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

Pós-graduação em Estatística

Regressão beta inflacionada: inferência e aplicações

Tarciana Liberal Pereira

Tese de Doutorado

Recife
17 de dezembro de 2010

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

Tarciana Liberal Pereira

Regressão beta inflacionada: inferência e aplicações

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: *Francisco Cribari-Neto*

Recife
17 de dezembro de 2010

Catalogação na fonte
Bibliotecária Jane Souto Maior, CRB4-571

Pereira, Tarciana Liberal
Regressão beta inflacionada: inferência e aplicações / Francisco Cribari Neto - Recife: O Autor, 2010.
xxii, 100 p. : il., fig., tab.,

Orientador: Francisco Cribari Neto.
Tese (doutorado) Universidade Federal de Pernambuco. CCEN. Estatística, 2010.

Inclui bibliografia, anexo e apêndice.

1. Análise de regressão. 2. Regressão beta inflacionada. 3. Máxima verossimilhança. I. Cribari Neto, Francisco (Orientador). II. Título.

519.536 CDD (22. ed.) MEI2011 – 004

Universidade Federal de Pernambuco
Pós-Graduação em Estatística

17 de dezembro de 2010
(data)

Nós recomendamos que a tese de doutorado de autoria de

Tarciana Liberal Pereira

intitulada

"Regressão beta inflacionada: inferência e aplicações"

seja aceita como cumprimento parcial dos requerimentos para o grau de Doutora em Estatística.

Coordenador da Pós-Graduação em Estatística

Banca Examinadora:

Francisco Cribari Neto orientador

Silvia Lopes de Paula Ferrari (USP)

Gauss Moutinho Cordeiro (UFRPE)

Patrícia Leone Espinheira Ospina

Raydonal Ospina

Este documento será anexado à versão final da dissertação.

Aos meus pais, Agamenon e Vilma, e ao meu marido, Luiz,
dedico esta tese com muito amor e carinho.

Agradecimentos

- A Deus, companheiro inseparável, pela oportunidade desta vida, pelas pessoas e ferramentas colocadas em meu caminho para superar os obstáculos e amenizar as dificuldades e por estar sempre ao meu lado me iluminando e me fortalecendo.

- Ao meu orientador, mestre e amigo, professor Francisco Cribari-Neto, pela oportunidade concedida, pela confiança em mim depositada, pelos ensinamentos, dedicação, paciência e excelente orientação na elaboração desta tese. Muito obrigada por todo aprendizado, pela convivência enriquecedora dos últimos anos e pelo exemplo de professor, orientador e pesquisador que sempre levarei comigo na minha vida profissional.

- Ao meu marido, Luiz, por seu amor incondicional, companheirismo e dedicação, por ter sido o apoio necessário para reerguer-me nos momentos difíceis e pelo permanente estímulo que foi determinante para concretização deste trabalho. Muito obrigada por ser meu porto seguro e por suportar os meus momentos de estresse com a paciência e dedicação de quem ama.

- Aos meus pais, Agamenon e Vilma, pelo amor e apoio incondicionais, pelo esforço contínuo e imensurável para me educar e por me fornecerem princípios básicos e fundamentais para minha formação e concretização dos meus objetivos. Ao meu pai, pelo exemplo de honestidade e perseverança. À minha mãe pelo exemplo de força, fé e grandioso espírito de solidariedade humana.

- Às minhas irmãs, Keyla e Dayse, por toda força, apoio e carinho que tanto me ajudaram. Por todos os momentos vivenciados que nos ajudaram a crescer e manter a nossa união. A Dayse, pelos conselhos e por nossa amizade. A Keyla pela preocupação e incentivo.

- Às minhas queridas sobrinhas, Bruna e Mylle, pelo carinho e amor que sempre tive. A Mylle, por todo carinho com que sempre me recebeu, pela sua alegria e simplicidade. A Bruna, pelos momentos de alegria proporcionados pelas suas vindas a Recife.

- Aos meus sogros, Maria José e Luizito, a vovó Rey e Polly por me acolherem com tanto carinho, por me incentivarem, apoiarem e sempre vibrarem nas minhas conquistas.

- A toda família Liberal e Araújo pelo apoio, preocupação, incentivo e amizade.

- À minha grande amiga e eterna companheira, Tati, por termos conseguido vencer juntas, desde a graduação, todos os desafios e obstáculos. Por todos os momentos que vivenciamos: momentos de alegria, de estresse, momentos difíceis, de superação, momentos de

companheirismo e cumplicidade. Começamos juntas, aprendemos, crescemos, amadurecemos, erramos, rimos bastante, choramos e fortalecemos cada vez mais nossa amizade.

- A Valéria, secretária e amiga, pela eficiência, organização, carinho e atenção que são qualidades extras ao seu excelente trabalho. Por ser essa amiga linda e maravilhosa que sempre me faz sorrir nos momentos de dificuldade, que vibra por cada passo e vitória conquistada e torce pelo meu crescimento.

- A Fábio pela sua amizade, por toda a sua imensa disponibilidade e atenção, pela convivência gratificante e enriquecedora, por todos os momentos e por sua incrível simplicidade que o torna essa pessoa tão especial.

- A todos os colegas de doutorado que conviveram durante esse curso e com os quais compartilhei momentos maravilhosos. A Tati, Hemílio, Marcelo, Daniel e Jane pelos momentos de estudo, trabalho e alegria compartilhados. A Tati e Fábio por terem sido verdadeiros parceiros, pelo apoio, longas horas de trabalho e discussão e pelos momentos de descontração que foram essenciais nessa caminhada. A Hemílio pela disponibilidade e ajuda, a Abrãao e Manoel pelas conversas amigáveis, a Luz pela preocupação, apoio e pelo seu sorriso sempre incentivador, a Laércio, Basílides, Rodrigo, Luiz, Lídia, Waldemar, Izabel e Andréa pelos momentos e experiências trocadas.

- Aos professores do Departamento de Estatística da Universidade Federal de Pernambuco pelos conhecimentos transmitidos desde a graduação, pelos conselhos e por suas contribuições à minha formação como estatística. Em especial aos professores Cristina Raposo, Cláudia Regina, Manoel Senna, Francisco Cribari e Klaus Vasconcellos pela preocupação demonstrada, incentivo e amizade.

- Aos colegas do Departamento de Estatística da Universidade Federal da Paraíba pela compreensão, apoio e incentivo dados durante o doutorado. Em especial ao professor Roberto Quirino pela confiança em mim depositada e pelo apoio que foi essencial para conclusão deste curso.

- Aos funcionários do Departamento de Estatística da UFPE, Jymmy, Neto, Maurício, Cícero e Tonho pela competência profissional, por todo apoio dado ao longo das minhas atividades acadêmicas, pela agradável convivência e momentos de descontração.

- Aos professores Paulo Justiniano Ribeiro Júnior e Raydonal Ospina Martínez pelas sugestões, discussões e suporte durante a realização deste trabalho.

- Aos colegas do mestrado, especialmente a Lutemberg, Marcelinho, Fernando, Marcelinha e Fernanda pelas conversas, momentos divertidos e incentivo.

- Aos queridos amigos, Gabi, Jú, Eufra, Mana, Suzy, Quel, Chi e Jú PE pela amizade sincera, apoio e compreensão por tantos momentos de ausência.

- À CAPES pelo apoio financeiro.

- Aos participantes da banca examinadora pelas sugestões.

Embora ninguém possa voltar atrás e fazer um novo começo,
qualquer um pode começar agora e fazer um novo fim.
—CHICO XAVIER

Resumo

Nesta tese são abordadas aplicações e inferências em modelos de regressão beta inflacionados tanto sob dispersão fixa quanto sob dispersão variável. Na primeira parte da tese, realizamos uma análise das eficiências administrativas dos municípios do estado de São Paulo com base em modelos de regressão beta e beta inflacionado com efeitos espaciais. Adicionalmente, uma comparação com os resultados obtidos a partir do modelo de regressão linear e com inferência realizada via quasi-verossimilhança é apresentada. O modelo de regressão beta inflacionado se mostrou mais adequado para explicar os escores de eficiência média dos municípios. Na segunda parte, a partir do teste RESET (Ramsey, 1969), desenvolvemos testes de erro de especificação para modelos de regressão beta inflacionados tanto sob dispersão fixa quanto sob dispersão variável. Em particular, nós propomos duas variantes do teste. Na primeira variante, nós apenas adicionamos variáveis de teste para o submodelo da média ao passo que na segunda variante, variáveis de teste são adicionadas em todos os submodelos. Nós consideramos diversos erros de especificação em nossa avaliação numérica: não-linearidade negligenciada, função de ligação incorreta, variáveis omitidas, correlação espacial negligenciada e dispersão variável não modelada. Os resultados de um estudo de Monte Carlo mostraram que nosso teste de especificação tipicamente apresenta bons poderes em amostras de tamanho pequeno a moderado, exceto quando a correlação espacial é negligenciada. Neste caso, tamanhos amostrais maiores são necessários para obter bons poderes. Por fim, na terceira parte, desenvolvemos novos ajustes para as estatísticas da razão de verossimilhanças usual e sinalizada em modelos de regressão beta inflacionados a partir da proposta de Skovgaard (2001). Os ajustes são de fácil obtenção pois só requerem cumulantes até segunda ordem da função de log-verossimilhança. Evidências obtidas a partir de um estudo de Monte Carlo mostraram que os testes propostos apresentaram melhores desempenhos do que os testes não modificados em pequenas amostras.

Palavras-chave: Dados de proporções, dados de taxas, eficiência técnica, máxima verossimilhança, regressão beta, regressão beta inflacionada, testes de hipótese.

Abstract

Beta regression models are useful for modeling random variables that assume values in the standard unit interval, such as rates and proportions. Such models cannot be used, however, when the data contain zeros and/or ones, i.e., when some observations equal one or both interval limits. Ospina (2008) proposed a class of inflated beta regressions models (with fixed dispersion) to handle data that contain zeros and/or ones. The proposed model contains two submodels, namely: one for the condicional mean (μ) and a second submodel for the probability that the variate equals zero or one. In the present dissertation, we consider an empirical application of the class of inflated beta regressions and also develop new inferential procedures for these models under both fixed and variable dispersion. In the first part of our work, we analyze data on administrative efficiency of public administrations of São Paulo countries. To that end, we use spatially dependent beta and inflated beta regressions. The results are compared to those obtained using the classical linear regression model and also to those obtained via quasi-likelihood inference. The inflated beta regression inference was clearly superior to the competing inferences. In the second part of the dissertation, we propose misspecification tests for inflated beta regressions under both fixed and variable dispersion. In particular, we propose two tests. In the first test, only the mean submodel is augmented whereas in the second test all submodels are augmented with testing variables. In our numerical evaluation, we consider several sources of misspecification, namely: neglected nonlinearities, incorrect link functions, omitted regressors, neglected spatial correlation and neglected variable dispersion. The Monte Carlo results show that our misspecification tests typically have good power in small to moderate samples, except when the practitioner fails to account for spatial correlation, in which case larger sample sizes are needed to achieve good power. In the third and final part of the dissertation, we develop finite-sample adjustments for likelihood ratio test statistics (standard and signed) in inflated beta regression models. The proposed adjustments are easily computed since they only require first and second order log-likelihood cumulants. Monte Carlo evidence shows that the proposed tests outperform the unmodified tests in small samples.

Keywords: Beta regression, hypothesis tests, inflated beta regression, maximum likelihood, proportions, rates, technical efficiency.

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

Modelos de regressão são geralmente usados para analisar dados que estão relacionados a outras variáveis. A análise de regressão convencional, baseada na suposição de erros normais, é amplamente usada nestas aplicações. Entretanto, tal modelagem pode ser inapropriada em situações em que a variável dependente é limitada, como por exemplo, quando desejamos modelar taxas e proporções. Uma vez que nestas situações a variável resposta é, em geral, restrita ao intervalo $(0, 1)$, os valores ajustados para a variável de interesse obtidos via modelagem clássica de regressão podem exceder os limites do intervalo. Uma solução é transformar a variável dependente de forma que ela assuma valores na reta real. Entretanto, algumas desvantagens podem surgir quando se usa tal prática. Por exemplo, os parâmetros do modelo podem não ser facilmente interpretados em termos da resposta original. Adicionalmente, as distribuições de taxas e proporções são tipicamente assimétricas e assim, inferências baseadas na suposição de normalidade podem conduzir a conclusões errôneas.

Kieschnick e McCullough (2003) apresentaram diferentes estratégias de modelagem que são comumente usadas para lidar com dados de proporções distribuídos no intervalo unitário padrão. Os modelos abordados foram o modelo normal linear, o modelo normal não-linear, o modelo normal censurado, o modelo logito e os modelos que usam a distribuição beta, simplex e de quasi-verossimilhança. Segundo os autores, dado que a variável resposta assume valores num intervalo limitado, a média da resposta deve ser não-linear nos parâmetros de regressão. Adicionalmente, modelos de regressão envolvendo taxas e proporções são geralmente heteroscedásticos, ou seja, a variância deve se aproximar de zero à medida que a média da resposta se aproxima dos limites do intervalo. Com base nos resíduos e no critério AICc, que é uma variação do critério de seleção AIC, os autores concluíram que o modelo que usa a distribuição beta para a variável resposta e função de ligação logito para a esperança condicional superou os outros seis modelos.

Existem diferentes especificações para a regressão beta na literatura, tais como, Paolino (2001), Kieschnick e McCullough (2003), Ferrari e Cribari-Neto (2004) e Vasconcellos e Cribari-Neto (2005). O modelo proposto por Ferrari e Cribari-Neto apresenta algumas vantagens das quais é possível destacar: a especificação usada é similar à usada na classe de modelos lineares generalizados (McCullagh e Nelder, 1989); modela-se diretamente a média ao invés dos dois parâmetros que indexam a distribuição; a função de ligação que relaciona a resposta média ao preditor linear é bastante geral; os autores desenvolveram estimação pontual e intervalar, testes de hipóteses e medidas de diagnóstico; é possível utilizar os pacotes `betareg` e `gamlss` da linguagem de programação `R` (ver `http://www.R-project.org`) para obter resultados de aplicações e simulações. Adicionalmente, Ospina et al. (2006) derivaram os vieses de segunda ordem dos estimadores de máxima verossimilhança dos parâmetros do modelo. Espi-

nheira et al. (2008a, 2008b) propuseram novos resíduos para a regressão beta e desenvolveram ferramentas de diagnóstico para avaliar a influência de observações na classe de modelos de regressão beta. Simas et al. (2010) propuseram uma variante do modelo de regressão beta que introduz não-linearidade e dispersão variável.

Em uma ampla variedade de problemas que envolvem taxas, frações e proporções, a variável de interesse pode conter zeros e/ou uns. Por exemplo, quando se deseja avaliar a taxa de mortalidade ou infecção para uma determinada doença, a proporção da renda familiar gasta com educação dos filhos ou a proporção de pessoas que apresentam uma determinada característica. Nessas situações, a função de log-verossimilhança do modelo de regressão beta se torna ilimitada e não é adequado considerar que os dados provêm de uma distribuição absolutamente contínua. Novamente, é possível fazer uso de transformações de forma que a variável transformada assuma valores na reta real. Contudo, como foi citado anteriormente, as transformações a dados na forma de frações ou proporções modificam a natureza real dos dados e não possibilitam a interpretação direta dos parâmetros envolvidos no modelo. Uma solução mais adequada é considerar um modelo estatístico que permita adicionar à distribuição contínua da variável resposta um ponto de massa em um ou em ambos os extremos.

Modelos para dados contínuos que apresentam excesso de zeros ou uns têm sido usados na prática. Aitchison (1955, 1969) utilizou uma mistura de uma distribuição degenerada em zero com uma distribuição lognormal. Feuerverger (1979) introduziu uma mistura de uma distribuição degenerada em zero com uma distribuição gama. Yoo (2004) modelou despesas com telefonia móvel considerando um modelo de mistura entre uma distribuição degenerada em zero e uma outra com suporte nos reais positivos (por exemplo, gama, exponencial ou Weibull). Neste caso, os zeros representam os indivíduos que nada gastam com telefonia móvel e a distribuição contínua é usada para modelar as despesas dos indivíduos que têm algum gasto com telefonia móvel. Heller, Stasinopoulos e Rigby (2006) consideraram uma mistura entre uma distribuição degenerada em zero e a distribuição normal inversa para modelar dados de demanda em seguros. Cook, Kieschnick e McCullough (2006) propuseram um modelo de regressão para dados de financiamento em que a variável resposta segue distribuição de mistura entre uma distribuição degenerada em zero e uma distribuição beta. Hoff (2007) considerou um modelo de regressão beta inflacionado em um para avaliar escores de eficiência distribuídos no intervalo $(0,1]$. Para este modelo assume-se que a distribuição do escore de eficiência segue distribuição de mistura entre uma distribuição degenerada no ponto um e uma distribuição beta. A probabilidade de que o escore de eficiência é igual a um e a média da distribuição beta foram modeladas através de preditores lineares usando ligação logito. Ospina e Ferrari (2010) propuseram distribuições de mistura entre uma distribuição beta e uma distribuição de Bernoulli degenerada em zero e/ou um para acomodar dados observados nos intervalos $(0,1]$, $[0,1)$ e $[0,1]$. Estas distribuições são denominadas distribuições beta inflacionadas. Adicionalmente, Ospina (2008) propôs um modelo de regressão beta inflacionado em zero ou um, sugerindo uma parametrização que permite modelar de forma direta a média da distribuição beta e a massa de probabilidade em zero ou um usando preditores lineares e funções de ligação. Estes modelos, chamados modelos de regressão beta inflacionados (dispersão constante), são extensões naturais do modelo de regressão beta proposto por Ferrari e Cribari-Neto (2004), assumindo que a distribuição da variável dependente é beta inflacionada. Neste contexto, o autor supõe que o

componente contínuo é modelado pela distribuição beta, uma vez que esta é bastante flexível para ajustar dados no intervalo $(0,1)$, e o componente discreto, ou seja, o ponto de massa, é modelado através de uma distribuição degenerada. Adicionalmente, o autor apresentou resultados de inferência estatística sob o enfoque da teoria de verossimilhança, entre eles estimação pontual, construção de intervalos de confiança assintóticos e testes de hipóteses. Um modelo de regressão beta inflacionado em zero e um, obtido como uma extensão natural do modelo de regressão beta inflacionado em zero ou um também foi apresentado. Neste caso a distribuição da variável resposta é uma mistura entre uma distribuição beta e uma distribuição de Bernoulli. O modelo de regressão beta inflacionado em zero ou um (Ospina, 2008) será abordado neste trabalho.

Esta tese tem como objetivo aplicar a classe de regressão beta inflacionada na modelagem de dados de proporções, taxas ou frações na presença de zeros ou uns, bem como desenvolver novas estratégias inferenciais para esta classe de modelos. No que se refere à aplicabilidade dos modelos de regressão beta inflacionados, dados de escores de eficiência administrativa dos municípios do estado de São Paulo e do Brasil são analisados. Eficiência do gasto público significa que o governo consegue transformar recursos do orçamento em serviços prestados de forma a maximizar a oferta de serviços dadas as restrições orçamentárias enfrentadas. Considerando que os insumos são escassos e possuem usos alternativos, torna-se relevante realizar análises de avaliação de eficiência. Será exatamente a comparação entre a forma de produzir de cada município que possibilitará a avaliação de suas eficiências. Considerando que os dados de eficiência são limitados, ou seja, a resposta é restrita ao intervalo $(0,1]$, e que as observações tendem a apresentar dependência espacial devido à localização geográfica de cada município, nós utilizamos modelos de regressão beta e beta inflacionado com efeitos espaciais. Adicionalmente, uma comparação com os modelos de regressão linear e com inferência realizada via quasi-verossimilhança é apresentada.

Na prática, em geral, a forma funcional do modelo e os regressores são postulados com base em estudos anteriores. Caso a especificação utilizada seja errônea, inferências imprecisas podem ocorrer no que tange à estimação dos parâmetros, testes de hipótese e intervalos de confiança. Desta forma, é importante na modelagem de dados a realização de testes de hipóteses para verificar se o modelo está corretamente especificado. Um teste bastante utilizado, na classe de modelos de regressão linear, para verificar se a suposição de especificação correta é válida é o teste RESET (*regression specification error test*) proposto por Ramsey (1969). A mecânica do teste RESET consiste em adicionar uma forma não-linear ao modelo através de potências de variáveis de teste e em seguida testar, através de um teste F usual, a significância de tais variáveis. Nós propomos testes de erro de especificação para modelos de regressão beta inflacionados tanto sob dispersão fixa quanto sob dispersão variável. Os testes são baseados no teste RESET. Duas variantes do teste proposto foram apresentadas. Na primeira variante, é testado apenas se o submodelo da média está bem especificado, ao passo que na segunda a existência de erros de especificação é verificada nos três submodelos de regressão.

O teste da razão de verossimilhanças é bastante utilizado para realizar testes de hipóteses sobre os parâmetros de modelos de regressão. A estatística utilizada neste teste é proporcional à diferença entre o valor máximo do logaritmo da função de verossimilhança perfilada e o valor máximo desse logaritmo sob a hipótese nula. Um problema comum refere-se à qua-

lidade da aproximação usual da distribuição nula da estatística da razão de verossimilhanças pela distribuição qui-quadrado. Torna-se importante obter ajustes para a estatística da razão de verossimilhanças que melhorem a qualidade de tal aproximação. Ainda em relação a inferências nos modelos de regressão beta inflacionados, com o intuito de melhorar o desempenho em amostras finitas do teste da razão de verossimilhanças, nós derivamos, baseados na correção de Skovgaard (2001), ajustes para as estatísticas da razão de verossimilhanças usual e sinalizada na classe de modelos de regressão beta inflacionados. Os ajustes são de fácil obtenção, pois só requerem cumulantes até segunda ordem da função de log-verossimilhança e não requerem ortogonalidade entre os parâmetros de interesse e de pertubação.

A presente tese encontra-se organizada em cinco capítulos. Os Capítulos 2, 3 e 4 são autosuficientes, ou seja, cada capítulo pode ser lido separadamente, em qualquer ordem. Desta forma, algumas notações e resultados são apresentados mais de uma vez. No Capítulo 2 introduzimos os modelos de regressão beta, beta inflacionado e o processo de estimação via quasi-verossimilhança. Alguns conceitos de estatística espacial, necessários para incorporar a dependência espacial aos modelos de regressão, são apresentados. Avaliamos dados de eficiência administrativa municipal no estado de São Paulo via modelagem de regressão beta, beta inflacionada, linear e quasi-verossimilhança. Uma análise da eficiência pura, que é obtida expurgando-se das eficiências administrativas os efeitos condicionantes estimados a partir dos modelos de regressão, bem como algumas medidas descritivas, histogramas, gráficos de resíduos e mapas são apresentados. Uma comparação com os resultados obtidos para o Brasil e algumas conclusões finais também são apresentadas.

No Capítulo 3, apresentamos um variante do modelo de regressão beta inflacionado que incorpora dispersão variável. Nós apresentamos a função de log-verossimilhança do modelo, as funções escore e a matriz de informação de Fisher, uma vez que essas quantidades são necessárias para os testes de erro de especificação propostos, os quais também são introduzidos neste capítulo. Adicionalmente, avaliamos através de estudos de simulações de Monte Carlo os desempenhos dos testes propostos no que tange à capacidade de identificar alguns erros de especificação, tais como não-linearidade negligenciada, variáveis omitidas, escolha incorreta da função de ligação, correlação espacial negligenciada e dispersão variável não modelada. Em particular, nós propomos duas variantes do teste. Na primeira variante, nós apenas adicionamos variáveis de teste ao submodelo da média. A segunda variante segue da adição de variáveis de teste em todos os submodelos. Os novos testes são aplicados a dados reais. Algumas considerações são apresentadas no final do capítulo.

No quarto capítulo, dando continuidade ao desenvolvimento de alguns aspectos de inferência para a classe de modelos de regressão beta inflacionados, derivamos estatísticas ajustadas da razão de verossimilhanças e da razão de verossimilhanças sinalizada. Estas estatísticas ajustadas são baseadas na proposta de Skovgaard (1996, 2001) que tem ampla aplicabilidade e vem sendo utilizada para corrigir a estatística da razão de verossimilhanças em várias classes de modelos. Este ajuste tem várias vantagens, tais como, derivação mais simples da estatística ajustada do que a corrigida via Bartlett, aplicação direta à estatística da razão de verossimilhanças e não necessidade de ortogonalização entre os parâmetros de interesse e os de perturbação. As quantidades necessárias para a obtenção das estatísticas ajustadas na classe de modelos de regressão beta inflacionado com dispersão variável também são apresentadas. O comportamento

em amostras finitas de diferentes testes baseados na função de verossimilhança em modelos de regressão beta inflacionados são avaliados via simulação de Monte Carlo. Finalmente, algumas conclusões e detalhes técnicos, que encontram-se no Apêndice A, são apresentados.

Por fim, no último capítulo, apresentamos as considerações finais do trabalho que incluem conclusões e possíveis direcionamentos para trabalhos futuros.

Todas as avaliações numéricas, assim como os gráficos e 'scripts' desenvolvidos para estimação dos modelos de regressão foram produzidos utilizando a liguagem de programação `R` que se encontra disponível gratuitamente em `http://www.r-project.org`.

CAPÍTULO 2

Uma avaliação da eficiência de administrações municipais no estado de São Paulo

2.1 Introdução

A economia brasileira atravessou vários ciclos ao longo de sua história. Valendo-se de políticas econômicas desenvolvimentistas, o Brasil desenvolveu grande parte de sua infraestrutura em pouco tempo e alcançou elevadas taxas de crescimento econômico. Entretanto, o governo manteve suas contas em frequente desequilíbrio durante prolongados períodos, ampliando assim seu endividamento e desencadeando surtos inflacionários.

As disparidades e as desigualdades regionais são até hoje um problema no Brasil. Cada região possui características distintas devido a vários fatores, como história, desenvolvimento, população e economia. A região Centro-Sul, de todas as macrorregiões, é a mais desenvolvida, não só economicamente, mas também no que pertine a indicadores sociais (saúde, educação, renda, mortalidade infantil, analfabetismo, entre outros). São Paulo é o estado mais populoso do Brasil, com mais de quarenta milhões de habitantes, e é a terceira unidade administrativa mais populosa da América do Sul. É o mais rico dos estados brasileiros, sendo responsável por cerca de um terço do produto interno bruto do país, e é indubitavelmente um dos maiores pólos econômicos da América Latina. A riqueza produzida em São Paulo equivale a duas vezes o que é produzido na região Norte e a 94% do que é produzido por todos os estados do Nordeste.

Uma gestão administrativa eficiente é importante para que se possa maximizar a oferta de serviços públicos dada a restrição orçamentária vigente. Uma das alternativas para que o governo possa atuar efetivamente na promoção do desenvolvimento econômico é a melhoria do gasto público. Um conceito bastante utilizado neste contexto é o de eficiência produtiva. Eficiência do gasto público significa que o governo consegue transformar recursos do orçamento em serviços prestados de forma a maximizar a oferta de serviços dadas as restrições orçamentárias enfrentadas. Uma metodologia voltada especificamente para esta questão é a de *Data Envelopment Analysis* (DEA) que atribui a cada unidade um valor representativo de seu desempenho relativo. Geralmente, esses escores variam entre 0 e 1, e as unidades eficientes recebem valor igual a 1.

Uma questão importante em avaliações de eficiência é a determinação dos fatores que contribuem para que um município seja mais eficiente do que outro. A análise de regressão convencional, baseada em modelos com erros normais, pode ser inapropriada em situações em que a variável de interesse é limitada, por exemplo, quando a resposta é restrita ao intervalo $(0, 1]$. Uma solução bastante utilizada reside em aplicar uma transformação na variável dependente de forma que ela assuma valores na reta real. Contudo, esta transformação pode trazer

alguns inconvenientes. Por exemplo, os parâmetros do modelo podem não ser mais facilmente interpretados em termos da resposta original e os valores ajustados para a variável de interesse podem exceder os limites do intervalo.

Um modelo que pode ser utilizado e que em geral se apresenta mais adequado para este tipo de dado é o modelo tobit (James Tobin, 1958). Entretanto, como o método de máxima verossimilhança é, em geral, o método de estimação utilizado no modelo tobit, é necessário, para sua aplicação, assumir normalidade e homoscedasticidade. Em se tratando de medidas de proporção ou taxas, em geral, uma certa assimetria e heteroscedasticidade é exibida nos dados e, portanto, inferências baseadas na normalidade podem ser errôneas.

Sampaio de Souza, Cribari-Neto e Stosic (2005) estimaram índices de eficiência para municípios brasileiros e utilizaram regressão linear e regressão quantílica para identificar os fatores que explicam a variabilidade dos escores de eficiência obtidos. A variável dependente é o logaritmo natural dos escores de eficiência para quase cinco mil municípios brasileiros.

Visando encontrar alternativas mais precisas e confiáveis para modelar dados de eficiência administrativa, apresentaremos uma nova modelagem a partir dos modelos de regressão beta. Ferrari e Cribari-Neto (2004) propuseram um modelo de regressão adequado a situações em que a variável dependente é medida continuamente no intervalo unitário padrão. Esta classe de modelos de regressão considera que a resposta segue distribuição beta e sua média está relacionada, através de uma função de ligação, a um preditor linear definido por regressores e parâmetros de regressão desconhecidos.

A distribuição beta é útil para modelar variáveis distribuídas de forma contínua no intervalo $(0, 1)$. Entretanto, ao lidar com dados de eficiência que podem apresentar valores iguais a um, não é mais adequado considerar que os dados provêm de uma distribuição contínua. Para acomodar dados observados nos intervalos $(0, 1]$, $[0, 1)$ e $[0, 1]$, Ospina e Ferrari (2010) propuseram distribuições de mistura entre uma distribuição beta e uma distribuição de Bernoulli degenerada em zero e/ou um. Tais distribuições são denominadas distribuições beta inflacionadas. Adicionalmente, Ospina (2008) propôs modelos de regressão beta inflacionados que são extensões naturais do modelo de regressão beta proposto por Ferrari e Cribari-Neto (2004), assumindo que a distribuição da variável dependente é beta inflacionada.

