



Universidade Federal de Pernambuco
Centro de Ciências Exatas e da Natureza
Departamento de Estatística

Pós-Graduação em Estatística

Modelagem e Inferência em Regressão Beta

Fábio Mariano Bayer

Tese de Doutorado

Recife
Outubro de 2011

Universidade Federal de Pernambuco
Centro de Ciências Exatas e da Natureza
Departamento de Estatística

Fábio Mariano Bayer

Modelagem e Inferência em Regressão Beta

Trabalho apresentado ao Programa de Pós-Graduação em Estatística do Departamento de Estatística da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Doutor em Estatística.

Orientador: *Prof. Dr. Francisco Cribari Neto*

Recife
Outubro de 2011

Catálogo na fonte
Bibliotecária Jane Souto Maior, CRB4-571

Bayer, Fábio Mariano
Modelagem e inferência em regressão beta / Fábio
Mariano Bayer - Recife: O Autor, 2011.
xiii, 115 folhas : il., fig., tab.

Orientador: Francisco Cribari Neto.
Tese (doutorado) - Universidade Federal de Pernambuco.
CCEN, Estatística, 2011.

Inclui bibliografia e apêndice.

1. Estatística. 2. Estatística Matemática. I. Cribari Neto,
Francisco (orientador). II. Título.

310

CDD (22. ed.)

MEI2011 – 156

*Dedico esta tese ao meu pai, Arlei João Bayer, e à minha
esposa, Débora.*

Agradecimentos

Ao meu orientador, Francisco Cribari Neto, pela disponibilidade, paciência, valiosas sugestões e pela orientação constante e cuidadosa. Mais do que um ótimo professor, o professor Cribari é um pesquisador com competência, dedicação e ética a serem seguidas e admiradas.

À minha esposa, Débora, pelo amor, carinho, paciência e companheirismo nesta etapa de tantas privações. Obrigado por me ajudar e estar sempre ao meu lado!

Às minhas grandes amigas Tatiene e Tarciana, pela acolhida, pelas conversas acadêmicas e pelos ótimos momentos de descontração. Sou eternamente grato pelo carinho de vocês.

Aos grandes amigos Marcus, Renata e sua filhinha Mariana. Muito obrigado por nos adotarem em Recife. Certamente criamos vínculos eternos de amor e amizade.

À Valéria, secretária do programa, não só pela imensa competência mas principalmente pelo carinho e amizade.

Ao professor Renato Cintra, pelas oportunidades, discussões e colaborações científicas. O professor Renato não colaborou diretamente neste texto, mas também o considero como orientador. Muito obrigado pelos ensinamentos, confiança e amizade.

Ao professor Gauss Cordeiro, pela importante contribuição dada ao desenvolvimento desta tese.

Ao professor Raydonal Ospina, pela amizade e pelas grandes sugestões e colaborações feitas a este trabalho.

Ao meu amigo Hemílio, pela acolhida em Recife.

Ao meu colega e amigo Abraão e aos demais colegas do Programa de Pós-Graduação em Estatística da UFPE. Muito obrigado pelos momentos compartilhados.

Aos amigos Carla e Alexandre, pela acolhida, amizade e inesquecíveis momentos de descontração que passamos juntos.

À minha família e à família da Débora por estarem sempre presentes.

À todos os professores do Departamento de Estatística da UFPE, em especial aos professores Francisco Cribari-Neto, Renato Cintra, Klaus Leite Pinto Vasconcelos, Gauss Cordeiro, Getúlio Amaral, Francisco Cysneiros, Raydonal Ospina e Patrícia Espinheira Ospina. Muito obrigado pelos ensinamentos.

Aos funcionários do Departamento de Estatística da UFPE, em especial ao Jimmy e à Valéria.

Aos meus colegas da UFFS, em especial aos amigos da “sala 1”.

Ao Programa de Pós-Graduação em Estatística da UFPE, pela oportunidade concedida.

À CAPES, pelo apoio financeiro.

Aos participantes da banca examinadora.

Resumo

Esta tese aborda aspectos de modelagem e inferência em regressão beta, mais especificamente melhoramentos do teste de razão de verossimilhanças e proposição e investigação de critérios de seleção de modelos. O modelo de regressão beta foi proposto por Ferrari e Cribari-Neto [2004. Beta regression for modeling rates and proportions. *J. Appl. Statist.* 31, 799–815] para modelar variáveis contínuas no intervalo $(0,1)$, como taxas e proporções. No primeiro capítulo, abordamos o problema de inferência em pequenas amostras. Focamos no melhoramento do teste da razão de verossimilhanças. Consideramos correções de segunda ordem para a estatística da razão de verossimilhanças em regressão beta em duas abordagens. Determinamos, por meio de uma abordagem matricial, o fator de correção de Bartlett e também uma correção de Bartlett Bootstrap. Comparamos os testes baseados nas estatísticas corrigidas com o teste da razão de verossimilhanças usual e com o teste que utiliza o ajuste de Skovgaard, que já está proposto na literatura. Os resultados numéricos evidenciam que as correções de Bartlett são mais acuradas do que a estatística não corrigida e do que o ajuste de Skovgaard. No segundo e terceiro capítulos, expandimos o modelo de regressão beta proposto por Ferrari e Cribari-Neto, considerando um modelo que assume que o parâmetro de dispersão, assim como o parâmetro de média, varia ao longo das observações e pode ser modelado por meio de uma estrutura de regressão. Com isso, surge o problema da seleção de variáveis, tanto para a estrutura da média quanto para a da dispersão. Esse assunto é tratado em dois capítulos independentes e auto-contidos, porém, ambos relacionados. No Capítulo 2 propomos critérios de seleção para modelos com dispersão variável e investigamos, por meio de simulação de Monte Carlo, os desempenhos destes e de outros critérios de seleção em amostras de tamanho finito. Percebemos que o processo de seleção conjunta de regressores para a média e para a dispersão não é uma boa prática e propomos um esquema de seleção em duas etapas. A seleção de modelos com o esquema proposto, além de requerer um menor custo computacional, apresentou melhor desempenho do que o método usual de seleção. Dentre os critérios investigados encontra-se o critério de informação de Akaike (AIC). O AIC é, sem dúvida, o critério mais conhecido e aplicado em diferentes classes de modelos. Baseados no AIC diversos critérios têm sido propostos, dentre eles o SIC, o HQ e o AICc. Com o objetivo de estimar o valor esperado da log-verossimilhança, que é uma medida de discrepância entre o modelo verdadeiro e o modelo candidato estimado, Akaike obtém o AIC como uma correção assintótica para a log-verossimilhança esperada. No entanto, em pequenas amostras, ou quando o número de parâmetros do modelo é grande relativamente ao tamanho amostral, o AIC se torna viesado e tende a selecionar modelos com alta dimensionalidade. Ao considerarmos uma estrutura de regressão também para o parâmetro de dispersão introduzimos um maior número de parâmetros a serem estimados no modelo. Isso pode diminuir o desempenho dos critérios de seleção quando o tamanho amostral é pequeno ou moderado. Para contornar esse problema

propomos no Capítulo 3 novos critérios de seleção para serem usados em pequenas amostras, denominados *bootstrap likelihood quasi-CV* (BQCV) e sua modificação 632QCV. Comparamos os desempenhos dos critérios propostos, do AIC e de suas diversas variações que utilizam log-verossimilhança bootstrap por meio de um extensivo estudo de simulação. Os resultados numéricos evidenciam o bom desempenho dos critérios propostos.

Palavras-chave: AIC, ajustes para pequenas amostras, bootstrap, correção de Bartlett, critérios de seleção de modelos, dispersão variável, regressão beta, teste da razão de verossimilhanças

Abstract

This doctoral dissertation deals with beta regression modeling and inferences. More specifically, we aim at obtaining improved variants of the likelihood ratio test and proposing and evaluating model selection criteria. The beta regression model was proposed by Ferrari and Cribari-Neto [2004. Beta regression for modeling rates and proportions. *J. Appl. Statist.* 31, 799–815] to model continuous variates that assume values in $(0,1)$, such as rates and proportions. In the first chapter, we focus on testing inference in small samples. We derive a Bartlett correction to the likelihood ratio test statistic and also consider a Bootstrap Bartlett correction. Using Monte Carlo simulations we compare the finite sample performance of the corrected tests to that of the standard likelihood ratio test and also to the adjusted test obtained by using Skovgaard's correction; the latter is already available in the literature. The numerical evidence favors the corrected tests we propose. In the next two chapters we consider the beta regression model with varying dispersion. The model consists of two submodels: one for the mean and one for the dispersion. In Chapter 2, we propose model selection criteria for varying dispersion beta regression. Our Monte Carlo simulation results show that joint model selection performs poorly when the sample size is not large. We then propose a two-step model selection strategy. The finite sample behavior of the proposed criteria is evaluated numerically and contrasted to that of Akaike's Information Criterion (AIC) and some of its variants. The AIC is the most commonly used criterion. It was obtained as an asymptotic correction to the expected log-likelihood. When the number of observations is not large and/or the number of parameters is large, the AIC tends to select models that are overparametrized. The problem thus becomes more serious when we add a dispersion submodel to the beta regression model. In Chapter 3, we propose two new model selection criteria: bootstrap likelihood quasi-CV (BQCV) and its 632QCV variant. We compare the finite sample merits of different model selection criteria in varying dispersion beta regression. The numerical evidence shows that the proposed criteria lead to accurate model selection.

Keywords: AIC, Bartlett correction, beta regression, bootstrap, likelihood ratio test, model selection criteria, small sample adjustments, variable dispersion

Sumário

1	Teste da Razão de Verossimilhanças Corrigido em Regressão Beta	1
	Resumo do capítulo	1
1.1	Introdução	1
1.2	O modelo de regressão beta	4
1.3	Correções do teste da razão de verossimilhanças	6
1.3.1	Correção de Bartlett em forma matricial	7
1.3.2	Correção de Bartlett bootstrap	9
1.3.3	Ajuste de Skovgaard	9
1.4	Avaliação numérica	10
1.5	Aplicação	18
1.6	Conclusões	20
2	Modelos de Regressão Beta com Dispersão Variável: Avaliação de Critérios de Seleção de Modelos em Amostras de Tamanho Finito	21
	Resumo do capítulo	21
2.1	Introdução	21
2.2	Definição e estimação do modelo	24
2.3	Seleção de modelos	27
2.3.1	Critérios de seleção usuais	28
2.3.2	Critérios de seleção para modelos com dispersão variável	30
2.4	Avaliação numérica	31
2.5	Aplicação	42
2.6	Conclusões	46
3	Variações Bootstrap do AIC em Regressão Beta	48
	Resumo do capítulo	48
3.1	Introdução	48
3.2	Critério de informação de Akaike e variações bootstrap	50
3.2.1	Extensões bootstrap do AIC	52
3.2.2	Verossimilhança bootstrap e validação cruzada	53
3.2.3	Verossimilhança bootstrap e <i>quasi</i> -CV	54
3.3	O modelo de regressão beta	55
3.4	Avaliação numérica	57
3.4.1	Critérios de seleção de modelos em regressão beta com dispersão constante	57

3.4.2	Critérios de seleção de modelos em regressão beta com dispersão variável	63
3.5	Aplicação	70
3.6	Conclusões	72
4	Considerações Finais	74
4.1	Conclusões	74
4.2	Trabalhos futuros	75
A	Cumulantes para o Fator de Correção de Bartlett	76
B	Correção de Bartlett (implementação em R)	82
C	Comparação dos testes baseados nas estatísticas LR_{b3} e LR_{sk1}	91
D	Função Escore para o Modelo de Regressão Beta com Dispersão Variável	96
E	Matriz de Informação de Fisher para o Modelo de Regressão Beta com Dispersão Variável	97
F	Avaliação dos Testes para Dispersão Variável	98
G	Análise dos Resíduos (implementação em R)	101
H	Tabelas de Resultados Complementares	105
	Referências Bibliográficas	111

Lista de Figuras

1.1	Densidades beta para diferentes valores de μ (indicados no gráfico), com $\phi = 10$ (a) e $\phi = 90$ (b).	2
1.2	Gráfico Q-Q considerando $\phi = 10$ e $q = 1$.	16
1.3	Densidade χ_q^2 (linha sólida) e densidades nulas estimadas das estatísticas de teste para $\phi = 5$ e $q = 3$.	17
2.1	Densidades estimadas dos estimadores do parâmetro de inclinação sob dispersão variável, com (linha cheia) e sem (linha tracejada) modelagem da dispersão.	23
2.2	Densidades beta para diferentes valores de σ (indicados no gráfico), com $\mu = 0.25$ (a) e $\mu = 0.5$ (b).	25
2.3	Porcentagem (%) de seleção correta dos submodelos da média e dispersão conjuntamente para os quatro critérios com melhores desempenhos.	34
2.4	Porcentagem (%) de seleção correta dos regressores para a dispersão com submodelo da média corretamente especificado para os quatro critérios com melhores desempenhos.	36
2.5	Porcentagem (%) de seleção correta dos regressores para a média com submodelo da dispersão corretamente especificado para os três critérios com melhores desempenhos.	38
2.6	Porcentagem (%) de seleção correta dos regressores para a média supondo dispersão constante para os três critérios com melhores desempenhos.	40
2.7	Comparação entre os melhores resultados do esquema de duas etapas proposto e da seleção conjunta.	43
2.8	Boxplot dos dados de habilidade de leitura.	44
2.9	Gráficos de resíduos para o modelo ajustado.	46
3.1	Gráficos dos resíduos para o modelo ajustado.	71
C.1	Gráfico Q-Q para $n = 15$ e diferentes valores de ϕ e q , considerando simulações com 50000 réplicas de Monte Carlo.	93
C.2	Densidade χ_q^2 (linha sólida) e densidades nulas estimadas das estatísticas de teste para $n = 15$ e $q = 3$, considerando 50000 réplicas de Monte Carlo.	94
C.3	Densidade χ_q^2 (linha sólida) e densidades nulas estimadas das estatísticas de teste para $n = 15$ e $q = 3$, considerando 50000 réplicas de Monte Carlo.	95

Lista de Tabelas

1.1	Taxas de rejeição nula (percentuais) do teste $\mathcal{H}_0 : \beta_2 = 0$ ($q = 1$).	12
1.2	Taxas de rejeição nula (percentuais) do teste $\mathcal{H}_0 : \beta_2 = \beta_3 = 0$ ($q = 2$).	13
1.3	Taxas de rejeição nula (percentuais) do teste $\mathcal{H}_0 : \beta_2 = \beta_3 = \beta_4 = 0$ ($q = 3$).	14
1.4	Momentos e quantis estimados das estatísticas de teste, considerando $q = 2$, $\phi = 30$ e $n = 20$.	15
1.5	Taxas de rejeição não-nula (percentuais) dos testes corrigidos para $n = 20$ e $\alpha = 5\%$.	18
2.1	Valores dos parâmetros considerados nas simulações.	31
2.2	Porcentagem (%) de seleção correta dos submodelos da média e dispersão conjuntamente.	33
2.3	Porcentagem (%) de seleção correta dos regressores para a dispersão com submodelo da média corretamente especificado.	35
2.4	Porcentagem (%) de seleção correta dos regressores para a média com submodelo da dispersão corretamente especificado.	37
2.5	Porcentagem (%) de seleção correta dos regressores para a média supondo dispersão constante.	39
2.6	Porcentagem (%) de seleção correta dos regressores da média e dispersão utilizando o esquema de seleção proposto, para diferentes critérios em cada etapa.	42
2.7	Estimativas dos parâmetros do modelo de regressão beta com dispersão variável para dados de habilidade de leitura.	45
3.1	Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão constante, considerando grande dispersão e modelo facilmente identificável (modelo (3.8)).	59
3.2	Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão constante, considerando grande dispersão e modelo fracamente identificável (modelo (3.9)).	60
3.3	Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão constante, considerando pequena dispersão e modelo facilmente identificável (modelo (3.10)).	61
3.4	Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão constante, considerando pequena dispersão e modelo fracamente identificável (modelo (3.11)).	62

3.5	Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, selecionando regressores da média e dispersão conjuntamente em um modelo facilmente identificável (modelo (3.14)).	64
3.6	Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, selecionando regressores da média e dispersão conjuntamente em um modelo fracamente identificável (modelo (3.15)).	65
3.7	Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, considerando conhecida a estrutura de regressão da dispersão e selecionando regressores da média em um modelo facilmente identificável (modelo (3.14)).	66
3.8	Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, considerando conhecida a estrutura de regressão da dispersão e selecionando regressores da média em um modelo fracamente identificável (modelo (3.15)).	67
3.9	Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, considerando conhecida a estrutura de regressão da média e selecionando regressores da dispersão em um modelo facilmente identificável (modelo (3.14)).	68
3.10	Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, considerando conhecida a estrutura de regressão da média e selecionando regressores da dispersão em um modelo fracamente identificável (modelo (3.15)).	69
3.11	Estimativas dos parâmetros do modelo de regressão beta com dispersão variável para dados de gastos com alimentação.	70
C.1	Taxas de rejeição nula (percentuais) dos testes da razão de verossimilhanças original e corrigidos.	92
F.1	Taxas de rejeição nula (percentuais) do teste $\mathcal{H}_0 : \gamma_2 = \gamma_3 = 0$.	99
F.2	Taxas de rejeição não-nula (percentuais) dos testes, considerando a hipótese alternativa $\gamma_2 = -\gamma_3 = \delta$.	100
F.3	Taxas de rejeição não-nula (percentuais) dos testes, considerando a hipótese alternativa $\gamma_2 = \delta, \gamma_3 = 0$.	100
H.1	Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, selecionando regressores da média e dispersão conjuntamente em um modelo facilmente identificável (modelo (3.12)).	105
H.2	Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, selecionando regressores da média e dispersão conjuntamente em um modelo fracamente identificável (modelo (3.13)).	106

H.3	Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, considerando conhecida a estrutura de regressão da dispersão e selecionando regressores da média em um modelo facilmente identificável (modelo (3.12)).	107
H.4	Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, considerando conhecida a estrutura de regressão da dispersão e selecionando regressores da média em um modelo fracamente identificável (modelo (3.13)).	108
H.5	Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, considerando conhecida a estrutura de regressão da média e selecionando regressores da dispersão em um modelo facilmente identificável (modelo (3.12)).	109
H.6	Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, considerando conhecida a estrutura de regressão da média e selecionando regressores da dispersão em um modelo fracamente identificável (modelo (3.13)).	110

Teste da Razão de Verossimilhanças Corrigido em Regressão Beta

Resumo do capítulo

Regressão beta é uma importante classe de modelos usada para modelar variáveis aleatórias contínuas no intervalo $(0,1)$, como taxas e proporções. Consideramos o problema de inferência em pequenas amostras e derivamos correções de Bartlett para a estatística da razão de verossimilhanças. Determinamos o fator de correção de Bartlett em forma matricial e a correção de Bartlett bootstrap. Utilizando simulação de Monte Carlo, comparamos o desempenho do teste da razão de verossimilhanças corrigido com o do teste não corrigido e o do teste que utiliza o ajuste de Skovgaard, que se encontra proposto na literatura. Os resultados de simulação evidenciam que as correções propostas diminuem distorções de tamanho do teste da razão de verossimilhanças, melhorando as inferências em pequenas amostras. Em termos comparativos, a correção de Bartlett se mostrou mais acurada do que o ajuste de Skovgaard em modelos de regressão beta. Apresentamos também uma aplicação a dados reais.

Palavras-chave: Bootstrap, correção de Bartlett, regressão beta, teste da razão de verossimilhanças.

1.1 Introdução

A análise de regressão é uma técnica estatística utilizada para investigar e modelar, com base em um banco de dados, a relação entre uma variável de interesse e um conjunto de variáveis explicativas. O modelo de regressão normal linear é bastante utilizado em análises empíricas. No entanto, tal modelo torna-se inapropriado em situações em que a variável resposta é restrita ao intervalo $(0,1)$, como ocorre com taxas e proporções. Para essas situações Ferrari e Cribari-Neto (2004) propuseram a classe de modelos de regressão beta, em que a variável resposta (y) possui distribuição beta. A estrutura e procedimentos inferenciais do modelo são similares aos dos modelos lineares generalizados (McCullagh e Nelder, 1989).

A densidade beta é dada por

$$\pi(y; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1} (1-y)^{q-1}, \quad 0 < y < 1,$$

em que $p > 0$ e $q > 0$ são parâmetros que indexam a distribuição e $\Gamma(\cdot)$ é a função gama:

$$\Gamma(p) = \int_0^{\infty} y^{p-1} e^{-y} dy.$$

No modelo de regressão proposto em Ferrari e Cribari-Neto (2004) é utilizada uma reparametrização da densidade beta, sendo $\mu = p/(p+q)$ e $\phi = p+q$, isto é, $p = \mu\phi$ e $q = (1-\mu)\phi$. Com isso, a função de densidade é

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1}, \quad 0 < y < 1, \quad (1.1)$$

e segue que

$$E(y) = \mu, \quad \text{var}(y) = \frac{V(\mu)}{1+\phi},$$

em que $V(\mu) = \mu(1-\mu)$ denota a função de variância e ϕ pode ser interpretado como um parâmetro de precisão. A distribuição beta é bastante flexível para modelar proporções. Dependendo dos valores dos dois parâmetros que a indexam, a densidade assume formas bem variadas, acomodando distribuições simétricas, assimétricas, em formas de J e de J invertido. Isso é verificado pelas diversas densidades beta apresentadas na Figura 1.1.

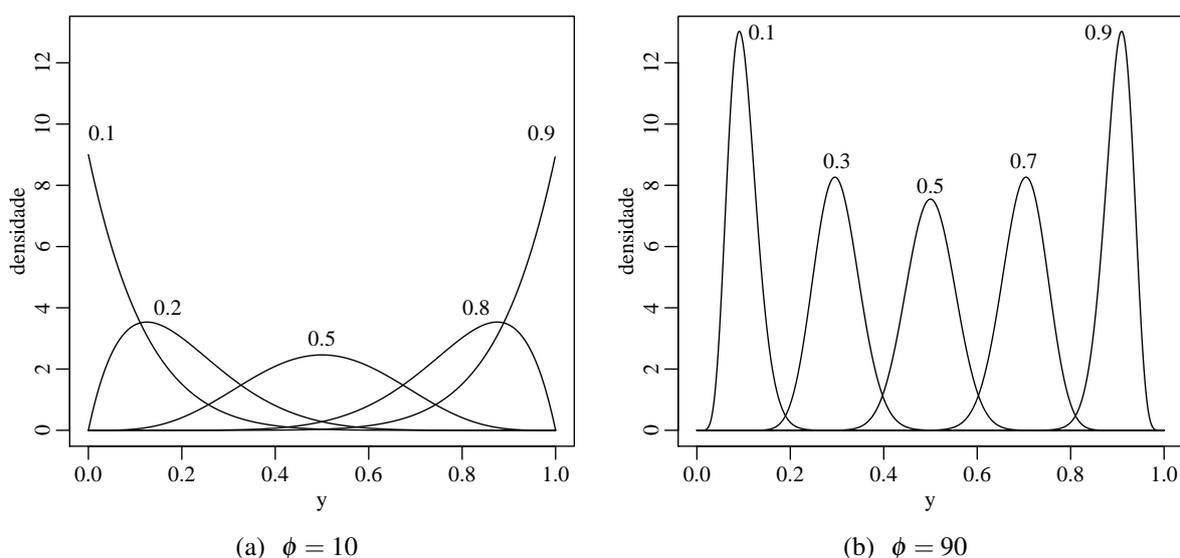


Figura 1.1 Densidades beta para diferentes valores de μ (indicados no gráfico), com $\phi = 10$ (a) e $\phi = 90$ (b).

A classe de modelos de regressão beta tem como objetivo permitir a modelagem de respostas que pertencem ao intervalo $(0,1)$, por meio de uma estrutura de regressão que contém uma função de ligação, covariáveis e parâmetros desconhecidos. Muitos estudos, em diferentes áreas do conhecimento, como em Brehm e Gates (1993), Hancox *et al.* (2010), Kieschnick e McCullough (2003), Smithson e Verkuilen (2006) e Zucco (2008), utilizam regressão beta ou outras abordagens para examinar como um conjunto de covariáveis se relaciona com alguma porcentagem ou proporção.

Detalhes sobre inferências em grandes amostras e análise de diagnóstico nessa classe de modelos podem ser encontrados em Espinheira *et al.* (2008a,b). Melhoramentos em estimação pontual e intervalar são apresentados por Ospina *et al.* (2006). Em Simas *et al.* (2010)

apresenta-se uma generalização do modelo de regressão beta, considerando uma estrutura de regressão para o parâmetro de precisão e modelos não-lineares. Simas *et al.* também obtêm correções analíticas de viés para os estimadores de máxima verossimilhança, generalizando os resultados de Ospina *et al.* (2006). Uma discussão a respeito de modelagem de regressão beta no sistema \mathbb{R} é apresentada com detalhes por Cribari-Neto e Zeileis (2010). Modelos de regressão beta inflacionados, que acomodam dados que contêm zeros e/ou uns, são tratados como extensões do modelo de regressão beta em Ospina (2008), Ospina e Ferrari (2011) e Pereira (2010).

Com o objetivo de fazer inferências sobre os parâmetros do modelo de regressão beta pode-se utilizar testes da razão de verossimilhanças, escore ou Wald. Tradicionalmente o teste da razão de verossimilhanças é o que tem merecido maior atenção. A estatística do teste da razão de verossimilhanças tem, em problemas regulares e sob a hipótese nula, distribuição limite qui-quadrado. Contudo, em amostras pequenas a distribuição nula limite (χ^2) pode fornecer uma aproximação pobre à distribuição nula exata da estatística da razão de verossimilhanças, implicando distorção do tamanho do teste. Com o objetivo de melhorar a aproximação da distribuição da estatística de teste pela distribuição χ^2 uma estratégia usual é utilizar correções de Bartlett (Bartlett, 1937) para a estatística da razão de verossimilhanças. A correção de Bartlett depende de cumulantes e cumulantes mistos de até quarta ordem das derivadas da função de log-verossimilhança, o que, em alguns modelos, pode ser oneroso ou inviável. Para o modelo de regressão beta a determinação do termo de correção de Bartlett, na notação de Lawley (1956), é, como citado em Ferrari e Pinheiro (2011), substancialmente complicada, pois ao contrário do que ocorre nos modelos lineares generalizados os parâmetros do modelo de regressão beta não são ortogonais.

Uma abordagem útil para melhorar as inferências em pequenas amostras, nos casos em que a correção de Bartlett é de difícil obtenção analítica, é o ajuste de Skovgaard (Skovgaard, 2001). O ajuste de Skovgaard é mais direto do que a correção de Bartlett, requerendo apenas derivadas de segunda ordem da função de log-verossimilhança e independentemente da ortogonalidade dos parâmetros do modelo. Nos modelos de regressão beta com dispersão variável e de regressão beta inflacionados são derivados, respectivamente, em Ferrari e Pinheiro (2011) e Pereira (2010) o ajuste de Skovgaard para a estatística da razão de verossimilhanças. Em Ferrari e Cysneiros (2008) esse ajuste é derivado na classe de modelos não-lineares da família exponencial. Os resultados numéricos desses trabalhos mostram que o teste baseado na estatística corrigida de Skovgaard possui melhor desempenho do que o teste baseado na estatística não corrigida.

Em uma abordagem assintótica diferente do ajuste de Skovgaard (2001), a correção de Bartlett possui acurácia de mais alta ordem. Com isso, torna-se de interesse obter correções de Bartlett para melhorar as inferências em pequenas amostras, além de obter resultados comparativos entre as duas abordagens para as estatísticas da razão de verossimilhanças melhoradas. Aliado a isso, apesar do resultado dado por Lawley (1956) ser de difícil obtenção em modelos de regressão beta, há abordagens matriciais na qual a obtenção do fator de correção de Bartlett é de fácil implementação, dependendo apenas da obtenção de cumulantes de até quarta ordem e de simples operações matriciais. Neste sentido, Cordeiro (1993) apresenta uma expressão matricial alternativa às fórmulas de Lawley, podendo ser facilmente implementada em uma linguagem matricial de programação. Outra alternativa para modelos em que a correção de

Bartlett é de difícil derivação analítica é a utilização da correção de Bartlett bootstrap (Rocke, 1989). Neste caso o fator de correção de Bartlett é determinado utilizando o método de reamostragens bootstrap (Efron, 1979).

Nosso objetivo nesse capítulo é melhorar o desempenho em pequenas amostras do teste da razão de verossimilhanças em modelos de regressão beta. Para isso, utilizamos duas abordagens da correção de Bartlett. Primeiramente, utilizamos uma correção analítica em notação matricial, apresentando os cumulantes de até quarta ordem das derivadas da função de log-verossimilhança. Em segundo lugar, consideramos a correção de Bartlett bootstrap para obter uma estatística de teste cuja distribuição exata nula é melhor aproximada pela distribuição limite. O ajuste de Skovgaard apresentado em Ferrari e Pinheiro (2011) é considerado e comparado numericamente às correções de Bartlett obtidas. Percebemos que as distribuições nulas das estatísticas da razão de verossimilhanças corrigidas por Bartlett em geral estão mais próximas da distribuição nula limite (χ^2) do que da distribuição nula da estatística não corrigida e da estatística ajustada por Skovgaard. Apresentamos evidências numéricas de que o teste da razão de verossimilhanças corrigido por Bartlett possui distorção de tamanho menor do que outros testes.

Este capítulo está organizado da seguinte forma. A próxima seção apresenta o modelo de regressão beta proposto em Ferrari e Cribari-Neto (2004), explicitando a função de log-verossimilhança e os estimadores de máxima verossimilhança, o vetor escore e a matriz de informação. A Seção 1.3 apresenta o teste da razão de verossimilhanças, assim como as correções da estatística da razão de verossimilhanças consideradas e determinadas neste trabalho. A avaliação numérica dos testes da razão de verossimilhanças é apresentada na Seção 1.4. Além dos resultados de simulação de Monte Carlo para avaliação do tamanho e do poder dos testes corrigidos, apresentamos também resultados que evidenciam a boa aproximação das estatísticas corrigidas por Bartlett pela distribuição de referência qui-quadrado. Na Seção 1.5 apresentamos uma aplicação a dados reais. As conclusões do capítulo são apresentadas na última seção e os cumulantes de até quarta ordem obtidos podem ser encontrados no Apêndice A.

1.2 O modelo de regressão beta

Seja $y = (y_1, \dots, y_n)^\top$ um vetor com n variáveis aleatórias independentes, em que cada y_i , $i = 1, \dots, n$, tem densidade na forma (1.1) com média μ_i e parâmetro de precisão desconhecido ϕ . Com isso, o modelo de regressão beta pode ser escrito como

$$g(\mu_i) = \sum_{j=1}^p x_{ij} \beta_j = \eta_i, \quad (1.2)$$

em que $\beta = (\beta_1, \dots, \beta_p)^\top$ é o vetor de parâmetros desconhecidos e x_{i1}, \dots, x_{ip} são observações de p variáveis independentes conhecidas ($p < n$). Quando o intercepto é incluído no modelo, temos que $x_{i1} = 1$, para $i = 1, \dots, n$. Finalmente, $g(\cdot)$ é uma função de ligação estritamente monótona e duas vezes diferenciável, com domínio em $(0, 1)$ e imagem em \mathbb{R} . Para o modelo de regressão beta, podemos escolher diferentes funções de ligação como, por exemplo, a função $\text{logit } g(\mu) = \log\{\mu/(1 - \mu)\}$.

Para a estimação do vetor paramétrico k -dimensional $\theta = (\beta^\top, \phi)^\top$, em que $k = (p + 1)$, é utilizado o método da máxima verossimilhança. O logaritmo da função de verossimilhança para n variáveis aleatórias independentes é

$$\ell(\theta) = \ell(\theta; y) = \sum_{i=1}^n \ell_i(\mu_i, \phi), \quad (1.3)$$

em que

$$\begin{aligned} \ell_i(\mu_i, \phi) &= \log \Gamma(\phi) - \log \Gamma(\mu_i \phi) - \log \Gamma((1 - \mu_i)\phi) + (\mu_i \phi - 1) \log y_i \\ &\quad + \{(1 - \mu_i)\phi - 1\} \log(1 - y_i). \end{aligned}$$

A função escore $U(\theta)$ é obtida pela diferenciação da função de log-verossimilhança em relação aos parâmetros desconhecidos. Com isso, as funções escore para β e ϕ são dadas, respectivamente, por

$$\begin{aligned} U_\beta(\theta) &= \phi X^\top T(y^* - \mu^*), \\ U_\phi(\theta) &= \sum_{i=1}^n \mu_i(y_i^* - \mu^*) + \log(1 - y_i) - \psi((1 - \mu_i)\phi) + \psi(\phi), \end{aligned}$$

em que X é uma matriz $n \times p$ cuja i -ésima linha é x_i^\top , $T = \text{diag}\{1/g'(\mu_1), \dots, 1/g'(\mu_n)\}$, $y^* = \{y_1^*, \dots, y_n^*\}^\top$, $\mu^* = \{\mu_1^*, \dots, \mu_n^*\}^\top$, $y_i^* = \log\left(\frac{y_i}{1-y_i}\right)$, $\mu_i^* = \psi(\mu_i \phi) - \psi((1 - \mu_i)\phi)$ e $\psi(\cdot)$ é a função digama¹.

Os estimadores de máxima verossimilhança são obtidos pela solução do seguinte sistema:

$$\begin{cases} U_\beta(\theta) = 0 \\ U_\phi(\theta) = 0 \end{cases}.$$

A solução deste sistema não possui forma fechada, fazendo-se necessário o uso de algoritmos de otimização não-linear, como o algoritmo quasi-Newton BFGS. Para detalhes, ver Press *et al.* (1992).

A matriz de informação de Fisher conjunta para β e ϕ é dada por

$$K = K(\theta) = \begin{pmatrix} K_{(\beta, \beta)} & K_{(\beta, \phi)} \\ K_{(\phi, \beta)} & K_{(\phi, \phi)} \end{pmatrix},$$

em que $K_{(\beta, \beta)} = \phi X^\top W X$, $K_{(\beta, \phi)} = (K_{(\phi, \beta)})^\top = X^\top T c$ e $K_{(\phi, \phi)} = \text{tr}(D)$. Ainda, temos que $w_i = \phi \{\psi'(\mu_i \phi) + \psi'((1 - \mu_i)\phi)\} \frac{1}{\{g'(\mu_i)\}^2}$, $c_i = \phi \{\psi'(\mu_i \phi)\mu_i - \psi'((1 - \mu_i)\phi)(1 - \mu_i)\}$, $d_i = \psi'(\mu_i \phi)\mu_i^2 + \psi'((1 - \mu_i)\phi)(1 - \mu_i)^2 - \psi'(\phi)$, em que $W = \text{diag}\{w_1, \dots, w_n\}$, $c = (c_1, \dots, c_n)^\top$ e $D = \text{diag}\{d_1, \dots, d_n\}$. Maiores detalhes e derivadas de mais alta ordem da função de log-verossimilhança são apresentados no Apêndice A.

¹A função poligama é definida, para $m = 0, 1, \dots$, como $\psi^{(m)}(x) = (d^{m+1}/dx^{m+1}) \log \Gamma(x)$, $x > 0$. A função digama é obtida usando $m = 0$.

Sob certas condições de regularidade, temos que para tamanhos amostrais grandes, a distribuição conjunta de $\hat{\beta}$ e $\hat{\phi}$ é aproximadamente normal k -multivariada, de forma que

$$\begin{pmatrix} \hat{\beta} \\ \hat{\phi} \end{pmatrix} \sim \mathcal{N}_k \left(\begin{pmatrix} \beta \\ \phi \end{pmatrix}, K^{-1} \right),$$

aproximadamente, sendo $\hat{\beta}$ e $\hat{\phi}$ os estimadores de máxima verossimilhança de β e ϕ , respectivamente.

1.3 Correções do teste da razão de verossimilhanças

Considere o modelo paramétrico introduzido em (1.2) com a correspondente função de log-verossimilhança em (1.3), em que $\theta = (\theta_1^\top, \theta_2^\top)^\top$ é o vetor k -dimensional de parâmetros do modelo, θ_1 sendo um vetor q -dimensional e θ_2 sendo um vetor com dimensão $k - q$. Suponha que o interesse reside em testar a hipótese nula

$$\mathcal{H}_0: \theta_1 = \theta_1^0$$

versus a hipótese alternativa

$$\mathcal{H}_1: \theta_1 \neq \theta_1^0,$$

em que θ_1^0 é um vetor de escalares conhecidos de mesma dimensão que θ_1 . Assim, o vetor θ_2 é o vetor de parâmetros de incômodo e q é o número de restrições impostas por \mathcal{H}_0 .

A estatística da razão de verossimilhanças para esse teste pode ser escrita como

$$LR = 2\{\ell(\hat{\theta}; y) - \ell(\tilde{\theta}; y)\},$$

em que o vetor $\tilde{\theta}$ é o estimador de máxima verossimilhança restrito de θ obtido sob hipótese nula, ou seja, $\tilde{\theta} = (\theta_1^{0\top}, \tilde{\theta}_2^\top)^\top$.

A estatística da razão de verossimilhanças LR tem distribuição aproximadamente χ_q^2 sob \mathcal{H}_0 com erro de ordem n^{-1} . Em pequenas amostras a aproximação da distribuição nula exata da estatística LR pela distribuição χ_q^2 pode ser pobre. Como consequência, os testes da razão de verossimilhanças podem apresentar tamanhos notavelmente distorcidos.

Com o objetivo de melhorar o teste da razão de verossimilhanças surge a correção de Bartlett, originalmente introduzida por Bartlett (1937) e posteriormente generalizada por Lawley (1956). A ideia básica da correção de Bartlett é, ao invés de utilizar a aproximação χ^2 para a estatística LR , utilizar a seguinte aproximação

$$\frac{LR}{c} \sim \chi_q^2,$$

em que $c = E(LR)/q$ é conhecido como fator de correção de Bartlett. Contudo, apesar de, em geral, não ser possível avaliar $E(LR)$, o fator c também pode ser consistentemente estimado sob \mathcal{H}_0 utilizando-se cumulantes (ou momentos) das derivadas da função de log-verossimilhança (Skovgaard, 2001). Essa correção de Bartlett leva a uma aproximação assintótica da estatística de teste pela distribuição qui-quadrado com erro de segunda ordem (Barndorff-Nielsen e Cox, 1984; Bartlett, 1937; Rocke, 1989). Temos que $\Pr(LR \leq x) = \Pr(\chi_q^2 \leq x) + O(n^{-1})$ e $\Pr\left(\frac{LR}{c} \leq x\right) = \Pr(\chi_q^2 \leq x) + O(n^{-2})$.

1.3.1 Correção de Bartlett em forma matricial

O fator de correção de Bartlett pode ser escrito como

$$c = 1 + \frac{\varepsilon_k - \varepsilon_{k-q}}{q},$$

pois, utilizando a expansão de Lawley, o valor esperado da estatística da razão de verossimilhanças pode ser escrito como

$$E(LR) = q + \varepsilon_k - \varepsilon_{k-q} + O(n^{-2}),$$

em que

$$\varepsilon_k = \sum_{\theta} (\lambda_{rstu} - \lambda_{rstuvw}), \quad (1.4)$$

$$\lambda_{rstu} = \kappa^{rs} \kappa^{tu} \left\{ \frac{\kappa_{rstu}}{4} - \kappa_{rst}^{(u)} + \kappa_{rt}^{(su)} \right\}, \quad (1.5)$$

$$\lambda_{rstuvw} = \kappa^{rs} \kappa^{tu} \kappa^{vw} \left\{ \kappa_{rtv} \left(\frac{\kappa_{suw}}{6} - \kappa_{sw}^{(u)} \right) + \kappa_{rtu} \left(\frac{\kappa_{svw}}{4} - \kappa_{sw}^{(v)} \right) + \kappa_{rt}^{(v)} \kappa_{sw}^{(u)} + \kappa_{rt}^{(u)} \kappa_{sw}^{(v)} \right\} \quad (1.6)$$

e

$$\kappa_{rs} = E \left(\frac{\partial^2 \ell(\theta)}{\partial \theta_r \partial \theta_s} \right), \quad \kappa_{rst} = E \left(\frac{\partial^3 \ell(\theta)}{\partial \theta_r \partial \theta_s \partial \theta_t} \right), \quad \kappa_{rs}^{(t)} = \frac{\partial \kappa_{rs}}{\partial \theta_t}, \text{ etc.} \quad (1.7)$$

Ainda, $-\kappa^{rs}$ é o elemento (r,s) da inversa da matriz de informação de Fisher, K^{-1} , e o somatório em (1.4) cobre todos os componentes de θ , isto é, os índices r, s, t, u, v e w varrem todos os k parâmetros. A expressão ε_{k-q} é obtida de (1.4) quando o somatório cobre somente os elementos do vetor de parâmetros de incômodo θ_2 . Todos os κ 's são momentos sob a distribuição amostral e são, em geral, de ordem n , ε_k e ε_{k-q} sendo de ordem $O(n^{-1})$.

Obter a correção de Bartlett por meio dessa notação de Lawley pode ser muito difícil, pois a fórmula envolve o produto de cumulantes que não são invariantes por permutação (Cordeiro, 1993). Em especial, nos modelos de regressão beta em que os parâmetros não são ortogonais, ou seja, a matriz de informação de Fisher não é bloco diagonal, a derivação da correção de Bartlett na notação de Lawley se torna especialmente complicada ou de obtenção impossível. Para contornar esse problema, utilizamos uma expressão matricial geral apresentada em Cordeiro (1993).

Para expressar ε_k na forma matricial primeiramente definimos as seguintes matrizes de ordem $k \times k$: $A^{(tu)}$, $P^{(t)}$ e $Q^{(u)}$, para $t, u = 1, \dots, k$. Os elementos (r,s) dessas matrizes são dados por

$$A^{(tu)} = \left\{ \frac{\kappa_{rstu}}{4} - \kappa_{rst}^{(u)} + \kappa_{rt}^{(su)} \right\}, \quad P^{(t)} = \{ \kappa_{rst} \}, \quad Q^{(u)} = \{ \kappa_{su}^{(r)} \},$$

para $t, u = 1, \dots, k$. Usando notação matricial, podemos escrever

$$\sum_{\theta} \lambda_{rstu} = \text{tr}(K^{-1}L), \quad (1.8)$$

$$\begin{aligned} & \sum_{\theta} \kappa^{rs} \kappa^{tu} \kappa^{vw} \left\{ \frac{1}{6} \kappa_{rtv} \kappa_{suw} - \kappa_{rtv} \kappa_{sw}^{(u)} + \kappa_{rt}^{(v)} \kappa_{sw}^{(u)} \right\} \\ &= -\frac{1}{6} \text{tr}(K^{-1}M_1) + \text{tr}(K^{-1}M_2) - \text{tr}(K^{-1}M_3), \end{aligned} \quad (1.9)$$

$$\begin{aligned} & \sum_{\theta} \kappa^{rs} \kappa^{tu} \kappa^{vw} \left\{ \frac{1}{4} \kappa_{rtu} \kappa_{svw} - \kappa_{rtu} \kappa_{sw}^{(v)} + \kappa_{rt}^{(u)} \kappa_{sw}^{(v)} \right\} \\ &= -\frac{1}{4} \text{tr}(K^{-1}N_1) + \text{tr}(K^{-1}N_2) - \text{tr}(K^{-1}N_3), \end{aligned} \quad (1.10)$$

em que os elementos (r,s) das matrizes L , M_1 , M_2 , M_3 , N_1 , N_2 e N_3 são dados por

$$\begin{aligned} L &= \left\{ \text{tr}(K^{-1}A^{(rs)}) \right\}, \\ M_1 &= \left\{ \text{tr}(K^{-1}P^{(r)}K^{-1}P^{(s)}) \right\}, \\ M_2 &= \left\{ \text{tr}(K^{-1}P^{(r)}K^{-1}Q^{(s)\top}) \right\}, \\ M_3 &= \left\{ \text{tr}(K^{-1}Q^{(r)}K^{-1}Q^{(s)}) \right\}, \\ N_1 &= \left\{ \text{tr}(P^{(r)}K^{-1})\text{tr}(P^{(s)}K^{-1}) \right\}, \\ N_2 &= \left\{ \text{tr}(P^{(r)}K^{-1})\text{tr}(Q^{(s)}K^{-1}) \right\}, \\ N_3 &= \left\{ \text{tr}(Q^{(r)}K^{-1})\text{tr}(Q^{(s)}K^{-1}) \right\}. \end{aligned}$$

Assim, de (1.4)–(1.10) podemos escrever

$$\varepsilon_k = \text{tr} \left[K^{-1}(L - M - N) \right], \quad (1.11)$$

em que $M = -\frac{1}{6}M_1 + M_2 - M_3$ e $N = -\frac{1}{4}N_1 + N_2 - N_3$.

O termo em (1.11) é facilmente computado utilizando uma linguagem matricial de programação, como Ox (Doornik, 2007) e R (R Development Core Team, 2009), requerendo apenas a obtenção de $(k+1)^2$ matrizes de ordem k , nomeadamente: K^{-1} , k matrizes $P^{(t)}$, k matrizes $Q^{(u)}$ e k^2 matrizes $A^{(tu)}$. Todas as outras matrizes são facilmente obtidas apenas com operações matriciais elementares a partir dessas $(k+1)^2$ matrizes. Com isso, para a obtenção do fator de correção de Bartlett c são necessárias $(k+1)^2$ matrizes de ordem k e $(k-q+1)^2$ matrizes de ordem $k-q$. O custo computacional pode ser diminuído substancialmente considerando as matrizes de ordem $k-q$ como submatrizes das matrizes de ordem k . Para obtenção das matrizes $P^{(t)}$, $Q^{(u)}$ e $A^{(tu)}$ necessita-se dos cumulantes até quarta ordem de derivadas da função de log-verossimilhança, como apresentados em (1.7). Esses cumulantes, para o modelo de regressão beta introduzido na Seção 1.2, foram computados e estão apresentados no Apêndice A. A implementação computacional em linguagem R , para determinação do fator de correção de Bartlett no modelo de regressão beta, encontra-se no Apêndice B

A estatística da razão de verossimilhanças corrigida usual é dada por LR/c . Contudo, há algumas modificações dessa estatística de teste. Neste texto consideramos as seguintes

variações da estatística da razão de verossimilhanças corrigida por Bartlett:

$$\begin{aligned} LR_{b1} &= \frac{LR}{c}, \\ LR_{b2} &= LR \exp \left\{ -\frac{(\varepsilon_k - \varepsilon_{k-q})}{q} \right\}, \\ LR_{b3} &= LR \left\{ 1 - \frac{(\varepsilon_k - \varepsilon_{k-q})}{q} \right\}. \end{aligned}$$

As estatísticas LR_{b1} , LR_{b2} e LR_{b3} são equivalentes até ordem $O(n^{-1})$ (Lemonte *et al.*, 2010), sendo que a estatística LR_{b2} tem a vantagem de nunca assumir valores negativos.

1.3.2 Correção de Bartlett bootstrap

Rocke (1989) introduz uma alternativa à determinação analítica da correção de Bartlett em que o fator de correção é determinado via método bootstrap (Efron, 1979). A obtenção da correção de Bartlett bootstrap, por meio de bootstrap paramétrico, pode ser descrita brevemente como segue. Reamostras bootstrap são usadas para estimar o valor esperado da estatística da razão de verossimilhanças diretamente da amostra observada $y = (y_1, \dots, y_n)^\top$. Para isso, são geradas, sob \mathcal{H}_0 , B reamostras bootstrap $(y^{*1}, y^{*2}, \dots, y^{*B})$ do modelo, substituindo os parâmetros do modelo pelas estimativas de máxima verossimilhança restritas usando a amostra original y (bootstrap paramétrico). Para cada pseudo amostra y^{*b} são obtidas as estatísticas LR bootstrap, LR^{*b} , com $b = 1, 2, \dots, B$. A estatística LR^{*b} é calculada, em cada iteração, da seguinte maneira:

$$LR^{*b} = 2\{\ell(\hat{\theta}^{*b}; y^{*b}) - \ell(\tilde{\theta}^{*b}; y^{*b})\},$$

em que $\hat{\theta}^{*b}$ e $\tilde{\theta}^{*b}$ são os estimadores de máxima verossimilhança de θ obtidos por meio da maximização de $\ell(\theta; y^{*b})$ sob \mathcal{H}_1 e \mathcal{H}_0 , respectivamente. Assim sendo, a estatística da razão de verossimilhanças corrigida pelo fator de correção de Bartlett bootstrap é dada por

$$LR_{boot} = \frac{LRq}{\overline{LR^*}},$$

em que $\overline{LR^*} = \frac{1}{B} \sum_{b=1}^B LR^{*b}$.

A correção de Bartlett bootstrap tem vantagens computacionais quando comparada ao esquema bootstrap usual para determinação de valores críticos mais acurados. Na abordagem bootstrap usual são necessárias no mínimo 1000 reamostras bootstrap, pois tem-se o interesse nas caudas da distribuição (Efron, 1986, 1987). Já para a correção de Bartlett bootstrap, segundo simulações prévias omitidas, valores de B maiores que 200 conduzem a melhoramentos negligíveis para o teste Bartlett bootstrap. Adicionalmente, em Rocke (1989) o autor afirma que a correção de Bartlett bootstrap utilizando $B = 100$ tipicamente fornece resultados equivalentes ao método bootstrap usual com $B = 700$.

1.3.3 Ajuste de Skovgaard

Em uma outra abordagem, Skovgaard (2001) generaliza os resultados de Skovgaard (1996) apresentando um ajuste mais simples para a estatística da razão de verossimilhanças do que

a correção de Bartlett. Esta estatística vem sendo derivada para diversas classes de modelos, como em Ferrari e Cysneiros (2008), Ferrari e Pinheiro (2011), Melo *et al.* (2009) e Pereira (2010). Esses trabalhos têm mostrado numericamente que testes de hipóteses baseados na estatística de Skovgaard têm melhor desempenho em termos de tamanho do teste do que o teste não corrigido.

Para introduzirmos o ajuste de Skovgaard para a estatística da razão de verossimilhanças, derivado em Ferrari e Pinheiro (2011), considere o seguinte. Lembrando que $\theta = (\theta_1^\top, \theta_2^\top)^\top$, em que θ_1 e θ_2 são os parâmetros de interesse e incômodo, respectivamente, sejam J a matriz de informação observada e J_{11} a matriz $q \times q$ de informação observada correspondente a θ_1 . Considere ainda a seguinte notação: $\hat{J} = J(\hat{\theta})$, $\tilde{J} = J(\tilde{\theta})$, $\hat{K} = K(\hat{\theta})$, $\tilde{K} = K(\tilde{\theta})$ e $\tilde{U} = U(\tilde{\theta})$.

