise de influência local em modelos lineares generalizados com erros nas variáveis. Alguns trabalhos recentes na área foram desenvolvidos por Russo et al. (2009) em modelos não lineares elípticos para dados correlacionados, Venezuela et al. (2011) em modelos de equações de estimação generalizadas e Riquelme (2012) em modelos mistos com erros nas variáveis. Trabalhos de análise de influência local relacionados com a regressão beta incluem Espinheira et al. (2008*a*) em modelos de regressão beta com dispersão constante, Ferrari et al. (2011) em modelos de regressão beta com dispersão variável, Ospina (2008) em modelos de regressão beta inflacionados e Venezuela (2008) em modelos de regressão beta para dados longitudinais.

Seja $l(\theta)$ o logaritmo da função de verossimilhança do modelo proposto, em que θ é o vetor de parâmetros desconhecidos do modelo com dimensão p. Seja ω um vetor de perturbações de dimensão $s \times 1$ restrito a algum conjunto aberto Ω de $\mathbb{R}^s$, em que s geralmente é igual a n, que é o número de observações da amostra. Seja $l(\theta|\omega)$ o logaritmo da função de verossimilhança do modelo perturbado e ω_0 o vetor de não perturbação tal que $l(\theta|\omega_0) = l(\theta)$. Cook (1986) propôs estudar o efeito de ω sobre θ a partir do afastamento pela verossimilhança (*likelihood displacement*) definido como

$$\mathrm{LD}(\omega) = 2\{l(\hat{\theta}) - l(\hat{\theta}_\omega)\}, \tag{4.7}$$

em que $\hat{\theta}$ e $\hat{\theta}_\omega$ são os estimadores de máxima verossimilhança de θ para o modelo proposto e para o modelo perturbado, respectivamente. A ideia da influência local consiste basicamente em estudar o comportamento da função $\mathrm{LD}(\omega)$ em torno de ω_0. Para isso, Cook (1986) utiliza curvaturas normais (Bates e Watts, 1980). O autor sugere avaliar a curvatura do gráfico $\mathrm{LD}_{\omega_0+ad} \times a$ em que $a \in \mathbb{R}$ e d é uma direção de norma igual a 1. Cook (1986) mostrou que a curvatura normal na direção d pode ser expressa como

$$C_d(\theta) = 2|d^\top \Delta^\top \ddot{l}^{-1} \Delta d|, \tag{4.8}$$

em que $-\ddot{l}$ é a matriz de informação observada para θ e Δ é uma matriz de dimensão $p \times s$ com elemento (j, i) dado por $\Delta_{ji} = \partial^2 l(\theta|\omega)/\partial\theta_j \partial\omega_i$, ambas as matrizes avaliadas em $\theta = \hat{\theta}$ e $\omega = \omega_0$. O autor propõe analisar a direção $d_{\max}$, direção correspondente à maior curvatura $C_{\max}$. Desta

forma, $C_{\max}$ é o maior autovalor da matriz $\Delta^\top \ddot{l}^{-1}\Delta$ e $d_{\max}$ é o correspondente autovetor de norma igual a 1. Para um determinado esquema de perturbação, se o componente de posição i do vetor $d_{\max}$ tiver um valor relativamente grande em relação aos valores da maioria dos demais componentes de $d_{\max}$, então é possível que a observação de posição i da amostra tenha uma influência desproporcional quando pequenas perturbações são incorporadas no modelo ou nos dados. Assim, gráficos de $d_{\max}$ em função da ordem das observações permitem identificar observações que podem ser localmente influentes sob determinado esquema de perturbação.

4.2.1 Influência local no modelo RBIZUT

Baseado no que foi descrito, pode-se notar que é necessária apenas a obtenção das matrizes $\ddot{l}$ e Δ e do vetor ω_0 para a realização de análise de influência local em modelos de regressão paramétricos que contêm apenas efeitos fixos e observações independentes. No caso do modelo RBIZUT, a matriz de informação observada, que é igual a $-\ddot{l}$ foi obtida usando os cálculos do Apêndice B.3 e pode ser expressa como

$$K^*(\theta) = \begin{pmatrix} K_D^*(\gamma_0,\gamma_1,\gamma_c) & 0 \\ 0 & K_C^*(\gamma_0,\gamma_1,\gamma_c,\beta_1,\beta_2) \end{pmatrix}, \tag{4.9}$$

em que

$$K_D^*(\gamma_0,\gamma_1,\gamma_c) = \begin{pmatrix} K^*_{\gamma_0\gamma_0} & K^*_{\gamma_0\gamma_1} & K^*_{\gamma_0\gamma_c} \\ K^*_{\gamma_1\gamma_0} & K^*_{\gamma_1\gamma_1} & K^*_{\gamma_1\gamma_c} \\ K^*_{\gamma_c\gamma_0} & K^*_{\gamma_c\gamma_1} & K^*_{\gamma_c\gamma_c} \end{pmatrix}, \; K_C^*(\gamma_0,\gamma_1,\gamma_c,\beta_1,\beta_2) = \begin{pmatrix} K^*_{\beta_1\beta_1} & K^*_{\beta_1\beta_2} \\ K^*_{\beta_2\beta_1} & K^*_{\beta_2\beta_2} \end{pmatrix}, \tag{4.10}$$

$$\begin{aligned}
K^*_{\gamma_0\gamma_0} &= Z_0^\top[\Delta_0^* T_0^2 + \Delta_1^* T_{10}^2 + \Delta_c^* T_{c0}^2 + \Delta^*_{(c,1)}(T_0 + T_{10} + T_{c0})^2 + \Delta^*_{00}]Z_0, \\
K^*_{\gamma_1\gamma_1} &= Z_1^\top[\Delta_0^* T_{01}^2 + \Delta_1^* T_1^2 + \Delta_c^* T_{c1}^2 + \Delta^*_{(c,1)}(T_{01} + T_1 + T_{c1})^2 + \Delta^*_{11}]Z_1, \\
K^*_{\gamma_c\gamma_c} &= Z_c^\top[\Delta_0^* T_{0c}^2 + \Delta_1^* T_{1c}^2 + \Delta_c^* T_c^2 + \Delta^*_{(c,1)}(T_{0c} + T_{1c} + T_c)^2 + \Delta^*_{cc}]Z_c, \\
K^*_{\gamma_0\gamma_1} &= (K^*_{\gamma_1\gamma_0})^\top = Z_0^\top[\Delta_0^* T_0 T_{01} + \Delta_1^* T_{10} T_1 + \Delta_c^* T_{c0} T_{c1} \\
&\quad + \Delta^*_{(c,1)}(T_0 + T_{10} + T_{c0})(T_{01} + T_1 + T_{c1}) + \Delta^*_{01}]Z_1, \\
K^*_{\gamma_0\gamma_c} &= (K^*_{\gamma_c\gamma_0})^\top = Z_0^\top[\Delta_0^* T_0 T_{0c} + \Delta_1^* T_{10} T_{1c} + \Delta_c^* T_{c0} T_c \\
&\quad + \Delta^*_{(c,1)}(T_0 + T_{10} + T_{c0})(T_{0c} + T_{1c} + T_c) + \Delta^*_{0c}]Z_c, \\
K^*_{\gamma_1\gamma_c} &= (K^*_{\gamma_c\gamma_1})^\top = Z_1^\top[\Delta_0^* T_{01} T_{0c} + \Delta_1^* T_1 T_{1c} + \Delta_c^* T_{c1} T_c \\
&\quad + \Delta^*_{(c,1)}(T_{01} + T_1 + T_{c1})(T_{0c} + T_{1c} + T_c) + \Delta^*_{1c}]Z_c, \\
K^*_{\beta_1\beta_1} &= X_1^\top[\Delta^{**}_{(c,1)} W T_\mu^2 + \Delta^*_{\mu\mu}]X_1, \\
K^*_{\beta_2\beta_2} &= X_2^\top[\Delta^{**}_{(c,1)} B T_\phi^2 + \Delta^*_{\phi\phi}]X_2, \\
K^*_{\beta_1\beta_2} &= K^*_{\beta_2\beta_1}{}^\top = X_1^\top[\Delta^{**}_{(c,1)} A^* T_\mu T_\phi]X_2,
\end{aligned} \tag{4.11}$$

$\Delta_0^* = \mathrm{diag}\{I_{\{0\}}(y_1)/\delta_{10}^2, \ldots, I_{\{0\}}(y_n)/\delta_{n0}^2\}$, $\Delta_1^* = \mathrm{diag}\{I_{\{1\}}(y_1)/\delta_{11}^2, \ldots, I_{\{1\}}(y_n)/\delta_{n1}^2\}$, $\Delta_c^* = \mathrm{diag}\{I_{\{c_1\}}(y_1)/\delta_{1c}^2, \ldots, I_{\{c_n\}}(y_n)/\delta_{nc}^2\}$, $\Delta^*_{(c,1)} = \mathrm{diag}\{I_{(c_1,1)}(y_1), \ldots, I_{(c_n,1)}(y_n)\}$, $\Delta^*_{00} = \mathrm{diag}\{\Delta^*_{100},$

$\ldots, \Delta^*_{n00}\}$, $\Delta^*_{11} = \text{diag}\{\Delta^*_{111}, \ldots, \Delta^*_{n11}\}$, $\Delta^*_{cc} = \text{diag}\{\Delta^*_{1cc}, \ldots, \Delta^*_{ncc}\}$, $\Delta^*_{01} = \text{diag}\{\Delta^*_{101}, \ldots, \Delta^*_{n01}\}$, $\Delta^*_{0c} = \text{diag}\{\Delta^*_{10c}, \ldots, \Delta^*_{n0c}\}$, $\Delta^*_{1c} = \text{diag}\{\Delta^*_{11c}, \ldots, \Delta^*_{n1c}\}$, $\Delta^*_{i00} = a^*_{i0}(\partial^2\delta_{i0}/\partial\zeta^2_{i0}) + a^*_{i1}(\partial^2\delta_{i1}/\partial\zeta^2_{i0}) + a^*_{ic}(\partial^2\delta_{ic}/\partial\zeta^2_{i0})$, $\Delta^*_{i11} = a^*_{i0}(\partial^2\delta_{i0}/\partial\zeta^2_{i1}) + a^*_{i1}(\partial^2\delta_{i1}/\partial\zeta^2_{i1}) + a^*_{ic}(\partial^2\delta_{ic}/\partial\zeta^2_{i1})$, $\Delta^*_{icc} = a^*_{i0}(\partial^2\delta_{i0}/\partial\zeta^2_{ic}) + a^*_{i1}(\partial^2\delta_{i1}/\partial\zeta^2_{ic}) + a^*_{ic}(\partial^2\delta_{ic}/\partial\zeta^2_{ic})$, $\Delta^*_{i01} = a^*_{i0}(\partial^2\delta_{i0}/\partial\zeta_{i0}\partial\zeta_{i1}) + a^*_{i1}(\partial^2\delta_{i1}/\partial\zeta_{i0}\partial\zeta_{i1}) + a^*_{ic}(\partial^2\delta_{ic}/\partial\zeta_{i0}\partial\zeta_{i1})$, $\Delta^*_{i0c} = a^*_{i0}(\partial^2\delta_{i0}/\partial\zeta_{i0}\partial\zeta_{ic}) + a^*_{i1}(\partial^2\delta_{i1}/\partial\zeta_{i0}\partial\zeta_{ic}) + a^*_{ic}(\partial^2\delta_{ic}/\partial\zeta_{i0}\partial\zeta_{ic})$, $\Delta^*_{i1c} = a^*_{i0}(\partial^2\delta_{i0}/\partial\zeta_{i1}\partial\zeta_{ic}) + a^*_{i1}(\partial^2\delta_{i1}/\partial\zeta_{i1}\partial\zeta_{ic}) + a^*_{ic}(\partial^2\delta_{ic}/\partial\zeta_{i1}\partial\zeta_{ic})$, $a^*_{i0} = -I_{\{0\}}(y_i)/\delta_{i0} + I_{(c_i,1)}(y_i)/(1-\alpha_i)$, $a^*_{i1} = -I_{\{1\}}(y_i)/\delta_{i1} + I_{(c_i,1)}(y_i)/(1-\alpha_i)$, $a^*_{ic} = -I_{\{c_i\}}(y_i)/\delta_{ic} + I_{(c_i,1)}(y_i)/(1-\alpha_i)$, $\Delta^{**}_{(c,1)} = \text{diag}\{I_{(c_1,1)}(y_1), \ldots, I_{(c_n,1)}(y_n)\}$, $\Delta^*_{\mu\mu} = \text{diag}\{\Delta^*_{1\mu\mu}, \ldots, \Delta^*_{n\mu\mu}\}$, $\Delta^*_{\phi\phi} = \text{diag}\{\Delta^*_{1\phi\phi}, \ldots, \Delta^*_{n\phi\phi}\}$, $A^* = \text{diag}\{a^*_1, \ldots, a^*_n\}$, $\Delta_{i\mu\mu} = [\phi_i/(1-c_i)](y^*_i - \mu^*_i)(\partial^2\mu_i/\partial\eta^2_{i1})$, $\Delta_{i\phi\phi} = [1/(1-c_i)][\mu_i(y^*_i - \mu^*_i) + y^\bullet_i - \mu^\bullet_i](\partial^2\phi_i/\partial\eta^2_{i2})$, $a^*_i = [-1/(1-c_i)](y^*_i - \mu^*_i) - [\phi_i/(1-c_i)]\{[(1-\mu_i)/(1-c_i)]\psi'[\phi_i(1-\mu_i)/(1-c_i)] - [(\mu_i - c_i)/(1-c_i)]\psi'[\phi_i(\mu_i - c_i)/(1-c_i)]\}$ e demais escalares e matrizes definidos em (3.2), (3.8), (3.9), (3.10), (3.11) e (3.21).

A matriz Δ e o vetor ω_0 variam de acordo com o esquema de perturbação utilizado. Os esquemas de perturbação mais utilizados são: ponderação de casos, perturbação da variável resposta e perturbação individual de variáveis preditoras. A perturbação da variável resposta só faz sentido se a mesma for estritamente contínua, já que nesse caso perturba-se y_i de forma que $y_i(\omega) = y_i + h(\omega_i)$, em que geralmente $h(\omega_i) = \omega_i$ ou $h(\omega_i) = \omega_i s(y_i)$, sendo $s(y_i)$ um estimador do desvio padrão de y_i. Assim, no caso de variáveis resposta que não são estritamente contínuas, esse esquema de perturbação não é razoável, já que $y_i(\omega)$ pode assumir valores que a variável resposta não pode assumir. Como a variável resposta no modelo RBIZUT não é estritamente contínua, consideramos neste trabalho apenas os seguintes esquemas de perturbação: ponderação de casos e perturbação individual de variáveis preditoras.

Ponderação de casos

O primeiro esquema de perturbação considerado é a ponderação de casos, que consiste em perturbar o logaritmo da função de verossimilhança individual da forma

$$l(\theta|\omega) = \sum_{i=1}^{n} \omega_i l_i(\theta), \tag{4.12}$$

em que $0 \leq \omega_i \leq 1$, para $i = 1, 2, \ldots, n$. Das equações (3.4) e (3.5) obtém-se que, para o modelo RBIZUT,

$$l(\theta|\omega) = l_1(\gamma_0, \gamma_1, \gamma_c|\omega) + l_2(\beta_1, \beta_2|\omega), \tag{4.13}$$

em que

$$\begin{aligned} l_1(\gamma_0, \gamma_1, \gamma_c|\omega) &= \sum_{i=1}^{n} \omega_i l_{i1}(\delta_{i0}, \delta_{i1}, \delta_{ic}), \\ l_2(\beta_1, \beta_2|\omega) &= \sum_{i:y_i \in (c_i,1)} \omega_i l_{i2}(\mu_i, \phi_i), \end{aligned} \tag{4.14}$$

$l_{i1}(\delta_{i0}, \delta_{i1}, \delta_{ic})$ e $l_{i2}(\mu_i, \phi_i)$ definidos em (3.5). Para a ponderação de casos, $\omega_0 = (1, 1, \ldots, 1)^\top$ e a matriz Δ nesse esquema de perturbação para o modelo RBIZUT pode ser segmentada de forma que

$\Delta = (\Delta_{\gamma_0}, \Delta_{\gamma_1}, \Delta_{\gamma_c}, \Delta_{\beta_1}, \Delta_{\beta_2})^\top$, em que cada uma das matrizes que compõem Δ tem n colunas e número de linhas igual à dimensão do correspondente vetor de parâmetros. No Apêndice C.1, prova-se que essas matrizes podem ser expressas como

$$\begin{aligned}
\Delta_{\gamma_0} &= Z_0^\top(\widehat{T}_0\widehat{\mathcal{E}}_0 + \widehat{T}_{10}\widehat{\mathcal{E}}_1 + \widehat{T}_{c0}\widehat{\mathcal{E}}_c),\\
\Delta_{\gamma_1} &= Z_1^\top(\widehat{T}_{01}\widehat{\mathcal{E}}_0 + \widehat{T}_1\widehat{\mathcal{E}}_1 + \widehat{T}_{c1}\widehat{\mathcal{E}}_c),\\
\Delta_{\gamma_c} &= Z_c^\top(\widehat{T}_{0c}\widehat{\mathcal{E}}_0 + \widehat{T}_{1c}\widehat{\mathcal{E}}_1 + \widehat{T}_c\widehat{\mathcal{E}}_c),\\
\Delta_{\beta_1} &= X_1^\top\widehat{\phi}^*\mathrm{mdiag}(y_{(c,1)})\widehat{T}_\mu\widehat{\mathcal{E}}_\mu,\\
\Delta_{\beta_2} &= X_2^\top c^*\mathrm{mdiag}(y_{(c,1)})\widehat{T}_\phi\widehat{\mathcal{E}}_\phi,
\end{aligned} \tag{4.15}$$

em que $\mathcal{E}_0 = \mathrm{diag}\{-a^*_{10}, \ldots, -a^*_{n0}\}$, $\mathcal{E}_1 = \mathrm{diag}\{-a^*_{11}, \ldots, -a^*_{n1}\}$, $\mathcal{E}_c = \mathrm{diag}\{-a^*_{1c}, \ldots, -a^*_{nc}\}$, a^*_{i0}, a^*_{i1} e a^*_{ic} definidos em (4.11), $\mathcal{E}_\mu = \mathrm{diag}\{y^*_1 - \mu^*_1, \ldots, y^*_n - \mu^*_n\}$, $\mathcal{E}_\phi = \mathrm{diag}\{\mu_1(y^*_1 - \mu^*_1) + y^\bullet_1 - \mu^\bullet_1, \ldots, \mu_n(y^*_n - \mu^*_n) + y^\bullet_n - \mu^\bullet_n\}$, mdiag(.) é o operador que transforma um vetor em uma matriz diagonal e as demais matrizes como definidas em (3.11).

Perturbação individual de variáveis preditoras

As cinco matrizes de variáveis preditoras do modelo RBIZUT podem ser iguais, podem conter apenas algumas variáveis em comum ou podem ser completamente diferentes. Na prática, é razoável imaginar que o mais comum seja que algumas variáveis estejam presentes em todas as 5 equações do modelo RBIZUT, enquanto outras estejam presentes em apenas uma ou em apenas algumas das equações do modelo. A matriz Δ para a perturbação individual de uma variável preditora vai variar em função das equações do modelo RBIZUT em que a variável está presente. Consideramos neste trabalho três cenários: perturbação de uma variável preditora presente apenas em uma das equações do componente discreto do modelo, perturbação de uma variável preditora presente apenas na equação para μ e perturbação de uma variável preditora presente em uma das equações do componente discreto e na equação para μ. Sem perda de generalidade, assumimos que no primeiro e no terceiro casos, a variável perturbada está presente em Z_0. Por motivos semelhantes aos discutidos anteriormente, para a perturbação da variável resposta, a perturbação individual de variáveis preditoras só é razoável se a variável for contínua.

Perturbação de uma variável presente apenas em Z_0

Admitamos que a variável que está na posição m de Z_0 seja perturbada de forma que $z_{im0}(\omega) = z_{im0} + \omega_i Sz_{m0}$, em que Sz_{m0} é o desvio padrão amostral da variável perturbada. Nesse caso, $\omega_0 = (0, 0, \ldots, 0)^\top$ e o logaritmo da função de verossimilhança do modelo perturbado é dado por

$$l(\theta|\omega) = l_1(\gamma_0, \gamma_1, \gamma_c|\omega) + l_2(\beta_1, \beta_2), \tag{4.16}$$

em que

$$
\begin{aligned}
l_1(\gamma_0,\gamma_1,\gamma_c|\omega) &= \sum_{i=1}^{n} l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic}|\omega),\\
l_2(\beta_1,\beta_2) &= \sum_{i:y_i\in(c_i,1)} l_{i2}(\mu_i,\phi_i),
\end{aligned}
\tag{4.17}
$$

com $l_{i2}(\mu_i,\phi_i)$ definido em (3.5) e $l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic}|\omega)$ tendo a mesma expressão de $l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})$ definido em (3.5), mas em $l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic}|\omega)$ os termos δ_{i0}, δ_{i1} e δ_{ic} são funções de ω_i. Para esse caso, a matriz $\Delta = (\Delta_{\gamma_0},\Delta_{\gamma_1},\Delta_{\gamma_c})^\top$ e prova-se no Apêndice C.2 que seus componentes são dados por

$$
\begin{aligned}
\Delta_{\gamma_0} &= Sz_{m0}(\widehat{\gamma}_{m0}Z_0^\top\widehat{\mathcal{M}}_1 + P_0\widehat{\mathcal{M}}_2),\\
\Delta_{\gamma_1} &= Sz_{m0}\widehat{\gamma}_{m0}Z_1^\top\widehat{\mathcal{M}}_3,\\
\Delta_{\gamma_c} &= Sz_{m0}\widehat{\gamma}_{m0}Z_c^\top\widehat{\mathcal{M}}_4,
\end{aligned}
\tag{4.18}
$$

em que $\mathcal{M}_1 = \text{diag}\{m_{11},\ldots,m_{n1}\}$, $\mathcal{M}_2 = \text{diag}\{m_{12},\ldots,m_{n2}\}$, $\mathcal{M}_3 = \text{diag}\{m_{13},\ldots,m_{n3}\}$, $\mathcal{M}_4 = \text{diag}\{m_{14},\ldots,m_{n4}\}$,

$$
\begin{aligned}
m_{i1} &= \frac{-I_{\{0\}}(y_i)}{\delta_{i0}^2}\left(\frac{\partial\delta_{i0}}{\partial\zeta_{i0}}\right)^2 + \frac{-I_{\{1\}}(y_i)}{\delta_{i1}^2}\left(\frac{\partial\delta_{i1}}{\partial\zeta_{i0}}\right)^2 + \frac{-I_{\{c_i\}}(y_i)}{\delta_{ic}^2}\left(\frac{\partial\delta_{ic}}{\partial\zeta_{i0}}\right)^2\\
&\quad + \frac{-I_{(c_i,1)}(y_i)}{(1-\alpha_i)^2}\left(\frac{\partial\delta_{i0}}{\partial\zeta_{i0}} + \frac{\partial\delta_{i1}}{\partial\zeta_{i0}} + \frac{\partial\delta_{ic}}{\partial\zeta_{i0}}\right)^2 - \Delta^*_{i00},\\
m_{i2} &= -a^*_{i0}\frac{\partial\delta_{i0}}{\partial\zeta_{i0}} - a^*_{i1}\frac{\partial\delta_{i1}}{\partial\zeta_{i0}} - a^*_{ic}\frac{\partial\delta_{ic}}{\partial\zeta_{i0}},\\
m_{i3} &= \frac{-I_{\{0\}}(y_i)}{\delta_{i0}^2}\frac{\partial\delta_{i0}}{\partial\zeta_{i0}}\frac{\partial\delta_{i0}}{\partial\zeta_{i1}} + \frac{-I_{\{1\}}(y_i)}{\delta_{i1}^2}\frac{\partial\delta_{i1}}{\partial\zeta_{i0}}\frac{\partial\delta_{i1}}{\partial\zeta_{i1}} + \frac{-I_{\{c_i\}}(y_i)}{\delta_{ic}^2}\frac{\partial\delta_{ic}}{\partial\zeta_{i0}}\frac{\partial\delta_{ic}}{\partial\zeta_{i1}}\\
&\quad + \frac{-I_{(c_i,1)}(y_i)}{(1-\alpha_i)^2}\left(\frac{\partial\delta_{i0}}{\partial\zeta_{i0}} + \frac{\partial\delta_{i1}}{\partial\zeta_{i0}} + \frac{\partial\delta_{ic}}{\partial\zeta_{i0}}\right)\left(\frac{\partial\delta_{i0}}{\partial\zeta_{i1}} + \frac{\partial\delta_{i1}}{\partial\zeta_{i1}} + \frac{\partial\delta_{ic}}{\partial\zeta_{i1}}\right) - \Delta^*_{i01},\\
m_{i4} &= \frac{-I_{\{0\}}(y_i)}{\delta_{i0}^2}\frac{\partial\delta_{i0}}{\partial\zeta_{i0}}\frac{\partial\delta_{i0}}{\partial\zeta_{ic}} + \frac{-I_{\{1\}}(y_i)}{\delta_{i1}^2}\frac{\partial\delta_{i1}}{\partial\zeta_{i0}}\frac{\partial\delta_{i1}}{\partial\zeta_{ic}} + \frac{-I_{\{c_i\}}(y_i)}{\delta_{ic}^2}\frac{\partial\delta_{ic}}{\partial\zeta_{i0}}\frac{\partial\delta_{ic}}{\partial\zeta_{ic}}\\
&\quad + \frac{-I_{(c_i,1)}(y_i)}{(1-\alpha_i)^2}\left(\frac{\partial\delta_{i0}}{\partial\zeta_{i0}} + \frac{\partial\delta_{i1}}{\partial\zeta_{i0}} + \frac{\partial\delta_{ic}}{\partial\zeta_{i0}}\right)\left(\frac{\partial\delta_{i0}}{\partial\zeta_{ic}} + \frac{\partial\delta_{i1}}{\partial\zeta_{ic}} + \frac{\partial\delta_{ic}}{\partial\zeta_{ic}}\right) - \Delta^*_{i0c},
\end{aligned}
$$

P_0 é uma matriz $p_0 \times n$ de zeros exceto a linha de posição m que é composta por uns e as demais matrizes e escalares estão definidos em (3.2), (3.11) e (4.11).