Uma outra alternativa que pode ser usada para modelar dados de eficiência administrativa é a utilização da função de quasi-verossimilhança para a estimação dos parâmetros. Em vez de especificar uma distribuição de probabilidade para os dados, apenas uma relação entre a média e a variância é especificada na forma de uma função de variância. A função de quasi-verossimilhança incorpora ao modelo a variância da variável resposta como uma função da média e permite incluir um fator multiplicativo conhecido como parâmetro de dispersão. Os modelos de quasi-verossimilhança podem ser ajustados usando uma extensão simples dos algoritmos utilizados para ajustar modelos lineares generalizados.

Em se tratando de dados de eficiência administrativa de municípios, as observações tendem a apresentar dependência espacial devido à localização geográfica de cada município. A necessidade da quantificação da dependência espacial, bem como as limitações das ferramentas usuais da estatística para lidar com dados dessa natureza, levaram ao desenvolvimento da estatística espacial. A característica fundamental da estatística espacial é que a referência geográfica, isto é, as coordenadas espaciais são utilizadas no processo de coleta, descrição e análise

dos dados. Por incorporarem a dimensão espacial as técnicas de estatística espacial tipicamente produzem informações não reveladas pelas ferramentas tradicionais da estatística.

Dada a importância econômica do estado de São Paulo e a necessidade de analisar a eficiência administrativa de municipalidades levando em consideração a dependência espacial presente nos dados, pretendemos, neste capítulo, abordar modelos de regressão beta e beta inflacionados com efeitos espaciais. Pretendemos ainda comparar os resultados obtidos a partir de tais modelos com aqueles oriundos de abordagens baseadas em regressão linear e quasi-verossimilhança. Uma comparação com os resultados obtidos para o Brasil também será apresentada.

O presente capítulo encontra-se organizado da seguinte forma. A Seção 2.2 introduz os modelos de regressão beta, beta inflacionado e o processo de estimação via quasi-verossimilhança. A incorporação da dependência espacial aos modelos considerados neste capítulo é tratada na Seção 2.3. A análise de dados de eficiência administrativa municipal é apresentada na Seção 2.4. Finalmente, algumas conclusões e comentários finais encontram-se na Seção 2.5.

2.2 Alguns modelos de regressão para dados limitados ao intervalo $(0,1]$

2.2.1 O modelo de regressão beta

Um modelo que pode ser utilizado em situações em que a variável resposta é distribuída continuamente no intervalo $(0,1)$ e pode ser explicada por outras variáveis através de uma estrutura de regressão é o modelo de regressão beta proposto por Ferrari e Cribari-Neto (2004). Para obter uma estrutura de regressão para a média da variável resposta incluindo um parâmetro de precisão, Ferrari e Cribari-Neto (2004) definiram uma parametrização diferente da usual para a densidade beta. Sob essa nova parametrização, a densidade de y pode ser escrita como

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

em que $0<\mu<1$ e $\phi>0$. Aqui, $\Gamma(\cdot)$ é a função gama. Dizemos que y tem distribuição beta com média μ e precisão ϕ e escrevemos $y \sim \mathcal{B}(\mu,\phi)$. A média e a variância de y são dadas, respectivamente, por $\mathbb{E}(y)=\mu$ e $\mathrm{Var}(y)=\frac{V(\mu)}{1+\phi}$, em que $V(\mu)=\mu(1-\mu)$. É importante notar que a variância da variável resposta é uma função de μ, ou seja, dos valores das covariáveis, implicando que respostas de variâncias não-constantes são facilmente acomodadas dentro do modelo.

Sejam $y_1,\ldots,y_n$ variáveis aleatórias independentes, em que $y_t \sim \mathcal{B}(\mu_t,\phi)$, $t=1,\ldots,n$. O modelo de regressão beta é definido supondo que a média de y_t satisfaz uma relação funcional da forma

$$g(\mu_t) = \sum_{i=1}^{k} x_{ti}\beta_i = \eta_t, \tag{2.2.1}$$

em que $\beta=(\beta_1,\ldots,\beta_k)^\top$ é um vetor de parâmetros de regressão desconhecidos ($\beta \in \mathbb{R}^k$), η_t é um preditor linear e $x_{t1},\ldots,x_{tk}$ são observações de k covariáveis ($k<n$), as quais são as-

sumidas conhecidas e fixas. A função de ligação $g : (0,1) \to \mathbb{R}$ é estritamente monótona e duas vezes diferenciável. Entre as funções de ligação mais utilizadas estão a função logit com $g(\mu) = \log(\mu/(1-\mu))$, a função de ligação probit com $g(\mu) = \Phi^{-1}(\mu)$, em que $\Phi(\cdot)$ é a função de distribuição normal padrão, a especificação log-log complementar com $g(\mu) = \log(-\log(1-\mu))$, a ligação log-log com $g(\mu) = -\log(-\log(\mu))$ e a função de ligação Cauchy com $g(\mu) = \tan(\pi(\mu - 0.5))$.

A função de log-verossimilhança para uma amostra de n observações independentes é da forma

$$\begin{aligned}\ell(\beta,\phi) &= \sum_{t=1}^{n} \ell_t(\mu_t,\phi) \\ &= \sum_{t=1}^{n}[\log\Gamma(\phi) - \log\Gamma(\mu_t\phi) - \log\Gamma((1-\mu_t)\phi) + \\ &\quad (\mu_t\phi - 1)\log y_t + ((1-\mu_t)\phi - 1)\log(1-y_t)], \qquad (2.2.2)\end{aligned}$$

em que $\ell_t(\mu_t,\phi)$ é a função de log-densidade de y_t e $\mu_t = g^{-1}(\eta_t)$, como definida em (2.2.1), é função de β. Os estimadores de máxima verossimilhança de β e ϕ são obtidos através da maximização numérica de (2.2.2) usando um algoritmo de otimização não-linear, tal como um algoritmo de Newton ou um algoritmo quasi-Newton. Sob certas condições de regularidade, quando o tamanho da amostra é grande,

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

aproximadamente, em que $\hat\beta$ e $\hat\phi$ são os estimadores de máxima verossimilhança de β e ϕ, respectivamente, e

$$K^{-1} = K^{-1}(\beta,\phi) = \begin{pmatrix} K^{\beta\beta} & K^{\beta\phi} \\ K^{\phi\beta} & K^{\phi\phi}\end{pmatrix},$$

em que

$$K^{\beta\beta} = \frac{1}{\phi}(X^\top W^* X)^{-1}\left\{I_k + \frac{X^\top T c c^\top T^\top X (X^\top W^* X)^{-1}}{\varrho\phi}\right\},$$

$$K^{\beta\phi} = (K^{\phi\beta})^\top = -\frac{1}{\varrho\phi}(X^\top W^* X)^{-1} X^\top T c$$

e

$$K^{\phi\phi} = \varrho^{-1},$$

em que X é uma matriz $n \times k$ cuja t-ésima linha é $x_t^\top$, $\varrho = \mathrm{tr}(D) - \phi^{-1} c^\top T^\top X (X^\top W^* X)^{-1} X^\top T c$, $D = \mathrm{diag}\{d_1,\ldots,d_n\}$, com $d_t = \psi'(\mu_t\phi)\mu_t^2 + \psi'((1-\mu_t)\phi)(1-\mu_t)^2 - \psi'(\phi)$, $W^* = \mathrm{diag}\{w_1^*,\ldots,w_n^*\}$, com $w_t^* = \phi\{\psi'(\mu_t\phi) + \psi'((1-\mu_t)\phi)\}\{d\mu_t/d\eta_t\}^2$, $c = (c_1,\ldots,c_n)^\top$, com $c_t = \phi\{\psi'(\mu_t\phi)\mu_t - \psi'((1-\mu_t)\phi)(1-\mu_t)\}$ e $T = \mathrm{diag}\{d\mu_1/d\eta_1,\ldots,d\mu_n/d\eta_n\}$. Aqui, $\psi'(\cdot)$ é a função trigama, $\mathrm{tr}(\cdot)$ é o operador traço e I_k é a matriz identidade de dimensão k.

2.2.2 O modelo de regressão beta inflacionado

Na prática, dados na forma de taxas e proporções podem incluir zeros e/ou uns. Nesta situação, não é mais possível supor que os dados provêm de uma distribuição contínua e, desta forma, o modelo de regressão beta não é mais adequado. Ospina (2008) desenvolveu uma modelagem estatística para dados distribuídos de forma contínua no intervalo $(0,1)$, mas que incluem observações em um ou em ambos extremos. Uma vez que dados de eficiência assumem valores no intervalo $(0,1]$ abordaremos apenas o modelo de regressão beta inflacionado em zero ou um.

A função de distribuição acumulada do modelo beta inflacionado é dada por

$$\mathrm{BI}_c(y;\alpha,\mu,\phi) = \alpha\,\mathbb{1}_{\{c\}}(y) + (1-\alpha)F(y;\mu,\phi),$$

em que $\mathbb{1}_{\{c\}}(y)$ é uma função indicadora que assume valor 1 se $y = c$ e 0 caso contrário, $F(\cdot;\mu,\phi)$ é a função de distribuição acumulada beta $\mathcal{B}(\mu,\phi)$ e $0 < \alpha < 1$ é o parâmetro de mistura da distribuição dado por $\alpha = \Pr(y = c)$. A função BI_c tem um ponto de massa em $y = c$, não sendo assim absolutamente contínua.

A função densidade de probabilidade de y é dada por

$$\mathrm{bi}_c(y;\alpha,\mu,\phi) = \begin{cases} \alpha, & y = c, \\ (1-\alpha)f(y;\mu,\phi), & y \in (0,1), \end{cases} \tag{2.2.3}$$

em que $0 < \alpha < 1$, $0 < \mu < 1$, $\phi > 0$ e $f(y;\mu,\phi)$ é a função de densidade $\mathcal{B}(\mu,\phi)$. Para esta distribuição, $\mathbb{E}(y) = \alpha c + (1-\alpha)\mu$ e $\mathrm{Var}(y) = (1-\alpha)\frac{\mathrm{V}(\mu)}{\phi+1} + \alpha(1-\alpha)(c-\mu)^2$, em que $V(\mu) = \mu(1-\mu)$. Desta forma, no modelo beta inflacionado a resposta é estimada através de $\hat{\mathbb{E}}(y) = \hat{\alpha}c + (1-\hat{\alpha})\hat{\mu}$. Se $c = 0$, a distribuição (2.2.3) é denominada distribuição beta inflacionada no ponto zero e escrevemos $y \sim \mathrm{BEZI}(\alpha,\mu,\phi)$. Se $c = 1$, a distribuição (2.2.3) é denominada distribuição beta inflacionada no ponto um e escrevemos $y \sim \mathrm{BEOI}(\alpha,\mu,\phi)$.

Ospina (2008) sugeriu uma parametrização que permite modelar de forma direta a média da distribuição beta e a massa da probabilidade em zero ou um usando preditores lineares e funções de ligação adequadas, estendendo assim o modelo de regressão beta inflacionado em um (Hoff, 2007) e o modelo de regressão beta inflacionado em zero (Cook, Kieschnick e McCullough, 2006). Sejam $y_1,\ldots,y_n$ variáveis aleatórias independentes com distribuição beta inflacionada no ponto c ($c = 0$ ou $c = 1$). O modelo de regressão beta inflacionado em c é definido supondo que a média condicional de y_t e a massa de probabilidade em c satisfazem as seguintes relações funcionais:

$$h(\alpha_t) = \sum_{i=1}^{M} z_{ti}\gamma_i = \zeta_t, \tag{2.2.4}$$

$$g(\mu_t) = \sum_{i=1}^{m} x_{ti}\beta_i = \eta_t, \tag{2.2.5}$$

em que $\gamma = (\gamma_1,\ldots,\gamma_M)^\top$ e $\beta = (\beta_1,\ldots,\beta_m)^\top$ são vetores de parâmetros de regressão desconhecidos, tais que $\gamma \in \mathbb{R}^M$ e $\beta \in \mathbb{R}^m$, $x_{t1},\ldots,x_{tm}$ e $z_{t1},\ldots,z_{tM}$ são observações de covariáveis conhecidas ($m + M < n$) que podem coincidir total ou parcialmente. As funções de ligação $h : (0,1) \to \mathbb{R}$

e $g : (0,1) \to \mathbb{R}$ são estritamente monótonas e duas vezes diferenciáveis. Para as funções de ligação h e g é possível utilizar, por exemplo, a especificação logit, a função de ligação probit, a especificação log-log complementar e a função de ligação log-log definidas na Seção 2.2.1 para o modelo de regressão beta.

A função de log-verossimilhança para o modelo de regressão beta inflacionado em c é da forma

$$\ell(\theta) = \ell_1(\gamma) + \ell_2(\beta, \phi),$$

em que

$$\begin{aligned} \ell_1(\gamma) &= \sum_{t=1}^{n} \ell_t(\alpha_t), \\ \ell_2(\beta, \phi) &= \sum_{t:y_t \in (0,1)} \ell_t(\mu_t, \phi), \end{aligned}$$

sendo

$$\begin{aligned} \ell_t(\alpha_t) &= \mathbb{1}_{\{c\}}(y_t) \log \alpha_t + (1 - \mathbb{1}_{\{c\}}(y_t)) \log(1 - \alpha_t), \\ \ell_t(\mu_t, \phi) &= \log\Gamma(\phi) - \log\Gamma(\mu_t\phi) - \log\Gamma((1-\mu_t)\phi) + (\mu_t\phi - 1)\log y_t \\ &\quad + \{(1-\mu_t)\phi - 1\}\log(1 - y_t). \end{aligned}$$

Os parâmetros μ_t e α_t são definidos como funções de γ e β, através de (2.2.4) e (2.2.5), ou seja, $\alpha_t = h^{-1}(\zeta_t)$ e $\mu_t = g^{-1}(\eta_t)$. Temos que $\ell_1(\gamma)$ é a função de log-verossimilhança de um modelo linear generalizado com resposta binária e $\ell_2(\beta, \phi)$ é a função de log-verossimilhança de um modelo de regressão beta em que a variável dependente é restrita ao intervalo aberto $(0, 1)$.

Dada a separabilidade dos vetores de parâmetros γ e $(\beta^\top, \phi)^\top$, é possível obter de forma independente as funções escore para γ e para $(\beta^\top, \phi)^\top$. Os vetores escore para γ, β e ϕ podem ser escritos, em forma matricial, como

$$\begin{aligned} U_\gamma(\gamma) &= Z^\top P G(y^c - \alpha^*), \\ U_\beta(\beta, \phi) &= \phi X^\top T H(y^* - \mu^*), \\ U_\phi &= \operatorname{tr}(HD^*), \end{aligned} \tag{2.2.6}$$

em que Z é uma matriz $n \times M$ de valores fixos conhecidos cuja t-ésima linha é $z_t^\top = (z_{t1}, \ldots, z_{tM})$, X é uma matriz $n \times m$ de valores fixos conhecidos cuja t-ésima linha é $x_t^\top = (x_{t1}, \ldots, x_{tm})$, $y^* = (y_1^*, \ldots, y_n^*)^\top$, $\mu^* = (\mu_1^*, \ldots, \mu_n^*)^\top$, $y^c = \left(\mathbb{1}_{\{c\}}(y_1), \ldots, \mathbb{1}_{\{c\}}(y_n)\right)^\top$ e $\alpha^* = (\alpha_1, \ldots, \alpha_n)^\top$. Com

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

e

$$\mu_t^* = \mathbb{E}(y_t^* | \mathbb{1}_{\{c\}}(y_t) = 0) = \psi(\mu_t\phi) - \psi((1-\mu_t)\phi).$$

Sejam $P = \text{diag}\{1/[\alpha_1(1-\alpha_1)], \ldots, 1/[\alpha_n(1-\alpha_n)]\}$, $H = \text{diag}\{1 - \mathbb{1}_{\{c\}}(y_1), \ldots, 1 - \mathbb{1}_{\{c\}}(y_n)\}$, $G = \text{diag}\{d\alpha_1/d\zeta_1, \ldots, d\alpha_n/d\zeta_n\}$ e $D^* = \text{diag}\{d_1^*, \ldots, d_n^*\}$ matrizes diagonais, com $d_t^* = \mu_t(y_t^* - \mu_t^*) + s(y_t) + \psi(\phi) - \psi((1-\mu_t)\phi)$ e

$$s(y_t) = \begin{cases} \log(1-y_t), & y_t \in (0,1), \\ 0, & y = c. \end{cases}$$

O estimador de máxima verossimilhança de γ é obtido como solução do sistema não-linear $U_\gamma(\gamma) = 0$. Por outro lado, o estimador de máxima verossimilhança de $(\beta^\top, \phi)^\top$ é obtido como solução do sistema não-linear $(U_\beta(\beta,\phi)^\top, U_\phi(\beta,\phi)^\top) = 0$. Tais estimadores não possuem forma fechada e devem ser obtidos numericamente pela maximização da função de log-verossimilhança usando um algoritmo de otimização não-linear, tal como um algoritmo de Newton ou um algoritmo quasi-Newton.

A matriz de informação de Fisher para o modelo beta inflacionado pode ser escrita como

$$K(\theta) = \begin{pmatrix} K_\gamma(\gamma) & 0 \\ 0 & K_\vartheta(\vartheta) \end{pmatrix}, \tag{2.2.7}$$

em que $K_\gamma(\gamma) = K_{\gamma\gamma}$ é a matriz de informação de Fisher de γ dada por

$$K_{\gamma\gamma} = Z^\top QZ,$$

$Q = GPG = \text{diag}\{q_1, \ldots, q_n\}$ sendo uma matriz diagonal e $q_t = p_t(d\alpha_t/d\zeta)^2$.

Adicionalmente, tem-se que

$$K_\vartheta(\vartheta) = \begin{pmatrix} K_{\beta\beta} & K_{\beta\phi} \\ K_{\phi\beta} & K_{\phi\phi} \end{pmatrix}$$

é a matriz de informação de Fisher de $\vartheta = (\beta^\top, \phi)$, com $K_{\beta\beta} = \phi^2 X^\top \Delta T W T X$, $K_{\beta\phi} = K_{\phi\beta}^\top = X^\top \Delta T c$ e $K_{\phi\phi} = \text{tr}(\Delta D)$. Aqui, $\Delta = \text{diag}\{\delta_1, \ldots, \delta_n\}$ e $W = \text{diag}\{w_1, \ldots, w_n\}$ Para $t = 1, \ldots, n$, $\delta_t = 1 - \alpha_t$ e $w_t = \psi'(\mu_t\phi) + \psi'((1-\mu_t)\phi)$.

2.2.3 Quasi-Verossimilhança

Modelos de quasi-verossimilhança são baseados nas especificações de como a resposta média está relacionada às variáveis explicativas (função de ligação) e como a variância está relacionada à média (função de variância). Um método para calcular as estimativas de máxima quasi-verossimilhança foi obtido por Wedderburn (1974) através da generalização do método de Gauss-Newton para cálculo de estimativas de mínimos quadrados não-lineares.

A construção da função de quasi-verossimilhança é similar à construção da função de log-verossimilhança em modelos lineares generalizados. A parte sistemática relaciona a esperança de uma observação ao preditor linear via uma função de ligação. A parte aleatória especifica como a variância é relacionada à média através de uma função de variância que depende de um parâmetro de dispersão.

Considere um vetor de respostas independentes $y = (y_1, \ldots, y_n)^\top$ com vetor de médias $\mu = (\mu_1, \ldots, \mu_n)^\top$ e matriz de covariância $\sigma^2 V(\mu)$. Assuma que μ é uma função das covariáveis e dos parâmetros de regressão $\beta = (\beta_1, \ldots, \beta_m)^\top$. Em geral, σ^2 é desconhecido e não depende de β. Supondo que os $y's$ são independentes, $V(\mu)$ deve ser diagonal, ou seja, $V(\mu) = \text{diag}\{V_1(\mu), \ldots, V_n(\mu)\}$. Outra suposição necessária é que $V_t(\mu)$ depende apenas do t-ésimo componente de μ.

Para cada observação (aqui eliminamos o subscrito t), a função de quasi-verossimilhança pode ser definida pela relação

$$Q(y;\mu) = \int_y^\mu \frac{y-s}{\sigma^2 V(s)} ds.$$

A função $Q(y;\mu)$ deve comportar-se como a função de log-verossimilhança e, caso uma família exponencial uniparamétrica seja especificada para a variável resposta, a função $Q(y;\mu)$ torna-se a função de log-verossimilhança da distribuição.

Uma vez que os componentes do vetor y são independentes por suposição, a função de quasi-verossimilhança para todas as observações é a soma das contribuições individuais. As equações de estimação de quasi-verossimilhança para os parâmetros β são obtidas pela diferenciação da função de quasi-verossimilhança e podem ser escritas como

$$U(\beta) = B^\top V^{-1}(y-\mu)/\sigma^2,$$

em que B é uma matriz $n \times m$ com elementos $B_{tr} = \partial\mu_t/\partial\beta_r$, $r = 1, \ldots, m$ e a função $U(\beta)$ é chamada de função quasi-escore. A sequência de estimativas pode ser gerada pelo método de Newton-Raphson com escore de Fisher.

2.3 Incorporando dependência espacial aos modelos de regressão

A independência é uma suposição muito conveniente e que torna muitas modelagens estatísticas tratáveis, contudo, modelos que envolvem dependência estatística são frequentemente mais realistas (Cressie, 1993). Enquanto as ferramentas usuais da estatística supõem que as variações são aleatórias no espaço, a estatística espacial considera existir uma dependência da variação com o espaço de amostragem.

Uma das áreas da estatística espacial, cujas ferramentas podem ser utilizadas para analisar dados de eficiência de municípios, é a análise de dados de área. Esses dados ocorrem com muita frequência quando se lida, por exemplo, com eventos agregados por municípios, bairros, setores censitários ou zonas climáticas dos quais não se dispõe da localização exata dos eventos, mas de um valor por área. Isto é, a análise de dados de áreas usa atributos que não variam continuamente, mas que possuem valores específicos para subáreas que compõem uma dada região em estudo. Em geral, os dados são associados a levantamentos populacionais tais como censos, estatísticas de saúde e cadastramento de imóveis. Alguns exemplos de dados de área são: taxa de mortalidade ou taxa de incidência relacionadas a alguma doença, percentual de adultos analfabetos, renda média e mediana etária. Este tipo de dado pode geralmente ser visualizado em mapas em que o espaço é particionado em áreas contíguas e disjuntas, coloridas de acordo

com alguma variável. Em cada uma dessas áreas, medem-se uma ou mais variáveis aleatórias e possivelmente covariáveis de interesse que podem afetar a distribuição de probabilidade das variáveis medidas. Assim, para uma melhor representação dos dados de área, considere uma região A, que é em geral uma região limitada e com área finita, particionada em n sub-regiões disjuntas A_i, $i = 1,\ldots,n$. Os dados de área são representados por um vetor $n \times 1$ aleatório $\mathbf{Z} = (z_1,\ldots,z_n)^\top$, em que se associa a variável aleatória z_i à região A_i.

Em estatística clássica, o conceito de correlação diz respeito à relação entre duas variáveis e pode ser traduzido pelo índice usual de correlação que mede o grau de associação entre as variáveis. A autocorrelação espacial mede o quanto o valor observado de uma variável em uma região está relacionado aos valores desta mesma variável nas localidades vizinhas. A noção de autocorrelação espacial está associada à idéia de que valores observados em áreas geográficas adjacentes se mostram mais parecidos ou diferentes do que o esperado sob a hipótese de que a distribuição das variáveis é invariante por permutação dos índices que localizam as áreas no espaço (Assunção, 2001). Assim, a dependência espacial pode ocorrer quando áreas próximas tendem a apresentar valores similares (autocorrelação positiva), isto é, a presença de um determinado fenômeno em uma região influencia as regiões próximas com o mesmo fenômeno, ou quando áreas tendem a apresentar valores dissimilares (autocorrelação negativa), ou seja, a presença deste fenômeno dificulta a sua aparição em regiões vizinhas.

Uma estatística padrão que costuma ser utilizada para medir a associação espacial entre unidades de área é o índice de Moran. O índice de Moran (I) foi introduzido pelo estatístico australiano Moran em 1950 e é a medida de autocorrelação espacial mais utilizada. Considere que $z_i,\ldots,z_n$ são as variáveis aleatórias medidas nas n áreas. O índice de Moran é calculado como

$$\mathrm{I} = \frac{n \sum_{i \neq j} w_{ij}(z_i - \overline{z})(z_j - \overline{z})}{S_0 \sum_{i=1}^{n}(z_i - \overline{z})^2}, \tag{2.3.1}$$

em que S_0 é a soma $\sum_{i \neq j} w_{ij}$ e w_{ij} são os elementos da matriz quadrada, $n \times n$, W. Essa matriz, denominada matriz de proximidade espacial (matriz de vizinhanças), é uma ferramenta essencial para descrever o arranjo espacial dos dados. Os elementos w_{ij} ($w_{ij} \geq 0$ e $w_{ii} = 0$) da matriz W refletem a intensidade da interdependência existente entre as regiões A_i e A_j, ou seja, representam o peso ou o grau de proximidade espacial entre as áreas A_i e A_j. A escolha dos elementos w_{ij} é arbitrária e pode ser feita considerando-se vários critérios. No presente trabalho a escolha dos elementos da matriz de proximidades,denominada aqui de $W1$, é feita da seguinte forma: para $i \neq j$, $w_{ij} = 1$ se a distância entre os municípios i e j não excede 50 quilômetros e $w_{ij} = 0$, caso contrário. É importante salientar que também foi considerada uma outra especificação dos elementos da matriz de proximidades: para $i \neq j$, w_{ij} é igual ao quociente entre a distância entre os municípios i e j e a distância máxima encontrada. Contudo, por apresentar resultados similares aos obtidos utilizando-se a matriz $W1$, esses resultados não serão apresentados aqui.

Na análise de regressão clássica, a hipótese padrão é que as observações são não-correlacionadas e, consequentemente, os resíduos do modelo são não-correlacionados com a variável dependente, que deve apresentar distribuição normal com média zero e variância constante. No caso de dados espaciais, quando há autocorrelação espacial, é pouco provável que a suposição

de que as observações são não-correlacionadas seja satisfeita. Uma vez que essa dependência entre as observações altera o poder explicativo do modelo, modelos de regressão espacial devem ser utilizados para que as estimativas incorporem essa estrutura espacial. Uma maneira de considerar a dependência espacial nos modelos de regressão é através da classe de modelos espaciais auto regressivos *Spatial Auto Regresive* (SAR) introduzida por Whittle (1954). A dependência espacial é considerada através da adição ao modelo de regressão de um novo termo na forma de uma relação espacial para a variável dependente. A expressão geral desse modelo é

$$y = X\beta + \rho Wy + \varepsilon,$$

em que β é um vetor de parâmetros desconhecidos, X é uma matriz que contém as observações das covariáveis utilizadas no modelo, ρ é o coeficiente espacial autorregressivo e W é a matriz de proximidade espacial. O produto Wy representa a dependência espacial em y e ε denota um erro aleatório com média zero e variância constante.

Nesse trabalho, a correlação espacial existente nos dados é incorporada aos modelos de regressão utilizados através da adição de uma variável explicativa que considera esta dependência espacial. De forma similar ao modelo SAR, a variável utilizada é o produto entre a matriz de vizinhanças e a variável dependente. A estimação dos parâmetros é realizada seguindo os passos usuais para os modelos de regressão como descrito na Seção 2.2.

2.4 Avaliação da eficiência administrativa de municípios do estado de São Paulo

Na presente seção, nós apresentamos resultados empíricos para o estado de São Paulo obtidos a partir de modelagens beta e beta inflacionada. Resultados obtidos via quasi-verossimilhança e regressão linear são também apresentados. A análise leva em consideração efeitos de natureza espacial. Os resultados apresentados conduzem a conclusões importantes, como será visto. Nós utilizamos a linguagem de programação `R` (`http://www.r-project.org`) para realizar as estimações. Os modelos de regressão beta e beta inflacionados foram ajustados usando as distribuições BE e BEOI do pacote `gamlss`. A análise espacial foi realizada usando os pacotes `spdep` e `shapefiles`.