A estatística de teste ajustada por Skovgaard é dada por

$$LR_{sk1} = LR - 2 \log \xi,$$

em que

$$\xi = \frac{\{|\tilde{K}| |\hat{K}| |\tilde{J}_{11}|\}^{1/2} \{\tilde{U}^\top \tilde{Y}^{-1} \hat{K} \hat{J}^{-1} \tilde{Y} \tilde{K}^{-1} \tilde{U}\}^{q/2}}{|\tilde{Y}| \{|\tilde{K} \tilde{Y}^{-1} \hat{J} \tilde{K}^{-1} \tilde{Y}\}_{11}|^{1/2} LR^{q/2-1} \tilde{U}^\top \tilde{Y}^{-1} \tilde{v}}.$$

Aqui, \tilde{Y} e \tilde{v} são dados por $\tilde{Y} = E_\theta[U(\theta)U^\top(\theta_2)]$ e $\tilde{v} = E_\theta[U(\theta)(\ell(\theta) - \ell(\theta_2))]$, substituindo θ por $\hat{\theta}$ e θ_2 por $\tilde{\theta}$ depois de calculados os valores esperados.

Uma versão assintoticamente equivalente a LR_{sk1} é dada por

$$LR_{sk2} = LR \left(1 - \frac{1}{LR} \log \xi \right)^2.$$

Sob \mathcal{H}_0 as estatísticas LR_{sk1} e LR_{sk2} seguem distribuição χ_q^2 com alto grau de acurácia (Skovgaard, 2001; Ferrari e Pinheiro, 2011). Para maiores detalhes e derivações em forma matricial de \tilde{Y} e \tilde{v} na classe de modelos de regressão beta, ver Ferrari e Pinheiro (2011). Em Ferrari e Pinheiro (2011) o ajuste de Skovgaard é derivado para um modelo de regressão beta mais geral, considerando o modelo não-linear e também a modelagem do parâmetro de precisão.

1.4 Avaliação numérica

Nesta seção apresentamos os resultados de simulações de Monte Carlo para a avaliação de desempenho em pequenas amostras do teste da razão de verossimilhanças (LR) em regressão beta e de seis versões corrigidas desse teste, nomeadamente: correção de Bartlett (LR_{b1}) e suas duas versões assintoticamente equivalentes LR_{b2} e LR_{b3} , correção de Bartlett bootstrap (LR_{boot}) e duas versões assintoticamente equivalentes do teste com ajuste de Skovgaard, LR_{sk1} e LR_{sk2} . O experimento computacional é semelhante ao apresentado em Ferrari e Pinheiro (2011) para avaliação numérica do ajuste de Skovgaard em regressão beta com dispersão constante. O número de réplicas de Monte Carlo foi fixado em 10000 e para a correção de Bartlett bootstrap utilizamos bootstrap paramétrico com 500 réplicas. Todas as simulações foram realizadas utilizando a linguagem de programação R (R Development Core Team, 2009).

Para a avaliação numérica foi considerado o modelo (1.2) com o componente sistemático da média dado por

$$\text{logit}(\mu_i) = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i},$$

em que os valores das covariáveis são determinados aleatoriamente da distribuição uniforme $\mathcal{U}(-0.5, 0.5)$ e são mantidos constantes durante todo o experimento. Consideramos três valores diferentes para o parâmetro de precisão ϕ , são eles: 30, 10 e 5.

Para obter a taxa de rejeição nula dos testes sobre o vetor de parâmetros β consideramos tamanhos amostrais iguais a 15, 20, 30 e 40 e os seguintes níveis nominais para os testes: $\alpha = 10\%$, 5% e 1% . Foram consideradas três hipóteses nulas, nomeadamente $\mathcal{H}_0 : \beta_2 = 0$ ($q = 1$), $\mathcal{H}_0 : \beta_2 = \beta_3 = 0$ ($q = 2$) e $\mathcal{H}_0 : \beta_2 = \beta_3 = \beta_4 = 0$ ($q = 3$), que são testadas contra hipóteses alternativas bilaterais. Para o primeiro caso, quando $q = 1$, consideramos $\beta_1 = 1$, $\beta_2 = 0$, $\beta_3 = 1$, $\beta_4 = 5$ e $\beta_5 = -4$. Para o segundo caso os valores dos parâmetros foram $\beta_1 = 1$, $\beta_2 = \beta_3 = 0$, $\beta_4 = 5$ e $\beta_5 = -4$. Finalmente, para $q = 3$, temos $\beta_1 = 1$, $\beta_2 = \beta_3 = \beta_4 = 0$ e $\beta_5 = -4$.

As Tabelas 1.1, 1.2 e 1.3 apresentam as taxas de rejeição (em porcentagem) da hipótese nula para as três restrições sobre o vetor paramétrico β . Observa-se que o teste da razão de verossimilhanças é bastante liberal, com taxas de rejeição chegando a ser oito vezes maior do que o nível nominal, como observado na Tabela 1.2 para $\phi = 5$, $\alpha = 1\%$ e $n = 15$. Observa-se que, em geral, os tamanhos amostrais maiores e os maiores valores de ϕ estão associados a taxas de rejeição nula mais próximas dos níveis nominais.

Os resultados de simulação para o caso em que se testa apenas uma restrição, apresentados na Tabela 1.1, mostram o bom desempenho dos testes corrigidos. Dentre as versões corrigidas destaca-se a correção de Bartlett LR_{b3} , seguida pelo ajuste de Skovgaard LR_{sk1} e da correção de Bartlett bootstrap LR_{boot} . O teste baseado estatística LR_{boot} apresenta desempenho superior aos demais quando $\phi = 30$. Considerando o caso em que $\phi = 30$ e $\alpha = 10\%$, as taxas são, respectivamente para os quatro tamanhos amostrais, 10.2, 10.3, 10.6 e 10.0 para LR_{b3} , enquanto que as taxas de rejeição nula para LR_{sk1} são 10.2, 10.3, 10.8 e 10.2. O bom desempenho do teste LR_{b3} pode ser observado em todos os casos considerados na simulação. Para $\phi = 30$ e $\alpha = 5\%$ na Tabela 1.1, por exemplo, percebe-se que os testes corrigidos LR_{b3} e LR_{boot} apresentam as taxas de rejeição mais próximas aos valores nominais.

Os resultados para os casos em que testamos mais de uma restrição, nomeadamente $q = 2$ e $q = 3$, apresentados nas Tabelas 1.2 e 1.3, são semelhantes aos resultados do caso em que $q = 1$, evidenciando o bom desempenho dos testes corrigidos. Por exemplo, para $q = 2$, $\phi = 30$ e $\alpha = 5\%$ o teste da razão de verossimilhanças não corrigido apresentou probabilidade de erro tipo I igual a 14.4% para $n = 15$, enquanto que as estatísticas LR_{b3} e LR_{boot} alcançaram a taxa de 5.6%. Dentre os demais ajustes, LR_{sk1} apresentou o melhor resultado, com taxas de rejeição nula de 6.4%, bem acima das duas melhores correções obtidas neste trabalho. Já para $q = 3$, podemos destacar o caso $\phi = 30$ e $\alpha = 5\%$ em que LR apresentou taxa de rejeição nula igual a 14.6% para $n = 15$, enquanto as estatísticas LR_{b3} e LR_{boot} alcançaram a taxa de 5.0%. Para $\phi = 30$ e $\alpha = 1\%$ também destacam-se as estatísticas LR_{b3} , LR_{sk1} e LR_{boot} que alcançam taxas de rejeição nula muito próximas a 1.0%, enquanto o teste não corrigido apresentou taxas iguais a 4.8%, 3.3%, 2.4% e 1.8%, respectivamente, para os quatro tamanhos amostrais considerados.

Tabela 1.1 Taxas de rejeição nula (percentuais) do teste $\mathcal{H}_0 : \beta_2 = 0$ ($q = 1$).

ϕ	Estat	$\alpha = 10\%$				$\alpha = 5\%$				$\alpha = 1\%$				
		n	15	20	30	40	15	20	30	40	15	20	30	40
30	LR		19.5	16.8	14.8	13.0	12.1	9.7	8.0	7.0	4.2	2.7	2.3	1.6
	LR_{b1}		12.7	11.8	11.3	10.5	6.8	6.1	6.1	5.2	1.7	1.3	1.4	1.1
	LR_{b2}		11.7	11.2	11.0	10.2	6.1	5.6	6.0	5.1	1.4	1.1	1.2	1.1
	LR_{b3}		10.2	10.3	10.6	10.0	5.1	5.0	5.7	4.9	1.1	1.0	1.1	1.0
	LR_{sk1}		10.2	10.3	10.8	10.2	5.2	4.9	5.7	5.0	1.2	1.0	1.2	1.0
	LR_{sk2}		13.2	11.7	12.7	11.7	7.6	6.2	7.2	6.2	2.7	1.7	2.1	2.0
	LR_{boot}		10.2	10.2	10.6	10.3	4.9	5.0	5.6	4.9	1.1	1.0	1.2	1.1
10	LR		22.0	21.4	17.9	13.7	14.4	13.8	11.0	8.1	5.5	5.1	3.6	2.2
	LR_{b1}		15.2	15.9	14.2	11.4	8.6	9.1	8.2	6.3	2.4	2.5	2.2	1.4
	LR_{b2}		13.8	15.1	13.9	11.2	7.7	8.5	7.9	6.2	1.9	2.2	2.0	1.3
	LR_{b3}		12.0	14.2	13.5	11.0	6.3	7.8	7.6	6.0	1.5	1.9	1.9	1.2
	LR_{sk1}		12.1	14.6	13.9	11.0	6.4	8.0	7.8	5.9	1.5	2.0	2.0	1.2
	LR_{sk2}		14.9	17.3	16.3	13.0	8.7	10.2	9.9	7.6	2.8	3.5	3.6	2.6
	LR_{boot}		12.2	14.6	14.4	12.7	6.6	8.2	8.5	7.2	1.5	2.1	2.2	1.9
5	LR		19.1	16.2	15.4	12.7	12.2	9.6	8.7	6.8	4.3	3.0	2.5	1.8
	LR_{b1}		12.9	11.5	12.0	10.8	7.0	6.2	6.3	5.3	1.8	1.3	1.3	1.2
	LR_{b2}		12.1	11.0	11.6	10.7	6.3	5.8	6.0	5.2	1.4	1.2	1.3	1.1
	LR_{b3}		10.6	10.2	11.3	10.5	5.3	5.2	5.8	5.1	0.9	1.0	1.2	1.1
	LR_{sk1}		11.7	10.9	11.5	10.7	6.2	5.6	6.1	5.3	1.3	1.1	1.2	1.2
	LR_{sk2}		15.2	14.1	14.1	13.1	9.1	8.1	8.2	7.3	3.4	2.8	2.8	2.8
	LR_{boot}		13.9	10.4	11.7	12.3	7.8	5.5	6.2	6.6	2.5	1.1	1.4	1.8

Os resultados apresentados nas Tabelas 1.1, 1.2 e 1.3 deixam claro que, em termos de tamanho do teste, os testes corrigidos apresentam melhor desempenho em pequenas amostras do que o teste da razão de verossimilhanças não corrigido. Dentre as estatísticas consideradas, torna-se evidente a superioridade das seguintes estatísticas corrigidas: a correção de Bartlett LR_{b3} , o ajuste de Skovgaard LR_{sk1} e a correção de Bartlett bootstrap LR_{boot} . As taxas de rejeição nula desses testes estão mais próximas aos níveis nominais do que as do teste não corrigido e as dos outros testes corrigidos considerados. Em particular, destacamos os resultados do teste baseado na correção de Bartlett bootstrap para $\phi = 30$, pois, apesar da simplicidade de obtenção desta estatística corrigida, o teste obtém bons resultados. Para melhor comparação dos desempenhos dos testes baseados nas estatísticas LR_{b3} e LR_{sk1} , apresentamos no Apêndice C resultados de simulação considerando 50.000 réplicas de Monte Carlo e outros valores de ϕ . Esses resultados evidenciam a superioridade dos testes baseados na estatística LR_{b3} , principalmente nos menores tamanhos amostrais.

Para verificar a aproximação das distribuições nulas das estatísticas corrigidas pela distri-

Tabela 1.2 Taxas de rejeição nula (percentuais) do teste $\mathcal{H}_0 : \beta_2 = \beta_3 = 0$ ($q = 2$).

ϕ	Estat	n	$\alpha = 10\%$				$\alpha = 5\%$				$\alpha = 1\%$			
			15	20	30	40	15	20	30	40	15	20	30	40
30	LR		23.0	17.8	14.6	13.8	14.4	10.6	7.8	7.6	5.4	3.3	1.9	1.9
	LR_{b1}		13.7	11.7	10.2	10.7	7.8	6.0	4.8	5.4	2.0	1.4	1.0	1.0
	LR_{b2}		12.4	10.9	9.8	10.5	6.9	5.6	4.7	5.3	1.5	1.2	1.0	1.0
	LR_{b3}		10.7	10.1	9.4	10.2	5.6	5.0	4.5	5.2	1.1	1.0	1.0	0.9
	LR_{sk1}		11.2	10.3	9.6	10.3	6.4	5.1	4.5	5.3	1.8	1.0	1.0	0.9
	LR_{sk2}		12.2	10.9	9.9	10.5	7.1	5.4	4.7	5.4	1.9	1.2	1.1	1.0
	LR_{boot}		10.5	10.1	9.5	10.4	5.6	5.0	4.6	5.2	1.1	1.0	1.0	1.0
10	LR		26.0	19.1	16.0	15.2	17.4	11.8	9.1	8.4	7.0	3.7	2.7	2.4
	LR_{b1}		16.5	12.7	11.7	12.0	9.8	6.7	6.3	6.2	2.8	1.6	1.4	1.4
	LR_{b2}		15.1	12.0	11.3	11.8	8.9	6.3	6.0	6.0	2.3	1.4	1.3	1.4
	LR_{b3}		13.2	11.0	10.9	11.6	7.4	5.7	5.6	5.9	1.8	1.3	1.2	1.3
	LR_{sk1}		13.4	11.5	11.0	11.7	7.5	5.9	5.7	6.0	1.8	1.3	1.3	1.3
	LR_{sk2}		14.5	12.2	11.4	12.1	8.4	6.4	6.0	6.3	2.2	1.5	1.4	1.5
	LR_{boot}		13.6	11.0	11.1	12.8	7.8	5.6	5.8	6.8	2.0	1.2	1.3	1.7
5	LR		27.8	19.7	15.3	13.1	19.3	12.0	8.5	7.0	8.0	4.2	2.4	1.9
	LR_{b1}		18.6	13.1	11.2	10.1	11.0	7.1	5.8	5.5	3.6	1.8	1.2	1.2
	LR_{b2}		17.2	12.4	10.8	10.0	10.0	6.5	5.6	5.4	3.1	1.7	1.1	1.1
	LR_{b3}		14.9	11.5	10.5	9.8	8.4	6.0	5.4	5.2	2.3	1.5	1.0	1.0
	LR_{sk1}		14.4	12.0	11.2	10.0	7.9	6.2	5.6	5.2	2.2	1.6	1.1	1.2
	LR_{sk2}		16.0	12.8	11.5	10.4	9.1	6.7	5.9	5.6	2.7	1.8	1.2	1.3
	LR_{boot}		15.4	12.1	11.0	14.8	8.9	6.4	5.8	8.7	2.6	1.7	1.2	2.5

buição nula limite χ_q^2 apresentamos a Tabela 1.4 e as Figuras 1.2 e 1.3. Nas Figuras 1.2 e 1.3 são considerados apenas as três estatísticas corrigidas com melhores desempenhos, além da estatística LR .

A Tabela 1.4 apresenta os primeiros quatro momentos e os quantis estimados via simulação de Monte Carlo das estatísticas de teste comparativamente aos valores exatos da distribuição nula limite χ_q^2 . Essas medidas mostram a aproximação pobre da distribuição nula da estatística da razão de verossimilhanças pela distribuição χ_q^2 , evidenciada principalmente pela variância estimada. Por outro lado, os dois primeiros momentos e os quantis mostram a melhor aproximação das distribuições nulas das estatísticas corrigidas pela distribuição nula limite. A estatística LR_{b3} se destaca, sendo seguida pela estatística LR_{boot} . Para a média e variância, por exemplo, os valores estimados para a estatística LR_{b3} são, respectivamente, 1.9993 e 4.0729, muito próximos aos valores dois e quatro, que são o valor esperado e variância da distribuição χ_2^2 . Por outro lado, o ajuste LR_{sk2} apresenta a pior aproximação pela distribuição nula limite dentre as estatísticas corrigidas. As medidas de assimetria e curtose de LR_{sk2} mostram-se com

Tabela 1.3 Taxas de rejeição nula (percentuais) do teste $\mathcal{H}_0 : \beta_2 = \beta_3 = \beta_4 = 0$ ($q = 3$).

ϕ	Estat	$\alpha = 10\%$				$\alpha = 5\%$				$\alpha = 1\%$			
		15	20	30	40	15	20	30	40	15	20	30	40
30	LR	23.0	17.4	14.6	13.6	14.6	10.7	8.2	7.4	4.8	3.3	2.4	1.8
	LR_{b1}	13.1	11.2	10.3	10.4	7.0	5.9	5.5	5.2	1.8	1.3	1.1	1.2
	LR_{b2}	11.9	10.6	10.0	10.2	6.1	5.3	5.3	5.0	1.4	1.2	1.1	1.1
	LR_{b3}	10.3	9.8	9.6	10.0	5.0	4.8	5.1	5.0	1.0	1.0	1.0	1.1
	LR_{sk1}	10.2	9.9	9.7	10.1	5.1	4.8	5.1	5.1	1.1	1.0	1.0	1.1
	LR_{sk2}	10.2	10.3	9.9	10.3	5.7	5.1	5.2	5.2	1.3	1.1	1.0	1.2
	LR_{boot}	10.2	9.8	9.5	9.9	5.0	4.8	5.1	5.0	1.0	1.0	1.0	1.2
10	LR	22.1	18.6	15.3	13.6	13.7	11.2	8.7	7.5	4.6	3.2	2.2	1.8
	LR_{b1}	12.2	11.7	10.8	10.3	6.8	6.2	5.5	5.3	1.5	1.2	1.0	1.0
	LR_{b2}	11.2	11.2	10.5	10.1	6.0	5.7	5.3	5.2	1.3	1.1	1.0	1.0
	LR_{b3}	9.8	10.2	10.1	9.9	5.0	5.1	5.1	5.0	0.9	0.9	0.9	1.0
	LR_{sk1}	10.3	10.6	10.4	10.3	5.1	5.3	5.2	5.1	1.0	0.9	1.0	1.0
	LR_{sk2}	11.2	11.1	10.7	10.4	5.8	5.7	5.4	5.2	1.2	1.0	1.0	1.0
	LR_{boot}	9.6	10.1	10.2	10.0	4.7	5.1	5.0	5.0	0.9	0.9	1.0	1.0
5	LR	21.5	18.4	15.0	12.9	13.6	11.0	8.3	7.2	4.4	3.5	2.3	1.5
	LR_{b1}	12.5	11.6	10.6	9.9	6.5	6.0	5.4	5.1	1.5	1.4	1.2	0.8
	LR_{b2}	11.3	11.0	10.3	9.8	5.9	5.5	5.2	5.1	1.3	1.3	1.1	0.8
	LR_{b3}	9.7	10.2	10.0	9.7	4.8	5.0	5.0	4.9	1.0	1.0	1.0	0.8
	LR_{sk1}	10.1	10.8	10.5	10.0	5.1	5.4	5.4	5.1	1.0	1.1	1.2	0.8
	LR_{sk2}	11.2	11.3	10.7	10.3	5.8	5.7	5.5	5.3	1.2	1.3	1.2	0.9
	LR_{boot}	9.6	10.2	10.0	9.6	4.8	5.0	5.0	5.1	1.1	1.1	1.0	0.8

valores mais distorcidos do que as da estatística não corrigida LR , sendo iguais a 3.1049 e 27.7738, enquanto os valores assintóticos são dois e nove. Já a estatística LR_{sk1} , apesar de ter bom desempenho no que tange às taxas de rejeição nula apresentadas na Tabela 1.2, para $q = 2$, $\phi = 30$ e $n = 20$, não apresenta uma aproximação pela distribuição nula tão boa quanto as estatísticas LR_{b3} e LR_{boot} . Esse fato é evidenciado pelas medidas de variância (4.2331), assimetria (2.1816), curtose (11.5872) e pelo quantil 90 (4.6612), que são consideravelmente diferentes dos valores de referência da distribuição qui-quadrado.

As Figuras 1.2 e 1.3 mostram graficamente a aproximação das distribuições nulas das estatísticas de teste pela distribuição assintótica das mesmas. A Figura 1.2 apresenta o gráfico Q-Q (quantis empíricos exatos *versus* quantis assintóticos) para diferentes tamanhos amostrais no caso $\phi = 10$ e $q = 1$. Já a Figura 1.3 mostra as densidades empíricas das estatísticas para $\phi = 5$ e $q = 3$. Essas densidades foram estimadas utilizando o método kernel² com a

²Para maiores detalhes sobre estimação de densidades via o método não-paramétrico de kernel, ver Silverman (1986) e também Venables e Ripley (2002)

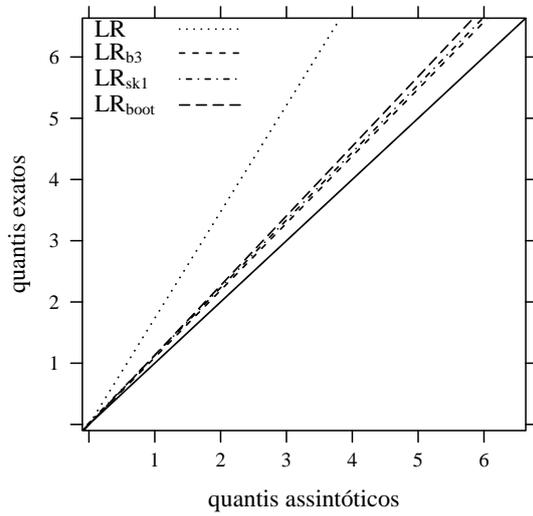
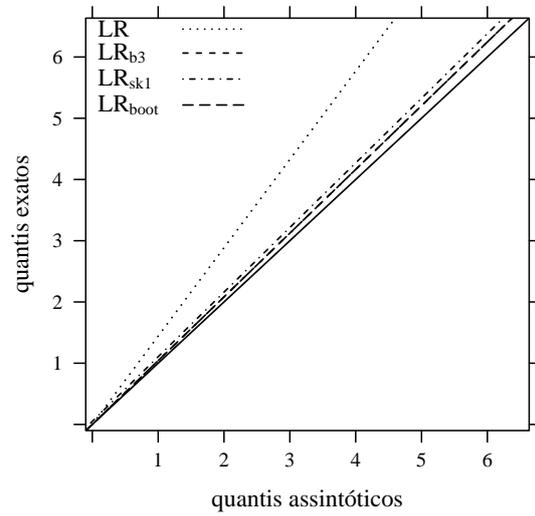
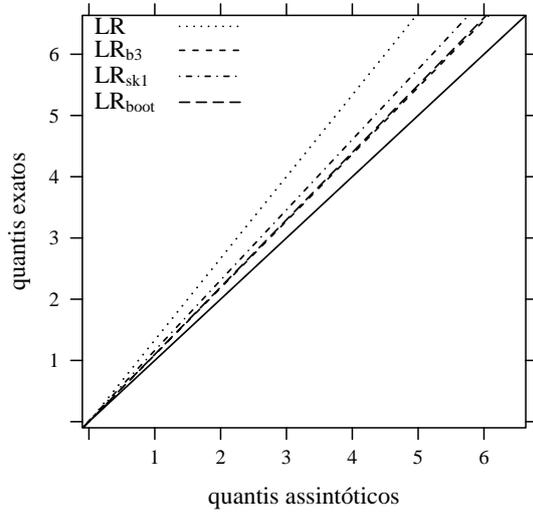
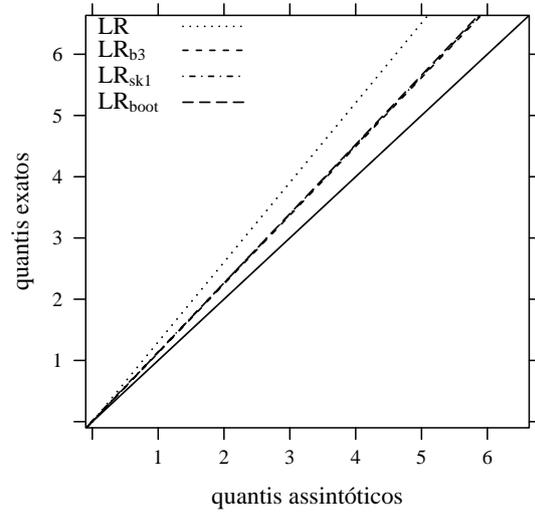
Tabela 1.4 Momentos e quantis estimados das estatísticas de teste, considerando $q = 2$, $\phi = 30$ e $n = 20$.

	Média	Variância	Assimetria	Curtose	Quantil 90	Quantil 95	Quantil 99
χ_q^2	2.0000	4.0000	2.0000	9.0000	4.6052	5.9915	9.2103
LR	2.6741	7.2829	2.0784	9.7003	6.1775	8.0134	12.2788
LR_{b1}	2.1353	4.6449	2.0799	9.7146	4.9319	6.4065	9.7992
LR_{b2}	2.0777	4.3982	2.0804	9.7182	4.7979	6.2333	9.5338
LR_{b3}	1.9993	4.0729	2.0810	9.7243	4.6147	5.9960	9.1731
LR_{sk1}	2.0127	4.2331	2.1816	11.5872	4.6612	6.0227	9.2845
LR_{sk2}	2.0906	4.6776	3.1049	27.7738	4.7836	6.2003	9.5926
LR_{boot}	2.0024	4.1168	2.1086	9.9347	4.6103	5.9856	9.2791

função kernel gaussiana. O gráfico Q-Q, Figura 1.2, deixa claro que as distribuições nulas das estatísticas corrigidas estão muito mais próximas da distribuição de referência do que a da estatística LR . Dentre as estatísticas corrigidas LR_{b3} apresenta melhor adequação entre quantis empíricos e assintóticos. As densidades empíricas, apresentadas na Figura 1.3, corroboram com os resultados anteriores, evidenciando a maior proximidade das distribuições nulas das estatísticas corrigidas à distribuição assintótica do que a da estatística LR . A distribuição nula de LR mostra-se pobremente aproximada pela distribuição χ_q^2 , confirmando a tendência da estatística da razão de verossimilhanças usual de rejeitar a hipótese nula com frequência superior ao nível nominal. O gráfico das densidades empíricas também deixa clara a melhor aproximação das distribuições nulas das estatísticas corrigidas por Bartlett relativamente à estatística ajustada por Skovgaard.

Após análise das taxas de rejeição nula dos testes e das aproximações das distribuições das estatísticas de teste pela distribuição nula limite, investigamos as taxas de rejeição não-nula, ou seja, os poderes dos testes. A Tabela 1.5 apresenta as taxas de rejeição obtidas sob as seguintes hipóteses alternativas: $\beta_2 = \delta$ ($q = 1$), $\beta_2 = \beta_3 = \delta$ ($q = 2$) e $\beta_2 = \beta_3 = \beta_4 = \delta$ ($q = 3$), para diferentes valores de δ . Para a comparação de poder consideramos apenas os testes corrigidos baseados nas estatísticas LR_{b3} , LR_{sk1} e LR_{boot} , pois esses foram os testes que se destacaram em termos de tamanho e aproximação da estatística de teste pela distribuição nula limite. Por ser notavelmente liberal o teste da razão de verossimilhanças também não foi considerado.

Analisando a Tabela 1.5 notamos que, como era de se esperar, à medida que o valor de δ se afasta de zero os testes se tornam mais poderosos. É possível notar que os testes corrigidos apresentam desempenho similar em termos de poder. No entanto, o teste baseado na LR_{sk1} se mostrou levemente mais poderoso, principalmente para $\delta > 0$. Para $\delta < 0$ observou-se que o teste LR_{b3} possui maior poder em alguns cenários, como, por exemplo, para $\phi = 5$ e $q = 3$, assim como quando $\phi = 10$ e $q = 1$. Quando os valores de δ são mais próximos de zero, considerando $\delta = -0.5$, $q = 1$, $\phi = 10$ e 5, por exemplo, o teste LR_{boot} torna-se mais poderoso do que os outros testes apresentados.

(a) $n = 15$ (b) $n = 20$ (c) $n = 30$ (d) $n = 40$ **Figura 1.2** Gráfico Q-Q considerando $\phi = 10$ e $q = 1$.

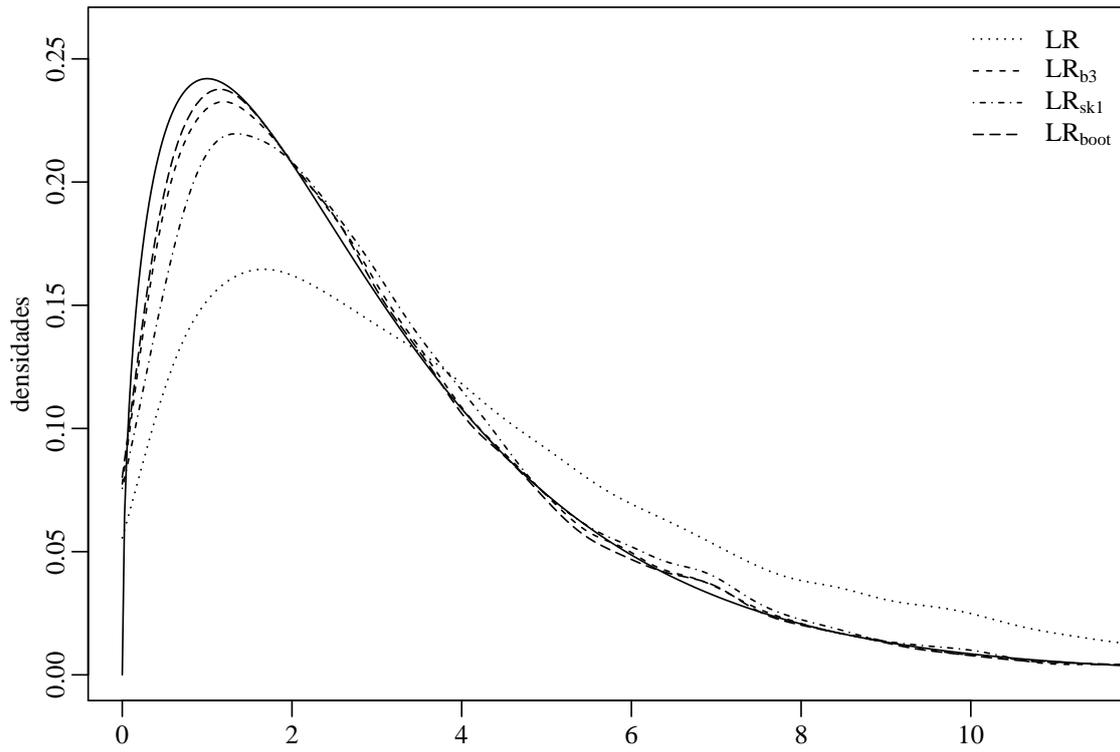
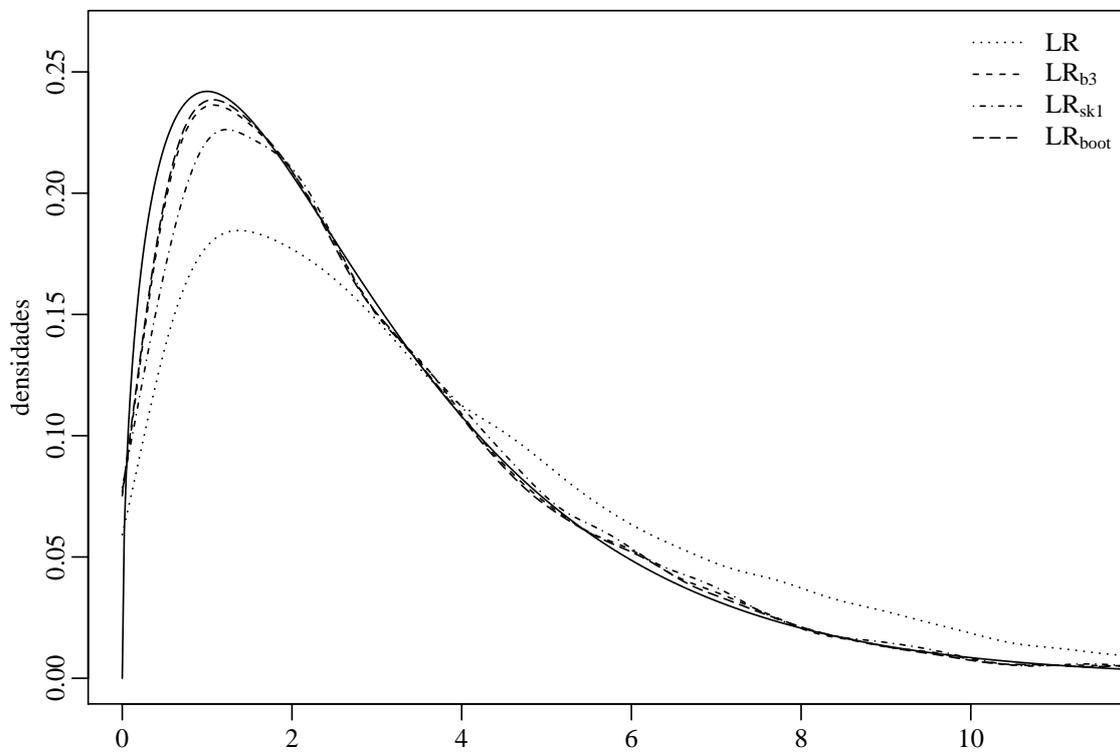
(a) $n = 15$ (b) $n = 20$

Figura 1.3 Densidade χ_q^2 (linha sólida) e densidades nulas estimadas das estatísticas de teste para $\phi = 5$ e $q = 3$.

Tabela 1.5 Taxas de rejeição não-nula (percentuais) dos testes corrigidos para $n = 20$ e $\alpha = 5\%$.

ϕ	Estat	δ											
		-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	0.5	1.0	1.5	2.0	2.5	3.0
$q = 1$													
30	LR_{b3}	100	100	99.8	96.5	71.1	24.8	25.2	71.8	96.3	99.8	100	100
	LR_{sk1}	100	100	99.8	96.4	70.7	24.7	25.4	72.3	96.4	99.8	100	100
	LR_{boot}	100	100	99.8	96.4	71.1	24.8	25.1	71.9	96.1	99.8	100	100
10	LR_{b3}	99.2	96.2	85.3	62.0	33.3	11.6	12.5	34.7	63.5	86.0	96.2	99.1
	LR_{sk1}	99.2	96.0	85.2	61.7	33.3	11.5	13.2	36.0	64.8	86.8	96.5	99.2
	LR_{boot}	99.2	96.3	85.6	62.4	33.7	12.0	12.7	34.9	63.6	85.8	96.2	99.1
5	LR_{b3}	90.8	79.6	61.2	39.6	20.4	9.0	9.6	21.5	41.4	62.1	81.2	91.8
	LR_{sk1}	91.1	79.8	61.5	39.8	20.7	9.2	10.6	23.2	43.6	64.4	82.8	92.6
	LR_{boot}	91.3	80.4	62.1	40.6	21.4	9.6	9.8	21.8	41.4	62.1	80.9	91.5
$q = 2$													
30	LR_{b3}	100	100	100	99.7	87.7	32.7	31.6	88.1	99.7	100	100	100
	LR_{sk1}	100	100	99.9	99.7	88.0	32.9	31.4	88.1	99.7	100	100	100
	LR_{boot}	98.6	100	99.9	99.7	87.7	32.8	31.6	88.0	99.6	100	100	100
10	LR_{b3}	100	99.7	97.6	82.2	47.2	15.3	15.6	47.5	82.5	97.4	99.7	100
	LR_{sk1}	100	99.8	97.8	82.9	47.8	15.5	15.8	48.0	82.9	97.6	99.7	100
	LR_{boot}	99.5	99.7	97.6	82.3	47.2	15.3	15.5	47.7	82.2	97.3	99.7	100
5	LR_{b3}	99.2	95.7	84.1	60.6	25.8	11.6	9.5	26.6	53.8	79.3	93.7	98.4
	LR_{sk1}	99.3	96.0	84.8	61.5	25.6	11.8	10.4	27.9	56.1	81.0	94.7	98.8
	LR_{boot}	99.0	95.8	84.5	61.4	26.2	11.9	9.7	26.2	53.5	78.8	93.5	98.3
$q = 3$													
30	LR_{b3}	100	100	100	100	96.2	41.9	41.6	95.1	99.9	100	100	100
	LR_{sk1}	100	100	100	100	96.0	41.8	41.5	94.9	100	100	100	100
	LR_{boot}	99.2	99.9	100	100	96.2	42.2	41.5	95.0	100	100	100	99.4
10	LR_{b3}	100	100	99.2	91.5	58.6	17.8	17.0	56.9	89.7	98.5	99.9	100
	LR_{sk1}	100	100	99.2	91.4	58.3	17.8	17.3	57.2	89.7	98.7	99.9	100
	LR_{boot}	99.7	100	99.2	91.4	58.3	17.7	17.2	56.9	89.6	98.5	99.9	99.9
5	LR_{b3}	99.6	98.0	90.4	69.4	35.8	11.3	11.8	35.6	68.0	88.7	97.4	99.3
	LR_{sk1}	99.6	97.9	90.2	69.3	35.7	11.3	12.3	36.1	68.5	89.6	97.7	99.5
	LR_{boot}	99.4	97.9	90.4	69.6	35.8	11.2	11.7	35.3	67.7	88.4	97.2	99.3

1.5 Aplicação

Esta seção contém uma aplicação dos testes da razão de verossimilhanças corrigidos apresentados e avaliados anteriormente. Consideramos apenas as estatísticas corrigidas LR_{b3} e LR_{boot} comparativamente à estatística da razão de verossimilhanças não corrigida (LR). Os dados

utilizados foram retirados de Griffiths *et al.*(1993, Tabela 15.4) e referem-se a gastos com alimentação, renda e número de pessoas por domicílio em uma grande cidade dos Estados Unidos. A modelagem destes dados também foi considerada no artigo seminal dos modelos de regressão beta (Ferrari e Cribari-Neto, 2004).

Ferrari e Cribari-Neto (2004) modelam a proporção de gastos com alimentação (y) como função da renda (x_2) e do número de pessoas (x_3) de cada residência. Contudo, consideramos ainda como covariáveis candidatas a interação entre renda e número de pessoas ($x_4 = x_2 \times x_3$), os quadrados das variáveis renda ($x_5 = x_2^2$) e número de pessoas ($x_6 = x_3^2$). Com isso, temos o seguinte modelo

$$\text{logit}(\mu_i) = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i}, \quad (1.12)$$

com $i = 1, \dots, 38$, ou seja, há 38 observações.

Primeiramente, desejamos fazer inferência sobre a significância da interação (x_4) no modelo, i.e., desejamos testar $\mathcal{H}_0 : \beta_4 = 0$. Para esse teste, a estatística da razão de verossimilhanças (LR) é igual a 3.859 (p -valor: 0.049) e as estatísticas corrigidas são $LR_{b3} = 3.208$ (p -valor: 0.073) e $LR_{boot} = 3.192$ (p -valor: 0.074). Esses resultados mostram que as inferências mudam quando são utilizadas as estatísticas corrigidas. Ao nível nominal de 5% o teste da razão de verossimilhanças rejeita a hipótese nula, enquanto que utilizando as estatísticas corrigidas as conclusões inferenciais são contrárias.

Seguindo as conclusões inferenciais dos testes corrigidos removemos a variável x_4 (interação) e estimamos o seguinte modelo:

$$\text{logit}(\mu_i) = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_5 x_{5i} + \beta_6 x_{6i}.$$

Para esse modelo, as estimativas pontuais (com erros-padrão entre parênteses) são: $\hat{\beta}_1 = 0.4861$ (0.5946), $\hat{\beta}_2 = -0.0495$ (0.0218), $\hat{\beta}_3 = 0.0172$ (0.1563), $\hat{\beta}_5 = 0.0003$ (0.0002), $\hat{\beta}_6 = 0.0129$ (0.0198) e $\hat{\phi} = 39.296$ (8.925). Os erros-padrão das estimativas são muito altos, tornando não significativas, aos níveis nominais usuais, as variáveis x_3 , x_5 e x_6 . Com isso, surge o interesse em testar $\mathcal{H}_0 : \beta_5 = \beta_6 = 0$. Para este teste os valores das estatísticas são $LR = 3.791$ (p -valor: 0.15), $LR_{b3} = 3.296$ (p -valor: 0.192) e $LR_{boot} = 3.210$ (p -valor: 0.201). Neste caso, os três testes não rejeitam a hipótese nula.

Os resultados inferenciais até aqui apontados nos levam a considerar o seguinte modelo reduzido:

$$\text{logit}(\mu_i) = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i}.$$

As estimativas pontuais (com erros-padrão entre parênteses) para esse modelo são: $\hat{\beta}_1 = -0.6225$ (0.224), $\hat{\beta}_2 = -0.0123$ (0.003), $\hat{\beta}_3 = 0.1185$ (0.035) e $\hat{\phi} = 35.61$ (8.080).

Por outro lado, voltando ao modelo (1.12) testemos diretamente $\mathcal{H}_0 : \beta_4 = \beta_5 = \beta_6 = 0$. Para esse teste temos os seguintes valores das estatísticas de testes com seus p -valores associados: $LR = 7.6501$ (p -valor: 0.054), $LR_{b3} = 6.554$ (p -valor: 0.088) e $LR_{boot} = 6.068$ (p -valor: 0.108). Esses últimos resultados também mostram o comportamento mais liberal do teste não corrigido. O p -valor do teste LR está próximo ao nível nominal de 5%, enquanto que os testes corrigidos possuem p -valores mais próximos ao nível nominal de 10%. O teste baseado em LR_{boot} , por exemplo, rejeita a hipótese nula ao nível de 10% de significância.

1.6 Conclusões

Regressão beta é uma importante classe de modelos de regressão para modelar variáveis que assumem valores no intervalo $(0,1)$. Nesses modelos assume-se que a variável resposta possui distribuição beta, a qual é bastante flexível, acomodando distribuições das mais diferentes formas. Por essa flexibilidade, e pela parametrização adotada em termos dos parâmetros de média e precisão, os modelos de regressão beta vêm sendo aplicados em larga escala.

Usualmente, após a estimação, tem-se interesse em realizar testes de hipóteses sobre os parâmetros do modelo. Contudo, as inferências nesses modelos baseiam-se em aproximações assintóticas, conduzindo a testes de hipóteses com tamanhos distorcidos em pequenas amostras. Neste sentido, este capítulo focou em correções para pequenas amostras do teste da razão de verossimilhanças em regressão beta.

A estatística da razão de verossimilhanças, apesar de ter boas propriedades assintóticas, tipicamente não tem distribuição nula exata bem aproximada pela distribuição nula limite qui-quadrado em amostras de tamanho reduzido, implicando distorções no teste de hipótese. Para corrigir esse problema consideramos duas abordagens da correção de Bartlett. Em primeiro lugar obtemos uma correção analítica em forma matricial que pode ser facilmente implementada em linguagens matriciais de programação. Consideramos ainda a correção de Bartlett bootstrap, na qual o fator de correção de Bartlett é determinado por meio do esquema bootstrap paramétrico. Comparativamente às correções propostas consideramos o ajuste de Skovgaard para o modelo de regressão beta já disponível na literatura.

Os resultados numéricos evidenciam que as distribuições nulas das estatísticas corrigidas por Bartlett são melhor aproximadas pela distribuição assintótica do que as distribuições nula da estatística de Skovgaard e a da estatística não corrigida. Em termos de tamanho de teste, o teste da razão de verossimilhanças corrigido por Bartlett se mostrou com menor distorção, alcançando taxas de rejeição da hipótese nula mais próximas dos níveis nominais do que os demais testes considerados. À medida que o tamanho amostral diminui, o desempenho superior do teste corrigido por Bartlett se torna mais evidente quando comparado ao teste que utiliza o ajuste de Skovgaard (cf. Tabela C.1). A correção de Bartlett bootstrap também apresentou bom desempenho em termos de taxa de rejeição nula, alcançando desempenho equivalente ou superior ao ajuste de Skovgaard em diversos cenários. Essa correção, que utiliza bootstrap, mostrou-se uma ótima alternativa às correções analíticas, principalmente para os maiores valores do parâmetro de precisão. Quanto às taxas de rejeição não-nula, as correções propostas não apresentam notável diferença de poder comparativamente à estatística de Skovgaard. Ainda, a aplicação a dados reais apresentada na seção anterior mostra que o uso do teste da razão de verossimilhanças corrigido pode inverter as conclusões inferenciais. A utilização das estatísticas corrigidas nos remete a inferências mais acuradas, devido ao comportamento bastante liberal do teste da razão de verossimilhanças usual.

Recomendamos a utilização do teste da razão de verossimilhanças corrigido baseado na estatística LR_{b3} para inferências em pequenas amostras. O teste baseado nessa estatística corrigida por Bartlett apresenta taxas de rejeição nula mais próximas aos níveis nominais do que outras estatísticas corrigidas consideradas. Adicionalmente, a estatística LR_{b3} se mostrou superior às demais em termos de aproximação em pequenas amostras pela distribuição nula assintótica.

Modelos de Regressão Beta com Dispersão Variável: Avaliação de Critérios de Seleção de Modelos em Amostras de Tamanho Finito

Resumo do capítulo

Seleção de modelos em análise de regressão consiste em escolher um conjunto de covariáveis úteis a partir de um conjunto mais amplo de regressores. Este capítulo tem o objetivo de investigar, por meio de simulação de Monte Carlo, os desempenhos de diferentes critérios de seleção em amostras de tamanho finito aplicados a modelos de regressão beta com dispersão variável. Tal modelo assume que a variável resposta possui distribuição beta com parâmetros de média e dispersão e, assim como a média, a dispersão é modelada em termos de covariáveis e parâmetros desconhecidos. Os resultados revelam que a seleção de covariáveis para média e dispersão conjuntamente não constitui boa prática. Nós propomos um esquema de seleção de modelos em duas etapas e critérios específicos para o modelo com dispersão variável. Além da estratégia de seleção proposta envolver menor custo computacional, os resultados numéricos mostram que sua utilização implica uma seleção de modelos mais acurada do que a seleção simultânea dos dois submodelos. Apresentamos uma aplicação a dados reais e verificamos o bom ajuste do modelo selecionado por meio de algumas ferramentas gráficas para análise de resíduos.

Palavras-chave: Critérios de seleção de modelos, dispersão variável, regressão beta.

2.1 Introdução

A análise de regressão é uma técnica estatística utilizada para investigar e modelar, com base em um banco de dados, a relação entre uma variável de interesse e um conjunto de variáveis explicativas. O modelo de regressão normal linear é bastante utilizado em análises empíricas. No entanto, tal modelo torna-se inapropriado em situações em que a variável resposta é restrita ao intervalo $(0,1)$, como ocorre com taxas e proporções. Nestes casos, o modelo de regressão normal linear pode prever valores da variável resposta fora do intervalo unitário padrão. Uma possível solução é transformar a variável dependente, porém, os parâmetros tornam-se de difícil interpretação em termos da variável resposta original. Adicionalmente, em geral, medidas de razão e/ou proporção apresentam comportamento distribucional assimétrico, não satisfazendo assim simetria implicada pela suposição de normalidade.

Muitos estudos, em diferentes áreas do conhecimento, examinam como um conjunto de variáveis se relaciona com algum tipo de porcentagem ou proporção, como em Brehm e Gates (1993), Kieschnick e McCullough (2003), Smithson e Verkuilen (2006) e Zucco (2008). Para estas situações, Ferrari e Cribari-Neto (2004) propuseram a classe de modelos de regressão beta, em que a variável resposta (y) segue lei beta.

A distribuição beta é bastante flexível para modelar proporções, uma vez que, dependendo dos valores dos dois parâmetros que a indexam, a densidade assume formas bem variadas. A densidade beta é dada por

$$\pi_1(y; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1} (1-y)^{q-1}, \quad 0 < y < 1,$$

em que $p > 0$ e $q > 0$ são parâmetros que indexam a distribuição e $\Gamma(\cdot)$ é a função gama:

$$\Gamma(p) = \int_0^\infty y^{p-1} e^{-y} dy.$$

A média e a variância da variável y são dadas, respectivamente, por

$$E(y) = \frac{p}{p+q}, \tag{2.1}$$

$$\text{var}(y) = \frac{pq}{(p+q)^2(p+q+1)}. \tag{2.2}$$

No modelo proposto em Ferrari e Cribari-Neto (2004) é utilizada uma reparametrização da densidade beta, sendo $\mu = p/(p+q)$ e $\phi = p+q$, isto é, $p = \mu\phi$ e $q = (1-\mu)\phi$. Com isso, a partir de (2.1) e (2.2), temos que

$$E(y) = \mu$$

e

$$\text{var}(y) = \frac{V(\mu)}{1+\phi},$$

em que $V(\mu) = \mu(1-\mu)$ denota a função de variância e ϕ pode ser interpretado como um parâmetro de precisão, uma vez que, para μ fixo, se o valor de ϕ aumenta, então a variância da variável resposta diminui. Deste modo, a densidade de y pode ser escrita como

$$\pi_2(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1}, \quad 0 < y < 1,$$

em que $0 < \mu < 1$ e $\phi > 0$.

Tais modelos de regressão beta apresentam grande aplicabilidade em situações de modelagem cujo objetivo é estudar a relação entre uma variável y , que assume continuamente valores no intervalo unitário padrão $(0,1)$, e outras variáveis que afetam seu comportamento através de uma estrutura de regressão. A resposta média é relacionada a um preditor linear, que incorpora covariáveis e parâmetros desconhecidos, através de uma função de ligação. Deste modo, a

modelagem e os procedimentos inferenciais dos modelos de regressão beta são similares aos modelos lineares generalizados (McCullagh e Nelder, 1989).

No modelo de regressão beta proposto em Ferrari e Cribari-Neto (2004), o parâmetro ϕ é uma constante. Isto implica que a dispersão (recíproco da precisão) é constante para todas as observações. Entretanto, as perdas de eficiência decorrentes de se tomar ϕ constante erroneamente podem ser substanciais. Este fato pode ser visualizado na Figura 2.1, que apresenta as densidades estimadas dos estimadores de máxima verossimilhança do parâmetro de inclinação ($\beta_2 = 1.5$) sob dispersão variável, com e sem modelagem da dispersão. As estimativas das densidades foram construídas a partir de uma simulação de Monte Carlo com cinco mil réplicas. O processo gerador dos dados incluía dispersão variável em um modelo de regressão beta com modelo da média $g(\mu_t) = \gamma_1 + \gamma_2 x_{1t}$, em que $g(\cdot)$ é a função logit.