Perturbação de uma variável presente apenas em X_1

Admitamos que a variável que está na posição m de X_1 seja perturbada de forma que $x_{im1}(\omega) = x_{im1} + \omega_i Sx_{m1}$, em que Sx_{m1} é o desvio padrão amostral da variável perturbada. Nesse caso, $\omega_0 = (0,0,\ldots,0)^\top$ e o logaritmo da função de verossimilhança do modelo perturbado é dado por

$$
l(\theta|\omega) = l_1(\gamma_0,\gamma_1,\gamma_c) + l_2(\beta_1,\beta_2|\omega),
\tag{4.19}
$$

em que

$$\begin{aligned} l_1(\gamma_0,\gamma_1,\gamma_c) &= \sum_{i=1}^{n} l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic}), \\ l_2(\beta_1,\beta_2|\omega) &= \sum_{i:y_i\in(c_i,1)} l_{i2}(\mu_i,\phi_i|\omega), \end{aligned} \tag{4.20}$$

em que $l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})$ foi definido em (3.5) e $l_{i2}(\mu_i,\phi_i|\omega)$ tem a mesma expressão de $l_{i2}(\mu_i,\phi_i)$ definido em (3.5), mas em $l_{i2}(\mu_i,\phi_i|\omega)$ os termos μ_i e ϕ_i são funções de ω_i. Para esse caso, a matriz $\Delta = (\Delta_{\beta_1},\Delta_{\beta_2})^\top$ e prova-se no Apêndice C.2 que seus componentes são dados por

$$\begin{aligned} \Delta_{\beta_1} &= Sx_{m1}(\widehat{\beta}_{m1}X_1^\top\widehat{\mathcal{M}}_5 + P_\mu\widehat{\mathcal{M}}_6), \\ \Delta_{\beta_2} &= Sx_{m1}\widehat{\beta}_{m1}X_2^\top\widehat{\mathcal{M}}_7, \end{aligned} \tag{4.21}$$

em que $\mathcal{M}_5 = \text{diag}\{m_{15},\ldots,m_{n5}\}$, $\mathcal{M}_6 = \text{diag}\{m_{16},\ldots,m_{n6}\}$, $\mathcal{M}_7 = \text{diag}\{m_{17},\ldots,m_{n7}\}$,

$$m_{i5} = \begin{cases} -w_i\left(\dfrac{\partial\mu_i}{\partial\eta_{i1}}\right)^2 + \dfrac{\phi_i}{1-c_i}(y_i^*-\mu_i^*)\dfrac{\partial^2\mu_i}{\partial\eta_{i1}^2}, & \text{se } y_i\in(c_i,1) \\ 0, & \text{caso contrário} \end{cases}$$

$$m_{i6} = \begin{cases} \dfrac{\phi_i}{1-c_i}(y_i^*-\mu_i^*)\dfrac{\partial\mu_i}{\partial\eta_{i1}}, & \text{se } y_i\in(c_i,1) \\ 0, & \text{caso contrário} \end{cases}$$

$$m_{i7} = \begin{cases} \dfrac{1}{1-c_i}(y_i^*-\mu_i^*)\dfrac{\partial\mu_i}{\partial\eta_{i1}}\dfrac{\partial\phi_i}{\partial\eta_{i2}} + \dfrac{\phi_i}{1-c_i} \\ \left\{\left(\dfrac{1-\mu_i}{1-c_i}\right)\psi'\left[\left(\dfrac{1-\mu_i}{1-c_i}\right)\phi_i\right] - \left(\dfrac{\mu_i-c_i}{1-c_i}\right)\psi'\left[\left(\dfrac{\mu_i-c_i}{1-c_i}\right)\phi_i\right]\right\}\dfrac{\partial\mu_i}{\partial\eta_{i1}}\dfrac{\partial\phi_i}{\partial\eta_{i2}}, & \text{se } y_i\in(c_i,1) \\ 0, & \text{caso contrário} \end{cases}$$

w_i definido em (3.21) (não confundir com ω_i), P_μ uma matriz $p_\mu \times n$ de zeros exceto a linha de posição m que é composta por uns e as demais matrizes e escalares estão definidos em (3.2), (3.9) (3.10) e (3.11).

Perturbação de uma variável presente em Z_0 e X_1

Suponha agora que a variável preditora perturbada está presente tanto em Z_0 quanto em X_1 e que a perturbação efetuada foi semelhante às dos 2 casos anteriores. Assim, $\omega_0 = (0,0,\ldots,0)^\top$ e $\Delta = (\Delta_{\gamma_0},\Delta_{\gamma_1},\Delta_{\gamma_c},\Delta_{\beta_1},\Delta_{\beta_2})^\top$. Sem perda de generalidade, admita que a variável perturbada esteja na posição m tanto em Z_0 quanto em X_1. Nesse caso, pelo fato dos parâmetros do componente discreto do modelo serem separáveis dos parâmetros do componente contínuo do modelo, é fácil perceber que os componentes de Δ são iguais aos apresentados em (4.18) e (4.21).

4.3 Conclusão

Neste capítulo discutimos análise de diagnóstico para os modelos RBIZUT. Para avaliar o componente discreto do modelo, sugerimos o uso de um resíduo e de um teste de qualidade de ajuste existentes na literatura. Para estudar o ajuste do componente contínuo do modelo, propusemos uma modificação de um resíduo proposto por Ferrari et al. (2011). Discutimos e apresentamos um estudo de simulação para justificar por que dois resíduos que em princípio seriam mais interessantes que os sugeridos não devem ser usados para os modelos RBIZUT. Desenvolvemos também análise de influência local para o modelo proposto neste trabalho. Dois esquemas de perturbação foram considerados: ponderação de casos e perturbação individual de variáveis preditoras.

Capítulo 5

Aplicação

Neste capítulo, apresentamos uma aplicação do modelo RBIZUT para dados reais de cartão de crédito. Ajustamos um modelo RBIZUT, conduzimos testes de hipóteses de Wald, selecionamos um modelo final, realizamos análise de diagnóstico e apresentamos as conclusões obtidas.

5.1 Aplicação a dados reais de cartão de crédito

Um cliente de cartão de crédito recebe mensalmente uma fatura que traz, entre outras informações, o valor total da fatura e o valor de pagamento mínimo. Ele pode optar por pagar qualquer valor entre o total e o mínimo, incluindo os 2 extremos. O cliente paga juros na fatura do mês seguinte em função do valor financiado, que é a diferença entre o valor total e o valor pago. Se dividirmos o valor pago pelo cliente pelo valor total da fatura, obtemos uma variável denominada Valor relativo do pagamento (VRP), conforme mencionado no Capítulo 1. O VRP tem probabilidade positiva de assumir os valores 0 e 1, porque muitos clientes não tem dinheiro suficiente para pagar nem mesmo o valor do pagamento mínimo na data de vencimento da fatura e muitos clientes pagam o valor total para evitar cobrança de juros. O VRP tem probabilidade positiva de assumir um terceiro valor que é dado pela razão entre o valor do pagamento mínimo e o valor total. Esse valor, c_i na notação que estamos utilizando, varia no intervalo $(0, 1)$ e não é fixo para todas as observações, já que o valor mínimo de pagamento como proporção do valor total da fatura não é constante para todos os clientes de cartão de crédito de uma instituição financeira. A variável pode ainda assumir qualquer valor no intervalo $(c_i, 1)$ e não pode assumir valor no intervalo $(0, c_i)$. Assim, o VRP é uma distribuição com parte discreta e parte contínua, sendo discreta nos pontos 0, 1 e c_i e contínua no intervalo $(c_i, 1)$. Dessa forma, em princípio, parece razoável imaginar que, dadas as variáveis que caracterizam o cliente e seu comportamento de uso do cartão de crédito, o VRP tenha distribuição BIZUT.

Nesta seção ajustamos um modelo RBIZUT para o VRP no cartão de crédito a partir de dados reais fornecidos por uma instituição financeira. A amostra disponível contém 5000 cartões de crédito de clientes não correntistas da instituição financeira com limite maior ou igual a 500 reais, fatura de no mínimo 100 reais e que tinham o cartão há pelo menos 3 meses em novembro de 2007 e não estavam inadimplentes. O valor c é variável com média 0,191, desvio padrão 0,134, mínimo 0,05, primeiro quartil e mediana 0,15 e terceiro quartil 0,2. Algumas poucas observações (0,03%) tinham valor de c igual a 1. Como o número de observações com essa característica era bem pequeno e

o modelo RBIZUT admite apenas valores de c no intervalo aberto $(0, 1)$, consideramos no ajuste do modelo $c = 0,9999$ para esses 14 clientes. Devido à necessidade da instituição financeira de manter a confidencialidade de algumas informações, a amostra foi gerada de forma que a proporção de clientes em cada um dos grupos (não pagou, pagou o mínimo, pagou o valor total e pagou um valor entre o mínimo e total) fosse diferente na amostra e na população.

A variável resposta do estudo é o VRP em novembro de 2007. Na amostra, 4,5% das observações assumiram o valor 0 (não pagamento), 10,2% assumiram o valor c_i (pagamento mínimo), 63,0% assumiram o valor 1 (pagamento total) e 22,3% assumiram valor no intervalo $(c_i, 1)$ (pagamento de valor maior que o mínimo e menor que o total). A Tabela 5.1 e a Figura 5.1 apresentam, respectivamente, medidas descritivas e histograma do VRP em novembro de 2007 para as observações com valor da variável no intervalo $(c_i, 1)$. Pode-se notar que, na amostra, o componente contínuo da variável resposta tem média 0,535, desvio padrão 0,274, mínimo 0,105 e máximo 0,999. Nota-se ainda que o componente contínuo do VRP parece ter distribuição assimétrica e uma concentração relativamente alta em valores próximos de 1. Como o suporte da variável resposta não é o mesmo para todas as observações, para uma melhor visualização da frequência de cartões com VRP próximo a c_i, para as observações com VRP no intervalo $(c_i, 1)$, realizamos também a análise descritiva de uma transformação da variável resposta denominada Valor relativo do pagamento padronizado (VRPP) e definida como $y_i^R = (y_i - c_i)/(1 - c_i)$ se $y_i \in (c_i, 1)$ e 0 caso contrário (Tabela 5.1 e Figura 5.2). Na amostra, o componente contínuo do VRPP tem média 0,418, desvio padrão 0,337, mínimo 0,001 e máximo 0,999. Pode-se notar ainda que há um grande número de observações próximas de 0 e 1 e, assim, o componente contínuo da variável resposta parece possuir uma alta concentração em valores bem próximos de c_i e 1.

Tabela 5.1: Medidas descritivas para o componente contínuo das variáveis VRP e VRPP.

Variável	Média	D. padrão	Mínimo	Q_1	Mediana	Q_3	Máximo
VRP	0,535	0,274	0,105	0,282	0,500	0,769	0,999
VRPP	0,418	0,337	0,001	0,096	0,364	0,706	0,999

As variáveis preditoras incluem informações do cliente (sexo, idade e estado do Brasil em que reside), informações do cartão de crédito e de seu uso (percentual de utilização do limite, fez ou não compra parcelada com juros, tempo que o cliente tem o cartão) e o VRP do cartão de crédito em outubro de 2007. Foram feitas diversas tabelas e gráficos não apresentados aqui relacionando as variáveis preditoras e funções da variável resposta para estudar a melhor forma de inserção de cada uma das variáveis independentes em cada uma das equações do modelo. A variável VRP do cartão de crédito em outubro de 2007, por exemplo, foi segmentada em 4 variáveis: indicadora de pagamento em atraso (VRP = 0 em outubro de 2007), indicadora de pagamento mínimo (VRP = c_i em outubro de 2007), indicadora de pagamento total (VRP = 1 em outubro de 2007) e VRPP em outubro de 2007. Pelo fato da análise descritiva para as variáveis idade do cliente e percentual de utilização do limite sugerir possíveis efeitos quadráticos dessas variáveis em pelo

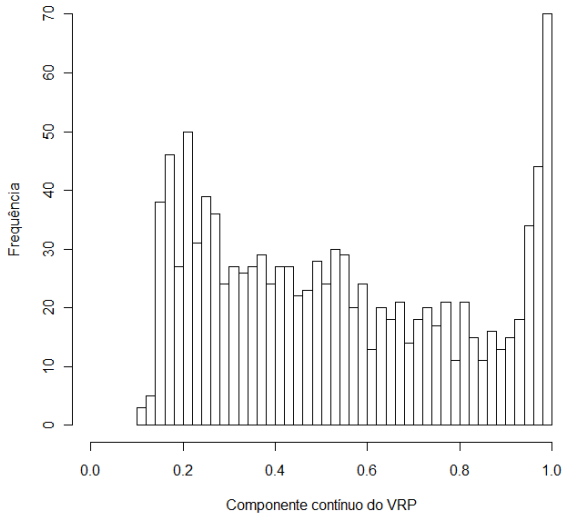

Figura 5.1: Histograma para o componente contínuo da variável VRP.

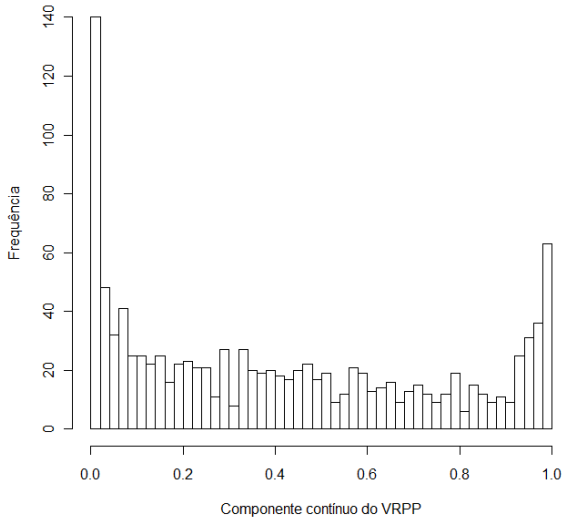

Figura 5.2: Histograma para o componente contínuo da variável VRPP.

menos uma das equações do modelo, essas variáveis foram incluídas no modelo de forma centralizada (valor observado menos a média amostral) para evitar problemas de multicolinearidade. As tabelas a seguir apresentam estatísticas descritivas para cada uma das variáveis preditoras, sendo que a Tabela 5.2 traz as variáveis quantitativas e a Tabela 5.3 traz as qualitativas e o VRP de Outubro de 2007 segmentado em 4 grupos. Pode-se notar que, na amostra, os clientes têm em média 43,4 anos, são predominantemente do sexo masculino (55,8%) e 18,6% deles residem nas regiões Norte ou Nordeste. Pode-se observar também que poucos clientes fazem compras parceladas com juros

(4,0%) e as médias do percentual de utilização do limite e do tempo que o cliente tem o cartão (em meses) são respectivamente 32,7 e 51,1. Conforme esperado, as distibuições do VRP e do VRPP em outubro de 2007 não parecem ser muito diferentes das mesmas em novembro de 2007.

Tabela 5.2: Medidas descritivas para as variáveis preditoras quantitativas do modelo para a variável VRP.

Variável	Média	D. padrão	Mínimo	Q_1	Mediana	Q_3	Máximo
Idade do cliente	43,4	12,3	18,0	34,0	42,0	52,0	105,0
Perc. de utiliz. do limite	32,7	29,7	0,4	10,6	22,5	45,7	247,6
Tempo com o cartão (meses)	51,1	48,0	3,0	16,0	36,0	73,0	382,0
VRPP em out/2007*	0,441	0,322	0,001	0,152	0,386	0,719	0,999

* Apenas para as observações tais que o VRP em Outubro de 2007 pertence ao intervalo $(c_i, 1)$.

Tabela 5.3: Distribuição de frequências para as variáveis preditoras qualitativas do modelo para a variável VRP.

Variável	Nível	Freq. abs.	Freq. rel. (%)
Sexo	Feminino	2211	44,2
	Masculino	2789	55,8
Região	Norte/Nordeste	931	18,6
	Outras	4069	81,4
Fez compras	Não	4799	96,0
parceladas com juros	Sim	201	4,0
Faixa de VRP	Pagamento em atraso	259	5,2
em out/2007	Pagamento mínimo	499	10,0
	Pagamento total	3194	63,9
	Pagamento entre o mínimo e o máximo	1048	21,0

O modelo ajustado foi o RLBIZUT conforme definido na Seção 3.1. As variáveis foram selecionadas para o modelo final da forma descrita na Seção 3.5. Em cada equação do modelo consideramos apenas duas possibilidades para as variáveis indicadoras referentes ao VRP de outubro de 2007: as três presentes ou todas elas ausentes. Como a instituição financeira considera o risco do cliente para a definição de c, preferiu-se não incluir a mesma como variável preditora para evitar problemas de endogeneidade, conforme discutido na Seção 3.3.2. De acordo com o argumentado na mesma seção, a não inclusão de c como variável preditora não deve prejudicar consideravelmente a qualidade do ajuste, já que para 81,3% das observações c está no intervalo [0,05, 0,20].

Obteve-se o modelo final e realizou-se análise de diagnóstico conforme discutido no Capítulo 4. A Figura 5.3 apresenta os gráficos do resíduo $r_i^{\{c\}}$ descritos na Seção 4.1 para o componente contínuo do modelo final obtido para o VRP. Pode-se notar que o envelope sugere que o ajuste do modelo é insatisfatório, já que há um número bem grande de pontos fora dos limites determinados pelo envelope. Pela Figura 5.2 pode-se observar que há um elevado número de observações da variável resposta bem próximas aos valores c_i e 1. Assim, uma possível explicação para a obtenção de ajuste insatisfatório pode ser devido ao VRP possuir uma concentração em valores próximos de c_i e 1

maior do que é esperada pelo modelo RLBIZUT.

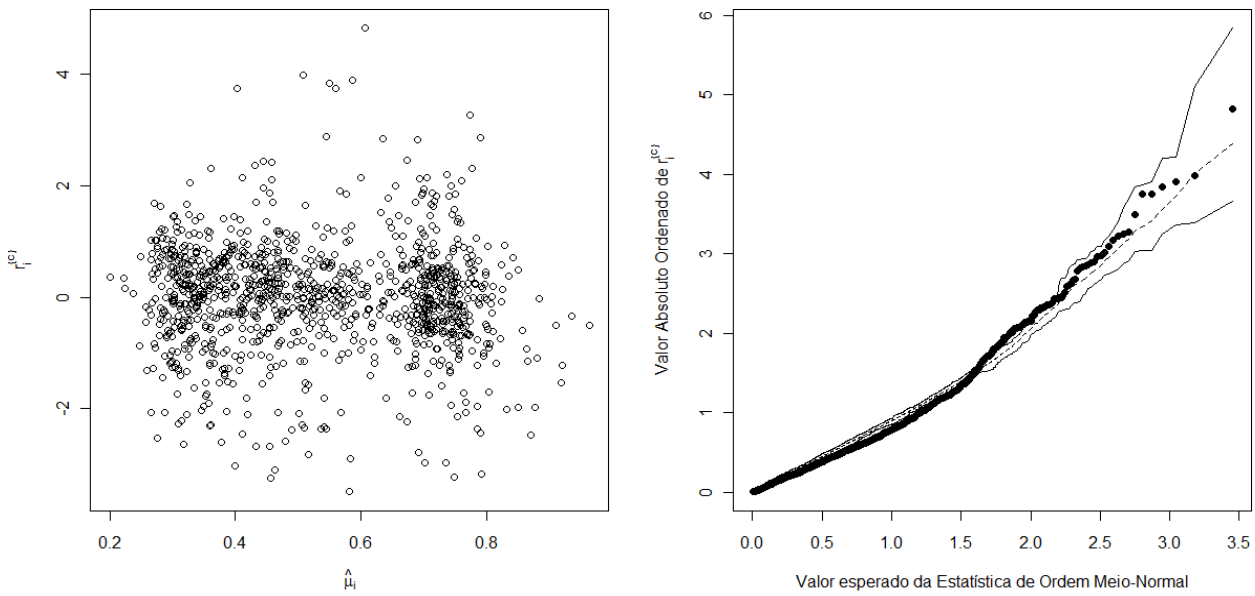

Figura 5.3: Gráficos de resíduos para o componente contínuo do modelo RLBIZUT para a variável VRP.

Para a instituição financeira pouca diferença faz se, dado um valor ϵ pequeno, $y_i = c_i$ ou $c_i < y_i \leq c_i + \epsilon$. Também é equivalente para a instituição se $y_i = 1$ ou $1 - \epsilon \leq y_i < 1$. O motivo é que, em ambas as situações a diferença de juros pago pelo cliente é bastante pequena. Em virtude disso, criamos uma transformação do VRP que ao mesmo tempo pode ser melhor ajustada pelo modelo RLBIZUT e é tão útil para a instituição financeira quanto o VRP. Denominamos a variável de intenção de valor de pagamento relativo (IVRP). Ela é definida como a razão entre a intenção de valor de pagamento e o valor da fatura, em que o primeiro é definido da seguinte forma:

- Se o valor do pagamento for inteiro e a diferença entre o valor do pagamento e o valor do pagamento mínimo for menor que 5 reais, então a intenção de valor de pagamento é igual ao valor do pagamento mínimo.

- Se o valor do pagamento for inteiro e a diferença entre o valor do pagamento e o valor da fatura for menor que 5 reais, então a intenção de valor de pagamento é igual ao valor da fatura.

- Se a diferença entre o valor do pagamento e o valor do pagamento mínimo for igual a 0,01 ou 0,03 ou 0,1 ou 0,3, então a intenção de valor de pagamento é igual ao valor do pagamento mínimo.

- Se a diferença entre o valor do pagamento e o valor da fatura for igual a 0,01 ou 0,03 ou 0,1 ou 0,3, então a intenção de valor de pagamento é igual ao valor da fatura.

Nos 2 primeiros casos, é possível que a intenção do cliente era pagar o pagamento mínimo ou total, mas por conveniência (facilidade de troco ou de digitação) ele pagou um valor inteiro. Nos 2 casos seguintes, também é razoável esperar que a intenção do cliente era pagar o pagamento mínimo ou total, mas por erro de digitação ele pagou um valor próximo disso. Nesse caso, foram escolhidas as diferenças 0,01, 0,03, 0,1 e 0,3 por questão de posicionamento dos números em um teclado comum.