Os dados utilizados fazem parte de um trabalho desenvolvido por Sampaio de Souza, Cribari-Neto e Stosic (2005), que estimaram índices de eficiência para municípios brasileiros utilizando, de forma combinada, o método bootstrap e o método de reamostragem jackknife para reduzir a influência de *outliers* e possíveis erros de medida no conjunto de dados. Os escores de eficiência dizem respeito à eficiência técnica de cada município no que se refere ao gerenciamento do gasto público. A técnica DEA, utilizada pelos autores para o cálculo dos escores de eficiência, é uma técnica não-paramétrica usada para obtenção de medidas relativas de eficiência que permitem a classificação das unidades em estudo como sendo eficientes ou ineficientes, quando comparadas com um conjunto de referência teórico. Esta técnica, introduzida por Charnes, Cooper e Rhodes (1978), teve seus fundamentos baseados no método proposto por Farrel (1957) para determinar uma medida de eficiência de uma organização comparando-a

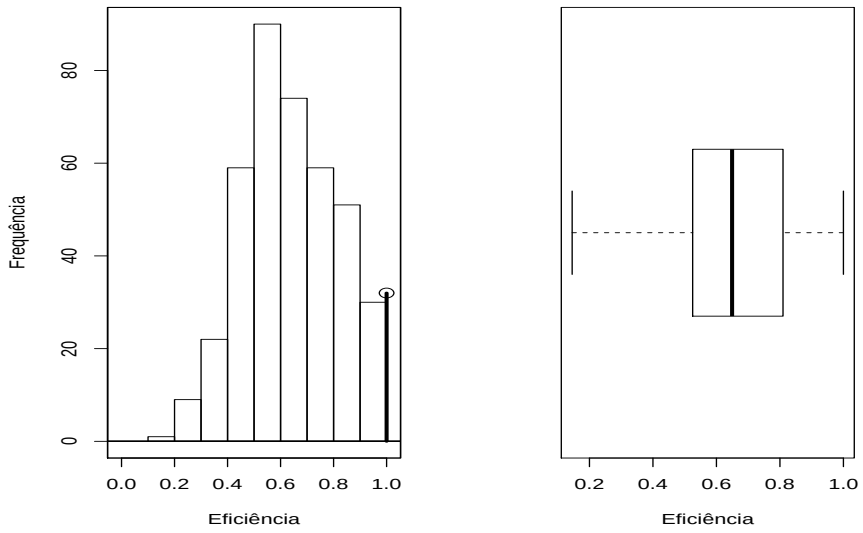

Figura 2.1 Histograma de frequências e gráfico box-plot do escore de eficiência para os dados de São Paulo.

com o melhor nível de eficiência observado. Os autores adotaram a suposição de rendimentos constantes de escala, conhecida como modelo CCR (Charnes, Cooper e Rhodes, 1978), e também a suposição de rendimentos variáveis de escala, conhecida como modelo BCC (Banker, Charnes e Cooper, 1984). Após o cálculo dos escores de eficiência, os autores utilizaram regressão linear e regressão quantílica para identificar os fatores que explicam a variabilidade dos escores de eficiência computados para os municípios brasileiros.

Santos, Cribari-Neto e Sampaio de Souza (2007) refizeram a avaliação citada acima considerando a idade dos municípios. As avaliações foram realizadas a partir dos fundamentos teóricos das técnicas DEA, regressão clássica e regressão quantílica.

Neste trabalho, os escores de eficiência baseados no modelo DEA-CCR foram utilizados como variável dependente nos modelos de regressão. A base de dados final utilizada contém 427 municípios do estado de São Paulo e as informações referem-se ao ano de 2000. As Tabelas 2.1 e 2.2 apresentam as descrições da variável dependente e das covariáveis utilizadas como regressoras, bem como algumas medidas descritivas que fornecem uma visão geral a respeito do comportamento de tais covariáveis. Na Tabela 2.2 as variáveis são do tipo *dummy*.

Na Figura 2.1 é apresentado o histograma, bem como o gráfico box-plot da variável dependente $EFIC$ para os municípios na amostra. É possível notar que a distribuição dessa variável é assimétrica. A barra vertical com um ponto acima no histograma representa a quantidade de uns na amostra, que corresponde a 7.49% dos dados.

Para verificar a existência de associação espacial entre os municípios foi utilizado o índice global de Moran (I). Considerando a hipótese nula de independência o valor obtido para o índice foi de I = 0.2104 e através do teste de permutação obteve-se p-valor = 0.001. Ou seja, a hipótese nula de independência é rejeitada aos níveis usuais de significância. Como o valor

Tabela 2.1 Variáveis contínuas para os dados de eficiência.

Variável	**Descrição**	**Média (Desvio Padrão)**
$EFIC$	Escores de eficiência calculados pelo método DEA (CCR)	0.66 (0.19)
$DESP$	Despesas com pessoal	1.95e+07 (9.32e+07)
SAL	Percentual de domicílios cujo chefe ganha até um salário mínimo	2.01 (1.84)
$REND$	Rendimento médio	700.15 (232.37)
CAD	Índice de atualização do cadastro predial	0.91 (0.12)
$DENS$	Densidade demográfica	306.89 (1146.85)
URB	Taxa de urbanização	82.45 (15.67)

Tabela 2.2 Variáveis *dummy* para os dados de eficiência.

Variável	Descrição	Frequência (%)
PFL	1 se o prefeito do município é do partido político PFL e 0 caso contrário	52 (12.18)
$PMDB$	1 se o prefeito do município é do partido político PMDB e 0 caso contrário	76 (17.80)
$PSDB$	1 se o prefeito do município é do partido político PSDB e 0 caso contrário	118 (27.63)
PT	1 se o prefeito do município é do partido político PT e 0 caso contrário	30 (7.02)
PPS	1 se o prefeito do município é do partido político PPS e 0 caso contrário	24 (5.62)
PPB	1 se o prefeito do município é do partido político PPB e 0 caso contrário	21 (4.92)
PTB	1 se o prefeito do município é do partido político PTB e 0 caso contrário	50 (11.71)
PDT	1 se o prefeito do município é do partido político PDT e 0 caso contrário	14 (3.28)
PCI	1 se o município participa de consórcios intermunicipais e 0 caso contrário	75 (17.56)
$INFO$	1 se o município é informatizado e 0 caso contrário	390 (91.33)
$PDCM$	1 se o município tem poder de decisão nos conselhos municipais e 0 caso contrário	217 (50.82)
$ALVO$	1 se o município participa do Projeto Alvorada e 0 caso contrário	3 (0.70)
$IDADE$	1 se a idade do município é menor ou igual a 8 anos e 0 caso contrário	36 (8.43)
MT	1 se o município é turístico e 0 caso contrário	72 (16.86)
ROY	1 se o município recebe mais de 10% da sua receita tributária a título de *royalties* e 0 caso contrário[a]	55 (12.88)

[a] Neste trabalho, *royalties* são uma compensação financeira decorrente da extração de recursos minerais.

do coeficiente é positivo, áreas próximas tendem a apresentar valores similares, ou seja, a presença de um município eficiente em uma região influencia positivamente as regiões próximas. A autocorrelação espacial ignorada é incorporada como componente do modelo através da variável $W1EFIC$, que é o produto entre a matriz de proximidades utilizada, $W1$, e a variável dependente.

Os modelos de regressão beta e beta inflacionado em um foram utilizados para modelar os dados. Contudo, como os escores de eficiência estão distribuídos no intervalo $(0,1]$ foi necessário, para o modelo beta, substituir os valores 1 dos escores de eficiência por 0.99999.

Foi realizado o teste da razão de verossimilhanças, descrito a seguir, para verificar se a hipótese de dispersão constante está sendo violada. Para a realização do teste é necessário considerar que o parâmetro de precisão varia ao longo das observações e, em adição à modelagem da média da variável resposta e do parâmetro de mistura α (no caso do modelo beta inflacionado), definir uma estrutura de regressão para o parâmetro de precisão, tal como

$$b(\phi_t) = \sum_{i=1}^{q} s_{ti}\lambda_i = \kappa_t, \tag{2.4.1}$$

em que $\lambda = (\lambda_1, \ldots, \lambda_q)^\top$ é um vetor de parâmetros desconhecidos, b é uma função real, positiva, estritamente monótona e duas vezes diferenciável e $s_{t1}, \ldots, s_{tq}$ são observações de q covariáveis conhecidas ($q < n$) que podem coincidir total ou parcialmente com as covariáveis consideradas na estrutura de regressão para a média e para α. A hipótese nula a ser testada é $\mathcal{H}_0 : \phi_1, \ldots, \phi_n = \phi$. Testar essa hipótese equivale a testar a hipótese nula $\mathcal{H}_0 : \lambda_2, \ldots, \lambda_q = 0$, no modelo 2.4.1 com $s_{t1} = 1$, $\forall t$. A estatística do teste da razão de verossimilhanças é dada por

$$RV = 2\left\{\ell(\hat{\theta}) - \ell(\tilde{\theta})\right\},$$

em que $\theta = (\gamma^\top, \beta^\top, \lambda^\top)^\top$, $\ell(\hat{\theta})$ é o valor maximizado do logaritmo da função de verossimilhança do modelo considerando uma estrutura de regressão para ϕ e $\ell(\tilde{\theta})$ é o valor restrito maximizado do logaritmo da função de verossimilhança obtido pela imposição da hipótese nula. Dado que, sob condições usuais de regularidade e sob a hipótese nula, RV converge em distribuição para $\chi^2_{(q-1)}$, o teste pode ser realizado com base em valores críticos aproximados obtidos dessa distribuição.

Para os dados de eficiência, os testes da razão de verossimilhança foram realizados considerando para a estrutura de regressão dos parâmetros de precisão as variáveis apresentadas nas Tabelas 2.3 e 2.4. A partir dos testes realizados obtivemos os valores 99.39 (p-valor < 0.01) e 19.21 (p-valor < 0.01) para a estatística RV nos modelos de regressão beta e beta inflacionado, respectivamente, rejeitando, assim, ao nível de 5% de significância as hipóteses de que os modelos beta e beta inflacionados têm dispersão constante.

Nos modelos de regressão beta e beta inflacionado a função de ligação logit foi utilizada para a média e para o parâmetro α, no caso do modelo beta inflacionado. Para o parâmetro de precisão foi utilizada a função de ligação logarítmica. É importante salientar que foram utilizadas outras funções de ligação para a modelagem dos dados contudo estas foram as que apresentaram os melhores resultados. As Tabelas 2.3 e 2.4 apresentam, para os modelos beta

Tabela 2.3 Estimativas dos parâmetros do modelo beta com dispersão variável usando os escores de eficiência para os dados de São Paulo.

Modelo para μ			
Variáveis	**Estimativa**	**Desvio Padrão**	**p-valor**
constante	0.1764	0.2281	0.4398
W1EFIC	0.0207	0.0098	0.0355
REND	−0.0006	0.0002	0.0037
PT	0.5133	0.1936	0.0083
URB	0.0105	0.0030	0.0006
MT	−0.2310	0.1336	0.0844
Modelo para ϕ			
Variáveis	**Estimativa**	**Desvio Padrão**	**p-valor**
constante	5.010e-01	5.123e-01	3.286e-01
W1EFIC	−3.728e-02	1.038e-02	3.681e-04
DESP	−1.667e-09	5.738e-10	3.861e-03
SAL	8.598e-02	3.091e-02	5.653e-03
REND	1.022e-03	2.676e-04	1.550e-04
PFL	2.550e-01	1.372e-01	6.371e-02
PPS	7.313e-01	1.902e-01	1.396e-04
PPB	5.697e-01	2.040e-01	5.479e-03
CAD	−9.524e-01	3.760e-01	1.169e-02
PCI	−5.973e-01	1.214e-01	1.260e-06
INFO	4.663e-01	1.622e-01	4.257e-03
DENS	1.491e-04	4.837e-05	2.199e-03
URB	−7.416e-03	3.489e-03	3.412e-02
MT	3.206e-01	1.410e-01	2.347e-02
ROY	−6.900e-01	1.569e-01	1.398e-05

e beta inflacionado selecionados, as estimativas de máxima verossimilhança e seus desvios-padrão correspondentes.

No que se refere às diferenças entre os ajustes desses modelos, é possível observar na Tabela 2.4 que cinco covariáveis foram consideradas significativas para modelar a probabilidade de se observar um (i.e., eficiência plena) na amostra, entre elas a variável despesas com pessoal($DESP$) e informatização do município ($INFO$). Desse modo, os escores de eficiência no modelo beta inflacionado são explicados pelas variáveis $W1EFIC$, $DESP$, SAL, $REND$, PT, PCI, $INFO$ e URB, ao passo que no modelo beta, os escores de eficiência são explicados apenas pelas variáveis $W1EFIC$, $REND$, PT, URB e MT.

A avaliação da adequacidade do modelo tem um papel importante em análise de regressão. Várias medidas da qualidade do ajuste para um contexto além dos modelos lineares têm sido propostas na literatura. Uma alternativa simples, baseada no logaritmo da verossimilhança, é o pseudo R^2 de McFadden (1974) definido por

Tabela 2.4 Estimativas dos parâmetros do modelo beta inflacionado com dispersão variável usando os escores de eficiência para os dados de São Paulo.

Modelo para μ			
Variáveis	**Estimativa**	**Desvio Padrão**	**p-valor**
constante	0.1403	0.2012	4.860e-01
W1EFIC	0.0142	0.0080	7.571e-02
REND	−0.0007	0.0001	6.047e-07
PT	0.2776	0.1495	6.408e-02
URB	0.0095	0.0026	3.346e-04
Modelo para α			
Variáveis	**Estimativa**	**Desvio Padrão**	**p-valor**
constante	7.774e-01	1.271e+00	0.5411
DESP	9.363e-09	4.915e-09	0.05747
SAL	−5.244e-01	2.267e-01	0.0212
REND	−2.835e-03	1.511e-03	0.06136
PCI	1.357e+00	4.296e-01	0.0017
INFO	−1.068e+00	5.248e-01	0.0425
Modelo para ϕ			
Variáveis	**Estimativa**	**Desvio Padrão**	**p-valor**
constante	2.0711	0.3028	2.806e-11
W1EFIC	−0.0384	0.0139	6.005e-03
REND	0.0012	0.0003	1.097e-04
INFO	−0.7416	0.2538	3.667e-03

$$PR^2 = 1 - \frac{\hat{\ell}_N}{\hat{\ell}_F},$$

em que $\hat{\ell}_F$ é a log-verossimilhança maximizada do modelo ajustado e $\hat{\ell}_N$ é a log-verossimilhança maximizada do modelo nulo, que é o modelo sem estrutura de regressão. Os pseudo R^2 obtidos das regressões beta e beta inflacionada foram, respectivamente, 0.31 e 0.86. Nota-se, assim, uma grande superioridade do modelo de regressão beta inflacionado no que se refere a essa medida de qualidade do ajuste.

Ospina (2008) definiu o resíduo quantil aleatorizado para o modelo de regressão beta inflacionado em zero ou um como

$$r_t^q = \Phi^{-1}(u_t), \quad t = 1, \ldots, n,$$

em que u_t é uma variável aleatória uniforme no intervalo $(a_t, b_t]$, sendo $a_t = \lim_{y \uparrow y_t} \mathrm{BI}_c(y; \hat{\alpha}, \hat{\mu}, \hat{\phi})$, $b_t = \mathrm{BI}_c(y_t; \hat{\alpha}, \hat{\mu}, \hat{\phi})$, $\mathrm{BI}_c(y_t; \alpha, \mu, \phi)$ a função de distribuição acumulada beta inflacionada e Φ^{-1} a função quantil normal padrão. Como esse resíduo pode variar de uma realização a outra, é aconselhável, em situações práticas, obter pelo menos quatro realizações dos resíduos para detectar padrões no seu comportamento.

Com o intuito de verificar possíveis afastamentos das suposições feitas para o modelo, a Figura 2.2 apresenta os gráficos dos resíduos quantis aleatorizados contra os índices das observações, bem como o gráfico normal de probabilidade com envelopes simulados. A dispersão dos resíduos é maior para o modelo beta do que para o modelo beta inflacionado. Uma vez que os resíduos, em geral, permanecem dentro das bandas de confiança dos envelopes simulados, é possível verificar que não há indícios claros de afastamento da suposição de que o modelo de regressão beta inflacionado é adequado para os dados. Por outro lado, para o modelo de regressão beta, há fortes indícios da falta de qualidade do ajuste do modelo.

De acordo com os resultados obtidos, o modelo de regressão beta inflacionado com dispersão variável foi escolhido para modelar os dados de eficiência dos municípios paulistas. Os parâmetros da regressão beta inflacionada têm importantes interpretações. Suponha que o valor da i-ésima variável independente aumenta em uma unidade e todas as outras variáveis independentes permaneçam constantes. Seja μ a média da variável dependente sob os valores originais das covariáveis e $\mu^\dagger$ a média da variável dependente sob os novos valores das covariáveis. Considerando a função de ligação logit para a média, a razão de chances (*odds ratio*) é dada por

$$\frac{\mu^\dagger/(1-\mu^\dagger)}{\mu/(1-\mu)} = \exp(\beta_i).$$

Ou seja, β_i pode ser interpretado como o logaritmo da razão de chances.

Através da análise dos coeficientes estimados para o modelo selecionado (ver Tabela 2.4) é possível verificar que a variável *PT* influencia positivamente a eficiência administrativa, ou seja, municípios gerenciados pelo partido dos trabalhadores tendem a apresentar um maior grau de eficiência administrativa. A variável que introduz a dependência espacial e a taxa de urbanização também exercem efeito positivo sobre a eficiência, ao passo que o rendimento

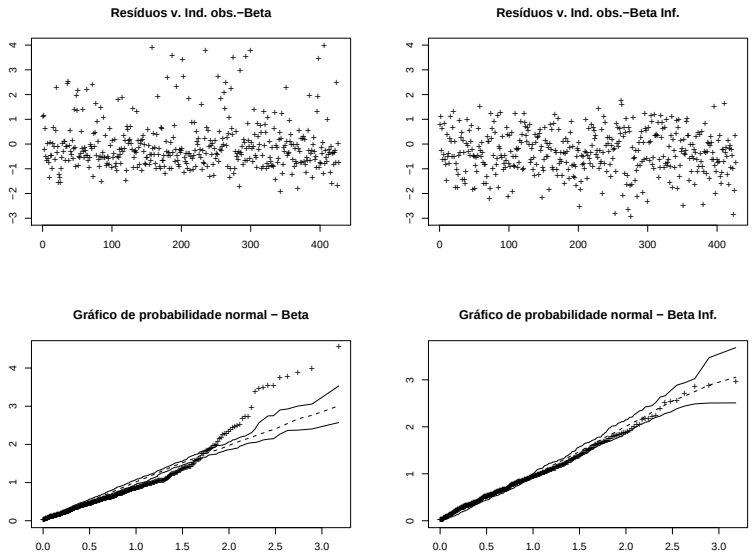

Figura 2.2 Gráficos de resíduos dos modelos beta e beta inflacionado para os dados de São Paulo.

médio exerce efeito negativo. Altos níveis de eficiência tendem a se propagar, pelo menos parcialmente, pelas localidades próximas. Os municípios mais urbanizados tendem a ser mais eficientes, dado que esta variável captura o fato de que recursos humanos e materiais tendem a ser mais escassos em áreas rurais e o custo médio de serviços públicos em áreas rurais pouco habitadas tende a ser alto. Dentre as variáveis significativas, a que menos exerce influência sobre a eficiência média do município é o rendimento médio.

Considerando a estrutura de regressão para o parâmetro de precisão, à medida que a variável $W1EFIC$ aumenta, a precisão diminui e, assim, aumenta a dispersão. Os municípios informatizados possuem escores de eficiência mais dispersos e os municípios cujos trabalhadores possuem maiores rendimentos médios possuem escores de eficiência menos dispersos.

O índice global de Moran (I) foi calculado para os resíduos obtidos a partir do modelo de regressão beta inflacionado. Considerando a hipótese nula de independência, o valor obtido para o índice foi de I = 0.0024 e através do teste de permutação obteve-se p-valor = 0.472. Ou seja, a hipótese nula de independência não é rejeitada aos níveis usuais de significância. Assim, os resíduos do modelo não apresentam dependência espacial, ou seja, a autocorrelação espacial incorporada através da variável $W1EFIC$ foi absorvida pelo modelo. Este fato revela a relevância de considerar o efeito da dependência espacial nos escores de eficiência.

Duas alternativas utilizadas aqui para modelar dados observados no intervalo $(0,1]$ são a utilização da função de quasi-verossimilhança para estimação dos parâmetros e a utilização do logaritmo das medidas de eficiência como variável dependente no modelo de regressão linear clássica. Para o ajuste considerando estimação por quasi-verossimilhança foi utilizada a função de ligação logit e a função de variância $V(\mu) = \mu(1-\mu)$.

Para o modelo linear, seis variáveis foram significativas, entre elas destacam-se as variáveis

Tabela 2.5 Estimativas dos parâmetros dos modelos linear e quasi-verossimilhança usando os escores de eficiência para os dados de São Paulo.

Regressão linear			**Estimação por quasi-verosimilhança**		
Variáveis	**Estimativa**	**Desvio Padrão**	**Variáveis**	**Estimativa**	**Desvio Padrão**
Constante	−5.623e-01	9.357e-02	Constante	0.1762	0.2485
REND	−2.080e-04	8.197e-05	W1EFIC	0.0178	0.0088
PT	1.609e-01	6.035e-02	REND	−0.0008	0.0002
INFO	−1.034e-01	5.460e-02	PT	0.4583	0.1803
URB	4.293e-03	1.134e-03	INFO	−0.2611	0.1547
IDADE	−1.051e-01	5.590e-02	URB	0.0131	0.0030
MT	−8.126e-02	4.589e-02	MT	−0.2279	0.1242

Tabela 2.6 Medidas descritivas das eficiências estimadas para os dados de São Paulo.

Medida	**Eficiência**	**Beta**	**Beta inflacionada**	**Quasi Veros.**	**Linear**
Mínimo	0.1455	0.4580	0.3856	0.3708	0.4029
$Q_{1/4}$	0.5251	0.6610	0.6450	0.6292	0.6027
Mediana	0.6493	0.6975	0.6682	0.6673	0.6395
Media	0.6625	0.6930	0.6690	0.6625	0.6341
$Q_{3/4}$	0.8097	0.7203	0.6926	0.6929	0.6626
Maximo	1.0000	0.8473	1.0000	0.8191	0.8071

$REND$, URB e $IDADE$. A variável que introduz a correlação espacial não foi significativa para o modelo de regressão linear. O R^2 do modelo linear é 0.07.[1] Fazendo estimação por quasi-verossimilhança, chega-se a um modelo com seis variáveis significativas, das quais se destacam as variáveis $INFO$, URB e $W1EFIC$.

A Tabela 2.6 apresenta algumas medidas descritivas das eficiências estimadas: mínimo, primeiro quartil ($Q_{1/4}$), mediana, média, terceiro quartil ($Q_{3/4}$) e máximo. O modelo beta inflacionado foi o único que forneceu escore de eficiência estimado acima de 0.85.

Na análise de dados espaciais a construção de mapas cumpre uma função importante na visualização dos dados. As Figuras 2.3, 2.4, 2.5 e 2.6 apresentam os mapas das eficiências estimadas a partir dos ajustes dos quatros modelos utilizados na análise. As áreas em branco referem-se aos municípios que não fizeram parte da amostra. É possível verificar que a cidade de São Paulo, que apresenta escore de eficiência igual a um, só apresenta eficiência estimada acima de 0.8 no mapa das eficiências estimadas a partir do modelo beta inflacionado. As eficiências estimadas para o município de São Paulo obtidas através dos ajustes dos modelos beta inflacionado, beta, quasi-verossimilhança e linear foram, respectivamente, 1.0000, 0.7336, 0.6691 e 0.6126.

A fim de analisar as qualidades das diferentes gestões administrativas no que tange ao ge-

[1] O pseudo R^2 obtido para este modelo foi 0.15.

Tabela 2.7 Medidas descritivas das eficiências brutas e puras para os dados de São Paulo.

Medida	Eficiência				
	Bruta	Pura			
		Beta Inflacionada	Beta	Quasi Veros.	Linear
Mínimo	0.1455	0.0000	0.0000	0.0000	0.0000
$Q_{1/4}$	0.5251	0.4056	0.4130	0.4005	0.3935
Mediana	0.6493	0.5414	0.5618	0.5390	0.5379
Media	0.6625	0.5549	0.5683	0.5471	0.5429
$Q_{3/4}$	0.8097	0.7137	0.7058	0.6813	0.6817
Maximo	1.0000	1.0000	1.0000	1.0000	1.0000

renciamento do gasto público calcularemos a seguir as 'eficiências técnicas puras' (ou 'eficiências puras') dos municípios. Para tanto, expurgaremos das eficiências brutas calculadas pelo método DEA os efeitos condicionantes estimados a partir dos modelos de regressão. As eficiências puras normalizadas são obtidas a partir da seguinte expressão:

$$\theta_t^{(p)} = \frac{r_t - \min(r_1, \ldots, r_n)}{\max\{r_1 - \min(r_1, \ldots, r_n), \ldots, r_n - \min(r_1, \ldots, r_n)\}}, \tag{2.4.2}$$

com

$$r_t = \theta_t - \hat{\theta}_t, \tag{2.4.3}$$

em que θ_t são as eficiências brutas DEA e $\hat{\theta}_t$ são as eficiências estimadas através do modelo de regressão considerado, $t = 1, \ldots, n$. As Tabelas 2.7 e 2.8 apresentam, respectivamente, algumas medidas descritivas das eficiências brutas e puras e os percentuais de municípios por intervalo de eficiência. Nota-se que as medidas de eficiência pura preditas pelos modelos de regressão se concentram mais em torno da média, isto é, municípios menos eficientes apresentam melhor desempenho e aqueles mais eficientes apresentam pior desempenho, quando comparadas com as medidas de eficiência bruta. Na Tabela 2.8 é possível verificar que apenas um município é apontado como plenamente eficiente: Corumbataí para os modelos beta inflacionado, linear e quasi-verossimilhança e Taubaté para o modelo beta.

A Tabela 2.9 apresenta as eficiências brutas e puras obtidas para os municípios de São Paulo e Lorena, que apresentaram eficiência bruta igual a 1, bem como para o município de Vinhedo que apresentou um valor baixo para a eficiência bruta. É importante notar a grande discrepância entre a eficiência pura estimada via modelo beta inflacionado e a eficiência bruta para o município de São Paulo. A discrepância entre a eficiência bruta e a eficiência pura foi de 0.4380 quando o expurgo dos fatores condicionantes é feito através do modelo beta inflacionado, sendo igual a 0.1196, 0.1128 e 0.0752 para os modelos beta, quasi-verossimilhança e linear. Nota-se, portanto, que o modelo beta inflacionado desconta muito mais fortemente o papel exercido por fatores condicionantes na eficiência administrativa desse importante município.

As Figuras 2.7, 2.8, 2.9, 2.10, 2.11 e 2.12 apresentam, respectivamente, os histogramas, o mapa das eficiências brutas obtidas pelo método DEA e os mapas das eficiências puras obtidas através dos modelos de regressão utilizados. Devido a maior concentração dos escores de

Tabela 2.8 Percentual de municípios de acordo com as medidas de eficiências brutas e puras para os dados de São Paulo.

Eficiência	**Eficiência**				
	Bruta	**Pura**			
		Beta Inflacionada	**Beta**	**Quasi Veros.**	**Linear**
0.0 |– 0.2	0.23	4.22	3.04	3.51	3.51
0.2 |– 0.4	7.26	19.67	18.97	21.55	22.95
0.4 |– 0.6	34.89	34.19	35.36	36.53	35.83
0.6 |– 0.8	31.15	28.57	27.87	26.70	25.53
0.8 |– 1.0	18.97	13.11	14.52	11.47	11.94
1.0	7.49	0.23	0.23	0.23	0.23

Tabela 2.9 Eficiências bruta e puras para os municípios de São Paulo, Lorena e Vinhedo.

Municípios	**Eficiência**				
	Bruta	**Pura**			
		Beta Inflacionada	**Beta**	**Quasi Veros.**	**Linear**
São Paulo	1.0000	0.5620	0.8804	0.8872	0.9248
Lorena	1.0000	0.9393	0.9063	0.8791	0.8892
Vinhedo	0.2846	0.2746	0.2201	0.2374	0.2234

Tabela 2.10 Medidas descritivas dos escores de eficiência para os dados de São Paulo, Brasil e demais estados.

Medida	**São Paulo**	**Brasil**	**Demais estados**
Mínimo	0.1455	0.0946	0.0946
$Q_{1/4}$	0.5251	0.3952	0.3873
Mediana	0.6493	0.5031	0.4915
Média	0.6625	0.5222	0.5083
$Q_{3/4}$	0.8097	0.6257	0.6048
Máximo	1.0000	1.0000	1.0000
p (%)	7.50	1.79	1.22

eficiência pura em torno da média, os histogramas das eficiências puras apresentam um formato mais simétrico em comparação ao histograma das eficiências brutas. Através da análise dos mapas é possível verificar uma redução nas eficiências puras em relação as eficiências brutas.

Com o intuito de avaliar a eficiência do gasto público municipal no Brasil, bem como verificar o impacto do estado de São Paulo no cenário brasileiro, no que se refere aos escores de eficiência, são apresentadas breves análises da eficiência de administrações públicas municipais para o Brasil e para os demais estados do Brasil, excluindo o estado de São Paulo.

As amostras contam com 4755 municípios para o Brasil e 4328 municípios para os demais estados, no agregado, excluindo o estado de São Paulo. O modelo de regressão beta inflacionado foi ajustado aos dados. Através dos testes da razão de verossimilhanças rejeitou-se a hipótese de dispersão constante. Os principais resultados obtidos estão descritos a seguir.

A Tabela 2.10 apresenta algumas medidas descritivas para os escores de eficiência, em que p representa o percentual de municípios eficientes na amostra. É possível destacar a superioridade do estado de São Paulo no tocante ao gerenciamento de recursos públicos. A eficiência média para o estado de São Paulo é de aproximadamente 0.66 e 7.50% dos municípios são considerados eficientes. Por outro lado, no Brasil, apenas 1.79% dos municípios são eficientes e a eficiência média é, aproximadamente, 0.52. Excluindo São Paulo, os demais estados possuem 1.22% dos municípios eficientes e a eficiência média é, aproximadamente, 0.51.

Os pseudo R^2 obtidos para os modelos de regressão beta inflacionados para o Brasil e para os demais estados foram, respectivamente, 0.38 e 0.31. Suspeita-se que os baixos valores para os pseudo R^2 possam ser devidos ao grande número de observações, bem como à grande heterogeneidade existente nos dados. A Figura 2.13 apresenta os gráficos dos resíduos quantis aleatorizados contra os índices das observações para o Brasil e para os demais estados.

A Tabela 2.11, a seguir, apresenta as estimativas dos parâmetros, com seus respectivos desvios-padrão. É importante notar que tanto para o Brasil como para os demais estados foram consideradas as variáveis CAP e $SECA$ que indicam, respectivamente, se o município é a capital do estado e se o município está localizado ou não no polígono da seca. Considerando a modelagem da média observa-se que ao se excluir o estado de São Paulo da análise, a variável CAP deixa de ser significativa e a variável CAD passa a ser significativa. Isto é,

excluindo o estado de São Paulo, para os demais estados, as capitais não apresentam diferenças significativas no que se refere à eficiência. Por outro lado o índice de atualização do cadastro predial passa a influenciar negativamente a eficiência administrativa nos demais estados. No que se refere à modelagem do parâmetro α as variáveis $ALVO$ e ROY tornam-se significativas. Na modelagem do parâmetro de dispersão as variáveis PTB e CAP passam a ser significativas e a variável $DENS$ deixa de ser significativa.