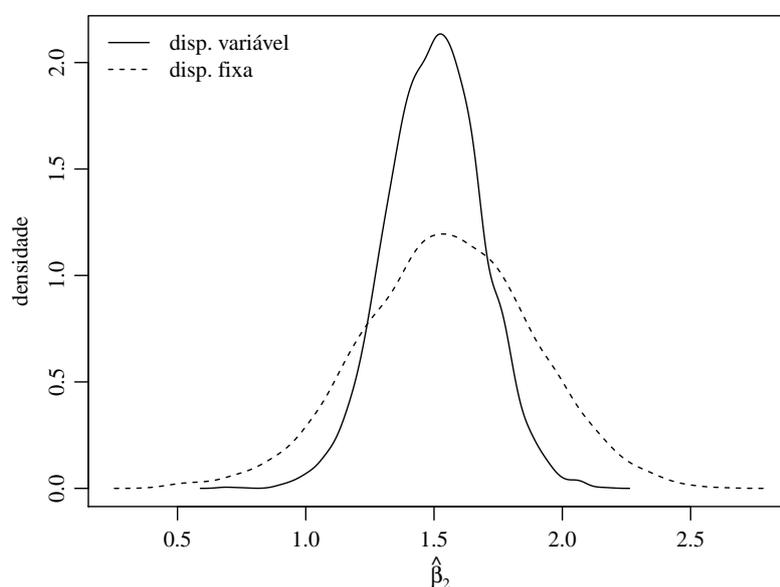


Figura 2.1 Densidades estimadas dos estimadores do parâmetro de inclinação sob dispersão variável, com (linha cheia) e sem (linha tracejada) modelagem da dispersão.

De fato, como ilustrado na Figura 2.1, podemos perceber que a estimação eficiente dos parâmetros em uma regressão depende da modelagem correta da dispersão. No contexto de modelos lineares generalizados (MLG), trabalhos como Nelder e Lee (1991) e Smyth e Verbyla (1999) definem os modelos lineares generalizados duplos, ou *joint generalized linear models*, permitindo a modelagem simultânea da média e da variância da variável resposta. Nesta perspectiva, Cuervo-Cepeda e Gamerman (2004), Smithson e Verkuilen (2006), Espinheira (2007) e Simas *et al.* (2010) sugerem o modelo de regressão beta com parâmetro de dispersão variável.

O modelo de regressão beta com dispersão variável pode ser entendido como uma extensão do modelo proposto em Ferrari e Cribari-Neto (2004) para situações em que o parâmetro de dispersão não é constante ao longo das observações. A dispersão é modelada em termos de covariáveis e de parâmetros desconhecidos, da mesma forma que a média condicional. Com isso, a seleção de modelos torna-se uma etapa importante no procedimento de modelagem, ou seja, a escolha correta, ou melhor escolha, das variáveis regressoras a serem incluídas nos dois

preditores lineares é fundamental.

No âmbito da análise de regressão linear, assim como em MLG, a seleção de modelos é reconhecidamente parte fundamental do processo de modelagem e constitui área de pesquisa amplamente explorada. O foco principal de pesquisa dentro da seleção de modelos se dá nos chamados critérios de seleção ou critérios de informação. No entanto, não são encontrados na literatura estudos sobre o uso de critérios de seleção em modelos de regressão beta com dispersão variável.

O objetivo deste trabalho é avaliar os desempenhos de diferentes critérios de seleção de modelos em amostras de tamanho finito em regressão beta. Os critérios são utilizados para determinar o melhor modelo para a média e para a dispersão do modelo de regressão beta com dispersão variável, em que a média e a dispersão são modeladas simultaneamente e os parâmetros dos dois submodelos são estimados por máxima verossimilhança. Para tanto, nós propomos critérios de seleção específicos para o modelo com dispersão variável e um esquema de seleção de modelos em duas etapas. O esquema proposto possui um menor custo computacional do que a abordagem usual de seleção conjunta, seguindo os seguintes passos: (i) primeiramente consideramos dispersão constante e selecionamos os regressores para a média; (ii) assumindo que o submodelo selecionado para a média é o verdadeiro, utilizamos critérios de seleção para determinar o submodelo da dispersão. Os resultados numéricos mostram que o esquema proposto conduz a um aumento na frequência de seleção do verdadeiro modelo quando comparado à seleção conjunta de regressores para os submodelos da média e da dispersão.

O presente trabalho encontra-se dividido da seguinte maneira. Na seção seguinte apresentamos o modelo de regressão beta com dispersão variável, descrevendo a estimação por máxima verossimilhança dos parâmetros do modelo, assim como um teste score para dispersão constante. Na Seção 2.3 são apresentados os critérios de seleção utilizados na análise. A Seção 2.4 apresenta os resultados numéricos, a discussão dos desempenhos dos critérios de seleção em amostras de tamanho finito e a proposta de seleção de modelos em duas etapas. Na Seção 2.5 apresentamos uma aplicação a dados reais de habilidade de leitura de crianças (Pammer e Kevan, 2007). Na última seção são apresentadas as principais conclusões e algumas recomendações.

2.2 Definição e estimação do modelo

Tome uma variável aleatória y com distribuição beta indexada pelos parâmetros μ e ϕ . O objetivo é definir um modelo de regressão para modelar a média e a dispersão da variável resposta y . Para isso, trabalhamos com uma parametrização diferente da densidade beta. Tomando $\mu = p/(p+q)$ e $\sigma^2 = 1/(1+\phi)$, segue que

$$E(y) = \mu,$$

$$\text{var}(y) = V(\mu)\sigma^2.$$

Com essa reparametrização, a densidade de y pode ser escrita como

$$f(y; \mu, \sigma) = \frac{\Gamma(\frac{1-\sigma^2}{\sigma^2})}{\Gamma(\mu(\frac{1-\sigma^2}{\sigma^2}))\Gamma((1-\mu)(\frac{1-\sigma^2}{\sigma^2}))} y^{\mu(\frac{1-\sigma^2}{\sigma^2})-1} (1-y)^{(1-\mu)(\frac{1-\sigma^2}{\sigma^2})-1}, \quad 0 < y < 1, \quad (2.3)$$

em que $0 < \mu < 1$ e $0 < \sigma < 1$. Desta forma, temos um parâmetro de dispersão (σ) e não mais um parâmetro de precisão (ϕ). Além disso, como σ pertence ao intervalo $(0,1)$, é fácil verificar situações de grande dispersão e pequena dispersão. A Figura 2.2 apresenta formas variadas assumidas pela densidade beta considerando diferentes valores para os parâmetros μ e σ .

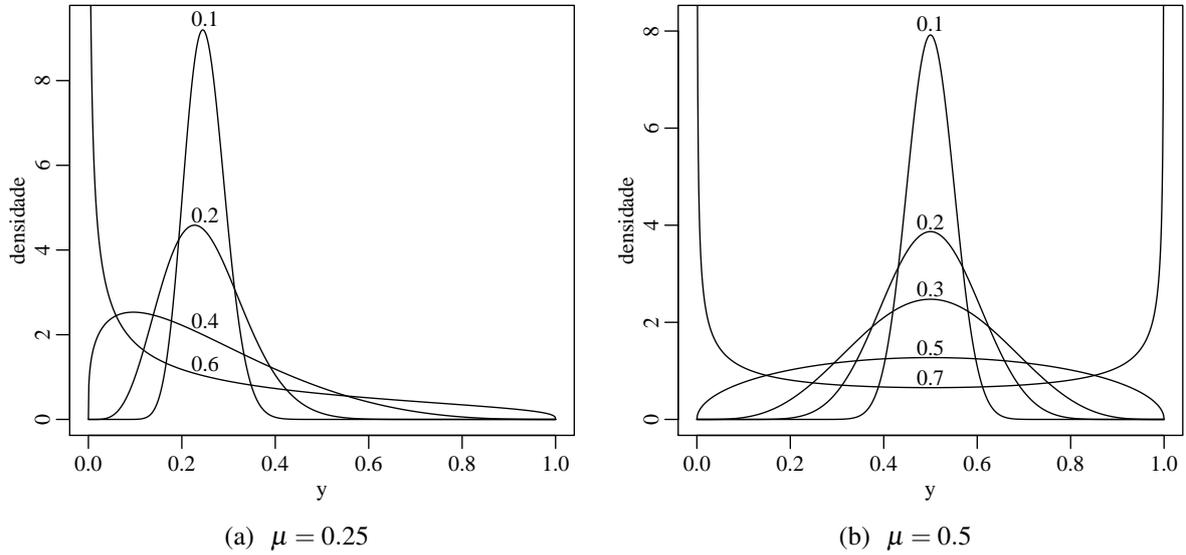


Figura 2.2 Densidades beta para diferentes valores de σ (indicados no gráfico), com $\mu = 0.25$ (a) e $\mu = 0.5$ (b).

Sejam y_1, \dots, y_n variáveis aleatórias independentes, em que cada y_t , $t = 1, \dots, n$, tem densidade na forma (2.3) com média μ_t e parâmetro de dispersão desconhecido σ_t . Com isso, o modelo de regressão beta com parâmetro de dispersão variável pode ser escrito como

$$\begin{aligned} g(\mu_t) &= \sum_{i=1}^r x_{ti} \beta_i = \eta_t, \\ h(\sigma_t) &= \sum_{i=1}^s z_{ti} \gamma_i = v_t, \end{aligned} \quad (2.4)$$

em que $\beta = (\beta_1, \dots, \beta_r)^\top$ e $\gamma = (\gamma_1, \dots, \gamma_s)^\top$ são vetores de parâmetros desconhecidos para a média e dispersão, respectivamente. Adicionalmente, x_{t1}, \dots, x_{tr} e z_{t1}, \dots, z_{ts} são observações de r e s variáveis independentes conhecidas, em que $r + s = k < n$. Quando interceptos são incluídos nos submodelos da média e da dispersão, temos que $x_{t1} = z_{t1} = 1$, para $t = 1, \dots, n$.

Finalmente, $g(\cdot)$ e $h(\cdot)$ são funções de ligação estritamente monótonas e duas vezes diferenciáveis, com domínio em $(0,1)$ e imagem em \mathbb{R} . Para o modelo de regressão beta, com a parametrização proposta, podemos escolher diferentes funções de ligação como, por exemplo, logit, probit, log-log, log-log complementar e Cauchy. Note que as mesmas funções de ligação do submodelo da média podem ser usadas com a dispersão. Uma discussão detalhada sobre essas e outras funções de ligação pode ser encontrada em McCullagh e Nelder (1989).

Para a estimação conjunta dos vetores de parâmetros β e γ é utilizado o método da máxima verossimilhança. O logaritmo da função de verossimilhança para n variáveis aleatórias independentes é

$$\ell(\beta, \gamma) = \sum_{t=1}^n \ell_t(\mu_t, \sigma_t),$$

em que

$$\begin{aligned} \ell_t(\mu_t, \sigma_t) &= \log \Gamma\left(\frac{1-\sigma_t^2}{\sigma_t^2}\right) - \log \Gamma\left(\mu_t \left(\frac{1-\sigma_t^2}{\sigma_t^2}\right)\right) - \log \Gamma\left((1-\mu_t) \left(\frac{1-\sigma_t^2}{\sigma_t^2}\right)\right) \\ &+ \left[\mu_t \left(\frac{1-\sigma_t^2}{\sigma_t^2}\right) - 1\right] \log y_t + \left[(1-\mu_t) \left(\frac{1-\sigma_t^2}{\sigma_t^2}\right) - 1\right] \log(1-y_t). \end{aligned}$$

A função escore é obtida pela diferenciação da função de log-verossimilhança em relação aos parâmetros desconhecidos. O vetor escore relativo a β é dado por

$$U_\beta(\beta, \gamma) = X^\top \Phi T(y^* - \mu^*),$$

em que X é uma matriz $n \times r$ cuja t -ésima linha é x_t , $T = \text{diag}\{1/g'(\mu_1), \dots, 1/g'(\mu_n)\}$, $\Phi = \text{diag}\left\{\frac{1-\sigma_1^2}{\sigma_1^2}, \dots, \frac{1-\sigma_n^2}{\sigma_n^2}\right\}$, $y^* = (y_1^*, \dots, y_n^*)^\top$, $\mu^* = (\mu_1^*, \dots, \mu_n^*)^\top$, $y_t^* = \log\left(\frac{y_t}{1-y_t}\right)$, $\mu_t^* = \psi\left(\mu_t \left(\frac{1-\sigma_t^2}{\sigma_t^2}\right)\right) - \psi\left((1-\mu_t) \left(\frac{1-\sigma_t^2}{\sigma_t^2}\right)\right)$ e $\psi(\cdot)$ é a função digamma, isto é, $\psi(u) = \frac{d \log \Gamma(u)}{du}$, para $u > 0$.

Considerando as derivadas da função de log-verossimilhança em relação ao vetor de parâmetros do submodelo da dispersão, γ , temos que o vetor escore é

$$U_\gamma(\beta, \gamma) = Z^\top H a,$$

em que Z é uma matriz $n \times s$ cuja t -ésima linha é z_t , $H = \text{diag}\{1/h'(\sigma_1), \dots, 1/h'(\sigma_n)\}$, $a_t = -\frac{2}{\sigma_t^3} \left\{ \mu_t (y_t^* - \mu_t^*) + \log(1-y_t) - \psi\left(\frac{(1-\mu_t)(1-\sigma_t^2)}{\sigma_t^2}\right) + \psi\left(\frac{(1-\sigma_t^2)}{\sigma_t^2}\right) \right\}$ e $a = (a_1, \dots, a_n)^\top$.

Os estimadores de máxima verossimilhança são obtidos pela solução do seguinte sistema:

$$\begin{cases} U_\beta(\beta, \gamma) = 0, \\ U_\gamma(\beta, \gamma) = 0. \end{cases}$$

A solução deste sistema não possui forma fechada, fazendo-se necessário o uso de algoritmos de otimização não-linear para a obtenção de estimativas de máxima verossimilhança.

Com o objetivo de verificar possíveis afastamentos das suposições feitas para o modelo, é útil visualizar graficamente os resíduos. Vários tipos de resíduos estão disponíveis para o

modelo de regressão beta. Amplos trabalhos a respeito de resíduos e análise de diagnóstico em regressão beta são encontrados em Espinheira *et al.* (2008a,b) e Ferrari *et al.* (2011). Dentre os diversos resíduos propostos o *resíduo ponderado padronizado 2* é o que apresenta as melhores propriedades. Este resíduos é dado por

$$r_{pp2} = \frac{y_t^* - \hat{\mu}_t^*}{\sqrt{\hat{v}_t(1 - h_{tt})}}, \quad (2.5)$$

em que $v_t^* = \psi' \left(\mu_t \left(\frac{1 - \sigma_t^2}{\sigma_t^2} \right) \right) + \psi' \left((1 - \mu_t) \left(\frac{1 - \sigma_t^2}{\sigma_t^2} \right) \right)$ e h_{tt} é o t -ésimo elemento da diagonal da matriz chapéu (para detalhes, ver Espinheira *et al.* (2008b) e Ferrari *et al.* (2011)).

Contudo, como temos o interesse em modelar a possível variabilidade da dispersão em termos de covariáveis e parâmetros desconhecidos, torna-se importante desenvolver uma estratégia pré-teste, a fim de testar a hipótese de dispersão constante. Baseados nos resultados de simulação apresentados no Apêndice F, em que avaliamos numericamente o tamanho e o poder dos testes da razão de verossimilhanças e escore, sugerimos a utilização do teste escore para determinar se a dispersão deve ser modelada.

Para tal teste, considere a seguinte hipótese nula de dispersão constante:

$$\mathcal{H}_0 : \sigma_1 = \sigma_2 = \dots = \sigma_n = \sigma,$$

ou, equivalentemente,

$$\mathcal{H}_0 : \gamma_i = 0, \quad i = 2, \dots, s,$$

para o modelo definido em (2.4) com $z_{t1} = 1$, para $t = 1, \dots, n$. A estatística escore é

$$S = \tilde{U}_{(s-1)\gamma}^\top \tilde{K}_{(s-1)(s-1)}^{-1} \tilde{U}_{(s-1)\gamma},$$

em que $\tilde{U}_{(s-1)\gamma}^\top$ é o vetor que contém os $s - 1$ elementos finais da função escore de γ avaliado no estimador de máxima verossimilhança restrito (sob \mathcal{H}_0) e $\tilde{K}_{(s-1)(s-1)}^{-1}$ é a matriz $(s - 1) \times (s - 1)$ formada pelas últimas $s - 1$ linhas e últimas $s - 1$ colunas da inversa da matriz de informação de Fisher $K(\beta, \gamma)$ avaliada no estimador restrito. A obtenção da matriz $K(\beta, \gamma)$ é apresentada em (E.1) no Apêndice E.

Sob condições usuais de regularidade e sob \mathcal{H}_0 , a estatística S converge em distribuição para $\chi_{(s-1)}^2$, de forma que o teste pode ser realizado usando valores críticos aproximados obtidos como quantis da distribuição $\chi_{(s-1)}^2$. Se a estatística de teste S é maior que o valor crítico selecionado a um nível de significância determinado, então rejeita-se \mathcal{H}_0 e decide-se modelar a dispersão.

2.3 Seleção de modelos

Uma das perguntas mais importantes na modelagem estatística, nas mais diversas áreas de aplicação, talvez seja “Quais variáveis explicativas são importantes para o modelo?”. Para responder a essa pergunta, é possível utilizar os critérios de seleção de modelos, que balanceiam a complexidade do modelo e sua capacidade explicativa.

Na literatura de regressão linear vários métodos têm sido propostos para seleção de modelos. O primeiro critério de seleção amplamente usado foi o R^2 ajustado. Tal critério utiliza o coeficiente de determinação (R^2), penalizando-o à medida que variáveis explicativas adicionais são incorporadas ao modelo. Outros critérios de seleção bastante difundidos em regressão linear são o AIC (Akaike, 1973, 1974), C_p de Mallows (Mallows, 1973), BIC (Akaike, 1978) ou SIC (Schwarz, 1978), HQ (Hannan e Quinn, 1979), entre outros. Subsequentemente, surgiram adaptações, ou correções, de alguns desses critérios para pequenas amostras. Uma boa referência sobre seleção de modelos em regressão linear é McQuarrie e Tsai (1998).

Em MLG existe uma larga aplicação de critérios de seleção derivados da função de log-verossimilhança maximizada, como os critérios AIC, BIC, entre outros. Além desses critérios baseados na função de log-verossimilhança, Hu e Shao (2008) propuseram uma classe de critérios consistentes baseados numa modificação da estatística R^2 . Um critério de seleção é consistente quando, assumindo que o modelo verdadeiro está entre os modelos candidatos de dimensão finita, ele identifica o modelo correto assintoticamente com probabilidade um (McQuarrie e Tsai, 1998).

Em modelos de regressão beta, todavia, não são encontrados estudos específicos no que tange a critérios de seleção. Contudo, muitos dos critérios de seleção originalmente desenvolvidos para regressão linear, modelos autorregressivos e MLG podem ser aplicados aos modelos de regressão beta. A seguir são apresentados os critérios de seleção considerados para avaliação de desempenho em amostras de tamanho finito quando aplicados em modelos de regressão beta com dispersão variável.

2.3.1 Critérios de seleção usuais

Primeiramente, dada uma medida global de variação explicada R_p^2 , podemos considerar o R_p^2 ajustado (\bar{R}_p^2) como um critério de seleção. Para isso, inclui-se um termo penalizador ao R_p^2 quando novos regressores são incluídos no modelo. No entanto, na literatura existem diferentes formulações para R_p^2 , que em muitos casos são chamadas de pseudo R^2 .

Ferrari e Cribari-Neto (2004) consideram uma medida global de ajuste para regressão beta definida como o quadrado do coeficiente de correlação amostral entre $g(y)$ e $\hat{\eta} = X\hat{\beta}$, em que $\hat{\beta}$ denota o estimador de máxima verossimilhança de β . Com isso, baseado no pseudo R^2 de Ferrari e Cribari-Neto (2004), consideramos o seguinte critério de seleção:

$$\bar{R}_{FC}^2 = 1 - (1 - R_{FC}^2) \frac{(n-1)}{(n-k)},$$

em que R_{FC}^2 é o quadrado do coeficiente de correlação amostral entre $\hat{\eta}$ e $g(y)$ e $k = r + s$ é o número de parâmetros estimados do modelo.

Seja L_{null} a função de verossimilhança maximizada do modelo sem regressores e L_{fit} a função de verossimilhança maximizada do modelo ajustado. Uma generalização do coeficiente de determinação dos modelos de regressão linear para modelos mais gerais é apresentada em Nagelkerke (1991) e Long (1997). Essa medida, conhecida como transformação da razão de verossimilhanças, é dada por

$$R_{RV}^2 = 1 - \left(\frac{L_{\text{null}}}{L_{\text{fit}}} \right)^{2/n}. \quad (2.6)$$

Baseado nesta medida de qualidade consideramos o critério de seleção dado por

$$\bar{R}_{RV}^2 = 1 - (1 - R_{RV}^2) \frac{(n-1)}{(n-k)}.$$

Hu e Shao (2008) propõem uma medida R^2 para MLG com base no coeficiente de determinação da regressão linear. Usando essa medida de qualidade de ajuste eles propuseram uma classe de critérios de seleção para MLG, que é dada por

$$\bar{R}_{HS}^2 = 1 - \frac{n-1}{n-\lambda_n k} \frac{\sum_{t=1}^n (y_t - \hat{\mu}_t)^2}{\sum_{t=1}^n (y_t - \bar{y})^2},$$

em que $\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t$ e $\hat{\mu}_t = g^{-1}(\hat{\eta}_t)$. Se $\lambda_n = 1$, então \bar{R}_{HS}^2 se reduz ao R^2 modificado dado em Mittlböck e Schemper (2002). Adicionalmente, se λ_n é tal que $\lambda_n = o(n)$ e $\lambda_n \rightarrow \infty$ quando $n \rightarrow \infty$, então o critério é consistente (Hu e Shao, 2008). Hu e Shao (2008) sugerem $\lambda_n = 1$, $\lambda_n = \log(n)$ e $\lambda_n = \sqrt{n}$.

Os critérios de seleção apresentados até o momento são baseados em medidas de qualidade de ajuste (R_p^2). Portanto, dentre os modelos candidatos, o melhor modelo é aquele que maximiza o critério escolhido. Contudo, outros critérios comumente utilizados são aqueles derivados do AIC, que é formulado a partir da função de log-verossimilhança maximizada. Da maneira em que estes critérios são apresentados, o melhor modelo será aquele que minimizar o critério de seleção escolhido.

O critério de informação de Akaike (Akaike, 1973, 1974) é dado por

$$\text{AIC} = -2\ell(\hat{\beta}, \hat{\gamma}) + 2k,$$

em que $\hat{\beta}$ e $\hat{\gamma}$ são as estimativas de máxima verossimilhança de β e γ , respectivamente.

Assumindo que o modelo verdadeiro possui dimensão infinita e que o conjunto de modelos candidatos não contém o modelo verdadeiro, Shibata (1980) define que um critério de seleção é assintoticamente eficiente se, em amostras grandes, o mesmo escolhe o modelo que minimiza o erro quadrático médio entre μ e $\hat{\mu} = g^{-1}(\hat{\eta})$. Neste sentido, o AIC é um critério assintoticamente eficiente. O AIC foi derivado com a intenção de determinar de forma aproximada um estimador assintoticamente não-viesado para a distância de Kullback-Leibler (Kullback e Leibler, 1951) entre o modelo verdadeiro e o modelo candidato estimado. No entanto, por ser uma aproximação assintótica, ele pode não apresentar bom desempenho em pequenas amostras. Com o objetivo de melhorar o comportamento do AIC em amostras de tamanho finito, Sugiura (1978) introduz um estimador exatamente não-viesado para a distância de Kullback-Leibler em regressão linear, o AICc. Hurvich e Tsai (1989) ampliam a aplicabilidade do AICc para regressão não-linear e modelos autorregressivos, mostrando que, apesar de não ser mais um estimador exatamente não-viesado, ele possui propriedades assintóticas equivalentes ao AIC e desempenho superior em pequenas amostras. O AICc é definido como

$$\text{AICc} = -2\ell(\hat{\beta}, \hat{\gamma}) + \frac{2nk}{n-k-1}.$$

Sob uma perspectiva bayesiana, Akaike (1978) e Schwarz (1978) introduziram um critério consistente para a seleção de modelos de regressão linear. O critério de informação de Schwarz

(SIC), também conhecido como critério de informação bayesiano (BIC), é dado por

$$\text{SIC} = -2\ell(\hat{\beta}, \hat{\gamma}) + k \log(n).$$

McQuarrie (1999), utilizando a relação entre o AIC e o AICc, deriva uma versão corrigida do SIC para pequenas amostras, o SICc. O SICc é consistente, sendo assintoticamente equivalente ao SIC. Esse critério é dado por

$$\text{SICc} = -2\ell(\hat{\beta}, \hat{\gamma}) + \frac{nk \log(n)}{n - k - 1}.$$

Outro critério consistente é o HQ, proposto por Hannan e Quinn (1979) para modelos autorregressivos de séries temporais. O critério HQ é dado por

$$\text{HQ} = -2\ell(\hat{\beta}, \hat{\gamma}) + 2k \log(\log(n)).$$

Assim como o AIC e o SIC, o critério HQ possui sua versão corrigida para pequenas amostras, o HQc (McQuarrie e Tsai, 1998). A versão corrigida do critério HQ é dada por

$$\text{HQc} = -2\ell(\hat{\beta}, \hat{\gamma}) + \frac{2nk \log(\log(n))}{n - k - 1}.$$

2.3.2 Critérios de seleção para modelos com dispersão variável

Os critérios de seleção usuais não possuem características específicas que consideram a estrutura de regressão para a dispersão. Em alguns deles é utilizada somente a proximidade entre y e $\hat{\mu}$ para estabelecer uma medida de qualidade que é penalizada pela inclusão de covariáveis no modelo, desconsiderando a modelagem de σ . Neste sentido, propomos dois critérios de seleção que considera a modelagem da dispersão.

Entendendo que as inclusões de covariáveis nos submodelos da dispersão ou da média podem interferir diferentemente na qualidade do ajuste, propomos, baseados na medida R_{RV}^2 , o critério de seleção da transformação da razão de verossimilhança ponderado, dado por

$$\bar{R}_{RVP}^2 = 1 - (1 - R_{RV}^2) \left(\frac{n - 1}{n - (1 + \alpha)r - (1 - \alpha)s} \right)^\delta,$$

em que $0 \leq \alpha \leq 1$ e $\delta > 0$. Percebemos que um caso particular do critério proposto, para $\alpha = 0$ e $\delta = 1$, é o \bar{R}_{RV}^2 dado em (2.6). Na avaliação numérica, Seção 2.4, apresentamos sugestões de uso para os valores de α e δ .

Sabendo que $\text{var}(y_t) = \sigma_t^2 \mu_t (1 - \mu_t)$ e que $\text{var}(y_t)$ pode ser aproximado por $(y_t - \hat{\mu}_t)^2$, definimos $\sigma_t^* = \sqrt{\frac{(y_t - \hat{\mu}_t)^2}{\hat{\mu}_t(1 - \hat{\mu}_t)}}$. Dessa forma, baseados no \bar{R}_{HS}^2 de Hu e Shao (2008), propomos o seguinte critério:

$$\bar{R}_D^2 = \alpha \left[1 - \frac{n - 1}{n - \lambda_n r} \frac{\sum_{t=1}^n (y_t - \hat{\mu}_t)^2}{\sum_{t=1}^n (y_t - \bar{y})^2} \right] + (1 - \alpha) \left[1 - \frac{n - 1}{n - \delta_n s} \frac{\sum_{t=1}^n (\sigma_t^* - \hat{\sigma}_t)^2}{\sum_{t=1}^n (\sigma_t^* - \bar{\sigma}^*)^2} \right],$$

em que $\bar{\sigma}^* = (1/n) \sum_{t=1}^n \sigma_t^*$, $0 \leq \alpha \leq 1$ e δ_n , assim como λ_n , é uma função de n , como, por exemplo, $\delta_n = 1$, $\delta_n = \log(n)$ e $\delta_n = \sqrt{n}$. O \bar{R}_D^2 é uma combinação convexa de duas medidas de qualidade dos ajustes da média e da dispersão penalizadas pelas quantidades de regressores considerados em cada submodelo.

2.4 Avaliação numérica

Como visto, alguns critérios de seleção possuem propriedades assintóticas como eficiência e consistência. No entanto, na prática os critérios podem apresentar comportamentos notadamente distintos, pois trabalha-se com amostras de tamanho finito. É importante, assim, avaliar numericamente os desempenhos dos diferentes critérios de seleção de modelos em amostras de tamanho finito.

O nosso objetivo aqui é investigar os desempenhos em amostras de tamanho finito dos critérios de seleção apresentados na Seção 2.3 quando utilizados em modelos de regressão beta com dispersão variável. O programa utilizado na implementação computacional foi escrito em R versão 2.9 (R Development Core Team, 2009), tendo sido feita a estimação dos parâmetros com o pacote GAMLSS (Stasinopoulos e Rigby, 2007).

O modelo de regressão beta considerado no estudo de simulação é dado por

$$g(\mu_t) = \beta_1 + x_{t2}\beta_2 + x_{t3}\beta_3 + x_{t4}\beta_4 + x_{t5}\beta_5, \quad (2.7)$$

$$h(\sigma_t) = \gamma_1 + z_{t2}\gamma_2 + z_{t3}\gamma_3 + z_{t4}\gamma_4 + z_{t5}\gamma_5, \quad (2.8)$$

$t = 1, \dots, n$, em que (2.7) é o submodelo da média, (2.8) é o submodelo da dispersão e $x_{ti} = z_{ti}$, para $i = 2, \dots, 5$ e $\forall t$. Nós geramos 5000 amostras deste modelo para diferentes valores do vetor de parâmetros $\theta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5)$ e para quatro diferentes tamanhos amostrais, a saber: $n = 25, 50, 100, 200$. Os valores para o vetor de parâmetros são apresentados na Tabela 2.1.

Tabela 2.1 Valores dos parâmetros considerados nas simulações.

Modelos	β_1	β_2	β_3	β_4	β_5	γ_1	γ_2	γ_3	γ_4	γ_5
Modelo 1	1.5	-1	-1	0	0	-1	-1	-1	0	0
Modelo 2	-1.5	1	1	0	0	-1	-1.25	-1/2	-1/4	0
Modelo 3	1	-3/4	-1/4	0	0	-1	-1	-1	0	0
Modelo 4	-1	3/4	1/4	0	0	-1	-1.25	-1/2	-1/4	0

No Modelo 1 temos o caso em que o modelo verdadeiro é facilmente identificável, uma vez que todos os parâmetros, a menos dos interceptos, possuem os mesmos valores. O Modelo 2 possui o submodelo da média facilmente identificável, enquanto que o submodelo para a dispersão é fracamente identificável. A fraca identificabilidade é caracterizada pela proximidade a zero dos valores absolutos dos parâmetros γ_i à medida que i cresce. No cenário de fraca identificabilidade as covariáveis possuem diferentes influências sobre o preditor linear. No Modelo 3 temos o submodelo da média fracamente identificável, ao passo que o submodelo da dispersão é facilmente identificável. O quarto modelo apresenta o caso em que o modelo verdadeiro é fracamente identificável. O termo “modelos facilmente identificáveis” está sendo usado no mesmo sentido do termo *easily identified models* apresentado em McQuarrie e Tsai (1998). A magnitude dos parâmetros do modelo influenciam na identificação do modelo verdadeiro, pois todas as covariáveis são obtidas como realizações da distribuição uniforme padrão $\mathcal{U}(0,1)$, as quais permanecem constantes durante todo o experimento.

A cada iteração da simulação são geradas ocorrências da variável resposta com distribuição beta de parâmetros μ_t e σ_t dados por (2.7) e (2.8). Para as funções de ligação consideramos a função logit para $g(\cdot)$ e $h(\cdot)$, ou seja,

$$\mu_t = \frac{\exp(\beta_1 + \sum_{i=2}^5 x_{ti}\beta_i)}{1 + \exp(\beta_1 + \sum_{i=2}^5 x_{ti}\beta_i)}, \quad \sigma_t = \frac{\exp(\gamma_1 + \sum_{i=2}^5 z_{ti}\gamma_i)}{1 + \exp(\gamma_1 + \sum_{i=2}^5 z_{ti}\gamma_i)}.$$

Para avaliação de desempenho dos critérios de seleção consideramos o procedimento descrito a seguir. O conjunto de modelos candidatos é composto por todos os modelos gerados pelas combinações de todas as variáveis regressoras, considerando sempre o modelo com intercepto. Ou seja, como são considerados quatro regressores, para o submodelo da média temos um total de $(2^4 + 1) = 17$ modelos candidatos, assim como para o submodelo da dispersão. Para seleção conjunta dos submodelos da média e da dispersão tem-se um total de $17 \times 17 = 289$ modelos candidatos. Dessa forma, uma vez que o modelo verdadeiro pertence ao conjunto de modelos candidatos, a avaliação dos critérios de seleção é feita pela proporção de escolha do verdadeiro modelo dentre as 5000 réplicas de Monte Carlo.

A avaliação se deu em cinco abordagens diferenciadas, como descritas a seguir:

1. utilizamos os critérios de seleção para escolher os regressores de μ_t e σ_t conjuntamente;
2. consideramos o submodelo da média corretamente especificado e utilizamos os critérios para selecionar os regressores de σ_t ;
3. o submodelo para σ_t foi corretamente especificado e utilizamos os critérios para selecionar o submodelo da média;
4. supusemos que o parâmetro de dispersão era fixo em todos os modelos e utilizamos os critérios de seleção para escolher somente os regressores para a média;
5. baseados nos resultados das quatro primeiras abordagens, propomos um esquema de seleção de modelos em duas etapas.

Para o critério \bar{R}_{HS}^2 consideramos $\lambda_n = 1$, $\lambda_n = \log(n)$ e $\lambda_n = \sqrt{n}$, mas são apresentados apenas os resultados em que $\lambda_n = \log(n)$ por ter conduzido aos melhores resultados. Para os critérios propostos na Seção 2.3.2 apresentamos os resultados das seguintes variantes, com suas respectivas nomenclaturas:

\bar{R}_{D1}^2 : usa $\alpha = 0.4$, $\lambda_n = \log(n)$ e $\delta_n = \log(n)$;

\bar{R}_{D2}^2 : usa $\alpha = 0.6$, $\lambda_n = \log(n)$ e $\delta_n = \log(n)$;

\bar{R}_{D3}^2 : usa $\alpha = 0.6$, $\lambda_n = \log(n)$ e $\delta_n = 1$;

\bar{R}_{D4}^2 : usa $\alpha = 0.5$, $\lambda_n = \log(n)$ e $\delta_n = 1$;

\bar{R}_{RVp1}^2 : usa $\alpha = 0$ e $\delta = 3$;

\bar{R}_{RVp2}^2 : usa $\alpha = 0$ e $\delta = 2$;

\bar{R}_{RVp3}^2 : usa $\alpha = 0$ e $\delta = 1.5$;

\bar{R}_{RVp4}^2 : usa $\alpha = 0.4$ e $\delta = 1$;

\bar{R}_{RVp5}^2 : usa $\alpha = 0.4$ e $\delta = 2$.

A determinação de α , δ , λ_n e δ_n para as variações dos critérios \bar{R}_D^2 e \bar{R}_{RVp}^2 foi baseada em resultados numéricos de simulações prévias. Para o critério \bar{R}_D^2 , por exemplo, usando valores de α maiores que 0.5 é dado maior peso ao ajuste do submodelo da média. Já para o critério \bar{R}_{RVp}^2 , valores de δ maiores do que um penalizam mais fortemente a inclusão de covariáveis no modelo, enquanto que a inclusão de covariáveis especificamente no submodelo da média é penalizada mais fortemente quando α é maior do que zero.

As porcentagens de escolha do verdadeiro modelo para cada critério de seleção, nas cinco abordagens mencionadas anteriormente, estão apresentadas nas Tabelas 2.2, 2.3, 2.4, 2.5 e 2.6, respectivamente. Os melhores resultados estão destacados em negrito.

Tabela 2.2 Porcentagem (%) de seleção correta dos submodelos da média e dispersão conjuntamente.

n	Modelo 1				Modelo 2				Modelo 3				Modelo 4			
	25	50	100	200	25	50	100	200	25	50	100	200	25	50	100	200
AIC	4.9	24.0	42.9	49.3	1.2	4.4	11.3	24.2	2.3	14.5	39.6	49.1	0.7	2.9	10.0	23.6
AICc	3.9	27.2	49.0	53.0	0.6	3.2	10.8	23.9	1.4	15.4	43.8	52.5	0.3	2.0	9.5	23.6
SIC	3.8	24.6	64.4	89.5	0.7	1.4	3.7	10.4	1.6	10.2	47.8	88.0	0.4	0.5	2.8	8.9
SICc	1.4	20.0	63.5	91.3	0.2	0.6	2.2	8.8	0.3	6.2	44.4	89.4	0.1	0.1	1.8	7.5
HQ	4.7	26.3	58.7	73.5	1.1	3.0	8.0	19.8	2.2	14.5	49.4	72.4	0.6	1.9	6.9	18.6
HQc	3.2	26.5	62.6	76.7	0.5	1.5	6.3	18.5	0.9	12.4	51.6	75.5	0.2	0.8	5.3	17.3
\bar{R}_{FC}^2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
\bar{R}_{RV}^2	4.2	14.7	19.8	22.0	1.5	5.4	11.5	17.7	2.2	10.5	19.8	21.8	0.9	4.2	10.9	18.8
\bar{R}_{HS}^2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
\bar{R}_{D1}^2	4.4	15.4	36.5	58.7	0.3	0.7	2.2	6.8	1.9	7.4	29.7	60.8	0.2	0.5	1.9	7.7
\bar{R}_{D2}^2	6.0	17.9	46.4	73.9	0.3	0.8	2.7	8.7	1.5	6.3	34.0	74.0	0.1	0.4	2.0	8.7
\bar{R}_{D3}^2	14.7	23.6	35.1	40.0	4.9	9.3	16.7	27.3	3.7	8.0	26.3	40.0	1.1	3.7	12.2	28.4
\bar{R}_{D4}^2	12.8	22.1	32.3	37.6	4.6	8.8	15.2	25.4	4.0	8.0	24.7	37.6	1.3	4.0	11.7	27.2
\bar{R}_{RVp1}^2	3.3	26.5	60.3	70.5	0.6	1.9	7.6	21.0	1.0	12.8	50.4	69.1	0.2	0.9	6.3	20.0
\bar{R}_{RVp2}^2	4.9	25.9	45.5	51.0	0.9	3.8	11.1	24.0	2.2	15.2	41.7	50.6	0.5	2.6	9.8	23.7
\bar{R}_{RVp3}^2	5.1	21.9	33.7	36.9	1.2	5.1	11.7	22.6	2.6	14.0	31.8	35.9	0.7	3.5	11.1	23.1
\bar{R}_{RVp4}^2	5.0	13.5	17.3	18.5	3.2	9.1	15.5	22.3	2.4	9.0	16.5	18.1	1.7	6.5	14.7	22.6
\bar{R}_{RVp5}^2	8.0	28.4	40.5	43.3	3.2	8.6	19.0	32.5	2.5	13.8	35.4	42.5	1.2	5.2	16.1	32.8

Os resultados da seleção conjunta dos regressores da média e dispersão, apresentados na Tabela 2.2, mostram que os critérios de seleção considerados apresentam desempenho pobre nos casos em que o submodelo da dispersão é fracamente identificável e em pequenas amostras. Este fato pode ser observado pela baixa proporção de seleção do verdadeiro modelo

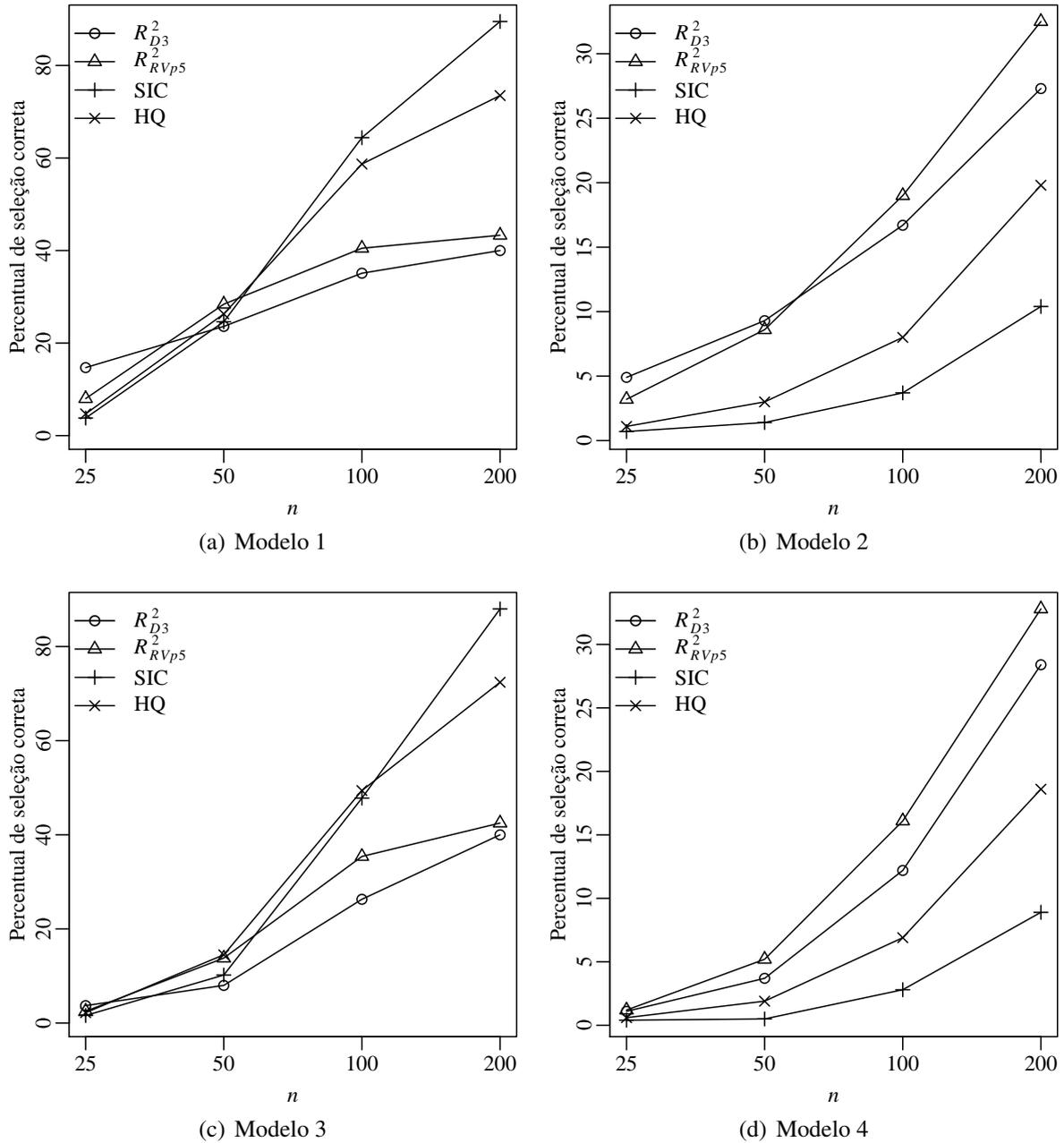


Figura 2.3 Porcentagem (%) de seleção correta dos submodelos da média e dispersão conjuntamente para os quatro critérios com melhores desempenhos.

para $n = 25$, principalmente, nos Modelos 2 e 4. O critério \bar{R}^2_{D3} apresenta bom desempenho nas amostras menores, independentemente do modelo considerado. Nos Modelos 1 e 3, com submodelo da média facilmente identificável, o critério SIC se destaca. Outro critério que se destaca em alguns cenários é o \bar{R}^2_{RVp5} . Os critérios \bar{R}^2_{D3} e \bar{R}^2_{RVp5} apresentam bons desempenhos nos cenários de fraca identificabilidade do submodelo da média e em amostras pequenas, contudo, possuem desempenho fraco nos outros cenários. Já o SIC apresenta bons resultados

Tabela 2.3 Porcentagem (%) de seleção correta dos regressores para a dispersão com submodelo da média corretamente especificado.

n	Modelo 1				Modelo 2				Modelo 3				Modelo 4			
	25	50	100	200	25	50	100	200	25	50	100	200	25	50	100	200
AIC	10.9	38.7	64.1	70.5	4.5	7.4	17.2	34.8	11.0	38.3	64.4	70.6	4.4	7.4	16.9	34.8
AICc	6.2	37.5	68.0	73.4	1.2	4.4	15.0	33.3	6.9	38.3	67.1	72.8	1.4	4.5	14.6	33.5
SIC	6.9	29.4	70.5	94.5	1.7	1.9	4.1	11.0	7.6	28.8	69.0	94.0	2.0	1.8	3.8	9.5
SICc	1.9	22.0	67.4	95.2	0.2	0.7	2.4	9.2	2.0	20.9	66.4	94.6	0.3	0.5	2.5	7.8
HQ	9.8	36.2	72.4	86.1	3.5	4.2	10.1	23.3	10.0	36.7	71.1	85.6	3.6	4.2	9.6	22.3
HQc	4.6	32.8	73.6	87.7	0.8	2.0	7.5	21.3	5.0	32.2	72.2	87.5	0.8	2.1	7.2	20.3
\bar{R}_{FC}^2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
\bar{R}_{RV}^2	13.1	33.7	45.1	46.9	6.9	13.2	25.1	40.4	13.4	33.6	45.0	47.4	7.0	13.8	25.4	40.9
\bar{R}_{HS}^2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
\bar{R}_{D1}^2	7.0	20.4	51.4	81.3	0.4	1.0	3.3	9.6	7.2	20.1	52.0	81.5	0.6	1.3	3.1	9.5
\bar{R}_{D2}^2	7.0	19.8	51.3	81.3	0.3	0.9	3.1	9.5	7.4	19.4	51.4	81.3	0.5	1.1	3.0	9.4
\bar{R}_{D3}^2	17.4	26.2	38.5	43.9	6.0	10.2	18.6	30.1	17.3	26.1	39.0	43.7	5.1	10.7	18.8	30.6
\bar{R}_{D4}^2	16.9	26.2	38.3	43.8	6.3	10.5	18.6	30.2	17.0	26.0	38.9	43.7	5.8	10.9	18.8	30.8
\bar{R}_{RVp1}^2	5.1	33.3	73.0	84.4	1.0	2.4	9.3	25.4	5.7	33.2	71.7	84.0	1.1	2.5	8.7	24.4
\bar{R}_{RVp2}^2	9.4	38.6	65.7	71.8	2.7	5.8	16.3	34.0	9.5	38.6	65.7	71.7	2.8	6.0	15.8	34.2
\bar{R}_{RVp3}^2	11.4	38.8	57.4	61.2	4.5	9.0	20.4	37.9	11.6	38.3	57.4	60.8	4.4	9.4	20.4	38.4
\bar{R}_{RVp4}^2	11.6	25.6	31.0	32.1	10.2	17.1	27.0	38.7	12.2	24.9	30.6	32.1	10.2	17.5	27.5	39.0
\bar{R}_{RVp5}^2	12.4	35.9	50.4	53.3	6.0	11.2	23.2	39.6	12.5	36.1	50.7	53.0	5.9	11.7	23.1	40.0

em grandes amostras e quando o submodelo da média é facilmente identificável, apresentando desempenho fraco em amostras pequenas. Com isso, o critério HQ se destaca, pois ele se coloca com bons resultados em quase todos os cenários e tamanhos amostrais, apesar de não apresenta o melhor desempenho em nenhum deles.

A Figura 2.3 apresenta, de forma gráfica, os desempenhos dos critérios \bar{R}_{D3}^2 , \bar{R}_{RVp5}^2 , SIC e HQ. Pelo gráfico fica clara a superioridade de desempenho dos critérios \bar{R}_{D3}^2 e \bar{R}_{RVp5}^2 nos modelos em que o submodelo da dispersão é fracamente identificável, Figuras 2.3(b) e 2.3(d), e nos casos em que $n = 25$ para todos os cenários. O desempenho do HQ acompanha o bom desempenho do SIC nos Modelos 1 e 3 para amostras de tamanho 100 e 200 e, adicionalmente, também possui desempenhos aceitáveis para os Modelos 2 e 4. Com base nessas análises, para seleção conjunta dos submodelos da média e dispersão, sugerimos o uso do \bar{R}_{D3}^2 ou \bar{R}_{RVp5}^2 quando $n \leq 50$ e o uso do HQ para as amostras maiores.

Os resultados da aplicação dos critérios de seleção para a escolha conjunta dos submodelos da média e da dispersão indicam que os desempenhos dos critérios de seleção estão fortemente ligados à identificabilidade dos submodelos verdadeiros. A baixa proporção de escolha do modelo correto nessa abordagem pode indicar que os melhores critérios para a seleção dos submodelos da média não são, necessariamente, os critérios com melhor desempenho para

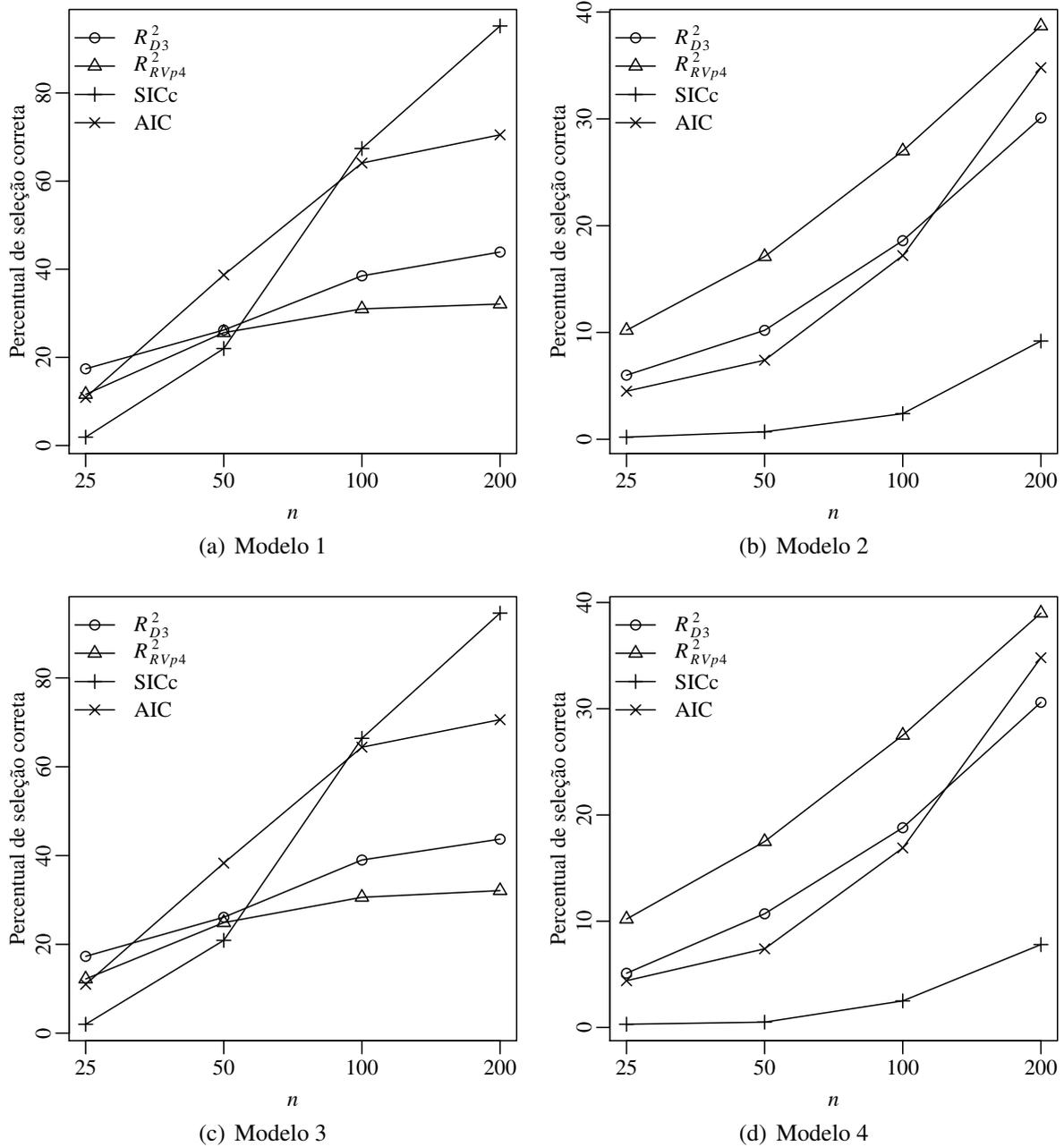


Figura 2.4 Porcentagem (%) de seleção correta dos regressores para a dispersão com submodelo da média corretamente especificado para os quatro critérios com melhores desempenhos.

a seleção dos submodelos da dispersão. Isto pode ser melhor notado a partir dos resultados apresentados nas Tabelas 2.3 e 2.4.