A Figura 5.4 apresenta os gráficos de resíduos para o componente contínuo do modelo final obtido para o IVRP. Os gráficos sugerem um ajuste satisfatório para a parte contínua do modelo. No gráfico da esquerda os resíduos parecem se distribuir sem nenhum padrão, conforme é esperado quando o modelo está bem ajustado. Além disso, poucas observações têm resíduo, em módulo, superior a 3 e nenhuma tem resíduo superior a 4. O número de observações com resíduo, em módulo, superior a 3 parece estar de acordo com o que era esperado em termos de frequência e magnitude ao observarmos o gráfico à direita, já que os valores de todos esses resíduos estão no interior do envelope construído.

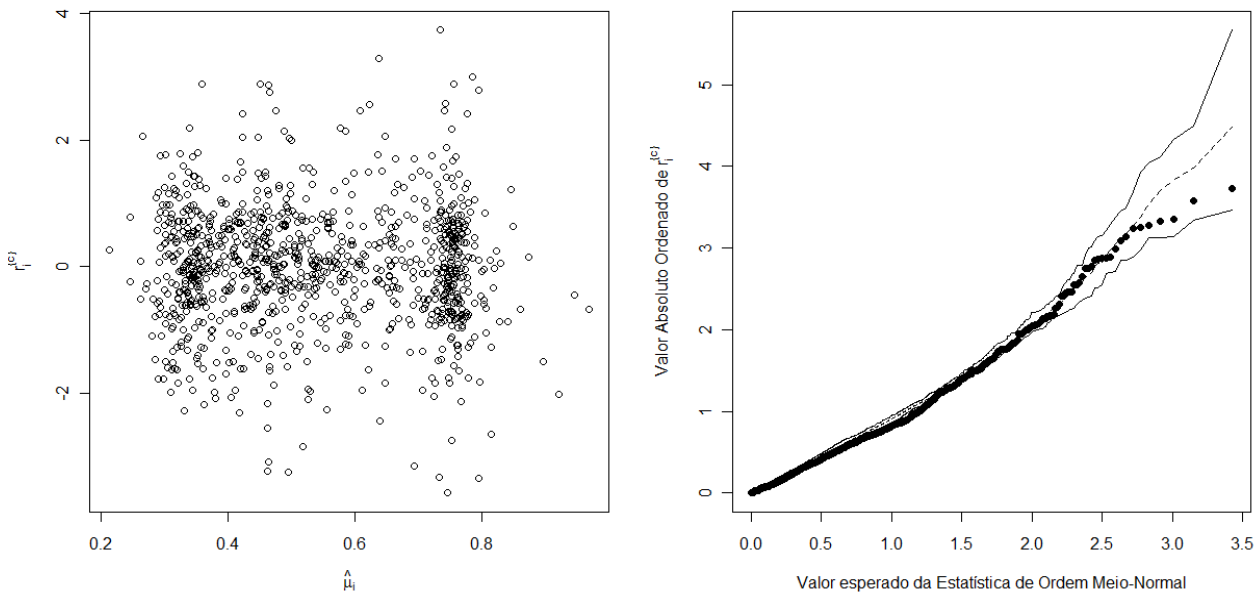

Figura 5.4: Gráficos de resíduos para o componente contínuo do modelo RLBIZUT para a variável IVRP.

As Figuras 5.5 e 5.6 apresentam os gráficos de resíduos para o componente discreto e o parâmetro de mistura do modelo final obtido para o IVRP. Os gráficos dos componentes $r_i^{\{d0\}}$, $r_i^{\{d1\}}$ e $r_i^{\{dc\}}$ da Figura 5.5 estão de acordo com o que geralmente é observado em modelos de regressão logística multinomial bem ajustados e nenhuma observação apresenta $r_i^{\{d0\}}$, $r_i^{\{d1\}}$ e $r_i^{\{dc\}}$ muito maior que as demais. Muitas observações apresentam valores bem grandes de $r_i^{\{d0\}}$, $r_i^{\{d1\}}$ ou $r_i^{\{dc\}}$. No entanto, o número de observações com essas características e os seus respectivos valores parecem estar de acordo com o que é esperado, já que os maiores valores de $r_i^{\{d0\}}$, $r_i^{\{d1\}}$ e $r_i^{\{dc\}}$ estão todos no interior do envelope apresentado na Figura 5.6. Para o $r_i^{\{dq\}}$ há uma observação com valor bem maior que

as demais. No entanto, isso parece não ser um problema, já que o valor obtido está bem distante dos limites determinados pelo envelope inferior direito da Figura 5.6.

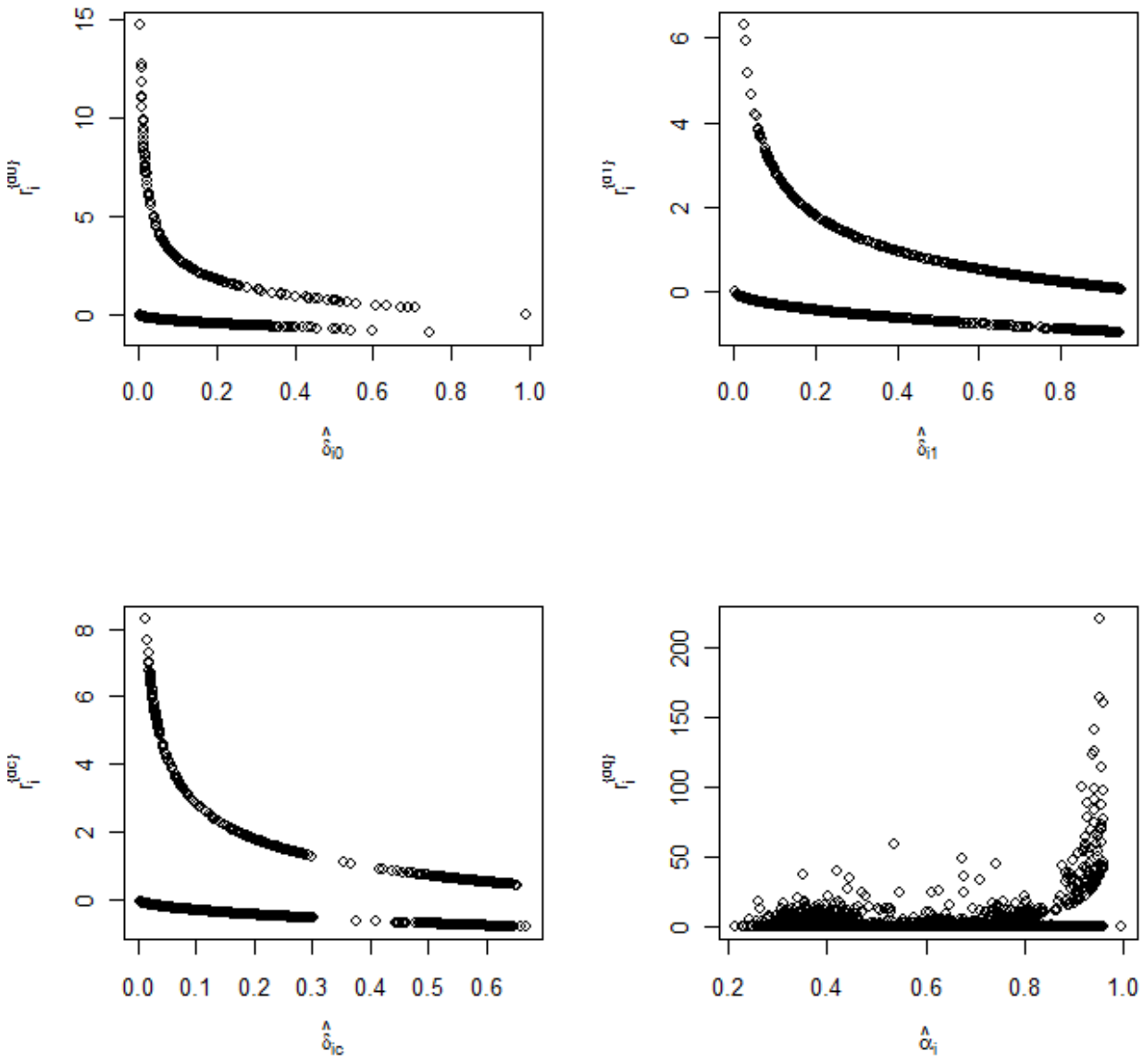

Figura 5.5: Diagramas de dispersão dos resíduos para o componente discreto e o parâmetro de mistura do modelo RLBIZUT para a variável IVRP.

Conforme discutido na Seção 4.1, para avaliar o ajuste do componente discreto do modelo e o parâmetro de mistura é mais interessante a realização de um teste de qualidade de ajuste do que a análise de resíduos. A Tabela 5.4 traz, para cada grupo, os valores observados e esperados supondo que o componente discreto do modelo e o parâmetro de mistura estejam bem ajustados, conforme descrito na Seção 4.1. Pode-se notar que as frequências estimadas e observadas estão próximas entre

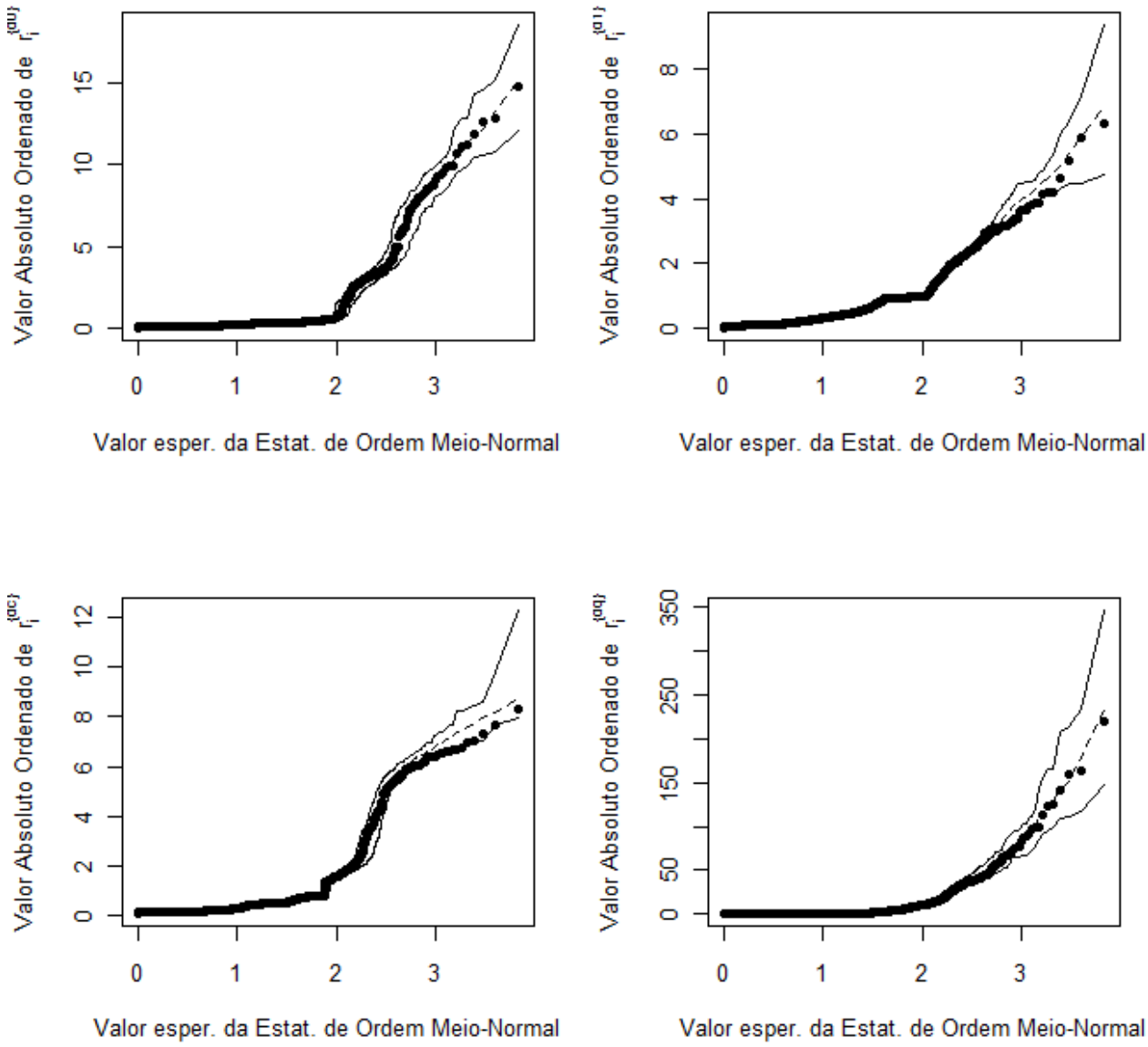

Figura 5.6: Gráfico de probabilidade meio-normal com envelope simulado para os resíduos para o componente discreto e o parâmetro de mistura do modelo RLBIZUT para a variável IVRP.

si. Além disso, o valor-p do teste de qualidade de ajuste é 0,101. A Tabela 5.4 e o valor-p sugerem que a parte discreta do modelo e que o parâmetro de mistura também foram ajustados de forma satisfatória.

A Tabela 5.5 apresenta os resultados do modelo RLBIZUT final para a variável intenção de pagamento relativo. Para todas as variáveis os sinais das estimativas dos parâmetros ficaram de acordo com o que era esperado. Como foram usadas funções de ligações logarítmica e logito, a relação entre as variáveis presentes no modelo e o IVRP de novembro de 2007 podem ser melhor

Tabela 5.4: Valores observados e esperados em cada decil de $1 - \hat{\alpha}_i$ no modelo final para a variável IVRP.

Grupo	$y = 0$		$y = c_i$		$y \in (c_i; 1)$		$y = 1$	
	Obser.	Esper.	Obser.	Esper.	Obser.	Esper.	Obser.	Esper.
1	6	5,2	6	9,6	19	21,7	469	463,5
2	2	4,3	7	10,4	25	23,5	466	461,9
3	4	4,5	7	11,5	27	25,7	462	458,3
4	9	5,2	18	13,7	35	29,5	438	451,7
5	8	6,5	15	14,7	41	35,6	436	443,2
6	11	8,7	26	20,3	46	46,6	417	424,4
7	32	41,1	196	193,9	85	91,7	187	173,3
8	69	64,7	147	150,0	135	154,0	149	131,4
9	39	46,5	72	74,7	287	259,1	102	119,6
10	44	37,3	86	81,1	314	326,8	56	54,8

interpretadas a partir de exponenciais das estimativas dos parâmetros (apresentadas na última coluna da Tabela 5.5) ou de exponenciais de funções das estimativas dos parâmetros.

O IVRPP em outubro de 2007 (definido de forma análoga ao VRPP), bem como as variáveis indicadoras relacionadas com o IVRP em outubro de 2007 parecem ser bastante importantes no ajuste do IVRP em novembro de 2007. Essas variáveis são as únicas presentes nas 5 equações do modelo e as únicas presentes na equação que ajusta o parâmetro de precisão do componente contínuo da distribuição BIZUT. Pode-se notar ainda que os valores-p dos testes de Wald associados a essas variáveis são, em geral, bem pequenos. Observando-se os sinais das estimativas dos parâmetros, conclui-se que, conforme esperado, quanto maior a intenção de pagamento relativo padronizado em outubro de 2007, maior é a média da intenção de pagamento relativo em novembro de 2007, quando a intenção do cliente é pagar um valor entre o pagamento mínimo e o máximo. Observa-se também que quanto maior o IVRPP em outubro de 2007, menor é a chance do cliente ter a intenção de não pagar nada ou de efetuar o pagamento mínimo e maior é a chance dele ter a intenção de fazer o pagamento total em novembro de 2011 (todas essas chances em relação à intenção de pagar um valor maior que o mínimo e menor que o máximo). Estima-se, por exemplo, que a cada aumento de 0,1 na intenção de pagamento relativo padronizado em outubro de 2011, a chance do cliente ter a intenção de realizar o pagamento total em novembro de 2011 (em relação à intenção de realizar pagamento entre o mínimo e o total) aumenta 31% ($\exp(0,1 \times 2,6845) - 1$), mantidas as demais variáveis constantes. Pode-se notar ainda que clientes que apresentaram um determinado padrão de intenção de pagamento em Outubro de 2011 têm maior chance de ter esse padrão em Novembro de 2011 do que clientes que tinham uma intenção de pagamento entre o valor mínimo e o total em Outubro de 2011. Por exemplo, estima-se que a chance de um cliente ter a intenção de pagar o valor mínimo em Novembro de 2011 (em relação à intenção de pagar entre o mínimo e total) é 477% maior entre os clientes que tiveram a intenção de pagar o valor mínimo em Outubro de 2011 do que entre aqueles que tiveram a intenção de pagar um valor pouco superior ao mínimo, mantidas as demais

Tabela 5.5: Modelo RLBIZUT final para a variável IVRP em Nov/2007.

Equação	Variável	Estimativa	E.P.(1)	Valor-p	Exp(2)
μ	Intercepto	-1,4129	0,0800	< 0,0001	0,2434
	Pagamento em atraso Out/07	0,9032	0,1373	< 0,0001	2,4675
	Inten. de pagamento mínimo Out/07	0,1024	0,1378	0,4573	1,1078
	Inten. de pagamento total Out/07	2,2708	0,1170	< 0,0001	9,6871
	IVRPP Out/07	2,4304	0,1540	< 0,0001	11,3634
	Perc. de utiliz. do limite centraliz. (linear)	-0,0038	0,0011	0,0006	0,9962
ϕ	Intercepto	1,3403	0,0895	< 0,0001	3,8202
	Pagamento em atraso Out/07	-0,5496	0,1555	0,0004	0,5772
	Inten. de pagamento mínimo Out/07	-0,2556	0,1602	0,1107	0,7745
	Inten. de pagamento total Out/07	-0,7487	0,1247	< 0,0001	0,4730
	IVRPP Out/07	-0,7432	0,1699	< 0,0001	0,4756
δ_0	Intercepto	-1,6505	0,2365	< 0,0001	0,1920
	Tempo com o cartão (meses)	-0,0082	0,0024	0,0006	0,9918
	Pagamento em atraso Out/07	1,2042	0,2634	< 0,0001	3,3341
	Inten. de pagamento mínimo Out/07	0,9276	0,2681	0,0005	2,5284
	Inten. de pagamento total Out/07	0,1907	0,2797	0,4954	1,2101
	IVRPP Out/07	-0,6756	0,4549	0,1375	0,5089
	Idade do cliente centralizado	-0,0121	0,0063	0,0534	0,9880
	Perc. de utiliz. do limite centraliz. (linear)	-0,0067	0,0042	0,1048	0,9933
	Perc. de utiliz. do limite centraliz. (quadrático)	0,0002	0,0000	0,0004	1,0002
δ_c	Intercepto	-0,5795	0,1566	0,0002	0,5602
	Tempo com o cartão (meses)	-0,0039	0,0013	0,0033	0,9961
	Pagamento em atraso Out/07	0,3440	0,2183	0,1150	1,4106
	Inten. de pagamento mínimo Out/07	1,7519	0,1805	< 0,0001	5,7655
	Inten. de pagamento total Out/07	0,0094	0,1882	0,9600	1,0094
	IVRPP Out/07	-1,8960	0,3807	< 0,0001	0,1502
	Região Norte/Nordeste	-0,3367	0,1360	0,0133	0,7141
δ_1	Intercepto	-2,1369	0,1954	< 0,0001	0,1180
	Fez compras parceladas com juros	-0,6699	0,2022	0,0009	0,5118
	Pagamento em atraso Out/07	1,6873	0,2749	< 0,0001	5,4049
	Inten. de pagamento mínimo Out/07	1,8081	0,2534	< 0,0001	6,0988
	Inten. de pagamento total Out/07	4,7385	0,2084	< 0,0001	114,2627
	IVRPP Out/07	2,6845	0,2886	< 0,0001	14,6509
	Perc. de utiliz. do limite centraliz. (linear)	-0,0165	0,0017	< 0,0001	0,9836
	Região Norte/Nordeste	-0,4532	0,1125	0,0001	0,6356

(1) Erro padrão; (2) Exponencial da estimativa.

variáveis preditoras constantes. O parâmetro de dispersão do componente contínuo do IVRP em Novembro de 2011 diminui à medida que aumenta a intenção de pagamento relativo padronizada em Outubro de 2011 e é maior para clientes que tiveram a intenção de pagar um valor pouco superior ao mínimo do que para os demais. É importante ressaltar que, na interpretação dos parâmetros das variáveis indicadoras relacionadas ao IVRP de Outubro de 2011, a comparação deve ser feita sempre em relação aos clientes que tiveram a intenção de pagar um valor pouco superior ao mínimo em Outubro de 2011. Isso deve ser feito porque a interpretação sempre é feita considerando-se que todas as demais variáveis são mantidas constantes. Duas observações só tem os valores de todas as variáveis iguais, com exceção de uma das indicadoras relacionadas ao IVRP de Outubro de 2011, se o IVRPP de Outubro de 2011 for igual a 0. Como isso não é possível para clientes que tiveram a intenção de pagar valor entre o mínimo e o máximo, consideramos que o IVRPP é um valor bem próximo de 0, que ocorre quando a intenção de pagar for um pouco superior ao valor mínimo.

As demais variáveis estão presentes em no máximo 3 equações do modelo. O percentual de utilização do limite é a única variável presente na equação para a média do componente contínuo do IVRP de novembro de 2011 e a única presente em pelo menos 3 equações que não está relacionada com o IVRP de outubro de 2011. Estima-se que a cada aumento de um ponto percentual na utilização do limite do cartão de crédito, a chance do componente contínuo do IVRP diminui em 0,38%, mantidas as demais variáveis preditoras constantes. Assim, se o cliente decidir pagar um valor maior que o mínimo e menor que o máximo da fatura, quanto maior o percentual de utilização do cartão de crédito, menor tende a ser o percentual da fatura pago. Conclui-se ainda que quanto maior o percentual de utilização do limite, menor é a chance do cliente ter a intenção de pagar o valor máximo e que o percentual de utilização do limite tem um efeito quadrático na chance dele ter a intenção de pagar o valor mínimo (nos dois casos em relação a ele ter intenção de pagar um valor entre o mínimo e o máximo). Portanto, quanto mais o cliente gasta com o cartão (em relação ao seu limite) mais provável que ele não tenha intenção de pagar o valor total e se ele decidir não pagar o valor total, maior é em média, o valor que ele deixa para pagar no mês seguinte.

O tempo que o cliente tem o cartão e a variável indicadora de cliente das regiões norte e nordeste estão presentes em duas equações do componente discreto do modelo. Quanto maior o tempo que o cliente tem o cartão, menor é a chance de ele ter a intenção de não pagar nenhum valor ou pagar o valor mínimo (em relação a ele ter a intenção de pagar um valor entre o mínimo e o máximo). Assim, o risco de crédito associado ao cliente e a receita da instituição com juros pagos por ele diminui à medida que aumenta o tempo de experiência do cliente com o cartão de crédito. Estima-se, por exemplo, que a chance do cliente ter a intenção de não pagar a fatura (em relação a ele ter a intenção de pagar um valor entre o mínimo e o máximo), diminui 0,82% a cada aumento de um mês no tempo que o cliente possui o cartão, mantidas as demais variáveis preditoras constantes. Já os clientes da regiões Norte e Nordeste do Brasil têm menor chance de ter a intenção de pagar o valor mínimo ou o máximo (em relação a ele ter a intenção de pagar um valor entre o mínimo e o máximo) do que os clientes das demais regiões do país.

As variáveis idade do cliente e indicadora de pagamento parcelado com juros estão presentes

em apenas uma equação do componente discreto do modelo. Estima-se que a cada aumento de 1 ano de idade, a chance do cliente ter a intenção de não pagar a fatura (em relação a ele ter a intenção de pagar um valor entre o mínimo e o máximo), diminui 1,2%, mantidas as demais variáveis preditoras constantes. Estima-se ainda que clientes que fizeram compras parceladas com juros têm uma chance de pagar o valor total da fatura (em relação a eles terem a intenção de pagar um valor entre o mínimo e o máximo) 49% menor do que clientes que não fizeram compras parceladas com juros. Portanto, quanto mais velho o cliente menor o risco de crédito associado a ele e clientes que aceitam pagar juros em suas compras parceladas têm uma maior probabilidade de decidir pagar juros também no cartão de crédito.

A única variável analisada que não está presente em nenhum modelo é o sexo do cliente. Dessa forma, não há indícios de que a distribuição da intenção de valor de pagamento relativo varie em função do sexo.