As variáveis que mais influenciam a eficiência administrativa dos municípios brasileiros são as variáveis CAP e ROY. As capitais dos estados têm, em geral, maiores graus de eficiência e os municípios que recebem *royalties* tendem a apresentar menores graus de eficiência tanto para o Brasil como para os demais estados. Esta é a variável que mais contribui para a diminuição da eficiência tanto para o cenário brasileiro como para os outros estados. Este fato exige um pouco de atenção dado que esta receita adicional deveria ser canalizada em prol de uma melhor qualidade de vida medida pelo acesso aos serviços públicos. Contudo, o recebimento de *royalties* gera gastos indevidos e, por consequência, despesas ineficientes. A variável $W1EFIC$ foi significativa para os dois cenários considerados e também exerce efeito positivo na eficiência. Tanto para o Brasil como para os demais estados, os municípios novos tendem a não apresentar bom desempenho no tocante à administração dos gastos públicos, dado que a falta de experiência em gestão pública e a falta de estrutura administrativa afetam a eficiência dos municípios.

Comparando-se os resultados obtidos para o Brasil com os resultados obtidos para o estado de São Paulo, verifica-se que o modelo obtido para o estado de São Paulo apresenta uma quantidade consideravelmente menor de variáveis significativas. As variáveis idade do município ($IDADE$) e recebimento de *royalties* (ROY), que são variáveis consideradas importantes em estudos anteriores para explicar escores de eficiência administrativa de municípios, deixaram de ser significativas. Uma constatação importante é que a variável PT (prefeito filiado ao PT), que foi a variável que mais influenciou a eficiência administrativa na análise do estado de São Paulo, não foi significativa para o Brasil. Estes resultados podem ser traduzidos pela superioridade econômica do estado de São Paulo perante os demais estados da federação. Não é exagero comparar a magnitude econômica do estado de São Paulo à de uma nação inteira, dado que o estado é o mais rico da federação, é o pólo econômico da América do Sul, possui o mais amplo parque industrial do país e um mercado de trabalho caracterizado pela qualificação de sua mão-de-obra. Dessa forma, por se destacar no contexto nacional pela expressiva participação na economia, é de se esperar que São Paulo apresente um comportamento diverso dos demais estados do Brasil no que se refere à eficiência administrativa de municípios. E isso ocorre.

2.5 Conclusões

Com o objetivo de avaliar a eficiência do gasto público municipal no estado de São Paulo, nós utilizamos neste capítulo modelos de regressão para lidar com dados em forma de frações, taxas ou proporções correlacionados espacialmente e medidos continuamente no intervalo aberto $(0,1)$, mas que podem conter zeros e/ou uns. A análise foi focada na comparação dos modelos de regressão beta e beta inflacionado.

Baseado no pseudo R^2 obtido e no gráfico dos resíduos o modelo de regressão beta infla-

Tabela 2.11 Estimativas dos parâmetros do modelos beta inflacionado com dispersão variável usando os escores de eficiência para os dados do Brasil e demais estados.

Brasil			**Demais estados**		
Modelo para μ			**Modelo para μ**		
Variáveis	**Estimativa**	**Desvio Padrão**	**Variáveis**	**Estimativa**	**Desvio Padrão**
constante	−0.5187	4.606e-02	Constante	−0.3824	8.074e-02
W1EFIC	0.0065	1.577e-03	W1EFIC	0.0054	$1.574e-03$
REND	−0.0002	6.414e-05	REND	−0.0003	6.870e−05
PMDB	−0.0550	2.116e-02	PMDB	−0.0604	2.119e−02
PTB	0.0756	3.478e-02	PTB	0.0744	3.810e-02
PCI	−0.1396	1.907e-02	CAD	−0.1609	7.120e-02
INFO	0.1400	2.103e-02	PCI	−0.1066	1.962e-02
PDCM	0.0817	1.789e-02	INFO	0.1387	2.104e-02
DENS	0.0002	3.598e-05	PDCM	0.0895	1.821e-02
URB	0.0104	4.677e-04	DENS	0.0004	5.232e-05
ALVO	0.1093	2.943e-02	URB	0.0098	4.758e-04
SECA	−0.1452	2.715e-02	ALVO	0.1366	2.959e-02
IDADE	−0.1345	2.607e-02	SECA	−0.1391	2.707e-02
CAP	0.5066	1.845e-01	IDADE	−0.1240	2.650e-02
ROY	−0.2561	3.136e-02	ROY	−0.2583	3.192e-02
Modelo para α			**Modelo para α**		
Variáveis	**Estimativa**	**Desvio Padrão**	**Variáveis**	**Estimativa**	**Desvio Padrão**
constante	−7.1387	4.912e-01	Constante	−7.6640	0.6508
DENS	0.0002	8.347e-05	DENS	0.0006	0.0001
URB	0.0429	6.108e-03	URB	0.0426	0.0078
IDADE	0.8770	2.939e-01	ALVO	0.5880	0.3087
			E25	1.2320	0.3330
			R2	−1.2867	0.6508
Modelo para ϕ			**Modelo para ϕ**		
Variáveis	**Estimativa**	**Desvio Padrão**	**Variáveis**	**Estimativa**	**Desvio Padrão**
constante	2.892e+00	1.907e-01	Constante	2.8457	0.2034
SAL	1.286e-02	2.968e-03	SAL	0.0116	0.0030
REND	4.735e-04	1.547e-04	REND	0.0006	0.0002
CAD	−6.020e-01	1.666e-01	PTB	−0.1550	0.0827
PCI	2.221e-01	4.274e-02	CAD	−0.4356	0.1747
INFO	1.511e-01	4.750e-02	PCI	0.1282	0.0452
DENS	−8.873e-05	4.267e-05	INFO	0.1973	0.0489
URB	−9.606e-03	1.100e-03	PDCM	−0.1025	0.0421
ALVO	−1.481e-01	6.541e-02	URB	−0.0091	0.0011
SECA	3.246e-01	6.154e-02	ALVO	−0.2049	0.0667
IDADE	−3.114e-01	5.643e-02	SECA	0.3262	0.0617
MT	−1.968e-01	8.316e-02	IDADE	−0.3489	0.0583
			MT	−0.4003	0.0936
			CAP	−0.6225	0.2765

cionado foi escolhido para modelar os dados. Como era de se esperar, a variável que introduz a correlação espacial, $W1EFIC$, foi significativa na estrutura de regressão para a média e para o parâmetro de precisão no modelo beta inflacionado, sendo assim importante considerar o efeito da dependência espacial em se tratando de dados de avaliações municipais. Municípios mais urbanizados e aqueles gerenciados por administrador filiado ao PT tendem a ter um maior grau de eficiência administrativa. Por outro lado, baseado na renda média dos trabalhadores, constatou-se que municípios mais pobres tendem a administrar melhor seus recursos.

Duas alternativas foram utilizadas para lidar com dados observados no intervalo $(0,1]$: a função de quasi-verossimilhança para estimação dos parâmetros e o logaritmo das medidas de eficiência como variável dependente no modelo de regressão linear clássica. O R^2 obtido para o modelo linear foi inferior ao pseudo R^2 obtido para o modelo beta inflacionado e a variável que introduz a dependência espacial não foi significativa para o modelo de regressão linear. O modelo beta inflacionado foi o único que forneceu escore de eficiência estimado acima de 0.85.

Com o intuito de avaliar o gasto público municipal no Brasil, bem como verificar o impacto do estado de São Paulo no cenário brasileiro, foram apresentadas breves análises da eficiência de administrações públicas municipais para o Brasil e para os demais estados do Brasil excluindo o estado de São Paulo. Os resultados apontaram para a grande superioridade do estado de São Paulo no que tange à eficiência administrativa de municípios. A eficiência média do estado superou em mais de 26% a eficiência média do Brasil e o percentual de municípios eficientes foi bem maior para o estado de São Paulo do que para o Brasil e para os demais estados, excluindo o estado de São Paulo. Para o Brasil, as capitais dos estados tendem a apresentar um maior grau de eficiência e os municípios que recebem *royalties* tendem a apresentar um menor grau tanto para o Brasil como para os outros estados. Esta estratégia administrativa para o recebimento de *royalties* merece mais atenção, pois o recebimento desta receita adicional ao invés de incentivar a utilização ótima dos recursos e promover a eficiência pode contribuir para o aumento da ineficiência administrativa do município. Municípios novos tendem a apresentar menores escores de eficiência e, assim, não apresentam bom desempenho no tocante à administração dos gastos públicos.

Comparando-se os resultados obtidos para o Brasil com os resultados obtidos para o estado de São Paulo, constatamos que variáveis como idade do município e recebimento de *royalties* deixaram de ser significativas. Ou seja, para o estado de São Paulo, o recebimento de *royalties* e a criação de novos municípios não interferem na eficiência administrativa. A vinculação partidária dos prefeitos dos municípios ao partido político PT não foi significativa para o Brasil. Entretanto para São paulo essa foi a variável que mais influenciou a eficiência administrativa. Por ser um estado à parte, com desempenho econômico bastante peculiar e superior aos dos outros estados, algumas variáveis que parecem interferir na eficiência não apresentaram efeito significativo para o estado de São Paulo.

De forma geral, o modelo de regressão beta inflacionado mostrou-se o mais adequado para explicar os escores de eficiência, visto que não foi necessária transformação para lidar com os dados, os escores de eficiência iguais a um foram facilmente acomodados no modelo, foi possível modelar a dispersão variável presente neste tipo de dado e foram obtidos resultados mais confiáveis no que se refere à qualidade do ajuste. A partir dos resultados obtidos é possível fornecer uma visão geral de como os municípios têm se comportado com relação ao gerencia-

mento de recursos públicos, avaliar os desempenhos de administrações municipais e fornecer instrumentos que podem ser usados para avaliar os governos locais.

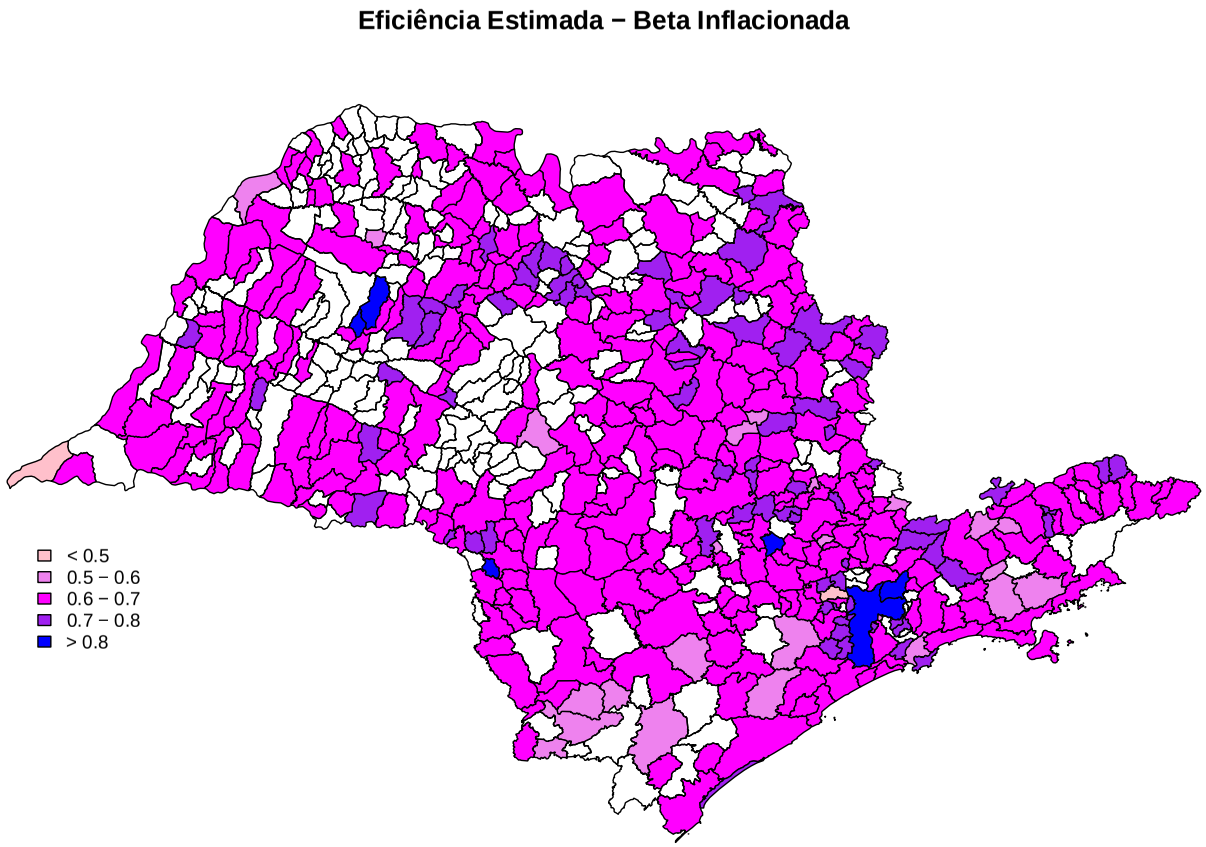

Figura 2.3 Mapa das eficiências estimadas para os dados de São Paulo - Beta Inflacionada.

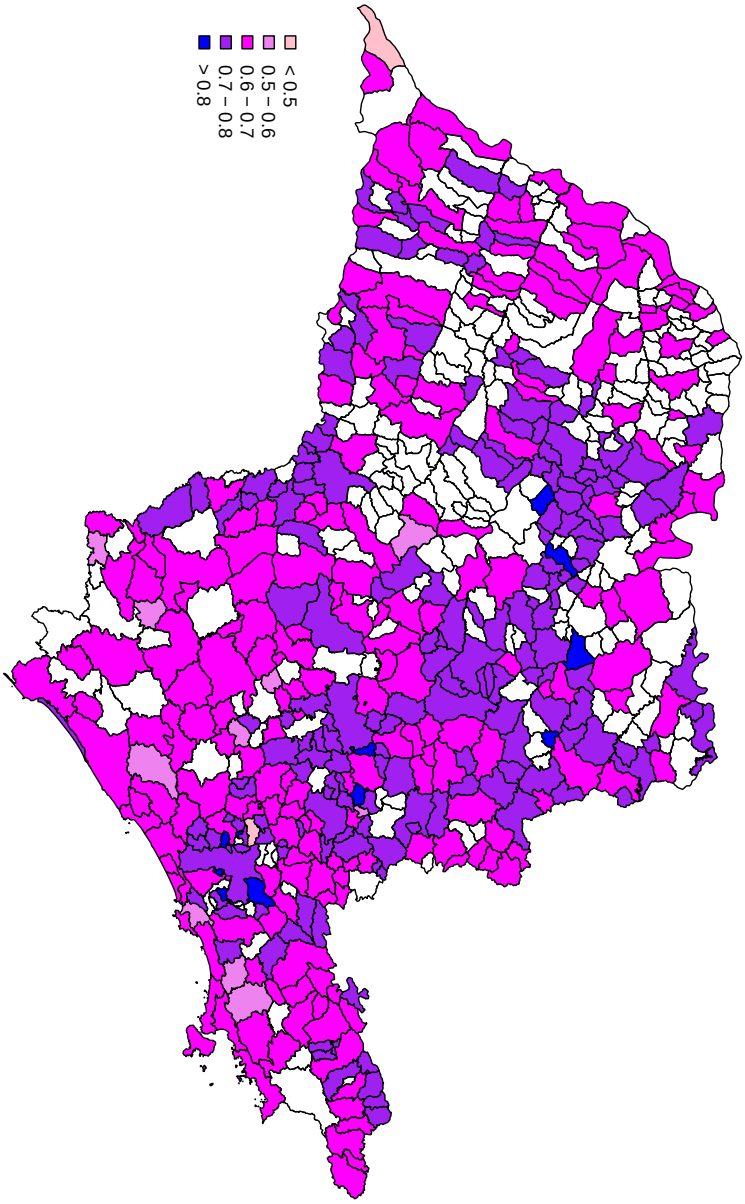

Figura 2.4 Mapa das eficiências estimadas para os dados de São Paulo - Beta.

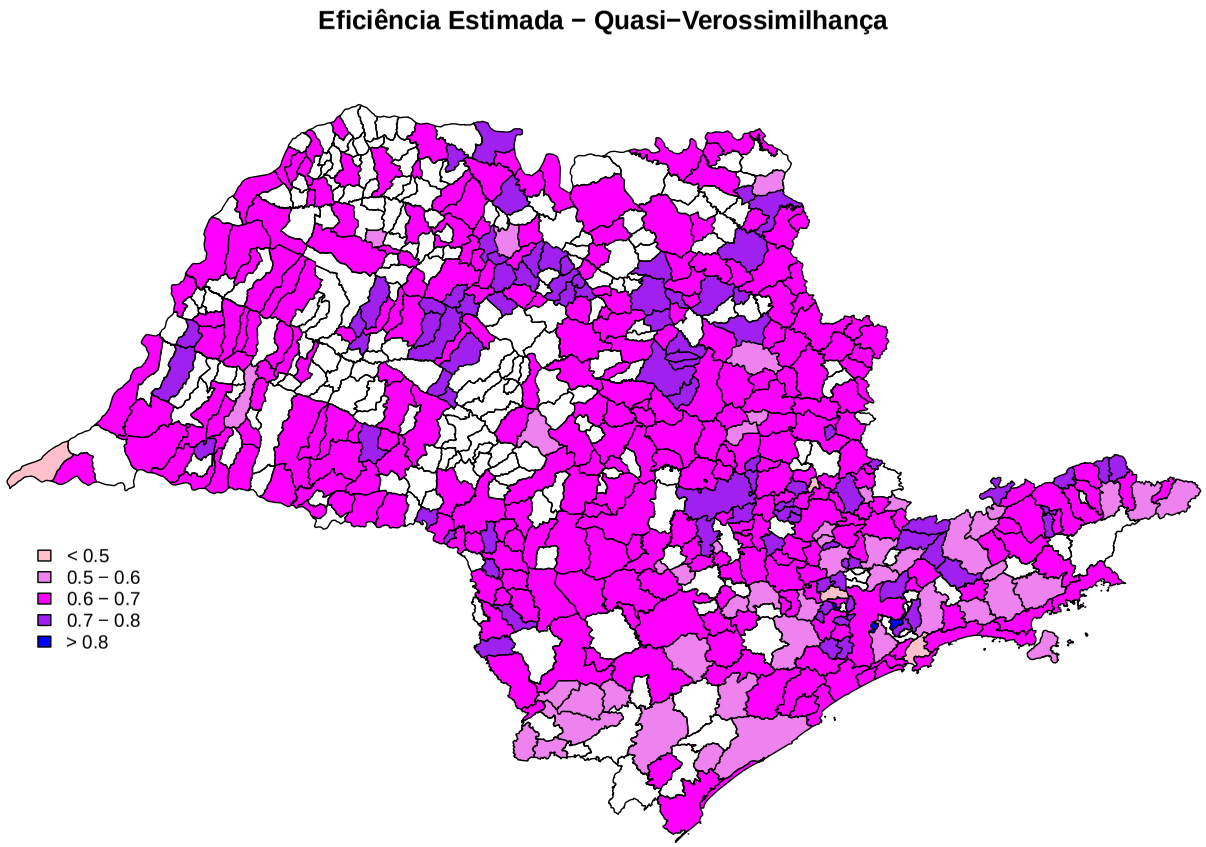

Figura 2.5 Mapa das eficiências estimadas para os dados de São Paulo - Quasi-Verossimilhança.

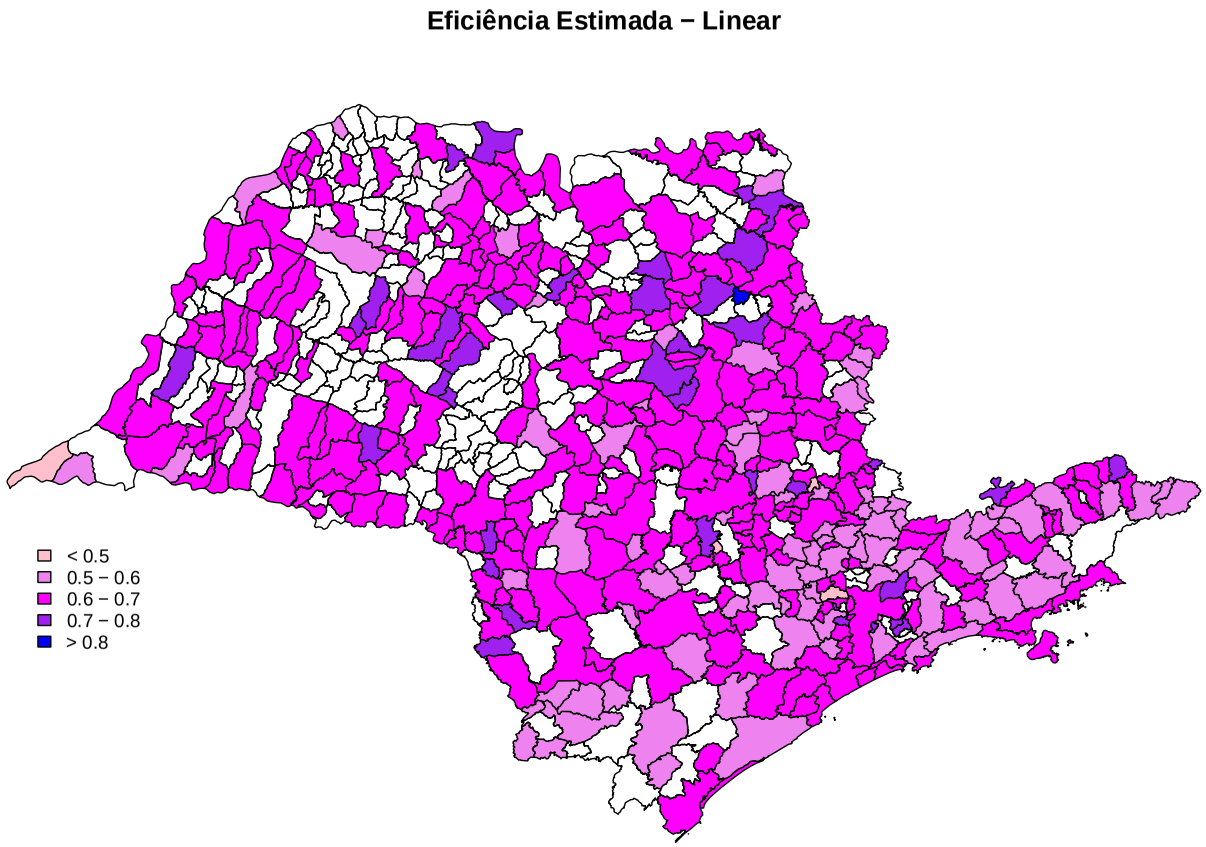

Figura 2.6 Mapa das eficiências estimadas para os dados de São Paulo - Linear.

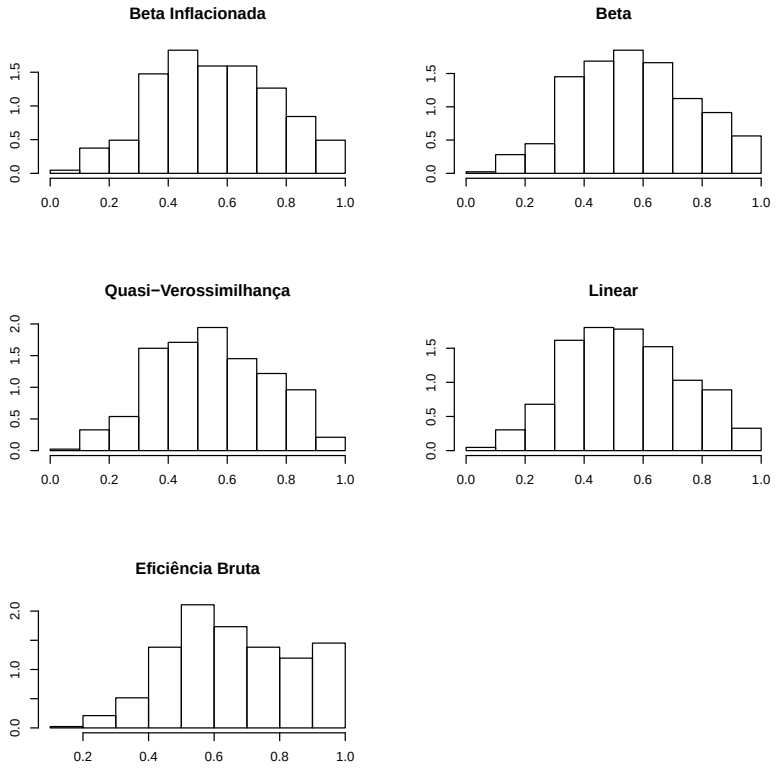

Figura 2.7 Histogramas das medidas de eficiência bruta e das medidas de eficiência pura geradas pelos modelos de regressão beta inflacionado, beta, linear e quasi-verossimilhança para os dados de São Paulo.

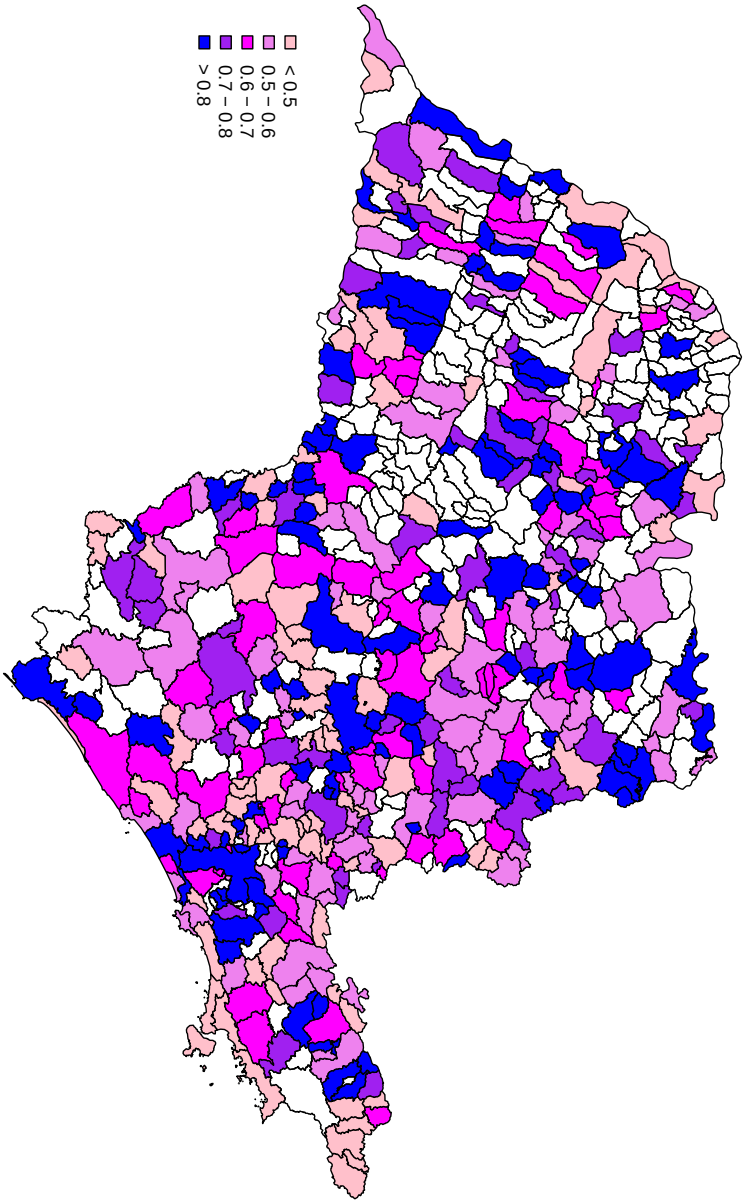

Figura 2.8 Mapa das eficiências brutas para os dados de São Paulo.