A Tabela 2.3 apresenta os resultados da segunda abordagem, em que o submodelo para μ_t está corretamente especificado e os critérios de seleção são utilizados para escolher os regressores de σ_t . Nesta abordagem, o critério \bar{R}_{RVp4}^2 foi o que conduziu aos melhores resultados nos Modelos 2 e 4, em que o submodelo de σ_t é fracamente identificável. Para os Modelos 1 e 3,

Tabela 2.4 Porcentagem (%) de seleção correta dos regressores para a média com submodelo da dispersão corretamente especificado.

n	Modelo 1				Modelo 2				Modelo 3				Modelo 4			
	25	50	100	200	25	50	100	200	25	50	100	200	25	50	100	200
AIC	51.1	62.6	67.1	69.5	43.9	61.6	67.4	68.9	25.1	39.8	61.8	69.8	21.1	40.6	62.2	68.7
AICc	73.6	73.2	72.4	71.8	68.0	74.3	73.7	71.6	23.9	41.8	65.3	72.2	20.6	43.1	66.2	71.5
SIC	67.4	85.1	91.6	95.0	58.3	83.9	91.7	94.5	23.6	36.7	70.0	93.6	21.1	37.3	72.6	94.0
SICc	83.4	92.2	94.1	96.1	75.7	92.3	94.6	95.9	14.0	30.0	67.9	94.4	12.5	30.3	70.8	94.8
HQ	56.3	73.4	81.4	85.3	48.5	72.7	81.6	85.3	25.2	40.7	69.7	85.0	21.2	42.1	71.0	85.3
HQc	78.6	83.1	85.7	87.4	72.2	83.7	85.9	87.4	21.1	39.2	71.8	86.7	18.5	39.1	72.7	86.9
\bar{R}_{FC}^2	68.5	57.6	62.4	63.0	65.9	57.9	60.3	61.0	33.0	38.5	54.4	60.0	31.1	36.6	52.9	61.4
\bar{R}_{RV}^2	38.9	44.0	45.1	47.0	34.7	42.8	45.4	44.7	22.3	33.1	42.6	46.4	20.4	32.9	44.0	45.9
\bar{R}_{HS}^2	0.7	2.1	0.0	0.0	0.0	4.0	0.0	0.0	0.2	15.2	28.7	16.5	0.0	9.0	31.5	22.2
\bar{R}_{D1}^2	33.2	67.1	69.6	72.1	20.1	55.8	65.5	70.6	19.9	32.1	57.1	74.4	15.2	29.3	57.7	79.3
\bar{R}_{D2}^2	66.7	89.0	89.6	90.9	47.6	84.7	86.5	89.3	23.4	32.6	67.0	91.2	18.5	29.8	65.1	92.1
\bar{R}_{D3}^2	77.7	91.1	90.9	91.7	66.7	88.7	89.0	90.5	23.5	32.3	67.5	91.7	19.0	28.8	65.5	92.6
\bar{R}_{D4}^2	64.9	86.6	85.4	85.4	52.6	82.2	82.1	83.8	23.4	32.7	64.5	86.2	18.3	30.2	63.9	88.6
\bar{R}_{RVP1}^2	74.5	81.2	82.8	83.5	66.9	81.3	83.1	83.3	22.0	39.5	70.5	82.7	19.5	40.1	71.9	83.3
\bar{R}_{RVP2}^2	61.6	67.4	69.5	70.5	54.9	67.2	70.1	70.2	25.9	40.7	63.4	70.7	22.2	42.0	64.1	70.0
\bar{R}_{RVP3}^2	51.6	56.9	59.7	60.7	46.0	57.7	60.0	59.0	25.8	38.6	55.3	59.3	21.7	39.5	55.8	59.9
\bar{R}_{RVP4}^2	49.7	55.0	57.2	58.3	43.6	54.9	57.4	56.6	25.7	38.1	53.6	57.2	22.2	38.7	53.6	57.5
\bar{R}_{RVP5}^2	72.7	79.2	80.9	81.3	64.6	78.9	81.2	81.7	23.3	40.4	69.6	80.9	21.0	41.2	70.9	81.4

em que o submodelo para σ_t é facilmente identificável, o critério \bar{R}_{D3}^2 foi o que apresentou o melhor desempenho com $n = 25$; o AIC, o \bar{R}_{RVP2}^2 e o \bar{R}_{RVP3}^2 forneceram os melhores resultados quando $n = 50$; para $n = 100$, o melhor critério foi o HQc; para o tamanho amostral $n = 200$, o SICc conduziu aos melhores resultados. De maneira geral, com base nos resultados da Tabela 2.3, podemos destacar o AIC, pois este critério se colocou sempre entre aqueles com bom desempenho e não apresentou desempenho pobre em nenhum cenário. A Figura 2.4 apresenta os gráficos comparativos das porcentagens de seleção do modelo correto dos quatro critérios que se destacaram, nomeadamente, \bar{R}_{D3}^2 , \bar{R}_{RVP4}^2 , SICc e AIC.

Para a terceira abordagem, que seleciona o submodelo da média dado que o submodelo da dispersão está corretamente especificado, os resultados estão dispostos na Tabela 2.4. Percebemos que o SICc é o critério de seleção com melhor desempenho quando o submodelo da média é facilmente identificável (Modelos 1 e 2) independentemente do tamanho amostral. Já para os casos em que o submodelo verdadeiro de μ_t é fracamente identificável (Modelos 3 e 4), temos que o critério \bar{R}_{FC}^2 conduziu aos melhores resultados quando $n = 25$; para $n = 50$, o AICc apresentou o melhor desempenho; para $n = 100$, o melhor critério foi o HQc; em amostras maiores, como $n = 200$, o SICc mostrou-se o melhor critério de seleção. Contudo, nesta abordagem podemos destacar, além do SICc, o critério \bar{R}_{D3}^2 pela regularidade e bom desem-

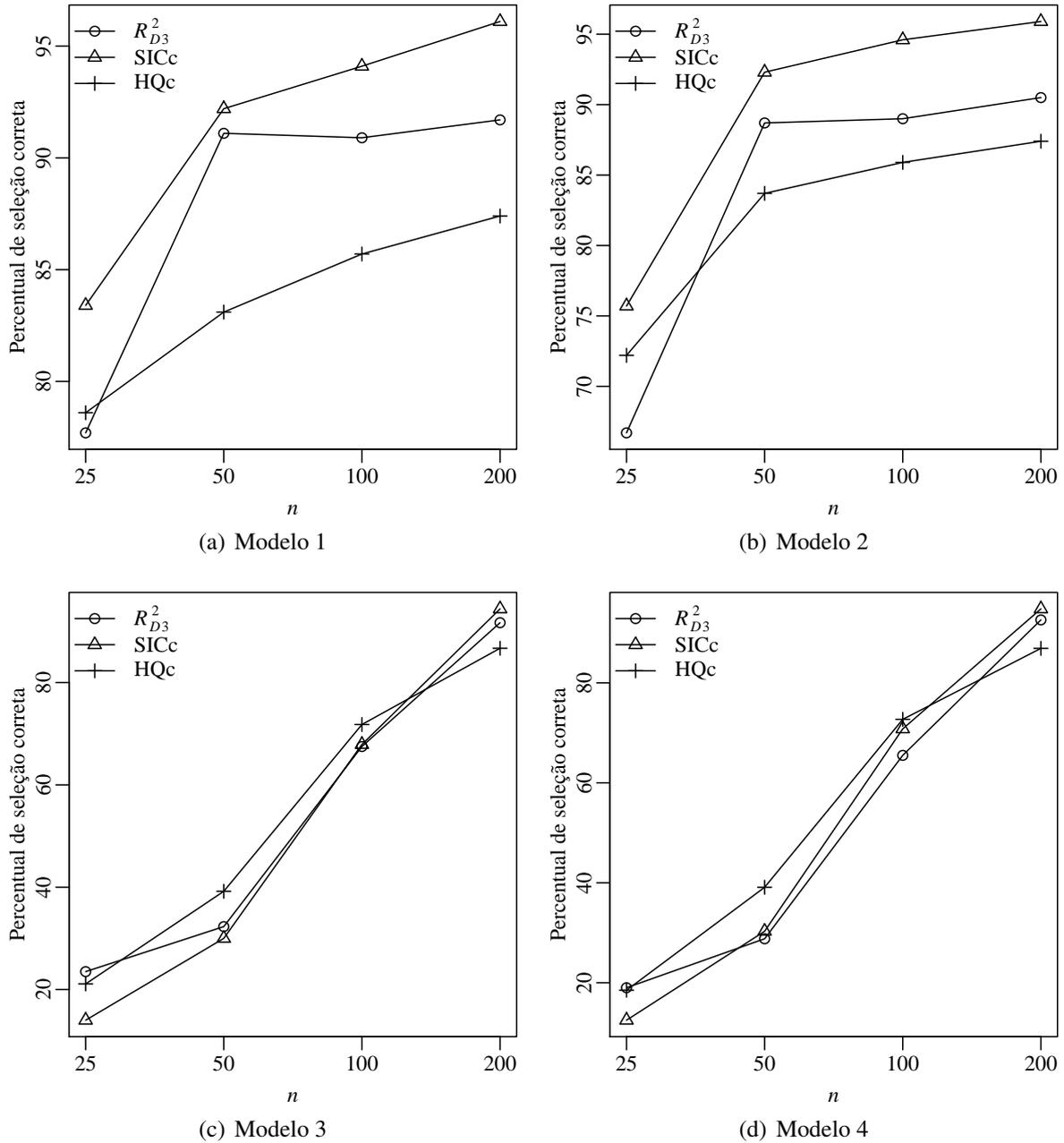


Figura 2.5 Porcentagem (%) de seleção correta dos regressores para a média com submodelo da dispersão corretamente especificado para os três critérios com melhores desempenhos.

penho apresentado em todos os cenários considerados. Com isso, nessa terceira abordagem, destacamos os critérios \bar{R}^2_{D3} , SICc e HQc e apresentamos seus desempenhos na seleção do submodelo verdadeiro da média nos gráficos da Figura 2.5.

Na quarta abordagem, o interesse reside em avaliar os desempenhos dos critérios de seleção na escolha do submodelo da média ao assumirmos que o parâmetro de dispersão é constante ao longo das observações. Os resultados desta abordagem encontram-se na Tabela 2.5 e na

Tabela 2.5 Porcentagem (%) de seleção correta dos regressores para a média supondo dispersão constante.

n	Modelo 1				Modelo 2				Modelo 3				Modelo 4			
	25	50	100	200	25	50	100	200	25	50	100	200	25	50	100	200
AIC	70.8	68.4	73.2	74.0	66.5	66.0	70.4	72.8	32.3	40.5	63.8	71.1	30.8	37.9	61.6	73.4
AICc	83.2	76.1	76.9	75.7	80.5	74.3	73.8	74.6	31.4	40.3	66.6	72.7	29.4	38.2	63.7	74.9
SIC	83.2	89.4	94.6	96.7	80.9	87.3	93.6	96.9	28.3	31.8	64.4	91.7	26.6	29.4	63.8	91.1
SICc	89.2	93.1	95.8	97.2	88.0	92.2	95.2	97.3	20.1	28.2	63.2	91.7	19.3	25.3	62.2	91.0
HQ	75.0	79.4	85.9	89.1	71.4	77.6	84.8	88.4	32.0	39.2	68.9	85.8	30.8	36.2	67.4	86.9
HQc	86.1	85.9	88.7	90.1	83.8	83.5	87.5	89.5	28.5	37.2	69.2	87.2	26.6	34.1	68.1	87.9
\bar{R}_{FC}^2	54.4	45.7	50.3	49.2	49.8	44.3	48.7	49.0	30.5	32.8	45.3	48.2	29.1	31.6	43.7	50.0
\bar{R}_{RV}^2	54.3	46.8	50.1	50.7	49.7	44.7	48.5	50.0	30.6	33.2	46.0	48.4	29.5	32.6	44.6	50.7
\bar{R}_{HS}^2	8.3	0.6	0.0	0.0	6.1	0.2	0.0	0.0	11.6	15.4	27.6	20.4	11.2	15.1	29.5	20.6
\bar{R}_{D1}^2	85.4	92.3	91.4	90.9	84.1	90.4	89.3	91.1	20.2	27.5	61.0	87.9	19.1	24.2	57.8	86.3
\bar{R}_{D2}^2	89.0	94.6	95.3	96.2	87.5	93.4	94.5	96.2	18.4	26.0	59.8	90.1	18.5	23.7	59.2	89.4
\bar{R}_{D3}^2	89.2	94.6	95.4	96.2	87.6	93.5	94.5	96.2	18.4	26.0	59.7	90.2	18.4	23.5	59.1	89.5
\bar{R}_{D4}^2	88.3	94.0	94.0	94.4	86.8	92.7	92.9	94.8	18.7	26.7	60.6	89.5	18.5	23.6	58.8	88.7
\bar{R}_{RVp1}^2	85.8	85.3	86.7	87.1	83.3	82.9	85.4	86.2	27.0	37.2	69.0	84.0	25.6	34.0	67.7	85.0
\bar{R}_{RVp2}^2	76.6	71.9	74.7	74.9	72.9	70.2	72.0	73.6	32.8	40.6	65.2	71.8	31.2	38.2	62.6	74.4
\bar{R}_{RVp3}^2	68.1	61.4	64.6	64.7	63.7	59.4	62.6	63.6	33.3	39.5	57.9	62.0	31.5	37.1	55.8	64.2
\bar{R}_{RVp4}^2	67.5	59.8	62.6	62.7	63.1	57.7	60.3	61.5	33.6	39.0	56.3	60.1	31.8	36.9	54.3	61.7
\bar{R}_{RVp5}^2	85.5	84.1	84.9	85.6	83.3	81.6	84.0	84.3	28.0	38.3	69.2	82.2	26.4	35.2	67.6	83.5

Figura 2.6.

Analisando os resultados contidos na Tabela 2.5, percebe-se que os desempenhos dos critérios de seleção aplicados na escolha do submodelo da média dependem fracamente da modelagem correta da dispersão. Essa conclusão é evidenciada pela similaridade dos resultados das Tabelas 2.4 e 2.5. Percebemos que a seleção do submodelo da média, supondo σ_t constante, em muitos casos foi mais precisa do que ao assumirmos o submodelo da dispersão correto, principalmente em amostras menores ($n = 25, 50$) e nos modelos facilmente identificáveis, Modelos 1 e 2. Um exemplo pode ser visto na comparação dos resultados do Modelo 2, com $n = 25$, nas Tabelas 2.4 e 2.5. Para este caso, assumindo σ_t constante, o SICc apresentou frequência de escolha do verdadeiro modelo aproximadamente 15% maior do que assumindo o submodelo da dispersão corretamente especificado, passando de 75.7% para 88.0%.

Na seleção do submodelo da média supondo dispersão constante podemos destacar o desempenho dos critérios SICc, \bar{R}_{D3}^2 e \bar{R}_{RVp2}^2 . Ainda, por apresentar bom desempenho nos diferentes cenários verificados, também destacamos o HQc. A comparação gráfica dos desempenhos desses critérios está apresentada na Figura 2.6.

Pela análise dos resultados apresentados, percebemos que os melhores critérios para a seleção do submodelo da média são, muitas vezes, diferentes dos critérios de seleção com melhor

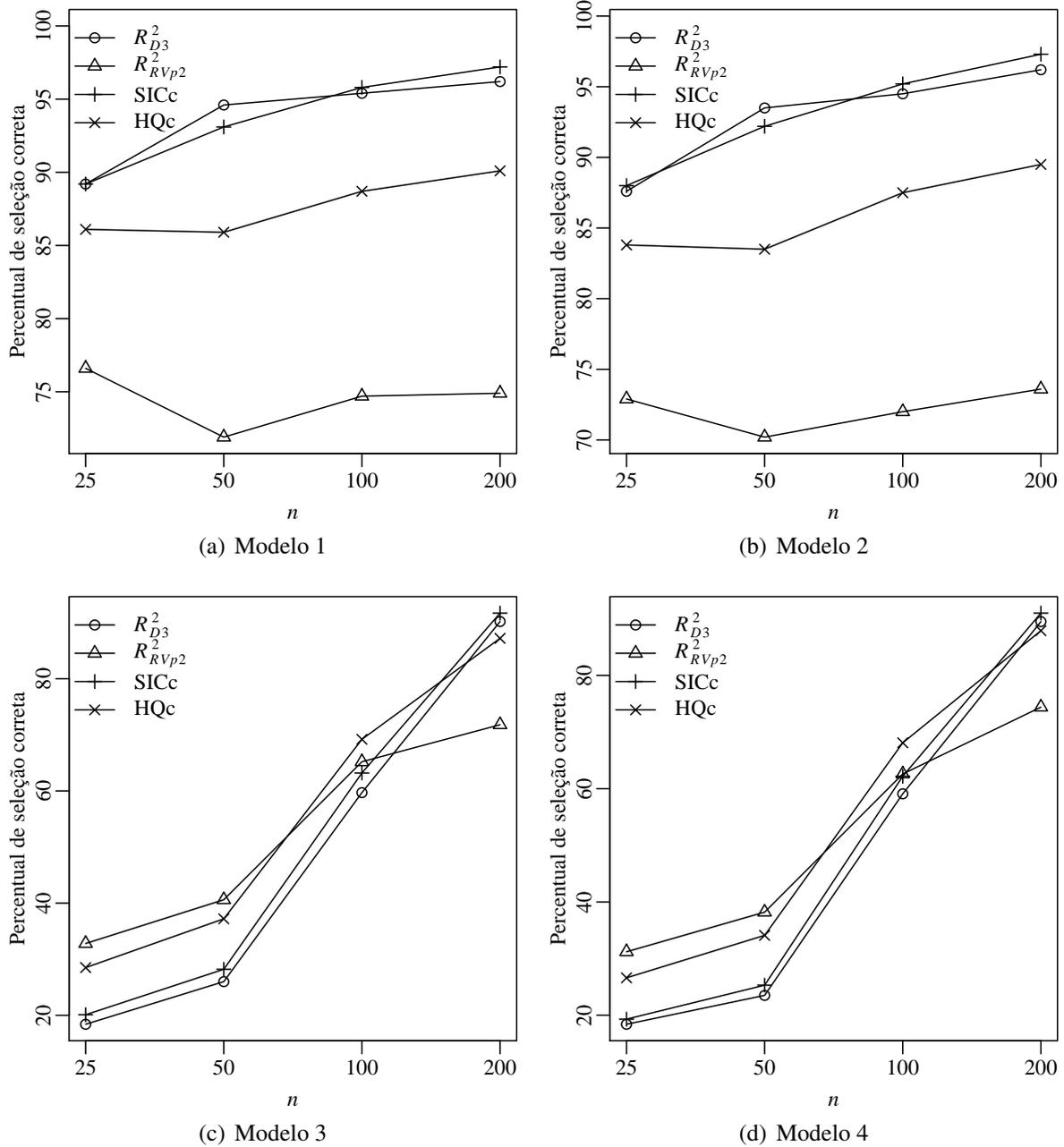


Figura 2.6 Porcentagem (%) de seleção correta dos regressores para a média supondo dispersão constante para os três critérios com melhores desempenhos.

desempenho na escolha do submodelo da dispersão. Este fato pode explicar o desempenho pobre dos critérios de seleção aplicados na escolha conjunta dos regressores da média e da dispersão; ver Tabela 2.2. Adicionalmente, o bom desempenho dos critérios de seleção na escolha dos submodelos da média ao assumirmos dispersão constante indica que o procedimento mais promissor para escolha dos regressores para média e para a dispersão é fazer esta seleção por etapas. Nesse sentido, propomos o seguinte esquema de seleção de modelos em duas etapas:

- (1) assumimos que a dispersão é constante e selecionamos o submodelo da média;
- (2) assumindo que o modelo selecionado na etapa (1) é o correto, utilizamos os critérios de seleção para escolher o submodelo da dispersão.

De imediato percebemos uma grande vantagem do esquema proposto quando comparado à seleção conjunta dos dois submodelos, que é o menor custo computacional. Ao considerarmos m covariáveis candidatas para a média e para a dispersão, na seleção conjunta temos um total de $(2^m + 1)^2$ modelos candidatos a serem estimados. Já pelo esquema de duas etapas proposto temos um total de $2 \times (2^m + 1)$ modelos a serem estimados. Por exemplo, consideremos um caso de aplicação com 10 covariáveis candidatas para os submodelos da média e da dispersão. Teríamos assim um total de $(2^{10} + 1)^2 = 1050625$ modelos candidatos ao considerarmos a seleção conjunta e apenas $2 \times (2^{10} + 1) = 2050$ modelos candidatos considerando o esquema em duas etapas. Percebemos que em problemas práticos pode ser inviável estimar $(2^m + 1)^2$ modelos candidatos, até mesmo para valores razoavelmente pequenos de m .

No estudo de simulação para verificarmos o desempenho do esquema de seleção proposto consideramos várias combinações de critérios para as etapas (1) e (2). Para isso, utilizamos os resultados indicados pelas abordagens anteriores. Contudo, dentre as situações estudadas, as seguintes variantes do esquema proposto (EP) alcançaram os melhores resultados e serão apresentadas:

EP1: usa SICc para etapa (1) e \bar{R}_{RVp4}^2 para etapa (2);

EP2: usa SICc para etapa (1) e \bar{R}_{D3}^2 para etapa (2);

EP3: usa SICc para etapa (1) e SICc para etapa (2);

EP4: usa HQc para etapa (1) e HQc para etapa (2);

EP5: usa AIC para etapa (1) e \bar{R}_{RVp4}^2 para etapa (2);

EP6: usa \bar{R}_{RVp4}^2 para etapa (1) e \bar{R}_{D3}^2 para etapa (2);

EP7: usa \bar{R}_{RVp5}^2 para etapa (1) e \bar{R}_{RVp5}^2 para etapa (2).

Considerando essas sete variações do esquema proposto apresentamos, na Tabela 2.6, seus desempenhos na seleção de modelos de regressão beta com dispersão variável.

Comparando os melhores resultados das Tabelas 2.6 e 2.2 percebemos que a abordagem proposta proporciona um aumento na proporção de escolha do modelo verdadeiro em quase todos os cenários. A Figura 2.7 mostra graficamente a comparação dos melhores resultados de cada uma das duas abordagens, evidenciando o melhor desempenho do esquema de duas etapas.

Os resultados de desempenhos das variantes do esquema proposto mostram que EP1 apresentou melhores resultados quando o submodelo da dispersão é fracamente identificável, Modelos 2 e 4. Já para os modelos em que a dispersão é facilmente identificável, Modelos 1 e 3, verificamos que EP6 apresentou melhor desempenho quando $n = 25$; EP7 se destacou para $n = 50$; para os tamanhos amostrais de 100 e 200 os esquemas que apresentaram melhores desempenhos foram, respectivamente, EP4 e EP3. Contudo, assim como em outras abordagens,

Tabela 2.6 Porcentagem (%) de seleção correta dos regressores da média e dispersão utilizando o esquema de seleção proposto, para diferentes critérios em cada etapa.

n	Modelo 1				Modelo 2				Modelo 3				Modelo 4			
	25	50	100	200	25	50	100	200	25	50	100	200	25	50	100	200
EP1	10.5	24.2	29.6	31.2	9.0	15.6	25.7	37.7	2.4	7.0	19.7	29.5	1.7	5.2	16.6	35.5
EP2	15.8	24.6	36.7	42.5	5.2	9.5	17.7	29.5	2.8	6.3	24.2	40.2	1.0	2.9	11.5	27.9
EP3	1.6	20.7	64.6	92.5	0.2	0.6	2.3	8.9	0.2	5.0	40.8	86.8	0.0	0.1	1.6	7.1
EP4	3.8	28.4	65.4	79.1	0.6	1.6	6.6	18.9	1.1	11.4	49.9	76.2	0.2	0.7	5.1	17.6
EP5	8.5	17.6	22.5	24.0	6.7	11.4	19.1	28.6	4.0	10.2	20.1	22.7	3.2	7.1	16.6	28.8
EP6	12.8	15.9	23.6	27.6	3.9	6.2	11.4	19.3	5.3	9.7	21.8	26.0	1.4	4.2	10.5	19.2
EP7	10.9	30.9	42.9	45.6	4.9	9.3	19.6	33.4	3.4	13.5	35.3	43.2	1.4	4.0	15.8	33.5

os critérios que apresentam bons desempenhos em alguns cenários possuem fraco desempenho em outros. Dessa forma, identificamos algumas variantes do esquema em duas etapas que possuem bons comportamentos médios e os sugerimos para uso, são eles: EP1 ou EP5 para $n = 25, 50$ e EP4 para $n = 100, 200$.

Vale destacar que os resultados numéricos evidenciam que a seleção correta do submodelo da dispersão é a parte mais crítica na seleção de modelo de regressão beta com dispersão variável. Esse fato é notável ao analisarmos os resultados da Tabela 2.6, por exemplo, em que para os Modelos 2 e 4, onde o submodelo da dispersão é fracamente identificável, o percentual de seleção dos modelos corretos é inferior aos resultados dos Modelos 1 e 3.

2.5 Aplicação

Para aplicação do esquema de duas etapas em modelos de regressão beta com dispersão variável, utilizamos os dados de Pammer e Kevan (2007) referentes a um estudo da habilidade de leitura em um grupo de 44 crianças australianas da escola primária. Estes dados também foram utilizados por Smithson e Verkuilen (2006) e Ferrari *et al.* (2011).

Smithson e Verkuilen (2006) modelam a habilidade de leitura das crianças (y) em função da contribuição relativa do QI (quociente de inteligência) não-verbal (x_2) e da presença de dislexia (x_3). A variável resposta y é o escore de um teste de habilidade de leitura e as variáveis independentes x_2 e x_3 são, respectivamente, um escore padronizado de QI não-verbal e uma variável *dummy* para a presença (1) ou ausência (-1) de dislexia. Como a variável resposta pode assumir valores no intervalo $[a, b]$ foram feitas transformações nesta variável para que ela pertença ao intervalo aberto $(0, 1)$. Para melhor entender as variáveis em estudo apresentamos a Figura 2.8. Os gráficos mostram que as crianças com ausência de dislexia tendem a ter maior habilidade de leitura e também QI mais elevado.

Ainda em Smithson e Verkuilen (2006), os autores consideram uma terceira covariável (x_4) dada pela interação entre x_2 e x_3 , ou seja, $x_4 = x_2 \times x_3$. Primeiramente, os autores utilizam modelos de regressão linear para modelar y em função dessas covariáveis; posteriormente, uti-

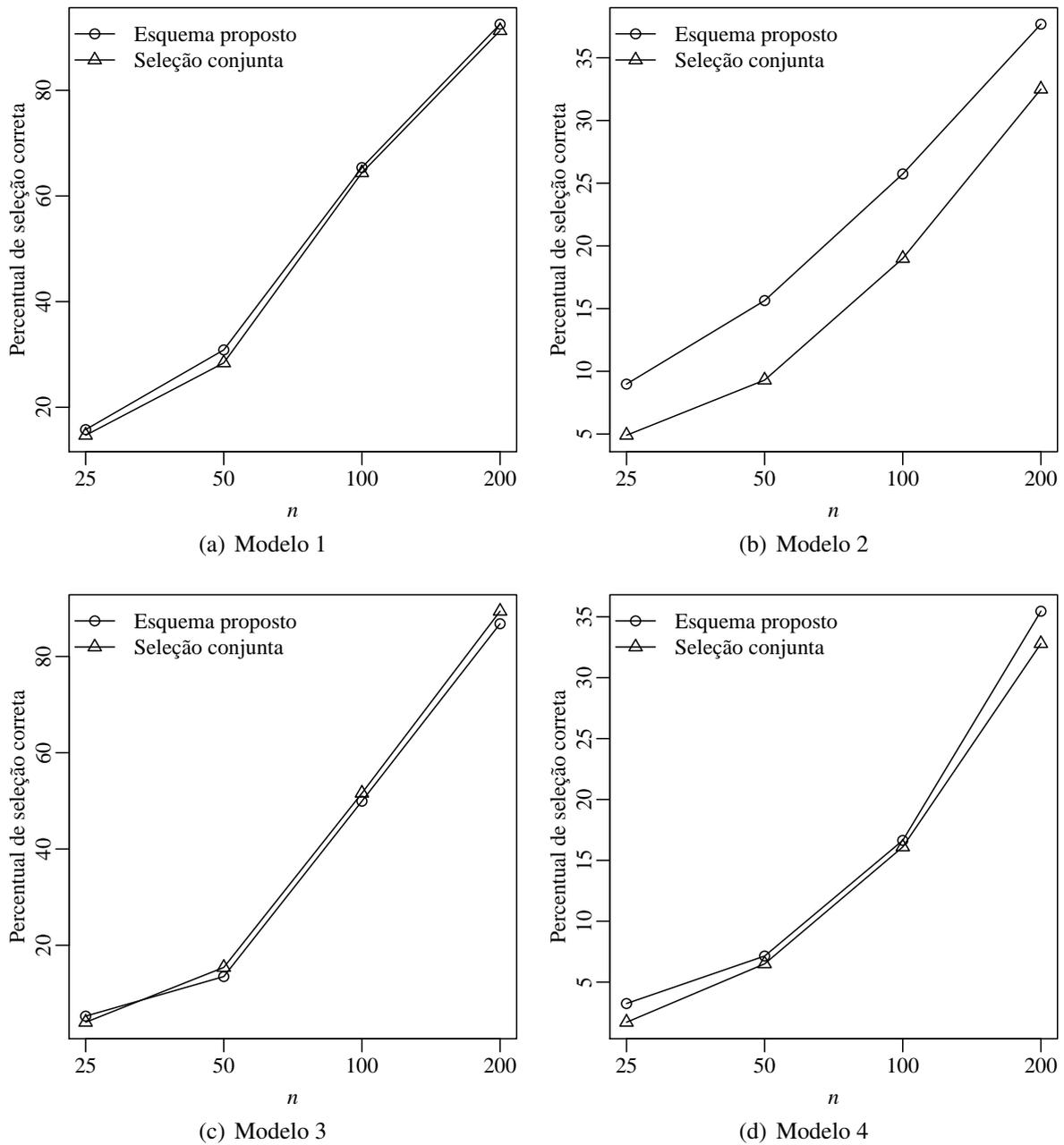


Figura 2.7 Comparação entre os melhores resultados do esquema de duas etapas proposto e da seleção conjunta.

lizam um modelo de regressão beta exclusivamente para a média, assumindo que o parâmetro de dispersão é constante. No entanto, eles concluem que os resultados inferenciais alcançados a partir desse modelo beta podem não ser adequados dada a não-constância do parâmetro de dispersão. Assim sendo, os autores utilizam um modelo de regressão beta com regressores para a média μ_t assim como para o parâmetro de precisão ϕ_t .

Em nossa aplicação consideramos um modelo de regressão beta com parâmetro de dis-

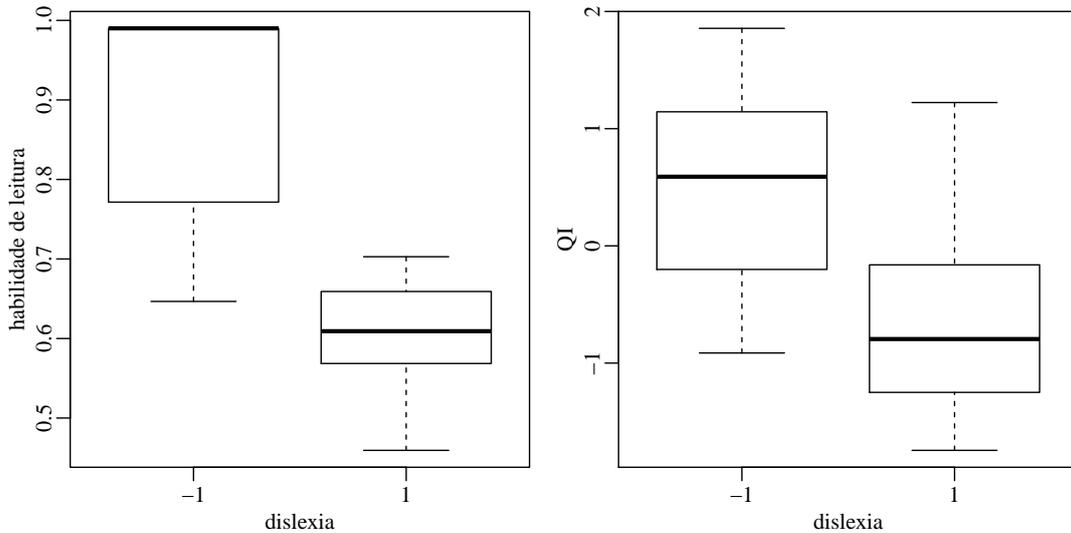


Figura 2.8 Boxplot dos dados de habilidade de leitura.

persão variável para modelar os dados. São consideradas funções de ligação logit para μ_t e para o parâmetro de dispersão σ_t . Incluímos ainda duas outras covariáveis, nomeadamente: $x_5 = x_2^2$ e $x_6 = x_3 \times x_5$. Como temos cinco covariáveis candidatas a regressoras, tanto para a média quanto para a dispersão, e considerando os modelos sempre com interceptos, temos um total de $2 \times (2^5 + 1) = 66$ modelos a serem estimados utilizando o esquema proposto. Utilizando a seleção conjunta dos regressores da média e da dispersão teríamos que estimar $(2^5 + 1)^2 = 1089$ modelos candidatos.

Em primeiro lugar testamos a constância da dispersão por meio de um teste escore, conforme apresentado na Seção 2.2 e no Apêndice F. Para esse teste, consideramos o modelo da média com as covariáveis candidatas x_2 , x_3 e x_4 e a hipótese nula $\mathcal{H}_0 : \gamma_2 = \gamma_3 = \gamma_4 = 0$, em que $\text{logit}(\sigma_t) = \gamma_1 + \gamma_2 x_2 + \gamma_3 x_3 + \gamma_4 x_4$. Para esse teste a estatística escore foi igual a 18.069, com p -valor igual a 0.0004. Com isso, rejeita-se fortemente a hipótese nula de dispersão constante e resolve-se modelar a dispersão.

Como temos uma amostra de tamanho 44, que é um tamanho amostral próximo de 50, os resultados numéricos apresentados na seção anterior indicam que podemos usar EP1 ou EP5. Utilizando EP1 é considerada apenas a covariável x_3 nos submodelos da média e da dispersão. A análise de resíduos desse modelo indica um ajuste inadequado, principalmente o envelope simulado. Já ao utilizar EP5, com SICc na etapa (1) e \bar{R}_{RVp4}^2 na etapa (2), o modelo considera as covariáveis x_3 , x_5 e x_6 no submodelo da média e as covariáveis x_2 , x_3 , x_4 e x_5 no submodelo da dispersão. Todas as covariáveis se mostram significativas, aos níveis nominais usuais, como pode ser verificado no modelo ajustado apresentado na Tabela 2.7. Os gráficos de resíduos desse modelo encontra-se na Figura 2.9.

Percebe-se que há uma grande diferença entre as medidas de qualidade $R_{FC}^2 = 0.63$ e $R_{RV}^2 = 0.88$ do modelo selecionado. Isto acontece pelo fato de R_{FC}^2 ser menos sensível ao ajuste correto do submodelo da dispersão do que R_{RV}^2 , que assume valores sensivelmente maiores quando o submodelo da dispersão é corretamente selecionado. Em modelos com dispersão

Tabela 2.7 Estimativas dos parâmetros do modelo de regressão beta com dispersão variável para dados de habilidade de leitura.

Parâmetros	Estimativa	Erro padrão	Estat. z	p-valor
Submodelo de μ				
β_1 (Constante)	1.0494	0.1605	6.539	0.0000
β_3 (Dislexia)	-0.8587	0.1587	-5.411	0.0000
β_5 (QI ²)	0.4524	0.0580	7.804	0.0000
β_6 (Dislexia \times QI ²)	-0.3866	0.0576	-6.720	0.0000
Submodelo de σ				
γ_1 (Constante)	-1.0072	0.1828	-5.509	0.0000
γ_2 (QI)	-0.9259	0.1498	-6.180	0.0000
γ_3 (Dislexia)	-0.9047	0.1603	-5.645	0.0000
γ_4 (Dislexia \times QI)	-0.8559	0.2633	-3.251	0.0025
γ_5 (QI ²)	-1.1005	0.2065	-5.328	0.0000
$R_{FC}^2 = 0.63$				
$R_{RV}^2 = 0.88$				

constante as duas medidas de qualidade do ajuste tendem a apresentar valores mais próximos. Portanto, sugerimos utilizar R_{RV}^2 como medida de qualidade do ajuste para modelos de regressão beta com dispersão variável.

Para os gráficos dos resíduos, apresentados na Figura 2.9, foi considerado o *resíduo padronizado 2* dado em (2.5). Uma implementação em R que retorna os quatro gráficos apresentados na Figura 2.9, com opção de diferentes resíduos, é apresentada no Apêndice G. As Figuras 2.9(a) e 2.9(b) indicam que os resíduos estão aleatoriamente distribuídos em torno de zero. Percebemos que não há observações atípicas, uma vez que todos os resíduos estão no intervalo $(-2, 2)$. A Figura 2.9(c) também evidencia um bom ajuste do modelo, mostrando pontos simetricamente distribuídos em torno da reta identidade. O envelope simulado, Figura 2.9(d), indica que o modelo está bem ajustado, apesar de apresentar alguns pontos fora dos limites. Seria indicada uma análise de influência, para identificação de possíveis pontos influentes com objetivo de melhorar, ou explicar, alguma pequena distorção dos resíduos, evidenciada pelo envelope simulado. Para maiores detalhes sobre análise de influência em regressão beta, ver Ferrari *et al.* (2011).

Por fim, o modelo de regressão beta apresentado na Tabela 2.7 difere do modelo para os dados de habilidade apresentado em Smithson e Verkuilen (2006). Estes autores modelam o parâmetro de precisão ϕ (e não o parâmetro de dispersão σ) utilizando a função de ligação $-\ln(\cdot)$ para o submodelo de ϕ . O modelo selecionado em Smithson e Verkuilen (2006) para estes dados considera ainda as covariáveis x_2 , x_3 e x_4 como regressoras para a média e x_2 e x_3 para o submodelo de ϕ . Esse mesmo modelo de Smithson e Verkuilen (2006) é também considerado em Ferrari *et al.* (2011).

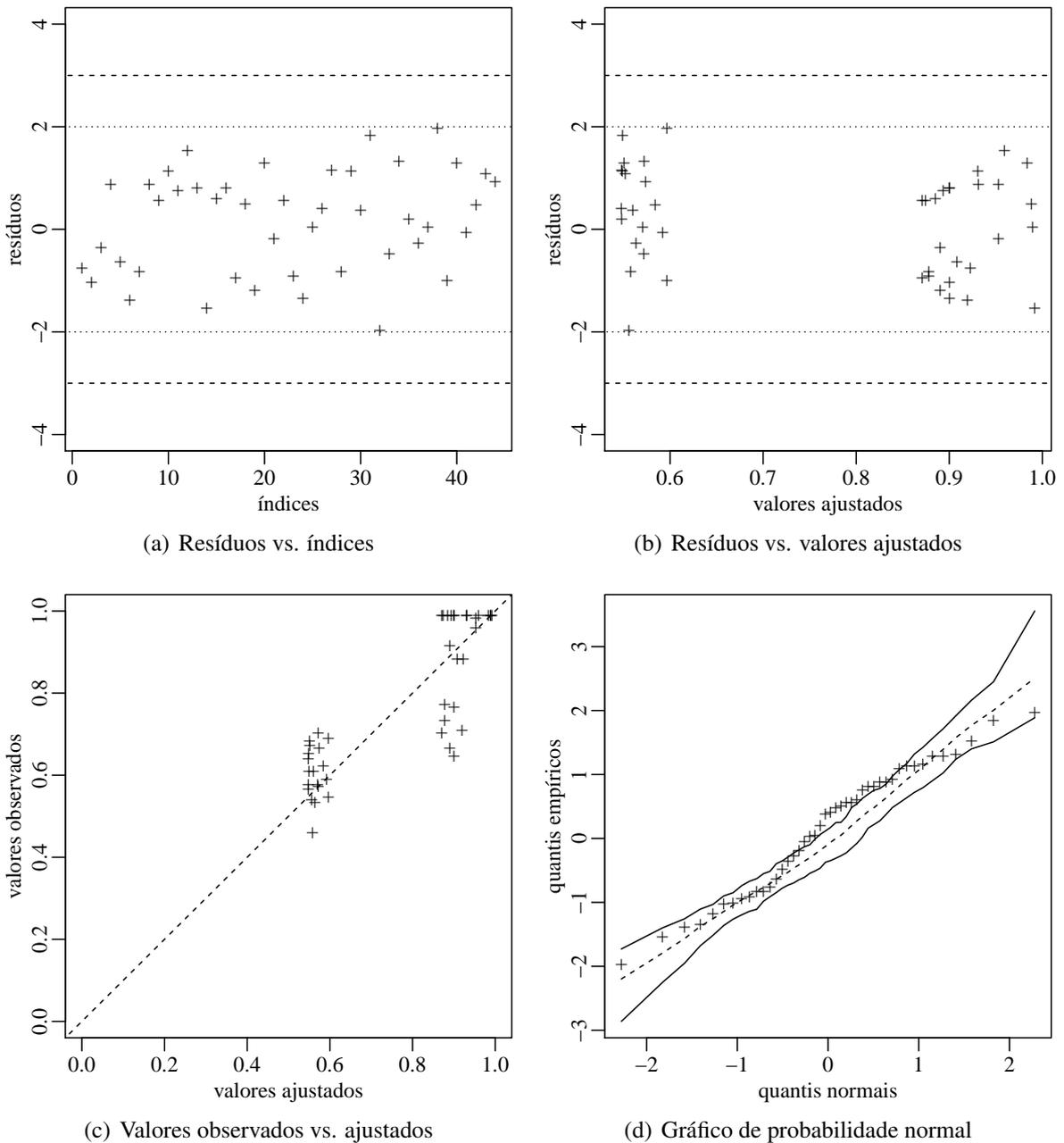


Figura 2.9 Gráficos de resíduos para o modelo ajustado.

2.6 Conclusões

Neste capítulo analisamos os desempenhos em amostras de tamanho finito de critérios de seleção em modelos de regressão beta com dispersão variável, propomos novos critérios de seleção para esses modelos e introduzimos um novo esquema de seleção em duas etapas. No modelo com dispersão variável o parâmetro de dispersão depende de covariáveis e parâmetros desconhecidos, da mesma forma que a média condicional. Os critérios de seleção são aplicados

para a escolha das variáveis regressoras para a média e para o parâmetro de dispersão.

Foram considerados diversos critérios de seleção no estudo de simulação, sendo aplicados a diferentes modelos e tamanhos amostrais. Além do uso de critérios de seleção baseados na função de log-verossimilhança maximizada, como aqueles da família do AIC, e dos critérios propostos em Hu e Shao (2008), que são critérios originalmente desenvolvidos para MLG, propusemos critérios de seleção ajustados baseados em diferentes medidas de qualidade de ajuste e específicos para modelos que consideram a modelagem da dispersão.

Notamos que os desempenhos dos critérios de seleção dependem fortemente da identificabilidade dos modelos. Além disso, concluímos que a utilização de critérios de seleção para selecionar o modelo da média e da dispersão conjuntamente não constitui boa prática. Com o intuito de obter uma seleção de modelos mais acurada e com menor custo computacional propusemos um esquema de seleção de modelos em duas etapas. Sugerimos primeiramente supor que o parâmetro de dispersão é constante e, então, utilizar os critério de seleção para selecionar o submodelo da média. Assumindo que o submodelo encontrado para média é o verdadeiro, então seleciona-se o submodelo da dispersão. Os resultados desta seleção por etapas mostraram-se superiores à seleção conjunta dos regressores da média e dispersão.

Nessa perspectiva, deve ser dado destaque aos critérios \bar{R}_{D3}^2 , \bar{R}_{RVp4}^2 , SICc e HQc. Para a seleção de modelos com dispersão variável investigamos diversas variantes do esquema proposto e, a partir de nossos resultados de simulação, sugerimos o uso dos esquemas EP1 ou EP5 para tamanhos amostrais menores ou iguais a 50 e o uso do esquema EP4 quando o tamanho amostral é maior que 50. O esquema EP1 utiliza o SICc para a etapa (1) e o \bar{R}_{RVp4}^2 para etapa (2); EP5 usa o AIC para etapa (1) e o \bar{R}_{RVp4}^2 para etapa (2) e EP4 usa o HQc para as duas etapas. Esses esquemas propostos foram empregados na seleção do modelo apresentado no estudo de aplicação aos dados de habilidade de leitura, o qual se mostrou bem ajustado aos dados.

Adicionalmente, por meio da utilização dos critérios de seleção baseados nas medidas de qualidade de ajuste e também por estudos preliminares, podemos concluir que a medida R_{RV}^2 , em (2.6), é a mais indicada para modelos de regressão beta com dispersão variável. Esta se mostrou mais sensível à modelagem correta do parâmetro de dispersão do que a medida de qualidade de ajuste R_{FC}^2 . Já a medida R_{HS}^2 não é indicada para modelos de regressão beta, uma vez que pode assumir valores negativos.

Variações Bootstrap do AIC em Regressão Beta

Resumo do capítulo

O critério de informação de Akaike (AIC) é um critério de seleção de modelos largamente utilizado em aplicações práticas. O AIC é um estimador do valor esperado da log-verossimilhança, sendo uma medida de discrepância entre o modelo verdadeiro e o modelo candidato estimado. No entanto, em pequenas amostras o AIC é viesado e tende a selecionar modelos com alta dimensionalidade. Para contornar esse problema nós propomos novos critérios de seleção para serem usados em pequenas amostras, denominados *bootstrap likelihood quasi-CV* (BQCV) e sua modificação 632QCV. Comparamos os desempenhos dos critérios propostos, do AIC e de suas diversas variações que utilizam log-verossimilhança bootstrap por meio de um extensivo estudo de simulação. O estudo numérico considera inúmeros cenários para a seleção em pequenas amostras de modelos de regressão beta com dispersão constante e com dispersão variável. Os resultados mostram que as variações bootstrap do AIC em regressão beta são boas alternativas para a seleção de modelos em pequenas amostras. Pode-se verificar que o uso da log-verossimilhança bootstrap diminui o problema da sobre-especificação do AIC na seleção de modelos. Dentre as diversas variações do AIC investigadas os critérios propostos, BQCV e 632QCV, destacaram-se, apresentando os melhores desempenhos na seleção de modelos em muitos cenários considerados. Apresentamos uma aplicação a dados reais, em que a seleção do modelo de regressão beta com dispersão variável é feita utilizando os critérios de seleção propostos.

Palavras-chave: AIC, bootstrap, critérios de seleção, dispersão variável, regressão beta, validação cruzada.

3.1 Introdução

Em análise de regressão, geralmente há o interesse em selecionar o melhor modelo a partir de uma ampla classe de modelos candidatos. Desta forma, a seleção de modelos é uma etapa essencial na análise estatística de dados e constitui uma importante área de pesquisa. O principal foco de pesquisa dentro da seleção de modelos se encontra nos critérios de seleção, ou critérios de informação. O critério de informação de Akaike (AIC) (Akaike, 1973) é, sem dúvida, o critério de seleção de modelos mais conhecido e aplicado. Muitas abordagens competidoras e/ou derivadas do AIC têm sido propostas, estimuladas pelo trabalho seminal de Akaike. Entre os critérios propostos estão o SIC (Schwarz, 1978), HQ (Hannan e Quinn, 1979) e AICc (Hurvich e Tsai, 1989).

O AIC foi proposto com o objetivo de estimar (menos duas vezes) a log-verossimilhança esperada. Usando expansão em série de Taylor e a propriedade da normalidade assintótica do estimador de máxima verossimilhança Akaike mostra que a função de log-verossimilhança maximizada é um estimador positivamente viesado para a log-verossimilhança esperada. Dessa forma, determinando um termo de ajuste para o viés, Akaike obtém o AIC como uma correção assintoticamente aproximada para a log-verossimilhança esperada. Contudo, em pequenas amostras o AIC é viesado e tende a escolher modelos com alta dimensionalidade (Hurvich e Tsai, 1989).

Focados em algumas classes de modelos, diversos pesquisadores têm investigado variações do AIC que objetivam melhorá-lo, reduzindo seu viés e melhorando seu desempenho na seleção de modelos em pequenas amostras. A primeira correção do AIC, o AICc, foi proposta em Sugiura (1978) para modelos de regressão linear. Posteriormente, Hurvich e Tsai (1989) ampliaram a aplicabilidade do AICc para regressão não-linear e modelos autorregressivos. Eles mostraram que tal critério possui propriedades assintóticas equivalentes ao AIC, mas desempenho superior em pequenas amostras. No entanto, correções analíticas para o AIC, como o AICc, podem ser de difícil obtenção em algumas classes de modelos (Shibata, 1997). As dificuldades analíticas advêm de resultados distribucionais e assintóticos, além de certas hipóteses restritivas. Com o intuito de contornar dificuldades analíticas e de obter correções mais acuradas em pequenas amostras surgem as variações bootstrap do AIC.