5.2 Conclusão

Neste capítulo ajustamos um modelo RBIZUT para a variável VRP a partir de um banco de dados reais fornecido por uma instituição financeira. O VRP do cartão de crédito é bastante útil para administradoras de cartões de crédito, já que ele permite obter periodicamente medidas de risco e retorno associadas a cada um dos seus clientes. Para isso, a administradora de cartão de crédito deve inicialmente ajustar um modelo RBIZUT para o VRP usando uma amostra de seus clientes em determinado mês, de forma semelhante ao que fizemos neste capítulo. Em seguida, utilizando-se as estimativas dos parâmetros do modelo obtidas para essa amostra de clientes, a instituição financeira pode estimar mensalmente os parâmetros da distribuição BIZUT para cada um dos seus clientes. As variáveis usadas como preditoras no modelo estão disponíveis 10 dias antes da data do vencimento da fatura. Portanto, a administradora de cartão de crédito pode estimar, com 10 dias de antecedência ao vencimento, medidas de risco e retorno associadas a cada um dos seus clientes. Por exemplo, $\hat{\delta}_{i0}$ é o estimador de uma medida de risco, porque é um estimador da probabilidade do cliente não pagar nenhum valor da fatura. Por outro lado, $[(1-\hat{\alpha}_i)(1-\hat{\mu}_i)+\hat{\delta}_{ic}(1-c_i)]$(valor total da fatura i) é o estimador de uma medida de retorno, porque é um estimador da esperança do valor financiado e o cliente paga juros em função dessa parte do valor da fatura que ele deixou de pagar em um dado mês. A instituição pode ainda ajustar modelos não para obter medidas de risco e retorno para cada cliente associadas ao pagamento que ele realizará em 10 dias, mas para obter essas mesmas estimativas associadas ao pagamento que será realizado em 6 meses, 12 meses ou qualquer outro prazo de interesse.

A administradora de cartão de crédito pode ainda ajustar modelos RBIZUT para o VRP para auxiliar na seleção de possíveis clientes. Muitas vezes, a instituição possui grandes bancos de dados de potenciais clientes e deve selecionar aqueles que vão receber uma oferta de cartão de crédito. Para alguns tipos de produto, o cliente que sempre paga a fatura total não é muito rentável, em virtude dos programas em que o cliente troca pontos obtidos em função do gasto no cartão de crédito por inúmeros diferentes produtos ou serviços. Assim, a instituição financeira pode ajustar modelos RBIZUT para escolher clientes que ao mesmo tempo tenham uma alta probabilidade de não se

tornarem inadimplentes e de não pagarem o valor total de suas faturas. Para isso, a administradora de cartão de crédito deve utilizar uma amostra de clientes que adquiriram o cartão de crédito da instituição e estavam presentes nesse banco de dados de potenciais clientes.

Para o banco de dados disponível não foi possível obter um bom ajuste para o VRP. No entanto, pôde-se ajustar adequadamente uma transformação dessa variável que é igualmente útil para a instituição financeira. Além disso, é possível que o próprio VRP possa ser bem ajustado para outros segmentos de cartão de crédito da instituição financeira que forneceu os dados ou para dados de outras administradoras de cartão de crédito. Assim, acredita-se que a introdução do modelo RBIZUT possa ser bastante útil na área de cartão de crédito.

Capítulo 6

Conclusão

O objetivo deste trabalho foi desenvolver aspectos inferenciais e de diagnóstico para ajustar variáveis do tipo proporção que não podem assumir valores no intervalo $(0, c)$ e assumem o valor c com probabilidade positiva. Variáveis com essas características podem ser encontradas na área de econometria quando o objetivo é estudar algum tipo de pagamento limitado entre dois valores em relação ao seu valor máximo. Alguns exemplos dessas variáveis são o pagamento do cartão de crédito em relação ao valor total da fatura e o valor recebido como seguro desemprego em relação ao máximo valor do benefício.

Para ajustar essas variáveis, introduzimos inicialmente a distribuição beta inflacionada truncada (BIZUT), que é uma mistura de uma distribuição beta com suporte no intervalo $(c, 1)$ e uma distribuição trinomial que assume os valores zero, um e c. Obtivemos os momentos da distribuição, o vetor escore e a matriz de informação de Fisher e provamos que ela pertence à família exponencial de posto completo. Ainda discutimos estimação dos seus parâmetros por máxima verossimilhança e pelo método dos momentos, conduzimos estudos de simulação de Monte Carlo e apresentamos uma aplicação a dados de seguro desemprego.

Em seguida introduzimos os modelos de regressão beta inflacionados truncados (RBIZUT) para ajustar variáveis que, dadas as variáveis preditoras, têm distribuição BIZUT. O modelo proposto admite que todos os parâmetros da distribuição podem variar em função de variáveis preditoras. Além disso, o modelo permite que o parâmetro conhecido c varie entre as unidades populacionais. Obtivemos formas fechadas para a função escore e a matriz de informação de Fisher, discutimos estimação dos parâmetros por máxima verossimilhança e apresentamos expressões para intervalos de confiança e testes de hipóteses. Obtivemos ainda alguns resultados para os modelos RBIZUT com c variável, conduzimos estudos de simulação de Monte Carlo e discutimos seleção de modelos e endogeneidade.

Discutimos também aspetos de diagnóstico para os modelos RBIZUT. Inicialmente sugerimos o uso de alguns resíduos e um teste de qualidade de ajuste existentes na literatura para avaliar a qualidade do ajuste de um modelo RBIZUT. Em seguida, desenvolvemos análise de influência local para essa classe de modelos.

Para finalizar o trabalho, apresentamos uma aplicação dos modelos RBIZUT a dados reais de cartão de crédito e discutimos como os modelos RBIZUT podem ser úteis para instituições que

administram esse produto financeiro. Além das instituições financeiras, a administração federal de diversos países também pode se beneficiar da introdução dos modelos RBIZUT. Previdência pública e seguro desemprego são dois benefícios importantes que governos federais costumam conceder, mas que ao mesmo tempo representam gastos públicos relevantes. Dessa forma, estudar esses benefícios a partir dos modelos RBIZUT pode ser interessante tanto para um melhor entendimento desses gastos pelo governo como para auxiliar na previsão de valores gastos com esses benefícios. Assim, esperamos que este trabalho possa motivar trabalhos aplicados na área de econometria e também trabalhos na área de Estatística como os que são citados a seguir.

6.1 Trabalhos futuros

Dentre os estudos relacionados a este trabalho que podem ser desenvolvidos, destacamos:

- Extensão dos modelos RBIZUT para dados longitudinais utilizando equações de estimação generalizadas ou modelos mistos. As variáveis citadas neste trabalho que possivelmente podem ser ajustadas utilizando os modelos RBIZUT são geralmente medidas mensalmente e portanto é razoável o ajuste dessas variáveis através de equações de estimação generalizadas ou modelos mistos.

- Desenvolvimento de modelos RBIZUT não lineares. O objetivo neste caso seria estender o modelo proposto neste trabalho para obter melhores ajustes para variáveis de interesse.

- Desenvolvimento para os modelos RBIZUT de um resíduo que possa ser utilizado na análise de diagnóstico tanto do componente contínuo do modelo quanto do componente discreto e de um resíduo para o componente contínuo do modelo que leve em consideração um estimador de $\mathrm{P}(y_i \in (c_i, 1))$. O desenvolvimento desses resíduos permitiria uma melhor análise de diagnóstico em modelos RBIZUT.

- Desenvolvimento de estimadores para os parâmetros do modelo RBIZUT para situações em que uma ou mais variáveis do modelo são endógenas. Esse desenvolvimento permitiria incluir c como variável preditora do modelo em todas as situações em que o valor c varia entre as unidades populacionais.

Apêndice A

Provas dos resultados do Capítulo 2

Neste apêndice são apresentadas as provas de alguns resultados apresentados no Capítulo 2.

A.1 Prova dos resultados dos momentos da distribuição beta

O objetivo é provar que

$$\mathrm{E}(Y^r) = \mu_r = \mathrm{E}[(Y-a)^r] + \sum_{i=1}^{r}(-1)^{i+1}\binom{r}{i}a^i\mu_{r-i},$$

em que $\mathrm{E}[(Y-a)^r] = (b-a)^r\prod_{i=1}^{r}\left(\frac{p+r-i}{\phi+r-i}\right)$ e $p = \frac{(\mu-a)\phi}{(b-a)}$, com $\mu_0 = 1$ e $\mu_1 = E(Y) = \mu$.

Tem-se que

$$\mathrm{E}[(Y-a)^r] = E(Y^r) + \sum_{i=1}^{r}(-1)^{i}\binom{r}{i}a^i\mu_{r-i}.$$

Logo,

$$\mu_r = \mathrm{E}[(Y-a)^r] + \sum_{i=1}^{r}(-1)^{i+1}\binom{r}{i}a^i\mu_{r-i}.$$

Utilizando-se a parametrização da distribuição beta dada em (2.1),

$$\mathrm{E}[(Y-a)^r] = \int_a^b (y-a)^r\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\frac{(y-a)^{p-1}(b-y)^{q-1}}{(b-a)^{p+q-1}}dy$$

$$= \frac{(b-a)^r\Gamma(p+q)\Gamma(p+r)}{\Gamma(p)\Gamma(p+q+r)}\int_a^b\frac{\Gamma(p+r+q)}{\Gamma(p+r)\Gamma(q)}\frac{(y-a)^{p+r-1}(b-y)^{q-1}}{(b-a)^{p+r+q-1}}dy = (b-a)^r\prod_{i=1}^{r}\left(\frac{p+r-i}{\phi+r-i}\right).$$

A variância da distribuição beta é obtida utilizando-se (2.3). Assim

$$\mathrm{E}(Y^2) = \frac{(b-a)^2p(p+1)}{(p+q)(p+q+1)} + \frac{2a(b-a)p}{p+q} + a^2 \text{ e } [\mathrm{E}(Y)]^2 = \frac{(b-a)^2p^2}{(p+q)^2} + \frac{2a(b-a)p}{p+q} + a^2$$

e, portanto,

$$\mathrm{Var}(Y) = \frac{(b-a)^2pq}{(p+q)^2(p+q+1)} = \frac{(\mu-a)(b-\mu)}{1+\phi}.$$

A.2 Prova dos resultados dos momentos da distribuição BIZUT

Os resultados dos momentos da distribuição BIZUT são facilmente provados utilizando-se esperança e variância condicional e assim

$$\mathrm{E}(Y^r) = \mathrm{E}[\mathrm{E}(Y^r|I_{\{0,c,1\}}(Y))] = c^r\delta_c + \delta_1 + (1-\alpha)\mu_r.$$

Tem-se também que

$$V_1 = \mathrm{Var}(Y|I_{\{0,c,1\}}(Y)=1) = \mathrm{E}(Y^2|I_{\{0,c,1\}}(Y)=1) - [\mathrm{E}(Y|I_{\{0,c,1\}}(Y)=1)]^2$$

$$= (c^2\delta_c + \delta_1)/\alpha - [(c\delta_c + \delta_1)/\alpha]^2 = (c^2\delta_c(\alpha-\delta_c) + \delta_1(\alpha-\delta_1) - 2c\delta_c\delta_1)/\alpha^2$$

e utilizando-se (2.8) e (2.9) que

$$\mathrm{E}[\mathrm{Var}(Y|I_{\{0,c,1\}}(Y))] = \alpha V_1 + (1-\alpha)V_2 \quad \text{e}$$

$$\mathrm{Var}[\mathrm{E}(Y|I_{\{0,c,1\}}(Y))] = [(c\delta_c + \delta_1)/\alpha]^2\alpha + \mu^2(1-\alpha) - [c\delta_c + \delta_1 + \mu(1-\alpha)]^2$$

$$= (c\delta_c + \delta_1)^2(1/\alpha - 1) + \mu^2(1-\alpha)\alpha - 2\mu(c\delta_c + \delta_1)(1-\alpha) = (1-\alpha)\left[\frac{c\delta_c + \delta_1}{\sqrt{\alpha}} - \sqrt{\alpha}\mu\right]^2.$$

Como $\mathrm{Var}(Y) = \mathrm{E}[\mathrm{Var}(Y|I_{\{0,c,1\}}(Y))] + \mathrm{Var}[\mathrm{E}(Y|I_{\{0,c,1\}}(Y))]$, então

$$\mathrm{Var}(Y) = \alpha V_1 + (1-\alpha)V_2 + (1-\alpha)\left[\frac{c\delta_c + \delta_1}{\sqrt{\alpha}} - \sqrt{\alpha}\mu\right]^2.$$

A.3 Matriz de informação de Fisher da distribuição BIZUT

Para a obtenção dos componentes da matriz de informação de Fisher é necessário o cálculo de $\mathrm{E}(t_k(y))$, $k = 1,2,3,4,5$. As estatísticas $t_1(y)$, $t_2(y)$ e $t_3(y)$ têm distribuição de Bernoulli e portanto $\mathrm{E}(t_1(y)) = \delta_0$, $\mathrm{E}(t_2(y)) = \delta_1$ e $\mathrm{E}(t_3(y)) = \delta_c$. Utilizando-se que, na família exponencial de posto completo, $\mathrm{E}[t_k(y)] = \partial A(\eta)/\partial\eta_k$, lembrando-se que $\psi(j) = \Gamma'(j)/\Gamma(j)$ e usando-se a definição de $C(\theta)$ apresentada em (2.12), tem-se que

$$\begin{aligned}
\mathrm{E}[t_4(y)] =& \frac{\partial A(\eta)}{\partial\eta_4} = -\psi(\eta_4+\eta_5) + \frac{\Gamma'(\eta_4)\Gamma(\eta_5)(1-c)^{\eta_4+\eta_5-1} + (e^{\eta_1}+e^{\eta_2}+e^{\eta_3})\Gamma'(\eta_4+\eta_5)}{\Gamma(\eta_4)\Gamma(\eta_5)(1-c)^{\eta_4+\eta_5-1} + (e^{\eta_1}+e^{\eta_2}+e^{\eta_3})\Gamma(\eta_4+\eta_5)} + \\
& \frac{\Gamma(\eta_4)\Gamma(\eta_5)(1-c)^{\eta_4+\eta_5-1}\log(1-c)}{\Gamma(\eta_4)\Gamma(\eta_5)(1-c)^{\eta_4+\eta_5-1} + (e^{\eta_1}+e^{\eta_2}+e^{\eta_3})\Gamma(\eta_4+\eta_5)} \\
=& -\psi(\phi) + \frac{\Gamma'\left[\left(\frac{\mu-c}{1-c}\right)\phi\right]\Gamma\left[\left(\frac{1-\mu}{1-c}\right)\phi\right](1-c)^{\phi-1} + \frac{\alpha\Gamma'(\phi)}{C(\theta)}}{\Gamma\left[\left(\frac{\mu-c}{1-c}\right)\phi\right]\Gamma\left[\left(\frac{1-\mu}{1-c}\right)\phi\right](1-c)^{\phi-1} + \frac{\alpha\Gamma(\phi)}{C(\theta)}} + \\
& \frac{\Gamma\left[\left(\frac{\mu-c}{1-c}\right)\phi\right]\Gamma\left[\left(\frac{1-\mu}{1-c}\right)\phi\right](1-c)^{\phi-1}\log(1-c)}{\Gamma\left[\left(\frac{\mu-c}{1-c}\right)\phi\right]\Gamma\left[\left(\frac{1-\mu}{1-c}\right)\phi\right](1-c)^{\phi-1} + \frac{\alpha\Gamma(\phi)}{C(\theta)}}
\end{aligned}$$

$$
\begin{aligned}
&= -\psi(\phi) + \frac{(1-\alpha)\Gamma(\phi)\psi\left[\left(\frac{\mu-c}{1-c}\right)\phi\right] + (1-\alpha)\Gamma(\phi)\log(1-c) + \alpha\Gamma'(\phi)}{\Gamma(\phi)} \\
&=(1-\alpha)\left\{\psi\left[\left(\frac{\mu-c}{1-c}\right)\phi\right] - \psi(\phi) + \log(1-c)\right\} \quad \text{e} \\
\mathrm{E}[t_5(y)] =& \frac{\partial A(\eta)}{\partial \eta_4} = -\psi(\eta_4+\eta_5) + \frac{\Gamma(\eta_4)\Gamma'(\eta_5)(1-c)^{\eta_4+\eta_5-1} + (e^{\eta_1}+e^{\eta_2}+e^{\eta_3})\Gamma'(\eta_4+\eta_5)}{\Gamma(\eta_4)\Gamma(\eta_5)(1-c)^{\eta_4+\eta_5-1} + (e^{\eta_1}+e^{\eta_2}+e^{\eta_3})\Gamma(\eta_4+\eta_5)} + \\
& \frac{\Gamma(\eta_4)\Gamma(\eta_5)(1-c)^{\eta_4+\eta_5-1}\log(1-c)}{\Gamma(\eta_4)\Gamma(\eta_5)(1-c)^{\eta_4+\eta_5-1} + (e^{\eta_1}+e^{\eta_2}+e^{\eta_3})\Gamma(\eta_4+\eta_5)} \\
&= -\psi(\phi) + \frac{\Gamma\left[\left(\frac{\mu-c}{1-c}\right)\phi\right]\Gamma'\left[\left(\frac{1-\mu}{1-c}\right)\phi\right](1-c)^{\phi-1} + \frac{\alpha\Gamma'(\phi)}{C(\theta)}}{\Gamma\left[\left(\frac{\mu-c}{1-c}\right)\phi\right]\Gamma\left[\left(\frac{1-\mu}{1-c}\right)\phi\right](1-c)^{\phi-1} + \frac{\alpha\Gamma(\phi)}{C(\theta)}} + \\
& \frac{\Gamma\left[\left(\frac{\mu-c}{1-c}\right)\phi\right]\Gamma\left[\left(\frac{1-\mu}{1-c}\right)\phi\right](1-c)^{\phi-1}\log(1-c)}{\Gamma\left[\left(\frac{\mu-c}{1-c}\right)\phi\right]\Gamma\left[\left(\frac{1-\mu}{1-c}\right)\phi\right](1-c)^{\phi-1} + \frac{\alpha\Gamma(\phi)}{C(\theta)}} \\
&= -\psi(\phi) + \frac{(1-\alpha)\Gamma(\phi)\psi\left[\left(\frac{1-\mu}{1-c}\right)\phi\right] + (1-\alpha)\Gamma(\phi)\log(1-c) + \alpha\Gamma'(\phi)}{\Gamma(\phi)} \\
&=(1-\alpha)\left\{\psi\left[\left(\frac{1-\mu}{1-c}\right)\phi\right] - \psi(\phi) + \log(1-c)\right\}.
\end{aligned}
$$

Os componentes da matriz de informação de Fisher são assim obtidos a partir da equação $k_{\theta_i\theta_j} = -\mathrm{E}[\partial^2 \log[f(y;\mu,\delta_0,\delta_1,\delta_c,\mu,\phi,c)]/\partial\theta_i\theta_j]$. A primeira derivada em relação a cada um dos parâmetros pode ser obtida considerando-se (2.15) com uma única observação e, fazendo-se a segunda derivada, obtém-se as expressões apresentadas em (2.18).

A.4 Distribuição assintótica do estimador de máxima verossimilhança de $\mathbf{E}(Y)$

Como $\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{D} N_5(0, K(\theta)^{-1})$ e $\mathrm{E}(Y) = c\delta_c + \delta_1 + (1-\alpha)\mu$ é uma função diferenciável de θ, então pelo método delta (Lehmann e Casella, 1998, página 61) $\sqrt{n}(\widehat{\mathrm{E}(Y)} - \mathrm{E}(Y)) \xrightarrow{D} N(0,v)$ em que $v = B^\top K(\theta)^{-1}B$ e $B = (\partial\mathrm{E}(Y)/\partial\delta_0, \partial\mathrm{E}(Y)/\partial\delta_1, \partial\mathrm{E}(Y)/\partial\delta_c, \partial\mathrm{E}(Y)/\partial\mu, \partial\mathrm{E}(Y)/\partial\phi)^\top = (-\mu, 1-\mu, c-\mu, 1-\alpha, 0)^\top$. Sendo $k^{\theta_i\theta_j}$ o termo de posição (i,j) de $K(\theta)^{-1}$ então

$$
\begin{aligned}
v &= B^\top \begin{pmatrix} k^{\delta_0\delta_0} & k^{\delta_0\delta_1} & k^{\delta_0\delta_c} & 0 & 0 \\ k^{\delta_0\delta_1} & k^{\delta_1\delta_1} & k^{\delta_1\delta_c} & 0 & 0 \\ k^{\delta_0\delta_c} & k^{\delta_1\delta_c} & k^{\delta_c\delta_c} & 0 & 0 \\ 0 & 0 & 0 & k^{\mu\mu} & k^{\mu\phi} \\ 0 & 0 & 0 & k^{\mu\phi} & k^{\phi\phi} \end{pmatrix} B \\
&= \mu^2 k^{\delta_0\delta_0} + (1-\mu)^2 k^{\delta_1\delta_1} + (c-\mu)^2 k^{\delta_c\delta_c} + (1-\alpha)^2 k^{\mu\mu} - 2\mu(1-\mu)k^{\delta_0\delta_1} - 2\mu(c-\mu)k^{\delta_0\delta_c} + \\
&\quad 2(1-\mu)(c-\mu)k^{\delta_1\delta_c}
\end{aligned}
$$

Assim, para amostras grandes, um intervalo de confiança aproximado para $\mathrm{E}(Y)$ com coeficiente de confiança de $(100 \times \gamma)\%$ é dado por $\mathrm{IC}(\mathrm{E}(Y); \gamma) = \widehat{\mathrm{E}(Y)} \pm z_{(1+\gamma)/2}(\hat{v}/n)^{1/2}$, em que $\hat{v}$ é obtido substituindo-se em v os parâmetros θ_k por seus respectivos estimadores $\hat{\theta}_k$.

Apêndice B

Provas dos resultados do Capítulo 3

Neste apêndice são apresentadas as provas de alguns resultados apresentados no Capítulo 3 e as tabelas dos estudos de simulação realizados no mesmo capítulo.

B.1 Momentos de y_i^* e $y_i^\bullet$

A distribuição BIZUT pertence à família exponencial de posto completo conforme foi provado no Capítulo 2 e portanto sua função densidade de probabilidade pode ser escrita na forma $f(y;\eta) = \exp[\eta^\top T(y) - A(\eta)]h(y)$, em que $\eta = (\eta_1, \eta_2, \eta_3, \eta_4, \eta_5)^\top$, $T(y) = (t_1(y), t_2(y), t_3(y), t_4(y), t_5(y))^\top$. Em uma das formas de escrever a função densidade de probabilidade da distribuição BIZUT tem-se que $\eta_4 = \mu\phi/(1-c)$, $\eta_5 = \phi/(1-c)$, $t_4(y) = y_i^*$ e $t_5(y) = y_i^\bullet$ e $A(\eta) = \log\Gamma[\eta_4 - c\eta_5] + \log\Gamma[\eta_5 - \eta_4] - \log\Gamma[(1-c)\eta_5] + [(1-c)\eta_5 - 1]\log(1-c)$. Pelas propriedades da família exponencial de posto completo (Lehmann e Casella, 1998, página 27), $\mathrm{E}[t_i(y)] = \partial A(\eta)/\partial\eta_i$ e $\mathrm{Cov}[t_i(y), t_j(y)] = \partial^2 A(\eta)/\partial\eta_i\partial\eta_j$. Usando esses resultados, obtém-se facilmente os primeiros momentos de y_i^* e $y_i^\bullet$.