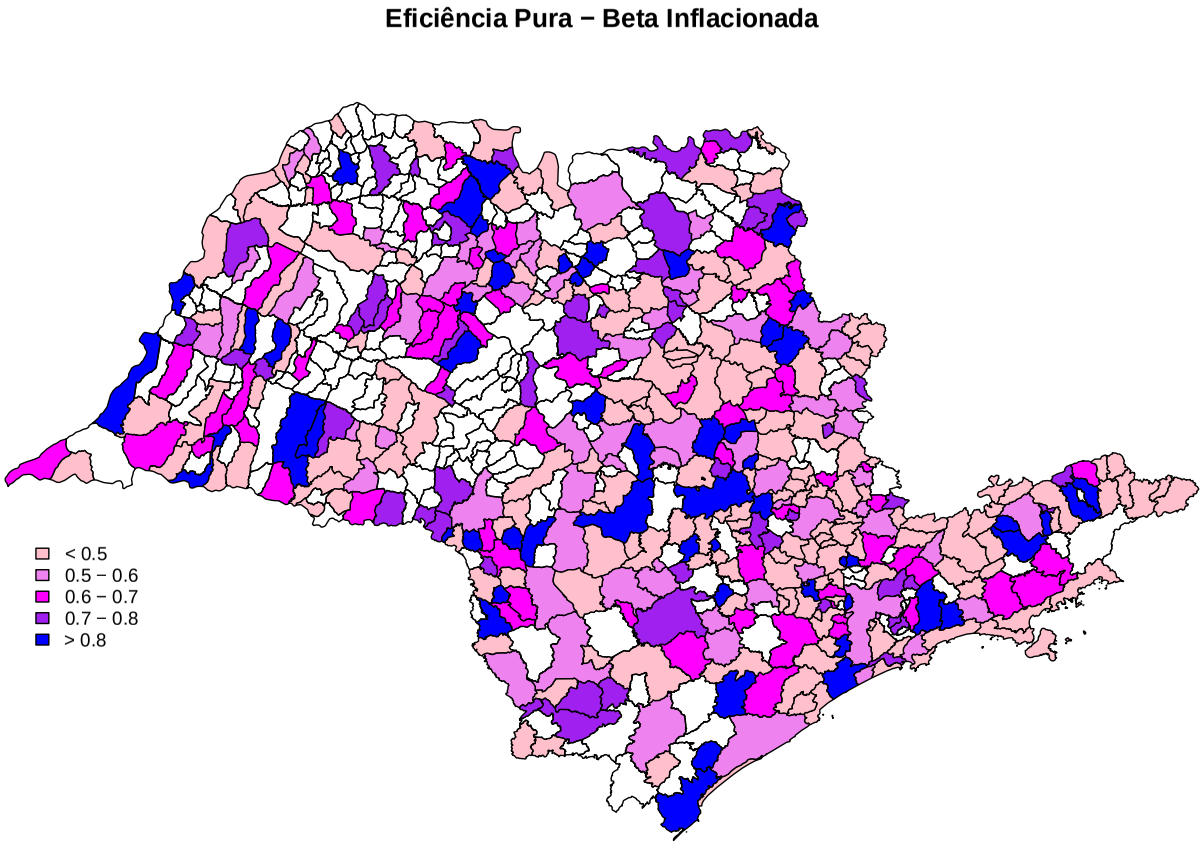

Figura 2.9 Mapa das eficiências puras para os dados de São Paulo - Beta Inflacionada.

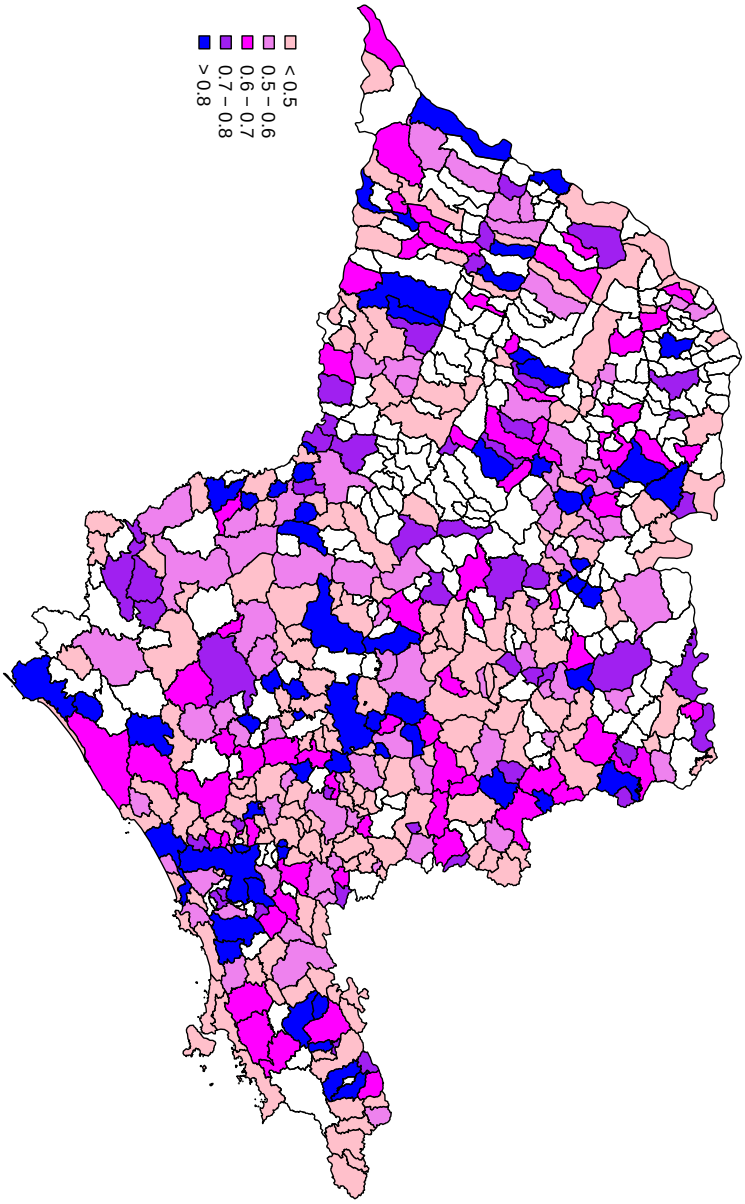

Figura 2.10 Mapa das eficiências puras para os dados de São Paulo - Beta.

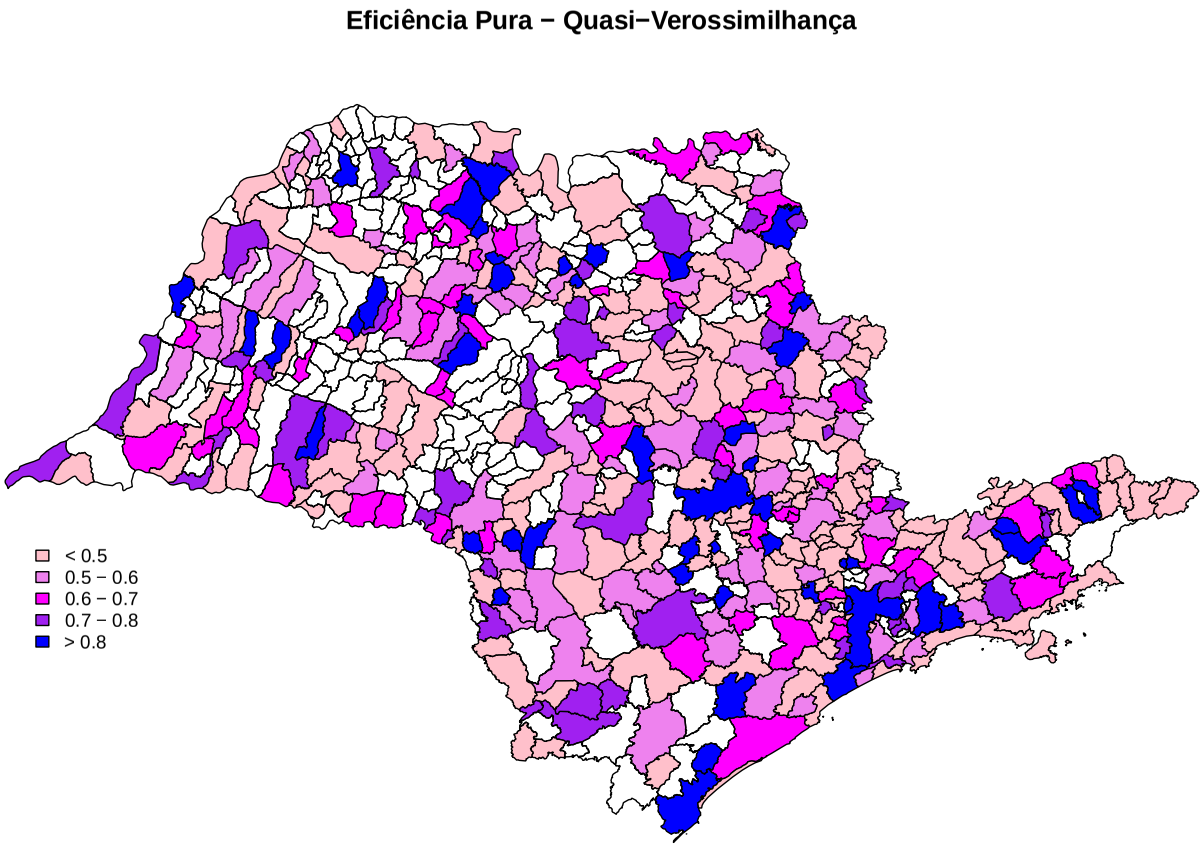

Figura 2.11 Mapa das eficiências puras para os dados de São Paulo - Quasi-Verossimilhança.

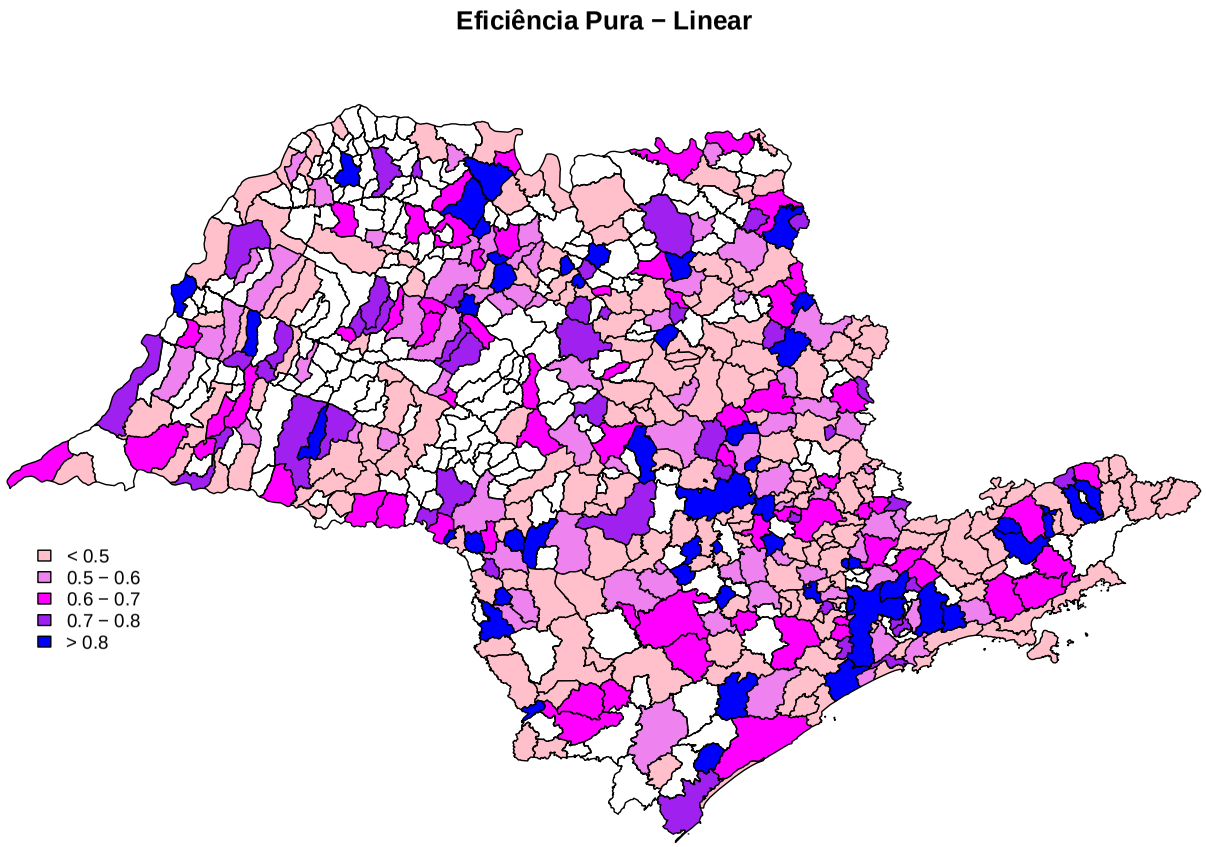

Figura 2.12 Mapa das eficiências puras para os dados de São Paulo - Linear.

Figura 2.13 Gráficos de resíduos do modelo beta inflacionado para os dados do Brasil e demais estados.

CAPÍTULO 3

Testes de erro de especificação para Regressão Beta Inflacionada

3.1 Introdução

Frequentemente estamos interessados em modelar a dependência de taxas, proporções e outras variáveis, que assumem valores no intervalo unitário padrão, em um conjunto de variáveis explicativas. Para tais situações, o modelo de regressão beta proposto por Ferrari e Cribari-Neto (2004) é particularmente útil. Esta classe de modelos de regressão considera que a resposta segue distribuição beta e sua média está relacionada, através de uma função de ligação, a um preditor linear definido por regressores e parâmetros de regressão desconhecidos. No entanto, alguns conjuntos de dados podem incluir observações iguais aos limites do intervalo, ou seja, observações iguais a zero e/ou um. Ospina e Ferrari (2010) introduziram uma família de distribuições, conhecidas como distribuições beta inflacionadas, que são misturas de uma distribuição beta e uma distribuição de Bernoulli degenerada em zero e/ou um para permitir que os usuários modelem dados que assumem valores em $[0, 1)$, $(0, 1]$ ou $[0, 1]$. Ospina (2008) introduziu uma classe de modelos de regressão beta inflacionados (dispersão fixa) em que a resposta segue uma distribuição beta inflacionada. O modelo inclui um submodelo de regressão para a probabilidade de que a variável dependente é igual a um dos limites do intervalo.

Ao postularmos uma estrutura paramétrica de regressão não é possível saber de fato se o modelo considerado retrata adequadamente a realidade do fenômeno em estudo. Caso a especificação utilizada seja errônea, inferências imprecisas podem ocorrer no que se refere à estimação dos parâmetros, intervalos de confiança e testes de hipóteses. Testes de especificação em modelos de regressão formam uma importante área de pesquisa. Ramsey (1969) introduziu o teste de erro de especificação (RESET) em análise de regressão linear para detectar forma funcional inapropriada e variáveis omitidas. O teste foi desenvolvido comparando-se a distribuição dos resíduos sob a hipótese de que a especificação do modelo é correta com a distribuição dos resíduos produzidos sob a hipótese alternativa de que existe um erro de especificação. Sob a hipótese nula de ausência de erro de especificação existirá um estimador eficiente, consistente e assintoticamente normal. Contudo, sob a hipótese alternativa de má especificação, esse estimador será viesado e inconsistente (Hausman, 1978). Ramsey e Schmidt (1976) mostraram que o teste RESET baseado no uso dos resíduos de mínimos quadrados é equivalente ao teste originalmente proposto por Ramsey (1969). Dessa forma, o procedimento do teste se reduz a incluir uma forma não-linear ao modelo, através de potências de variáveis adicionais, chamadas de variáveis de teste, e por meio de valores críticos da distribuição F, testar a exclusão de tais variáveis. A intuição por trás do teste é que se essas variáveis de teste têm algum poder em

explicar a variável dependente, então o modelo está mal especificado.

Alguns autores estudaram as propriedades do teste RESET. Ramsey e Gilbert (1972) sugeriram usar como variáveis de teste a segunda, terceira e quarta potências do valor ajustado. Thursby e Schmidt (1977) recomendaram usar a segunda, terceira e quarta potências das variáveis independentes. Shukur e Edgerton (2002) generalizaram o teste RESET para cobrir sistemas de equações. Eles aplicaram os testes Wald, multiplicador de Lagrange e razão de verossimilhanças para testar a exclusão das variáveis de teste. Mantalos e Shukur (2007) investigaram a robustez do teste RESET para erros não-normais. Os tamanhos e poderes de várias generalizações do teste RESET usando valores críticos assintóticos e obtidos via bootstrap também foram investigados. Cribari-Neto e Lima (2007) propuseram um teste de erro de especificação para modelos de regressão beta. O teste é útil para identificar não-linearidade negligenciada e função de ligação incorretamente especificada.

Neste capítulo, uma forma geral de teste de erro especificação para modelos de regressão beta inflacionados é apresentada. Os tamanhos e os poderes dos testes são investigados via simulação de Monte Carlo. O teste é baseado no teste RESET e a inferência é baseada em valores críticos assintóticos. Em particular, nós propomos duas variantes do teste. Na primeira variante, nós apenas adicionamos variáveis de teste para o submodelo da média. A segunda variante segue da adição de variáveis de teste para todos os submodelos. Na avaliação numérica nós consideramos erros de especificação devido a não-linearidade no preditor linear, variáveis omitidas, escolha incorreta da função de ligação, correlação espacial negligenciada e dispersão variável não modelada. Diferentes escolhas das variáveis de teste são consideradas e avaliadas. Os resultados indicam que o teste proposto pode ser um tanto útil para detectar erro de especificação em modelos de regressão beta inflacionados.

O capítulo contém quatro seções. A primeira descreve uma variante do modelo de regressão beta inflacionado que incorpora dispersão variável. Nós apresentamos a função de log-verossimilhança do modelo, as funções escore e a matriz de informação de Fisher, uma vez que essas quantidades são necessárias para os testes de erro de especificação propostos, os quais são introduzidos na Seção 3.3. Na Seção 3.4, com base em simulações de Monte Carlo, o desempenho dos testes propostos, no que tange à sua capacidade de identificar alguns erros de especificação, é avaliado. Uma aplicação a dados reais é apresentada na Seção 3.5. A Seção 3.6 contém algumas conclusões deste trabalho.

3.2 O modelo de regressão beta inflacionado em zero ou um

Ospina e Ferrari (2010) propuseram distribuições de mistura entre uma distribuição beta e uma distribuição de Bernoulli degenerada em zero e/ou um para acomodar dados observados nos intervalos $(0,1]$, $[0,1)$ e $[0,1]$. Os modelos propostos são ditos serem inflacionados no sentido de que a massa de probabilidade em alguns pontos (zero e/ou um) excede o que é permitido pela distribuição beta.

Sob a parametrização de Ferrari e Cribari-Neto (2004), a distribuição beta tem função de densidade

$$f(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, \quad 0 < y < 1, \tag{3.2.1}$$

em que $0 < \mu < 1$ e $\phi > 0$. Dizemos que y tem distribuição beta com média μ e precisão ϕ e escrevemos $y \sim \mathcal{B}(\mu,\phi)$. A média e a variância de y são dadas, respectivamente, por $\mathbb{E}(y) = \mu$ e $\text{Var}(y) = V(\mu)/(1+\phi)$, em que $V(\mu) = \mu(1-\mu)$.

A função de distribuição acumulada do modelo beta inflacionado em zero ou um é dada por

$$\text{BI}_c(y;\alpha,\mu,\phi) = \alpha \mathbb{1}_{\{c\}}(y) + (1-\alpha)F(y;\mu,\phi), \tag{3.2.2}$$

em que $\mathbb{1}_{\{c\}}(y)$ é uma função indicadora que assume valor 1 se $y = c$ e 0 caso contrário, $F(\cdot;\mu,\phi)$ é a função de distribuição acumulada beta $\mathcal{B}(\mu,\phi)$ e $0 < \alpha < 1$ é o parâmetro de mistura da distribuição especificado por $\alpha = \Pr(y = c)$. A função BI_c tem um ponto de massa em $y = c$, não sendo assim absolutamente contínua. Desta forma, com probabilidade α, a variável y é selecionada de uma distribuição degenerada no ponto c e, com probabilidade $(1-\alpha)$, a variável é selecionada de uma distribuição beta.

A função densidade de probabilidade do modelo beta inflacionado é dada por

$$\text{bi}_c(y;\alpha,\mu,\phi) = \begin{cases} \alpha, & y = c, \\ (1-\alpha)f(y;\mu,\phi), & y \in (0,1), \end{cases} \tag{3.2.3}$$

em que $0 < \alpha < 1$, $0 < \mu < 1$, $\phi > 0$ e $f(y;\mu,\phi)$ é a função de densidade $\mathcal{B}(\mu,\phi)$. A densidade (3.2.3) é uma distribuição beta inflacionada no ponto c, $c = 0$ ou $c = 1$. Para esta distribuição, $\mathbb{E}(y) = \alpha c + (1-\alpha)\mu$ e $\text{Var}(y) = (1-\alpha)\text{V}(\mu)/(\phi+1) + \alpha(1-\alpha)(c-\mu)^2$. Se $c = 0$, a distribuição (3.2.3) é denominada distribuição beta inflacionada no ponto zero e escrevemos $y \sim \text{BEZI}(\alpha,\mu,\phi)$. Se $c = 1$, a distribuição (3.2.3) é denominada distribuição beta inflacionada no ponto um e escrevemos $y \sim \text{BEOI}(\alpha,\mu,\phi)$.

Ospina (2008) propôs modelos de regressão beta inflacionado com dispersão constante, os quais são extensões naturais do modelo de regressão beta introduzido por Ferrari e Cribari-Neto (2004). O autor assume que a distribuição da resposta é beta inflacionada. A parte contínua dos dados é modelada pela distribuição beta e a parte discreta, isto é, o ponto de massa, é modelada através de uma distribuição degenerada no valor conhecido c, em que c igual a zero ou um. Apresentaremos a seguir uma extensão do modelo de regressão beta inflacionado em zero ou um (Ospina, 2008) que incorpora dispersão variável.

Sejam $y_1,\ldots,y_n$ variáveis aleatórias independentes com distribuição beta inflacionada no ponto c ($c = 0$ ou $c = 1$). O modelo de regressão beta inflacionado em c é definido supondo que a média de y_t, a massa de probabilidade em c e o parâmetro de precisão satisfazem às seguintes relações funcionais:

$$h(\alpha_t) = \sum_{i=1}^{M} z_{ti}\gamma_i = \zeta_t, \tag{3.2.4}$$

$$g(\mu_t) = \sum_{i=1}^{m} x_{ti}\beta_i = \eta_t, \tag{3.2.5}$$

$$b(\phi_t) = \sum_{i=1}^{q} s_{ti}\lambda_i = \kappa_t, \tag{3.2.6}$$

$t = 1,\ldots,n$, em que $\gamma = (\gamma_1,\ldots,\gamma_M)^\top$, $\beta = (\beta_1,\ldots,\beta_m)^\top$ e $\lambda = (\lambda_1,\ldots,\lambda_q)^\top$ são vetores de parâmetros de regressão desconhecidos, com $\gamma \in \mathbb{R}^M$, $\beta \in \mathbb{R}^m$ e $\lambda \in \mathbb{R}^q$, $x_{t1},\ldots,x_{tm}$, $z_{t1},\ldots,z_{tM}$ e $s_{t1},\ldots,s_{tq}$ são observações de covariáveis conhecidas ($m + M + q < n$) que podem coincidir total ou parcialmente. As funções de ligação $h : (0,1) \to \mathbb{R}$, $g : (0,1) \to \mathbb{R}$ e $b : (0,\infty) \to \mathbb{R}$ são estritamente monótonas e duas vezes diferenciáveis. Diferentes funções de ligação podem ser usadas, tais como logit, probit, log-log complementar e log-log para μ e α e logarítmica ou raiz quadrada para ϕ. Aqui, μ_t é a média de y_t condicional em $y_t \in (0,1)$.

A função de verossimilhança para o modelo de regressão beta inflacionado em c, considerando o vetor de parâmetros $\theta = (\gamma^\top,\beta^\top,\lambda^\top)^\top$, é da forma

$$L(\theta) = \prod_{t=1}^{n} \mathrm{bi}_c(y_t;\alpha_t,\mu_t,\phi_t) = L_1(\gamma)L_2(\beta,\lambda),$$

em que

$$L_1(\gamma) = \prod_{t=1}^{n} \alpha_t^{\mathbb{1}_{\{c\}}(y_t)}(1-\alpha_t)^{1-\mathbb{1}_{\{c\}}(y_t)},$$
$$L_2(\beta,\lambda) = \prod_{t:y_t\in(0,1)} f(y_t;\mu_t,\phi_t).$$

Os parâmetros α_t, μ_t e ϕ_t são definidos como funções de γ, β e λ, respectivamente, através de (3.2.4), (3.2.5) e (3.2.6), ou seja, $\alpha_t = h^{-1}(\zeta_t)$, $\mu_t = g^{-1}(\eta_t)$ e $\phi_t = b^{-1}(\kappa_t)$.

A função de log-verossimilhança para $\theta = (\gamma^\top,\beta^\top,\lambda^\top)^\top$ é da forma

$$\ell(\theta) = \ell_1(\gamma) + \ell_2(\beta,\lambda),$$

em que

$$\ell_1(\gamma) = \sum_{t=1}^{n} \ell_t(\alpha_t) \quad \text{e} \quad \ell_2(\beta,\lambda) = \sum_{t:y_t\in(0,1)} \ell_t(\mu_t,\phi_t),$$

com

$$\ell_t(\alpha_t) = \mathbb{1}_{\{c\}}(y_t)\log\alpha_t + (1-\mathbb{1}_{\{c\}}(y_t))\log(1-\alpha_t),$$
$$\ell_t(\mu_t,\phi_t) = \log\Gamma(\phi_t) - \log\Gamma(\mu_t\phi_t) - \log\Gamma((1-\mu_t)\phi_t) + (\mu_t\phi_t - 1)\log y_t + \{(1-\mu_t)\phi_t - 1\}\log(1-y_t).$$

Para $t = 1,\ldots,n$, a variável aleatória $\mathbb{1}_{\{c\}}(y_t)$ é do tipo Bernoulli, em que $\Pr(\mathbb{1}_{\{c\}}(y_t) = 1) = \alpha_t$ é associada às covariáveis através de um preditor linear ζ_t e de uma função de ligação h. Dessa forma, $\ell_1(\gamma)$ é a função de log-verossimilhança de um modelo linear generalizado com resposta binária e $\ell_2(\beta,\lambda)$ é a função de log-verossimilhança de um modelo de regressão beta com dispersão variável em que a variável dependente é restrita ao intervalo aberto $(0,1)$.

Dado que a função de verossimilhança pode ser fatorada em dois termos, os parâmetros são separáveis e inferência por máxima verossimilhança sobre $(\beta^\top,\lambda^\top)^\top$ pode ser realizada de forma independente do vetor de parâmetros γ e vice-versa. Pela separabilidade dos vetores de parâmetros γ e $(\beta^\top,\lambda^\top)^\top$, é possível obter de forma independente as funções escore para γ e $(\beta^\top,\lambda^\top)^\top$. Para $R = 1,\ldots,M$, a função escore para γ é

$$U_R = \frac{\partial \ell_1(\gamma)}{\partial \gamma_R} = \sum_{t=1}^{n} \frac{\partial \ell_t(\alpha_t)}{\partial \alpha_t}\frac{d\alpha_t}{d\zeta_t}\frac{\partial \zeta_t}{\partial \gamma_R},$$

com

$$\frac{d\alpha_t}{d\zeta_t} = \frac{dh^{-1}(\zeta_t)}{d\zeta_t} = \frac{1}{h'(\alpha_t)}$$

e

$$\frac{\partial \ell_t(\alpha_t)}{\partial \alpha_t} = \frac{\mathbb{1}_{\{c\}}(y_t)}{\alpha_t} - \frac{1-\mathbb{1}_{\{c\}}(y_t)}{(1-\alpha_t)}.$$

Portanto,

$$U_R = \sum_{t=1}^{n} \frac{\mathbb{1}_{\{c\}}(y_t)-\alpha_t}{\alpha_t(1-\alpha_t)}\frac{1}{h'(\alpha_t)} z_{tR}.$$

Para $r = 1\ldots,m$, a função escore para β é dada por

$$U_r = \frac{\partial \ell_2(\beta,\lambda)}{\partial \beta_r} = \sum_{t:y_t\in(0,1)} \frac{\partial \ell_t(\mu_t,\phi_t)}{\partial \mu_t}\frac{d\mu_t}{d\eta_t}\frac{\partial \eta_t}{\partial \beta_r},$$

com

$$\frac{d\mu_t}{d\eta_t} = \frac{dg^{-1}(\eta_t)}{d\eta_t} = \frac{1}{g'(\mu_t)}$$

e

$$\frac{\partial \ell_t(\mu_t,\phi_t)}{\partial \mu_t} = \phi_t\left\{\log\left(\frac{y_t}{1-y_t}\right) - [\psi(\mu_t\phi_t) - \psi((1-\mu_t)\phi_t)]\right\}.$$

Seja

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

e

$$\mu_t^* = \mathbb{E}(y_t^*|\mathbb{1}_{\{c\}}(y_t) = 0) = \psi(\mu_t\phi) - \psi((1-\mu_t)\phi). \tag{3.2.7}$$

Então

$$U_r = \sum_{t=1}^{n}(1-\mathbb{1}_{\{c\}}(y_t))\phi_t(y_t^* - \mu_t^*)\frac{1}{g'(\mu_t)}x_{tr}.$$

Considere agora a função escore para λ. Segue que, para $p = 1\ldots, q$,

$$U_p = \frac{\partial \ell_2(\beta,\lambda)}{\partial \lambda_p} = \sum_{t:y_t\in(0,1)} \frac{\partial \ell_t(\mu_t,\phi_t)}{\partial \phi_t}\frac{d\phi_t}{d\kappa_t}\frac{\partial \kappa_t}{\partial \lambda_p},$$

com

$$\frac{d\phi_t}{d\kappa_t} = \frac{db^{-1}(\kappa_t)}{d\kappa_t} = \frac{1}{b'(\phi_t)}$$

e

$$\begin{aligned}\frac{\partial \ell_t(\mu_t,\phi_t)}{\partial \phi_t} = \mu_t\left\{\log\left(\frac{y_t}{1-y_t}\right) - [\psi(\mu_t\phi_t) - \psi((1-\mu_t)\phi_t)]\right\} + \log(1-y_t)\\ - \psi((1-\mu_t)\phi_t) + \psi(\phi_t).\end{aligned}$$

Nós temos então que

$$U_p = \sum_{t=1}^{n}(1 - \mathbb{1}_{\{c\}}(y_t))a_t\frac{1}{b'(\phi_t)}s_{tp},$$

com $a_t = \mu_t\left\{\log\left(\frac{y_t}{1-y_t}\right) - [\psi(\mu_t\phi_t) - \psi((1-\mu_t)\phi_t)]\right\} + s(y_t) - \psi((1-\mu_t)\phi_t) + \psi(\phi_t)$ e $s(y_t) = \log(1-y_t)$, se $y_t \in (0,1)$ e $s(y_t) = 0$, se $y_t = 1$.