Extensões do AIC que utilizam métodos bootstrap (Efron, 1979), para determinar um melhor termo de ajuste do viés da log-verossimilhança maximizada, têm sido introduzidas e exploradas em diferentes classes de modelos. Alguns exemplos são os trabalhos Cavanaugh e Shumway (1997), Ishiguro e Sakamoto (1991), Ishiguro *et al.* (1997), Seghouane (2010), Shang e Cavanaugh (2008) e Shibata (1997), os quais introduzem critérios conhecidos como o WIC, AICb, EIC, entre outras denominações. Essas extensões bootstrap possuem vantagens sobre o AIC. Além de ser um processo com melhor acurácia em pequenas amostras do que aproximações analíticas, Shibata (1997) afirma que provavelmente a vantagem mais importante das variações bootstrap é que as mesmas são facilmente calculadas. Para tanto, apenas simulações de Monte Carlo são necessárias, ao passo que aproximações assintóticas podem se tornar muito complicadas para serem calculadas analiticamente.

Como visto, o AIC e suas extensões bootstrap têm o objetivo de estimar a log-verossimilhança esperada utilizando uma correção de viés para a log-verossimilhança maximizada. No entanto, seguindo a ideia de Pan (1999), propomos uma alternativa de prover diretamente um estimador para a log-verossimilhança esperada sem requerer um termo de ajuste de viés. Neste sentido, Pan (1999) utiliza bootstrap não-paramétrico e validação cruzada (CV) para determinar o critério chamado *bootstrap likelihood CV* (BCV). Utilizando bootstrap paramétrico e um método que denominamos por *quasi-CV* definimos uma nova variação do AIC para pequenas amostras. Para esse critério utilizamos a denominação *bootstrap likelihood quasi-CV* (BQCV) e também propomos a modificação denominada 632QCV.

Dado que critérios de seleção baseados na log-verossimilhança bootstrap têm sido explorados e aplicados satisfatoriamente em seleção de modelos autorregressivos (Ishiguro *et al.*, 1997), modelos de estado de tempo (Bengtsson e Cavanaugh, 2006; Cavanaugh e Shumway, 1997), modelos mistos (Shang e Cavanaugh, 2008), modelos de regressão linear (Pan, 1999; Seghouane, 2010) e modelos de regressão logística e regressão Cox (Pan, 1999), surge o inte-

resse em investigar os desempenhos de tais critérios em modelos de regressão beta. Essa classe de modelos, proposta originalmente por Ferrari e Cribari-Neto (2004), tem como objetivo permitir a modelagem de respostas que pertencem ao intervalo $(0,1)$, como taxas e proporções. Consideraremos o modelo de regressão beta com dispersão variável, em uma abordagem semelhante à de Simas *et al.* (2010), que é uma generalização do modelo para dispersão constante proposto em Ferrari e Cribari-Neto (2004). No modelo com dispersão variável a dispersão é modelada ao longo das observações por meio de uma estrutura de regressão que contém uma função de ligação, covariáveis e parâmetros desconhecidos, da mesma forma que a média condicional. Com isso, os critérios de seleção tornam-se importantes na seleção dos regressores da média e da dispersão no modelo de regressão beta.

O objetivo deste trabalho é propor novos critérios de seleção de modelos e investigar numericamente os seus desempenhos em pequenas amostras na seleção de modelos de regressão beta. Também investigamos os desempenhos das variações do AIC encontradas na literatura, que servem como métrica de comparação aos resultados do BQCV e do 632QCV. A avaliação numérica se dá na classe de modelos de regressão beta, nos casos de dispersão constante e dispersão variável. Apresentamos resultados de extensivas simulações que confirmam o bom desempenho dos critérios propostos.

Este capítulo encontra-se dividido da seguinte maneira. Na seção seguinte apresentamos uma breve introdução ao AIC, as extensões bootstrap para regressão beta utilizadas na avaliação numérica, assim como os critérios de seleção propostos. Na Seção 3.3 é apresentado o modelo de regressão beta com dispersão variável, sendo o modelo de dispersão constante um caso particular do mesmo; comentamos ainda sobre a estimação por máxima verossimilhança dos parâmetros do modelo. A Seção 3.4 apresenta os resultados do estudo de simulação e também uma discussão sobre os desempenhos do AIC e de suas variações para o caso de regressão beta com dispersão constante, em 3.4.1, e para o caso com dispersão variável, em 3.4.2. Na Seção 3.5 apresentamos uma aplicação a dados reais. Nessa aplicação, utilizamos os critérios propostos para selecionar um modelo de regressão beta com dispersão variável para dados de gastos com alimentação (Griffiths *et al.*, 1993, Tabela 15.4). Na última seção são apresentadas as principais conclusões e recomendações.

3.2 Critério de informação de Akaike e variações bootstrap

Segundo Kullback (1968) uma boa medida de separação entre duas densidades é a informação de Kullback-Leibler (KL), também conhecida como entropia ou discrepância (Cavanaugh, 1997). A informação KL pode ser usada como ferramenta para escolher um modelo estimado que seja o mais próximo possível do verdadeiro modelo. Utilizando a estratégia de minimizar a distância, em termos de KL, entre a densidade correspondente ao modelo estimado e a do modelo verdadeiro, Akaike (1973) introduziu o AIC. Seguindo notação semelhante à apresentada em Bengtsson e Cavanaugh (2006) formalizamos a seguir a noção de seleção de um modelo ajustado a partir de uma classe de modelos candidatos.

Suponha que o vetor n -dimensional Y é amostrado de acordo com uma densidade desconhecida $f(Y|\theta_{k_0})$, em que θ_{k_0} representa um vetor de dimensão k_0 que contém os parâmetros verdadeiros. A respectiva família paramétrica de densidades pode ser denotada por

$\mathcal{F}(k_i) = \{f(Y|\theta_{k_i}) | \theta_{k_i} \in \Theta_{k_i}\}$, em que Θ_{k_i} é o espaço paramétrico k_i -dimensional. Ainda, denotamos por $\hat{\theta}_{k_i}$ a estimativa de máxima verossimilhança do parâmetro θ_{k_i} . Essa estimativa é obtida maximizando $f(Y|\theta_{k_i})$ em Θ_{k_i} , em que $f(Y|\hat{\theta}_{k_i})$ representa a função de verossimilhança maximizada.

O objetivo do AIC é selecionar dentre a classe de famílias $\mathcal{F} = \{\mathcal{F}(k_1), \mathcal{F}(k_2), \dots, \mathcal{F}(k_L)\}$ para o modelo estimado $f(Y|\hat{\theta}_{k_i})$, com $i = 1, 2, \dots, L$, a que melhor se aproxima de $f(Y|\theta_{k_0})$. Por simplicidade de notação não consideraremos famílias diferentes na classe \mathcal{F} que possuam a mesma dimensão. Isto posto, sendo $f(Y|\theta_k)$ o modelo candidato dizemos que $f(Y|\hat{\theta}_k)$ é corretamente especificado se $f(Y|\theta_{k_0}) \in \mathcal{F}(k)$, em que $\mathcal{F}(k)$ é a família de menor dimensão que contém $f(Y|\theta_{k_0})$. Se $f(Y|\theta_{k_0}) \in \mathcal{F}(k)$ mas famílias com dimensões menores também contêm $f(Y|\theta_{k_0})$, então dizemos que $f(Y|\hat{\theta}_k)$ é sobre-especificado. Por outro lado, $f(Y|\hat{\theta}_k)$ será subespecificado se $f(Y|\theta_{k_0}) \notin \mathcal{F}(k)$.

Para determinar dentre os modelos estimados $f(Y|\hat{\theta}_{k_1}), f(Y|\hat{\theta}_{k_2}), \dots, f(Y|\hat{\theta}_{k_L})$ qual é o mais próximo de $f(Y|\theta_{k_0})$ considera-se a medida KL entre o modelo verdadeiro $f(Y|\theta_{k_0})$ e o modelo candidato $f(Y|\theta_k)$, dada por

$$d(\theta_{k_0}, \theta_k) = E_0 \left[\log \left\{ \frac{f(Y|\theta_{k_0})}{f(Y|\theta_k)} \right\} \right],$$

em que $E_0(\cdot)$ denota a esperança sob $f(Y|\theta_{k_0})$. Definindo

$$\delta(\theta_{k_0}, \theta_k) = E_0 \{-2 \log f(Y|\theta_k)\}, \quad (3.1)$$

temos que $2d(\theta_{k_0}, \theta_k) = \delta(\theta_{k_0}, \theta_k) - \delta(\theta_{k_0}, \theta_{k_0})$. Como $\delta(\theta_{k_0}, \theta_{k_0})$ não depende de θ_k percebemos que minimizar $2d(\theta_{k_0}, \theta_k)$, ou $d(\theta_{k_0}, \theta_k)$, equivale a minimizar a discrepância $\delta(\theta_{k_0}, \theta_k)$. Portanto, o modelo $f(Y|\theta_k)$ que minimiza menos duas vezes a log-verossimilhança esperada, $\delta(\theta_{k_0}, \theta_k)$, é o modelo mais próximo, em termos da informação de Kullback-Leibler, do modelo verdadeiro.

Consideremos agora um dado conjunto de estimativas de máxima verossimilhança para θ_k , $\hat{\theta}_k$, e definamos

$$\delta(\theta_{k_0}, \hat{\theta}_k) = E_0 \{-2 \log f(Y|\theta_k)\} |_{\theta_k = \hat{\theta}_k}.$$

Desta forma, $\delta(\theta_{k_0}, \hat{\theta}_k)$ fornece uma medida de separação entre o modelo real e o modelo candidato estimado. No entanto, avaliar $\delta(\theta_{k_0}, \hat{\theta}_k)$ não é possível, pois requer o conhecimento da densidade $f(Y|\theta_{k_0})$. Akaike (1973) propõe uma solução aproximada na qual $-2 \log f(Y|\hat{\theta}_k)$ é um estimador viesado de $\delta(\theta_{k_0}, \hat{\theta}_k)$, cujo viés

$$B = E_0 \{-2 \log f(Y|\hat{\theta}_k) - \delta(\theta_{k_0}, \hat{\theta}_k)\} \quad (3.2)$$

pode ser assintoticamente aproximado por $-2k$, lembrando que k é a dimensão de θ_k .

Desta forma, o valor esperado do critério de informação de Akaike, definido por

$$\text{AIC} = -2 \log f(Y|\hat{\theta}_k) + 2k,$$

é assintoticamente igual ao valor esperado de $\delta(\theta_{k_0}, \hat{\theta}_k)$, definido por

$$\Delta(\theta_{k_0}, k) = E_0 \{\delta(\theta_{k_0}, \hat{\theta}_k)\}.$$

Notamos então que o termo de qualidade de ajuste do AIC, $-2\log f(Y|\hat{\theta}_k)$, é uma estimativa viesada para menos duas vezes a log-verossimilhança esperada, enquanto que o termo penalizador do AIC, $2k$, é um termo de ajuste para o viés dado em (3.2).

Por ser uma aproximação para grandes amostras, a utilidade do AIC em pequenas amostras é limitada (Bengtsson e Cavanaugh, 2006). Neste sentido, muitos trabalhos têm sido desenvolvidos com o objetivo de propor melhores estimadores para a log-verossimilhança esperada em pequenas amostras ou para casos em que o número de parâmetros do modelo é grande quando comparado ao tamanho amostral. Para modelos de regressão linear Sugiura (1978) desenvolveu o AICc, que, para essa classe de modelos, é um estimador exatamente não-viesado de $\Delta(\theta_{k_0}, k)$, ou seja, $E_0\{\text{AICc}\} = \Delta(\theta_{k_0}, k)$. Baseados em Sugiura (1978), Hurvich e Tsai (1989) estendem o uso do AICc para regressão não-linear e para modelos autorregressivos. Os autores mostram que o AICc não é mais exatamente não-viesado para esses modelos, mas é assintoticamente equivalente ao AIC, sendo $E_0\{\text{AICc}\} + o(1) = \Delta(\theta_{k_0}, k)$, e tipicamente possui melhor desempenho do que o AIC em pequenas amostras.

Segundo Cavanaugh (1997), a vantagem do AICc sobre o AIC é que em pequenas amostras o AICc estima a discrepância esperada com menor viés do que o AIC. Esse fato indica um desempenho pobre do AIC na seleção de modelos em pequenas amostras. Por outro lado, a vantagem do AIC sobre o AICc é que o AIC é universalmente aplicável, independentemente da classe de modelos, enquanto que a derivação AICc depende da forma do modelo candidato. Com isso, na busca de critérios de seleção de modelos com melhores desempenhos em pequenas amostras e na tentativa de determinar variações para o AIC que não dependam da classe de modelos na qual se está trabalhando é que a abordagem bootstrap é utilizada.

3.2.1 Extensões bootstrap do AIC

Denominaremos de extensões bootstrap do AIC (EIC) os critérios que utilizam estimadores bootstrap para o termo de viés B dado em (3.2). O objetivo é obter estimativas bootstrap de B que sejam mais acuradas do que $-2k$ em pequenas amostras, conduzindo a critérios com melhores desempenhos na seleção de modelos.

Utilizaremos cinco estimadores bootstrap diferentes, B_i , para B , com $i = 1, \dots, 5$. Esses estimadores B_i definem cinco extensões bootstrap do AIC, que denominaremos, respectivamente, de EIC $_i$ para $i = 1, \dots, 5$. Com isso, as extensões bootstrap do AIC utilizadas para a seleção de modelos de regressão beta em pequenas amostras possuem a seguinte forma:

$$\text{EIC}_i = -2\log f(Y|\hat{\theta}_k) + B_i, \quad i = 1, \dots, 5,$$

em que os B_i 's serão definidos a seguir.

Considere Y^* representando a amostra bootstrap, paramétrica ou não-paramétrica, e E_* a esperança com respeito à distribuição de Y^* . Considere W amostras bootstrap $Y^*(i)$ e o conjunto $\{\hat{\theta}_k^*(i)\}$, com $i = 1, 2, \dots, W$, das W réplicas bootstrap de $\hat{\theta}_k$, i.e., a estimativa $\hat{\theta}_k^*(i)$ é o valor de θ_k que maximiza a função de verossimilhança $f(Y^*(i)|\theta_k)$.

Em Ishiguro *et al.* (1997) os autores propõem uma extensão bootstrap do AIC, que é denominada EIC. O EIC é o caso particular do WIC, proposto em Ishiguro e Sakamoto (1991), para dados independente e identicamente distribuídos (i.i.d.). No presente trabalho denominaremos

este critério de EIC1, em que a estimativa bootstrap para o viés (3.2) é dada por

$$B_1 = E_* \{2 \log f(Y^* | \hat{\theta}_k^*) - 2 \log f(Y | \hat{\theta}_k^*)\}.$$

Outra extensão bootstrap é apresentada em Cavanaugh e Shumway (1997), para seleção de modelos de estado de tempo. Tal critério, que denominaremos por EIC2, possui o seguinte estimador para (3.2):

$$B_2 = 2E_* \{2 \log f(Y | \hat{\theta}_k) - 2 \log f(Y | \hat{\theta}_k^*)\}.$$

Os critérios EIC1 e EIC2 são utilizados sob as denominações de AICb1 e AICb2 em Shang e Cavanaugh (2008) para a seleção de modelos mistos e utilizando bootstrap paramétrico.

Shibata (1997), além de mostrar que B_1 e B_2 são assintoticamente equivalentes, propõe as seguintes três estimativas bootstrap de (3.2):

$$B_3 = 2E_* \{2 \log f(Y^* | \hat{\theta}_k^*) - 2 \log f(Y^* | \hat{\theta}_k)\},$$

$$B_4 = 2E_* \{2 \log f(Y^* | \hat{\theta}_k) - 2 \log f(Y | \hat{\theta}_k^*)\},$$

$$B_5 = 2E_* \{2 \log f(Y^* | \hat{\theta}_k^*) - 2 \log f(Y | \hat{\theta}_k)\}.$$

Destes três termos penalizadores são derivadas então outras três extensões bootstrap do AIC, que denominamos EIC3, EIC4 e EIC5, respectivamente. Em Seghouane (2010), considerando bootstrap paramétrico, são propostas duas versões corrigidas do AIC para modelos de regressão linear como aproximações assintóticas de EIC1, EIC2, EIC3, EIC4 e EIC5.

3.2.2 Verossimilhança bootstrap e validação cruzada

Os critérios de seleção apresentados até agora objetivam estimar a log-verossimilhança esperada utilizando uma correção de viés para a log-verossimilhança maximizada. Por outro lado, na tentativa de prover diretamente um estimador para a log-verossimilhança esperada, em Pan (1999) é introduzido um critério de seleção baseado em validação cruzada (CV) e bootstrap.

A técnica de CV é largamente utilizada na estimação da taxa de erro de predição de modelos (Efron, 1983; Efron e Tibshirani, 1997). No âmbito da seleção de modelos, segundo Davies *et al.* (2005), o primeiro critério baseado em CV foi o PRESS (Allen, 1974). Por outro lado, o uso de técnicas de bootstrap para melhorar o desempenho de regras de seleção de modelos foi sugerido em Efron (1986). Finalmente, em Breiman e Spector (1992) e Hjorth (1994) é discutida a utilização de CV e bootstrap na seleção de modelos.

Segundo Efron (1983) e Efron e Tibshirani (1997), CV tem a propriedade de remover o viés, mas, por outro lado, conduz a estimadores com alta variância. Contudo, usando bootstrap pode-se reduzir essa variabilidade. Com isso, seguindo a mesma ideia de Efron (1983), mas no contexto de seleção de modelos, Pan (1999) apresenta um método baseado na combinação de bootstrap não-paramétrico e CV, o *bootstrap likelihood CV* (BCV). BCV fornece um estimador de (3.1) que não necessita de correção de viés. Considerando uma amostra Y de tamanho n , o BCV é definido por

$$\text{BCV} = E_* \left\{ -2 \log f(Y^- | \hat{\theta}_k^*) \frac{n}{m^*} \right\},$$

em que Y^* é a amostra bootstrap não-paramétrica, $Y^- = Y - Y^*$, ou seja, $Y = Y^- \cup Y^*$ e $Y^- \cap Y^* = \emptyset$ e $m^* > 0$ é o número de elementos de Y^- . Dessa forma, nenhuma observação de Y é usada duas vezes: cada observação ou está em Y^* ou está em Y^- .

Seguindo Efron (1983), Pan argumenta que o BCV pode sobre-estimar (3.1) e, por outro lado, $-2 \log f(Y|\hat{\theta}_k)$ subestima o mesmo. Com isso, seguindo a regra 632+ (Efron e Tibshirani, 1997), Pan introduz o 632CV como a seguinte média ponderada:

$$632CV = 0.368 \{-2 \log f(Y|\hat{\theta}_k)\} + 0.632BCV.$$

3.2.3 Verossimilhança bootstrap e quasi-CV

A seguir, introduziremos novos critérios de seleção de modelos que incorporam correções para pequenas amostras. Assim como o BCV, estes critérios fornecem estimadores diretos para a log-verossimilhança esperada.

Sejam F a função de distribuição da amostra observada $Y = (y_1, \dots, y_n)$ e \hat{F} a função de distribuição estimada, ou seja, \hat{F} é a função de distribuição F avaliada na estimativa $\hat{\theta}$. Dessa forma, definimos

$$Y_p^* = (y_1^*, y_2^*, \dots, y_n^*) \sim \hat{F} \quad \text{amostra de estimação (ou treinamento),} \quad (3.3)$$

$$Y = (y_1, y_2, \dots, y_n) \sim F \quad \text{amostra de validação.} \quad (3.4)$$

A partir de W pseudo-amostras Y_p^* segundo \hat{F} e sendo $\{\hat{\theta}_k^{p^*}(i), i = 1, 2, \dots, W\}$ o conjunto das W réplicas bootstrap de $\hat{\theta}_k$, definimos o critério *bootstrap likelihood quasi-CV* (BQCV) da seguinte maneira:

$$BQCV = E_{p^*} \{-2 \log f(Y|\hat{\theta}_k^{p^*})\},$$

em que E_{p^*} é a esperança com respeito à distribuição de Y_p^* .

Pela lei forte dos grandes números temos que

$$\frac{1}{W} \sum_{i=1}^W \{-2 \log f(Y|\hat{\theta}_k^{p^*}(i))\} \xrightarrow{q.c.} E_{p^*} \{-2 \log f(Y|\hat{\theta}_k^{p^*})\}.$$

em que $\xrightarrow{q.c.}$ denota convergência quase certa.

Com isso, podemos resumir o procedimento para obtenção do BQCV pelo seguinte algoritmo:

1. Usando a amostra $Y = (y_1, \dots, y_n)$ estima-se o vetor de parâmetros θ ;
2. Utilizando a distribuição estimada \hat{F} , que é a distribuição F avaliada em $\hat{\theta}$, geram-se W pseudo-amostras Y_p^* ;
3. Para cada $Y_p^*(i)$, com $i = 1, \dots, W$, determina-se $\hat{\theta}_k^{p^*}(i)$ e calcula-se $-2 \log f(Y|\hat{\theta}_k^{p^*}(i))$;
4. Com as W réplicas de Monte Carlo de $-2 \log f(Y|\hat{\theta}_k^{p^*}(i))$ calcula-se

$$BQCV = \frac{1}{W} \sum_{i=1}^W \{-2 \log f(Y|\hat{\theta}_k^{p^*}(i))\}.$$

Baseados em experimentos computacionais, sugerimos $W = 200$.

Notamos que o processo não satisfaz a uma validação cruzada genuína, justificando a denominação *quasi-CV*, pois não particiona a amostra Y , mas sim trata as amostras Y_p^* e Y como partições de um mesmo conjunto de dados. Com isso, a cada realização da amostra de estimação do BQCV temos um procedimento que se assemelha ao *2-fold CV*. Neste caso, a amostra de treinamento é a pseudo-amostra do esquema bootstrap paramétrico, Y_p^* , e a amostra de validação é a amostra observada, Y .

Seguindo o mesmo procedimento utilizado por Pan (1999) para a determinação do 632CV, determinamos um outro critério de seleção, denominado 632QCV, como uma variação do BQCV. Esse novo critério de seleção é determinado pela seguinte média ponderada:

$$632QCV = 0.368 \{ -2 \log f(Y | \hat{\theta}_k) \} + 0.632BQCV.$$

3.3 O modelo de regressão beta

Muitos estudos, em diferentes áreas do conhecimento, como em Brehm e Gates (1993), Hancox *et al.* (2010), Kieschnick e McCullough (2003), Smithson e Verkuilen (2006) e Zucco (2008), examinam como um conjunto de covariáveis se relaciona com uma variável resposta que assume valores no intervalo $(0,1)$, como taxas e proporções. Para estas situações Ferrari e Cribari-Neto (2004) propuseram os modelos de regressão beta, em que a variável resposta (y) segue lei beta. A distribuição beta é bastante flexível para modelar proporções, uma vez que, dependendo dos valores dos dois parâmetros que a indexam, a densidade assume formas bem variadas. Em uma parametrização diferente da proposta em Ferrari e Cribari-Neto (2004), considerando os parâmetros de média μ e de dispersão σ , a densidade de y é dada por

$$\pi(y|\mu, \sigma) = \frac{\Gamma(\frac{1-\sigma^2}{\sigma^2})}{\Gamma(\mu(\frac{1-\sigma^2}{\sigma^2}))\Gamma((1-\mu)(\frac{1-\sigma^2}{\sigma^2}))} y^{\mu(\frac{1-\sigma^2}{\sigma^2})-1} (1-y)^{(1-\mu)(\frac{1-\sigma^2}{\sigma^2})-1}, \quad 0 < y < 1, \quad (3.5)$$

em que $0 < \mu < 1$, $0 < \sigma < 1$, $\Gamma(\cdot)$ é a função gama e $V(\mu) = \mu(1-\mu)$ denota a função de variância. A média e a variância de y são dadas, respectivamente, por

$$\begin{aligned} E(y) &= \mu, \\ \text{var}(y) &= V(\mu)\sigma^2. \end{aligned}$$

Seja $Y = (y_1, \dots, y_n)$ um vetor de variáveis aleatórias independentes, em que cada y_t , $t = 1, \dots, n$, tem densidade na forma (3.5) com média μ_t e parâmetro de dispersão desconhecido σ_t . Com isso, o modelo de regressão beta com dispersão variável pode ser escrito como

$$g(\mu_t) = \sum_{i=1}^r x_{ti} \beta_i = \eta_t, \quad (3.6)$$

$$h(\sigma_t) = \sum_{i=1}^s z_{ti} \gamma_i = \nu_t, \quad (3.7)$$

em que $\beta = (\beta_1, \dots, \beta_r)^\top$ e $\gamma = (\gamma_1, \dots, \gamma_s)^\top$ são vetores de parâmetros desconhecidos para a média e dispersão, respectivamente. Ainda, $x_t = (x_{t1}, \dots, x_{tr})$ e $z_t = (z_{t1}, \dots, z_{ts})$ são observações de r e s variáveis independentes conhecidas, em que $r + s = k < n$. A matriz de regressores para a média, X , é uma matriz $n \times r$ cuja t -ésima linha é x_t e a matriz de regressores para a dispersão, Z , é uma matriz $n \times s$ cuja t -ésima linha é z_t . Quando interceptos são incluídos nos submodelos da média e dispersão, temos que $x_{t1} = z_{t1} = 1$, para $t = 1, \dots, n$. O modelo de regressão beta com dispersão constante é o caso particular em que $s = 1$, $z_{t1} = 1$ e h é a ligação identidade. Finalmente, $g(\cdot)$ e $h(\cdot)$ são funções de ligação estritamente monótonas e duas vezes diferenciáveis, com domínio em $(0,1)$ e imagem em \mathbb{R} . Na parametrização que consideramos, as mesmas funções de ligação usadas para média podem ser usadas para a dispersão. Para o modelo de regressão beta, com a parametrização proposta, podemos escolher diferentes funções de ligação como, por exemplo, a função logit ou a função probit. Uma discussão detalhada sobre essas e outras funções de ligação pode ser encontrada em McCullagh e Nelder (1989).

Para a estimação conjunta dos vetores de parâmetros β e γ é utilizado o método da máxima verossimilhança. Considerando $\theta_k = (\beta_1, \dots, \beta_r, \gamma_1, \dots, \gamma_s)^\top$ e o vetor Y , com n variáveis aleatórias independentes, o logaritmo da função de verossimilhança $f(Y|\theta_k)$ é

$$\log f(Y|\theta_k) = \sum_{t=1}^n \log f(y_t|\mu_t, \sigma_t),$$

em que

$$\begin{aligned} \log f(y_t|\mu_t, \sigma_t) &= \log \Gamma\left(\frac{1 - \sigma_t^2}{\sigma_t^2}\right) - \log \Gamma\left(\mu_t \left(\frac{1 - \sigma_t^2}{\sigma_t^2}\right)\right) - \log \Gamma\left((1 - \mu_t) \left(\frac{1 - \sigma_t^2}{\sigma_t^2}\right)\right) \\ &\quad + \left[\mu_t \left(\frac{1 - \sigma_t^2}{\sigma_t^2}\right) - 1\right] \log y_t + \left[(1 - \mu_t) \left(\frac{1 - \sigma_t^2}{\sigma_t^2}\right) - 1\right] \log(1 - y_t). \end{aligned}$$

A função escore é obtida pela diferenciação da função de log-verossimilhança ($\log f(Y|\theta_k)$) em relação aos parâmetros desconhecidos, conforme o Apêndice D. No Apêndice E encontra-se a expressão analítica da matriz de informação de Fisher, necessária para inferências em grandes amostras. Sendo $U_\beta(\beta, \gamma)$ a função escore para o vetor de parâmetros β e $U_\gamma(\beta, \gamma)$ a função escore para o vetor de parâmetros γ , os estimadores de máxima verossimilhança são obtidos pela solução do seguinte sistema:

$$\begin{cases} U_\beta(\beta, \gamma) = 0, \\ U_\gamma(\beta, \gamma) = 0. \end{cases}$$

Tal solução não possui forma fechada, fazendo-se necessário o uso de algoritmos de otimização não-linear para a obtenção de estimativas de máxima verossimilhança.

Como medida de qualidade de ajuste do modelo, sugerimos a transformação da razão de verossimilhanças (Nagelkerke, 1991), dada por

$$R_{RV}^2 = 1 - \left(\frac{L_{\text{null}}}{L_{\text{fit}}}\right)^{2/n},$$

em que L_{null} é a função de verossimilhança maximizada do modelo sem regressores e L_{fit} é a função de verossimilhança maximizada do modelo ajustado. Uma outra medida que pode ser considerada é o quadrado do coeficiente de correlação amostral entre $g(y)$ e $\hat{\eta} = X\hat{\beta}$, em que $\hat{\beta}$ denota o estimador de máxima verossimilhança de β . Essa medida de qualidade de ajuste, que denominamos por R_{FC}^2 , foi proposta por Ferrari e Cribari-Neto (2004) para os modelos de regressão beta com dispersão constante.

3.4 Avaliação numérica

Nesta seção investigaremos os desempenhos do AIC e de suas variações bootstrap para pequenas amostras aplicados na seleção de modelos de regressão beta. Para tanto, utilizamos simulações de Monte Carlo com implementação na linguagem matricial de programação Ox versão 5.1 (Doornik, 2007). A estimação pelo método da máxima verossimilhança foi feita utilizando o algoritmo de otimização *quasi*-Newton BFGS¹ com primeiras derivadas analíticas.

Os modelos de regressão beta considerados no estudo de simulação possuem submodelo para média na forma apresentada em (3.6) e submodelo para dispersão na forma apresentada em (3.7). Consideramos 1000 amostras para diferentes valores do vetor de parâmetros e utilizamos $W = 200$ reamostras bootstrap para a log-verossimilhança bootstrap. Segundo simulações prévias omitidas neste texto, valores de W superiores a 200 conduzem a melhoramentos negligíveis no desempenho dos critérios de seleção. Para as extensões bootstrap do AIC investigamos o uso de bootstrap paramétrico, $EICi_p$, assim como o uso de bootstrap não-paramétrico, $EICi_{np}$. Os valores das covariáveis são obtidos como realizações da distribuição uniforme padrão $\mathcal{U}(0,1)$, os quais permanecem constantes durante todo o experimento. A cada iteração da simulação são geradas ocorrências da variável resposta com distribuição beta de parâmetros μ_t e σ_t dados por (3.6) e (3.7). Em todos os casos consideramos a função de ligação logit para $g(\cdot)$ e $h(\cdot)$.

Além do AIC e de suas variações bootstrap já apresentadas, nós consideramos outros critérios de seleção para avaliação de desempenho. Foram considerados o AICc (Hurvich e Tsai, 1989), SIC (Schwarz, 1978), SICc (McQuarrie, 1999), HQ (Hannan e Quinn, 1979) e HQc (McQuarrie e Tsai, 1998). O uso destes critérios em modelos de regressão beta é feito de maneira *ad hoc*, uma vez que alguns destes critérios não se estendem de maneira direta a essa classe de modelos.

O estudo numérico encampa dois casos. Primeiramente, na Seção 3.4.1, consideramos dispersão constante, utilizando $r = 7$, $s = 1$ e a presença de interceptos. Posteriormente, na Seção 3.4.2, são apresentados os resultados para o modelo de regressão beta com dispersão variável, considerando $r = 6$, $s = 6$ e também com interceptos.

3.4.1 Critérios de seleção de modelos em regressão beta com dispersão constante

Para o modelo de regressão beta com dispersão constante nós conduzimos o estudo de simulação seguindo a abordagem de Davies *et al.* (2005), Hurvich e Tsai (1989), Pan (1999)

¹Para detalhes do algoritmo BFGS, ver Press *et al.* (1992)

e Seghouane (2010) no âmbito de regressão linear. A dimensão dos modelos verdadeiros é $k_0 = 4$, que são dados por

$$\text{logit}(\mu_t) = -1.5 + x_{t2} + x_{t3}, \quad \text{logit}(\sigma_t) = -1, \quad (3.8)$$

$$\text{logit}(\mu_t) = 1 - 0.75x_{t2} - 0.25x_{t3}, \quad \text{logit}(\sigma_t) = -1, \quad (3.9)$$

$$\text{logit}(\mu_t) = -1.5 + x_{t2} + x_{t3}, \quad \text{logit}(\sigma_t) = -2.2, \quad (3.10)$$

$$\text{logit}(\mu_t) = 1 - 0.75x_{t2} - 0.25x_{t3}, \quad \text{logit}(\sigma_t) = -2.2. \quad (3.11)$$

Consideramos tamanhos amostrais iguais a $n = 15, 20, 30, 40$ e a inclusão de seis covariáveis candidatas obtidas como realizações da distribuição $\mathcal{U}(0,1)$. Com isso, juntamente com a coluna referente ao intercepto, temos uma matriz X , com dimensão $n \times 7$, de covariáveis para o submodelo da média. Por simplicidade, assim como em Davies *et al.* (2005), Hurvich e Tsai (1989), Pan (1999) e Seghouane (2010), os modelos candidatos são sequencialmente aninhados, ou seja, o modelo candidato com dimensão k consiste no modelo com as $1, 2, \dots, k-1$ primeiras colunas de X . O modelo verdadeiro consiste nas três primeiras colunas de X . Desta forma, os critérios de informação são utilizados para selecionar o valor de k .

Notamos que os quatro modelos/cenários estudados possuem as características elencadas a seguir. Os dois primeiros, apresentados em (3.8) e (3.9), possuem grande dispersão, enquanto que os dois últimos, (3.10) e (3.11), possuem pequena dispersão. Adicionalmente, os modelos (3.8) e (3.10) são facilmente identificáveis, uma vez que todos os parâmetros, a menos dos interceptos, possuem os mesmos valores ($\beta_2 = \beta_3 = 1$). Já os modelos dados em (3.9) e (3.11) são fracamente identificáveis. A fraca identificabilidade é caracterizada pela proximidade a zero dos valores absolutos dos parâmetros β_i à medida que i cresce. No cenário de fraca identificabilidade as covariáveis possuem diferentes influências sobre a variável resposta. Por exemplo, nos modelos (3.9) e (3.11) a variável x_3 tem menos influência sobre a variável resposta do que a variável x_2 . O termo “modelos facilmente identificáveis” está sendo usado no mesmo sentido do termo *easily identified models* apresentado em McQuarrie e Tsai (1998).

Os resultados numéricos do estudo de simulação, com os desempenhos dos critérios de seleção em regressão beta com dispersão constante, são apresentados nas Tabelas 3.1, 3.2, 3.3 e 3.4 para os modelos dados em (3.8), (3.9), (3.10) e (3.11), respectivamente. Para cada critério apresentamos a frequência de escolha da verdadeira ordem do modelo ($= k_0$), assim como a frequência de escolha de modelos subespecificados ($< k_0$) e sobre-especificados ($> k_0$).

Os resultados na Tabela 3.1 mostram o bom desempenho do BQCV. Para os menores tamanhos amostrais, $n = 15$ e $n = 20$, o BQCV foi o critério que apresentou o melhor desempenho, selecionando, respectivamente, 41% e 55% das vezes o modelo correto. O 632QCV alcançou bom desempenho apenas para $n = 15$, perdendo desempenho frente aos demais à medida que o tamanho amostral cresce. Para $n = 30$ o critério AICc teve o melhor desempenho. Já para $n = 40$ a extensão $\text{EIC}_{2,p}$ foi o critério que apresentou a maior frequência de seleção do modelo verdadeiro. Percebemos ainda que os EIC_i 's que utilizam bootstrap paramétrico apresentaram desempenhos melhores do que as extensões bootstrap do AIC que utilizam bootstrap não-paramétrico. Pelos resultados dos critérios BCV e 632CV, na Tabela 3.1, notamos o fraco desempenho dos mesmos na seleção do modelo verdadeiro. Contudo, percebe-se uma superioridade significativa do 632CV frente ao BCV. Dentre os critérios que não utilizam correções bootstrap, destaca-se o desempenho do AICc. Notamos que o AICc apresentou desempenho razoável em quase todos os tamanhos amostrais considerados na Tabela 3.1.

Tabela 3.1 Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão constante, considerando grande dispersão e modelo facilmente identificável (modelo (3.8)).

	$n = 15$			$n = 20$			$n = 30$			$n = 40$		
	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$
AIC	277	284	439	209	433	358	175	501	324	110	603	287
AICc	630	318	52	384	518	98	264	596	140	163	653	184
SIC	412	309	279	353	475	172	354	543	103	315	601	84
SICc	780	208	12	594	370	36	496	470	34	405	559	36
HQ	276	284	440	232	454	314	237	543	220	182	617	201
HQc	627	320	53	430	492	78	332	580	88	256	644	100
BQCV	424	409	167	212	548	240	115	524	361	50	552	398
632QCV	246	406	348	122	452	426	47	370	583	20	358	622
EIC1 _p	626	308	66	407	494	99	280	550	170	179	622	199
EIC2 _p	759	223	18	501	442	57	317	579	104	198	663	139
EIC3 _p	414	364	222	268	510	222	209	546	245	127	632	241
EIC4 _p	754	212	34	504	420	76	352	494	154	225	579	196
EIC5 _p	423	325	252	291	437	272	241	448	311	147	507	346
EIC1 _{np}	842	155	3	523	435	42	351	561	88	221	631	148
EIC2 _{np}	936	63	1	685	302	13	430	525	45	265	654	81
EIC3 _{np}	505	361	134	309	498	193	227	539	234	136	618	246
EIC4 _{np}	926	73	1	650	335	15	453	490	57	274	598	128
EIC5 _{np}	508	368	124	325	463	212	235	476	289	154	523	323
BCV	969	31	0	806	187	7	629	349	22	460	503	37
632CV	912	88	0	629	352	19	428	515	57	254	641	105

A Tabela 3.2 apresenta os resultados numéricos para o modelo fracamente identificável e com grande dispersão. Este cenário é o mais crítico e talvez o mais encontrado em aplicações a dados reais em pequenas amostras. Neste cenário podemos perceber o ótimo desempenho dos critérios propostos. O BQCV apresentou as maiores frequências de seleção do modelo verdadeiro em quase todos os tamanhos amostrais considerados, perdendo somente para o 632QCV quando $n = 15$ e empatando quando $n = 20$. Além do BQCV e do 632QCV, também podemos destacar o bom desempenho dos critérios EIC3 e EIC5 tanto com bootstrap paramétrico quanto não-paramétrico. Assim como os resultados da Tabela 3.1 podemos perceber o fraco desempenho do BCV e a superioridade das extensões do AIC que utilizam o bootstrap paramétrico. Dentre os critérios que não utilizam reamostragem de bootstrap o AIC foi o que obteve destaque.

Os resultados numéricos correspondentes ao terceiro cenário, em que o modelo é facilmente identificável com pequena dispersão, estão apresentados na Tabela 3.3. Nessa situação destaca-se o desempenho do SICc, que apresentou o melhor desempenho em praticamente todos os tamanhos amostrais. Verificamos que o BQCV não alcançou os bons desempenhos

Tabela 3.2 Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão constante, considerando grande dispersão e modelo fracamente identificável (modelo (3.9)).

	$n = 15$			$n = 20$			$n = 30$			$n = 40$		
	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$
AIC	559	111	330	540	127	333	624	154	222	608	200	192
AICc	903	70	27	830	99	71	765	139	96	708	183	109
SIC	716	99	185	785	94	121	838	98	64	850	122	28
SICc	953	40	7	942	45	13	924	61	15	919	72	9
HQ	557	110	333	608	125	267	717	137	146	722	170	108
HQc	903	69	28	864	87	49	832	108	60	801	157	42
BQCV	750	146	104	604	175	221	514	202	284	435	238	327
632QCV	521	181	298	378	175	447	294	176	530	235	184	581
EIC1 _p	886	86	28	831	96	73	737	159	104	686	200	114
EIC2 _p	940	50	10	886	74	40	802	139	59	745	178	77
EIC3 _p	745	126	129	691	130	179	687	150	163	642	208	150
EIC4 _p	929	52	19	875	86	39	766	160	74	699	206	95
EIC5 _p	684	132	184	686	126	188	633	178	189	589	228	183
EIC1 _{np}	969	30	1	903	80	17	790	152	58	721	205	74
EIC2 _{np}	991	9	0	953	43	4	864	113	23	785	173	42
EIC3 _{np}	815	106	79	727	140	133	689	158	153	633	209	158
EIC4 _{np}	988	12	0	942	54	4	827	138	35	749	197	54
EIC5 _{np}	793	128	79	679	161	160	636	191	173	582	234	184
BCV	998	2	0	980	19	1	935	59	6	880	112	8
632CV	988	12	0	939	57	4	841	124	35	763	195	42

evidenciados nos cenários anteriores. Neste cenário, contrariamente aos anteriores, as extensões bootstrap do AIC que utilizam bootstrap não-paramétrico apresentaram, em geral, os melhores desempenhos. Outra mudança em relação aos resultados das Tabelas 3.1 e 3.2 ocorreu no desempenho do EIC2. Percebemos que tanto o EIC2_p quanto o EIC2_{np} apresentaram melhores desempenhos do que as demais extensões bootstrap do AIC. Outro fato a se destacar é a melhora no desempenho do BCV e do 632CV neste cenário. Principalmente nos tamanhos amostrais $n = 20, 30, 40$, o BCV e o 632CV apresentaram desempenhos equiparáveis com os demais. De forma geral, podemos dizer que este cenário, com resultados apresentados na Tabela 3.3, é o cenário mais favorável, com pequena dispersão e com o modelo facilmente identificável. Notamos que os resultados neste cenário diferem visivelmente dos demais apresentados.

Os resultados apresentados na Tabela 3.4 se assemelham aos contidos nas Tabelas 3.1 e 3.2. A Tabela 3.4 apresenta os desempenhos dos critérios de seleção quando o modelo é fracamente identificável e com pequena dispersão. Notamos, mais uma vez, o bom desempenho do BQCV, evidenciado pelos resultados correspondentes aos tamanhos amostrais $n = 15$ e $n = 20$.

Tabela 3.3 Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão constante, considerando pequena dispersão e modelo facilmente identificável (modelo (3.10)).

	$n = 15$			$n = 20$			$n = 30$			$n = 40$		
	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$
AIC	1	506	493	0	572	428	0	667	333	0	683	317
AICc	6	907	87	2	865	133	0	846	154	0	803	197
SIC	2	650	348	2	764	234	0	888	112	0	907	93
SICc	31	929	40	2	936	62	0	954	46	0	958	42
HQ	0	506	494	1	608	391	0	761	239	0	793	207
HQc	6	907	87	2	886	112	0	905	95	0	895	105
BQCV	1	774	225	0	682	318	0	609	391	0	558	442
632QCV	0	607	393	0	472	528	0	388	612	0	331	669
EIC1 _p	15	855	130	1	834	165	0	796	204	0	748	252
EIC2 _p	36	903	61	3	895	102	0	867	133	0	832	168
EIC3 _p	2	744	254	0	701	299	0	727	273	0	740	260
EIC4 _p	52	745	203	2	866	132	0	778	222	0	720	280
EIC5 _p	3	643	354	0	653	347	0	638	362	0	612	388
EIC1 _{np}	156	824	20	12	913	75	0	854	146	0	804	196
EIC2 _{np}	386	613	1	33	938	29	0	929	71	0	908	92
EIC3 _{np}	8	792	200	2	727	271	0	747	253	0	716	284
EIC4 _{np}	398	600	2	53	910	37	3	887	110	0	815	185
EIC5 _{np}	23	776	201	2	701	297	0	653	347	0	578	422
BCV	587	413	0	137	843	20	8	935	57	1	924	75
632CV	296	699	5	30	931	39	1	908	91	0	867	133

Já para os maiores tamanhos amostrais considerados, $n = 30$ e $n = 40$, o critério que apresentou o melhor desempenho foi o AICc. Dentre as extensões bootstrap do AIC destacamos os desempenhos do EIC3_p e do EIC3_{np}. Neste cenário os desempenhos do BCV e do 632CV voltam a ser pobres na seleção do modelo correto. Também percebemos que as extensões bootstrap do AIC que utilizam bootstrap paramétrico apresentaram superioridade de desempenho perante as extensões que utilizam o bootstrap não-paramétrico.

De maneira geral, analisando as Tabelas 3.1, 3.2, 3.3 e 3.4, podemos verificar alguns padrões nos desempenhos das variações bootstrap do AIC na seleção de modelos de regressão beta com dispersão constante. Primeiramente, verificamos o bom desempenho do critério BQCV proposto. Constatamos numericamente que o BQCV possui o melhor desempenho em modelos fracamente identificáveis e também em modelos com grande dispersão. Já o 632QCV mostrou bom desempenho quando $n = 15$. Também verificamos a superioridade das extensões bootstrap do AIC que utilizam bootstrap paramétrico frente às extensões que usam bootstrap não-paramétrico. Dentre as extensões bootstrap podemos destacar o desempenho do EIC3, principalmente a versão com bootstrap paramétrico. Os resultados mostraram semelhanças

Tabela 3.4 Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão constante, considerando pequena dispersão e modelo fracamente identificável (modelo (3.11)).

	$n = 15$			$n = 20$			$n = 30$			$n = 40$		
	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$
AIC	371	242	387	310	341	349	281	450	269	239	455	306
AICc	693	264	43	538	361	101	392	481	127	304	529	167
SIC	502	265	233	492	338	170	487	418	95	450	471	79
SICc	783	198	19	720	258	22	626	345	29	553	401	46
HQ	370	241	389	348	346	306	350	459	191	314	508	178
HQc	689	267	44	574	349	77	478	440	82	397	508	95
BQCV	523	339	138	325	405	270	213	452	335	149	401	450
632QCV	362	327	331	183	335	482	108	327	565	68	258	674
EIC1 _p	679	271	50	510	367	123	396	455	149	285	487	228
EIC2 _p	765	216	19	604	337	59	452	451	97	344	516	140
EIC3 _p	517	313	170	400	356	244	325	473	202	273	487	240
EIC4 _p	727	217	56	587	336	77	437	436	127	297	473	230
EIC5 _p	462	297	241	353	375	272	297	445	258	232	434	334
EIC1 _{np}	838	159	3	656	312	32	463	440	97	330	507	163
EIC2 _{np}	917	82	1	769	226	5	555	398	47	411	502	87
EIC3 _{np}	567	326	107	431	380	189	336	472	192	265	481	254
EIC4 _{np}	910	89	1	757	233	10	519	419	62	380	481	139
EIC5 _{np}	574	321	105	429	373	198	324	410	266	233	426	341
BCV	957	43	0	853	146	1	678	307	15	549	405	46
632CV	899	100	1	737	253	10	533	405	62	388	497	115

de desempenho entre os critérios EIC3 e EIC5, assim como entre os critérios EIC2 e EIC4. Essas semelhanças são verificadas pelas similaridades dos mesmos nas frequências de seleção do modelo correto nos diversos cenários investigados. Também verificamos, pelos resultados numéricos, o desempenho pobre do BCV na seleção de modelos, assim como a superioridade do 632CV frente ao mesmo.

Com isso, concluímos que as variações bootstrap do AIC em regressão beta com dispersão constante são boas alternativas para a seleção de modelos em pequenas amostras. Pode-se verificar facilmente que o uso de variações bootstrap do AIC diminui o problema da sobre-especificação na seleção de modelos. Dentre as diversas variações do AIC investigadas destacam-se o critério BQCV proposto e o EIC3. Isso é verificado pela superioridade de seus desempenhos relativamente ao AIC.

3.4.2 Critérios de seleção de modelos em regressão beta com dispersão variável

O estudo de simulação para o desempenho de critérios de seleção em modelos de regressão beta com dispersão variável foi conduzido considerando os quatro modelos a seguir:

$$\text{logit}(\mu_t) = -1.5 + x_{t2} + x_{t3}, \quad \text{logit}(\sigma_t) = -0.7 - 0.6x_{t2} - 0.6x_{t3}, \quad (3.12)$$

$$\text{logit}(\mu_t) = 1 - 0.75x_{t2} - 0.25x_{t3}, \quad \text{logit}(\sigma_t) = -0.7 - 0.5x_{t2} - 0.3x_{t3}, \quad (3.13)$$

$$\text{logit}(\mu_t) = -1.5 + x_{t2} + x_{t3}, \quad \text{logit}(\sigma_t) = -1.1 - 1.1x_{t2} - 1.1x_{t3}, \quad (3.14)$$

$$\text{logit}(\mu_t) = 1 - 0.75x_{t2} - 0.25x_{t3}, \quad \text{logit}(\sigma_t) = -1.45 - 1x_{t2} - 0.5x_{t3}. \quad (3.15)$$

Os modelos/cenários possuem características elencadas a seguir. Os dois primeiros modelos, equações (3.12) e (3.13), consideram cenários com grande dispersão, enquanto que os dois últimos, (3.14) e (3.15), possuem pequena dispersão. Considerando os valores dos parâmetros, notamos que os modelos apresentados em (3.12) e (3.14) são facilmente identificáveis e os modelos (3.13) e (3.15) são fracamente identificáveis. Contudo, os resultados numéricos evidenciam similaridades nos desempenhos dos critérios de seleção entre modelos com grande dispersão e pequena dispersão. Por esse motivo, apresentaremos nesta seção somente os resultados dos modelos com pequena dispersão, (3.14) e (3.15). Os resultados para os modelos com grande dispersão, dados nas equações (3.12) e (3.13), encontram-se nas tabelas do Apêndice H.

Para os modelos considerados, percebemos que a dimensão dos modelos verdadeiros é $k_0 = 6$, sendo três parâmetros no submodelo da média e três na estrutura de regressão para a dispersão. Consideramos tamanhos amostrais $n = 25, 30, 40, 50$ e a inclusão de cinco covariáveis candidatas, tanto para o submodelo da média quanto para o submodelo da dispersão.

Os modelos candidatos são sequencialmente aninhados para o submodelo da média, ou seja, o modelo candidato com r parâmetros na estrutura de regressão da média consiste no submodelo com os $1, 2, \dots, r$ primeiros parâmetros. Dentro de cada submodelo da média candidato, também são sequencialmente aninhados os submodelos da dispersão. Com isso, temos que para cada valor de r variamos s de 1 até 6, totalizando $6 \times 6 = 36$ modelos candidatos.

Uma vez que o modelo verdadeiro pertence ao conjunto de modelos candidatos, a avaliação dos critérios de seleção é feita pelo número de vezes em que cada critério seleciona a ordem do modelo verdadeiro (k_0, r_0 ou s_0) nas 1000 réplicas de Monte Carlo. Essa avaliação se deu em três abordagens diferenciadas nesta seção. Primeiramente, utilizamos os critérios de seleção para escolher os regressores de μ_t e σ_t conjuntamente, conforme resultados nas Tabelas 3.5 e 3.6, para cada modelo. Posteriormente, consideramos o submodelo da dispersão corretamente especificado e utilizamos os critérios para selecionar os regressores de μ_t , com resultados expostos nas Tabelas 3.7 e 3.8. Na terceira e última abordagem, cujos resultados se encontram nas Tabelas 3.9 e 3.10, o submodelo para μ_t foi corretamente especificado e utilizamos os critérios para selecionar o submodelo da dispersão. As tabelas de resultados apresentam as frequências de escolha da verdadeira ordem do modelo, assim como as frequências de escolha de modelos subespecificados e sobre-especificados, com os melhores resultados destacados.