B.2 Obtenção da função escore para o modelo RBIZUT

O termo do vetor escore (3.6) associado à γ_{j0}, parâmetro de posição j do vetor γ_0, é obtido derivando-se o logaritmo da função de verossimilhança (3.4) em relação à γ_{j0}. Assim, esse termo é dado por

$$\begin{aligned}
\frac{\partial l(\theta)}{\partial \gamma_{j0}} &= \frac{\partial l_1(\gamma_0, \gamma_1, \gamma_c)}{\partial \gamma_{j0}} = \sum_{i=1}^{n} \frac{\partial l_{i1}(\delta_{i0}, \delta_{i1}, \delta_{ic})}{\partial \gamma_{j0}} \\
&= \sum_{i=1}^{n} \left\{ \frac{\partial l_{i1}(\delta_{i0}, \delta_{i1}, \delta_{ic})}{\partial \delta_{i0}} \frac{\partial \delta_{i0}}{\partial \zeta_{i0}} \frac{\partial \zeta_{i0}}{\partial \gamma_{j0}} + \frac{\partial l_{i1}(\delta_{i0}, \delta_{i1}, \delta_{ic})}{\partial \delta_{i1}} \frac{\partial \delta_{i1}}{\partial \zeta_{i0}} \frac{\partial \zeta_{i0}}{\partial \gamma_{j0}} + \frac{\partial l_{i1}(\delta_{i0}, \delta_{i1}, \delta_{ic})}{\partial \delta_{ic}} \frac{\partial \delta_{ic}}{\partial \zeta_{i0}} \frac{\partial \zeta_{i0}}{\partial \gamma_{j0}} \right\} \\
&= \sum_{i=1}^{n} \left\{ \left[\frac{I_{\{0\}}(y_i)}{\delta_{i0}} - \frac{I_{(c_i,1)}(y_i)}{1-\alpha_i} \right] \frac{\partial \delta_{i0}}{\partial \zeta_{i0}} z_{ij0} \right\} + \sum_{i=1}^{n} \left\{ \left[\frac{I_{\{1\}}(y_i)}{\delta_{i1}} - \frac{I_{(c_i,1)}(y_i)}{1-\alpha_i} \right] \frac{\partial \delta_{i1}}{\partial \zeta_{i0}} z_{ij0} \right\} + \\
&\quad \sum_{i=1}^{n} \left\{ \left[\frac{I_{\{c_i\}}(y_i)}{\delta_{ic}} - \frac{I_{(c_i,1)}(y_i)}{1-\alpha_i} \right] \frac{\partial \delta_{ic}}{\partial \zeta_{i0}} z_{ij0} \right\}
\end{aligned}$$

e

$$U_{\gamma_0}(\gamma_0, \gamma_1, \gamma_c) = \left(\frac{\partial l(\theta)}{\partial \gamma_{10}}, \ldots, \frac{\partial l(\theta)}{\partial \gamma_{p_0 0}} \right)^\top$$

e pode ser escrito na forma matricial como na equação (3.11). Equivalentemente,

$$\begin{aligned}
\frac{\partial l(\theta)}{\partial \gamma_{j1}} &= \frac{\partial l_1(\gamma_0,\gamma_1,\gamma_c)}{\partial \gamma_{j1}} = \sum_{i=1}^{n} \frac{\partial l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})}{\partial \gamma_{j1}} \\
&= \sum_{i=1}^{n} \left\{ \frac{\partial l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})}{\partial \delta_{i0}} \frac{\partial \delta_{i0}}{\partial \zeta_{i1}} \frac{\partial \zeta_{i1}}{\partial \gamma_{j1}} + \frac{\partial l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})}{\partial \delta_{i1}} \frac{\partial \delta_{i1}}{\partial \zeta_{i1}} \frac{\partial \zeta_{i1}}{\partial \gamma_{j1}} + \frac{\partial l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})}{\partial \delta_{ic}} \frac{\partial \delta_{ic}}{\partial \zeta_{i1}} \frac{\partial \zeta_{i1}}{\partial \gamma_{j1}} \right\} \\
&= \sum_{i=1}^{n} \left\{ \left[\frac{I_{\{0\}}(y_i)}{\delta_{i0}} - \frac{I_{(c_i,1)}(y_i)}{1-\alpha_i} \right] \frac{\partial \delta_{i0}}{\partial \zeta_{i1}} z_{ij1} \right\} + \sum_{i=1}^{n} \left\{ \left[\frac{I_{\{1\}}(y_i)}{\delta_{i1}} - \frac{I_{(c_i,1)}(y_i)}{1-\alpha_i} \right] \frac{\partial \delta_{i1}}{\partial \zeta_{i1}} z_{ij1} \right\} + \\
&\quad \sum_{i=1}^{n} \left\{ \left[\frac{I_{\{c_i\}}(y_i)}{\delta_{ic}} - \frac{I_{(c_i,1)}(y_i)}{1-\alpha_i} \right] \frac{\partial \delta_{ic}}{\partial \zeta_{i1}} z_{ij1} \right\},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial l(\theta)}{\partial \gamma_{jc}} &= \frac{\partial l_1(\gamma_0,\gamma_1,\gamma_c)}{\partial \gamma_{jc}} = \sum_{i=1}^{n} \frac{\partial l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})}{\partial \gamma_{jc}} \\
&= \sum_{i=1}^{n} \left\{ \frac{\partial l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})}{\partial \delta_{i0}} \frac{\partial \delta_{i0}}{\partial \zeta_{ic}} \frac{\partial \zeta_{ic}}{\partial \gamma_{jc}} + \frac{\partial l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})}{\partial \delta_{i1}} \frac{\partial \delta_{i1}}{\partial \zeta_{ic}} \frac{\partial \zeta_{ic}}{\partial \gamma_{jc}} + \frac{\partial l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})}{\partial \delta_{ic}} \frac{\partial \delta_{ic}}{\partial \zeta_{ic}} \frac{\partial \zeta_{ic}}{\partial \gamma_{jc}} \right\} \\
&= \sum_{i=1}^{n} \left\{ \left[\frac{I_{\{0\}}(y_i)}{\delta_{i0}} - \frac{I_{(c_i,1)}(y_i)}{1-\alpha_i} \right] \frac{\partial \delta_{i0}}{\partial \zeta_{ic}} z_{ijc} \right\} + \sum_{i=1}^{n} \left\{ \left[\frac{I_{\{1\}}(y_i)}{\delta_{i1}} - \frac{I_{(c_i,1)}(y_i)}{1-\alpha_i} \right] \frac{\partial \delta_{i1}}{\partial \zeta_{ic}} z_{ijc} \right\} + \\
&\quad \sum_{i=1}^{n} \left\{ \left[\frac{I_{\{c_i\}}(y_i)}{\delta_{ic}} - \frac{I_{(c_i,1)}(y_i)}{1-\alpha_i} \right] \frac{\partial \delta_{ic}}{\partial \zeta_{ic}} z_{ijc} \right\}
\end{aligned}$$

e os vetores

$$U_{\gamma_1}(\gamma_0,\gamma_1,\gamma_c) = \left(\frac{\partial l(\theta)}{\partial \gamma_{11}}, \ldots, \frac{\partial l(\theta)}{\partial \gamma_{p_1 1}} \right)^{\top} \text{ e } U_{\gamma_c}(\gamma_0,\gamma_1,\gamma_c) = \left(\frac{\partial l(\theta)}{\partial \gamma_{1c}}, \ldots, \frac{\partial l(\theta)}{\partial \gamma_{p_c c}} \right)^{\top}$$

podem ser escritos na forma matricial conforme apresentado em (3.11).

Para os casos em que a função de ligação H é a mesma dos modelos RLBIZUT, os componentes de $U_D(\gamma_0,\gamma_1,\gamma_c)$ simplificam-se consideravelmente. Nesse caso,

$$\delta_{i0} = \frac{e^{\zeta_{i0}}}{1+e^{\zeta_{i0}}+e^{\zeta_{i1}}+e^{\zeta_{ic}}}, \delta_{i1} = \frac{e^{\zeta_{i1}}}{1+e^{\zeta_{i0}}+e^{\zeta_{i1}}+e^{\zeta_{ic}}} \text{ e } \delta_{ic} = \frac{e^{\zeta_{ic}}}{1+e^{\zeta_{i0}}+e^{\zeta_{i1}}+e^{\zeta_{ic}}}$$

e assim

$$\frac{\partial \delta_{i0}}{\partial \zeta_{i0}} = \frac{e^{\zeta_{i0}}(1+e^{\zeta_{i1}}+e^{\zeta_{ic}})}{(1+e^{\zeta_{i0}}+e^{\zeta_{i1}}+e^{\zeta_{ic}})^2}, \frac{\partial \delta_{i1}}{\partial \zeta_{i0}} = \frac{-e^{\zeta_{i1}+\zeta_{0i}}}{(1+e^{\zeta_{i0}}+e^{\zeta_{i1}}+e^{\zeta_{ic}})^2} \text{ e } \frac{\partial \delta_{ic}}{\partial \zeta_{i0}} = \frac{-e^{\zeta_{ic}+\zeta_{0i}}}{(1+e^{\zeta_{i0}}+e^{\zeta_{i1}}+e^{\zeta_{ic}})^2}.$$

Definindo $f_\zeta(\zeta_i) = 1 + e^{\zeta_{i0}} + e^{\zeta_{i1}} + e^{\zeta_{ic}}$ temos que

$$\begin{aligned}
\frac{\partial l(\theta)}{\partial \gamma_{j0}} &= \sum_{i=1}^{n} \left\{ [I_{\{0\}}(y_i) - e^{\zeta_{i0}} I_{(c_i,1)}(y_i)] \left[\frac{1 + e^{\zeta_{i1}} + e^{\zeta_{ic}}}{f_\zeta(\zeta_i)} \right] z_{ij0} \right\} \\
&+ \sum_{i=1}^{n} \left\{ [I_{\{1\}}(y_i) - e^{\zeta_{i1}} I_{(c_i,1)}(y_i)] \left[\frac{-e^{\zeta_{i0}}}{f_\zeta(\zeta_i)} \right] z_{ij0} \right\} \\
&+ \sum_{i=1}^{n} \left\{ [I_{\{c_i\}}(y_i) - e^{\zeta_{ic}} I_{(c_i,1)}(y_i)] \left[\frac{-e^{\zeta_{i0}}}{f_\zeta(\zeta_i)} \right] z_{ij0} \right\} \\
&= \sum_{i=1}^{n} \left\{ \left[\frac{z_{ij0}}{f_\zeta(\zeta_i)} \right] [I_{\{0\}}(y_i)(1 + e^{\zeta_{i1}} + e^{\zeta_{ic}}) - e^{\zeta_{i0}}(1 - I_{\{0\}}(y_i))] \right\} \\
&= \sum_{i=1}^{n} \left\{ z_{ij0} I_{\{0\}}(y_i) - \frac{e^{\zeta_{i0}} z_{ij0}}{f_\zeta(\zeta_i)} \right\} \\
&= \sum_{i=1}^{n} \left\{ z_{ij0}(I_{\{0\}}(y_i) - \delta_{i0}) \right\}.
\end{aligned}$$

Logo, $U_{\gamma_0}(\gamma_0, \gamma_1, \gamma_c)$ pode ser escrito como na equação (3.14). De forma análoga, obtêm-se $U_{\gamma_1}(\gamma_0, \gamma_1, \gamma_c)$ e $U_{\gamma_c}(\gamma_0, \gamma_1, \gamma_c)$ para esta particular função de ligação H.

Para a obtenção de $U_{\beta_1}(\beta_1, \beta_2)$ e $U_{\beta_2}(\beta_1, \beta_2)$ utilizamos y_i^* e $y_i^\bullet$ conforme definidos em (3.8) e (3.9), respectivamente, e suas esperanças condicionais definidas em (3.10). Assim,

$$\begin{aligned}
\frac{\partial l(\theta)}{\partial \beta_{j1}} &= \frac{\partial l_2(\beta_1, \beta_2)}{\partial \beta_{j1}} = \sum_{i:y_i \in (c_i,1)} \frac{\partial l_{i2}(\mu_i, \phi_i)}{\partial \beta_{j1}} = \sum_{i:y_i \in (c_i,1)} \left\{ \frac{\partial l_{i2}(\mu_i, \phi_i)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_{i1}} \frac{\partial \eta_{i1}}{\partial \beta_{j1}} \right\} \\
&= \sum_{i:y_i \in (c_i,1)} \frac{\phi_i}{1 - c_i} \left\{ y_i^* + \psi \left[\left(\frac{1 - \mu_i}{1 - c_i} \right) \phi_i \right] - \psi \left[\left(\frac{\mu_i - c_i}{1 - c_i} \right) \phi_i \right] \right\} \frac{\partial \mu_i}{\partial \eta_{i1}} x_{ij1} \\
&= \sum_{i=1}^{n} \frac{\phi_i}{1 - c_1} I_{(c_i,1)}(y_i)(y_i^* - \mu_i^*) \frac{\partial \mu_i}{\partial \eta_{i1}} x_{ij1}
\end{aligned}$$

e, portanto,

$$U_{\beta_1}(\beta_1, \beta_2) = \left(\frac{\partial l(\theta)}{\partial \beta_{11}}, \ldots, \frac{\partial l(\theta)}{\partial \beta_{p_\mu 1}} \right)^\top$$

pode ser escrito conforme apresentado em (3.11). De forma semelhante,

$$\begin{aligned}
\frac{\partial l(\theta)}{\partial \beta_{j2}} &= \frac{\partial l_2(\beta_1,\beta_2)}{\partial \beta_{j2}} = \sum_{i:y_i\in(c_i,1)} \frac{\partial l_{i2}(\mu_i,\phi_i)}{\partial \beta_{j2}} = \sum_{i:y_i\in(c_i,1)} \left\{ \frac{\partial l_{i2}(\mu_i,\phi_i)}{\partial \phi_i} \frac{\partial \phi_i}{\partial \eta_{i2}} \frac{\partial \eta_{i2}}{\partial \beta_{j2}} \right\} \\
&= \sum_{i:y_i\in(c_i,1)} \left\{ \left(\frac{\mu_i}{1-c_i}\right)(y_i^* - \mu_i^*) + \left(\frac{1}{1-c_i}\right) y_i^\bullet + \psi(\phi_i) + \left(\frac{c_i}{1-c_i}\right)\psi\left[\left(\frac{\mu_i - c_i}{1-c_i}\right)\phi_i\right] - \right. \\
&\quad \left. \left(\frac{1}{1-c_i}\right)\psi\left[\left(\frac{1-\mu_i}{1-c_i}\right)\phi_i\right] - \log(1-c_i)\right\} \frac{\partial \phi_i}{\partial \eta_{i2}} x_{ij2} \\
&= \sum_{i=1}^{n} \frac{I_{(c_i,1)}(y_i)}{1-c_i} [\mu_i(y_i^* - \mu_i^*) + y_i^\bullet - \mu_i^\bullet] \frac{\partial \phi_i}{\partial \eta_{i2}} x_{ij2}
\end{aligned}$$

e, portanto,

$$U_{\beta_2}(\beta_1,\beta_2) = \left(\frac{\partial l(\theta)}{\partial \beta_{12}}, \ldots, \frac{\partial l(\theta)}{\partial \beta_{p_\phi 2}}\right)^\top$$

pode ser representado como produto de matrizes como dado em (3.11).

B.3 Obtenção da matriz de informação de Fisher

Cada um dos termos da matriz de informação de Fisher pode ser obtido multiplicando-se por -1 a esperança matemática da segunda derivada do logaritmo da função de verossimilhança (3.6) em relação a cada um dos parâmetros. Para uma melhor organização das demonstrações, em algumas situações, denotaremos $l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})$ simplesmente como l_{i1} e $l_{i2}(\mu_i,\phi_i)$ como l_{i2}. Utilizando-se as derivadas já apresentadas na Seção B.2 temos que

$$\begin{aligned}
\frac{\partial^2 l(\theta)}{\partial \gamma_{j0}\gamma_{l0}} &= \frac{\partial^2 l_1(\gamma_0,\gamma_1,\gamma_c)}{\partial \gamma_{j0}\gamma_{l0}} = \sum_{i=1}^{n} \frac{\partial^2 l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})}{\partial \gamma_{j0}\gamma_{l0}} \\
&= \sum_{i=1}^{n} \left(\frac{\partial^2 l_{i1}}{\partial \delta_{i0}^2} \frac{\partial \delta_{i0}}{\partial \zeta_{i0}} \frac{\partial \zeta_{i0}}{\partial \gamma_{l0}} + \frac{\partial^2 l_{i1}}{\partial \delta_{i0}\delta_{i1}} \frac{\partial \delta_{i1}}{\partial \zeta_{i0}} \frac{\partial \zeta_{i0}}{\partial \gamma_{l0}} + \frac{\partial^2 l_{i1}}{\partial \delta_{i0}\delta_{ic}} \frac{\partial \delta_{ic}}{\partial \zeta_{i0}} \frac{\partial \zeta_{i0}}{\partial \gamma_{l0}} \right) \frac{\partial \delta_{i0}}{\partial \zeta_{i0}} \frac{\partial \zeta_{i0}}{\partial \gamma_{j0}} \\
&+ \sum_{i=1}^{n} \frac{\partial l_{i1}}{\partial \delta_{i0}} \frac{\partial^2 \delta_{i0}}{\partial \zeta_{i0}^2} \frac{\partial \zeta_{i0}}{\partial \gamma_{l0}} \frac{\partial \zeta_{i0}}{\partial \gamma_{j0}} \\
&+ \sum_{i=1}^{n} \left(\frac{\partial^2 l_{i1}}{\partial \delta_{i1}\delta_{i0}} \frac{\partial \delta_{i0}}{\partial \zeta_{i0}} \frac{\partial \zeta_{i0}}{\partial \gamma_{l0}} + \frac{\partial^2 l_{i1}}{\partial \delta_{i1}^2} \frac{\partial \delta_{i1}}{\partial \zeta_{i0}} \frac{\partial \zeta_{i0}}{\partial \gamma_{l0}} + \frac{\partial^2 l_{i1}}{\partial \delta_{i1}\delta_{ic}} \frac{\partial \delta_{ic}}{\partial \zeta_{i0}} \frac{\partial \zeta_{i0}}{\partial \gamma_{l0}} \right) \frac{\partial \delta_{i1}}{\partial \zeta_{i0}} \frac{\partial \zeta_{i0}}{\partial \gamma_{j0}} \\
&+ \sum_{i=1}^{n} \frac{\partial l_{i1}}{\partial \delta_{i1}} \frac{\partial^2 \delta_{i1}}{\partial \zeta_{i0}^2} \frac{\partial \zeta_{i0}}{\partial \gamma_{l0}} \frac{\partial \zeta_{i0}}{\partial \gamma_{j0}} \\
&+ \sum_{i=1}^{n} \left(\frac{\partial^2 l_{i1}}{\partial \delta_{ic}\delta_{i0}} \frac{\partial \delta_{i0}}{\partial \zeta_{i0}} \frac{\partial \zeta_{i0}}{\partial \gamma_{l0}} + \frac{\partial^2 l_{i1}}{\partial \delta_{ic}\delta_{i1}} \frac{\partial \delta_{i1}}{\partial \zeta_{i0}} \frac{\partial \zeta_{i0}}{\partial \gamma_{l0}} + \frac{\partial^2 l_{i1}}{\partial \delta_{ic}^2} \frac{\partial \delta_{ic}}{\partial \zeta_{i0}} \frac{\partial \zeta_{i0}}{\partial \gamma_{l0}} \right) \frac{\partial \delta_{ic}}{\partial \zeta_{i0}} \frac{\partial \zeta_{i0}}{\partial \gamma_{j0}} \\
&+ \sum_{i=1}^{n} \frac{\partial l_{i1}}{\partial \delta_{ic}} \frac{\partial^2 \delta_{ic}}{\partial \zeta_{i0}^2} \frac{\partial \zeta_{i0}}{\partial \gamma_{l0}} \frac{\partial \zeta_{i0}}{\partial \gamma_{j0}}.
\end{aligned}$$

Temos ainda que

$$
\begin{aligned}
\mathrm{E}\left(\frac{\partial l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})}{\partial \delta_{i0}}\right) &= \mathrm{E}\left(\frac{I_{\{0\}}(y_i)}{\delta_{i0}} - \frac{I_{(c_i,1)}(y_i)}{1-\alpha_i}\right) = 0,\\
\mathrm{E}\left(\frac{\partial l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})}{\partial \delta_{i1}}\right) &= \mathrm{E}\left(\frac{I_{\{1\}}(y_i)}{\delta_{i1}} - \frac{I_{(c_i,1)}(y_i)}{1-\alpha_i}\right) = 0,\\
\mathrm{E}\left(\frac{\partial l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})}{\partial \delta_{ic}}\right) &= \mathrm{E}\left(\frac{I_{\{c_i\}}(y_i)}{\delta_{ic}} - \frac{I_{(c_i,1)}(y_i)}{1-\alpha_i}\right) = 0,\\
\mathrm{E}\left(\frac{\partial^2 l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})}{\partial \delta_{i0}^2}\right) &= \mathrm{E}\left(-\frac{I_{\{0\}}(y_i)}{\delta_{i0}^2} - \frac{I_{(c_i,1)}(y_i)}{(1-\alpha_i)^2}\right) = -\frac{1}{\delta_{i0}} - \frac{1}{1-\alpha_i},\\
\mathrm{E}\left(\frac{\partial^2 l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})}{\partial \delta_{i1}^2}\right) &= \mathrm{E}\left(-\frac{I_{\{1\}}(y_i)}{\delta_{i1}^2} - \frac{I_{(c_i,1)}(y_i)}{(1-\alpha_i)^2}\right) = -\frac{1}{\delta_{i1}} - \frac{1}{1-\alpha_i},\\
\mathrm{E}\left(\frac{\partial^2 l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})}{\partial \delta_{ic}^2}\right) &= \mathrm{E}\left(-\frac{I_{\{c_i\}}(y_i)}{\delta_{ic}^2} - \frac{I_{(c_i,1)}(y_i)}{(1-\alpha_i)^2}\right) = -\frac{1}{\delta_{ic}} - \frac{1}{1-\alpha_i},\\
\mathrm{E}\left(\frac{\partial^2 l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})}{\partial \delta_{i0}\delta_{i1}}\right) &= \mathrm{E}\left(-\frac{I_{(c_i,1)}(y_i)}{(1-\alpha_i)^2}\right) = -\frac{1}{1-\alpha_i},\\
\mathrm{E}\left(\frac{\partial^2 l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})}{\partial \delta_{i0}\delta_{ic}}\right) &= \mathrm{E}\left(-\frac{I_{(c_i,1)}(y_i)}{(1-\alpha_i)^2}\right) = -\frac{1}{1-\alpha_i},\\
\mathrm{E}\left(\frac{\partial^2 l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})}{\partial \delta_{i1}\delta_{ic}}\right) &= \mathrm{E}\left(-\frac{I_{(c_i,1)}(y_i)}{(1-\alpha_i)^2}\right) = -\frac{1}{1-\alpha_i}.
\end{aligned}
$$