Os vetores escore para γ, β e λ podem ser escritos, em forma matricial, como

$$\begin{aligned}U_\gamma(\gamma) &= Z^\top PG(y^c - \alpha^*),\\ U_\beta(\beta,\lambda) &= X^\top \Phi T H(y^* - \mu^*),\\ U_\lambda(\beta,\lambda) &= S^\top VHa,\end{aligned}$$

em que Z é uma matriz $n\times M$ cuja t-ésima linha é $z_t^\top = (z_{t1},\ldots,z_{tM})$, X é uma matriz $n\times m$ cuja t-ésima linha é $x_t^\top = (x_{t1},\ldots,x_{tm})$, S é uma matriz $n\times q$ cuja t-ésima linha é $s_t^\top = (s_{t1},\ldots,s_{tq})$, $y^* = (y_1^*,\ldots,y_n^*)^\top$, $\mu^* = (\mu_1^*,\ldots,\mu_n^*)^\top$, $y^c = \left(\mathbb{1}_{\{c\}}(y_1),\ldots,\mathbb{1}_{\{c\}}(y_n)\right)^\top$, $\alpha^* = (\alpha_1,\ldots,\alpha_n)^\top$ e $a = (a_1,\ldots,a_n)^\top$. Definimos as matrizes diagonais, $P = \text{diag}\{1/[\alpha_1(1-\alpha_1)],\ldots,1/[\alpha_n(1-\alpha_n)]\}$, $H = \text{diag}\{1 - \mathbb{1}_{\{c\}}(y_1),\ldots,1-\mathbb{1}_{\{c\}}(y_n)\}$, $G = \text{diag}\{d\alpha_1/d\zeta_1,\ldots,d\alpha_n/d\zeta_n\}$, $T = \text{diag}\{d\mu_1/d\eta_1,\ldots,d\mu_n/d\eta_n\}$, $V = \text{diag}\{d\phi_1/d\kappa_1,\ldots,d\phi_n/d\kappa_n\}$ e $\Phi = \text{diag}\{\phi_1,\ldots,\phi_n\}$.

Como consequência da ortogonalidade do vetor de parâmetros γ e do vetor de parâmetros $(\beta^\top,\lambda^\top)^\top$, o estimador de máxima verossimilhança de γ é assintoticamente independente dos estimadores de máxima verossimilhança de β e λ. Tais estimadores podem ser obtidos através de métodos iterativos, tal como o algoritmo de Newton ou um algoritmo quasi-Newton. O estimador de máxima verossimilhança de γ é obtido como solução do sistema não-linear $U_\gamma(\gamma) = 0$. O estimador de máxima verossimilhança de $(\beta^\top,\lambda^\top)^\top$ é obtido como solução do sistema não-linear $(U_\beta(\beta,\lambda)^\top, U_\lambda(\beta,\lambda)^\top) = 0$.

Para obtenção da matriz de informação de Fisher serão apresentados a seguir os valores esperados das derivadas de segunda ordem da função de log-verossimilhança. Para $R = 1,\ldots,M$

e $O = 1, \ldots, M$, temos que

$$\begin{aligned}
U_{RO} &= \frac{\partial^2 \ell_1(\gamma)}{\partial \gamma_R \gamma_O} = \sum_{t=1}^{n} \frac{\partial}{\partial \alpha_t}\left(\frac{\partial \ell_t(\alpha_t)}{\partial \alpha_t}\frac{d\alpha_t}{d\zeta_t}\frac{\partial \zeta_t}{\partial \gamma_R}\right)\frac{d\alpha_t}{d\zeta_t}\frac{\partial \zeta_t}{\partial \gamma_O} \\
&= \sum_{t=1}^{n}\left\{\frac{\partial^2 \ell_t(\alpha_t)}{\partial \alpha_t^2}\left(\frac{d\alpha_t}{d\zeta_t}\right)^2 + \frac{\partial \ell_t(\alpha_t)}{\partial \alpha_t}\left(\frac{\partial}{\partial \alpha_t}\frac{d\alpha_t}{d\zeta_t}\right)\frac{d\alpha_t}{d\zeta_t}\right\} z_{tO} z_{tR} \\
&= \sum_{t=1}^{n}\left\{\left(\frac{-\mathbb{1}_{(0,1)}(y_t)}{(1-\alpha_t)^2} - \frac{\mathbb{1}_{\{c\}}(y_t)}{\alpha_t^2}\right)\left(\frac{d\alpha_t}{d\zeta_t}\right)^2 + \left(\frac{\mathbb{1}_{\{c\}}(y_t)}{\alpha_t} - \frac{\mathbb{1}_{(0,1)}(y_t)}{(1-\alpha_t)}\right)\right. \\
&\left.\times\left(\frac{\partial}{\partial \alpha_t}\frac{d\alpha_t}{d\zeta_t}\right)\frac{d\alpha_t}{d\zeta_t}\right\} z_{tO} z_{tR}.
\end{aligned}$$

Uma vez que, sob as condições usuais de regularidade, $\mathbb{E}(\partial \ell_t(\alpha_t)/\partial \alpha_t) = 0$, segue-se que

$$\begin{aligned}
\mathbb{E}(U_{RO}) &= \mathbb{E}\left(\sum_{t=1}^{n}\left\{\left(\frac{-\mathbb{1}_{(0,1)}(y_t)}{(1-\alpha_t)^2} - \frac{\mathbb{1}_{\{c\}}(y_t)}{\alpha_t^2}\right)\left(\frac{d\alpha_t}{d\zeta_t}\right)^2\right\} z_{tO} z_{tR}\right) \\
&= \sum_{t=1}^{n}\left\{\frac{-1}{(1-\alpha_t)} - \frac{1}{\alpha_t}\right\}\left(\frac{d\alpha_t}{d\zeta_t}\right)^2 z_{tO} z_{tR}.
\end{aligned}$$

Seja $p_t = \frac{1}{\alpha_t(1-\alpha_t)}$. Então,

$$\mathbb{E}(U_{RO}) = -\sum_{t=1}^{n} p_t \left(\frac{1}{h'(\alpha_t)}\right)^2 z_{tO} z_{tR}.$$

Para $r = 1, \ldots, m$ e $o = 1, \ldots, m$,

$$\begin{aligned}
U_{ro} &= \frac{\partial^2 \ell_2(\beta, \lambda)}{\partial \beta_r \beta_o} = \sum_{t:y_t \in (0,1)} \frac{\partial}{\partial \mu_t}\left(\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t}\frac{d\mu_t}{d\eta_t}\frac{\partial \eta_t}{\partial \beta_r}\right)\frac{d\mu_t}{d\eta_t}\frac{\partial \eta_t}{\partial \beta_o} \\
&= \sum_{t:y_t \in (0,1)}\left\{\left(\frac{\partial^2 \ell_t(\mu_t, \phi_t)}{\partial \mu_t^2}\right)\left(\frac{d\mu_t}{d\eta_t}\right)^2 + \left(\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial(\mu_t)}\left(\frac{\partial}{\partial \mu_t}\frac{d\mu_t}{d\eta_t}\right)\frac{d\mu_t}{d\eta_t}\right)\right\} x_{to} x_{tr} \\
&= \sum_{t:y_t \in (0,1)}\left\{\left(-\phi_t \frac{\partial \mu_t^*}{\partial \mu_t}\left(\frac{d\mu_t}{d\eta_t}\right)^2\right) + \left(\phi_t(y_t^* - \mu_t^*)\left(\frac{\partial}{\partial \mu_t}\frac{d\mu_t}{d\eta_t}\right)\frac{d\mu_t}{d\eta_t}\right)\right\} x_{to} x_{tr}.
\end{aligned}$$

De (3.2.7) segue que

$$\frac{\partial \mu_t^*}{\partial \mu_t} = -\phi_t\{\phi_t[\psi'(\mu_t\phi_t) + \psi'((1-\mu_t)\phi_t)]\}.$$

Seja $w_t = [\psi'(\mu_t\phi_t) + \psi'((1-\mu_t)\phi_t)]$. Então,

$$U_{ro} = \sum_{t:y_t \in (0,1)}\left\{-\phi_t^2 w_t\left(\frac{d\mu_t}{d\eta_t}\right)^2 + \phi_t(y_t^* - \mu_t^*)\left(\frac{\partial}{\partial \mu_t}\frac{d\mu_t}{d\eta_t}\right)\frac{d\mu_t}{d\eta_t}\right\} x_{to} x_{tr}.$$

Para o caso dos cumulantes envolvendo os vetores de parâmetros β e λ pode-se usar o fato de que se $\tau : (0,1) \to \mathbb{R}$ é uma função contínua, então (Ospina, 2008)

$$\mathbb{E}\left(\sum_{t:y_t\in(0,1)} \tau(y_t)\right) = \sum_{t=1}^{n}(1-\alpha_t)\mathbb{E}\Big(\tau(y_t)|\mathbb{1}_{\{c\}}(y_t)=0\Big).$$

Considerando o fato de que $\mathbb{E}\Big(y_t^*|\mathbb{1}_{\{c\}}(y_t)=0\Big) = \mu_t^*$. Assim,

$$\begin{aligned}\mathbb{E}(U_{ro}) &= \mathbb{E}\left(\sum_{t:y_t\in(0,1)}\left\{-\phi_t^2 w_t\left(\frac{d\mu_t}{d\eta_t}\right)^2\right\}x_{to}x_{tr}\right)\\ &= -\sum_{t=1}^{n}\phi_t^2(1-\alpha_t)w_t\left(\frac{1}{g'(\mu_t)}\right)^2 x_{to}x_{tr}.\end{aligned}$$

Adicionalmente, para $p = 1,\ldots,q$ e $l = 1,\ldots,q$

$$\begin{aligned}U_{pl} = \frac{\partial^2\ell_2(\beta,\lambda)}{\partial\lambda_p\lambda_l} &= \sum_{t:y_t\in(0,1)}\frac{\partial}{\partial\phi_t}\left(\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\phi_t}\frac{d\phi_t}{d\kappa_t}\frac{\partial\kappa_t}{\partial\lambda_p}\right)\frac{d\phi_t}{d\kappa_t}\frac{\partial\kappa_t}{\partial\lambda_l}\\ &= \sum_{t:y_t\in(0,1)}\left\{\left(\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\phi_t^2}\left(\frac{d\phi_t}{d\kappa_t}\right)^2\right)+\left(\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial(\phi_t)}\left(\frac{\partial}{\partial\phi_t}\frac{d\phi_t}{d\kappa_t}\right)\frac{d\phi_t}{d\kappa_t}\right)\right\}s_{tl}s_{tp},\end{aligned}$$

em que

$$\begin{aligned}\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\phi_t^2} &= -\mu_t\frac{\partial\mu_t^*}{\partial\phi_t} - (1-\mu_t)\psi'((1-\mu_t)\phi_t)+\psi'(\phi_t)\\ &= -\Big\{(1-\mu_t)^2\psi'((1-\mu_t)\phi_t)+\mu_t^2\psi'(\mu_t\phi_t)-\psi'(\phi_t)\Big\}.\end{aligned}$$

Definindo $d_t = (1-\mu_t)^2\psi'((1-\mu_t)\phi_t)+\mu_t^2\psi'(\mu_t\phi_t)-\psi'(\phi_t)$. Temos que

$$U_{pl} = \sum_{t:y_t\in(0,1)}\left\{-d_t\left(\frac{d\phi_t}{d\kappa_t}\right)^2 + a_t\left(\frac{\partial}{\partial\phi_t}\frac{d\phi_t}{d\kappa_t}\right)\frac{d\phi_t}{d\kappa_t}\right\}s_{tl}s_{tp}.$$

Ospina e Ferrari (2010, Proposição 2.1) mostraram que a Densidade (3.2.3) pertence à família exponencial. Segue que $\mathbb{E}\Big(\log(1-y_t)|\mathbb{1}_{\{c\}}(y_t)=0\Big) = \psi((1-\mu_t)\phi_t)-\psi(\phi_t)$. Sob certas condições de regularidade, nós temos que $\mathbb{E}(\partial\ell_t(\mu_t,\phi_t)/\partial\phi_t) = 0$ e, portanto,

$$\begin{aligned}\mathbb{E}(U_{pl}) &= \mathbb{E}\left(\sum_{t:y_t\in(0,1)} -d_t\left(\frac{d\phi_t}{d\kappa_t}\right)^2 s_{tl}s_{tp}\right)\\ &= -\sum_{t=1}^{n}(1-\alpha_t)d_t\left(\frac{1}{b'(\phi_t)}\right)^2 s_{tl}s_{tp}.\end{aligned}$$

Por fim, para $r = 1,\ldots,m$ e $p = 1,\ldots,q$,

$$\begin{aligned}
U_{rp} = \frac{\partial^2 \ell_2(\beta,\lambda)}{\partial\beta_r\lambda_p} &= \sum_{t:y_t\in(0,1)} \frac{\partial}{\partial\phi_t}\left(\frac{\partial\ell_2(\beta,\lambda)}{\partial\beta_r}\right)\frac{d\phi_t}{d\kappa_t}\frac{\partial\kappa_t}{\partial\lambda_p} \\
&= \sum_{t:y_t\in(0,1)} \frac{\partial}{\partial\phi_t}\left(\phi_t(y_t^*-\mu_t^*)\frac{1}{g'(\mu_t)}x_{tr}\right)\frac{d\phi_t}{d\kappa_t}\frac{\partial\kappa_t}{\partial\lambda_p} \\
&= \sum_{t:y_t\in(0,1)} \{(y_t^*-\mu_t^*) - \phi_t\left[\psi'(\mu_t\phi_t)\mu_t - \psi'((1-\mu_t)\phi_t)(1-\mu_t)\right]\} \\
&\times \frac{1}{g'(\mu_t)}x_{tr}\frac{d\phi_t}{d\kappa_t}\frac{\partial\kappa_t}{\partial\lambda_p}.
\end{aligned}$$

Definindo $c_t = \phi_t\left[\mu_t w_t - \psi'((1-\mu_t)\phi_t)\right]$, podemos escrever

$$U_{rp} = \sum_{t:y_t\in(0,1)} -\{c_t - (y_t^*-\mu_t^*)\}\frac{1}{g'(\mu_t)}x_{tr}\frac{d\phi_t}{d\kappa_t}\frac{\partial\kappa_t}{\partial\lambda_p}.$$

Portanto,

$$\begin{aligned}
\mathbb{E}(U_{rp}) &= \mathbb{E}\left(\sum_{t:y_t\in(0,1)} -\{c_t - (y_t^*-\mu_t^*)\}\frac{1}{g'(\mu_t)}x_{tr}\frac{d\phi_t}{d\kappa_t}\frac{\partial\kappa_t}{\partial\lambda_p}\right) \\
&= -\sum_{t=1}^{n}(1-\alpha_t)c_t\frac{1}{g'(\mu_t)}\frac{1}{b'(\phi_t)}x_{tr}s_{tp}.
\end{aligned}$$

Pela separabilidade dos vetores de parâmetros γ e $(\beta^\top,\lambda^\top)$, temos que $U_{Rp} = U_{rR} = 0$.

A matriz de informação de Fisher para o modelo beta inflacionado com dispersão variável tem a forma

$$K(\theta) = \begin{pmatrix} K_\gamma(\gamma) & 0 \\ 0 & K_\vartheta(\vartheta) \end{pmatrix}.$$

Aqui, $K_\gamma(\gamma) = K_{\gamma\gamma}$ é a matriz de Informação de Fisher de γ dada por $K_{\gamma\gamma} = Z^\top QZ$, em que $Q = GPG = \text{diag}\{q_1,\ldots,q_n\}$ é uma matriz diagonal, com $q_t = p_t(d\alpha_t/d\zeta)^2$. Adicionalmente,

$$K_\vartheta(\vartheta) = \begin{pmatrix} K_{\beta\beta} & K_{\beta\lambda} \\ K_{\lambda\beta} & K_{\lambda\lambda} \end{pmatrix}$$

é a matriz de informação de Fisher para $\vartheta = (\beta^\top,\lambda^\top)$. Temos aqui, $K_{\beta\beta} = X^\top\Delta T\Phi W\Phi TX$, $K_{\beta\lambda} = K_{\lambda\beta}^\top = X^\top\Delta CTVS$ e $K_{\lambda\lambda} = S^\top V\Delta DVS$. Também, $\Delta = \text{diag}\{\delta_1,\ldots,\delta_n\}$, $W = \text{diag}\{w_1,\ldots,w_n\}$, $D = \text{diag}\{d_1,\ldots,d_n\}$ e $C = \text{diag}\{c_1,\ldots,c_n\}$. Para $t = 1,\ldots,n$, temos que $\delta_t = 1-\alpha_t$.

Note que se $\phi_t = \phi$, para $t = 1,\ldots,n$, então, as funções escores e a matriz de Fisher podem ser obtidas, respectivamente, de 2.2.6 e 2.2.7.

3.3 Testes de erro de especificação para a regressão beta inflacionada

Depois de especificada e estimada a regressão, a próxima etapa consiste na realização de testes que permitam avaliar a especificação do modelo. Com base na ideia original do teste RESET, nós propomos dois teste de erro de especificação para a classe de regressão beta inflacionada. No primeiro teste nós focamos no submodelo para a média e na segunda aproximação nós criamos uma estratégia na qual todos os três submodelos são testados. Considere o modelo de regressão beta inflacionado

$$h(\alpha) = Z\gamma,$$
$$g(\mu) = X\beta,$$

em que γ e β são vetores $M \times 1$ e $m \times 1$, respectivamente, Z e X são matrizes de regressores $n \times M$ e $n \times m$, respectivamente, e μ e α são vetores $n \times 1$. Adicionalmente, no caso do modelo beta inflacionado com dispersão variável, considere a estrutura de regressão para o parâmetro de precisão

$$b(\phi) = S\,\lambda,$$

em que λ é um vetor $q \times 1$, S é uma matriz $n \times q$ e ϕ é um vetor $n \times 1$.

O teste proposto é realizado em dois estágios. Primeiro, nós estimamos os vetores de parâmetros e selecionamos um conjunto de 'variáveis de teste', as quais são usadas para definir a seguinte 'regressão aumentada' para a média:

$$g(\mu) = X\beta + A_1\tau_1, \tag{3.3.1}$$

em que A_1 é uma matriz $n \times u_1$ de variáveis de teste e τ_1 é um vetor de parâmetros $u_1 \times 1$. Segundo, estimamos (3.3.1) e testamos a hipótese nula $\mathcal{H}_0 : \tau_1 = 0$ (o modelo está corretamente especificado). A ideia principal do teste é que se o modelo estiver bem especificado tais variáveis de teste não devem fornecer contribuição adicional para a qualidade do ajuste.

A implementação prática do teste depende de dois fatores. Primeiramente, deve-se selecionar as variáveis de teste que serão utilizadas na regressão aumentada. Segundo, é necessário decidir qual procedimento deverá ser utilizado para testar a exclusão dos termos adicionais. As variáveis de teste utilizadas são potências do preditor linear estimado e potências da resposta média estimada. Para testar a exclusão das variáveis adicionais na classe de regressão beta inflacionada é necessário recorrer a testes assintóticos. Os testes da razão de verossimilhanças (RV), escore e Wald serão utilizados para este fim.

Seja $\nu = (\tau_1^\top, \gamma^\top, \beta^\top, \lambda^\top)^\top$. Para testar a hipótese nula $\mathcal{H}_0 : \tau_1 = 0$ contra a hipótese alternativa de $\mathcal{H}_1 : \tau_1 \neq 0$, considere as estatísticas de teste apresentadas a seguir.

A estatística de teste da razão de verossimilhanças é

$$\xi_1 = 2\{\ell(\hat{\nu}) - \ell(\tilde{\nu})\}, \tag{3.3.2}$$

sendo $\hat{\nu} = (\hat{\tau}_1^\top, \hat{\gamma}^\top \hat{\beta}^\top, \hat{\lambda}^\top)^\top$ o estimador de máxima verossimilhança de ν e $\tilde{\nu} = (0^\top, \tilde{\gamma}^\top \tilde{\beta}^\top, \tilde{\lambda}^\top)^\top$ o estimador de máxima verossimilhança de ν obtido pela imposição da hipótese nula, ou seja,

o estimador de máxima verossimilhança restrito. A estatística de teste escore pode ser escrita como

$$\xi_2 = \tilde{U}_{1\beta}^{\top} \tilde{K}_{11}^{\beta\beta} \tilde{U}_{1\beta}, \tag{3.3.3}$$

em que $U_{1\beta}$ é o vetor de dimensão u_1 que contém os primeiros u_1 elementos do vetor escore $U_\beta(\beta, \lambda)$ e $K_{11}^{\beta\beta}$ é a matriz de dimensão $u_1 \times u_1$ formada pelas primeiras u_1 linhas e pelas primeiras u_1 colunas da inversa da matriz $K_{\beta\beta}$. O 'til' indica que as quantidades são avaliadas no estimador de máxima verossimilhança restrito. No caso do teste de Wald, a estatística de teste é dada por

$$\xi_3 = \hat{\tau}_1^{\top} \left(\hat{K}_{11}^{\beta\beta}\right)^{-1} \hat{\tau}_1. \tag{3.3.4}$$

Aqui, o 'chapéu' indica que as quantidades são avaliadas no estimador de máxima verossimilhança irrestrito.

Sob as condições usuais de regularidade e sob a hipótese nula, ξ_1, ξ_2 e ξ_3 convergem em distribuição para $\chi^2_{u_1}$. Desta forma, os testes podem ser realizados usando valores críticos aproximados obtidos desta distribuição.

Devemos agora introduzir o segundo teste de erro de especificação. Nós agora aumentamos os três submodelos, isto é, os submodelos de regressão para μ, α and ϕ. O modelo aumentado é portanto

$$\begin{aligned} g(\mu) &= X\beta + A_1\tau_1, \\ h(\alpha) &= Z\gamma + A_2\tau_2, \\ d(\phi) &= S\lambda + A_3\tau_3, \end{aligned}$$

em que A_1, A_2 e A_3 são, respectivamente, matrizes $n \times u_1$, $n \times u_2$ e $n \times u_3$ de variáveis de teste e τ_1, τ_2 e τ_3 são, respectivamente, vetores de parâmetros $u_1 \times 1$, $u_2 \times 1$ e $u_3 \times 1$. A hipótese nula a ser testada é $\mathcal{H}_0 : \tau_1 = \tau_2 = \tau_3 = 0$ (o modelo está corretamente especificado) contra a hipótese alternativa de que pelo menos um τ_i, $i = 1, 2, 3$, é diferente de zero (o modelo não está corretamente especificado).

A estatística de teste da razão de verossimilhanças é dada pela equação (3.3.2) e as estatísticas para os testes escore e Wald, dadas a seguir, são escritas como somas de três formas quadráticas:

$$\xi_2 = \tilde{U}_{1\gamma}^{\top} \tilde{K}_{11}^{\gamma\gamma} \tilde{U}_{1\gamma} + \tilde{U}_{1\beta}^{\top} \tilde{K}_{11}^{\beta\beta} \tilde{U}_{1\beta} + \tilde{U}_{1\lambda}^{\top} \tilde{K}_{11}^{\lambda\lambda} \tilde{U}_{1\lambda},$$

$$\xi_3 = \hat{\tau}_1^{\top} \left(\hat{K}_{11}^{\gamma\gamma}\right)^{-1} \hat{\tau}_1 + \hat{\tau}_2^{\top} \left(\hat{K}_{11}^{\beta\beta}\right)^{-1} \hat{\tau}_2 + \hat{\tau}_3^{\top} \left(\hat{K}_{11}^{\lambda\lambda}\right)^{-1} \hat{\tau}_3,$$

em que $U_{1\gamma}$ é o vetor de dimensão u_2 que contém os primeiros u_2 elementos do vetor escore $U_\gamma(\gamma)$ e $U_{1\lambda}$ é o vetor de dimensão u_3 que contém os primeiros u_3 elementos do vetor escore $U_\lambda(\beta, \lambda)$. De forma análoga, define-se $K_{11}^{\gamma\gamma}$ como a matriz de dimensão $u_2 \times u_2$ formada pelas primeiras u_2 linhas e pelas primeiras u_2 colunas da inversa da matriz $K_{\gamma\gamma}$ e $K_{11}^{\lambda\lambda}$ como a matriz de dimensão $u_3 \times u_3$ formada pelas primeiras u_3 linhas e pelas primeiras u_3 colunas da inversa da matriz $K_{\lambda\lambda}$. Sob as condições usuais de regularidade e sob a hipótese nula, ξ_1, ξ_2 e ξ_3 convergem em distribuição para $\chi^2_{u_1+u_2+u_3}$. Portanto, os testes podem ser realizados usando valores críticos aproximados desta distribuição.

Os dois testes apresentados acima são realizados aqui em duas versões. Na primeira considera-se que o erro de especificação está presente apenas na estrutura de regressão da média. Na segunda, supõe-se que as três estruturas de regressão apresentam erro de especificação.

3.4 Avaliação Numérica

O teste RESET é bastante utilizado em modelos de regressão linear clássica para verificar se a suposição de especificação correta é válida. Nesta seção serão apresentados os resultados de um estudo de Monte Carlo para examinar o comportamento dos testes de erro de especificação propostos para modelos de regressão beta inflacionados. Os critérios de avaliação são o tamanho empírico (probabilidade de rejeitar a hipótese nula quando ela é verdadeira) e o poder empírico (probabilidade de rejeitar a hipótese nula quando ela é falsa).

O experimento de Monte Carlo é realizado considerando o seguinte modelo de regressão beta inflacionado em um com dispersão variável:

$$\begin{aligned} h(\alpha_t) &= \gamma_0 + \gamma_1 z_{t1}, \\ g(\mu_t) &= \beta_0 + \beta_1 x_{t1}, \\ b(\phi_t) &= \lambda_0 + \lambda_1 s_{t1}, \end{aligned} \tag{3.4.1}$$

em que $t = 1,\ldots,n$, $h(\cdot)$, $g(\cdot)$ e $b(\cdot)$ são funções de ligação e z_{t1}, x_{t1} e s_{t1} são as covariáveis utilizadas. Para cada covariável são geradas aleatoriamente 40 observações da distribuição uniforme padrão e então repetidas 2, 3 e 5 vezes para obter amostras de tamanho 80, 120 e 200, mantendo assim a mesma estrutura de variação dos dados. Todas simulações foram realizadas usando o R (ver `http://www.R-project.org`) e foram baseadas em 10,000 réplicas de Monte Carlo. Para cada réplica é gerada uma amostra aleatória da variável dependente $y = (y_1,\ldots,y_n)^\top$ com $y_t \sim \text{BEOI}(\alpha_t,\mu_t,\phi_t)$, em que $\alpha_t = h^{-1}(\zeta_t)$, $\mu_t = g^{-1}(\eta_t)$ e $\phi_t = b^{-1}(\kappa_t)$. Os erros de especificação considerados na simulação de poder são devidos a não-linearidade negligenciada, variáveis omitidas, escolha incorreta da função de ligação, correlação espacial negligenciada e dispersão variável não modelada. Os testes são realizados considerando como variáveis de teste os quadrados dos preditores lineares estimados bem como os quadrados dos valores preditos; nós denotamos o vetor de valores preditos por $\mu^\dagger$. Nós consideramos outras variáveis de teste (incluindo potências dos regressores) mas por brevidade tais resultados não são apresentados.

Todos os testes são realizados considerando três níveis nominais, a saber: 1%, 5% e 10%. Com o intuito de que as comparações de poder se dêem entre testes que possuem o mesmo tamanho, os valores críticos usados nas simulações de poder são obtidos das simulações de tamanho. Ou seja, utilizamos valores críticos exatos (estimados) ao invés de valores críticos assintóticos.

O teste em que acrescentamos as variáveis de teste apenas na estrutura de regressão da média será chamado de R_μ e aquele que usa variáveis nas três estruturas de regressão será chamado R_θ, com $\theta = (\mu,\alpha,\phi)$ representando o vetor com os três parâmetros.

3.4.1 Simulações de Tamanho

Nas simulações de tamanho, a resposta é gerada de acordo com (3.4.1). Por brevidade, serão apresentados apenas os resultados das simulações baseadas na função de ligação logit para μ e α e função de ligação logarítmica para ϕ. Os verdadeiros valores dos parâmetros no Modelo (3.4.1) são $\gamma_0 = -2.0$, $\gamma_1 = 1.5$, $\beta_0 = -1.0$, $\beta_1 = 3.5$, $\lambda_0 = 5.1$ e $\lambda_1 = -2.8$. A Tabela 3.1 contém os percentuais (%) de rejeição da hipótese nula de que o modelo está corretamente especificado. Nós apresentamos resultados para as implementações da razão de verossimilhanças, escore e Wald do teste R_μ e também para a implementação da razão de verossimilhanças do teste R_θ, que foi o melhor procedimento encontrado para executar o teste.

No que tange ao teste R_μ, evidências numéricas mostram que a implementação Wald do teste RESET apresenta o pior desempenho. Por exemplo, quando $n = 120$ e o teste para μ é baseado em $\hat{\eta}^2$, os tamanhos empíricos dos testes da razão de verossimilhanças, escore e Wald no nível nominal de 5% são, respectivamente, 5.58%, 5.05% e 6.37%. No geral, a implementação escore do teste RESET apresenta melhores desempenhos. Para $n = 80$ e quando o teste é baseado em $\hat{\mu}^{\dagger 2}$, o tamanho estimado dos testes da razão de verossimilhanças, escore e Wald no nível nominal de 1% são, respectivamente, 1.40%, 1.01% e 2.11%. Entretanto, à medida que o tamanho amostral aumenta as diferenças entre os tamanhos efetivos dos testes escore e da razão de verossimilhanças decrescem. Nós também notamos que a melhor estratégia é usar $\hat{\eta}^2$ como variável de teste.