A Tabela 3.5 evidencia os bons resultados dos critérios propostos na seleção conjunta dos regressores da média e dispersão em modelos facilmente identificáveis. Percebemos que para os tamanhos amostrais $n = 25$ e $n = 30$ 632QCV foi o critério com melhor desempenho na seleção da ordem do verdadeiro modelo. Já para $n = 40$ e $n = 50$ o critério BQCV apresentou

Tabela 3.5 Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, selecionando regressores da média e dispersão conjuntamente em um modelo facilmente identificável (modelo (3.14)).

	$n = 25$			$n = 30$			$n = 40$			$n = 50$		
	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$
AIC	195	100	705	229	169	602	181	273	546	156	345	499
AICc	618	167	215	532	233	235	329	373	298	242	464	294
SIC	476	122	402	557	190	253	518	312	170	439	423	138
SICc	883	72	45	864	99	37	718	237	45	607	337	56
HQ	274	121	605	325	191	484	288	333	379	253	436	311
HQc	734	128	138	671	192	137	507	349	144	386	457	157
BQCV	861	107	32	640	267	93	309	466	225	187	506	307
632QCV	678	234	88	387	371	242	151	420	429	80	362	558
EIC1 _p	964	28	8	950	28	22	893	77	30	886	73	41
EIC2 _p	980	19	1	920	63	17	722	249	29	515	419	66
EIC3 _p	856	129	15	521	368	111	267	422	311	203	429	368
EIC4 _p	215	6	779	314	0	686	424	7	569	704	11	285
EIC5 _p	93	9	898	97	11	892	505	53	442	821	103	76
EIC1 _{np}	991	9	0	955	36	9	799	183	18	486	425	89
EIC2 _{np}	997	3	0	985	11	4	921	76	3	675	300	25
EIC3 _{np}	463	133	404	438	273	289	275	399	326	174	433	393
EIC4 _{np}	998	2	0	981	15	4	894	99	7	674	293	33
EIC5 _{np}	281	78	641	379	243	378	229	355	416	151	365	484
BCV	999	1	0	993	6	1	948	50	2	795	193	12
632CV	997	3	0	978	17	5	890	104	6	649	308	43

melhores resultados. Dentre as extensões (EIC's) do AIC o critério que se destaca é o EIC3 em suas duas versões, tanto com bootstrap paramétrico quanto com bootstrap não-paramétrico. Neste cenário, o AICc se sobressai quando comparado com os demais critérios que não utilizam a log-verossimilhança bootstrap. Esses resultados, apresentados na Tabela 3.5, também chamam a atenção para o desempenho pobre dos critérios BCV, 632CV e das extensões do AIC diferentes do EIC3. Também cabe ressaltar que quando o tamanho amostral aumenta os desempenhos dos EIC's não-paramétricos melhoram, tornando-se semelhantes. Por outro lado, esse mesmo comportamento não é verificado para os EIC's paramétricos, nos quais o EIC1_p e o EIC4_p apresentam baixa performance em todos os tamanhos amostrais. Por fim, verifica-se claramente que os critérios propostos, assim como algumas extensões do AIC, atenuam o problema de sobre-especificação do AIC.

Para o modelo com fraca identificabilidade se tornam ainda mais evidentes os bons desempenhos do critério BQCV e principalmente de 632QCV, verificados na Tabela 3.6. O critério 632QCV foi o de melhor desempenho para $n = 25, 30, 40$. Já para $n = 50$ BQCV foi o critério com a melhor performance de seleção da ordem do modelo verdadeiro. Percebe-se que

Tabela 3.6 Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, selecionando regressores da média e dispersão conjuntamente em um modelo fracamente identificável (modelo (3.15)).

	$n = 25$			$n = 30$			$n = 40$			$n = 50$		
	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$
AIC	317	69	614	439	100	461	517	136	347	484	157	359
AICc	778	75	147	748	118	134	736	112	152	676	150	174
SIC	662	56	282	778	78	144	861	65	74	889	67	44
SICc	963	18	19	957	33	10	966	25	9	957	30	13
HQ	424	71	505	572	101	327	685	105	210	694	133	173
HQc	867	58	75	856	86	58	867	74	59	849	95	56
BQCV	866	82	52	791	145	64	699	161	140	544	229	227
632QCV	726	158	116	573	237	190	452	218	330	313	225	462
EIC1 _p	960	30	10	931	50	19	901	61	38	827	113	60
EIC2 _p	984	14	2	966	26	8	949	33	18	887	80	33
EIC3 _p	885	88	27	721	213	66	657	154	189	587	180	233
EIC4 _p	326	7	667	484	8	508	641	15	344	804	44	152
EIC5 _p	11	0	989	29	0	971	228	9	763	630	128	242
EIC1 _{np}	994	6	0	977	17	6	951	42	7	856	117	27
EIC2 _{np}	1000	0	0	995	5	0	977	23	0	930	62	8
EIC3 _{np}	593	91	316	672	160	168	636	172	192	582	169	249
EIC4 _{np}	999	1	0	994	4	2	968	32	0	912	76	12
EIC5 _{np}	362	34	604	588	148	264	572	208	220	496	204	300
BCV	1000	0	0	1000	0	0	991	9	0	969	28	3
632CV	999	1	0	994	6	0	975	25	0	911	76	13

para $n = 25, 30$ o desempenho do 632QCV é 200% superior a qualquer critério de seleção que não utiliza bootstrap. Para a seleção conjunta dos regressores da média e da dispersão em modelos fracamente identificáveis mais uma vez percebe-se o bom desempenho do EIC3 frente as demais extensões bootstrap. Evidencia-se também o fraco desempenho do BCV e do 632CV. Percebemos que dentre os critérios que não utilizam log-verossimilhança bootstrap as versões corrigidas melhoram o problema da sobre-especificação de suas versões não corrigidas. Para $n = 25$, por exemplo, percebemos que houve sobre-especificação em 614 replicações utilizando o AIC na seleção de modelos, baixando para 147 quando se utiliza o AICc.

Agora analisemos os desempenhos dos critérios na seleção dos regressores do submodelo da média. Essa segunda abordagem considera o submodelo da dispersão corretamente especificado, em que os critérios selecionam apenas o submodelo da média, conforme resultados das Tabelas 3.7 e 3.8. A Tabela 3.7 apresenta os resultados dessa abordagem quando o modelo é facilmente identificável. A partir destes resultados, verificamos novamente o bom desempenho dos critérios propostos. Percebemos que para $n = 25, 30$ 632QCV possui grande superioridade de desempenho quando comparado com os demais critérios estudados. Já para $n = 40$ BQCV

Tabela 3.7 Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, considerando conhecida a estrutura de regressão da dispersão e selecionando regressores da média em um modelo facilmente identificável (modelo (3.14)).

	$n = 25$			$n = 30$			$n = 40$			$n = 50$		
	$< r_0$	$= r_0$	$> r_0$	$< r_0$	$= r_0$	$> r_0$	$< r_0$	$= r_0$	$> r_0$	$< r_0$	$= r_0$	$> r_0$
AIC	120	360	520	98	426	476	70	531	399	44	617	339
AICc	326	519	155	209	589	202	114	654	232	68	739	193
SIC	278	461	261	233	538	229	192	652	156	144	754	102
SICc	559	390	51	459	486	55	335	613	52	239	714	47
HQ	160	402	438	136	496	368	105	607	288	75	722	203
HQc	383	501	116	291	579	130	190	677	133	121	769	110
BQCV	487	510	3	279	699	22	115	776	109	56	801	143
632QCV	316	668	16	168	779	53	65	705	230	27	673	300
EIC1 _p	932	68	0	904	94	2	869	110	21	827	157	16
EIC2 _p	790	210	0	560	438	2	297	680	23	147	823	30
EIC3 _p	543	456	1	279	694	27	112	705	183	58	737	205
EIC4 _p	404	4	592	408	5	587	818	8	174	959	17	24
EIC5 _p	313	2	685	722	4	274	968	9	23	968	13	19
EIC1 _{np}	970	30	0	927	72	1	603	384	13	300	634	66
EIC2 _{np}	974	26	0	958	41	1	725	268	7	471	495	34
EIC3 _{np}	325	502	173	222	624	154	126	677	197	76	722	202
EIC4 _{np}	974	26	0	956	43	1	723	269	8	463	497	40
EIC5 _{np}	324	426	250	248	558	194	162	584	254	83	616	301
BCV	976	24	0	965	35	0	783	210	7	582	396	22
632CV	975	25	0	953	47	0	689	302	9	419	540	41

se mostrou o melhor critério e para $n = 50$ EIC2_p foi o critério que se sobressaiu. Novamente, dentre as extensões do AIC o critério EIC3 se destaca e, por outro lado, BCV e 632CV apresentam desempenhos pobres na seleção de modelos. Ainda analisando a Tabela 3.7, notamos que os critérios AICc e HQc se destacam dentre aqueles que não utilizam a log-verossimilhança bootstrap.

Na Tabela 3.8 estão apresentadas as frequências de seleção correta do submodelo da média quando o modelo é fracamente identificável. Os critérios que se destacam são os mesmos da situação anterior. Para $n = 25, 30, 40$ 632QCV foi o que alcançou melhor desempenho, enquanto que para $n = 50$ BQCV se destacou. Nesta Tabela 3.8, assim como na Tabela 3.6, fica evidente o problema de sobre-especificação do EIC5_p para os menores tamanhos amostrais. Nota-se que esse critério não cumpre o papel das demais correções bootstrap do AIC, que é o de melhorar a acurácia na seleção de modelos diminuindo o problema da sobre-especificação do AIC.

Consideramos agora a terceira e última abordagem na seleção de modelos em regressão

Tabela 3.8 Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, considerando conhecida a estrutura de regressão da dispersão e selecionando regressores da média em um modelo fracamente identificável (modelo (3.15)).

	$n = 25$			$n = 30$			$n = 40$			$n = 50$		
	$< r_0$	$= r_0$	$> r_0$	$< r_0$	$= r_0$	$> r_0$	$< r_0$	$= r_0$	$> r_0$	$< r_0$	$= r_0$	$> r_0$
AIC	369	173	458	421	202	377	437	256	307	453	269	278
AICc	711	171	118	702	199	99	614	250	136	598	259	143
SIC	626	153	221	729	164	107	722	196	82	772	180	48
SICc	892	86	22	893	93	14	850	131	19	851	135	14
HQ	449	173	378	535	204	261	572	246	182	611	246	143
HQc	783	145	72	791	152	57	724	204	72	727	211	62
BQCV	846	153	1	779	211	10	600	330	70	546	325	129
632QCV	743	247	10	651	307	42	451	376	173	378	323	299
EIC1 _p	941	58	1	908	90	2	836	139	25	796	168	36
EIC2 _p	958	42	0	918	82	0	808	182	10	765	221	14
EIC3 _p	856	142	2	737	243	20	582	297	121	551	288	161
EIC4 _p	583	17	400	625	21	354	848	60	92	899	75	26
EIC5 _p	52	1	947	238	4	758	764	89	147	796	101	103
EIC1 _{np}	982	17	1	985	15	0	905	93	2	827	166	7
EIC2 _{np}	983	16	1	988	12	0	945	54	1	875	123	2
EIC3 _{np}	696	175	129	717	218	65	618	278	104	576	290	134
EIC4 _{np}	983	16	1	990	10	0	940	59	1	873	124	3
EIC5 _{np}	652	149	199	690	219	91	579	300	121	549	292	159
BCV	985	14	1	992	8	0	958	41	1	907	91	2
632CV	985	14	1	989	11	0	934	65	1	863	133	4

beta com dispersão variável. A estrutura de regressão da média é assumida conhecida e investigamos os desempenhos dos critérios de seleção somente na escolha do submodelo da dispersão. Os resultados apresentados na Tabela 3.9 confirmam a qualidade do critério 632QCV na seleção dos regressores da dispersão quando o modelo é facilmente identificável. Percebe-se que, neste cenário, o critério proposto apresentou os melhores resultados em todos os tamanhos amostrais considerados. É importante salientar que, dentre as variações bootstrap consideradas, 632QCV foi o único critério que forneceu resultados melhores do que os critérios que não utilizam log-verossimilhança bootstrap para o menor tamanho amostral ($n = 25$). Percebe-se ainda que para os demais tamanhos amostrais somente BQCV e EIC3_p foram superiores em desempenho aos critérios que não consideram a log-verossimilhança bootstrap.

A Tabela 3.10 apresenta os resultados da comparação de desempenho dos critérios de informação na seleção dos submodelos da dispersão, considerando o submodelo da média corretamente especificado, quando o modelo é fracamente identificável. Percebemos que esse foi o único cenário considerado em que 632QCV não apresentou os melhores desempenhos

Tabela 3.9 Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, considerando conhecida a estrutura de regressão da média e selecionando regressores da dispersão em um modelo facilmente identificável (modelo (3.14)).

	$n = 25$			$n = 30$			$n = 40$			$n = 50$		
	$< s_0$	$= s_0$	$> s_0$	$< s_0$	$= s_0$	$> s_0$	$< s_0$	$= s_0$	$> s_0$	$< s_0$	$= s_0$	$> s_0$
AIC	328	230	442	295	313	392	254	401	345	246	466	288
AICc	632	252	116	523	342	135	396	452	152	339	499	162
SIC	567	221	212	553	303	144	519	394	87	535	409	56
SICc	838	143	19	785	183	32	679	297	24	650	331	19
HQ	398	247	355	378	332	290	361	430	209	357	480	163
HQc	705	223	72	620	301	79	508	421	71	467	465	68
BQCV	882	118	0	783	216	1	471	493	36	348	571	81
632QCV	734	266	0	574	413	13	304	582	114	214	572	214
EIC1 _p	974	26	0	941	59	0	810	174	16	716	227	57
EIC2 _p	994	6	0	969	31	0	800	196	4	664	327	9
EIC3 _p	842	158	0	625	371	4	348	511	141	299	517	184
EIC4 _p	617	21	362	622	15	363	701	55	244	809	116	75
EIC5 _p	155	11	834	313	37	650	541	140	319	535	174	291
EIC1 _{np}	1000	0	0	995	5	0	870	127	3	736	246	18
EIC2 _{np}	1000	0	0	999	1	0	952	47	1	883	114	3
EIC3 _{np}	564	236	200	509	355	136	358	468	174	290	480	230
EIC4 _{np}	1000	0	0	1000	0	0	947	52	1	872	126	2
EIC5 _{np}	533	183	284	514	303	183	382	388	230	318	388	294
BCV	1000	0	0	1000	0	0	962	38	0	922	76	2
632CV	1000	0	0	998	2	0	935	62	3	852	145	3

para os tamanhos amostrais $n = 25, 30$, tendo contudo desempenhos razoáveis. Para tamanhos amostrais maiores, $n = 40, 50$, o critério proposto foi o que teve os melhores resultados. Para $n = 25$ HQ foi o critério de melhor performance na seleção da ordem do modelo verdadeiro e para $n = 50$ EIC3_p se destacou.

Como uma análise geral dos desempenhos dos critérios de seleção para pequenas amostras em modelos de regressão beta com dispersão variável podemos fazer os seguintes destaques:

- Os critérios propostos são ótimas alternativas para seleção em pequenas amostras, conduzindo à seleção do verdadeiro modelo de forma superior aos demais critérios investigados. Em especial, 632QCV apresentou melhor desempenho em amostras menores e BQCV melhora em desempenho quando o tamanho amostral aumenta.
- Dentre os critérios que não utilizam a log-verossimilhança bootstrap o AICc e o HQc conduziram aos melhores resultados. O critério AICc se destacou nos menores tamanhos amostrais considerados, enquanto que HQc tem destaque nas amostras maiores.

Tabela 3.10 Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, considerando conhecida a estrutura de regressão da média e selecionando regressores da dispersão em um modelo fracamente identificável (modelo (3.15)).

	$n = 25$			$n = 30$			$n = 40$			$n = 50$		
	$< s_0$	$= s_0$	$> s_0$	$< s_0$	$= s_0$	$> s_0$	$< s_0$	$= s_0$	$> s_0$	$< s_0$	$= s_0$	$> s_0$
AIC	455	110	435	487	156	357	520	201	279	531	214	255
AICc	812	98	90	762	135	103	689	182	129	669	206	125
SIC	738	89	173	782	109	109	807	122	71	818	147	35
SICc	947	38	15	939	50	11	913	69	18	900	87	13
HQ	537	114	349	606	152	242	658	186	156	685	192	123
HQc	876	70	54	848	104	48	809	131	60	787	164	49
BQCV	975	25	0	928	70	2	800	175	25	690	243	67
632QCV	911	88	1	832	161	7	619	290	91	530	293	177
EIC1 _p	991	9	0	967	33	0	900	93	7	833	139	28
EIC2 _p	999	1	0	997	3	0	950	49	1	911	83	6
EIC3 _p	943	57	0	812	180	8	665	237	98	610	239	151
EIC4 _p	729	18	253	778	6	216	829	26	145	900	78	22
EIC5 _p	45	2	953	181	16	803	559	147	294	566	188	246
EIC1 _{np}	1000	0	0	1000	0	0	967	33	0	892	101	7
EIC2 _{np}	1000	0	0	1000	0	0	985	15	0	950	47	3
EIC3 _{np}	723	110	167	742	151	107	636	224	140	606	228	166
EIC4 _{np}	1000	0	0	1000	0	0	987	13	0	945	51	4
EIC5 _{np}	643	88	269	711	132	157	595	231	174	546	262	192
BCV	1000	0	0	1000	0	0	994	6	0	971	27	2
632CV	1000	0	0	1000	0	0	980	20	0	933	62	5

- Dentre as extensões do AIC (EIC's) EIC3 se destacou. Sua versão com bootstrap não-paramétrico (EIC3_{np}) apresentou os melhores desempenhos nas amostras menores e EIC3_p alcançou resultados favoráveis nos maiores tamanhos amostrais.
- Percebemos ainda que as performances dos critérios de informação na seleção do submodelo da média são bastante superiores aos seus desempenhos na seleção do submodelo da dispersão. Esse fato pode ser verificado comparando a Tabela 3.7 com a Tabela 3.9, assim como as Tabelas 3.8 e 3.10.
- Fica clara a superioridade dos critérios baseados na log-verossimilhança bootstrap para seleção de modelos de regressão beta em pequenas amostras. O estudo numérico investigou diversos cenários, abordagens e tamanhos amostrais, tornando os resultados bem gerais, validando principalmente a aplicabilidade dos critérios propostos na seleção de modelos com dispersão variável.

3.5 Aplicação

Para a aplicação a dados reais consideramos os dados de Griffiths *et al.* (1993, Tabela 15.4) referentes aos gastos com alimentação, renda e número de pessoas em 38 domicílios de uma grande cidade dos Estados Unidos da América. Estes dados de gastos com alimentação foram modelados utilizando regressão beta com dispersão constante por Ferrari e Cribari-Neto (2004). Para a seleção do modelo utilizamos o esquema de seleção em duas etapas, proposto no Capítulo 2 desta tese, juntamente como os critérios BQCV e 632QCV propostos neste capítulo. Nesse esquema, primeiramente considera-se fixa a dispersão e seleciona-se o submodelo da média; posteriormente, fixando-se o submodelo da média selecionado na etapa anterior, seleciona-se o submodelo da dispersão. Conforme mostrado no Capítulo 2, esse esquema de seleção tende a ter um desempenho igual ou superior à seleção conjunta dos submodelos da média e da dispersão, possuindo um custo computacional bastante inferior.

Ferrari e Cribari-Neto (2004) modelam a proporção de gastos com alimentação (y) como função da renda (x_2) e do número de pessoas (x_3) de cada residência. Nesta aplicação, consideramos a função de ligação logit para os submodelos da média e da dispersão. Incluímos ainda como possíveis covariáveis, tanto para média quanto para a dispersão, a interação entre renda e número de pessoas ($x_4 = x_2 \times x_3$), o quadrado da variável renda ($x_5 = x_2^2$) e uma transformação quadrática na variável x_3 , ou seja, $x_6 = x_3^2$.

Supondo dispersão constante, o submodelo da média selecionado, tanto pelo BQCV quanto pelo 632QCV, considera as covariáveis x_3 e x_4 . Assumindo que esse é o submodelo correto para média selecionamos agora o submodelo da dispersão. O submodelo da dispersão selecionado pelos critérios BQCV e 632QCV inclui apenas a covariável x_3 . As estimativas dos parâmetros do modelo de regressão beta com dispersão variável selecionado para os dados de gastos com alimentação estão apresentadas na Tabela 3.11.

Tabela 3.11 Estimativas dos parâmetros do modelo de regressão beta com dispersão variável para dados de gastos com alimentação.

Parâmetros	Estimativa	Erro padrão	Estat. z	Valor p
Submodelo de μ				
β_1 (Constante)	-1.3040	0.1103	-11.826	0.0000
β_3 (Nº de pessoas)	0.2890	0.0754	3.835	0.0005
β_4 (Interação)	-0.0031	0.0011	-2.975	0.0054
Submodelo de σ				
γ_1 (Constante)	-2.4825	0.3720	-6.673	0.0000
γ_3 (Nº de pessoas)	0.2011	0.1118	1.798	0.0813
$R_{FC}^2 = 0.4586$				
$R_{ML}^2 = 0.5448$				

Com o objetivo de verificar possíveis afastamentos das suposições feitas para o modelo, é útil visualizar graficamente os resíduos do modelo ajustado, conforme Figura 3.1. Amplos trabalhos a respeito de resíduos e análise de diagnóstico em regressão beta são encontrados em Espinheira *et al.* (2008a,b) e Ferrari *et al.* (2011). Dentre os diversos resíduos propostos,

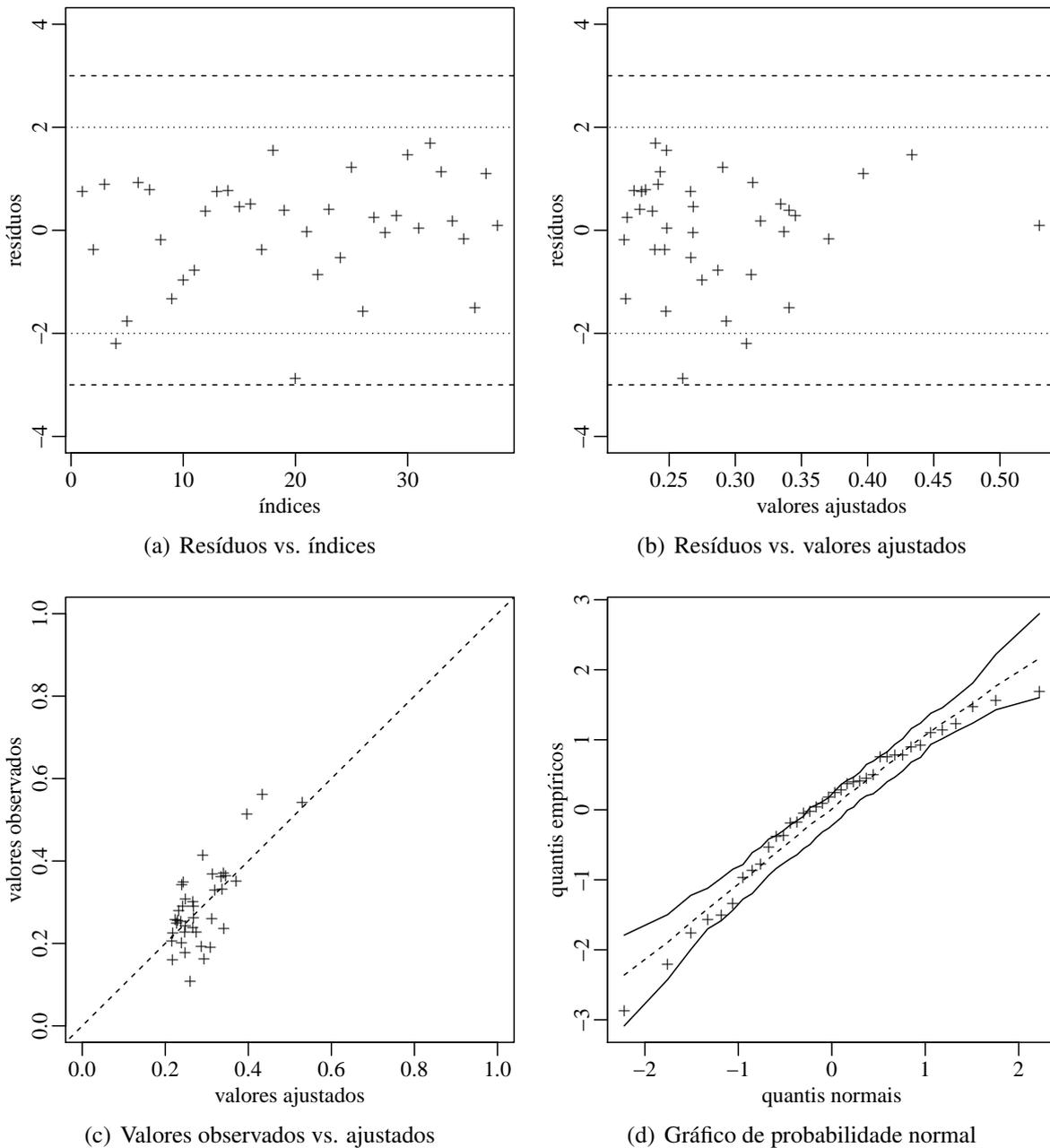


Figura 3.1 Gráficos dos resíduos para o modelo ajustado.

consideramos o *resíduo ponderado padronizado 2*, por apresentar as melhores propriedades. Este resíduo é dado por

$$r_{pp2} = \frac{y_t^* - \hat{\mu}_t^*}{\sqrt{\hat{v}_t(1 - h_{tt})}},$$

em que $v_t^* = \psi' \left(\mu_t \left(\frac{1 - \sigma_t^2}{\sigma_t^2} \right) \right) + \psi' \left((1 - \mu_t) \left(\frac{1 - \sigma_t^2}{\sigma_t^2} \right) \right)$ e h_{tt} é o t -ésimo elemento da diagonal

da matriz chapéu (para detalhes, ver Espinheira *et al.* (2008b) e Ferrari *et al.* (2011)).

Para a geração dos gráficos da Figura 3.1 foi utilizada uma implementação em \mathbb{R} (R Development Core Team, 2009), para análise de diagnóstico em regressão beta com dispersão variável, que está apresentada no Apêndice G. As Figuras 3.1(a) e 3.1(b) indicam que os resíduos estão aleatoriamente distribuídos em torno de zero. Percebemos apenas duas observações atípicas, com resíduo um pouco abaixo de -2 , mas ainda dentro do intervalo $(-3,3)$. A Figura 3.1(c) também evidencia um bom ajuste do modelo, mostrando pontos simetricamente distribuídos em torno da reta identidade. O envelope simulado, Figura 3.1(d), apresenta todos os pontos dentro dos limites, indicando que o modelo ajustado representa adequadamente os dados.

A análise de resíduos valida o modelo, mostrando que ele está bem ajustado aos dados. As estimativas dos parâmetros mostram uma relação positiva entre a média da variável resposta e o número de pessoas em cada residência, assim como uma relação negativa com a interação (x_4). O submodelo da dispersão mostra uma relação positiva entre o número de pessoas de cada residência e a dispersão da variável resposta. O bom desempenho do modelo de regressão beta com dispersão variável frente ao modelo com parâmetro de dispersão constante pode ser observado pelo valor de $R_{FC}^2 = 0.4586$, notavelmente superior ao valor de $R_{FC}^2 = 0.3878$ do modelo encontrado por Ferrari e Cribari-Neto (2004). Uma melhora ainda mais significativa é percebida em relação ao $R_{ML}^2 = 0.4088$, do modelo encontrado por Ferrari e Cribari-Neto (2004), contra o $R_{ML}^2 = 0.5448$, para o modelo da Tabela 3.11 que considera dispersão variável.

3.6 Conclusões

Neste capítulo propomos dois novos critérios para a seleção de modelos em pequenas amostras. Esses novos critérios são propostos como variações bootstrap do AIC, que fornecem estimadores diretos para a log-verossimilhança esperada. Tais critérios baseiam-se no método bootstrap e em um procedimento que chamamos de *quasi-CV* induzindo ao nome de *bootstrap likelihood quasi-CV* (BQCV) e sua modificação 632QCV.

A avaliação numérica dos desempenhos dos critérios propostos foi conduzida na classe dos modelos de regressão beta. Essa classe de modelos foi proposta por Ferrari e Cribari-Neto (2004) para a modelagem de taxas e proporções. Aqui, consideramos uma extensão do modelo que inclui uma estrutura de regressão também para o parâmetro de dispersão e não só para a média condicional. Essa modelagem da dispersão é importante para melhorar as inferências no modelo de regressão beta. Com isso, a aplicabilidade dos critérios de seleção se torna ainda mais importante, além da estrutura de regressão para a média os regressores da dispersão também devem ser selecionados. Essa abordagem aumenta o número de parâmetros a serem considerados no modelo, valorizando ainda mais o desenvolvimento e a aplicabilidade de critérios de seleção mais acurados em pequenas amostras.

Além dos critérios propostos investigamos outros critérios corrigidos para pequenas amostras. Fizemos uma ampla revisão bibliográfica, identificando diferentes variações bootstrap do AIC que vêm sendo propostas e utilizadas nas mais diversas classes de modelos. A comparação dos desempenhos dos critérios propostos com as diferentes variações bootstrap do AIC consideradas foi realizada em diversas abordagens e cenários de modelos de regressão beta

com dispersão constante e variável.

Os resultados numéricos evidenciam os bons desempenhos dos critérios propostos. BQCV e 632QCV se destacaram na maioria dos cenários considerados. BQCV apresentou os melhores resultados no modelo de regressão beta com dispersão constante, enquanto que 632QCV se destacou na seleção de modelos com dispersão variável. Os critérios baseados na log-verossimilhança bootstrap se mostraram bastante úteis na seleção de modelos em pequenas amostras, atenuando o problema da sobre-especificação do AIC em amostras menores.

Considerações Finais

4.1 Conclusões

Esta tese abordou aspectos de modelagem e inferências em regressão beta. Essa importante classe de modelos é utilizada para modelar, por meio de uma estrutura de regressão, variáveis que assumem valores no intervalo $(0,1)$, como taxas e proporções. Inicialmente abordamos o modelo seminal, com dispersão constante, proposto por Ferrari e Cribari-Neto (2004). Posteriormente, ampliamos o modelo com dispersão constante, considerando que, assim como a média, o parâmetro de dispersão é variável ao longo das observações e pode ser modelado por meio de uma estrutura de regressão. Nesses modelos investigamos aspectos de inferência em pequenas amostras e da utilização de critérios de seleção de modelos em amostras de tamanho finito. A abordagem se deu em três capítulos independentes e auto-contidos; as principais contribuições e conclusões de cada um deles estão listadas abaixo.

Capítulo 1: Utilizando uma notação matricial para o fator de correção de Bartlett (Cordeiro, 1993) e uma correção de Bartlett bootstrap (Rocke, 1989) propomos correções para a estatística da razão de verossimilhanças em modelos de regressão beta. A avaliação dos testes corrigidos foi realizada via simulação de Monte Carlo. Os desempenhos dos testes corrigidos propostos foram comparados com o teste da razão de verossimilhanças usual e com o teste que utiliza o ajuste de Skovgaard. Os resultados numéricos conduziram a conclusões importantes relativas ao comportamento dos testes em pequenas amostras. As distribuições nulas das estatísticas corrigidas por Bartlett são melhor aproximadas pela distribuição nula limite χ^2 , implicando testes menos distorcidos. O uso do teste da razão de verossimilhanças usual deve ser evitado em pequenas amostras, pois mostra-se bastante liberal. A correção de Bartlett derivada é de fácil implementação computacional e conduz a resultados mais confiáveis do que o teste não corrigido e também do que os testes que utilizam o ajuste de Skovgaard e a correção de Bartlett bootstrap.

Capítulo 2: Consideramos o problema da seleção de variáveis regressoras nos modelos de regressão beta com dispersão variável. Ampliamos o modelo com dispersão constante, proposto por Ferrari e Cribari-Neto (2004), considerando uma estrutura de regressão também para o parâmetro de dispersão. Nesse modelo, além de selecionar o submodelo da média é preciso selecionar as variáveis regressoras para a estrutura de regressão da dispersão. Avaliamos, por meio de um extensivo estudo simulação, os desempenhos de diversos critérios de seleção de modelos em diferentes cenários. Apresentamos critérios de seleção de modelos que consideram a modelagem da dispersão e propomos um esquema de seleção em duas etapas que envolver um menor custo computacional do que

a abordagem de seleção conjunta usual. Os resultados numéricos evidenciam que o esquema de seleção proposto conduz a melhores resultados do que a seleção conjunta e recomendamos o seu uso. Uma aplicação da estratégia proposta a dados reais de habilidade de leitura também é apresentada. A aplicação é apresentada com razoável detalhe, utilizando e recomendando algumas ferramentas de diagnóstico e medidas de qualidade.

Capítulo 3: O critério de seleção de Akaike (AIC) é um estimador do valor esperado da log-verossimilhança, sendo uma medida de discrepância entre o modelo verdadeiro e o modelo candidato estimado. No entanto, em pequenas amostras o AIC é viesado e tende a selecionar modelos com alta dimensionalidade. Para contornar esse problema propomos novos critérios de seleção para pequenas amostras. Esses novos critérios são propostos como variações bootstrap do AIC, que fornecem estimadores diretos para a log-verossimilhança esperada. Tais critérios baseiam-se no método bootstrap e em um procedimento que chamamos de *quasi-CV* induzindo ao nome de *bootstrap likelihood quasi-CV* (BQCV) e sua modificação 632QCV. Comparamos, através de simulações de Monte Carlo, os desempenhos em pequenas amostras dos critérios propostos, do AIC e de suas diversas variações bootstrap na seleção de modelos de regressão beta com dispersão constante e variável. Os resultados numéricos evidenciam os bons desempenhos dos critérios propostos. BQCV e 632QCV se destacaram na maioria dos cenários considerados. BQCV apresentou os melhores resultados no modelo de regressão beta com dispersão constante, enquanto que 632QCV se destacou na seleção de modelos com dispersão variável. Recomendamos o uso dos critérios baseados na log-verossimilhança bootstrap, pois eles se mostraram bastante úteis na seleção de modelos em pequenas amostras, atenuando o problema da sobre-especificação do AIC em amostras menores.

4.2 Trabalhos futuros

Em decorrência do desenvolvimento desta tese alguns tópicos correlatos já estão sendo desenvolvidos ou são linhas a serem desenvolvidas futuramente. Dentre esses, os principais pontos para as nossas pesquisas são:

1. Extensão do trabalho apresentado no Capítulo 1. Pretendemos obter o fator de correção de Bartlett para a estatística da razão de verossimilhanças no modelo de regressão beta não-linear e com dispersão variável;
2. Ajuste de Skovgaard para o teste da razão de verossimilhanças nos modelos beta autor-regressivos e de médias móveis (β ARMA). Esses são modelos dinâmicos propostos em Rocha e Cribari-Neto (2009) para modelar variáveis aleatórias contínuas pertencentes ao intervalo $(0,1)$ ao longo do tempo;
3. Investigar generalizações, demais melhoramentos nas inferências em pequenas amostras e utilização de critérios de seleção nos modelos β ARMA.

Cumulantes para o Fator de Correção de Bartlett

Neste apêndice apresentamos as derivadas até quarta ordem da função de log-verossimilhança (1.3) com respeito aos parâmetros desconhecidos. Apresentamos ainda os cumulantes até quarta ordem e derivadas até segunda ordem dos cumulantes necessários para a obtenção do fator de correção de Bartlett. Os cumulantes até terceira ordem são apresentados também em Ospina *et al.* (2006).

Primeiramente, definimos as seguintes quantidades:

$$\begin{aligned}\omega_i &= \psi'(\mu_i\phi) + \psi'((1-\mu_i)\phi), \\ m_i &= \psi''(\mu_i\phi) - \psi''((1-\mu_i)\phi), \\ a_i &= 3 \left(\frac{\partial}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} \right) \left(\frac{d\mu_i}{d\eta_i} \right)^2, \\ b_i &= \frac{d\mu_i}{d\eta_i} \left[\left(\frac{\partial^2}{\partial \mu_i^2} \frac{d\mu_i}{d\eta_i} \right) \frac{d\mu_i}{d\eta_i} + \left(\frac{\partial}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} \right)^2 \right], \\ c_i &= \phi [\mu_i \omega_i - \psi'((1-\mu_i)\phi)] = \phi \frac{\partial \mu_i^*}{\partial \phi}, \\ d_i &= (1-\mu_i)^2 \psi'((1-\mu_i)\phi) + \mu_i^2 \psi'(\mu_i\phi) - \psi'(\phi), \\ s_i &= (1-\mu_i)^3 \psi''((1-\mu_i)\phi) + \mu_i^3 \psi''(\mu_i\phi) - \psi''(\phi), \\ u_i &= -\phi \left[2\omega_i + \phi \frac{\partial \omega_i}{\partial \phi} \right], \\ r_i &= \left[2 \frac{\partial \mu_i^*}{\partial \phi} + \phi \frac{\partial^2 \mu_i^*}{\partial \phi^2} \right] \frac{d\mu_i}{d\eta_i},\end{aligned}$$

em que $\partial \omega_i / \partial \phi$, $\partial \mu_i^* / \partial \phi$ e $\partial^2 \mu_i^* / \partial \phi^2$ são definidos a seguir, juntamente com as derivadas de outras quantidades. Adicionalmente, temos que

$$\begin{aligned}\frac{d\mu_i}{d\eta_i} &= \frac{1}{g'(\mu_i)}, \\ \frac{\partial}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} &= \frac{-g''(\mu_i)}{(g'(\mu_i))^2}, \\ \frac{\partial}{\partial \mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^2 &= \frac{-2g''(\mu_i)}{(g'(\mu_i))^3}, \\ \frac{\partial}{\partial \mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^3 &= \frac{-3g''(\mu_i)}{(g'(\mu_i))^4}, \\ \frac{\partial^2}{\partial \mu_i^2} \left(\frac{d\mu_i}{d\eta_i} \right) &= \frac{-g'''(\mu_i)g'(\mu_i) + 2(g''(\mu_i))^2}{(g'(\mu_i))^3}.\end{aligned}$$

Em particular, considerando

$$g(\mu_i) = \text{logit}(\mu_i) = \log \frac{\mu_i}{(1 - \mu_i)},$$

como na avaliação numérica e na aplicação do Capítulo 1, temos

$$\begin{aligned} g'(\mu_i) &= \frac{1}{\mu_i(1 - \mu_i)}, \\ g''(\mu_i) &= \frac{2\mu_i - 1}{\mu_i^2(1 - \mu_i)^2}, \\ g'''(\mu_i) &= \frac{2(1 - 4\mu_i + 6\mu_i^2 - 3\mu_i^3)}{\mu_i^3(1 - \mu_i)^4}. \end{aligned}$$

Adicionalmente, temos as seguintes derivadas:

$$\begin{aligned} \frac{\partial \mu_i^*}{\partial \mu_i} &= \phi \psi'(\mu_i \phi) + \phi \psi'((1 - \mu_i) \phi) = \phi \omega_i, \\ \frac{\partial \mu_i^*}{\partial \phi} &= \mu_i \psi'(\mu_i \phi) - (1 - \mu_i) \psi'((1 - \mu_i) \phi) = \frac{c_i}{\phi}, \\ \frac{\partial \omega_i}{\partial \mu_i} &= \phi \psi''(\mu_i \phi) - \phi \psi''((1 - \mu_i) \phi) = \phi m_i, \\ \frac{\partial \omega_i}{\partial \phi} &= \mu_i \psi''(\mu_i \phi) + (1 - \mu_i) \psi''((1 - \mu_i) \phi), \\ \frac{\partial m_i}{\partial \mu_i} &= \phi \psi'''(\mu_i \phi) + \phi \psi'''((1 - \mu_i) \phi), \\ \frac{\partial m_i}{\partial \phi} &= \mu_i \psi'''(\mu_i \phi) - (1 - \mu_i) \psi'''((1 - \mu_i) \phi), \\ \frac{\partial a_i}{\partial \mu_i} &= 3 \left(\frac{d\mu_i}{d\eta_i} \right) \left[\left(\frac{\partial^2 d\mu_i}{\partial \mu_i^2 d\eta_i} \right) \left(\frac{d\mu_i}{d\eta_i} \right) + 2 \left(\frac{\partial d\mu_i}{\partial \mu_i d\eta_i} \right)^2 \right], \\ \frac{\partial b_i}{\partial \mu_i} &= \left(\frac{\partial d\mu_i}{\partial \mu_i d\eta_i} \right)^3 + \left(\frac{d\mu_i}{d\eta_i} \right) \left[\left(\frac{\partial^3 d\mu_i}{\partial \mu_i^3 d\eta_i} \right) \left(\frac{d\mu_i}{d\eta_i} \right) + 4 \left(\frac{\partial^2 d\mu_i}{\partial \mu_i^2 d\eta_i} \right) \left(\frac{\partial d\mu_i}{\partial \mu_i d\eta_i} \right) \right], \\ \frac{\partial^2 \mu_i^*}{\partial \phi^2} &= \mu_i^2 \psi''(\mu_i \phi) - (1 - \mu_i)^2 \psi''((1 - \mu_i) \phi), \\ \frac{\partial^2 \omega_i}{\partial \phi^2} &= \mu_i^2 \psi'''(\mu_i \phi) + (1 - \mu_i)^2 \psi'''((1 - \mu_i) \phi), \\ \frac{\partial^3 \mu_i^*}{\partial \phi^3} &= \mu_i^3 \psi'''(\mu_i \phi) - (1 - \mu_i)^3 \psi'''((1 - \mu_i) \phi), \\ \frac{\partial c_i}{\partial \mu_i} &= \phi \left(\omega_i + \phi \frac{\partial \omega_i}{\partial \phi} \right), \\ \frac{\partial c_i}{\partial \phi} &= \frac{\partial \mu_i^*}{\partial \phi} + \phi \frac{\partial^2 \mu_i^*}{\partial \phi^2}, \end{aligned}$$

$$\begin{aligned}
 \frac{\partial s_i}{\partial \mu_i} &= 3 \frac{\partial^2 \mu_i^*}{\partial \phi^2} + \phi \frac{\partial^3 \mu_i^*}{\partial \phi^3}, \\
 \frac{\partial s_i}{\partial \phi_i} &= \mu_i^4 \psi'''(\mu_i \psi) + (1 - \mu_i)^4 \psi'''((1 - \mu_i)\phi) - \psi'''(\phi), \\
 \frac{\partial u_i}{\partial \mu_i} &= -\phi^2 \left(3m_i + \phi \frac{\partial m_i}{\partial \phi} \right), \\
 \frac{\partial u_i}{\partial \phi_i} &= -2\omega_i - \phi \left(4 \frac{\partial \omega_i}{\partial \phi} + \frac{\partial^2 \omega_i}{\partial \phi^2} \right), \\
 \frac{\partial r_i}{\partial \mu_i} &= \left(2 \frac{\partial \mu_i^*}{\partial \phi} + \phi \frac{\partial^2 \mu_i^*}{\partial \phi^2} \right) \left(\frac{\partial}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} \right) + \left(2\omega_i + 4\phi \frac{\partial \omega_i}{\partial \phi} + \phi^2 \frac{\partial^2 \omega_i}{\partial \phi^2} \right) \left(\frac{d\mu_i}{d\eta_i} \right), \\
 \frac{\partial r_i}{\partial \phi_i} &= \left(3 \frac{\partial^2 \mu_i^*}{\partial \phi^2} + \phi \frac{\partial^3 \mu_i^*}{\partial \phi^3} \right) \frac{d\mu_i}{d\eta_i} = \frac{\partial s_i}{\partial \mu_i} \frac{d\mu_i}{d\eta_i}, \\
 \frac{\partial}{\partial \mu_i} \frac{\partial \omega_i}{\partial \phi} &= m_i + \phi \frac{\partial m_i}{\partial \phi}.
 \end{aligned}$$

Com essas quantidades, as expressões das derivadas de segunda, terceira e quarta ordens da função de log-verossimilhança podem ser escritas da seguinte maneira:

$$\begin{aligned}
 U_{rs} &= \sum_{i=1}^n \left\{ -\phi^2 \omega_i \left(\frac{d\mu_i}{d\eta_i} \right)^2 + \phi [y_i^* - \mu_i^*] \left(\frac{\partial}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} \right) \frac{d\mu_i}{d\eta_i} \right\} x_{ir} x_{is}, \\
 U_{r\phi} &= - \sum_{i=1}^n [c_i - (y_i^* - \mu_i^*)] \frac{d\mu_i}{d\eta_i} x_{ir}, \\
 U_{\phi\phi} &= - \sum_{i=1}^n d_i, \\
 U_{rst} &= -\phi \sum_{i=1}^n \left\{ \phi^2 m_i \left(\frac{d\mu_i}{d\eta_i} \right)^3 + \phi \omega_i a_i - [y_i^* - \mu_i^*] b_i \right\} x_{ir} x_{is} x_{it}, \\
 U_{rs\phi} &= \sum_{i=1}^n \left\{ \mu_i \left(\frac{d\mu_i}{d\eta_i} \right) + (y_i^* - \mu_i^*) \left(\frac{\partial}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} \right) - c_i \left(\frac{\partial}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} \right) \right\} \frac{d\mu_i}{d\eta_i} x_{ir} x_{is}, \\
 U_{r\phi\phi} &= - \sum_{i=1}^n r_i x_{ir}, \\
 U_{\phi\phi\phi} &= - \sum_{i=1}^n s_i, \\
 U_{rstu} &= -\phi \sum_{i=1}^n \left\{ \phi^2 \left[m_i \frac{\partial}{\partial \mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^3 + \frac{\partial m_i}{\partial \mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^3 \right] \right. \\
 &\quad \left. + \phi \left[\left(\frac{\partial a_i}{\partial \mu_i} + b_i \right) \omega_i + \frac{\partial \omega_i}{\partial \mu_i} a_i \right] - (y_i^* - \mu_i^*) \frac{\partial b_i}{\partial \mu_i} \right\} \frac{d\mu_i}{d\eta_i} x_{ir} x_{is} x_{it} x_{iu}, \\
 U_{rst\phi} &= \sum_{i=1}^n \left\{ -\phi \left[\phi \left(3m_i + \phi^2 \frac{\partial m_i}{\partial \phi} \right) \left(\frac{d\mu_i}{d\eta_i} \right)^3 + a_i \left(2\omega_i + \phi \frac{\partial \omega_i}{\partial \phi} \right) + b_i \frac{\partial \mu_i^*}{\partial \phi} \right] \right\}
 \end{aligned}$$

$$\begin{aligned}
 & + b_i (y_i^* - \mu_i^*) \} x_{ir} x_{is} x_{it}, \\
 U_{rs\phi\phi} &= - \sum_{i=1}^n \frac{\partial r_i}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} x_{ir} x_{is}, \\
 U_{r\phi\phi\phi} &= - \sum_{i=1}^n \frac{\partial s_i}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} x_{ir}, \\
 U_{\phi\phi\phi\phi} &= - \sum_{i=1}^n \frac{\partial s_i}{\partial \phi}.
 \end{aligned}$$

Tomando o valor esperado das derivadas acima, obtemos os seguintes cumulantes:

$$\begin{aligned}
 \kappa_{rs} &= -\phi^2 \sum_{i=1}^n \omega_i \left(\frac{d\mu_i}{d\eta_i} \right)^2 x_{ir} x_{is}, \\
 \kappa_{r\phi} &= - \sum_{i=1}^n c_i \frac{d\mu_i}{d\eta_i} x_{ir}, \\
 \kappa_{\phi\phi} &= - \sum_{i=1}^n d_i, \\
 \kappa_{rst} &= -\phi^2 \sum_{i=1}^n \left[\phi m_i \left(\frac{d\mu_i}{d\eta_i} \right)^3 + \omega_i a_i \right] x_{ir} x_{is} x_{it}, \\
 \kappa_{rs\phi} &= \sum_{i=1}^n \left[u_i \left(\frac{d\mu_i}{d\eta_i} \right) - c_i \left(\frac{\partial}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} \right) \right] x_{ir} x_{is}, \\
 \kappa_{r\phi\phi} &= - \sum_{i=1}^n r_i x_{ir}, \\
 \kappa_{\phi\phi\phi} &= - \sum_{i=1}^n s_i, \\
 \kappa_{rstu} &= -\phi^2 \sum_{i=1}^n \left\{ \phi \left[m_i \frac{\partial}{\partial \mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^3 + \frac{\partial m_i}{\partial \mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^3 + m_i a_i \right] \right. \\
 & \quad \left. + \omega_i \left(\frac{\partial a_i}{\partial \mu_i} + b_i \right) \right\} \frac{d\mu_i}{d\eta_i} x_{ir} x_{is} x_{it} x_{iu}, \\
 \kappa_{rst\phi} &= -\phi \sum_{i=1}^n \left[\phi \left(3m_i + \phi \frac{\partial m_i}{\partial \phi} \right) \left(\frac{d\mu_i}{d\eta_i} \right)^3 + a_i \left(2\omega_i + \phi \frac{\partial \omega_i}{\partial \phi} \right) + b_i \frac{c_i}{\phi} \right] x_{ir} x_{is} x_{it}, \\
 \kappa_{rs\phi\phi} &= - \sum_{i=1}^n \frac{\partial r_i}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} x_{ir} x_{is}, \\
 \kappa_{r\phi\phi\phi} &= - \sum_{i=1}^n \frac{\partial s_i}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} x_{ir}, \\
 \kappa_{\phi\phi\phi\phi} &= - \sum_{i=1}^n \frac{\partial s_i}{\partial \phi}.
 \end{aligned}$$

Definindo a quantidade

$$z_i = \frac{\partial \mu_i^*}{\partial \phi} + \phi \frac{\partial^2 \mu_i^*}{\partial \phi^2},$$

temos que

$$\frac{\partial z_i}{\partial \mu_i} = \omega_i + \phi \left(3 \frac{\partial \omega_i}{\partial \phi} + \phi \frac{\partial^2 \omega_i}{\partial \phi^2} \right)$$

e

$$\frac{\partial z_i}{\partial \phi_i} = 2 \frac{\partial^2 \mu_i^*}{\partial \phi^2} + \phi \frac{\partial^3 \mu_i^*}{\partial \phi^3}.$$