Assim, o termo (j,l) de $K_{\gamma_0\gamma_0}$ que compõe a matriz de informação de Fisher (3.19) é dado por

$$
\begin{aligned}
-\mathrm{E}\left(\sum_{i=1}^{n}\frac{\partial^2 l_{i1}}{\partial \gamma_{j0}\gamma_{l0}}\right) &= \sum_{i=1}^{n}\left[\left(\frac{1}{\delta_{i0}}+\frac{1}{1-\alpha_i}\right)\frac{\partial \delta_{i0}}{\partial \zeta_{i0}}z_{il0} + \frac{1}{1-\alpha_i}\frac{\partial \delta_{i1}}{\partial \zeta_{i0}}z_{il0} + \frac{1}{1-\alpha_i}\frac{\partial \delta_{ic}}{\partial \zeta_{i0}}z_{il0}\right]\frac{\partial \delta_{i0}}{\partial \zeta_{i0}}z_{ij0}\\
&+ \sum_{i=1}^{n}\left[\frac{1}{1-\alpha_i}\frac{\partial \delta_{i0}}{\partial \zeta_{i0}}z_{il0} + \left(\frac{1}{\delta_{i1}}+\frac{1}{1-\alpha_i}\right)\frac{\partial \delta_{i1}}{\partial \zeta_{i0}}z_{il0} + \frac{1}{1-\alpha_i}\frac{\partial \delta_{ic}}{\partial \zeta_{i0}}z_{il0}\right]\frac{\partial \delta_{i1}}{\partial \zeta_{i0}}z_{ij0}\\
&+ \sum_{i=1}^{n}\left[\frac{1}{1-\alpha_i}\frac{\partial \delta_{i0}}{\partial \zeta_{i0}}z_{il0} + \frac{1}{1-\alpha_i}\frac{\partial \delta_{i1}}{\partial \zeta_{i0}}z_{il0} + \left(\frac{1}{\delta_{ic}}+\frac{1}{1-\alpha_i}\right)\frac{\partial \delta_{ic}}{\partial \zeta_{i0}}z_{il0}\right]\frac{\partial \delta_{ic}}{\partial \zeta_{i0}}z_{ij0}\\
&= \sum_{i=1}^{n} z_{ij0}z_{il0}\left\{\frac{1}{\delta_{i0}}\left(\frac{\partial \delta_{i0}}{\partial \zeta_{i0}}\right)^2 + \frac{1}{\delta_{i1}}\left(\frac{\partial \delta_{i1}}{\partial \zeta_{i0}}\right)^2 + \frac{1}{\delta_{ic}}\left(\frac{\partial \delta_{ic}}{\partial \zeta_{i0}}\right)^2\right\}\\
&+ \sum_{i=1}^{n} z_{ij0}z_{il0}\left\{\frac{1}{1-\alpha_i}\left(\frac{\partial \delta_{i0}}{\partial \zeta_{i0}} + \frac{\partial \delta_{i1}}{\partial \zeta_{i0}} + \frac{\partial \delta_{ic}}{\partial \zeta_{i0}}\right)^2\right\}
\end{aligned}
$$

e, portanto, $K_{\gamma_0\gamma_0}$ pode ser escrito conforme apresentado em (3.22). De forma equivalente, obtêm-se $K_{\gamma_1\gamma_1}$ e $K_{\gamma_c\gamma_c}$. Também utilizando-se as derivadas já apresentadas na Seção B.2 obtém-se que

$$\begin{aligned}
\frac{\partial^2 l(\theta)}{\partial\gamma_{j0}\gamma_{l1}} &= \frac{\partial^2 l_1(\gamma_0,\gamma_1,\gamma_c)}{\partial\gamma_{j0}\gamma_{l1}} = \sum_{i=1}^{n}\frac{\partial^2 l_{i1}(\delta_{i0},\delta_{i1},\delta_{ic})}{\partial\gamma_{j0}\gamma_{l1}} \\
&= \sum_{i=1}^{n}\left(\frac{\partial^2 l_{i1}}{\partial\delta_{i0}^2}\frac{\partial\delta_{i0}}{\partial\zeta_{i1}}\frac{\partial\zeta_{i1}}{\partial\gamma_{l1}} + \frac{\partial^2 l_{i1}}{\partial\delta_{i0}\delta_{i1}}\frac{\partial\delta_{i1}}{\partial\zeta_{i1}}\frac{\partial\zeta_{i1}}{\partial\gamma_{l1}} + \frac{\partial^2 l_{i1}}{\partial\delta_{i0}\delta_{ic}}\frac{\partial\delta_{ic}}{\partial\zeta_{i1}}\frac{\partial\zeta_{i1}}{\partial\gamma_{l1}}\right)\frac{\partial\delta_{i0}}{\partial\zeta_{i0}}\frac{\partial\zeta_{i0}}{\partial\gamma_{j0}} \\
&+ \sum_{i=1}^{n}\frac{\partial l_{i1}}{\partial\delta_{i0}}\frac{\partial^2\delta_{i0}}{\partial\zeta_{i0}\zeta_{i1}}\frac{\partial\zeta_{i1}}{\partial\gamma_{l1}}\frac{\partial\zeta_{i0}}{\partial\gamma_{j0}} \\
&+ \sum_{i=1}^{n}\left(\frac{\partial^2 l_{i1}}{\partial\delta_{i1}\delta_{i0}}\frac{\partial\delta_{i0}}{\partial\zeta_{i1}}\frac{\partial\zeta_{i1}}{\partial\gamma_{l1}} + \frac{\partial^2 l_{i1}}{\partial\delta_{i1}^2}\frac{\partial\delta_{i1}}{\partial\zeta_{i1}}\frac{\partial\zeta_{i1}}{\partial\gamma_{l1}} + \frac{\partial^2 l_{i1}}{\partial\delta_{i1}\delta_{ic}}\frac{\partial\delta_{ic}}{\partial\zeta_{i1}}\frac{\partial\zeta_{i1}}{\partial\gamma_{l1}}\right)\frac{\partial\delta_{i1}}{\partial\zeta_{i0}}\frac{\partial\zeta_{i0}}{\partial\gamma_{j0}} \\
&+ \sum_{i=1}^{n}\frac{\partial l_{i1}}{\partial\delta_{i1}}\frac{\partial^2\delta_{i1}}{\partial\zeta_{i0}\zeta_{i1}}\frac{\partial\zeta_{i1}}{\partial\gamma_{l1}}\frac{\partial\zeta_{i0}}{\partial\gamma_{j0}} \\
&+ \sum_{i=1}^{n}\left(\frac{\partial^2 l_{i1}}{\partial\delta_{ic}\delta_{i0}}\frac{\partial\delta_{i0}}{\partial\zeta_{i1}}\frac{\partial\zeta_{i1}}{\partial\gamma_{l1}} + \frac{\partial^2 l_{i1}}{\partial\delta_{ic}\delta_{i1}}\frac{\partial\delta_{i1}}{\partial\zeta_{i1}}\frac{\partial\zeta_{i1}}{\partial\gamma_{l1}} + \frac{\partial^2 l_{i1}}{\partial\delta_{ic}^2}\frac{\partial\delta_{ic}}{\partial\zeta_{i1}}\frac{\partial\zeta_{i1}}{\partial\gamma_{l1}}\right)\frac{\partial\delta_{ic}}{\partial\zeta_{i0}}\frac{\partial\zeta_{i0}}{\partial\gamma_{j0}} \\
&+ \sum_{i=1}^{n}\frac{\partial l_{i1}}{\partial\delta_{ic}}\frac{\partial^2\delta_{ic}}{\partial\zeta_{i0}\zeta_{i1}}\frac{\partial\zeta_{i1}}{\partial\gamma_{l1}}\frac{\partial\zeta_{i0}}{\partial\gamma_{j0}}.
\end{aligned}$$

Assim, o termo (j,l) de $K_{\gamma_0\gamma_1}$ é dado por

$$\begin{aligned}
-\mathrm{E}\left(\sum_{i=1}^{n}\frac{\partial^2 l_{i1}}{\partial\gamma_{j0}\gamma_{l1}}\right) &= \sum_{i=1}^{n}\left[\left(\frac{1}{\delta_{i0}}+\frac{1}{1-\alpha_i}\right)\frac{\partial\delta_{i0}}{\partial\zeta_{i1}}z_{il1} + \frac{1}{1-\alpha_i}\frac{\partial\delta_{i1}}{\partial\zeta_{i1}}z_{il1} + \frac{1}{1-\alpha_i}\frac{\partial\delta_{ic}}{\partial\zeta_{i1}}z_{il1}\right]\frac{\partial\delta_{i0}}{\partial\zeta_{i0}}z_{ij0} \\
&+ \sum_{i=1}^{n}\left[\frac{1}{1-\alpha_i}\frac{\partial\delta_{i0}}{\partial\zeta_{i1}}z_{il1} + \left(\frac{1}{\delta_{i1}}+\frac{1}{1-\alpha_i}\right)\frac{\partial\delta_{i1}}{\partial\zeta_{i1}}z_{il1} + \frac{1}{1-\alpha_i}\frac{\partial\delta_{ic}}{\partial\zeta_{i1}}z_{il1}\right]\frac{\partial\delta_{i1}}{\partial\zeta_{i0}}z_{ij0} \\
&+ \sum_{i=1}^{n}\left[\frac{1}{1-\alpha_i}\frac{\partial\delta_{i0}}{\partial\zeta_{i1}}z_{il1} + \frac{1}{1-\alpha_i}\frac{\partial\delta_{i1}}{\partial\zeta_{i1}}z_{il1} + \left(\frac{1}{\delta_{ic}}+\frac{1}{1-\alpha_i}\right)\frac{\partial\delta_{ic}}{\partial\zeta_{i1}}z_{il1}\right]\frac{\partial\delta_{ic}}{\partial\zeta_{i0}}z_{ij0} \\
&= \sum_{i=1}^{n}z_{ij0}z_{il1}\left\{\frac{1}{\delta_{i0}}\frac{\partial\delta_{i0}}{\partial\zeta_{i0}}\frac{\partial\delta_{i0}}{\partial\zeta_{i1}} + \frac{1}{\delta_{i1}}\frac{\partial\delta_{i1}}{\partial\zeta_{i0}}\frac{\partial\delta_{i1}}{\partial\zeta_{i1}} + \frac{1}{\delta_{ic}}\frac{\partial\delta_{ic}}{\partial\zeta_{i0}}\frac{\partial\delta_{ic}}{\partial\zeta_{i1}}\right\} \\
&+ \sum_{i=1}^{n}z_{ij0}z_{il1}\left\{\frac{1}{1-\alpha_i}\left(\frac{\partial\delta_{i0}}{\partial\zeta_{i0}}+\frac{\partial\delta_{i1}}{\partial\zeta_{i0}}+\frac{\partial\delta_{ic}}{\partial\zeta_{i0}}\right)\left(\frac{\partial\delta_{i0}}{\partial\zeta_{i1}}+\frac{\partial\delta_{i1}}{\partial\zeta_{i1}}+\frac{\partial\delta_{ic}}{\partial\zeta_{i1}}\right)\right\}
\end{aligned}$$

e, logo, $K_{\gamma_0\gamma_1}$ pode ser escrito conforme apresentado em (3.22). De forma análoga, obtêm-se $K_{\gamma_0\gamma_c}$ e $K_{\gamma_1\gamma_c}$.

Assim como na função escore, para os casos em que a função de ligação H é a mesma dos modelos RLBIZUT, os componentes de $K_D(\gamma_0,\gamma_1,\gamma_c)$ simplificam-se significativamente. Nesse caso, as derivadas de δ_{i0}, δ_{i1} e δ_{ic} em relação à ζ_{i0} calculadas na Seção B.2 podem ser reescritas como

$$\frac{\partial\delta_{i0}}{\partial\zeta_{i0}} = \delta_{i0}(1-\delta_{i0}),\ \frac{\partial\delta_{i1}}{\partial\zeta_{i0}} = -\delta_{i0}\delta_{i1}\ \text{ e }\ \frac{\partial\delta_{ic}}{\partial\zeta_{i0}} = -\delta_{i0}\delta_{ic}.$$

Equivalentemente,

$$\frac{\partial \delta_{i0}}{\partial \zeta_{i1}} = -\delta_{i0}\delta_{i1}, \frac{\partial \delta_{i1}}{\partial \zeta_{i1}} = \delta_{i1}(1-\delta_{i1}) \text{ e } \frac{\partial \delta_{ic}}{\partial \zeta_{i1}} = -\delta_{i1}\delta_{ic}$$

e

$$\frac{\partial \delta_{i0}}{\partial \zeta_{ic}} = -\delta_{i0}\delta_{ic}, \frac{\partial \delta_{i1}}{\partial \zeta_{ic}} = -\delta_{i1}\delta_{ic} \text{ e } \frac{\partial \delta_{ic}}{\partial \zeta_{ic}} = \delta_{ic}(1-\delta_{ic}).$$

Logo, para essa particular função H, o termo (j,l) de $K_{\gamma_0\gamma_0}$ é dado por

$$\begin{aligned}
-\mathrm{E}\left(\sum_{i=1}^{n}\frac{\partial^2 l_{i1}}{\partial\gamma_{j0}\gamma_{l0}}\right) &= \sum_{i=1}^{n} z_{ij0}z_{il0}\left[\frac{1}{\delta_{i0}}\delta_{i0}^2(1-\delta_{i0})^2 + \frac{1}{\delta_{i1}}\delta_{i0}^2\delta_{i1}^2 + \frac{1}{\delta_{ic}}\delta_{i0}^2\delta_{ic}^2 + \frac{1}{1-\alpha_i}\delta_{i0}^2(1-\alpha_i)^2\right] \\
&= \sum_{i=1}^{n} z_{ij0}z_{il0}[\delta_{i0}(1-\delta_{i0})^2 + \delta_{i0}^2(1-\delta_{i0})] \\
&= \sum_{i=1}^{n} z_{ij0}z_{il0}[\delta_{i0}(1-\delta_{i0})]
\end{aligned}$$

e assim $K_{\gamma_0\gamma_0}$ pode ser escrito conforme apresentado em (3.26). De forma semelhante, o termo (j,l) de $K_{\gamma_0\gamma_1}$ é dado por

$$\begin{aligned}
-\mathrm{E}\left(\sum_{i=1}^{n}\frac{\partial^2 l_{i1}}{\partial\gamma_{j0}\gamma_{l1}}\right) &= \sum_{i=1}^{n} z_{ij0}z_{il1}\left[\frac{1}{\delta_{i0}}\delta_{i0}(1-\delta_{i0})(-\delta_{i0}\delta_{i1}) + \frac{1}{\delta_{i1}}(-\delta_{i0}\delta_{i1})\delta_{i1}(1-\delta_{i1})\right] \\
&+ \sum_{i=1}^{n} z_{ij0}z_{il1}\left[\frac{1}{\delta_{ic}}(-\delta_{i0}\delta_{ic})(-\delta_{i1}\delta_{ic}) + \frac{1}{1-\alpha_i}\delta_{i0}(1-\alpha_i)\delta_{i1}(1-\alpha_i)\right] \\
&= \sum_{i=1}^{n} z_{ij0}z_{il0}[-\delta_{i0}\delta_{i1}(1-\delta_{i0}) - \delta_{i0}\delta_{i1}(1-\delta_{i1}) + \delta_{i0}\delta_{i1}\delta_{ic} + \delta_{i0}\delta_{i1}(1-\alpha_i)] \\
&= \sum_{i=1}^{n} z_{ij0}z_{il0}[-\delta_{i0}\delta_{i1}]
\end{aligned}$$

e $K_{\gamma_0\gamma_1}$ pode ser escrito conforme apresentado em (3.26). De forma análoga, obtém-se os termos $K_{\gamma_1\gamma_1}$, $K_{\gamma_c\gamma_c}$, $K_{\gamma_0\gamma_c}$ e $K_{\gamma_1\gamma_c}$ para essa particular função H.

Para obter os termos da informação de Fisher que compõem $K_C(\gamma_0,\gamma_1,\gamma_c,\beta_1,\beta_2)$ será utilizada uma proposição adaptada de Ospina (2008, página 115) apresentada a seguir.

Proposição B.3.1. *Sejam $y_1, y_2, \ldots, y_n$ variáveis aleatórias independentes, em que y_i tem distribuição BIZUT$(\delta_{i0}, \delta_{i1}, \delta_{ic}, \mu_i, \phi_i, c_i)$ conforme definido em (2.6) e $\alpha_i = \delta_{i0} + \delta_{i1} + \delta_{ic}$. Se $\mathcal{F} : (0,1) \to \mathbb{R}$ é uma função contínua, então*

$$E\left(\sum_{i:y_i\in(c_i,1)}\mathcal{F}(y_i)\right) = \sum_{i=1}^{n}(1-\alpha_i)E(\mathcal{F}(y_i)|I_{(c_i,1)}(y_i)=1).$$

Demonstração. Seja

$$\mathcal{F}^*(y_i) = \begin{cases} 0 & \text{se } y_i \in \{0, 1, c_i\} \\ \mathcal{F}(y_i) & \text{se } y_i \in (c_i, 1). \end{cases}$$

Assim,

$$\begin{aligned} \mathrm{E}\left(\sum_{i:y_i\in(c_i,1)} \mathcal{F}(y_i)\right) &= \mathrm{E}\left(\sum_{i=1}^{n} \mathcal{F}^*(y_i)\right) = \sum_{i=1}^{n} \mathrm{E}(\mathcal{F}^*(y_i)) = \sum_{i=1}^{n} \mathrm{E}[\mathrm{E}(\mathcal{F}^*(y_i)|I_{(c_i,1)}(y_i))] \\ &= \sum_{i=1}^{n} [\mathrm{E}(\mathcal{F}^*(y_i)|I_{(c_i,1)}(y_i) = 0)\mathrm{P}(I_{(c_i,1)}(y_i) = 0) \\ &+ \mathrm{E}(\mathcal{F}^*(y_i)|I_{(c_i,1)}(y_i) = 1)\mathrm{P}(I_{(c_i,1)}(y_i) = 1)] \\ &= \sum_{i=1}^{n} (1 - \alpha_i)\mathrm{E}(\mathcal{F}(y_i)|I_{(c_i,1)}(y_i) = 1). \end{aligned}$$

□

Utilizando-se as derivadas apresentadas na Seção B.2 obtém-se que

$$\begin{aligned} \frac{\partial^2 l(\theta)}{\partial\beta_{j1}\beta_{l1}} &= \frac{\partial^2 l_2(\beta_1, \beta_2)}{\partial\beta_{j1}\beta_{l1}} = \sum_{i:y_i\in(c_i,1)} \frac{\partial^2 l_{i2}(\mu_i, \phi_i)}{\partial\beta_{j1}\beta_{l1}} \\ &= \sum_{i:y_i\in(c_i,1)} \left(\frac{\partial^2 l_{i2}}{\partial\mu_i^2} \left(\frac{\partial\mu_i}{\partial\eta_{i1}}\right)^2 \frac{\partial\eta_{i1}}{\partial\beta_{j1}} \frac{\partial\eta_{i1}}{\partial\beta_{l1}} + \frac{\partial l_{i2}}{\partial\mu_i} \frac{\partial^2\mu_i}{\partial\eta_{i1}^2} \frac{\partial\eta_{i1}}{\partial\beta_{j1}} \frac{\partial\eta_{i1}}{\partial\beta_{l1}} \right). \end{aligned}$$

Temos que

$$\frac{\partial^2 l_{i2}}{\partial\mu_i^2} = -w_i, \frac{\partial\eta_{i1}}{\partial\beta_{j1}} = x_{ij1} \text{ e } \frac{\partial\eta_{i1}}{\partial\beta_{l1}} = x_{il1},$$

em que w_i foi definido em (3.21). Como, sob condições de regularidade, $\mathrm{E}(\partial l_{i2}/\partial\mu_i|I_{(c_i,1)}(y_i) = 1) = 0$, utilizando-se a Proposição B.3.1 obtém-se que o termo (j, l) de $K_{\beta_1\beta_1}$ é dado por

$$\begin{aligned} -\mathrm{E}\left(\sum_{i:y_i\in(c_i,1)} \frac{\partial^2 l_{i2}}{\partial\beta_{j1}\beta_{l1}}\right) &= \mathrm{E}\left(\sum_{i:y_i\in(c_i,1)} w_i \left(\frac{\partial\mu_i}{\partial\eta_{i1}}\right)^2 x_{ij1}x_{il1}\right) \\ &= \sum_{i=1}^{n} (1 - \alpha_i) w_i \left(\frac{\partial\mu_i}{\partial\eta_{i1}}\right)^2 x_{ij1}x_{il1}. \end{aligned}$$

Logo, $K_{\beta_1\beta_1}$ pode ser escrito como foi apresentado em (3.22).

De forma semelhante,

$$\begin{aligned}\frac{\partial^2 l(\theta)}{\partial\beta_{j2}\beta_{l2}} &= \frac{\partial^2 l_2(\beta_1,\beta_2)}{\partial\beta_{j2}\beta_{l2}} = \sum_{i:y_i\in(c_i,1)} \frac{\partial^2 l_{i2}(\mu_i,\phi_i)}{\partial\beta_{j2}\beta_{l2}} \\ &= \sum_{i:y_i\in(c_i,1)} \left(\frac{\partial^2 l_{i2}}{\partial\phi_i^2}\left(\frac{\partial\phi_i}{\partial\eta_{i2}}\right)^2 \frac{\partial\eta_{i2}}{\partial\beta_{j2}}\frac{\partial\eta_{i2}}{\partial\beta_{l2}} + \frac{\partial l_{i2}}{\partial\phi_i}\frac{\partial^2\phi_i}{\partial\eta_{i2}^2}\frac{\partial\eta_{i2}}{\partial\beta_{j2}}\frac{\partial\eta_{i2}}{\partial\beta_{l2}}\right).\end{aligned}$$

Temos que

$$\frac{\partial^2 l_{i2}}{\partial\phi_i^2} = -b_i, \frac{\partial\eta_{i2}}{\partial\beta_{j2}} = x_{ij2} \text{ e } \frac{\partial\eta_{i2}}{\partial\beta_{l2}} = x_{il2},$$

em que b_i foi definido em (3.21). Como, sob condições de regularidade, $\mathrm{E}(\partial l_{i2}/\partial\phi_i | I_{(c_i,1)}(y_i) = 1) = 0$, utilizando-se a Proposição B.3.1 obtém-se que o termo (j,l) de $K_{\beta_2\beta_2}$ é dado por

$$\begin{aligned}-\mathrm{E}\left(\sum_{i:y_i\in(c_i,1)} \frac{\partial^2 l_{i2}}{\partial\beta_{j2}\beta_{l2}}\right) &= \mathrm{E}\left(\sum_{i:y_i\in(c_i,1)} b_i\left(\frac{\partial\phi_i}{\partial\eta_{i2}}\right)^2 x_{ij2}x_{il2}\right) \\ &= \sum_{i=1}^{n}(1-\alpha_i)b_i\left(\frac{\partial\phi_i}{\partial\eta_{i2}}\right)^2 x_{ij2}x_{il2}.\end{aligned}$$

Portanto, $K_{\beta_2\beta_2}$ pode ser escrito conforme apresentado em (3.22).

De forma análoga,

$$\frac{\partial^2 l(\theta)}{\partial\beta_{j1}\beta_{l2}} = \frac{\partial^2 l_2(\beta_1,\beta_2)}{\partial\beta_{j1}\beta_{l2}} = \sum_{i:y_i\in(c_i,1)} \frac{\partial^2 l_{i2}(\mu_i,\phi_i)}{\partial\beta_{j1}\beta_{l2}} = \sum_{i:y_i\in(c_i,1)} \frac{\partial^2 l_{i2}}{\partial\mu_i\partial\phi_i}\frac{\partial\mu_i}{\partial\eta_{i1}}\frac{\partial\phi_i}{\partial\eta_{i2}}\frac{\partial\eta_{i1}}{\partial\beta_{j1}}\frac{\partial\eta_{i2}}{\partial\beta_{l2}}.$$

Temos que

$$\begin{aligned}\frac{\partial^2 l_{i2}}{\partial\mu_i\partial\phi_i} &= \frac{1}{1-c_i}\left\{\log\left(\frac{y_i-c_i}{1-y_i}\right) + \psi\left[\left(\frac{1-\mu_i}{1-c_i}\right)\phi_i\right] - \psi\left[\left(\frac{\mu_i-c_i}{1-c_i}\right)\phi_i\right]\right\} \\ &+ \frac{\phi_i}{1-c_i}\left\{\left(\frac{1-\mu_i}{1-c_i}\right)\psi'\left[\left(\frac{1-\mu_i}{1-c_i}\right)\phi_i\right] - \left(\frac{\mu_i-c_i}{1-c_i}\right)\psi'\left[\left(\frac{\mu_i-c_i}{1-c_i}\right)\phi_i\right]\right\}.\end{aligned}$$

Do Apêndice A.3 obtemos ainda que

$$\begin{aligned}E(\log(y_i-c_i)|I_{(c_i,1)}(y_i)=1) &= \psi\left[\left(\frac{\mu_i-c_i}{1-c_i}\right)\phi_i\right] - \psi(\phi_i) + \log(1-c_i), \\ E(\log(1-y_i)|I_{(c_i,1)}(y_i)=1) &= \psi\left[\left(\frac{1-\mu_i}{1-c_i}\right)\phi_i\right] - \psi(\phi_i) + \log(1-c_i).\end{aligned}$$

Assim,

$$\mathrm{E}\left(\frac{\partial^2 l_{i2}}{\partial\mu_i\partial\phi_i}\middle| I_{(c_i,1)}(y_i)=1\right) = -a_i,$$

em que a_i foi definido em (3.21). Dessa forma, utilizando-se a Proposição B.3.1 obtém-se que o

termo (j,l) de $K_{\beta_1\beta_2}$ é dado por

$$\begin{aligned}
-\mathrm{E}\left(\sum_{i:y_i\in(c_i,1)}\frac{\partial^2 l_{i2}(\mu_i,\phi_i)}{\partial\beta_{j1}\beta_{l2}}\right) &= -\mathrm{E}\left(\sum_{i:y_i\in(c_i,1)}\frac{\partial^2 l_{i2}}{\partial\mu_i\partial\phi_i}\frac{\partial\mu_i}{\partial\eta_{i1}}\frac{\partial\phi_i}{\partial\eta_{i2}}\frac{\partial\eta_{i1}}{\partial\beta_{j1}}\frac{\partial\eta_{i2}}{\partial\beta_{l2}}\right)\\
&= \sum_{i=1}^{n}(1-\alpha_i)\mathrm{E}\left(\frac{\partial^2 l_{i2}}{\partial\mu_i\partial\phi_i}\frac{\partial\mu_i}{\partial\eta_{i1}}\frac{\partial\phi_i}{\partial\eta_{i2}}x_{ij1}x_{il2}\middle| I_{(c_i,1)}(y_i)=1\right)\\
&= \sum_{i=1}^{n}(1-\alpha_i)a_i\frac{\partial\mu_i}{\partial\eta_{i1}}\frac{\partial\phi_i}{\partial\eta_{i2}}x_{ij1}x_{il2}.
\end{aligned}$$

Logo, $K_{\beta_1\beta_2}$ pode ser escrito conforme apresentado em (3.22).