Para o teste R_θ, de acordo com a notação que nós usamos, $(\hat{\eta}^2, \hat{\zeta}^2, \hat{\kappa}^2)$ indica que a variável $\hat{\eta}^2$ é adicionada ao submodelo para μ, a variável $\hat{\zeta}^2$ é adicionada ao submodelo para α e a variável $\hat{\kappa}^2$ é adicionada ao submodelo para ϕ. Desse modo, $(\hat{\mu}^{\dagger 2}, \hat{\mu}^{\dagger 2}, \hat{\mu}^{\dagger 2})$ indica que $\hat{\mu}^{\dagger 2}$ foi usada como variável de teste nos três submodelos. É possível observar que os tamanhos estimados considerando diferentes variáveis de teste são similares.

No geral, evidências numéricas mostraram que o teste de tamanho menos distorcido é a implementação escore do teste R_μ com $\hat{\eta}^2$ usada como variável de teste.

3.4.2 Simulações de Poder

As simulações de poder para os testes R_μ e R_θ foram realizadas considerando que apenas o submodelo da média é mal especificado e considerando erro de especificação nos três submodelos (exceto sob dispersão variável negligenciada). Devemos agora apresentar resultados numéricos apenas para a implementação escore do teste R_μ que apresentou menores distorções de tamanho. Também, nós devemos apenas dispor resultados numéricos para a melhor e pior escolha das variáveis de teste para o teste R_θ, a saber $(\hat{\eta}^2, \hat{\eta}^2, \hat{\eta}^2)$ e $(\hat{\eta}^2, \hat{\zeta}^2, \hat{\kappa}^2)$ sob correlação espacial negligenciada e $(\hat{\eta}^2, \hat{\eta}^2, \hat{\eta}^2)$ e $(\hat{\mu}^{\dagger 2}, \hat{\mu}^{\dagger 2}, \hat{\mu}^{\dagger 2})$ nos demais casos.

3.4.2.1 Não-linearidade

Sob não-linearidade a resposta é gerada usando a seguinte especificação não-linear:

$$g(\mu_t) = (\beta_0 + \beta_1 x_{t1})^\delta,$$

Tabela 3.1 Taxas de rejeição da hipótese nula (%).

Teste R_μ													
Regressores adicionados	**Teste**	$n = 40$			$n = 80$			$n = 120$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$\hat{\eta}^2$	RV	1.88	7.55	13.44	1.30	6.15	11.45	1.26	5.58	10.81	1.21	4.96	10.22
	escore	0.78	5.36	10.93	0.98	5.20	10.45	1.01	5.05	10.15	0.95	4.57	9.75
	Wald	3.45	9.82	15.86	1.98	7.15	12.69	1.57	6.37	11.40	1.35	5.22	10.64
$\hat{\mu}^{\dagger 2}$	RV	2.01	7.83	14.10	1.40	6.15	11.70	1.44	5.82	10.94	1.02	5.75	10.65
	escore	0.98	5.87	11.46	1.01	5.27	10.34	1.17	5.26	10.22	0.83	5.44	10.34
	Wald	4.15	10.57	17.37	2.11	7.28	12.84	1.84	6.48	11.72	1.25	6.15	11.25
Teste R_θ													
Regressores adicionados	**Teste**	$n = 40$			$n = 80$			$n = 120$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$(\hat{\eta}^2, \hat{\zeta}^2, \hat{\kappa}^2)$	RV	3.37	10.34	17.69	1.70	6.99	12.88	1.37	6.31	11.90	1.15	5.68	10.67
$(\hat{\eta}^2, \hat{\eta}^2, \hat{\eta}^2)$	RV	2.40	8.77	15.39	1.63	6.73	12.86	1.32	6.11	11.55	1.06	5.72	10.77
$(\hat{\mu}^{\dagger 2}, \hat{\mu}^{\dagger 2}, \hat{\mu}^{\dagger 2})$	RV	2.60	8.71	15.21	1.57	6.92	13.12	1.38	6.34	12.24	1.19	6.05	10.97
$(\hat{\eta}^2, \hat{\eta}^2, \hat{\mu}^{\dagger 2})$	RV	2.06	8.31	14.91	1.72	6.79	12.38	1.22	6.27	11.49	1.12	5.59	10.58
$(\hat{\eta}^2, \hat{\mu}^{\dagger 2}, \hat{\mu}^{\dagger 2})$	RV	2.14	7.86	14.43	1.64	6.66	12.47	1.23	6.15	11.66	1.10	5.63	11.01
$(\hat{\eta}^2, \hat{\mu}^{\dagger 2}, \hat{\eta}^2)$	RV	2.29	8.23	15.12	1.52	6.66	12.52	1.32	5.95	11.57	0.97	5.60	10.96

com $\delta = 1.8$, $\beta_0 = 1.7$ and $\beta_1 = -1.7$. Quando ocorre erro de especificação também nos outros submodelos, nós geramos os dados usando, adicionalmente

$$h(\alpha_t) = (\gamma_0 + \gamma_1 z_{t1})^{\delta},$$
$$b(\phi_t) = (\lambda_0 + \lambda_1 s_{t1})^{\delta},$$

em que $\gamma_0 = 0.7$, $\gamma_1 = -0.7$, $\lambda_0 = 3.0$ and $\lambda_1 = -1.0$. O modelo é estimado sem considerar a não-linearidade no preditor linear ($\delta = 1$). Nosso interesse é avaliar se o teste proposto é capaz de detectar erro de especificação quando a não-linearidade é omitida.

A Tabela 3.2 apresenta as taxas de rejeição da hipótese nula sob a hipótese alternativa de não-linearidade. Considerando erro de especificação apenas no submodelo da média nós notamos que, em geral, poderes mais altos do teste R_μ são obtidos quando $\hat{\eta}^2$ é usada como variável de teste. Por exemplo, quando $n = 80$ e ao nível nominal de 5%, os percentuais de rejeição da hipótese nula para o teste escore é 96.85% quando $\hat{\eta}^2$ é usada como variável de teste e 35.39% quando a variável de teste é $\hat{\mu}^{\dagger 2}$. É interessante notar que quando o submodelo da média é o único submodelo incorretamente especificado, o teste R_μ com $\hat{\eta}^2$ como variável de teste é, em geral, mais poderoso do que o teste R_θ (no qual todos os três submodelos são aumentados).

Considerando erro de especificação nos três submodelos, o teste R_μ baseado em $\hat{\eta}^2$ e o teste R_θ baseado em $(\hat{\eta}^2, \hat{\eta}^2, \hat{\eta}^2)$ são igualmente poderosos. A escolha da variável de teste pode ter um impacto decisivo no poder do teste de erro de especificação. Como antes, o poder do teste R_μ quando $\hat{\mu}^{\dagger 2}$ é usada como variável de teste é substancialmente menor do que o poder do teste baseado em $\hat{\eta}^2$. Adicionalmente, a melhor escolha de variáveis de teste para R_θ é $(\hat{\eta}^2, \hat{\eta}^2, \hat{\eta}^2)$.

3.4.2.2 Função de Ligação Incorreta

A fim de considerar erro de especificação na função de ligação, dados são gerados através do Modelo (3.4.1) com função de ligação log-log complementar para μ. Aqui, $\beta_0 = -1.5$ e $\beta_1 = 2.5$. O modelo é estimado incorretamente usando função de ligação logit para μ. Quando todos os três submodelos são incorretamente especificados, adicionalmente ao submodelo gerado para a média, nós geramos os dados usando função de ligação log-log complementar para α e função de ligação inversa para ϕ. Aqui, $\gamma_0 = -1.7$, $\gamma_1 = 1.0$, $\lambda_0 = 0.03$ e $\lambda_1 = -0.02$. O modelo é estimado incorretamente usando função de ligação logit para α e função de ligação logarítmica para ϕ.

A Tabela 3.3 apresenta as taxas de rejeição da hipótese nula sob a hipótese alternativa de erro de especificação na função de ligação. O teste R_μ com $\hat{\eta}^2$ como variável de teste tem um desempenho melhor, no que se refere ao poder, do que o teste R_θ. Note que o poder mostra-se bastante superior quando incluímos $\hat{\eta}^2$ ao invés de $\hat{\mu}^{\dagger 2}$ como variável de teste no teste R_μ. O poder máximo obtido é de 29.03%, considerando $\hat{\mu}^{\dagger 2}$, e 99.95% considerando $\hat{\eta}^2$. Quando as três funções de ligação são incorretamente especificadas, o teste R_μ com $\hat{\eta}^2$ como variável de teste notavelmente tem uma performance melhor do que o teste R_θ.

Tabela 3.2 Taxas de rejeição da hipótese nula (%): não-linearidade negligenciada.

Erro de especificação no submodelo para a média													
TesteR_μ													
Regressores adicionados	**Tests**	$n = 40$			$n = 80$			$n = 120$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$\hat{\eta}^2$	escore	48.36	74.46	84.46	88.79	96.85	98.72	98.26	99.76	99.91	99.98	100.00	100.00
$\hat{\mu}^{\dagger 2}$	escore	6.87	19.22	28.14	19.08	35.39	47.06	28.52	48.69	60.48	47.24	69.19	79.29
Teste R_θ													
Regressores adicionados	**Tests**	$n = 40$			$n = 80$			$n = 120$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$(\hat{\eta}^2, \hat{\eta}^2, \hat{\eta}^2)$	RV	31.89	59.39	71.34	75.63	91.45	95.61	93.69	98.76	99.59	99.87	99.97	99.99
$(\hat{\mu}^{\dagger 2}, \hat{\mu}^{\dagger 2}, \hat{\mu}^{\dagger 2})$	RV	4.63	14.16	23.20	12.84	26.85	36.84	19.46	36.57	48.18	31.34	52.70	66.16
Erros de especificação nos três submodelos													
Teste R_μ													
Regressores adicionados	**Tests**	$n = 40$			$n = 80$			$n = 120$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$\hat{\eta}^2$	escore	68.69	86.79	92.09	99.16	99.80	99.91	99.98	100.00	100.00	100.00	100.00	100.00
$\hat{\mu}^{\dagger 2}$	escore	7.07	18.08	25.55	25.32	35.94	42.01	34.87	43.53	49.57	41.25	50.61	56.99
Teste R_θ													
Regressores adicionados	**Tests**	$n = 40$			$n = 80$			$n = 120$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$(\hat{\eta}^2, \hat{\eta}^2, \hat{\eta}^2)$	RV	59.29	86.55	92.84	99.43	99.85	99.91	99.99	100.00	100.00	100.00	100.00	100.00
$(\hat{\mu}^{\dagger 2}, \hat{\mu}^{\dagger 2}, \hat{\mu}^{\dagger 2})$	RV	5.54	17.92	26.59	26.63	36.83	43.64	34.77	44.42	50.66	41.98	49.89	55.58

Tabela 3.3 Taxas de rejeição da hipótese nula (%): função de ligação incorreta.

Erro de especficação no submodelo para a média													
Teste R_μ													
Regressores adicionados	**Testes**	$n = 40$			$n = 80$			$n = 120$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$\hat{\eta}^2$	escore	29.81	53.61	66.23	64.79	85.48	91.95	86.81	96.24	98.34	98.83	99.85	99.95
$\hat{\mu}^{\dagger 2}$	escore	2.19	8.14	14.99	3.55	11.75	19.67	4.24	14.00	23.06	8.53	18.86	29.03
Teste R_θ													
Regressores adicionados	**Testes**	$n = 40$			$n = 80$			$n = 120$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$(\hat{\eta}^2, \hat{\eta}^2, \hat{\eta}^2)$	RV	17.87	37.58	51.11	46.87	70.88	81.07	70.50	89.08	94.07	95.18	98.88	99.54
$(\hat{\mu}^{\dagger 2}, \hat{\mu}^{\dagger 2}, \hat{\mu}^{\dagger 2})$	RV	2.22	8.61	15.86	3.33	12.79	20.54	5.00	15.09	24.92	8.22	21.97	33.59
Erro de especificação nos três submodelos													
Teste R_μ													
Regressores adicionados	**Testes**	$n = 40$			$n = 80$			$n = 120$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$\hat{\eta}^2$	escore	22.14	45.73	57.82	48.78	74.81	84.11	75.10	90.45	94.55	95.58	98.68	99.33
$\hat{\mu}^{\dagger 2}$	escore	1.86	7.41	13.36	2.28	8.05	14.30	2.83	9.13	15.75	3.26	10.33	16.93
Teste R_θ													
Regressores adicionados	**Testes**	$n = 40$			$n = 80$			$n = 120$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$(\hat{\eta}^2, \hat{\eta}^2, \hat{\eta}^2)$	RV	10.44	29.16	40.76	29.72	55.94	68.34	55.80	77.34	85.94	86.52	95.06	97.74
$(\hat{\mu}^{\dagger 2}, \hat{\mu}^{\dagger 2}, \hat{\mu}^{\dagger 2})$	RV	1.93	8.05	15.40	3.35	12.12	19.49	4.81	14.42	22.93	7.02	19.55	29.95

3.4.2.3 Omissão de variável importante

Devemos agora considerar a situação em que o erro de especificação ocorre pela não inclusão de um regressor importante no modelo. Como antes, nós consideramos dois cenários. Primeiro, o erro de especificação ocorre apenas no submodelo da média, que é estimado considerando apenas a variável x_{t1} como regressor. O verdadeiro processo gerador dos dados é dado por

$$g(\mu_t) = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2},$$

com $\beta_0 = -0.1, \beta_1 = 1.5$ and $\beta_2 = -2.0$. Segundo, quando todos os três submodelos apresentam erro de especificação, nós temos, em adição, que o mecanismo de geração dos dados é tal que

$$h(\alpha_t) = \gamma_0 + \gamma_1 z_{t1} + \gamma_2 z_{t2},$$
$$d(\phi_t) = \lambda_0 + \lambda_1 s_{t1} + \lambda_2 s_{t2},$$

em que $\gamma_0 = -0.5$, $\gamma_1 = 0.5$, $\gamma_2 = -1.5$, $\lambda_0 = 3.5$, $\lambda_1 = 1.5$ e $\lambda_2 = -3.5$. O modelo é estimado omitindo as variáveis x_{t2}, z_{t2} e s_{t2}. Os resultados apresentados na Tabela 3.4 levam a importantes conclusões. Primeiro, o teste R_μ (no qual nós apenas aumentamos o submodelo da média) é consideravelmente mais poderoso quando $\hat{\eta}^2$ é usado como variável de teste. Segundo, novamente R_μ é mais poderoso do que R_θ quando o erro de especificação ocorre no submodelo da média e também em todos três submodelos, especialmente quando o tamanho da amostra é pequeno. Por exemplo, para $n = 40$ e ao nível nominal de 10%, os poderes das melhores performance dos testes R_μ e R_θ, quando todos os três submodelos contêm erros, são 59.83% e 36.83%, respectivamente. Terceiro, o teste R_θ tem um melhor desempenho quando $(\hat{\eta}^2, \hat{\eta}^2, \hat{\eta}^2)$ são usadas como variáveis de teste.

3.4.2.4 Correlação espacial negligenciada

É frequente o uso de modelos que envolvem dependência estatística. No estudo de fenômenos posicionados no espaço, observações de unidades vizinhas apresentam, em geral, correlação ou dependência espacial, isto é, valores observados em áreas geográficas adjacentes se mostram mais parecidos ou diferentes do que o esperado sob independência. Assim, torna-se de suma importância a avaliação do desempenho do teste RESET para modelos de regressão beta inflacionados quando a correlação espacial existente nos dados não é considerada. Neste cenário, os dados são gerados com correlação espacial. A ideia consiste em gerar efeitos latentes normais espacialmente correlacionados e introduzir esse efeito espacial no modelo através do preditor linear. Dessa forma, o preditor linear do modelo com correlação espacial será o preditor linear do modelo somado ao efeito latente. Os efeitos latentes foram gerados no `R` através da função `grf` do pacote `geoR` que simula campos gaussianos aleatórios para covariâncias especificadas. O modelo é estimado sem considerar a correlação espacial existente nos dados. Os valores dos parâmetros no submodelo para a média são $\beta_0 = -0.8$ e $\beta_1 = 3.5$. Quando todos os três submodelos contêm erros de especificação, nós temos, em adição, que $\gamma_0 = -2.0$ e $\gamma_1 = 1.5$ (submodelo para α) e $\lambda_0 = 6.0$ e $\lambda_1 = -3.5$ (submodelo para ϕ).

Os resultados da simulação são apresentados na Tabela 3.5. Importantes conclusões podem ser tiradas desses resultados. Primeiro, os poderes de todos os testes são mais baixos quando

Tabela 3.4 Taxas de rejeição da hipótese nula (%): variáveis omitidas.

Erro de especificação no submodelo para a média													
Teste R_μ													
Regressores adicionados	**Testes**	$n = 40$			$n = 80$			$n = 120$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$\hat{\eta}^2$	escore	59.87	80.03	89.77	94.14	98.76	99.55	99.52	99.96	99.98	99.99	100.00	100.00
$\hat{\mu}^{\dagger 2}$	escore	0.89	3.77	8.23	1.17	4.45	8.60	1.10	4.06	8.75	1.04	4.03	8.71
Teste R_θ													
Regressores adicionados	**Testes**	$n = 40$			$n = 80$			$n = 120$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$(\hat{\eta}^2, \hat{\eta}^2, \hat{\eta}^2)$	RV	38.34	66.95	79.14	84.58	95.62	97.92	97.97	99.60	99.88	99.98	99.99	100.00
$(\hat{\mu}^{\dagger 2}, \hat{\mu}^{\dagger 2}, \hat{\mu}^{\dagger 2})$	RV	0.99	4.63	9.61	1.33	4.84	9.60	1.34	4.90	9.09	1.15	4.78	9.51
Erro de especificação nos três submodelos													
Teste R_μ													
Regressores adicionados	**Testes**	$n = 40$			$n = 80$			$n = 120$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$\hat{\eta}^2$	escore	23.78	46.16	59.83	55.13	79.47	86.88	80.04	93.12	96.56	97.16	99.17	99.67
$\hat{\mu}^{\dagger 2}$	escore	1.17	5.30	10.50	1.69	5.97	11.20	1.76	6.33	11.68	2.28	7.97	13.73
Teste R_θ													
Regressores adicionados	**Testes**	$n = 40$			$n = 80$			$n = 120$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$(\hat{\eta}^2, \hat{\eta}^2, \hat{\eta}^2)$	RV	7.54	24.04	36.83	29.05	53.14	66.15	51.15	72.88	82.35	81.82	93.71	96.56
$(\hat{\mu}^{\dagger 2}, \hat{\mu}^{\dagger 2}, \hat{\mu}^{\dagger 2})$	RV	0.72	3.94	8.71	1.40	6.03	10.96	1.52	6.46	12.07	2.08	8.18	14.11

a amostra é pequena. Grandes amostras são necessárias para detectar confiavelmente a correlação espacial negligenciada. É por isso que agora relatamos resultados para amostras de tamanhos que variam de 40 a 1,000 observações. Segundo, novamente o teste R_μ foi consideravelmente mais poderoso quando o submodelo para a média foi aumentado usando $\hat{\eta}^2$. Quando $n = 200$ e ao nível nominal de 5%, o poder da implementação escore do teste R_μ baseado em $\hat{\eta}^2$ é 40.68%, ao passo que o poder correspondente baseado em $\hat{\mu}^{\dagger 2}$ é 6.62%. Terceiro, ao contrário dos cenários de má especificação anteriores, R_μ é menos poderoso do que R_θ. Por exemplo, quando $n = 200$ e ao nível nominal de 10%, os poderes das melhores performances dos testes R_μ e R_θ são, respectivamente, 62.22% e 96.31% quando o erro de especificação ocorre apenas no submodelo da média, e 35.06% e 96.23% sob erro de especificação em todos os três submodelos. No entanto, nós notamos que o teste R_μ apresenta bons poderes quando o tamanho da amostra é grande.

3.4.2.5 Dispersão variável negligenciada

Uma propriedade importante dos testes de erro de especificação que é avaliada neste trabalho é a capacidade de identificar a falta de ajuste do modelo quando a dispersão dos dados é variável e o modelo é estimado com dispersão constante. Nós temos realizado simulações para verificar se o teste proposto é capaz de detectar tal falta de ajuste. Aqui, nós apenas consideramos inferência baseada no teste R_μ. A resposta foi gerada considerando $\beta_0 = -0.8$, $\beta_1 = 3.5$, $\gamma_0 = -2.0$, $\gamma_1 = 1.5$, $\lambda_0 = 6.0$ e $\lambda_1 = -3.5$. A estimação dos parâmetros foi realizada assumindo que ϕ é constante. Os resultados apresentados na Tabela 3.6 mostram que a implementação escore oferece o teste mais poderoso. Os resultados também mostram que o teste baseado em $\hat{\eta}^2$ é novamente consideravelmente mais poderoso do que o teste baseado em $\hat{\mu}^{\dagger 2}$. Por exemplo, para $n = 80$ e ao nível de significância de 10%, o poder da implementação escore do teste R_μ é 89.50 quando $\hat{\eta}^2$ é usada como variável de teste e 67.93% quando $\hat{\mu}^{\dagger 2}$ é usada.

3.4.3 Simulações de tamanho e poder para o modelo de regressão beta inflacionado em zero ou um com dispersão fixa

Um estudo de simulação, semelhante ao apresentado para o modelo beta inflacionado com dispersão variável, é realizado para modelos de regressão beta inflacionados com dispersão constante. O modelo de regressão beta inflacionado com dispersão constante usado na simulação é

$$\begin{aligned} h(\alpha_t) &= \gamma_0 + \gamma_1 z_{t1}, \\ g(\mu_t) &= \beta_0 + \beta_1 x_{t1}, \end{aligned} \tag{3.4.2}$$

$t = 1, \ldots, n$, $h(\cdot)$ e $g(\cdot)$ são funções de ligação e z_{t1} e x_{t1} são as covariáveis utilizadas. Para todas as simulações, considera-se $\phi = 70$. Para cada covariável são geradas 40 observações da distribuição uniforme padrão e então repetidas 2, 3 e 5 vezes para obter amostras de tamanho 80, 120 e 200. Os resultados são baseados em 10,000 réplicas de Monte Carlo e em cada réplica é gerada uma amostra aleatória da variável dependente $y = (y_1, \ldots, y_n)^\top$ com $y_t \sim \text{BEOI}(\alpha_t, \mu_t, \phi)$. Os erros de especificação considerados na simulação de poder são devidos a não-linearidade

Tabela 3.5 Taxas de rejeição da hipótese nula (%): correlação espacial negligenciada.

Erro de especificação no submodelo para a média													
Teste R_μ													
Regressores adicionados	**Testes**	$n = 40$		$n = 80$		$n = 120$		$n = 200$		$n = 400$		$n = 1000$	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$\hat{\eta}^2$	escore	2.79	7.25	8.58	19.54	16.97	32.79	40.68	62.22	85.22	93.91	99.98	100.00
$\hat{\mu}^{\dagger 2}$	escore	1.37	3.45	2.02	5.91	2.93	8.46	6.62	15.38	20.09	34.44	68.30	82.89
Teste R_θ													
Regressores adicionados	**Testes**	$n = 40$		$n = 80$		$n = 120$		$n = 200$		$n = 400$		$n = 1000$	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$(\hat{\eta}^2, \hat{\eta}^2, \hat{\eta}^2)$	RV	11.51	21.39	32.37	48.74	57.56	73.69	89.84	96.31	99.96	100.00	100.00	100.00
$(\hat{\eta}^2, \hat{\zeta}^2, \hat{\kappa}^2)$	RV	3.90	9.12	14.04	25.27	26.34	41.47	57.24	73.52	95.66	98.51	100.00	100.00
Erro de especificação nos três submodelos													
Teste R_μ													
Regressores adicionados	**Testes**	$n = 40$		$n = 80$		$n = 120$		$n = 200$		$n = 400$		$n = 1000$	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$\hat{\eta}^2$	escore	1.45	4.68	4.18	10.67	7.87	18.22	19.04	35.06	51.47	69.53	96.46	99.01
$\hat{\mu}^{\dagger 2}$	escore	2.87	6.52	1.16	3.42	1.20	3.96	2.38	5.93	5.06	10.23	16.16	29.30
Teste R_θ													
Regressores adicionados	**Testes**	$n = 40$		$n = 80$		$n = 120$		$n = 200$		$n = 400$		$n = 1000$	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
$(\hat{\eta}^2, \hat{\eta}^2, \hat{\eta}^2)$	RV	16.18	26.33	39.95	55.08	63.84	77.27	91.22	96.23	99.96	99.99	100.00	100.00
$(\hat{\eta}^2, \hat{\zeta}^2, \hat{\kappa}^2)$	RV	1.65	4.91	9.51	18.17	16.04	27.10	32.90	48.29	73.82	84.58	99.69	99.85

Tabela 3.6 Taxas de rejeição da hipótese nula (%): dispersão variável negligenciada.

		Teste R_μ											
Regressores adicionados	**Testes**	$n = 40$			$n = 80$			$n = 120$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$\hat{\eta}^2$	RV	39.00	59.18	69.20	69.76	84.31	89.26	86.65	94.09	96.41	97.69	99.27	99.66
	escore	42.07	61.01	70.01	71.67	84.49	89.50	87.95	94.30	96.47	97.71	99.29	99.65
	Wald	32.49	56.10	68.08	66.77	83.70	88.97	85.63	94.04	96.33	97.51	99.30	99.66
$\hat{\mu}^{\dagger 2}$	RV	17.93	34.34	44.88	39.61	58.60	68.41	57.33	73.48	82.29	78.38	91.48	95.00
	escore	20.34	34.32	44.62	40.24	58.33	67.93	56.58	73.57	82.02	78.52	91.20	94.82
	Wald	14.06	32.46	43.98	36.79	57.54	68.01	56.12	73.00	82.06	77.92	91.39	95.02

no preditor linear, variáveis omitidas, escolha incorreta da função de ligação e correlação espacial existente nos dados. Os testes são realizados considerando como variáveis auxiliares os quadrados dos preditores lineares estimados, bem como os quadrados dos valores preditos.

Nas simulações de tamanho a resposta é gerada considerando função de ligação logit, $\beta_0 = -1.0, \beta_1 = 3.5, \gamma_0 = -1.5$ e $\gamma_1 = 1.0$. Sob não-linearidade a resposta é gerada usando a seguinte especificação:

$$g(\mu_t) = (\beta_0 + \beta_1 x_{t1})^{\delta},$$

com $\delta = 1.8$, $\beta_0 = 1.9$ e $\beta_1 = -1.9$. Quando os erros de especificação ocorrem em ambos os submodelos (para μ e α), nós temos, adicionalmente, que

$$h(\alpha_t) = (\gamma_0 + \gamma_1 z_{t1})^{\delta},$$

em que $\gamma_0 = 0.7$ e $\gamma_1 = -0.7$. Aqui, nós devemos usar a seguinte notação: $\theta' = (\mu, \alpha)^{\top}$. Portanto, $R_{\theta'}$ denota o teste em que nós aumentamos os dois submodelos (μ and α).

Considerando como erro de especificação a função de ligação incorreta os dados são gerados através do Modelo (3.4.2) supondo função de ligação log-log complementar para μ. Aqui, $\beta_0 = -1.5$ e $\beta_1 = 2.5$. O modelo é estimado incorretamente usando função de ligação logit para μ. Quando, em adição, o submodelo para α é incorretamente especificado, a geração dos dados usa a ligação log-log complementar e a estimação é baseada na ligação logit. Neste caso, $\gamma_0 = -1.7$ e $\gamma_1 = 1.0$.

Sob variáveis omitidas, o processo de geração dos dados inclui um segundo regressor, x_{t2}, no submodelo da média. Nesse caso, $\beta_0 = -0.1$, $\beta_1 = 1.5$ e $\beta_2 = -2.0$. O modelo é estimado omitindo-se a variável x_{t2}. Quando o erro de especificação também ocorre no submodelo para α, a geração dos dados inclui um regressor adicional, z_{t2}. Aqui, $\gamma_0 = -0.5$, $\gamma_1 = 0.5$ e $\gamma_2 = -1.5$. O modelo é estimado omitindo-se a variável z_{t2}.

Finalmente, nas simulações de poder, para o caso de correlação espacial negligenciada, nós geramos os dados com correlação espacial e estimamos os parâmetros da regressão sem considerar tal correlação. Aqui $\beta_0 = -1.0$ e $\beta_1 = 3.5$. Quando o submodelo para α também contém erro de especificação, nós usamos $\gamma_0 = -1.5$ e $\gamma_1 = 1.0$.

Por brevidade, nós apenas apresentamos os resultados para a implementação escore do teste R_μ usando $\hat{\eta}^2$ como variável de teste e para a implementação da razão de verossimilhanças do teste $R_{\theta'}$ baseado em $(\hat{\eta}^2, \hat{\eta}^2)$, isto é, quando $\hat{\eta}^2$ é usado para aumentar o submodelo para μ e o submodelo para α; essas são as melhores performance dos testes. Os resultados da simulação de Monte Carlo são apresentados na Tabela 3.7. Nós notamos que as distorções de tamanho dos dois testes são pequenas, com uma leve vantagem para o teste R_μ. Nós também concluímos que os testes são satisfatoriamente poderosos, exceto sob correlação espacial negligenciada, caso em que o tamanho da amostra deve ser grande para que a probabilidade do erro tipo II seja pequena. Curiosamente e ao contrário do que observamos sob dispersão variável, o teste R_μ é mais poderoso do que o teste $R_{\theta'}$ quando o erro de especificação é a correlação espacial negligenciada (lembrando que sob dispersão variável, o teste que nós chamamos de R_θ é equivalente ao teste $R_{\theta'}$ aqui). De fato, em ambos os cenários (considerando erro de especificação apenas no submodelo da média e nos dois submodelos (μ e α)), o teste R_μ normalmente supera o teste $R_{\theta'}$.

Tabela 3.7 Taxas de rejeição da hipótese nula (%) sob dispersão fixa.