Com isso, tomando derivadas de primeira ordem dos cumulantes com relação aos parâmetros desconhecidos, obtemos

$$\kappa_{rs}^{(u)} = -\phi^2 \sum_{i=1}^n \left[\phi m_i \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^3 + \frac{2}{3} \omega_i a_i \right] x_{ir} x_{is} x_{iu},$$

$$\kappa_{rs}^{(\phi)} = \sum_{i=1}^n \left[u_i \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \right] x_{is} x_{is},$$

$$\kappa_{r\phi}^{(u)} = -\sum_{i=1}^n \left\{ \left[\phi \left(\omega_i + \phi \frac{\partial \omega_i}{\partial \phi} \right) \right] \frac{d\mu_i}{d\eta_i} + c_i \left(\frac{\partial}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \right) \right\} \frac{d\mu_i}{d\eta_i} x_{ir} x_{iu},$$

$$\kappa_{r\phi}^{(\phi)} = -\sum_{i=1}^n z_i \frac{d\mu_i}{d\eta_i} x_{ir},$$

$$\kappa_{\phi\phi}^{(u)} = -\sum_{i=1}^n r_i x_{iu},$$

$$\kappa_{\phi\phi}^{(\phi)} = -\sum_{i=1}^n s_i,$$

$$\kappa_{rst}^{(u)} = -\phi^2 \sum_{i=1}^n \left\{ \phi \left[m_i \left(\frac{\partial}{\partial \mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^3 + a_i \right) + \left(\frac{d\mu_i}{d\eta_i} \right)^3 \frac{\partial m_i}{\partial \mu_i} \right] + \omega_i \frac{\partial a_i}{\partial \mu_i} \right\} \frac{d\mu_i}{d\eta_i} x_{ir} x_{is} x_{it} x_{iu},$$

$$\kappa_{rst}^{(\phi)} = -\phi \sum_{i=1}^n \left[\left(\frac{d\mu_i}{d\eta_i} \right)^3 \left(3\phi m_i + \phi^2 \frac{\partial m_i}{\partial \phi} \right) + a_i \left(2\omega_i + \phi \frac{\partial \omega_i}{\partial \phi} \right) \right] x_{ir} x_{is} x_{it},$$

$$\begin{aligned} \kappa_{rs\phi}^{(t)} = \sum_{i=1}^n \left\{ \frac{\partial u_i}{\partial \mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^2 + \mu_i \frac{\partial}{\partial \mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^2 - \frac{\partial c_i}{\partial \mu_i} \left(\frac{\partial}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} \right) \frac{d\mu_i}{d\eta_i} \right. \\ \left. - \phi \frac{\partial \mu_i^*}{\partial \phi} \left[\left(\frac{\partial^2}{\partial \mu_i^2} \frac{d\mu_i}{d\eta_i} \right) \left(\frac{d\mu_i}{d\eta_i} \right) + \left(\frac{\partial}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} \right)^2 \right] \right\} x_{ir} x_{is} x_{it}, \end{aligned}$$

$$\kappa_{rs\phi}^{(\phi)} = \sum_{i=1}^n \left[\left(\frac{d\mu_i}{d\eta_i} \right) \frac{\partial u_i}{\partial \phi} - \left(\frac{\partial}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} \right) z_i \right] \frac{d\mu_i}{d\eta_i} x_{ir} x_{is},$$

$$\kappa_{r\phi\phi}^{(s)} = - \sum_{i=1}^n \frac{\partial r_i}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} x_{ir} x_{is},$$

$$\kappa_{r\phi\phi}^{(\phi)} = - \sum_{i=1}^n \frac{\partial r_i}{\partial \phi} x_{ir},$$

$$\kappa_{\phi\phi\phi}^{(r)} = - \sum_{i=1}^n \frac{\partial s_i}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} x_{ir},$$

$$\kappa_{\phi\phi\phi}^{(\phi)} = - \sum_{i=1}^n \frac{\partial s_i}{\partial \phi}.$$

Tomando derivadas de segunda ordem dos cumulantes de segunda ordem obtemos

$$\begin{aligned} \kappa_{rs}^{(tu)} &= -\phi^2 \sum_{i=1}^n \left\{ \phi \left[m_i \left(\frac{\partial}{\partial \mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^3 + \frac{2}{3} a_i \right) + \left(\frac{d\mu_i}{d\eta_i} \right)^3 \frac{\partial m_i}{\partial \mu_i} \right] \right. \\ &\quad \left. + \frac{2}{3} \omega_i \frac{\partial a_i}{\partial \mu_i} \right\} \frac{d\mu_i}{d\eta_i} x_{ir} x_{is} x_{it} x_{iu}, \end{aligned}$$

$$\kappa_{rs}^{(tu)} = \sum_{i=1}^n \left[\frac{\partial u_i}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} + 2u_i \frac{\partial}{\partial \mu_i} \left(\frac{d\mu_i}{d\eta_i} \right) \right] \left(\frac{d\mu_i}{d\eta_i} \right)^2 x_{ir} x_{is} x_{it},$$

$$\kappa_{rs}^{(\phi\phi)} = \sum_{i=1}^n \frac{\partial \mu_i}{\partial \phi} \left(\frac{d\mu_i}{d\eta_i} \right)^2 x_{ir} x_{is},$$

$$\begin{aligned} \kappa_{r\phi}^{(st)} &= -\phi \sum_{i=1}^n \left\{ \frac{\partial \omega_i}{\partial \mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^2 + \omega_i \frac{\partial}{\partial \mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^2 \right. \\ &\quad \left. + \phi \left[\frac{\partial}{\partial \mu_i} \left(\frac{\partial \omega_i}{\partial \phi} \right) \left(\frac{d\mu_i}{d\eta_i} \right)^2 + \frac{\partial \omega_i}{\partial \phi} \frac{\partial}{\partial \mu_i} \left(\frac{d\mu_i}{d\eta_i} \right)^2 \right] + \left(\omega_i + \phi \frac{\partial \omega_i}{\partial \phi} \right) \left(\frac{\partial}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} \right) \frac{d\mu_i}{d\eta_i} \right. \\ &\quad \left. + \frac{\partial \mu_i^*}{\partial \phi} \left[\frac{\partial^2}{\partial \mu_i} \left(\frac{d\mu_i}{d\eta_i} \right) \left(\frac{d\mu_i}{d\eta_i} \right) + \left(\frac{\partial}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} \right)^2 \right] \right\} \frac{d\mu_i}{d\eta_i} x_{ir} x_{is} x_{it}, \end{aligned}$$

$$\kappa_{r\phi}^{(\phi s)} = - \sum_{i=1}^n \left[\frac{\partial z_i}{\partial \mu_i} \left(\frac{d\mu_i}{d\eta_i} \right) + z_i \left(\frac{\partial}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} \right) \right] \frac{d\mu_i}{d\eta_i} x_{ir} x_{is},$$

$$\kappa_{r\phi}^{(\phi\phi)} = - \sum_{i=1}^n \frac{\partial z_i}{\partial \phi} \frac{d\mu_i}{d\eta_i} x_{ir} x_{is},$$

$$\kappa_{\phi\phi}^{(rs)} = - \sum_{i=1}^n \frac{\partial r_i}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} x_{ir} x_{is},$$

$$\kappa_{\phi\phi}^{(\phi r)} = - \sum_{i=1}^n \frac{\partial s_i}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} x_{ir},$$

$$\kappa_{\phi\phi}^{(\phi\phi)} = - \sum_{i=1}^n \frac{\partial s_i}{\partial \phi}.$$

APÊNDICE B

Correção de Bartlett (implementação em R)

Este apêndice apresenta a função implementada em linguagem R para a determinação do termo ε_k , dado em (1.4), para a correção de Bartlett em modelos de regressão beta com dispersão constante.

```
#####  
#  
# PROGRAMA: Dados alguns argumentos de entrada para a funcao, que sao provenientes de  
#           um modelo de regressao beta com dispersao constante ajustado utilizando  
#           a funcao betareg do R, a funcao retorna o \epsilon_k da correcao de Bartlett.  
#           A funcao utiliza a expressao matricial geral para obtencao do fator  
#           de correcao de Bartlett apresentado em Cordeiro (1993).  
#  
# OBS:      Esta implementacao nao realiza simulacao de Monte Carlo. Em um estudo de  
#           simulacao esta funcao deve ser chamada dentro de um outro arquivo com a  
#           implementacao de interesse.  
#           Ao final ha um exemplo de utilizacao desta funcao.  
#  
# AUTOR:    Fábio Mariano Bayer  
#  
# E-MAIL:   fabiobayer@gmail.com  
#  
# DATA:    Outubro/2011  
#  
#####  
  
# funcao auxiliar que conta quantos zeros tem um vetor de entrada  
manyzeros<-function(r)  
{  
  return(sum(as.numeric(r==0)))  
}  
  
# funcao que determina \epsilon_k do fator de correcao de Bartlett  
bartlett<-function(mut,phit,K_1,X)  
{  
  #####  
  # Argumentos de entrada da funcao  
  # mut - vetor \mu ajustado  
  # phit - estimativa de \phi  
  # K_1 - inversa da matriz de informacao (vcov)  
  # X - matriz de regressores  
  #####  
  
  # Obs: em K_1 a ultima linha (e coluna) eh do phi  
  
  pf<-ncol(X) # numero de parametros do modelo da media (sem contar o phi)  
  kf<-pf+1 # numero de parametros total (contando o phi)  
  indice_k<-seq(1,kf) #c(seq(1:pf),0) # o ultimo corresponde ao phi  
  
  # lembrete  
  #psigamma(x,deriv=0) = digamma(x)  
  #psigamma(x,deriv=1) = trigamma(x)
```

```

# Definindo quantidades
wt<-psigamma(mut*phit,deriv=1) + psigamma((1-mut)*phit,deriv=1)
mt<-psigamma(mut*phit,deriv=2) - psigamma((1-mut)*phit,deriv=2)
DmuDmueta<- -g2lin(mut)/(glin(mut)^2)
DmuDeta<- 1/glin(mut)
at<- 3*DmuDmueta*(DmuDeta^2)
dt<- ((1-mut)^2)*psigamma((1-mut)*phit,deriv=1) + (mut^2)*psigamma(mut*phit,deriv=1)
      -psigamma(phit,deriv=1)
st<- ((1-mut)^3)*psigamma((1-mut)*phit,deriv=2) + (mut^3)*psigamma(mut*phit,deriv=2)
      -psigamma(phit,deriv=2)
ct<- phit*(mut*wt - psigamma((1-mut)*phit,deriv=1))
DwDphi<- mut*psigamma(mut*phit,deriv=2) + (1-mut)*psigamma((1-mut)*phit,deriv=2)
DmustarDphi<- ct/phit
D2mustarDphi<- (mut^2)*psigamma(mut*phit,deriv=2) - ((1-mut)^2)*psigamma((1-mut)*phit,deriv=2)
D3mustarDphi<- (mut^3)*psigamma(mut*phit,deriv=3) - ((1-mut)^3)*psigamma((1-mut)*phit,deriv=3)
ut<- -phit*(2*wt+phit*(DwDphi))
rt<- (2*(DmustarDphi) + phit*D2mustarDphi)*DmuDeta
zt<- DmustarDphi + phit*D2mustarDphi
Dmueta3<- -3*g2lin(mut)/(glin(mut)^4)
Dmueta2<- -2*g2lin(mut)/(glin(mut)^3)
D2mueta<- (-g3lin(mut)*glin(mut)+2*(g2lin(mut)^2))/(glin(mut)^3)
bt<- DmuDeta*(D2mueta*DmuDeta + (DmuDmueta^2))
DaDmu<- 3*DmuDeta*(D2mueta*DmuDeta+2*(DmuDmueta^2))
DmDphi<- mut*psigamma(mut*phit,deriv=3) - (1-mut)*psigamma((1-mut)*phit,deriv=3)
D2wDphi<- (mut^2)*psigamma(mut*phit,deriv=3) + ((1-mut)^2)*psigamma((1-mut)*phit,deriv=3)
DrDmu<- (2*DmustarDphi + phit*D2mustarDphi)*DmuDmueta + (2*wt+4*phit*DwDphi
      +(phit^2)*D2wDphi)*DmuDeta
DsDmu<- 3*D2mustarDphi + phit*D3mustarDphi
DsDphi<- (mut^4)*psigamma(mut*phit,deriv=3) + ((1-mut)^4)*psigamma((1-mut)*phit,deriv=3)
      - psigamma(phit,deriv=3)
DuDphi<- -2*wt - 4*phit*DwDphi - (phit^2)*D2wDphi
DrDphi<- DsDmu*DmuDeta
DcDmu<- phit*(wt + phit*DwDphi)
DuDmu<- -(phit^2)*(3*mt + phit*DmDphi)
DzDphi<- 2*D2mustarDphi + phit*D3mustarDphi
DzDmu<- wt + 3*phit*DwDphi + (phit^2)*D2wDphi
DmDmu<- phit*psigamma(mut*phit,deriv=3) + phit*psigamma((1-mut)*phit,deriv=3)
DmuDwphi<- mt+phit*DmDphi
DwDmu<- phit*mt

k4<-function(r,s,t,u) # k_rstu
{
  vetor_theta<-c(r,s,t,u) # vetor de indices do vetor theta (0 corresponde ao phi)
  ordem_indices<-order(vetor_theta,decreasing=TRUE) # do maior indice para o menor

  nphi<-manyzeros(vetor_theta) # quantos phi's

  if(nphi == 0) # k_rstu
  {
    kappa<- -phit^2*sum((phit*(mt*Dmueta3 + DmDmu*(DmuDeta)^3 + mt*at) +
      wt*(DaDmu + bt))*DmuDeta*X[,vetor_theta[ordem_indices[1]]]*
      X[,vetor_theta[ordem_indices[2]]]*X[,vetor_theta[ordem_indices[3]]]*
      X[,vetor_theta[ordem_indices[4]]])
  }
  if(nphi == 1) # k_rstphi
  {
    kappa<- -phit*sum((phit*(3*mt+phit*DmDphi)*DmuDeta^3 + at*(2*wt+phit*DwDphi) +
      bt*ct/phit)*X[,vetor_theta[ordem_indices[1]]]*X[,vetor_theta[ordem_indices[2]]]*
      X[,vetor_theta[ordem_indices[3]]])
  }
  if(nphi == 2) # k_rsphiphi
  {
    kappa<- -sum(DrDmu*DmuDeta*X[,vetor_theta[ordem_indices[1]]]*
      X[,vetor_theta[ordem_indices[2]]])
  }
  if(nphi == 3) # k_rphiphiphi

```

```

{
  kappa<- -sum(DsDmu*DmuDeta*X[,vetor_theta[ordem_indices[1]]])
}
if(nphi == 4) # k_phiphiphiphi
{
  kappa<- -sum(DsDphi)
}
return(kappa)
}

k3<-function(r,s,t) # k_rst
{
  vetor_theta<-c(r,s,t) # vetor de indices do vetor theta (0 corresponde ao phi)
  ordem_indices<-order(vetor_theta,decreasing=TRUE) # do maior indice para o menor

  nphi<-manyzeros(vetor_theta) # quantos phi's

  if(nphi == 0) # k_rst
  {
    kappa<- -phit^2*sum((phit*mt*DmuDeta^3 + wt*at)*
      X[,vetor_theta[ordem_indices[1]]]*X[,vetor_theta[ordem_indices[2]]]*
      X[,vetor_theta[ordem_indices[3]]])
  }
  if(nphi == 1) # k_rsphi
  {
    kappa<- sum((ut*DmuDeta - ct*DmuDmueta)*DmuDeta*X[,vetor_theta[ordem_indices[1]]]*
      X[,vetor_theta[ordem_indices[2]]])
  }
  if(nphi == 2) # k_rphiphi
  {
    kappa<- -sum(rt*X[,vetor_theta[ordem_indices[1]]])
  }
  if(nphi == 3) # k_phiphiphi
  {
    kappa<- -sum(st)
  }
  return(kappa)
}

k31<-function(r,s,t,U) # k_rst(u)
{
  vetor_theta<-c(r,s,t,U) # vetor de indices do vetor theta (0 corresponde ao phi)
  ordem_indices<-order(vetor_theta,decreasing=TRUE) # do maior indice para o menor

  nphi<-manyzeros(vetor_theta) # quantos phi's

  if(U == 0) # k_rst(phi)
  {
    if(nphi == 1) # k_rst(phi)
    {
      kappa<- -phit*sum(((DmuDeta^3)*(3*phit*mt + phit^2 *DmDphi) + at*(2*wt + phit*DwDphi))*
        X[,vetor_theta[ordem_indices[1]]]*X[,vetor_theta[ordem_indices[2]]]*
        X[,vetor_theta[ordem_indices[3]]])
    }
    if(nphi == 2) # k_rsphi(phi)
    {
      kappa<- sum((DmuDeta*DuDphi - DmuDmueta*zt)*DmuDeta*
        X[,vetor_theta[ordem_indices[1]]]*X[,vetor_theta[ordem_indices[2]]])
    }
    if(nphi == 3) # k_rphiphi(phi)
    {
      kappa<- -sum(DrDphi * X[,vetor_theta[ordem_indices[1]]])
    }
    if(nphi == 4) # k_phiphiphi(phi)
    {
      kappa<- -sum(DsDphi)
    }
  }
}

```

```

}else{ # k_rst(u), ie se U>0

  if(nphi == 0) # k_rst(u)
  {
    kappa<- -phit^2 *sum((phit*(mt*(Dmueta3 + at) + (DmuDeta^3)*DmDmu) + wt*DaDmu)*DmuDeta*
      X[,vetor_theta[ordem_indices[1]]]*X[,vetor_theta[ordem_indices[2]]]*
      X[,vetor_theta[ordem_indices[3]]]*X[,vetor_theta[ordem_indices[4]]])
  }
  if(nphi == 1) # k_rsphi(t)
  {
    kappa<-sum((DuDmu*DmuDeta^2 + ut*Dmueta2-DcDmu*DmuDmueta*DmuDeta-ct*
      (D2mueta*DmuDeta+DmuDmueta^2))*DmuDeta*X[,vetor_theta[ordem_indices[1]]]*
      X[,vetor_theta[ordem_indices[2]]]*X[,vetor_theta[ordem_indices[3]]])
  }
  if(nphi == 2) # k_rphiphi(s)
  {
    kappa<- -sum(DrDmu*DmuDeta*X[,vetor_theta[ordem_indices[1]]]*
      X[,vetor_theta[ordem_indices[2]]])
  }
  if(nphi == 3) # k_phiphiphi(r)
  {
    kappa<- -sum(DsDmu*DmuDeta*X[,vetor_theta[ordem_indices[1]]])
  }
}
return(kappa)
}

k22<-function(r,s,T,U) # k_rs(tu)
{
  vetor_theta<-c(r,s,T,U) # vetor de indices do vetor theta (0 corresponde ao phi)
  ordem_indices<-order(vetor_theta,decreasing=TRUE) # do maior indice para o menor

  nphi<-manyzeros(vetor_theta) # quantos phi's

  if(U == 0)
  {
    if(T == 0) # k_rs(phiphi)
    {
      if(nphi == 2) # k_rs(phiphi)
      {
        kappa<- sum(DuDphi*(DmuDeta^2)*X[,vetor_theta[ordem_indices[1]]]*
          X[,vetor_theta[ordem_indices[2]]])
      }
      if(nphi == 3) # k_rphi(phiphi)
      {
        kappa<- -sum(DzDphi * DmuDeta * X[,vetor_theta[ordem_indices[1]]])
      }
      if(nphi == 4) # k_phiphi(phiphi)
      {
        kappa<- -sum(DsDphi)
      }
    }else{ # k_rs(tphi)
      if(nphi == 1) # k_rs(tphi)
      {
        kappa<-sum((DuDmu*DmuDeta+2*ut*DmuDmueta)*(DmuDeta)^2*
          X[,vetor_theta[ordem_indices[1]]]*X[,vetor_theta[ordem_indices[2]]]*
          X[,vetor_theta[ordem_indices[3]]])
      }
      if(nphi == 2) # k_rphi(sphi)
      {
        kappa<- -sum((DzDmu*DmuDeta + zt*DmuDmueta)*DmuDeta*
          X[,vetor_theta[ordem_indices[1]]]*X[,vetor_theta[ordem_indices[2]]])
      }
      if(nphi == 3) # k_phiphi(rphi)
      {
        kappa<- -sum(DsDmu*DmuDeta*X[,vetor_theta[ordem_indices[1]]])
      }
    }
  }
}

```

```

}
}else{ # se U > 0
  if(T == 0) # k_rs(phit)
  {
    if(nphi == 1) # k_rs(phit)
    {
      kappa<- -phit*sum(((2/3)*at*(2*wt+phit*DwDphi) + phit*(3*mt+phit*DmDphi)*
        (DmuDeta^3))*X[,vetor_theta[ordem_indices[1]]]*X[,vetor_theta[ordem_indices[2]]]*
        X[,vetor_theta[ordem_indices[3]]])
    }
    if(nphi == 2) # k_rphi(phit)
    {
      kappa<- -sum((DzDmu*DmuDeta + zt*DmuDmueta)*DmuDeta*
        X[,vetor_theta[ordem_indices[1]]]*X[,vetor_theta[ordem_indices[2]]])
    }
    if(nphi == 3) # k_phiphi(phit)
    {
      kappa<- -sum(DsDmu*DmuDeta*X[,vetor_theta[ordem_indices[1]]])
    }
  }
}else{ # k_rs(tu)
  if(nphi == 0) # k_rs(tu)
  {
    kappa<- -(phit^2)*sum((phit*(mt*(Dmueta3+(2/3)*at) + (DmuDeta^3)*DmDmu) + (2/3)*wt*DaDmu)*
      DmuDeta*X[,vetor_theta[ordem_indices[1]]]*X[,vetor_theta[ordem_indices[2]]]*
      X[,vetor_theta[ordem_indices[3]]]*X[,vetor_theta[ordem_indices[4]]])
  }
  if(nphi == 1) # k_rphi(st)
  {
    kappa<- -phit*sum((DwDmu*DmuDeta^2+wt*Dmueta2+phit*(DmuDwphi*DmuDeta^2+
      DwDphi*Dmueta2) + (wt+phit*DwDphi)*DmuDmueta*DmuDeta+DmstarDphi*(D2mueta*DmuDeta+
      DmuDmueta^2))*DmuDeta*X[,vetor_theta[ordem_indices[1]]]*
      X[,vetor_theta[ordem_indices[2]]]*X[,vetor_theta[ordem_indices[3]]])
  }
  if(nphi == 2) # k_phiphi(rs)
  {
    kappa<- -sum(DrDmu*DmuDeta*X[,vetor_theta[ordem_indices[1]]]*
      X[,vetor_theta[ordem_indices[2]]])
  }
}
}
return(kappa)
}

k21<-function(r,s,T) # k_rs(t)
{
  vetor_theta<-c(r,s,T) # vetor de indices do vetor theta (0 corresponde ao phi)
  ordem_indices<-order(vetor_theta,decreasing=TRUE) # do maior indice para o menor

  nphi<-manyzeros(vetor_theta) # quantos phi's

  if(T == 0) # k_rs(phi)
  {
    if(nphi == 1) # k_rs(phi)
    {
      kappa<- sum((ut*(DmuDeta)^2)*X[,vetor_theta[ordem_indices[1]]]*
        X[,vetor_theta[ordem_indices[2]]])
    }
    if(nphi == 2) # k_rphi(phi)
    {
      kappa<- -sum(zt*DmuDeta*X[,vetor_theta[ordem_indices[1]]])
    }
    if(nphi == 3) # k_phiphi(phi)
    {
      kappa<- -sum(st)
    }
  }
}else{
  if(nphi == 0) # k_rs(t)

```

```

{
  kappa<- -(phit^2)*sum((phit*mt*(DmuDeta^3) + (2/3)*wt*at)*
    X[,vetor_theta[ordem_indices[1]]]*X[,vetor_theta[ordem_indices[2]]]*
    X[,vetor_theta[ordem_indices[3]]])
}
if(nphi == 1) # k_rphi(s)
{
  kappa<- -sum((phit*(wt+phit*DwDphi)*DmuDeta + ct*DmuDmueta)*DmuDeta*
    X[,vetor_theta[ordem_indices[1]]]*X[,vetor_theta[ordem_indices[2]]])
}
if(nphi == 2) # k_phiphi(r)
{
  kappa<- sum(-rt*X[,vetor_theta[ordem_indices[1]]])
}
}
return(kappa)
}

# inicizlizacao das matrizes k x k
mA<- array(rep(0,kf^4),dim=c(kf,kf,kf,kf))
mP<- array(rep(0,kf^3),dim=c(kf,kf,kf))
mQ<- mP

# gera a matriz mA
for(ti in indice_k)
{
  if(ti>pf)
  {
    tib<-0
  }else{
    tib<-ti
  }
  for(ui in indice_k)
  {
    if(ui>pf)
    {
      uib<-0
    }else{
      uib<-ui
    }
    for(ri in indice_k)
    {
      if(ri>pf)
      {
        rib<-0
      }else{
        rib<-ri
      }
      for(si in indice_k)
      {
        if(si>pf)
        {
          sib<-0
        }else{
          sib<-si
        }
        mA[ti,ui,ri,si]<- k4(rib,sib,tib,uib)/4 - k31(rib,sib,tib,uib) + k22(rib,tib,sib,uib)
      }
    }
  }
}

# gera a matriz mP
for(ti in indice_k)
{
  if(ti>pf)
  {

```

```

    tib<-0
  }else{
    tib<-ti
  }
  for(ri in indice_k)
  {
    if(ri>pf)
    {
      rib<-0
    }else{
      rib<-ri
    }
    for(si in indice_k)
    {
      if(si>pf)
      {
        sib<-0
      }else{
        sib<-si
      }
      mP[ti,ri,si]<- k3(rib,sib,tib)
    }
  }
}

#gera a matriz mQ
for(ui in indice_k)
{
  if(ui>pf)
  {
    uib<-0
  }else{
    uib<-ui
  }
  for(ri in indice_k)
  {
    if(ri>pf)
    {
      rib<-0
    }else{
      rib<-ri
    }
    for(si in indice_k)
    {
      if(si>pf)
      {
        sib<-0
      }else{
        sib<-si
      }
      mQ[ui,ri,si]<- k21(sib,uib,rib)
    }
  }
}

# Inicializa matrizes
L<-matrix(rep(0,kf^2),ncol=kf)
M1<-L
M2<-L
M3<-L
N1<-L
N2<-L
N3<-L

# gera a matriz L
for(ri in indice_k)

```

```

{
  for(si in indice_k)
  {
    L[ri,si]<- sum(diag( K_1 %*% mA[ri,si,,] ))
  }
}

# geram as matrizes Mi
for(ri in indice_k)
{
  for(si in indice_k)
  {
    M1[ri,si]<- sum(diag( K_1%*%mP[ri,,]%*%K_1%*%mP[si,,] ))
  }
}
for(ri in indice_k)
{
  for(si in indice_k)
  {
    M2[ri,si]<- sum(diag( K_1%*%mP[ri,,]%*%K_1%*%t(mQ[si,,] )))
  }
}
for(ri in indice_k)
{
  for(si in indice_k)
  {
    M3[ri,si]<- sum(diag( K_1%*%mQ[ri,,]%*%K_1%*%mQ[si,,] ))
  }
}

# geram as matrizes Ni
for(ri in indice_k)
{
  for(si in indice_k)
  {
    N1[ri,si]<- (sum(diag( mP[ri,,]%*%K_1 )))*(sum(diag( mP[si,,]%*%K_1 )))
  }
}
for(ri in indice_k)
{
  for(si in indice_k)
  {
    N2[ri,si]<- (sum(diag( mP[ri,,]%*%K_1 )))*(sum(diag( mQ[si,,]%*%K_1 )))
  }
}
for(ri in indice_k)
{
  for(si in indice_k)
  {
    N3[ri,si]<- (sum(diag( mQ[ri,,]%*%K_1 )))*(sum(diag( mQ[si,,]%*%K_1 )))
  }
}

M<- -M1/6 + M2 - M3
N<- -N1/4 + N2 - N3

return(sum(diag( K_1%*%(L - M - N) )))

} # fim da funcao

# A seguir ha um exemplo de utilizacao desta funcao

```

```
#####
#
#           EXEMPLO DE USO DA FUNCAO CRIADA - bartlett()
#
#####

## Sejam:
# y - vetor contendo n observacoes da variavel dependente
# mX - matriz de regressores (a primeira coluna eh um vetor de 1's, relativa ao intercepto)

## Chamando o pacote betareg
# library(betareg)

## Estimando o modelo irrestrito (full) e o modelo restrito (null, sob H0)
# estima_full <- betareg(y~mX[,2]+mX[,3]+mX[,4]+mX[,5]) # modelo sem restricoes
# estima_null <- betareg(y~mX[,4]+mX[,5]) # modelo sob hipotese nula

# indice_X_H0<-c(1,4,5) # covariaveis incluidas no modelo restrito.
# 0 1 representa o intercepto e y~mX[,4]+mX[,5] é o modelo.

## definindo quantidades uteis para a utilizacao da funcao bartlett()
# mu_hat_full <- estima_full$fitted.values # vetor \mu ajustado
# phi_hat_full <- estima_full$coefficients$precision # estimativa de \phi
# mcov_full <- estima_full$vcov # inversa da matriz de informacao
# loglik_full <- estima_full$loglik # funcao de log-verossimilhanca maximizada

# mu_hat_null <- estima_null$fitted.values # vetor \mu ajustado
# phi_hat_null <- estima_null$coefficients$precision # estimativa de \phi
# mcov_null <- estima_null$vcov # inversa da matriz de informacao
# loglik_null <- estima_null$loglik # funcao de log-verossimilhanca maximizada

## Estatistica da razao de verossimilhanças
# LR <- 2*(loglik_full-loglik_null) # estatistica LR

## USANDO A FUNCAO BARTLETT CRIADA
# epsilon_k <- bartlett(mu_hat_full,phi_hat_full,mcov_full,mX)
# epsilon_k_q <- bartlett(mu_hat_null,phi_hat_null,mcov_null,mX[,indice_X_H0])

# B <- epsilon_k-epsilon_k_q

## Estatisticas corrigidas (Bartlett)
# LRs1 <- LR/c # LR_b1
# LRs2 <- LR*exp(-B/q) # LR_b2
# LRs3 <- LR*(1-B/q) # LR_b3
```

Comparação dos testes baseados nas estatísticas LR_{b3} e LR_{sk1}

Neste apêndice apresentamos uma avaliação numérica dos desempenhos dos testes da razão de verossimilhanças corrigidos baseados nas estatísticas LR_{b3} e LR_{sk1} . O experimento computacional é semelhante, e complementar, ao apresentado na Seção 1.4 do Capítulo 1. Os resultados apresentados neste apêndice consideram simulações com 50000 réplicas de Monte Carlo e acrescentamos o caso em que $\phi = 100$.

A estatística LR_{b3} foi a correção de Bartlett que apresentou melhor desempenho na avaliação numérica da Seção 1.4 e o ajuste de Skovgaard LR_{sk1} , que se encontra proposto na literatura, mostrou-se o principal concorrente ao teste corrigido por Bartlett. Utilizamos 50000 réplicas de Monte Carlo, diferentemente das 10000 réplicas consideradas no Capítulo 1, com o objetivo de reduzir possíveis erros experimentais. Assim, temos conclusões mais fidedignas a respeito dos desempenhos da correção de Bartlett e do ajuste de Skovgaard do teste da razão de verossimilhanças em regressão beta. Focamos, neste apêndice, na avaliação de tamanho dos testes corrigidos e na aproximação das distribuições nulas das estatísticas LR_{b3} e LR_{sk1} pela distribuição assintótica das mesmas. Os resultados das simulações estão apresentados na Tabela C.1 e nas Figuras C.1, C.2 e C.3.

Analisando a Tabela C.1 percebemos que para tamanhos amostrais moderados, como $n = 30$ e $n = 40$, os testes baseados nas estatísticas LR_{b3} e LR_{sk1} possuem taxas de rejeição nula comparáveis. Contudo, à medida que o tamanho amostral decresce o desempenho do teste baseado na estatística LR_{sk1} diminui consideravelmente, enquanto que o teste corrigido por Bartlett mantém o desempenho. Esse fato se torna mais evidente para os maiores valores do parâmetro de precisão ϕ . Por exemplo, para o caso em que $q = 2$ e $\phi = 100$ as taxas de rejeição nula do teste baseado em LR_{sk1} , para $n = 15$, foram 15.5%, 10.0% e 3.7%, para os níveis nominais de 10%, 5% e 1%, respectivamente. Para o mesmo caso, as taxas do teste baseado na LR_{b3} foram 9.9%, 4.9% e 1.0%. Percebe-se claramente a distorção de tamanho do teste baseado na estatística com ajuste de Skovgaard para os menores tamanhos amostrais, enquanto que o teste corrigido por Bartlett possui taxas de rejeição nula próximas aos níveis nominais independentemente do tamanho amostral. Considerando $n = 15$, a Tabela C.1 evidencia que o desempenho do teste baseado em estatística LR_{b3} é sempre superior, ou igual, ao desempenho do teste baseado na estatística LR_{sk1} em termos de tamanho do teste.

O desempenho inferior do teste baseado na estatística LR_{sk1} frente ao desempenho do teste baseado na estatística LR_{b3} pode ser explicado pela aproximação pobre da distribuição nula da estatística LR_{sk1} pela distribuição χ^2 de referência para os menores tamanhos amostrais. Todos os gráficos das Figuras C.1, C.2 e C.3 evidenciam a melhor aproximação da distribuição nula da estatística LR_{b3} do que da distribuição nula da estatística LR_{sk1} pela distribuição nula limite. A Figura C.3(b), por exemplo, mostra que a densidade estimada da estatística LR_{sk1} apresenta um afastamento da distribuição de referência na cauda da distribuição. Essa aproximação pobre na cauda pode implicar grandes distorções de tamanho do teste.

Percebemos que em todos os casos considerados o ajuste de Skovgaard apresenta fraco desempenho à medida que o tamanho amostral diminui e que o parâmetro ϕ aumenta. Por outro lado, o desempenho da correção de Bartlett se mantém o mesmo nos diferentes cenários investigados.

Tabela C.1 Taxas de rejeição nula (percentuais) dos testes da razão de verossimilhanças original e corrigidos.

		$\alpha = 10\%$				$\alpha = 5\%$				$\alpha = 1\%$			
		15	20	30	40	15	20	30	40	15	20	30	40
ϕ	n												
	Estat												
$\mathcal{H}_0 : \beta_2 = 0 (q = 1)$													
100	LR	18.9	15.9	14.1	12.7	11.8	9.5	7.7	6.8	4.0	2.8	2.0	1.6
	LR_{b3}	9.9	10.0	10.2	10.0	5.0	5.0	5.1	4.8	1.0	1.0	1.0	1.0
	LR_{sk1}	13.7	10.8	10.1	10.1	8.8	5.9	5.0	4.9	3.3	1.7	1.0	1.0
30	LR	19.5	16.1	14.1	12.9	12.3	9.5	8.0	6.9	4.3	2.8	2.2	1.7
	LR_{b3}	10.4	10.0	10.3	10.1	5.2	5.0	5.2	5.0	1.1	1.0	1.1	1.0
	LR_{sk1}	10.7	10.2	10.4	10.3	5.6	5.1	5.2	5.6	1.5	1.0	1.2	1.0
10	LR	20.8	18.2	15.2	14.7	13.4	11.4	8.9	8.3	5.0	3.9	2.5	2.2
	LR_{b3}	11.1	11.8	11.3	11.6	5.8	6.3	5.8	6.2	1.3	1.6	1.3	1.4
	LR_{sk1}	11.8	12.0	11.5	11.7	6.2	6.4	5.9	6.2	1.5	1.5	1.3	1.5
5	LR	22.1	17.2	17.5	15.9	14.3	10.4	10.6	9.3	5.6	3.3	3.4	2.7
	LR_{b3}	11.6	10.9	13.0	13.0	6.1	5.6	7.1	7.1	1.4	1.2	1.7	1.7
	LR_{sk1}	12.7	11.4	12.9	12.9	6.8	5.8	7.0	7.2	1.7	1.3	1.8	1.7
$\mathcal{H}_0 : \beta_2 = \beta_3 = 0 (q = 2)$													
100	LR	21.5	17.5	14.7	13.3	13.6	10.3	8.2	7.2	4.5	3.1	2.2	1.8
	LR_{b3}	9.9	9.8	10.0	9.9	4.9	4.9	5.0	4.9	1.0	0.9	1.0	1.0
	LR_{sk1}	15.5	10.4	10.0	9.9	10.0	5.6	5.0	4.9	3.7	1.6	1.2	1.1
30	LR	22.2	18.1	15.2	14.4	14.1	10.8	8.5	8.1	4.8	3.4	2.2	2.1
	LR_{b3}	10.3	10.1	10.3	10.9	5.1	5.2	5.0	5.6	1.1	1.0	1.0	1.2
	LR_{sk1}	10.3	10.3	10.4	11.1	5.3	5.3	5.1	5.6	1.4	1.0	1.0	1.2
10	LR	22.3	20.1	16.3	15.2	14.3	12.5	9.4	8.7	5.2	4.1	2.6	2.4
	LR_{b3}	10.7	11.8	11.2	11.8	5.5	6.2	5.9	6.1	1.2	1.3	1.3	1.4
	LR_{sk1}	11.0	11.8	11.5	11.8	5.7	6.2	6.0	6.2	1.2	1.3	1.3	1.4
5	LR	21.8	18.2	15.6	12.9	13.9	11.0	8.9	7.0	5.3	3.4	2.5	1.9
	LR_{b3}	10.4	10.6	10.8	9.8	5.5	5.3	5.6	4.9	1.1	1.1	1.2	1.1
	LR_{sk1}	11.7	11.3	10.5	10.0	6.1	5.7	5.5	5.0	1.2	1.2	1.2	1.1
$\mathcal{H}_0 : \beta_2 = \beta_3 = \beta_4 = 0 (q = 3)$													
100	LR	22.6	18.5	15.0	13.5	14.2	11.0	8.5	7.4	4.7	3.2	2.1	1.8
	LR_{b3}	10.1	10.1	10.1	9.9	5.1	5.0	5.1	5.1	1.0	1.0	1.0	1.0
	LR_{sk1}	13.3	10.6	10.0	9.9	8.4	5.8	5.0	5.1	3.2	1.7	1.0	1.0
30	LR	22.5	18.5	15.0	13.6	14.2	11.1	8.5	7.5	4.8	3.4	2.2	1.8
	LR_{b3}	10.1	10.1	9.9	10.0	5.0	5.1	5.0	5.1	1.0	1.0	1.0	1.0
	LR_{sk1}	10.2	10.2	10.0	10.2	5.1	5.1	5.0	5.2	1.0	1.0	1.0	1.1
10	LR	23.1	18.8	15.0	13.5	14.5	11.1	8.4	7.4	5.0	3.3	2.2	1.9
	LR_{b3}	10.2	10.2	10.0	9.9	5.2	5.1	5.0	5.0	1.0	1.1	1.0	1.0
	LR_{sk1}	10.4	10.4	10.1	10.1	5.2	5.3	5.1	5.1	1.0	1.1	1.0	1.0
5	LR	22.9	18.5	14.9	13.2	14.5	11.1	8.2	7.2	5.0	3.5	2.2	1.8
	LR_{b3}	10.5	10.3	9.9	9.8	5.2	5.2	5.0	4.9	1.0	1.0	1.0	1.0
	LR_{sk1}	11.4	11.2	10.2	10.3	5.8	5.6	5.2	5.1	1.1	1.1	1.1	1.1

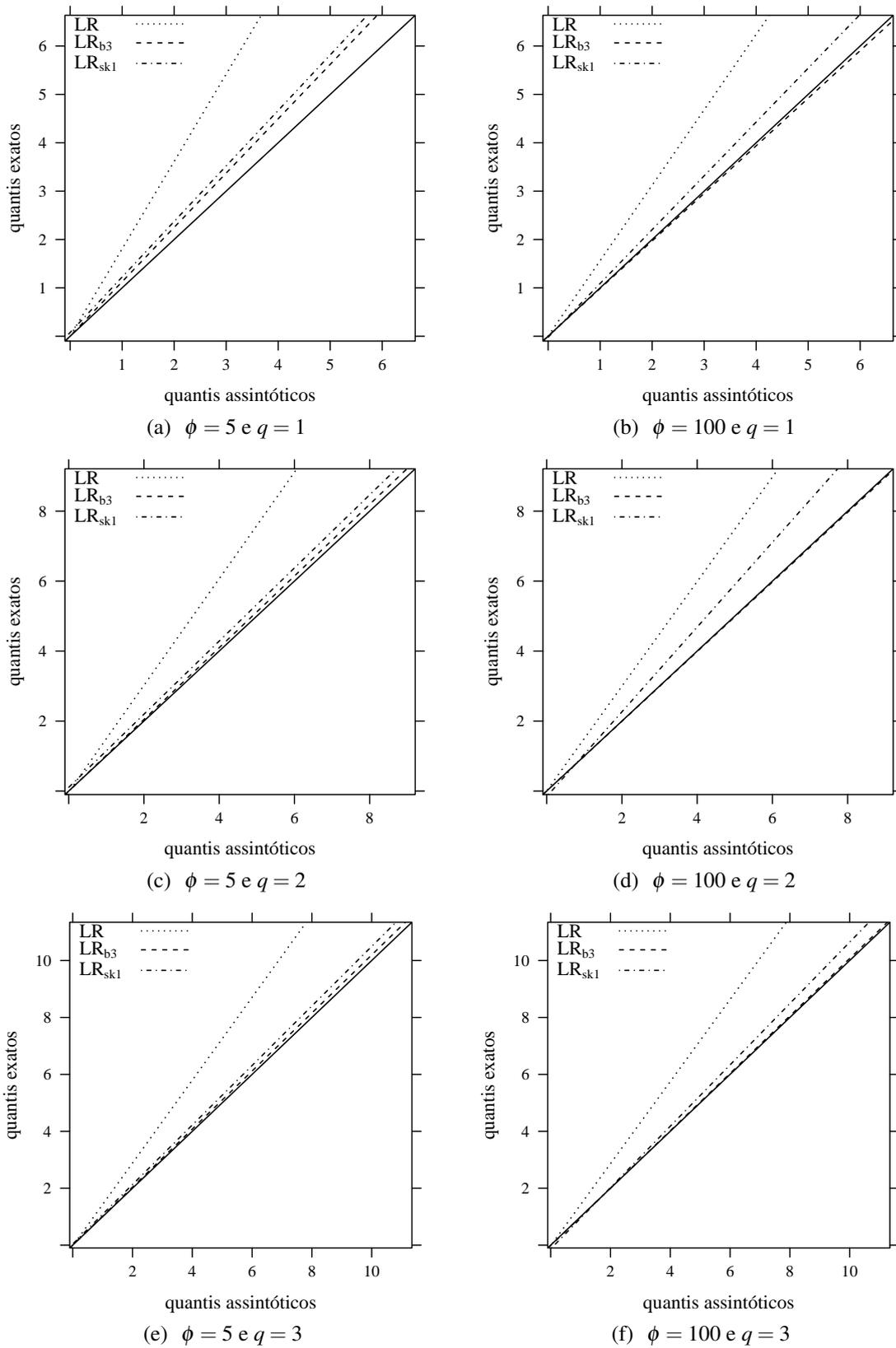


Figura C.1 Gráfico Q-Q para $n = 15$ e diferentes valores de ϕ e q , considerando simulações com 50000 réplicas de Monte Carlo.

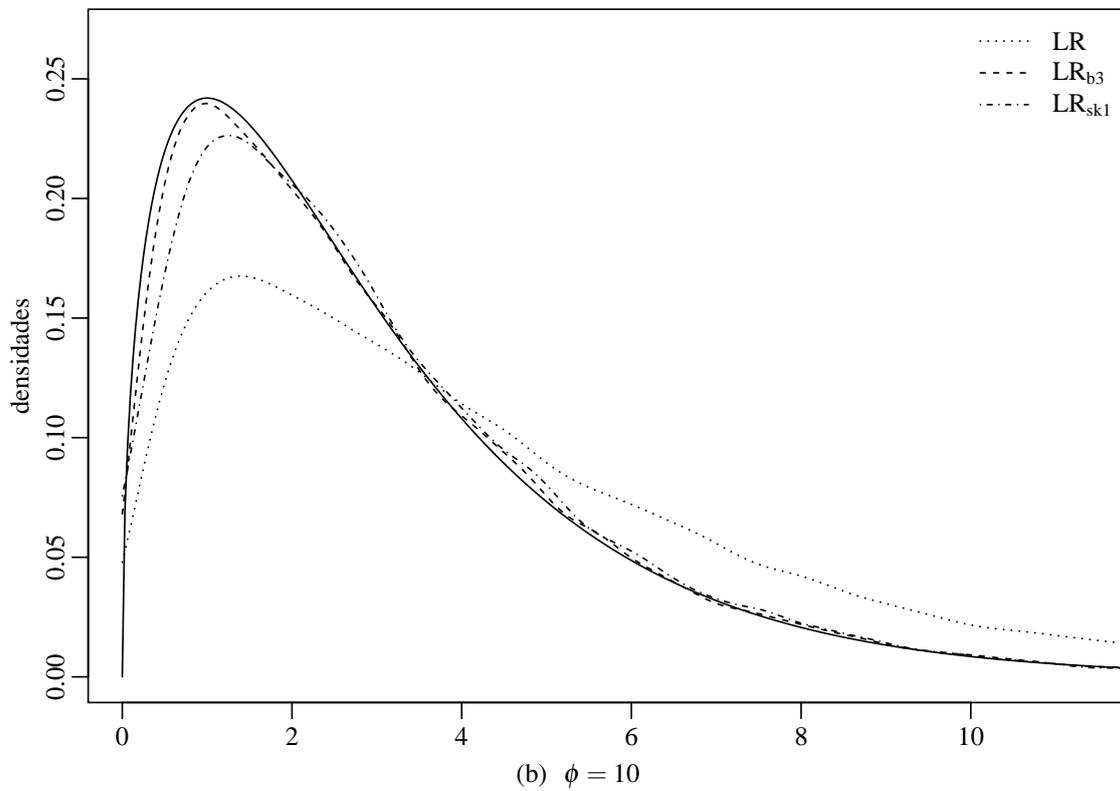
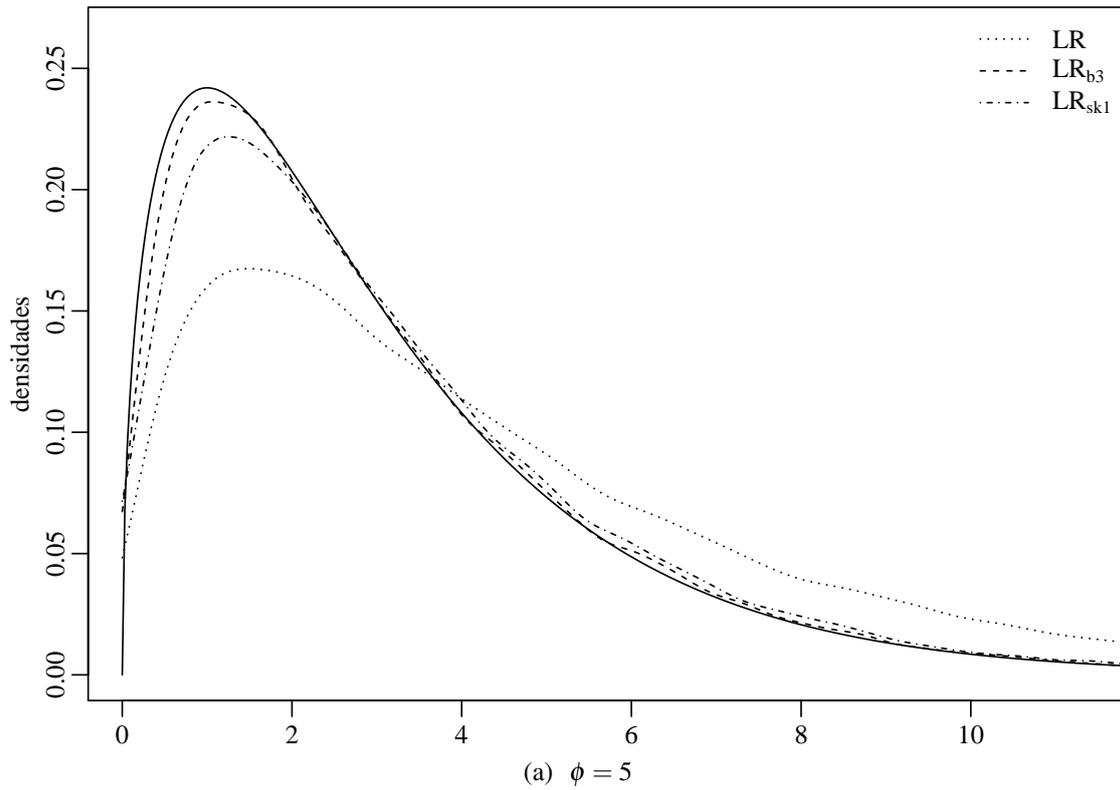
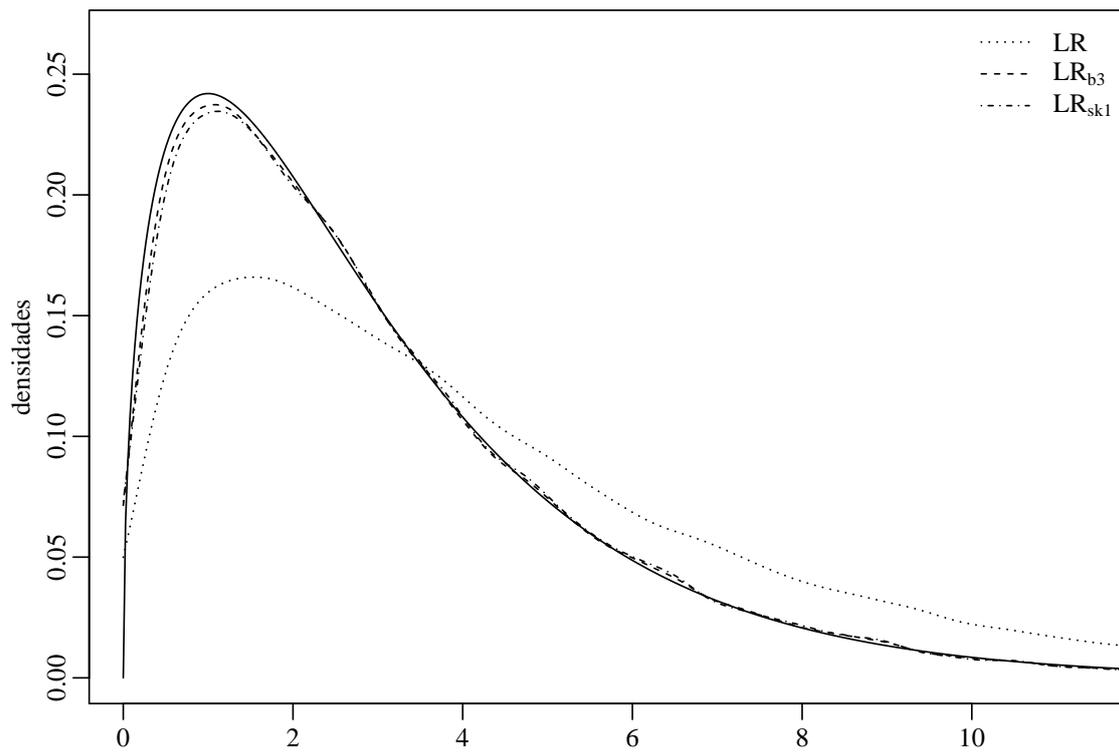
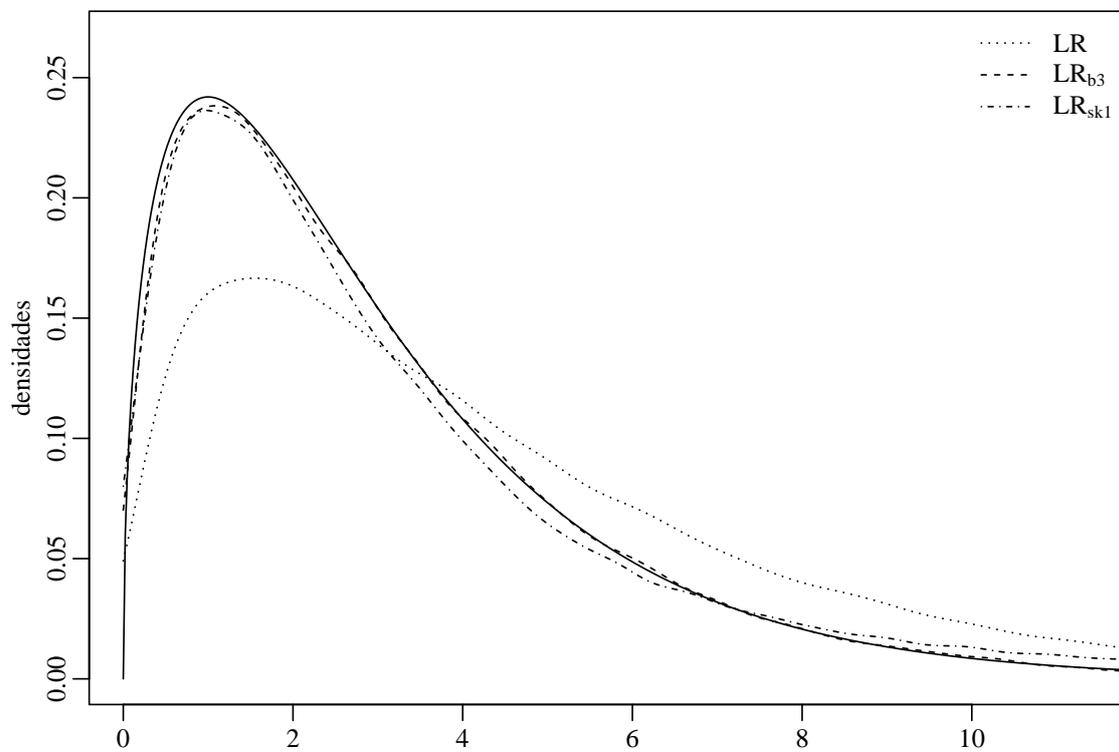


Figura C.2 Densidade χ_q^2 (linha sólida) e densidades nulas estimadas das estatísticas de teste para $n = 15$ e $q = 3$, considerando 50000 réplicas de Monte Carlo.