B.4 Obtenção da partição do vetor escore

O vetor $U_A(\theta)$ definido na Seção 3.2 e necessário para o cálculo das estatísticas escore e gradiente pode ser obtido substituindo em $U(\theta)$, cujos componentes foram apresentados em (3.11), Z_0, Z_1, Z_c, X_1 e X_2 por Z_{A0}, Z_{A1}, Z_{Ac}, X_{A1} e X_{A2}, respectivamente. Nesta seção provaremos que $U_{\gamma_{A0}}$ pode ser obtido substituindo Z_0 por Z_{A0} em U_{γ_0}. A prova para os demais termos de $U_A(\theta)$ é similar.

Sejam

$$\begin{aligned}
D_0 &= (d_{10},\ldots,d_{n0})^\top = T_0(\Delta_0 y_{\{0\}}-\Delta_{(c,1)}y_{(c,1)}),\\
D_1 &= (d_{11},\ldots,d_{n1})^\top = T_{10}(\Delta_1 y_{\{1\}}-\Delta_{(c,1)}y_{(c,1)}) \text{ e}\\
D_c &= (d_{1c},\ldots,d_{nc})^\top = T_{c0}(\Delta_c y_{\{c\}}-\Delta_{(c,1)}y_{(c,1)}).
\end{aligned}$$

Então, das equações (3.11) temos que $U_{\gamma_0}=Z_0^\top D_0+Z_0^\top D_1+Z_0^\top D_c$. Como

$$Z_0^\top=\begin{pmatrix} z_{110} & z_{210} & \cdots & z_{n10}\\ z_{120} & z_{220} & \cdots & z_{n20}\\ \vdots & \vdots & & \vdots\\ z_{1p_00} & z_{2p_00} & \cdots & z_{np_00}\end{pmatrix},$$

então

$$U_{\gamma_0}=\begin{pmatrix} z_{110}(d_{10}+d_{11}+d_{1c})+z_{210}(d_{20}+d_{21}+d_{2c})+\ldots+z_{n10}(d_{n0}+d_{n1}+d_{nc})\\ z_{120}(d_{10}+d_{11}+d_{1c})+z_{220}(d_{20}+d_{21}+d_{2c})+\ldots+z_{n20}(d_{n0}+d_{n1}+d_{nc})\\ \vdots\\ z_{1p_00}(d_{10}+d_{11}+d_{1c})+z_{2p_00}(d_{20}+d_{21}+d_{2c})+\ldots+z_{np_00}(d_{n0}+d_{n1}+d_{nc})\end{pmatrix}.$$

Sem perda de generalidade, suponha que os parâmetros de interesse do vetor γ_0 sejam os q_0 primeiros.

Nesse caso,

$$Z_{A0}^{\top} = \begin{pmatrix} z_{110} & z_{210} & \cdots & z_{n10} \\ z_{120} & z_{220} & \cdots & z_{n20} \\ \vdots & \vdots & & \vdots \\ z_{1q_00} & z_{2q_00} & \cdots & z_{nq_00} \end{pmatrix},$$

e $Z_{A0}^{\top}D_0 + Z_{A0}^{\top}D_1 + Z_{A0}^{\top}D_c$ é igual a

$$\begin{pmatrix} z_{110}(d_{10}+d_{11}+d_{1c}) + z_{210}(d_{20}+d_{21}+d_{2c}) + \ldots + z_{n10}(d_{n0}+d_{n1}+d_{nc}) \\ z_{120}(d_{10}+d_{11}+d_{1c}) + z_{220}(d_{20}+d_{21}+d_{2c}) + \ldots + z_{n20}(d_{n0}+d_{n1}+d_{nc}) \\ \vdots \\ z_{1q_00}(d_{10}+d_{11}+d_{1c}) + z_{2q_00}(d_{20}+d_{21}+d_{2c}) + \ldots + z_{nq_00}(d_{n0}+d_{n1}+d_{nc}) \end{pmatrix} = U_{\gamma_{A0}},$$

conforme queríamos demonstrar.

B.5 Prova dos resultados para o modelo RBIZUT com c variável

Nesta seção serão provadas as três proposições apresentadas na Seção 3.3 para o modelo RBIZUT com c variável.

B.5.1 Prova da Proposição 3.3.1

Com as funções de ligação consideradas nesta proposição, tem-se que $\mu_i = (c_i + e^{\eta_{i1}})/(1+e^{\eta_{i1}})$, $\phi_i = e^{\eta_{i2}}$, $(\mu_i - c_i)/(1-c_i) = e^{\eta_{i1}}/(1+e^{\eta_{i1}})$ e $(1-\mu_i)/(1-c_i) = 1/(1+e^{\eta_{i1}})$, em que η_{i1} e η_{i2} não são função de c_i. Tem-se ainda que $\partial\mu_i/\partial\eta_{i1} = [(1-c_i)e^{\eta_{i1}}]/(1+e^{\eta_{i1}})^2$. Definindo-se $w_i^* = (1-\alpha_i)w_i(\partial\mu_i/\partial\eta_{i1})^2$, em que w_i foi definido em (3.21), então $K_{\beta_1\beta_1} = X_1^{\top}\mathrm{diag}\{w_1^*, w_2^*, \ldots, w_n^*\}X_1$. Tem-se que

$$\begin{aligned} w_i^* &= \left(\frac{e^{\eta_{i2}}}{1-c_i}\right)^2 \left\{\psi'\left(\frac{e^{\eta_{i2}}}{1+e^{\eta_{i1}}}\right) + \psi'\left(\frac{e^{\eta_{i1}}e^{\eta_{i2}}}{1+e^{\eta_{i1}}}\right)\right\}(1-\alpha_i)\left[\frac{(1-c_i)e^{\eta_{i1}}}{(1+e^{\eta_{i1}})^2}\right]^2 \\ &= (1-\alpha_i)\left[\frac{e^{\eta_{i1}}e^{\eta_{i2}}}{(1+e^{\eta_{i1}})^2}\right]^2 \left\{\psi'\left(\frac{e^{\eta_{i2}}}{1+e^{\eta_{i1}}}\right) + \psi'\left(\frac{e^{\eta_{i1}}e^{\eta_{i2}}}{1+e^{\eta_{i1}}}\right)\right\}, \end{aligned}$$

que não é função de c_i. Logo, $K_{\beta_1\beta_1}$ também não é função de c_i. De forma equivalente, prova-se que $K_{\beta_2\beta_2}$ e $K_{\beta_1\beta_2}$ não são função de c_i e assim está provada a proposição.

B.5.2 Prova da Proposição 3.3.2

Temos que, quando a função de ligação para μ_i é a utilizada no modelo RLBIZUT,

$$\mu_i^R = \frac{\dfrac{c_i + e^{\eta_{i1}+\tau f_c(c_i)}}{1+e^{\eta_{i1}+\tau f_c(c_i)}} - c_i}{1-c_i} = \frac{\dfrac{e^{\eta_{i1}+\tau f_c(c_i)} - c_i e^{\eta_{i1}+\tau f_c(c_i)}}{1+e^{\eta_{i1}+\tau f_c(c_i)}}}{1-c_i} = \frac{e^{\eta_{i1}+\tau f_c(c_i)}}{1+e^{\eta_{i1}+\tau f_c(c_i)}}.$$

Assim, se $\tau = 0$, $\mu_i^R = \exp(\eta_{i1})/1 + \exp(\eta_{i1})$, que é constante em relação à c_i, provando o item II da Proposição 3.3.2.

Temos ainda que

$$\frac{\partial \mu_i^R}{\partial c_i} = \frac{e^{\eta_{i1}+\tau f_c(c_i)}\tau f_c'(c_i)(1+e^{\eta_{i1}+\tau f_c(c_i)}) - e^{\eta_{i1}+\tau f_c(c_i)}(e^{\eta_{i1}+\tau f_c(c_i)}\tau f_c'(c_i))}{(1+e^{\eta_{i1}+\tau f_c(c_i)})^2} = \frac{\tau f_c'(c_i)e^{\eta_{i1}+\tau f_c(c_i)}}{(1+e^{\eta_{i1}+\tau f_c(c_i)})^2}.$$

Como $f_c(c_i)$ é crescente, $f_c'(c_i) > 0$. Logo, se $\tau > 0$, $\partial\mu_i^R/\partial c_i > 0$ e se $\tau < 0$, $\partial\mu_i^R/\partial c_i < 0$. Portanto, valem os itens I e III da Proposição 3.3.2.

B.5.3 Prova da Proposição 3.3.3

Temos que, quando a função de ligação para μ_i é a utilizada no modelo RLBIZUT,

$$\begin{aligned} g_j(j) &= \frac{c_i + j + e^{\eta_{i1}+\tau f_c(c_i+j)}}{1+e^{\eta_{i1}+\tau f_c(c_i+j)}} - \frac{c_i + e^{\eta_{i1}+\tau f_c(c_i)}}{1+e^{\eta_{i1}+\tau f_c(c_i)}} \\ &= \frac{j + je^{\eta_{i1}+\tau f_c(c_i)} + c_i e^{\eta_{i1}}(e^{\tau f_c(c_i)} - e^{\tau f_c(c_i+j)}) + e^{\eta_{i1}+\tau f_c(c_i+j)} - e^{\eta_{i1}+\tau f_c(c_i)}}{[1+e^{\eta_{i1}+\tau f_c(c_i)}][1+e^{\eta_{i1}+\tau f_c(c_i+j)}]}. \end{aligned}$$

O denominador de $g_j(j)$ é positivo. Portanto, para analisar o sinal de $g_j(j)$, basta estudarmos o sinal do numerador do mesmo que denotaremos por $g_j^N(j)$. Assim,

$$\begin{aligned} g_j^N(j) &= j + je^{\eta_{i1}+\tau f_c(c_i)} + e^{\eta_{i1}}(c_i - 1)(e^{\tau f_c(c_i)} - e^{\tau f_c(c_i+j)}) \\ &= j + e^{\eta_{i1}}[je^{\tau f_c(c_i)} - (1-c_i)(e^{\tau f_c(c_i)} - e^{\tau f_c(c_i+j)})]. \end{aligned}$$

Como $f_c(.)$ é positiva e crescente no intervalo $(0,1)$, se $\tau \geq 0$, então $\exp[\tau f_c(c_i)] - \exp[\tau f_c(c_i+j)] \leq 0$ e, portanto, $g_j^N(j) > 0$. Se $\tau < 0$ temos que

$$g_j^N(j) > j - e^{\eta_{i1}}(e^{\tau f_c(c_i)} - e^{\tau f_c(c_i+j)}) > j - e^{\eta_{i1}}e^{\tau f_c(c_i)} > j - e^{\eta_{i1}}.$$

Assim, para j fixado, existe η_{i1} tal que $j - e^{\eta_{i1}} > 0$ e, portanto, existe η_{i1} tal que $g_j^N(j) > 0 \,\forall\, \tau$. Logo, a proposição 3.3.3 está provada.

Apêndice C

Provas dos resultados do Capítulo 4

Neste apêndice são obtidas as matrizes Δ para os dois esquemas de perturbação da análise de influência local discutida na Seção 4.2.1.

C.1 Obtenção de Δ para a ponderação de casos

O termo (j,i) de Δ_{γ_0} é dado por $\partial^2 l(\theta|\omega)/\partial\gamma_{j0}\partial\omega_i$. Utilizando-se (4.14) e os cálculos do Apêndice B.2 temos que

$$\begin{aligned}\frac{\partial l(\theta|\omega))}{\partial\gamma_{j0}} &= \sum_{i=1}^{n}\omega_i\left\{\left[\frac{I_{\{0\}}(y_i)}{\delta_{i0}}-\frac{I_{(c_i,1)}(y_i)}{1-\alpha_i}\right]\frac{\partial\delta_{i0}}{\partial\zeta_{i0}}z_{ij0}\right\}+\sum_{i=1}^{n}\omega_i\left\{\left[\frac{I_{\{1\}}(y_i)}{\delta_{i1}}-\frac{I_{(c_i,1)}(y_i)}{1-\alpha_i}\right]\frac{\partial\delta_{i1}}{\partial\zeta_{i0}}z_{ij0}\right\}\\ &\quad+\sum_{i=1}^{n}\omega_i\left\{\left[\frac{I_{\{c_i\}}(y_i)}{\delta_{ic}}-\frac{I_{(c_i,1)}(y_i)}{1-\alpha_i}\right]\frac{\partial\delta_{ic}}{\partial\zeta_{i0}}z_{ij0}\right\}.\end{aligned}$$

Assim,

$$\begin{aligned}\frac{\partial^2 l(\theta|\omega))}{\partial\gamma_{j0}\partial\omega_i} &= \left\{\left[\frac{I_{\{0\}}(y_i)}{\delta_{i0}}-\frac{I_{(c_i,1)}(y_i)}{1-\alpha_i}\right]\frac{\partial\delta_{i0}}{\partial\zeta_{i0}}z_{ij0}\right\}+\left\{\left[\frac{I_{\{1\}}(y_i)}{\delta_{i1}}-\frac{I_{(c_i,1)}(y_i)}{1-\alpha_i}\right]\frac{\partial\delta_{i1}}{\partial\zeta_{i0}}z_{ij0}\right\}\\ &\quad+\left\{\left[\frac{I_{\{c_i\}}(y_i)}{\delta_{ic}}-\frac{I_{(c_i,1)}(y_i)}{1-\alpha_i}\right]\frac{\partial\delta_{ic}}{\partial\zeta_{i0}}z_{ij0}\right\}\end{aligned}$$

e o termo (j,i) de Δ_{γ_0} pode ser escrito conforme apresentado em (4.15). De forma equivalente obtém-se Δ_{γ_1} e Δ_{γ_c}.

Para a obtenção de Δ_{β_1} e Δ_{β_2} também utilizamos (4.14) e os cálculos do Apêndice B.2. Tem-se que

$$\begin{aligned}\frac{\partial l(\theta|\omega))}{\partial\beta_{j1}} &= \sum_{i=1}^{n}\omega_i\frac{\phi_i}{1-c_1}I_{(c_i,1)}(y_i)(y_i^*-\mu_i^*)\frac{\partial\mu_i}{\partial\eta_{i1}}x_{ij1},\\ \frac{\partial^2 l(\theta|\omega))}{\partial\beta_{j1}\partial\omega_i} &= \frac{\phi_i}{1-c_1}I_{(c_i,1)}(y_i)(y_i^*-\mu_i^*)\frac{\partial\mu_i}{\partial\eta_{i1}}x_{ij1}\end{aligned}$$

e assim Δ_{β_1} pode ser escrito conforme apresentado em (4.15). De forma equivalente, obtém-se Δ_{β_2}.

C.2 Obtenção de Δ para a perturbação individual de variáveis preditoras

No caso da perturbação de uma variável presente apenas em Z_0, precisamos obter Δ_{γ_0}, Δ_{γ_1} e Δ_{γ_c}. O termo (j,i) de Δ_{γ_0} é dado por $\partial^2 l(\theta|\omega)/\partial\gamma_{j0}\partial\omega_i$. Utilizando-se (4.17) temos que, se $j=m$,

$$\begin{aligned}
\frac{\partial l(\theta|\omega))}{\partial\gamma_{j0}} &= \sum_{i=1}^{n}\left\{\frac{\partial l_{i1}(\delta_{i0}(\omega),\delta_{i1}(\omega),\delta_{ic}(\omega))}{\partial\delta_{i0}(\omega)}\frac{\partial\delta_{i0}(\omega)}{\partial\zeta_{i0}(\omega)}\frac{\partial\zeta_{i0}(\omega)}{\partial\gamma_{j0}}\right\}+\\
&\quad\sum_{i=1}^{n}\left\{\frac{\partial l_{i1}(\delta_{i0}(\omega),\delta_{i1}(\omega),\delta_{ic}(\omega))}{\partial\delta_{i1}(\omega)}\frac{\partial\delta_{i1}(\omega)}{\partial\zeta_{i0}(\omega)}\frac{\partial\zeta_{i0}(\omega)}{\partial\gamma_{j0}}\right\}+\\
&\quad\sum_{i=1}^{n}\left\{\frac{\partial l_{i1}(\delta_{i0}(\omega),\delta_{i1}(\omega),\delta_{ic}(\omega))}{\partial\delta_{ic}(\omega)}\frac{\partial\delta_{ic}(\omega)}{\partial\zeta_{i0}(\omega)}\frac{\partial\zeta_{i0}(\omega)}{\partial\gamma_{j0}}\right\}.
\end{aligned}$$

Assim, denotando $l_{i1}(\delta_{i0}(\omega),\delta_{i1}(\omega),\delta_{ic}(\omega))$ simplesmente como $l_{i1}(\omega)$, tem-se que

$$\begin{aligned}
\frac{\partial^2 l(\theta|\omega))}{\partial\gamma_{j0}\partial\omega_i} &= \left(\frac{\partial^2 l_{i1}(\omega)}{\partial\delta_{i0}^2(\omega)}\frac{\partial\delta_{i0}(\omega)}{\partial\zeta_{i0}(\omega)}\frac{\partial\zeta_{i0}(\omega)}{\partial\omega_i}+\frac{\partial^2 l_{i1}(\omega)}{\partial\delta_{i0}(\omega)\delta_{i1}(\omega)}\frac{\partial\delta_{i1}(\omega)}{\partial\zeta_{i0}(\omega)}\frac{\partial\zeta_{i0}(\omega)}{\partial\omega_i}\right)\frac{\partial\delta_{i0}(\omega)}{\partial\zeta_{i0}(\omega)}\frac{\partial\zeta_{i0}(\omega)}{\partial\gamma_{j0}}\\
&+ \left(\frac{\partial^2 l_{i1}(\omega)}{\partial\delta_{i0}(\omega)\delta_{ic}(\omega)}\frac{\partial\delta_{ic}(\omega)}{\partial\zeta_{i0}(\omega)}\frac{\partial\zeta_{i0}(\omega)}{\partial\omega_i}\right)\frac{\partial\delta_{i0}(\omega)}{\partial\zeta_{i0}(\omega)}\frac{\partial\zeta_{i0}(\omega)}{\partial\gamma_{j0}}\\
&+ \frac{\partial l_{i1}(\omega)}{\partial\delta_{i0}(\omega)}\frac{\partial^2\delta_{i0}(\omega)}{\partial\zeta_{i0}^2(\omega)}\frac{\partial\zeta_{i0}(\omega)}{\partial\omega_i}\frac{\partial\zeta_{i0}(\omega)}{\partial\gamma_{j0}}+\frac{\partial l_{i1}(\omega)}{\partial\delta_{i0}(\omega)}\frac{\partial\delta_{i0}(\omega)}{\partial\zeta_{i0}(\omega)}\frac{\partial^2\zeta_{i0}(\omega)}{\partial\gamma_{j0}\partial\omega_i}\\
&+ \left(\frac{\partial^2 l_{i1}(\omega)}{\partial\delta_{i1}(\omega)\delta_{i0}(\omega)}\frac{\partial\delta_{i0}(\omega)}{\partial\zeta_{i0}(\omega)}\frac{\partial\zeta_{i0}(\omega)}{\partial\omega_i}+\frac{\partial^2 l_{i1}(\omega)}{\partial\delta_{i1}^2(\omega)}\frac{\partial\delta_{i1}(\omega)}{\partial\zeta_{i0}(\omega)}\frac{\partial\zeta_{i0}(\omega)}{\partial\omega_i}\right)\frac{\partial\delta_{i1}(\omega)}{\partial\zeta_{i0}(\omega)}\frac{\partial\zeta_{i0}(\omega)}{\partial\gamma_{j0}}\\
&+ \left(\frac{\partial^2 l_{i1}(\omega)}{\partial\delta_{i1}(\omega)\delta_{ic}(\omega)}\frac{\partial\delta_{ic}(\omega)}{\partial\zeta_{i0}(\omega)}\frac{\partial\zeta_{i0}(\omega)}{\partial\omega_i}\right)\frac{\partial\delta_{i1}(\omega)}{\partial\zeta_{i0}(\omega)}\frac{\partial\zeta_{i0}(\omega)}{\partial\gamma_{j0}}\\
&+ \frac{\partial l_{i1}(\omega)}{\partial\delta_{i1}(\omega)}\frac{\partial^2\delta_{i1}(\omega)}{\partial\zeta_{i0}^2(\omega)}\frac{\partial\zeta_{i0}(\omega)}{\partial\omega_i}\frac{\partial\zeta_{i0}(\omega)}{\partial\gamma_{j0}}+\frac{\partial l_{i1}(\omega)}{\partial\delta_{i1}(\omega)}\frac{\partial\delta_{i1}(\omega)}{\partial\zeta_{i0}(\omega)}\frac{\partial^2\zeta_{i0}(\omega)}{\partial\gamma_{j0}\partial\omega_i}\\
&+ \left(\frac{\partial^2 l_{i1}(\omega)}{\partial\delta_{ic}(\omega)\delta_{i0}(\omega)}\frac{\partial\delta_{i0}(\omega)}{\partial\zeta_{i0}(\omega)}\frac{\partial\zeta_{i0}(\omega)}{\partial\omega_i}+\frac{\partial^2 l_{i1}(\omega)}{\partial\delta_{ic}(\omega)\delta_{i1}(\omega)}\frac{\partial\delta_{i1}(\omega)}{\partial\zeta_{i0}(\omega)}\frac{\partial\zeta_{i0}(\omega)}{\partial\omega_i}\right)\frac{\partial\delta_{ic}(\omega)}{\partial\zeta_{i0}(\omega)}\frac{\partial\zeta_{i0}(\omega)}{\partial\gamma_{j0}}\\
&+ \left(\frac{\partial^2 l_{i1}(\omega)}{\partial\delta_{ic}^2(\omega)}\frac{\partial\delta_{ic}(\omega)}{\partial\zeta_{i0}(\omega)}\frac{\partial\zeta_{i0}(\omega)}{\partial\omega_i}\right)\frac{\partial\delta_{ic}(\omega)}{\partial\zeta_{i0}(\omega)}\frac{\partial\zeta_{i0}(\omega)}{\partial\gamma_{j0}}\\
&+ \frac{\partial l_{i1}(\omega)}{\partial\delta_{ic}(\omega)}\frac{\partial^2\delta_{ic}(\omega)}{\partial\zeta_{i0}^2(\omega)}\frac{\partial\zeta_{i0}(\omega)}{\partial\omega_i}\frac{\partial\zeta_{i0}(\omega)}{\partial\gamma_{j0}}+\frac{\partial l_{i1}(\omega)}{\partial\delta_{ic}(\omega)}\frac{\partial\delta_{ic}(\omega)}{\partial\zeta_{i0}(\omega)}\frac{\partial^2\zeta_{i0}(\omega)}{\partial\gamma_{j0}\partial\omega_i}.
\end{aligned}$$

Calculando-se as derivadas e rearranjando-se os termos, obtém-se que

$$
\begin{aligned}
\frac{\partial^2 l(\theta|\omega))}{\partial\gamma_{j0}\partial\omega_i} &= \gamma_{j0} S z_{m0}(z_{ijo}+\omega_i S z_{m0})\left\{\frac{-I_{\{0\}}(y_i)}{\delta_{i0}^2(\omega)}\left(\frac{\partial\delta_{i0}(\omega)}{\partial\zeta_{i0}(\omega)}\right)^2 - \frac{I_{\{1\}}(y_i)}{\delta_{i1}^2(\omega)}\left(\frac{\partial\delta_{i1}(\omega)}{\partial\zeta_{i0}(\omega)}\right)^2\right\} \\
&+ \gamma_{j0} S z_{m0}(z_{ijo}+\omega_i S z_{m0})\left\{\frac{-I_{\{c\}}(y_i)}{\delta_{ic}^2(\omega)}\left(\frac{\partial\delta_{ic}(\omega)}{\partial\zeta_{i0}(\omega)}\right)^2\right\} \\
&+ \gamma_{j0} S z_{m0}(z_{ijo}+\omega_i S z_{m0})\left\{\frac{-I_{(c_i,1)}(y_i)}{(1-\alpha_i(\omega))^2}\left(\frac{\partial\delta_{i0}(\omega)}{\partial\zeta_{i0}(\omega)}+\frac{\partial\delta_{i1}(\omega)}{\partial\zeta_{i0}(\omega)}+\frac{\partial\delta_{ic}(\omega)}{\partial\zeta_{i0}(\omega)}\right)^2\right\} \\
&+ \gamma_{j0} S z_{m0}(z_{ijo}+\omega_i S z_{m0})\left\{\frac{I_{\{0\}}(y_i)}{\delta_{i0}(\omega)} - \frac{I_{(c_i,1)}(y_i)}{1-\alpha_i(\omega)}\right\}\frac{\partial^2\delta_{i0}(\omega)}{\partial\zeta_{i0}^2(\omega)} \\
&+ \gamma_{j0} S z_{m0}(z_{ijo}+\omega_i S z_{m0})\left\{\frac{I_{\{1\}}(y_i)}{\delta_{i1}(\omega)} - \frac{I_{(c_i,1)}(y_i)}{1-\alpha_i(\omega)}\right\}\frac{\partial^2\delta_{i1}(\omega)}{\partial\zeta_{i0}^2(\omega)} \\
&+ \gamma_{j0} S z_{m0}(z_{ijo}+\omega_i S z_{m0})\left\{\frac{I_{\{c_i\}}(y_i)}{\delta_{ic}(\omega)} - \frac{I_{(c_i,1)}(y_i)}{1-\alpha_i(\omega)}\right\}\frac{\partial^2\delta_{ic}(\omega)}{\partial\zeta_{i0}^2(\omega)} \\
&+ S z_{m0}\left\{\frac{I_{\{0\}}(y_i)}{\delta_{i0}(\omega)} - \frac{I_{(c_i,1)}(y_i)}{1-\alpha_i(\omega)}\right\}\frac{\partial\delta_{i0}(\omega)}{\partial\zeta_{i0}(\omega)} \\
&+ S z_{m0}\left\{\frac{I_{\{1\}}(y_i)}{\delta_{i1}(\omega)} - \frac{I_{(c_i,1)}(y_i)}{1-\alpha_i(\omega)}\right\}\frac{\partial\delta_{i1}(\omega)}{\partial\zeta_{i0}(\omega)} \\
&+ S z_{m0}\left\{\frac{I_{\{c_i\}}(y_i)}{\delta_{ic}(\omega)} - \frac{I_{(c_i,1)}(y_i)}{1-\alpha_i(\omega)}\right\}\frac{\partial\delta_{ic}(\omega)}{\partial\zeta_{i0}(\omega)}.
\end{aligned}
$$

Pode-se notar a partir das derivadas apresentadas que os 3 últimos termos são iguais a zero se $j \neq m$. Assim, avaliando os parâmetros em seus respectivos estimadores de máxima verossimilhança e em $\omega_i = 0$, pode-se perceber que Δ_{γ_0} para a perturbação de uma variável preditora presente apenas em Z_0 pode ser expressa conforme apresentado em (4.18). De forma semelhante, obtém-se Δ_{γ_1} e Δ_{γ_c}.