(a) $\phi = 30$



(b) $\phi = 100$

Figura C.3 Densidade χ_q^2 (linha sólida) e densidades nulas estimadas das estatísticas de teste para $n = 15$ e $q = 3$, considerando 50000 réplicas de Monte Carlo.

Função Escore para o Modelo de Regressão Beta com Dispersão Variável

Este apêndice apresenta a função escore para o modelo de regressão beta com dispersão variável apresentado nas Seções 2.2 e 3.3. A função escore é obtida pela diferenciação da função de log-verossimilhança em relação aos parâmetros desconhecidos. A função escore para o vetor de parâmetros β é dada por

$$U_{\beta}(\beta, \gamma) = X^{\top} \Phi T (y^* - \mu^*),$$

em que X é uma matriz $n \times r$ com as covariáveis da média, $\Phi = \text{diag} \left\{ \frac{1-\sigma_1^2}{\sigma_1^2}, \dots, \frac{1-\sigma_n^2}{\sigma_n^2} \right\}$, $T = \text{diag} \left\{ \frac{1}{g'(\mu_1)}, \dots, \frac{1}{g'(\mu_n)} \right\}$, $y^* = (y_1^*, \dots, y_n^*)^{\top}$, $\mu^* = (\mu_1^*, \dots, \mu_n^*)^{\top}$, $y_t^* = \log \left(\frac{y_t}{1-y_t} \right)$, $\mu_t^* = \psi \left(\mu_t \left(\frac{1-\sigma_t^2}{\sigma_t^2} \right) \right) - \psi \left((1-\mu_t) \left(\frac{1-\sigma_t^2}{\sigma_t^2} \right) \right)$ e $\psi(\cdot)$ é a função digamma, isto é, $\psi(u) = \frac{\partial \log \Gamma(u)}{\partial u}$, para $u > 0$.

Considerando as derivadas da função de log-verossimilhança em relação ao vetor de parâmetros que modelam a dispersão, γ , temos que

$$U_{\gamma}(\beta, \gamma) = Z^{\top} H a,$$

em que Z é uma matriz $n \times s$ com as covariáveis da dispersão, $H = \text{diag} \left\{ \frac{1}{h'(\sigma_1)}, \dots, \frac{1}{h'(\sigma_n)} \right\}$, $a_t = -\frac{2}{\sigma_t^3} \left\{ \mu_t (y_t^* - \mu_t^*) + \log(1 - y_t) - \psi \left((1 - \mu_t) \left(\frac{1 - \sigma_t^2}{\sigma_t^2} \right) \right) + \psi \left(\mu_t \left(\frac{1 - \sigma_t^2}{\sigma_t^2} \right) \right) \right\}$ e $a = (a_1, \dots, a_n)^{\top}$.

Matriz de Informação de Fisher para o Modelo de Regressão Beta com Dispersão Variável

Apresentamos neste apêndice a matriz de informação de Fisher para o modelo de regressão beta com dispersão variável apresentado nas Seções 2.2 e 3.3. A matriz de informação conjunta para β e γ é dada por

$$K(\beta, \gamma) = \begin{pmatrix} K_{(\beta, \beta)} & K_{(\beta, \gamma)} \\ K_{(\gamma, \beta)} & K_{(\gamma, \gamma)} \end{pmatrix}, \quad (\text{E.1})$$

em que $K_{(\beta, \beta)} = X^\top \Phi W X$, $K_{(\beta, \gamma)} = (K_{(\gamma, \beta)})^\top = X^\top C T H Z$ e $K_{(\gamma, \gamma)} = Z^\top D Z$. Ainda, temos que $W = \text{diag}\{w_1, \dots, w_n\}$, $C = \text{diag}\{c_1, \dots, c_n\}$ e $D = \text{diag}\{d_1, \dots, d_n\}$, em que

$$\begin{aligned} w_t &= \frac{(1 - \sigma_t^2)}{\sigma_t^2} \left[\psi' \left(\frac{\mu_t(1 - \sigma_t^2)}{\sigma_t^2} \right) + \psi' \left(\frac{(1 - \mu_t)(1 - \sigma_t^2)}{\sigma_t^2} \right) \right] \frac{1}{[g'(\mu_t)]^2}, \\ c_t &= \frac{(2 - 2\sigma_t^2)}{\sigma_t^5} \left[\mu_t \psi' \left(\frac{\mu_t(1 - \sigma_t^2)}{\sigma_t^2} \right) - (1 - \mu_t) \psi' \left(\frac{(1 - \mu_t)(1 - \sigma_t^2)}{\sigma_t^2} \right) \right], \\ d_t &= \frac{4}{\sigma_t^6} \left[\mu_t^2 \psi' \left(\frac{\mu_t(1 - \sigma_t^2)}{\sigma_t^2} \right) - (1 - \mu_t)^2 \psi' \left(\frac{(1 - \mu_t)(1 - \sigma_t^2)}{\sigma_t^2} \right) - \psi' \left(\frac{(1 - \sigma_t^2)}{\sigma_t^2} \right) \right] \frac{1}{[h'(\sigma_t)]^2}. \end{aligned}$$

Avaliação dos Testes para Dispersão Variável

Neste apêndice apresentamos uma avaliação numérica dos desempenhos dos testes de razão de verossimilhanças e escore para testar se a dispersão é constante no modelo de regressão beta. Ou seja, para o modelo de regressão beta com dispersão variável definido nas equações (2.3), (1.2) e (2.4) na Seção 2.2, temos interesse em testar a seguinte hipótese de dispersão constante:

$$\mathcal{H}_0 : \sigma_1 = \sigma_2 = \dots = \sigma_n = \sigma,$$

ou, equivalentemente,

$$\mathcal{H}_0 : \gamma_i = 0, \quad i = 2, \dots, s,$$

com $z_{t1} = 1$, para $t = 1, \dots, n$.

Para testar essa hipótese nula, considerando que $\theta = (\beta^\top, \gamma^\top)^\top$, as estatísticas de razão de verossimilhanças e escore são dadas, respectivamente, por

$$LR = 2\{\ell(\hat{\theta}) - \ell(\tilde{\theta})\},$$

e

$$S = \tilde{U}_{(s-1)\gamma}^\top \tilde{K}_{(s-1)(s-1)}^{-1} \tilde{U}_{(s-1)\gamma},$$

em que o vetor $\tilde{\theta}$ é o estimador de máxima verossimilhança restrito de θ (sob \mathcal{H}_0), $\hat{\theta}$ é o estimador de máxima verossimilhança irrestrito de θ , $\tilde{U}_{(s-1)\gamma}^\top$ é o vetor que contém os $s - 1$ elementos finais da função escore de γ avaliado em $\tilde{\theta}$ e $\tilde{K}_{(s-1)(s-1)}^{-1}$ é a matriz $(s - 1) \times (s - 1)$ formada pelas últimas $s - 1$ linhas e últimas $s - 1$ colunas da inversa da matriz de informação de Fisher $K(\beta, \gamma)$ avaliada em $\tilde{\theta}$. A obtenção da matriz $K(\beta, \gamma)$ é apresentada em (E.1) no Apêndice E.

Esses testes se baseiam na consistência e normalidade assintótica do estimador de máxima verossimilhança. Sob condições usuais de regularidade e sob \mathcal{H}_0 as estatísticas LR e S convergem em distribuição para $\chi_{(s-1)}^2$, de forma que os testes podem ser realizados usando valores críticos aproximados obtidos como quantis da distribuição $\chi_{(s-1)}^2$. Se as estatísticas de teste forem maiores que o valor crítico selecionado a um nível de significância determinado, então rejeita-se \mathcal{H}_0 e decide-se modelar a dispersão.

Utilizamos simulações de Monte Carlo para avaliar o desempenho, em termos de tamanho e poder, dos testes LR e S . O número de réplicas de Monte Carlo foi fixado em 10000 e todas as simulações foram realizadas utilizando a linguagem de programação R (R Development Core Team, 2009). Para este estudo, consideramos o modelo de regressão beta dado por

$$\text{logit}(\mu_t) = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} \quad \text{e} \quad \text{logit}(\sigma_t) = \gamma_1 + \gamma_2 x_{t2} + \gamma_3 x_{t3},$$

em que $t = 1, \dots, n$ e os valores das covariáveis são determinados aleatoriamente da distribuição $\mathcal{U}(0,1)$ e são mantidos constantes durante todo o experimento. Os valores dos parâmetros do submodelo da média são $\beta_1 = -1$, $\beta_2 = 1$ e $\beta_3 = -1$ e o intercepto do submodelo da dispersão é $\gamma_1 = -1$.

Para obter a taxa de rejeição nula dos testes consideramos tamanhos amostrais iguais a 25, 50 e 100 e os seguintes níveis nominais para os testes: $\alpha = 10\%$, 5% e 1% . Consideramos a hipótese nula $\mathcal{H}_0 : \gamma_2 = \gamma_3 = 0$ contra a hipótese alternativa bilateral. A Tabela F.1 apresenta os resultados da simulação para o tamanho dos testes.

Tabela F.1 Taxas de rejeição nula (percentuais) do teste $\mathcal{H}_0 : \gamma_2 = \gamma_3 = 0$.

		$\alpha = 10\%$			$\alpha = 5\%$			$\alpha = 1\%$		
		25	50	100	25	50	100	25	50	100
Estat	n									
	LR	17.47	12.86	11.42	10.38	6.68	6.00	3.11	1.64	1.17
	S	8.16	8.77	9.67	3.78	4.08	4.66	0.68	0.73	0.88

Os resultados da Tabela F.1 evidenciam a característica liberal do teste da razão de verossimilhanças. Para o caso em que $\alpha = 1\%$ e $n = 25$, por exemplo, percebe-se que a taxa de rejeição nula do teste LR é mais de três vezes maior do que o nível nominal do teste. Por outro lado, o teste S mostra-se mais adequado em termos de controle da probabilidade de erro tipo I. Suas taxas de rejeição nula mostram-se sempre mais próximas ao nível nominal quando comparadas às taxas do teste LR .

Para a avaliação do poder dos testes também consideramos o interesse em testar $\mathcal{H}_0 : \gamma_2 = \gamma_3 = 0$. Para obtenção das taxas de rejeição não-nula consideramos dois casos, gerando os dados sob os seguintes valores de γ_2 e γ_3 : (i) $\gamma_2 = -\gamma_3 = \delta$ e (ii) $\gamma_2 = \delta$, $\gamma_3 = 0$, para diferentes valores de δ . Os resultados para esses dois casos estão apresentados na Tabela F.2 e na Tabela F.3. O segundo caso, que considera $\gamma_3 = 0$, tem o objetivo de avaliar o impacto do erro cometido ao se considerar um modelo irrestrito com dimensão maior do que o modelo verdadeiro. Em situações práticas, ao não conhecer o modelo verdadeiro, essa situação pode se tornar bastante comum.

Como já era de se esperar, pela tendência liberal do teste da razão de verossimilhanças, o teste LR se mostra mais poderoso do que o teste S . Essa diferença entre as taxas de rejeição não-nula dos testes LR e S é mais evidente nas amostras menores. As taxas de rejeição não-nula apresentadas na Tabela F.3 mostram-se inferiores às taxas apresentadas na Tabela F.2. Isso mostra a perda de desempenho em termos de poder dos testes quando é considerado um modelo irrestrito maior do que o modelo verdadeiro.

Tabela F.2 Taxas de rejeição não-nula (percentuais) dos testes, considerando a hipótese alternativa $\gamma_2 = -\gamma_3 = \delta$.

α	δ	$n = 25$		$n = 50$		$n = 100$	
		<i>LR</i>	<i>S</i>	<i>LR</i>	<i>S</i>	<i>LR</i>	<i>S</i>
10%	0.5	25.77	15.95	41.64	34.51	60.68	56.89
	1.0	51.65	37.05	88.17	81.40	98.71	97.68
	1.5	76.89	60.53	99.33	97.75	100.0	100.0
	2.0	92.05	78.41	99.99	99.79	100.0	100.0
	3.0	99.50	94.39	100.0	100.0	100.0	100.0
5%	0.5	17.02	9.16	30.29	23.07	48.23	44.82
	1.0	40.02	26.73	81.26	71.69	97.28	95.39
	1.5	67.88	49.10	98.78	95.45	100.0	99.99
	2.0	87.75	69.00	99.93	99.54	100.0	100.0
	3.0	99.05	90.16	100.0	100.0	100.0	100.0
1%	0.5	6.24	2.86	12.92	8.72	26.54	23.29
	1.0	21.42	11.60	62.14	48.33	91.45	86.06
	1.5	48.50	29.25	95.20	85.67	99.99	99.57
	2.0	75.14	48.69	99.75	97.42	100.0	99.99
	3.0	97.49	76.75	100.0	99.93	100.0	100.0

Tabela F.3 Taxas de rejeição não-nula (percentuais) dos testes, considerando a hipótese alternativa $\gamma_2 = \delta, \gamma_3 = 0$.

α	δ	$n = 25$		$n = 50$		$n = 100$	
		<i>LR</i>	<i>S</i>	<i>LR</i>	<i>S</i>	<i>LR</i>	<i>S</i>
10%	0.5	20.22	10.55	23.34	19.00	33.82	31.61
	1.0	29.39	19.05	52.82	46.60	78.54	75.65
	1.5	42.59	30.33	80.58	74.65	97.10	95.87
	2.0	57.77	44.84	94.92	91.70	99.75	99.59
	3.0	79.52	67.15	99.71	99.21	100.0	99.98
5%	0.5	12.15	5.20	15.22	11.59	23.41	21.23
	1.0	20.27	11.73	40.92	34.53	68.88	64.82
	1.5	31.43	19.77	71.12	63.31	94.60	92.56
	2.0	46.24	32.68	91.26	85.38	99.58	99.07
	3.0	70.47	55.63	99.21	97.65	100.0	99.97
1%	0.5	3.71	1.13	5.07	3.46	8.95	8.44
	1.0	7.90	3.77	20.71	15.61	45.28	41.84
	1.5	17.91	7.31	48.72	38.98	85.34	80.99
	2.0	26.15	15.34	77.46	66.14	98.20	96.42
	3.0	48.79	31.61	97.01	90.99	99.97	99.84

APÊNDICE G

Análise dos Resíduos (implementação em R)

Este apêndice apresenta a função implementada em linguagem R para análise dos resíduos do modelo de regressão beta com dispersão variável, dada na Seção 2.2.

```
#####  
#  
# PROGRAMA: Essa funcao usa um modelo de regressao beta com dispersao variavel  
#           ajustado como o pacote GAMLSS e traca graficos de analise de diagnostico.  
#           O usuario pode escolher, na chamada da funcao, qual residuo deseja.  
#           As opcoes de residuos sao dados abaixo:  
#           - resid=0 -> residuo quantil aleatorizado (padrao da Gamlss)  
#           - resid=1 -> residuo padronizado ordinario (Ferrari e Cribari-Neto (2004))  
#           - resid=2 -> residuo ponderado padronizado 2 (Espinheira et al (2008))  
#           O usuario pode escolher imprimir os graficos diretamente em pdf. Para isso,  
#           basta utilizar pdf=1 na chamada da funcao.  
#  
# OBS:      Posteriormente sera implementada a distancia de Cook.  
#           Ao final ha um exemplo de utilizacao desta funcao.  
#  
# AUTOR:    Fábio Mariano Bayer  
#  
# E-MAIL:   fabiobayer@gmail.com  
#  
# DATA:    Outubro/2011  
#  
#####  
  
gamlss.diag<-function(fit1,resid=0,pdf=0)  
{  
  y1<-fit1$y  
  muhat1<-fit1$mu.fv  
  sigmahat1<-(fit1$sigma.fv)  
  sigmahat21<-sigmahat1^2  
  phihat1<-((1-sigmahat21)/sigmahat21)  
  
  X<-fit1$mu.x  
  X<-as.matrix(X)  
  Xs<-fit1$sigma.x  
  Xs<-as.matrix(Xs)  
  
  n <- length(y1)  
  j<-seq(n)  
  
  yhatrb1<-fit1$mu.fv  
  
  if(resid==0)  
  {  
    print("Usando residuo quantil aleatorizado (padrao da GAMLSS)",quote=F)  
    rqbe<-fit1$residuals #residuals(fit1, what = "z-scores")  
  }  
  if(resid==1)  
  {  
    print("Usando residuo padronizado ordinario (Ferrari e Cribari-Neto (2004))",quote=F)  
    vary1<-muhat1*(1-muhat1)/(1+phihat1)  
  }  
}
```

```

    rqbe<- (y1-yhatrb1)/sqrt(vary1)
  }
if(resid==2)
{
  print("Usando residuo ponderado padronizado 2 (Espinheira, Ferrari and
        Cribari-Neto(2008))",quote=F)
  ystar1<-log(y1/(1-y1))
  a1 <- muhat1*phihat1
  b1 <- (1-muhat1)*phihat1
  mustar1 <- digamma(a1) - digamma(b1)
  vt1<- trigamma(a1) + trigamma(b1)
  vt1<-as.vector(vt1)
  wt1<-phihat1*vt1*((muhat1^2)*(1-muhat1)^2)
  wt1<-as.vector(wt1)
  W1<-diag(wt1)
  PHI1<-diag(phihat1)
  Hstar1<-((W1**PHI1)^(1/2))**X**%solve(t(X)**PHI1**W1**X)**t(X)**(PHI1**W1)^(1/2)
  htt1<-diag(Hstar1)
  rqbe<- (ystar1-mustar1)/sqrt(vt1*(1-htt1))
  rqbe<-as.vector(rqbe)
}

sim = 100 # numero de replicas de MC para envelope simulado
e<- matrix(0,n,sim)

e1<- numeric(n)
e2<- numeric(n)
e3<- numeric(n)

for(i in 1:sim) #laco de Monte Carlo
{
  resp <- rbeta(n, muhat1*phihat1, (1-muhat1)*phihat1)
  resp<-ifelse(resp==1.00000000,0.99999999,resp)
  con<-gamlss.control(n.cyc=30,mu.step=0.05,sigma.step=0.05,trace=FALSE)
  fit <- gamlss(resp~(X-1),sigma.formula=~(Xs-1),family=BE,method=RS(),
               silent=FALSE,control=con)

if(resid==0)
{
  rqbebeta<-fit$residuals
}
if(resid==1)
{
  y<-fit$y
  muhat<-fit$mu.fv
  sigmahat<-(fit$sigma.fv)
  sigmahat2<-sigmahat^2
  phihat<-((1-sigmahat2)/sigmahat2)

  vary<-muhat*(1-muhat)/(1+phihat)
  rqbebeta<- (y-fit$mu.fv)/sqrt(vary)
}
if(resid==2)
{
  y<-fit$y
  muhat<-fit$mu.fv
  sigmahat<-fit$sigma.fv
  sigmahat2<-sigmahat^2
  phihat<-((1-sigmahat2)/(sigmahat2))
  ystar<-log(y/(1-y))
  a <- muhat*phihat
  b <- (1-muhat)*phihat
  mustar <- digamma(a) - digamma(b)
  vt<- trigamma(a) + trigamma(b)
  vt<-as.vector(vt)
  wt<-phihat*vt*((muhat^2)*(1-muhat)^2)

```

```

    wt<-as.vector(wt)
    W<-diag(wt)
    PHI<-diag(phihat)
    Hstar<-((W%*%PHI)^(1/2))%*%X%*%solve(t(X)%*%PHI%*%W%*%X)%*%t(X)%*%(PHI%*%W)^(1/2)
    htt<-diag(Hstar)
    rqbebeta<-(ystar-mustar)/sqrt(vt*(1-htt))
    rqbebeta<-as.vector(rqbebeta)
  }

  e[,i]<-sort(rqbebeta)
}

for(i in 1:n)
{
  quant<-5
  eo<-sort(e[i,])
  q5<-round(quant*sim/100)
  if(q5<1) q5<-1
  e1[i] <- eo[q5]
  e2[i] <- mean(eo)
  e3[i] <- eo[round((100-quant)*sim/100)]
}

# parametros para os graficos
# par(pty="m") # m para grafico de tamanho maximo
par(family="Times") # fonte dos graficos
par(mar=c(2.7, 2.5, 2, 1)) # margens c(baixo,esq,cima,direia)
par(mgp=c(1.5, 0.8, 0))

plot(yhatrb1,y1,main="Valores observados vs. Valores ajustados",
      xlim=c(0,1),ylim=c(0,1),xlab="",ylab="", pch = "+")
lines(c(-0.2,1.2),c(-0.2,1.2),lty=2)

t<-seq(-5,n+6,by=1)
plot(yhatrb1,rqbe,main="Residuos vs. Valores ajustados",xlab="",ylab="", pch = "+",
      ylim=c(-4,4))
lines(t,rep(-3,n+12),lty=2,col=1)
lines(t,rep(3,n+12),lty=2,col=1)
lines(t,rep(-2,n+12),lty=3,col=1)
lines(t,rep(2,n+12),lty=3,col=1)

plot(j, rqbe,main="Residuos vs. Índices",xlab="",ylab="", pch = "+",ylim=c(-4,4))
lines(t,rep(-3,n+12),lty=2,col=1)
lines(t,rep(3,n+12),lty=2,col=1)
lines(t,rep(-2,n+12),lty=3,col=1)
lines(t,rep(2,n+12),lty=3,col=1)

#par(pty="s") # s para grafico quadrado (squared)

faixa<-range(rqbe,e1,e3)
qqnorm(rqbe, pch = "+",ylim=faixa,main="Gráfico normal de probabilidade",xlab="",ylab="")
par(new=TRUE)
qqnorm(e1, axes=F, type="l",ylim=faixa,main=" ",xlab="",ylab="")
par(new=TRUE)
qqnorm(e2, axes=F, type="l",lty=2,ylim=faixa,main=" ",xlab="",ylab="")
par(new=TRUE)
qqnorm(e3, axes=F, type="l",ylim=faixa,main=" ",xlab="",ylab="")

if(pdf==1)
{
  mar_b<-2.5
  mar_e<-2.5
  mar_c<-0.5
  mar_d<-0.5
  dist_text<-1.5
  dist_tick<-0.5
}

```

```

pdf(file = "obs_v_fit.pdf",width = 4, height = 4,family = "Times")
par(mar=c(mar_b, mar_e, mar_c, mar_d)) # margens c(baixo,esq,cima,direia)
par(mgp=c(dist_text, dist_tick, 0))
plot(yhatrb1,y1,main=" ",xlab="valores ajustados",ylab="valores observados",
      xlim=c(0,1),ylim=c(0,1),pch = "+")
lines(c(-0.2,1.2),c(-0.2,1.2),lty=2)
dev.off()

pdf(file = "resid_v_fit.pdf",width = 4, height = 4,family = "Times")
par(mar=c(mar_b, mar_e, mar_c, mar_d)) # margens c(baixo,esq,cima,direia)
par(mgp=c(dist_text, dist_tick, 0))
plot(yhatrb1,rqbe,main=" ",xlab="valores ajustados",ylab="resíduos", pch = "+",
      ylim=c(-4,4))
lines(t,rep(-3,n+12),lty=2,col=1)
lines(t,rep(3,n+12),lty=2,col=1)
lines(t,rep(-2,n+12),lty=3,col=1)
lines(t,rep(2,n+12),lty=3,col=1)
dev.off()

pdf(file = "resid_v_ind.pdf",width = 4, height = 4,family = "Times")
par(mar=c(mar_b, mar_e, mar_c, mar_d)) # margens c(baixo,esq,cima,direia)
par(mgp=c(dist_text, dist_tick, 0))
plot(j, rqbe,main=" ",xlab="índices",ylab="resíduos", pch = "+",ylim=c(-4,4))
lines(t,rep(-3,n+12),lty=2,col=1)
lines(t,rep(3,n+12),lty=2,col=1)
lines(t,rep(-2,n+12),lty=3,col=1)
lines(t,rep(2,n+12),lty=3,col=1)
dev.off()

pdf(file = "envelope.pdf",width = 4, height = 4,family = "Times")
par(mar=c(mar_b, mar_e, mar_c, mar_d)) # margens c(baixo,esq,cima,direia)
par(mgp=c(dist_text, dist_tick, 0))
qqnorm(rqbe, pch = "+",ylim=faixa,main="",xlab="quantis normais",
       ylab="quantis empíricos")
par(new=TRUE)
qqnorm(e1,axes=F,type="l",ylim=faixa,main=" ",xlab="",ylab="")
par(new=TRUE)
qqnorm(e2,axes=F,type="l",lty=2,ylim=faixa,main=" ",xlab="",ylab="")
par(new=TRUE)
qqnorm(e3,axes=F,type="l",ylim=faixa,main=" ",xlab="",ylab="")
dev.off()

} # fim do pdf
} # fim da funcao

#####
#
#           EXEMPLO DE USO DA FUNCAO CRIADA - gamlss.diag()
#
#####

# library(gamlss) # chamando o pacote GAMLSS
# source("gamlss.diag.r") # lendo a funcao criada que esta no arquivo gamlss.diag.r

## Lendo dados de dislexia
# data<-read.table("dislex2.txt",h=T)
# attach(data) # transformando colunas em objetos independentes

## estimando o modelo com dispersao variavel usando o pacote GAMLSS
# fit = gamlss(resp~QI+dislex+intera, sigma.formula=~QI+dislex+QI2, family=BE)

## analise de diagnostico do modelo usando a funcao criada
# gamlss.diag(fit,resid=2,pdf=0) # com residuo pond. pad. 2 e sem gerar pdf
# gamlss.diag(fit,resid=0,pdf=1) # com residuo quantil aleatorizado e gerando pdf

```

APÊNDICE H

Tabelas de Resultados Complementares

Este apêndice apresenta resultados complementares do estudo de simulação desenvolvido no Capítulo 3. Essas tabelas apresentam os desempenhos dos critérios de seleção nos modelos de regressão beta com dispersão variável considerando grande dispersão. Estes resultados se assemelham bastante com aqueles contidos nas tabelas apresentadas na Seção 3.4.2 que considera pequena dispersão.

Tabela H.1 Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, selecionando regressores da média e dispersão conjuntamente em um modelo facilmente identificável (modelo (3.12)).

	$n = 25$			$n = 30$			$n = 40$			$n = 50$		
	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$
AIC	384	60	556	437	107	456	515	133	352	436	205	359
AICc	799	57	144	781	80	139	741	120	139	615	202	183
SIC	695	42	263	787	53	160	875	64	61	850	94	56
SICc	968	14	18	967	22	11	961	31	8	941	48	11
HQ	473	64	463	572	94	334	689	113	198	635	183	182
HQc	871	41	88	872	60	68	881	70	49	806	126	68
BQCV	889	71	40	860	90	50	715	182	103	476	287	237
632QCV	770	129	101	693	184	123	482	257	261	264	279	457
EIC _{1p}	971	23	6	956	31	13	939	49	12	910	59	31
EIC _{2p}	987	10	3	978	19	3	947	42	11	859	109	32
EIC _{3p}	895	75	30	831	121	48	655	188	157	516	224	260
EIC _{4p}	431	3	566	487	9	504	723	21	256	899	46	55
EIC _{5p}	232	18	750	370	48	582	644	93	263	765	117	118
EIC _{1np}	993	7	0	979	17	4	946	45	9	813	154	33
EIC _{2np}	999	0	0	992	7	1	984	14	2	914	69	17
EIC _{3np}	651	55	294	712	130	158	657	171	172	491	237	272
EIC _{4np}	998	2	0	993	6	1	969	28	3	873	110	17
EIC _{5np}	408	46	546	548	134	318	573	217	210	414	241	345
BCV	1000	0	0	1000	0	0	994	6	0	958	39	3
632CV	996	4	0	992	7	1	974	23	3	887	96	17

Tabela H.2 Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, selecionando regressores da média e dispersão conjuntamente em um modelo fracamente identificável (modelo (3.13)).

	$n = 25$			$n = 30$			$n = 40$			$n = 50$		
	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$	$< k_0$	$= k_0$	$> k_0$
AIC	507	47	446	590	64	346	655	93	252	682	86	232
AICc	899	38	63	887	32	81	866	60	74	840	66	94
SIC	810	30	160	902	17	81	956	21	23	967	17	16
SICc	986	4	10	994	3	3	997	3	0	991	5	4
HQ	616	44	340	732	43	225	830	59	111	864	49	87
HQc	938	26	36	946	18	36	955	24	21	946	30	24
BQCV	894	59	47	860	64	76	778	105	117	677	140	183
632QCV	742	139	119	674	139	187	526	196	278	439	174	387
EIC1 _p	977	17	6	971	15	14	961	21	18	935	37	28
EIC2 _p	990	6	4	985	9	6	978	11	11	949	33	18
EIC3 _p	903	65	32	848	83	69	790	88	122	751	92	157
EIC4 _p	604	4	392	719	6	275	870	9	121	949	26	25
EIC5 _p	158	2	840	225	7	768	721	55	224	799	78	123
EIC1 _{np}	995	2	3	985	11	4	970	24	6	932	43	25
EIC2 _{np}	997	2	1	996	3	1	993	6	1	967	21	12
EIC3 _{np}	747	72	181	816	79	105	750	109	141	698	114	188
EIC4 _{np}	997	2	1	993	5	2	989	10	1	961	28	11
EIC5 _{np}	526	42	432	729	81	190	718	122	160	663	137	200
BCV	1000	0	0	1000	0	0	999	1	0	997	3	0
632CV	997	2	1	994	4	2	987	12	1	962	26	12

Tabela H.3 Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, considerando conhecida a estrutura de regressão da dispersão e selecionando regressores da média em um modelo facilmente identificável (modelo (3.12)).

	$n = 25$			$n = 30$			$n = 40$			$n = 50$		
	$< r_0$	$= r_0$	$> r_0$	$< r_0$	$= r_0$	$> r_0$	$< r_0$	$= r_0$	$> r_0$	$< r_0$	$= r_0$	$> r_0$
AIC	316	261	423	317	298	385	264	415	321	232	450	318
AICc	633	268	99	525	345	130	393	463	144	329	496	175
SIC	567	251	182	571	297	132	530	388	82	504	432	64
SICc	844	130	26	794	189	17	704	274	22	634	341	25
HQ	370	286	344	404	313	283	362	441	197	356	463	181
HQc	705	230	65	632	293	75	520	416	64	465	457	78
BQCV	752	247	1	593	393	14	374	531	95	254	559	187
632QCV	590	388	22	438	498	64	238	533	229	149	479	372
EIC1 _p	974	26	0	946	51	3	940	48	12	918	66	16
EIC2 _p	924	76	0	824	176	0	634	351	15	492	469	39
EIC3 _p	669	327	4	505	448	47	362	487	151	278	497	225
EIC4 _p	735	8	257	769	15	216	966	14	20	972	19	9
EIC5 _p	539	9	452	841	19	140	955	22	23	955	23	22
EIC1 _{np}	983	16	1	957	43	0	799	192	9	604	352	44
EIC2 _{np}	983	16	1	965	35	0	876	118	6	744	239	17
EIC3 _{np}	584	300	116	540	370	90	392	478	130	328	493	179
EIC4 _{np}	984	15	1	965	35	0	869	125	6	726	257	17
EIC5 _{np}	578	268	154	527	355	118	379	440	181	315	427	258
BCV	986	14	0	974	26	0	916	82	2	819	168	13
632CV	985	15	0	966	34	0	856	138	6	689	285	26

Tabela H.4 Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, considerando conhecida a estrutura de regressão da dispersão e selecionando regressores da média em um modelo fracamente identificável (modelo (3.13)).

	$n = 25$			$n = 30$			$n = 40$			$n = 50$		
	$< r_0$	$= r_0$	$> r_0$	$< r_0$	$= r_0$	$> r_0$	$< r_0$	$= r_0$	$> r_0$	$< r_0$	$= r_0$	$> r_0$
AIC	530	110	360	581	109	310	631	121	248	633	135	232
AICc	858	72	70	838	75	87	796	105	99	763	118	119
SIC	803	63	134	849	60	91	896	60	44	885	80	35
SICc	958	30	12	959	30	11	968	27	5	946	45	9
HQ	621	97	282	697	93	210	754	101	145	772	107	121
HQc	910	51	39	900	55	45	900	63	37	861	95	44
BQCV	928	71	1	883	100	17	786	154	60	677	171	152
632QCV	857	134	9	766	182	52	610	194	196	491	183	326
EIC1 _p	976	23	1	964	29	7	932	58	10	905	72	23
EIC2 _p	986	14	0	965	32	3	949	48	3	883	93	24
EIC3 _p	901	97	2	828	126	46	754	132	114	716	135	149
EIC4 _p	797	12	191	827	12	161	947	38	15	940	47	13
EIC5 _p	307	9	684	663	33	304	865	64	71	867	63	70
EIC1 _{np}	985	13	2	986	14	0	952	48	0	910	77	13
EIC2 _{np}	985	13	2	989	11	0	969	31	0	949	48	3
EIC3 _{np}	843	89	68	847	110	43	793	130	77	742	137	121
EIC4 _{np}	985	13	2	988	12	0	965	35	0	940	54	6
EIC5 _{np}	790	94	116	820	124	56	768	143	89	693	168	139
BCV	986	13	1	990	10	0	981	19	0	966	33	1
632CV	986	13	1	990	10	0	964	36	0	939	53	8

Tabela H.5 Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, considerando conhecida a estrutura de regressão da média e selecionando regressores da dispersão em um modelo facilmente identificável (modelo (3.12)).

	$n = 25$			$n = 30$			$n = 40$			$n = 50$		
	$< s_0$	$= s_0$	$> s_0$	$< s_0$	$= s_0$	$> s_0$	$< s_0$	$= s_0$	$> s_0$	$< s_0$	$= s_0$	$> s_0$
AIC	540	121	339	574	138	288	593	197	210	590	203	207
AICc	853	84	63	802	123	75	763	163	74	741	163	96
SIC	794	89	117	828	99	73	861	105	34	893	90	17
SICc	974	20	6	964	28	8	945	50	5	942	56	2
HQ	629	115	256	675	131	194	723	167	110	770	137	93
HQc	896	66	38	880	84	36	856	114	30	869	109	22
BQCV	981	19	0	939	61	0	811	176	13	743	196	61
632QCV	905	95	0	830	163	7	659	274	67	553	276	171
EIC1 _p	996	4	0	986	14	0	953	37	10	923	55	22
EIC2 _p	1000	0	0	1000	0	0	969	31	0	944	55	1
EIC3 _p	925	75	0	823	173	4	721	212	67	679	197	124
EIC4 _p	819	6	175	846	1	153	916	20	64	944	33	23
EIC5 _p	336	25	639	544	52	404	684	87	229	687	87	226
EIC1 _{np}	1000	0	0	999	1	0	969	31	0	929	65	6
EIC2 _{np}	1000	0	0	1000	0	0	990	10	0	973	27	0
EIC3 _{np}	777	113	110	779	147	74	683	213	104	637	195	168
EIC4 _{np}	1000	0	0	1000	0	0	992	8	0	961	38	1
EIC5 _{np}	718	105	177	764	139	97	653	217	130	604	202	194
BCV	1000	0	0	1000	0	0	994	6	0	986	14	0
632CV	1000	0	0	1000	0	0	987	13	0	956	44	0

Tabela H.6 Frequências das ordens dos modelos selecionados em 1000 realizações para os modelos com dispersão variável, considerando conhecida a estrutura de regressão da média e selecionando regressores da dispersão em um modelo fracamente identificável (modelo (3.13)).

	$n = 25$			$n = 30$			$n = 40$			$n = 50$		
	$< s_0$	$= s_0$	$> s_0$	$< s_0$	$= s_0$	$> s_0$	$< s_0$	$= s_0$	$> s_0$	$< s_0$	$= s_0$	$> s_0$
AIC	594	82	324	676	89	235	669	121	210	714	101	185
AICc	878	60	62	886	59	55	825	96	79	825	89	86
SIC	820	56	124	902	42	56	918	48	34	947	36	17
SICc	983	8	9	976	17	7	973	25	2	984	15	1
HQ	672	83	245	791	79	130	794	92	114	842	80	78
HQc	932	31	37	945	33	22	913	56	31	925	49	26
BQCV	992	8	0	967	32	1	896	92	12	852	101	47
632QCV	957	43	0	909	87	4	750	177	73	673	172	155
EIC1 _p	998	2	0	994	6	0	961	36	3	945	37	18
EIC2 _p	1000	0	0	1000	0	0	981	18	1	980	19	1
EIC3 _p	944	56	0	901	92	7	783	140	77	786	104	110
EIC4 _p	871	3	126	917	1	82	953	11	36	972	20	8
EIC5 _p	316	19	665	613	39	348	737	79	184	724	94	182
EIC1 _{np}	1000	0	0	999	1	0	989	11	0	962	37	1
EIC2 _{np}	1000	0	0	1000	0	0	997	3	0	987	13	0
EIC3 _{np}	824	71	105	854	95	51	779	136	85	744	120	136
EIC4 _{np}	1000	0	0	1000	0	0	997	3	0	981	19	0
EIC5 _{np}	745	76	179	843	95	62	756	132	112	715	151	134
BCV	1000	0	0	1000	0	0	997	3	0	990	10	0
632CV	1000	0	0	1000	0	0	997	3	0	979	21	0

Referências Bibliográficas

- Akaike H (1973). “Information theory and an extension of the maximum likelihood principle.” In PB N, C F (eds.), *Proc. of the 2nd Int. Symp. on Information Theory*, pp. 267–281.
- Akaike H (1974). “A new look at the statistical model identification.” *IEEE Transactions on Automatic Control*, **19**(6), 716–723.
- Akaike H (1978). “A Bayesian analysis of the minimum AIC procedure.” *Annals of the Institute of Statistical Mathematics*, **30**(1), 9–14.
- Allen D (1974). “The relationship between variable selection and data augmentation and a method for prediction.” *Technometrics*, **16**, 125–127.
- Barndorff-Nielsen OE, Cox DR (1984). “Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator.” *Journal of the Royal Statistical Society, B*, **46**(3), pp. 483–495.
- Bartlett MS (1937). “Properties of sufficiency and statistical tests.” *Royal Society of London Proceedings Series A*, **160**, 268–282.
- Bengtsson T, Cavanaugh J (2006). “An improved Akaike information criterion for state-space model selection.” *Computational Statistics & Data Analysis*, **50**(10), 2635–2654.
- Brehm J, Gates S (1993). “Donut shops and speed traps: Evaluating models of supervision on police behavior.” *American Journal of Political Science*, **37**(2), 555–581.
- Breiman L, Spector P (1992). “Submodel selection and evaluation in regression: The X-random case.” *International Statistical Review*, **60**, 291–319.
- Cavanaugh J (1997). “Unifying the derivations for the Akaike and corrected Akaike information criteria.” *Statistics & Probability Letters*, **33**(2), 201–208.
- Cavanaugh JE, Shumway RH (1997). “A Bootstrap Variant of AIC for State-Space Model Selection.” *Statistica Sinica*, **7**, 473–496.
- Cordeiro GM (1993). “General matrix formulae for computing Bartlett corrections.” *Statistics & Probability Letters*, **16**(1), 11–18.
- Cribari-Neto F, Zeileis A (2010). “Beta Regression in R.” *Journal of Statistical Software*, **34**(2).

- Cuervo-Cepeda E, Gamerman D (2004). “Bayesian modeling of joint regressions for the mean and covariance matrix.” *Biometrical Journal*, **46**(4), 430–440.
- Davies S, Neath A, Cavanaugh J (2005). “Cross validation model selection criteria for linear regression based on the Kullback-Leibler discrepancy.” *Statistical Methodology*, **2**(4), 249–266.
- Doornik J (2007). *An Object-Oriented Matrix Language Ox 5*. London: Timberlake Consultants Press. URL <http://www.doornik.com/>.
- Efron B (1979). “Bootstrap methods: Another look at the Jackknife.” *The Annals of Statistics*, **7**(1), 1–26.
- Efron B (1983). “Estimating the error rate of a prediction rule: Improvement on cross-validation.” *Journal of the American Statistical Association*, **78**(382), 316–331.
- Efron B (1986). “How biased is the apparent error rate of a prediction rule?” *Journal of the American Statistical Association*, **81**(393), 461–470.
- Efron B (1987). “Better Bootstrap confidence intervals.” *Journal of the American Statistical Association*, **82**(397), 171–185.
- Efron B, Tibshirani R (1997). “Improvements on cross-validation: The .632+ Bootstrap method.” *Journal of the American Statistical Association*, **92**(438), 548–560.
- Espinheira PL (2007). *Regressão beta*. Tese de Doutorado, Instituto de Matemática e Estatística, Universidade de São Paulo (USP).
- Espinheira PL, Ferrari SLP, Cribari-Neto F (2008a). “Influence diagnostics in beta regression.” *Computational Statistics & Data Analysis*, **52**(9), 4417–4431.
- Espinheira PL, Ferrari SLP, Cribari-Neto F (2008b). “On beta regression residuals.” *Journal of Applied Statistics*, **35**(4), 407–419.
- Ferrari SLP, Cribari-Neto F (2004). “Beta regression for modelling rates and proportions.” *Journal of Applied Statistics*, **31**(7), 799–815.
- Ferrari SLP, Cysneiros AH (2008). “Skovgaard’s adjustment to likelihood ratio tests in exponential family nonlinear models.” *Statistics & Probability Letters*, **78**(17), 3047–3055.
- Ferrari SLP, Espinheira PL, Cribari-Neto F (2011). “Diagnostic tools in beta regression with varying dispersion.” *Statistica Neerlandica*, **65**(3), 337–351.
- Ferrari SLP, Pinheiro EC (2011). “Improved likelihood inference in beta regression.” *Journal of Statistical Computation and Simulation*, **81**(4), 431–443.
- Griffiths WE, Hill RC, Judge GG (1993). *Learning and practicing econometrics*. New York: Wiley.

- Hancox D, Hoskin CJ, Wilson RS (2010). “Evening up the score: Sexual selection favours both alternatives in the colour-polymorphic ornate rainbowfish.” *Animal Behaviour*, **80**(5), 845–851.
- Hannan EJ, Quinn BG (1979). “The determination of the order of an autoregression.” *Journal of the Royal Statistical Society. Series B*, **41**(2), 190–195.
- Hjorth JSU (1994). *Computer intensive statistical methods: Validation, model selection and Bootstrap*. Chapman and Hall.
- Hu B, Shao J (2008). “Generalized linear model selection using R^2 .” *Journal of Statistical Planning and Inference*, **138**(12), 3705–3712.
- Hurvich CM, Tsai CL (1989). “Regression and time series model selection in small samples.” *Biometrika*, **76**(2), 297–307.
- Ishiguro M, Sakamoto Y (1991). “WIC: an estimation-free information criterion.” *Research memorandum, Institute of Statistical Mathematics, Tokyo*.
- Ishiguro M, Sakamoto Y, Kitagawa G (1997). “Bootstrapping log likelihood and EIC, an extension of AIC.” *Annals of the Institute of Statistical Mathematics*, **49**(3), 411–434.
- Kieschnick R, McCullough BD (2003). “Regression analysis of variates observed on (0, 1): Percentages, proportions and fractions.” *Statistical Modelling*, **3**(3), 193–213.
- Kullback S (1968). *Information theory and statistics*. Dover.
- Kullback S, Leibler RA (1951). “On information and sufficiency.” *The Annals of Mathematical Statistics*, **22**(1), 79–86.
- Lawley DN (1956). “A general method for approximating to the distribution of likelihood ratio criteria.” *Biometrika*, **43**, 295–303.
- Lemonte AJ, Ferrari SLP, Cribari-Neto F (2010). “Improved likelihood inference in Birnbaum-Saunders regression.” *Computational Statistics & Data Analysis*, **54**, 1307–1316.
- Long JS (1997). *Regression models for categorical and limited dependent variables*. 2nd edition. SAGE Publications.
- Mallows CL (1973). “Some comments on C_p .” *Technometrics*, **15**, 661–675.
- McCullagh P, Nelder J (1989). *Generalized linear models*. 2nd edition. Chapman and Hall.
- McQuarrie AD (1999). “A small-sample correction for the Schwarz SIC model selection criterion.” *Statistics & Probability Letters*, **44**(1), 79–86.
- McQuarrie AD, Tsai CL (1998). *Regression and time series model selection*. Singapore: World Scientific.

- Melo TF, Vasconcellos KL, Lemonte AJ (2009). “Some restriction tests in a new class of regression models for proportions.” *Computational Statistics & Data Analysis*, **53**(12), 3972–3979.
- Mittlböck M, Schemper M (2002). “Explained variation for logistic regression - Small sample adjustments, confidence intervals and predictive precision.” *Biometrical Journal*, **44**(3), 263–272.
- Nagelkerke NJD (1991). “A note on a general definition of the coefficient of determination.” *Biometrika*, **78**(3), 691–692.
- Nelder JA, Lee Y (1991). “Generalized linear models for the analysis of taguchi-type experiments.” *Applied Stochastic Models and Data Analysis*, **7**(1), 107–120.
- Ospina R (2008). *Modelos de regressão beta inflacionados*. Tese de Doutorado, Instituto de Matemática e Estatística, Universidade de São Paulo (USP).
- Ospina R, Cribari-Neto F, Vasconcellos KL (2006). “Improved point and interval estimation for a beta regression model.” *Computational Statistics & Data Analysis*, **51**(2), 960–981.
- Ospina R, Ferrari SLP (2011). “A general class of zero-or-one inflated beta regression models.” *Pre-print*. URL <http://arxiv.org/abs/1103.2372>.
- Pammer K, Kevan A (2007). “The contribution of visual sensitivity, phonological processing, and nonverbal IQ to children’s reading.” *Scientific Studies of Reading*, **11**(1), 33–53.
- Pan W (1999). “Bootstrapping likelihood for model selection with small samples.” *Journal of Computational and Graphical Statistics*, **8**(4), 687–698.
- Pereira TL (2010). *Regressão beta inflacionada: Inferência e aplicações*. Tese de Doutorado, Universidade Federal de Pernambuco (UFPE).
- Press W, Teukolsky S, Vetterling W, Flannery B (1992). *Numerical recipes in C: The art of scientific computing*. 2nd edition. Cambridge University Press.
- R Development Core Team (2009). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>.
- Rocha AV, Cribari-Neto F (2009). “Beta autoregressive moving average models.” *Test*, **18**(3), 529–545.
- Rocke DM (1989). “Bootstrap Bartlett adjustment in seemingly unrelated regression.” *Journal of the American Statistical Association*, **84**(406), 598–601.
- Schwarz G (1978). “Estimating the dimension of a model.” *The Annals of Statistics*, **6**(2), 461–464.

- Seghouane AK (2010). “Asymptotic bootstrap corrections of AIC for linear regression models.” *Signal Processing*, **90**, 217–224.
- Shang J, Cavanaugh J (2008). “Bootstrap variants of the Akaike information criterion for mixed model selection.” *Computational Statistics & Data Analysis*, **52**(4), 2004–2021.
- Shibata R (1980). “Asymptotically efficient selection of the order of the model for estimating parameters of a linear process.” *The Annals of Statistics*, **8**(1), 147–164.
- Shibata R (1997). “Bootstrap estimate of Kullback-Leibler information for model selection.” *Statistica Sinica*, **7**, 375–394.
- Silverman BW (1986). *Density estimation for statistics and data analysis*. Chapman and Hall.
- Simas AB, Barreto-Souza W, Rocha AV (2010). “Improved estimators for a general class of beta regression models.” *Computational Statistics & Data Analysis*, **54**(2), 348–366.
- Skovgaard IM (1996). “An explicit large-deviation approximation to one-parameter tests.” *Bernoulli*, **2**(2), 145–165.
- Skovgaard IM (2001). “Likelihood asymptotics.” *Scandinavian Journal of Statistics*, **28**(1), 3–32.
- Smithson M, Verkuilen J (2006). “A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables.” *Psychological Methods*, **11**(1), 54–71.
- Smyth GK, Verbyla AP (1999). “Adjusted likelihood methods for modelling dispersion in generalized linear models.” *Environmetrics*, **10**(6), 695–709.
- Stasinopoulos DM, Rigby RA (2007). “Generalized additive models for location scale and shape (GAMLSS) in R.” *Journal of Statistical Software*, **23**, 1–46.
- Sugiura N (1978). “Further analysts of the data by Akaike’s information criterion and the finite corrections - Further analysts of the data by Akaike’s.” *Communications in Statistics - Theory and Methods*, **7**(1), 13–26.
- Venables WN, Ripley BD (2002). *Modern applied statistics with S*. 4th edition. Springer.
- Zucco C (2008). “The president’s “new” constituency: Lula and the pragmatic vote in Brazil’s 2006 presidential elections.” *Journal of Latin American Studies*, **40**(1), 29–49.