No caso da perturbação de uma variável presente apenas em X_1, precisamos obter Δ_{β_1}, e Δ_{β_2}. O termo (j,i) de Δ_{β_1} é dado por $\partial^2 l(\theta|\omega)/\partial\beta_{j1}\partial\omega_i$. Utilizando-se (4.20) temos que, se $j = m$,

$$
\frac{\partial l(\theta|\omega)}{\partial\beta_{j1}} = \sum_{i:y_i\in(c_i,1)}\left\{\frac{\partial l_{i2}(\mu_i(\omega),\phi_i(\omega))}{\partial\mu_i(\omega)}\frac{\partial\mu_i(\omega)}{\partial\eta_{i1}(\omega)}\frac{\partial\eta_{i1}(\omega)}{\partial\beta_{j1}}\right\}.
$$

Se $y_i \notin (c_i,1)$, $\partial^2 l(\theta|\omega)/\partial\beta_{j1}\partial\omega_i = 0$. Denotando $l_{i2}(\mu_i(\omega),\phi_i(\omega))$ simplesmente como $l_{i2}(\omega)$, se

$y_i \in (c_i, 1)$, tem-se que

$$\begin{aligned}
\frac{\partial^2 l(\theta|\omega)}{\partial \beta_{j1}\omega_i} &= \frac{\partial^2 l_{i2}(\omega)}{\partial \mu_i^2(\omega)}\left(\frac{\partial \mu_i(\omega)}{\partial \eta_{i1}(\omega)}\right)^2 \frac{\partial \eta_{i1}(\omega)}{\partial \beta_{j1}}\frac{\partial \eta_{i1}(\omega)}{\partial \omega_i} + \frac{\partial l_{i2}(\omega)}{\partial \mu_i(\omega)}\frac{\partial^2 \mu_i(\omega)}{\partial \eta_{i1}^2(\omega)}\frac{\partial \eta_{i1}(\omega)}{\partial \omega_i}\frac{\partial \eta_{i1}(\omega)}{\partial \beta_{j1}} \\
&+ \frac{\partial l_{i2}(\omega)}{\partial \mu_i(\omega)}\frac{\partial \mu_i(\omega)}{\partial \eta_{i1}(\omega)}\frac{\partial^2 \eta_{i1}(\omega)}{\partial \beta_{j1}\partial \omega_i} \\
&= -w_i\left(\frac{\partial \mu_i(\omega)}{\partial \eta_{i1}(\omega)}\right)^2 x_{ij1}\beta_{j1}Sx_{m1} + \frac{\phi_i}{1-c_i}(y_i^* - \mu_i^*)\left(\frac{\partial^2 \mu_i(\omega)}{\partial \eta_{i1}^2(\omega)}\right) x_{ij1}\beta_{j1}Sx_{m1} \\
&+ \frac{\phi_i}{1-c_i}(y_i^* - \mu_i^*)\left(\frac{\partial \mu_i(\omega)}{\partial \eta_{i1}(\omega)}\right) Sx_{m1}.
\end{aligned}$$

Pode-se perceber a partir das derivadas apresentadas que o último termo é igual a zero se $j \neq m$. Assim, avaliando os parâmetros em seus respectivos estimadores de máxima verossimilhança e em $\omega_i = 0$, pode-se perceber que Δ_{β_1} para a perturbação de uma variável preditora presente apenas em X_1 pode ser expressa conforme apresentado em (4.21). De forma semelhante, obtém-se Δ_{β_2}.

Referências Bibliográficas

Abdul-Muhmin, A. G. e Umar, Y. A. (2007), 'Credit card ownership and usage behaviour in saudi arabia: the impact of demographics and attitudes', *Journal of Financial Services Marketing* **12**, 219–234. 5

Aitchison, J. (1955), 'On the distribution of a positive random variable having a discrete probability mass at the origin', *Journal of the American Statistical Association* **50**, 901–908. 4

Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **19**(6), 716–723. 46

Akaike, H. (1983), 'Information measures and model selection', *Bulletin of the International Statistical Institute* **50**, 277–290. 46

Andrade, A. G. C. (2007), Efeitos da especificação incorreta da função de ligação no modelo de regressão beta, Dissertação de mestrado em estatística, Instituto de Matemática e Estatística - Universidade de São Paulo. 3

Anholeto, T. (2010), Resíduos de pearson melhorados em modelos de regressão beta, Dissertação de mestrado em estatística, Instituto de Matemática e Estatística - Universidade de São Paulo. 49

Atkinson, A. C. (1985), *Plots, transformations and regression*, Clarendon Press, Oxford. 1, 50

Barndorff-Nielsen, O. E. e Jorgensen, B. (1991), 'Some parametric models on the simplex', *Journal of Multivariate Analysis* **39**, 106–116. 2

Bates, D. M. e Watts, D. G. (1980), 'Relative curvature measures of nonlinearity', *Journal of the Royal Statistical Society. Series B (Methodological)* **42**(1), 1–25. 56

Bayer, F. M. (2011), Modelagem e inferência em regressão beta, Tese de doutorado em estatística, Centro de Ciências Exatas e da Natureza - Universidade Federal de Pernambuco. 46, 47, 48

Branscum, A. J., Johnson, W. O. e Thurmond, M. C. (2007), 'Bayesian beta regression: applications to household expenditure data and genetic distance between foot-and-mouth disease viruses', *Australian and New Zealand Journal of Statistics* **49**(3), 287–301. 3

Cook, D. O., Kieschnick, R. e McCullough, B. D. (2008), 'Regression analysis of proportions in finance with self selection', *Journal of Empirical Finance* **15**(5), 860–867. 3, 4

Cook, R. D. (1986), 'Assessment of local influence (with discussion)', *Journal of the Royal Statistical Society Series B* **48**, 133–169. 55, 56

Cox, C. (1996), 'Nonlinear quasi-likelihood models: applications to continuous proportions', *Computational Statistics and Data Analysis* **21**, 449–461. 1, 3

Cox, D. R. e Reid, N. (1987), 'Parameter orthogonality and approximate conditional inference', *Journal of the Royal Statistical Society Series B* **49**(1), 1–39. 14

Cribari-Neto, F. e Souza, T. C. (2011), 'Testing inference in variable dispersion beta regressions', *Journal of Statistical Computation and Simulation,* **a aparecer**. 3

Cribari-Neto, F. e Zarkos, S. G. (2003), 'Econometric and statistical computing using ox', *Computational Economics* **21**, 277295. 15

Cribari-Neto, F. e Zeileis, A. (2010), 'Beta regression in r', *Journal of Statistical Software* **34**(2), 1–24. 3

Doornik, J. A. (2007), *An Object-Oriented Matrix Language: Ox 5*, Timberlake Consultants Press, London. 15

Dow, W. H. e Norton, E. C. (2003), 'Choosing between and interpreting the heckit and two-part models for corner solutions', *Health Services & Outcomes Research Methodology* **4**, 5–18. 5

Dunn, P. K. e Smyth, G. K. (1996), 'Randomized quantile residuals', *Journal of Computational and Graphical Statistics* **5**(3), 236–244. 54

Espinheira, P. L., Ferrari, S. L. P. e Cribari-Neto, F. (2008*a*), 'Influence diagnostics in beta regression regression', *Computational Statistics and Data Analysis* **52**(9), 4417–4431. 3, 56

Espinheira, P. L., Ferrari, S. L. P. e Cribari-Neto, F. (2008*b*), 'On beta regression residuals', *Journal of Applied Statistics* **35**(4), 407–419. 3, 49, 53, 54

Fagerland, M. W., Hosmer, D. W. e Bofin, A. M. (2008), 'Multinomial goodness-of-fit tests for logistic regression models', *Statistics in Medicine* **27**, 4238–4253. 52, 53, 55

Ferrari, S. L. P. e Cribari-Neto, F. (2004), 'Beta regression for modelling rates and proportions', *Journal of Applied Statistics* **31**(7), 799–815. 2, 8, 25, 29, 49, 50

Ferrari, S. L. P., Espinheira, P. L. e Cribari-Neto, F. (2011), 'Diagnostic tools in beta regression with varying dispersion', *Statistica Neerlandica* **65**(3), 337–351. 49, 50, 56, 62

Ferrari, S. L. P. e Pinheiro, E. C. (2011), 'Improved likelihood inference in beta regression', *Journal of Statistical Computation and Simulation* **81**(4), 431–443. 3

Feuerverger, A. (1979), 'On some methods of analysis for weather experiments', *Biometrika* **66**(3), 655–658. 4

Galea, M., Paula, G. A. e Bolfarine, H. (1997), 'Local influence in elliptical linear regression models', *Journal of the Royal Statistical Society: Series D (The Statistician)* **46**(1), 71–79. 56

Giampaoli, V., Raposo, M. C. F. e Torres, S. T. F. (2009), 'Critérios de seleção de modelos para o modelo de regressão beta', *Revista Brasileira de Estatística* **70**, 7–27. 46

Goeman, J. J. e le Cessie, S. (2006), 'A goodness-of-fit test for multinomial logistic regression', *Biometrics* **62**(4), 980–985. 52

Greene, W. H. (1994), Accounting for excess zeros and sample selection in poisson and negative binomial regression models, Working papers ec-94-10, Department of Economics, New York University. 4

Gupta, A. K., Kasturiratna, D., Nguyen, T. e Pardo, L. (2006), 'A new family of ban estimators in polytomous regression models based on ϕ-divergence measures', *Statistical Methods and Applications* **15**(2), 159–176. 51

Gupta, A. K. e Nadarajah, S. (2004), *Handbook of beta distribution and its applications*, Marcel Dekker, New York. 8

Gupta, A. K., Nguyen, T. e Pardo, L. (2008), 'Residuals for polytomous logistic regression models based on ϕ-divergences test statistics', *Statistics* **42**(6), 495–514. 51

Hall, D. B. (2000), 'Zero-inflated poisson and binomial regression with random effects: a case study', *Biometrics* **56**, 1030–1039. 4

Hamilton, R. e Khan, M. (2001), 'Revolving credit card holders: who are they and how can be identified', *The Services Industries Journal* **21**(3), 37–48. 5

Hardin, J. W. e Hilbe, J. M. (2003), *Generalized estimating equations*, Chapman and Hall, Boca Raton. 3

Heller, G., Stasinopoulos, M. e Rigby, B. (2006), The zero-adjusted inverse gaussian distribution as a model for insurance claims, *in* '21th International Workshop on Statistical Modelling', Galway:Ireland. 226-233. 4

Hermalin, B. E. e Wallace, N. E. (1994), 'The determinants of efficiency and solvency in saving and loans', *The RAND Journal of Economics* **25**(3), 361–381. 1

Hoff, A. (2007), 'Second stage dea: Comparison of approaches for modelling the dea escore', *European Journal of Operational Research* **181**, 425–435. 4

Hosmer, D. W. e Lemeshow, S. (1980), 'Second stage dea: Comparison of approaches for modelling the dea escore', *Communications in Statistics - Theory and Methods* **9**(10), 1043–1069. 52

Hosmer, D. W. e Lemeshow, S. (2000), *Applied logistic regression*, second edn, John Wiley and Sons, New York. 24

Johnson, N. L., Kotz, S. e Balakrishnan, N. (1995), *Continuous univariate distributions*, second edn, John Wiley and Sons, New York. 2, 8

Johnson, R. A. e Wichern, D. W. (2007), *Applied Multivariate Statistical Analysis*, 6 edn, Prentice Hall, Upper Saddle River. 14, 33

Jonsson, R. (2012), 'When does heckman´s two step-procedure for censored data work and when does it not?', *Statistical Papers* **53**, 33–49. 5

Jorgensen, B. (1997*a*), 'Proper dispersion models (with discussion)', *Brazilian Journal of Probability and Statistics* **11**(2), 89–140. 2

Jorgensen, B. (1997*b*), *The theory of dispersion models*, Chapman and Hall, London. 2

Kay, R. (1986), 'A markov model for analysing cancer markers and disease states in survival studies', *Biometrics* **42**(4), 855–865. 6

Kieschnick, R. e McCullough, B. D. (2003), 'Regression analysis of variates observed on (0,1): percentages, proportions and fractions', *Statistical Modelling* **3**, 193–213. 1, 2

Kim, H. e DeVaney, S. A. (2001), 'The determinants of outstanding balances among credit card revolvers', *Financial Counseling and Planning* **12**(1), 67–77. 5

Kumaraswamy, P. (1980), 'A generalized probability density function for double-bounded random processes', *Journal of Hydrology* **46**, 79–88. 2

Kutner, M., Nachtsheim, C., Neter, J. e Li, W. (2004), *Applied linear statistical models*, 5 edn, McGraw-Hill/Irwin, New York. 47

Lambert, D. (1992), 'Zero-inflated poisson regression, with an application to defects in manufacturing', *Technometrics* **34**, 1–14. 4

Lehmann, E. L. e Casella, G. (1998), *Theory of point estimation*, 2 edn, Springer-Verlag, New York. 12, 13, 80, 82

Lesaffre, E. e Albert, A. (1989), 'Multiple-group logistic regression diagnostics', *Journal of the Royal Statistical Society Series C (Applied Statistics)* **38**(3), 425–440. 50, 51

Lesaffre, E., Rizopoulos, D. e Tsonaka, R. (2007), 'The logistic transform for bounded outcome scores', *Biostatistics* **8**(1), 72–85. 4

Lesaffre, E. e Verbeke, G. (1998), 'Local influence in linear mixed models', *Biometrics* **54**(2), 570–582. 56

Maddala, G. S. (1983), *Limited-dependent and qualitative variables in econometric*, second edn, Cambridge University Press, Cambridge. 5

Maddala, G. S. (1986), *Limited-dependent and qualitative variables in econometrics*, Cambridge University Press, Cambridge. 3

McCullagh, P. e Nelder, J. A. (1989), *Generalized linear models*, second edn, Chapman and Hall, London. 2

McQuarrie, A. D. R. e Tsai, C. L. (1998), *Regression and times series model selection*, World Scientific Publishing Company, Singapore. 2, 48

Melo, T. F. N., Vasconcellos, K. L. P. e Lemonte, A. J. (2009), 'Some restriction tests in a new class of regression models for proportions', *Computational Statistics and Data Analysis* **53**(12), 39723979. 3

Miller, A. (2002), *Subset selection in regression*, 2 edn, Chapman and Hall, Boca Raton. 47

Mitnik, P. A. e Baek, S. (2012), 'The kumaraswamy distribution: median-dispersion reparameterizations for regression modeling and simulation-based estimation', *Statistical Papers,* **a aparecer**. 2

Miyashiro, E. S. (2008), Modelos de regressão beta e simplex para análise de proporções, Dissertação de mestrado em estatística, Instituto de Matemática e Estatística - Universidade de São Paulo. 2, 3

Moura, C. S. e Istake, M. (2007), Principais determinantes do valor do rendimento das aposentadorias no brasil e no paraná para o ano de 2004, *in* 'V Ecopar'. 6

Ngoc, T. M. (2011), Some perspectives on the problem of model selection, Tese de doutorado em estatística, National University of Singapore. 48

Nocedal, J. e Wright, S. J. (2006), *Numerical Optimization*, 2 edn, Springer, New York. 13, 28

Osius, G. e Rojek, D. (1992), 'Normal goodness-of-fit tests for multinomial models with large degrees of freedom', *Journal of the American Statistical Association* **87**, 1145–1152. 52

Ospina, R. (2008), Modelos de regressão beta inflacionados, Tese de doutorado em estatística, Instituto de Matemática e Estatística - Universidade de São Paulo. 1, 4, 6, 8, 25, 54, 56, 88

Ospina, R., Cribari-Neto, F. e Vasconcellos, K. L. P. (2006), 'Improved point and interval estimation for a beta regression model', *Computational Statistics and Data Analysis* **51**(2), 960–981. 3

Ospina, R., Cribari-Neto, F. e Vasconcellos, K. L. P. (2011), 'Erratum to improved point and interval estimation for a beta regression model [comput. statist. data anal. 51 (2006) 960–981]', *Computational Statistics and Data Analysis* **55**(7), 2445. 3

Ospina, R. e Ferrari, S. L. P. (2010), 'Inflated beta distributions', *Statistical Papers* **51**(1), 111–126. 4, 9, 11, 13

Ospina, R. e Ferrari, S. L. P. (2012), 'A general class of zero-or-one inflated beta regression models', *Computational Statistics and Data Analysis,* **56**(6), 1609–1623. 5, 53

Ouwens, M. J. N. M., Tan, F. E. S. e Berger, M. P. F. (2001), 'Local influence to detect influential data structures for generalized linear mixed models', *Biometrics* **57**(4), 1166–1172. 56

Pace, L. e Salvan, A. (1997), *Principles of Statistical Inference from a Neo-Fisherian Perspective*, World Scientific, Singapore. 12

Paolino, P. (2001), 'Maximum likelihood estimation of models with beta-distributed dependent variables', *Political Analysis* **9**(4), 325–346. 2

Papke, L. e Wooldridge, J. (1996), 'Econometric methods for fractional response variables with an application to 401(k) plan participation rates', *Journal of Applied Econometrics* **11**, 619–632. 1, 3, 4

Pereira, T. L. (2010), Regressão beta inflacionada: inferência e aplicações, Tese de doutorado em estatística, Universidade Federal de Pernambuco. 5

Piepho, H.-P. (2003), 'The folded exponential transformation for proportions', *Journal of the Royal Statistical Society: Series D (The Statistician)* **52**(4), 575–589. 3

Pigeon, J. G. e Heyse, J. F. (1999), 'An improved goodness of fit statistic for probability prediction models', *Biometrical Journal* **41**, 71–82. 52

Pregibon, D. (1981), 'Logistic regression diagnostics', *The Annals of Statistics* **9**(4), 705–724. 51

Ramalho, J. J. S. e Silva, J. V. (2009), 'A two-part fractional regression model for the financial leverage decisions of micro, small, medium and large firms', *Quantitative finance* **9**(5), 621–636. 3

Régis, D. E. e Artes, R. (2008), Modelo multi-estado de markov em cartões de crédito, Technical report, Insper. 6

Ridout, M., Demétrio, C. G. B. e Hinde, J. (1998), Models for count data with many zeros, *in* 'International Biometric Conference - IBC 98, Cape Town', Cape Town, África do Sul. 179-192. 4

Rigby, R. A. e Stasinopoulos, D. M. (2005), 'Generalized additive models for location, scale and shape', *Journal of the Royal Statistics Society: Series C (Applied Statistics)* **54**(3), 507–554. 25, 46

Riquelme, M. A. (2012), Modelos mistos com erros nas variáveis, Tese de doutorado em estatística, Instituto de Matemática e Estatística - Universidade de São Paulo. 56

Robb, C. A. e Sharpe, D. L. (2009), 'Effect of personal financial knowledge on college students credit card behavior', *Journal of Financial Counseling and Planning* **20**(1), 25–43. 5

Russo, C. M., Paula, G. A. e Aoki, R. (2009), 'Influence diagnostics in nonlinear mixed-effects elliptical models', *Computational Statistics and Data Analysis* **53**(12), 4143–4156. 56

Schwarz, G. (1978), 'Estimating the dimension of a model', *Annals of Statistics* **6**(2), 461–464. 46

Simas, A. B., Barreto-Souza, W. e Rocha, A. V. (2010), 'Improved estimators for a general class of beta regression models', *Computational Statistics and Data Analysis* **54**, 348–366. 3

Smithson, M. e Verkuilen, J. (2006), 'A better lemon squeezer? maximum likelihood regression with beta-distributed dependent variables', *Psychological Methods* **11**, 54–71. 3

Souza, D. F. (2011), Regressão beta multivariada com aplicações em pequenas áreas, Tese de doutorado em estatística, Instituto de Matemática - Universidade Federal do Rio de Janeiro. 3

Stephens, M. A. (1986), Tests based on edf statistics, *in* R. B. D´Agostino e M. A. Stephens, eds, 'Goodness-of-fit techniques', Marcel Dekker, New York. 21

Stock, J. H. e Watson, M. W. (2003), *Introduction to econometrics*, Addison Wesley, Boston. 7, 36, 38

Terrell, G. R. (2002), 'The gradient statistics', *Computing Science and Statistics* **34**, 206–215. 35

Toder, E., Thompson, L., Fraveault, M., Johnson, R., Perese, K., Ratcliffe, C., Smith, K., Uccello, C., Waidmann, T., Berk, J., Woldemariam, R., Burtless, G., Sahm, C. e Wolf, D. (2002), Modeling income in the near term: revised projections of retirement income through 2020 for the 1931-1960 birth cohorts, Relatório técnico, The Urban Institute. 6

Venezuela, M. K. (2008), Equação de estimação generalizada e influência local para modelos de regressão beta com medidas repetidas, Tese de doutorado em estatística, Instituto de Matemática e Estatística - Universidade de São Paulo. 3, 56

Venezuela, M. K., Sandoval, M. C. e Botter, D. A. (2011), 'Local influence in estimating equations', *Computational Statistics and Data Analysis* **55**(4), 1867–1883. 56

Vieira, A. M. C., Hinde, J. P. e Demétrio, C. G. B. (2000), 'Zero inflated proportion data models applied to a biological control assay', *Journal of Applied Statistics* **27**(3), 373–389. 4

Wedderburn, R. W. M. (1974), 'Quasi-likelihood function, generalized linear models, and the gauss-newton method', *Biometrika* **61**(3), 439–447. 1

Zhao, Y., Lee, A. H. e Hui, Y. V. (1994), 'Influence diagnostics for generalized linear measurement error models', *Biometrics* **50**(4), 1117–1128. 56