Tamanho										
Cenário	**Testes**	**Regressores adicionados**	$n = 40$		$n = 80$		$n = 120$		$n = 200$	
			5%	10%	5%	10%	5%	10%	5%	10%
Teste R_μ	escore	$\hat{\eta}^2$	5.68	11.07	5.37	10.57	5.16	10.08	4.77	9.97
Teste $R_{\theta'}$	RV	$(\hat{\eta}^2, \hat{\eta}^2)$	6.96	13.31	5.70	11.01	5.45	11.01	5.40	10.49
Poder: Não-linearidade										
Cenário	**Testes**	**Regressores adicionados**	$n = 40$		$n = 80$		$n = 120$		$n = 200$	
			5%	10%	5%	10%	5%	10%	5%	10%
Teste R_μ com erro em μ	escore	$\hat{\eta}^2$	89.19	94.19	99.58	99.90	99.99	100.00	100.00	100.00
Teste $R_{\theta'}$ com erro em μ	RV	$(\hat{\eta}^2, \hat{\eta}^2)$	82.90	90.51	99.25	99.64	99.98	99.99	100.00	100.00
Teste R_μ com erro em μ e α	escore	$\hat{\eta}^2$	65.56	76.68	93.28	96.70	99.16	99.65	99.98	99.99
Teste $R_{\theta'}$ com erro em μ e α	RV	$(\hat{\eta}^2, \hat{\eta}^2)$	58.86	71.40	90.32	94.89	98.49	99.35	99.98	99.99
Poder: Função de ligação incorreta										
Cenário	**Testes**	**Regressores adicionados**	$n = 40$		$n = 80$		$n = 120$		$n = 200$	
			5%	10%	5%	10%	5%	10%	5%	10%
Teste R_μ com erro em μ	escore	$\hat{\eta}^2$	58.00	70.63	88.06	93.46	97.30	98.92	99.83	99.97
Teste $R_{\theta'}$ com erro em μ	RV	$(\hat{\eta}^2, \hat{\eta}^2)$	46.70	60.41	81.83	89.14	94.50	97.34	99.58	99.81
Teste R_μ com erro em μ e α	escore	$\hat{\eta}^2$	58.24	70.83	88.33	93.34	97.38	99.00	99.85	99.95
Teste $R_{\theta'}$ com erro em μ e α	RV	$(\hat{\eta}^2, \hat{\eta}^2)$	47.16	60.76	82.19	89.24	84.87	94.61	99.51	99.81
Poder: Variável Omitida										
Cenário	**Testes**	**Regressores adicionados**	$n = 40$		$n = 80$		$n = 120$		$n = 200$	
			5%	10%	5%	10%	5%	10%	5%	10%
Teste R_μ com erro em μ	escore	$\hat{\eta}^2$	86.85	92.89	99.43	99.78	99.96	99.99	100.00	100.00
Teste $R_{\theta'}$ com erro em μ	RV	$(\hat{\eta}^2, \hat{\eta}^2)$	79.27	88.73	98.54	99.50	99.90	99.96	100.00	100.00
Teste R_μ com erro em μ e α	escore	$\hat{\eta}^2$	83.18	90.56	98.93	99.50	99.95	99.98	100.00	100.00
Teste $R_{\theta'}$ com erro em μ e α	RV	$(\hat{\eta}^2, \hat{\eta}^2)$	75.07	85.22	97.81	99.21	99.84	99.95	100.00	100.00
Poder: Correlação espacial negligenciada										
Cenário	**Testes**	**Regressores adicionados**	$n = 40$		$n = 80$		$n = 120$		$n = 200$	
			5%	10%	5%	10%	5%	10%	5%	10%
Teste R_μ com erro em μ	escore	$\hat{\eta}^2$	9.08	19.39	26.64	41.93	44.75	63.45	75.78	88.35
Teste $R_{\theta'}$ com erro em μ	RV	$(\hat{\eta}^2, \hat{\eta}^2)$	6.19	13.24	17.53	30.90	31.59	48.15	60.98	77.32
Teste R_μ com erro em μ e α	escore	$\hat{\eta}^2$	8.32	19.30	28.86	46.94	51.51	71.24	84.55	93.13
Teste $R_{\theta'}$ com erro em μ e α	RV	$\hat{\eta}^2, (\hat{\eta}^2)$	6.63	13.73	22.24	37.41	39.70	57.68	74.48	87.45

Tabela 3.8 Variáveis contínuas para os dados de eficiência do estado de São Paulo.

Variável	Descrição	Média (desvio padrão)
$EFIC$	Escores de eficiência DEA-CCR	0.66 (0.19)
$DESP$	Despesas com pessoal	1.95e+07 (9.32e+07)
SAL	% Chefes ganham até um salário	2.01 (1.84)
$REND$	Rendimento médio	700.15 (232.37)
CAD	Índice atualização cadastro predial	0.91 (0.12)
$DENS$	Densidade demográfica	306.89 (1146.85)
URB	Taxa de urbanização	82.45 (15.67)

3.5 Aplicação

Devemos agora aplicar o teste proposto a um conjunto de dados reais. Os dados utilizados referem-se a índices de eficiência para municípios do estado de São Paulo. Cribari-Neto e Pereira (2010) consideraram modelos de regressão beta e beta inflacionado com efeitos espaciais para lidar com os dados de eficiência. Adicionalmente, uma comparação com os modelos de regressão linear e com inferência realizada via quasi-verossimilhança também foi apresentada pelos autores. Uma vez que os dados contêm uns (os quais correspondem a administrações pública plenamente eficientes), os autores estimaram a regressão beta após substituir tais observações por um valor levemente menor. O modelo de regressão beta inflacionado se mostrou mais adequado para explicar os escores de eficiência. A base de dados utilizada contém 427 municípios do estado de São Paulo e as informações referem-se ao ano de 2000. As Tabelas 3.8 e 3.9 apresentam uma breve descrição das variáveis, juntamente com algumas medidas descritivas. Note que todas as variáveis na Tabela 3.9 são do tipo *dummy*, ou seja, elas apenas assumem valores 0 e 1. A Tabela 3.10 apresenta as estimativas de máxima verossimilhança dos parâmetros que indexam o modelo beta inflacionado escolhido usando a função de ligação logit para μ e α e a função de ligação logarítmica para ϕ. A variável dependente são os escores da eficiência dos municípios do estado de São Paulo.

O nosso objetivo aqui é testar se o modelo estimado está corretamente especificado. Para este fim, nós aplicamos os testes propostos neste capítulo para o modelo de regressão beta inflacionado estimado por Cribari-Neto e Pereira (2010). A implementação escore do teste R_μ baseado em $\hat{\eta}^2$ e a implementação da razão de verossimilhança do teste R_θ baseada em $(\hat{\eta}^2, \hat{\eta}^2, \hat{\eta}^2)$ foram usadas. A estatística de teste escore R_μ é 4.94 ($p - valor = 0.03$) e a estatística da razão de verossimilhanças no caso do teste R_θ é 6.70 ($p - valor = 0.08$). Portanto, a especificação correta do modelo não é rejeitada ao nível de 1% de significância quando nós usamos o teste R_μ e ao nível de 5% quando a inferência é baseada no teste R_θ.

Tabela 3.9 Variáveis *dummy* para os dados de eficiência do estado de São Paulo.

Variável	**Descrição**	**Frequência** (%)
PFL	Prefeito do PFL	52 (12.18)
$PMDB$	Prefeito do PMDB	76 (17.80)
$PSDB$	Prefeito do PSDB	118 (27.63)
PT	Prefeito do PT	30 (7.02)
PPS	Prefeito do PPS	24 (5.62)
PPB	Prefeito do PPB	21 (4.92)
PTB	Prefeito do PTB	50 (11.71)
PDT	Prefeito do PDT	14 (3.28)
PCI	Participação consórcios intermunicipais	75 (17.56)
$INFO$	Grau de informatização	390 (91.33)
$PDCM$	Poder decisão conselhos municipais	217 (50.82)
$ALVO$	Participação no Projeto Alvorada	3 (0.70)
$IDADE$	Idade do município	36 (8.43)
MT	Municípios turísticos	72 (16.86)
ROY	Recebimento de royalties	55 (12.88)

Tabela 3.10 Estimativas dos parâmetros do modelo beta inflacionado com dispersão variável usando os escores de eficiência para os dados de São Paulo.

Submodelo para μ		
Variáveis	**Estimativa**	**Erro-Padrão**
constante	0.1403	0.2012
W1EFIC	0.0142	0.0080
REND	−0.0007	0.0001
PT	0.2776	0.1495
URB	0.0095	0.0026
Submodelo para α		
Variáveis	**Estimativa**	**Erro-Padrão**
constante	7.774e-01	1.271e+00
DESP	9.363e-09	4.915e-09
SAL	−5.244e-01	2.267e-01
REND	−2.835e-03	1.511e-03
PCI	1.357e+00	4.296e-01
INFO	−1.068e+00	5.248e-01
Submodelo para ϕ		
Variáveis	**Estimativa**	**Erro-Padrão**
constant	2.0711	0.3028
W1EFIC	−0.0384	0.0139
REND	0.0012	0.0003
INFO	−0.7416	0.2538

3.6 Conclusões

Neste capítulo nós propomos testes de erro de especificação para modelos de regressão beta inflacionados com dispersão variável e fixa. Em particular, nós propomos duas variantes do teste. Na primeira variante nós apenas aumentamos o submodelo da média ao passo que na segunda todos os três submodelos são aumentados com variáveis de teste. A hipótese nula de que o modelo está corretamente especificado é rejeitada sempre que as variáveis adicionadas melhoram sensivelmente a qualidade do ajuste do modelo. Nós consideramos diferentes estratégias de teste e comparamos suas performances em amostras finitas usando simulações de Monte Carlo. Diferentes escolhas de erros de especificação foram consideradas, a saber: não-linearidade negligenciada, escolha incorreta da função de ligação, variáveis omitidas, correlação espacial negligenciada e dispersão variável não modelada. Os resultados mostraram que o teste proposto apresenta bons poderes em amostras de tamanho pequeno a moderado, exceto quando a correlação espacial é negligenciada, neste caso tamanhos amostrais maiores são necessários para obter menores probabilidades do erro tipo II. Evidências numéricas têm também mostrado que o mais simples dos dois testes, isto é, quando apenas o submodelo da média é aumentado, apresenta tipicamente o melhor desempenho. Este teste é realizado usando os quadrados dos preditores lineares estimados como variáveis de teste.

Nós acreditamos que o teste proposto pode ser bastante útil em situações práticas, uma vez que é capaz de detectar várias formas de erro de especificação em modelos de regressão beta inflacionados com e sem dispersão variável. Nós recomendamos aos usuários que desejam modelar dados usando regressão beta inflacionada usar o teste para verificar se os seus modelos estão corretamente especificados.

CAPÍTULO 4

Estatísticas da Razão de Verossimilhanças Modificadas para Modelos de Regressão Beta Inflacionados

4.1 Introdução

Modelos de regressão beta são comumente usados para modelar dados que assumem valores no intervalo unitário padrão $(0, 1)$, tais como taxas e proporções. Ferrari e Cribari-Neto (2004) propuseram uma classe de modelos de regressão beta em que a resposta média é relacionada a um preditor linear, que envolve covariáveis e parâmetros de regressão desconhecidos, através de uma função de ligação. O modelo também é indexado por um parâmetro de precisão e é baseado na suposição de que a variável dependente tem distribuição beta. Adicionalmente, o modelo é heteroscedástico e, facilmente, acomoda assimetrias que são características comumente observadas em dados que assumem valores no intervalo unitário padrão.

Em aplicações práticas é comum que os dados de taxas e proporções apresentem zeros e/ou uns. Uma vez que nessa situação a função de log-verossimilhança do modelo de regressão beta se torna ilimitada, o uso de tal modelo fica comprometido. Uma opção é considerar um modelo estatístico que permita adicionar à distribuição contínua da variável resposta um ponto de massa em um ou em ambos os extremos. Ospina (2008) propôs uma classe de modelos de regressão beta inflacionados, que é uma extensão natural do modelo de regressão beta proposto por Ferrari e Cribari-Neto (2004). Os modelos de regressão beta inflacionados permitem a modelagem de dados que contêm zeros e/ou uns. Tais modelos são baseados na distribuição de probabilidade beta inflacionada em zero e/ou um, em que a média da variável resposta e a probabilidade que a variável é igual a zero e/ou um são relacionados a preditores lineares por meio de funções de ligação. Pereira e Cribari-Neto (2010) propuseram testes de erro de especificação para modelos de regressão beta inflacionados com dispersão variável e fixa. Os autores apresentaram uma variante do modelo de regressão beta inflacionado que incorpora dispersão variável. Em particular, eles apresentaram a função de log-verossimilhança, as funções escore e a matriz de informação de Fisher, uma vez que essas quantidades são necessárias para obtenção das estatísticas propostas.

Após o ajuste do modelo de regressão é, em geral, de interesse do pesquisador realizar testes de hipóteses sobre os parâmetros do modelo. O teste da razão de verossimilhanças é um dos testes mais utilizados para este fim. Este teste é baseado na teoria assintótica de primeira ordem, uma vez que usa valores críticos provenientes de aproximações que são válidas em grandes amostras. Em problemas regulares, esta estatística tem, aproximadamente e sob a

hipótese nula, distribuição qui-quadrado. Entretanto, a distribuição qui-quadrado pode não ser uma boa aproximação para a distribuição nula exata da estatística da razão de verossimilhanças em amostras de tamanho pequeno ou moderado. Na classe de modelos de regressão beta inflacionados, o teste da razão de verossimilhanças usual pode apresentar tamanho distorcido se o tamanho da amostra não for suficientemente grande. Nossos resultados de simulação (ver Seção 4.4) mostram que o teste da razão de verossimilhanças é notavelmente liberal. Por exemplo, para $n = 30$ e ao nível de significância de 5%, o tamanho estimado do teste da razão de verossimilhanças para o parâmetro da estrutura de regressão da média é quase o dobro do nível nominal selecionado.

Várias correções têm sido propostas para melhorar a aproximação da distribuição da estatística da razão de verossimilhanças. Bartlett (1937, 1938, 1947, 1954) propôs um número considerável de fatores de correção visando obter uma estatística modificada com o primeiro momento igual ao da distribuição qui-quadrado de referência. Box (1949) mostrou que, para alguns tipos de critérios propostos por Bartlett, a estatística modificada segue distribuição nula mais próxima da distribuição qui-quadrado do que a estatística não modificada. Na presença de parâmetros de perturbação, inferências podem ser baseadas na função de verossimilhança perfilada. Esta função depende apenas do parâmetro de interesse, uma vez que é obtida da verossimilhança original substituindo os parâmetros de perturbação por seus estimadores de máxima verossimilhança. Wei et al. (1998) utilizaram a aproximação de Cox e Reid (1987) para a função de verossimilhança perfilada modificada e propuseram um teste que tipicamente conduz a inferências mais confiáveis em amostras finitas do que o teste da razão de verossimilhanças usual. Uma correção de Bartlett para este teste foi obtida por Cysneiros e Ferrari (2006). Ferrari e Cribari-Neto (2002) corrigiram o teste de heteroscedasticidade em modelos de regressão normais lineares proposto por Simonoff e Tsai (1994) e obtiveram um teste da razão de verossimilhanças perfiladas modificadas corrigido para o caso em que o parâmetro de interesse é unidimensional.

Atenção também tem sido dispensada à obtenção de correções para a estatística da razão de verossimilhanças sinalizada que é útil quando o parâmetro de interesse é escalar. A distribuição nula da estatística sinalizada pode ser aproximada pela distribuição normal padrão com erro de ordem $n^{-1/2}$. Barndorff-Nielsen (1986, 1991) propôs uma correção da estatística sinalizada que pode ser de difícil obtenção em problemas regulares, dado que envolve derivadas com respeito ao espaço amostral. Para contornar tal dificuldade, Skovgaard (1996) obteve uma aproximação para o ajuste de Barndorff-Nielsen. Estes resultados foram estendidos por Skovgaard (2001) para englobar a estatística da razão de verossimilanças, no caso em que o parâmetro de interesse é multidimensional. O ajuste de Skovgaard tem ampla aplicabilidade e vem sendo utilizado para corrigir a estatística da razão de verossimilhanças em várias classes de modelos. Este ajuste tem várias vantagens, tais como derivação mais simples da estatística ajustada do que a corrigida via Bartlett, aplicação direta à estatística da razão de verossimilhanças e não necessidade de ortogonalização entre os parâmetros de interesse e os de perturbação. Ferrari e Cysneiros (2008) usaram a aproximação de Skovgaard para obter uma estatística da razão de verossimilhanças ajustada para modelos não lineares da família exponencial. Ferrari e Pinheiro (2010) derivaram ajustes para a estatística da razão de verossimilhanças usual na classe de modelos de regressão beta.

Neste capítulo nós obtemos, baseados na correção de Skovgaard, ajustes para as estatísticas da razão de verossimilhanças usual e sinalizada na classe de modelos de regressão beta inflacionados com dispersão variável. Uma avaliação dos testes corrigidos é realizada via simulação de Monte Carlo. Os resultados numéricos mostram que os testes ajustados propostos aqui têm desempenhos superiores aos testes não modificados em pequenas amostras.

Na Seção 4.2 deste capítulo nós descrevemos o modelo de regressão beta inflacionado com dispersão variável. As estatísticas da razão de verossimilhanças e da razão de verossimilhanças sinalizada ajustadas são derivadas na Seção 4.3. Na Seção 4.4 nós avaliamos, via simulação de Monte Carlo, os comportamentos em amostras finitas de diferentes testes baseados na função de verossimilhança em modelos de regressão beta inflacionados. Finalmente, algumas conclusões são apresentadas na Seção 4.5. Detalhes técnicos encontram-se no Apêndice A.

4.2 Especificação do modelo de regressão beta inflacionado em zero ou um

Sejam $y_1, \ldots, y_n$ variáveis aleatórias independentes, em que cada y_t, $t = 1, \ldots, n$, segue a função de densidade

$$\mathrm{bi}_c(y; \alpha_t, \mu_t, \phi_t) = \left\{\alpha_t^{\mathbb{1}_{\{c\}}(y)}(1-\alpha_t)^{1-\mathbb{1}_{\{c\}}(y)}\right\}\left\{f(y;\mu_t,\phi_t)^{1-\mathbb{1}_{\{c\}}(y)}\right\}, \tag{4.2.1}$$

em que $\mathbb{1}_{\{c\}}(y)$ é uma função indicadora que assume valor 1 se $y = c$ e 0 caso contrário, $0 < \alpha_t < 1$ é o parâmetro de mistura da distribuição dado por $\alpha_t = \Pr(y_t = c)$, $0 < \mu_t < 1$ é a média de y_t condicional em $y_t \in (0,1)$, $\phi_t > 0$ é o parâmetro de precisão e $f(y;\mu_t,\phi_t)$ é a função de densidade beta. Sob a parametrização de Ferrari e Cribari-Neto (2004), a distribuição beta tem função de densidade

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

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

A densidade (4.2.1) é de uma variável aleatória beta inflacionada no ponto c, $c = 0$ ou $c = 1$. Para esta distribuição, $\mathbb{E}(y_t) = \alpha_t c + (1-\alpha_t)\mu_t$ e $\mathrm{Var}(y_t) = (1-\alpha_t)\mu_t(1-\mu_t)/(\phi_t+1) + \alpha_t(1-\alpha_t)(c-\mu_t)^2$.

Ospina (2008) propôs modelos de regressão beta inflacionados com dispersão constante. O autor assume que a distribuição da resposta é beta inflacionada. A parte contínua dos dados é modelada pela distribuição beta e a parte discreta, isto é, o ponto de massa, é modelada através de uma distribuição degenerada no valor conhecido c, em que c é igual a zero ou um. Apresentaremos a seguir uma classe de modelos de regressão beta inflacionados com dispersão variável.

O modelo de regressão beta inflacionado em c com dispersão variável é definido supondo que a média condicional de y_t, a massa de probabilidade em c e o parâmetro de precisão satis-

fazem às seguintes relações funcionais:

$$h(\alpha_t) = \sum_{i=1}^{M} z_{ti}\gamma_i = \zeta_t,$$
$$g(\mu_t) = \sum_{i=1}^{m} x_{ti}\beta_i = \eta_t,$$
$$b(\phi_t) = \sum_{i=1}^{q} s_{ti}\lambda_i = \kappa_t,$$

$t = 1,\ldots,n$, em que $\gamma = (\gamma_1,\ldots,\gamma_M)^\top$, $\beta = (\beta_1,\ldots,\beta_m)^\top$ e $\lambda = (\lambda_1,\ldots,\lambda_q)^\top$ são vetores de parâmetros de regressão desconhecidos, tais que $\gamma \in \mathbb{R}^M$, $\beta \in \mathbb{R}^m$ e $\lambda \in \mathbb{R}^q$, e $x_{t1},\ldots,x_{tm}$, $z_{t1},\ldots,z_{tM}$ e $s_{t1},\ldots,s_{tq}$ são observações de covariáveis conhecidas ($m + M + q < n$) que podem coincidir total ou parcialmente. As funções de ligação $h : (0,1) \to \mathbb{R}$, $g : (0,1) \to \mathbb{R}$ e $b : (0,\infty) \to \mathbb{R}$ são estritamente monótonas e duas vezes diferenciáveis. Diferentes funções de ligação podem ser usadas, tais como logit, probit, log-log complementar e log-log para μ e α e logarítimica ou raiz quadrada para ϕ.

A função de log-verossimilhança para $\theta = (\gamma^\top,\beta^\top,\lambda^\top)^\top$ é dada por

$$\ell(\theta) = \ell_1(\gamma) + \ell_2(\beta,\lambda),$$

em que

$$\ell_1(\gamma) = \sum_{t=1}^{n} \ell_t(\alpha_t) = \sum_{t=1}^{n} y_t^c\alpha_t^* + \log(1-\alpha_t),$$
$$\ell_2(\beta,\lambda) = \sum_{t:y_t\in(0,1)} \ell_t(\mu_t,\phi_t) = \sum_{t:y_t\in(0,1)} \log\Gamma(\phi_t) - \log\Gamma(\mu_t\phi_t) - \log\Gamma((1-\mu_t)\phi_t) + (\mu_t\phi_t - 1)y_t^* + (\phi_t - 2)y_t^\dagger,$$

com $\alpha_t^* = \log\left(\frac{\alpha_t}{1-\alpha_t}\right)$,

$$y_t^c = \begin{cases} 1, & y_t = c, \\ 0, & y_t \in (0,1), \end{cases}, \qquad y_t^* = \begin{cases} \log\left(\frac{y_t}{1-y_t}\right), & y_t \in (0,1), \\ 0, & y_t = c, \end{cases}$$

$$\text{e} \quad y_t^\dagger = \begin{cases} \log(1-y_t), & y_t \in (0,1), \\ 0, & y_t = c. \end{cases}$$

Escrevendo a densidade (4.2.1) na família exponencial de dimensão 3 de posto completo (Ospina e Ferrari, 2010) é possível obter os momentos de y_t^c, y_t^* e $y_t^\dagger$.

Assim, $\mu_t^c = \mathbb{E}(y_t^c) = \alpha_t$, $\mu_t^* = \mathbb{E}(y_t^*|y_t^c = 0) = \psi(\mu_t\phi_t) - \psi((1-\mu_t)\phi_t)$, $\mu_t^\dagger = \mathbb{E}(y_t^\dagger|y_t^c = 0) = \psi((1-\mu_t)\phi_t) - \psi(\phi_t)$, $v_t^c = \mathrm{Var}(y_t^c) = \alpha_t(1-\alpha_t)$, $v_t^* = \mathrm{Var}(y_t^*|y_t^c = 0) = \psi'(\mu_t\phi_t) + \psi'((1-\mu_t)\phi_t)$, $v_t^\dagger = \mathrm{Var}(y_t^\dagger|y_t^c = 0) = \psi'((1-\mu_t)\phi_t) - \psi'(\phi_t)$, $c_t = \mathrm{Cov}(y_t^*,y_t^\dagger|y_t^c = 0) = -\psi'((1-\mu_t)\phi_t)$ e $c_t^* = \mathrm{Cov}(y_t^c,y_t^*) = c_t^\dagger = \mathrm{Cov}(y_t^c,y_t^\dagger) = 0$.

A função de log-verossimilhança pode ser escrita em forma matricial como

$$\ell(\theta) = \{(y^c - \mu^c)^\top \alpha^* + a^\top + [(y^* - \mu^*)^\top (\Phi \mathcal{M} - \mathcal{I}) + (y^\dagger - \mu^\dagger)^\top (\Phi - 2\mathcal{I}) + b^\top] H\}\iota, \tag{4.2.2}$$

com $y^c = (y_1^c, \ldots, y_n^c)^\top$, $y^* = (y_1^*, \ldots, y_n^*)^\top$, $y^\dagger = (y_1^\dagger, \ldots, y_n^\dagger)^\top$, $\mu^c = (\mu_1^c, \ldots, \mu_n^c)^\top$, $\mu^* = (\mu_1^*, \ldots, \mu_n^*)^\top$, $\mu^\dagger = (\mu_1^\dagger, \ldots, \mu_n^\dagger)^\top$, $a = (a_1, \ldots, a_n)^\top$ e $b = (b_1, \ldots, b_n)^\top$. Com $a_t = \log(1-\alpha_t) + \mu_t^c \alpha_t^*$ e $b_t = \log\Gamma(\phi_t) - \log\Gamma(\mu_t\phi_t) - \log\Gamma((1-\mu_t)\phi_t) + (\mu_t\phi_t - 1)\mu_t^* + (\phi_t - 2)\mu_t^\dagger$. Na notação usada acima, $\alpha^* = \text{diag}\{\alpha_1^*, \ldots, \alpha_n^*\}$, $\mathcal{M} = \text{diag}\{\mu_1, \ldots, \mu_n\}$, $H = \text{diag}\{1-y_1^c, \ldots, 1-y_n^c\}$ e $\Phi = \text{diag}\{\phi_1, \ldots, \phi_n\}$ são matrizes diagonais $n \times n$, $\mathcal{I}$ é a matriz identidade $n \times n$ e ι é o vetor coluna de uns n-dimensional.

Dada a separabilidade dos vetores de parâmetros γ e $(\beta^\top, \lambda^\top)^\top$ é possível obter de forma independente as funções escore para γ e para $(\beta^\top, \lambda^\top)^\top$:

$$\begin{aligned} U_\gamma(\gamma) &= Z^\top P G (y^c - \mu^c), \\ U_\beta(\beta, \lambda) &= X^\top \Phi T H (y^* - \mu^*), \\ U_\lambda(\beta, \lambda) &= S^\top V H \left\{\mathcal{M}(y^* - \mu^*) + (y^\dagger - \mu^\dagger)\right\}, \end{aligned} \tag{4.2.3}$$

em que Z é uma matriz $n \times M$ cuja t-ésima linha é $z_t^\top = (z_{t1}, \ldots, z_{tM})$, X é uma matriz $n \times m$ cuja t-ésima linha é $x_t^\top = (x_{t1}, \ldots, x_{tm})$ e S é uma matriz $n \times q$ cuja t-ésima linha é $s_t^\top = (s_{t1}, \ldots, s_{tq})$. Aqui, temos que $P = \text{diag}\{1/[\alpha_1(1-\alpha_1)], \ldots, 1/[\alpha_n(1-\alpha_n)]\}$, $G = \text{diag}\{1/h'(\alpha_1), \ldots, 1/h'(\alpha_n)\}$, $T = \text{diag}\{1/g'(\mu_1), \ldots, 1/g'(\mu_n)\}$ e $V = \text{diag}\{1/b'(\phi_1), \ldots, 1/b'(\phi_n)\}$ são matrizes diagonais $n \times n$.

Como consequência da ortogonalidade entre o vetor de parâmetros γ e o vetor de parâmetros $(\beta^\top, \lambda^\top)^\top$, o estimador de máxima verossimilhança de γ é obtido como solução do sistema não-linear $U_\gamma(\gamma) = 0$ e o estimador de máxima verossimilhança de $(\beta^\top, \lambda^\top)^\top$ é obtido como solução do sistema não-linear $(U_\beta(\beta, \lambda)^\top, U_\lambda(\beta, \lambda)^\top) = 0$. Tais estimadores não possuem forma fechada e devem ser obtidos numericamente usando um algoritmo de otimização não-linear, tal como um algoritmo de Newton ou um algoritmo quasi-Newton.

Para obtenção das matrizes de informação de Fisher observada e esperada serão apresentadas a seguir as derivadas de segunda ordem da função de log-verossimilhança e seus respectivos valores esperados. Para $R = 1, \ldots, M$ e $O = 1, \ldots, M$, temos que

$$\begin{aligned} U_{RO} = \frac{\partial^2 \ell_1(\gamma)}{\partial \gamma_R \gamma_O} &= \sum_{t=1}^{n} \left\{ \frac{\partial^2 \ell_t(\alpha_t)}{\partial \alpha_t^2} \left(\frac{d\alpha_t}{d\zeta_t}\right)^2 + \frac{\partial \ell_t(\alpha_t)}{\partial \alpha_t} \left(\frac{\partial}{\partial \alpha_t} \frac{d\alpha_t}{d\zeta_t}\right) \frac{d\alpha_t}{d\zeta_t} \right\} z_{tO} z_{tR} \\ &= \sum_{t=1}^{n} -\left\{ \frac{y_t^c(1-2\alpha_t) + \alpha_t^2}{(\alpha_t(1-\alpha_t))^2} \left(\frac{1}{h'(\alpha_t)}\right)^2 + \left(\frac{y_t^c - \mu_t^c}{\alpha_t(1-\alpha_t)}\right) \left(\frac{h''(\alpha_t)}{(h'(\alpha_t))^2}\right) \right. \\ &\quad \left. \times \frac{1}{h'(\alpha_t)} \right\} z_{tO} z_{tR}. \end{aligned}$$

Uma vez que $\mathbb{E}(y_t^c) = \mu_t^c$, segue que

$$\mathbb{E}(U_{RO}) = \mathbb{E}\left(\sum_{t=1}^{n} -\left\{\frac{y_t^c(1-2\alpha_t)+\alpha_t^2}{(\alpha_t(1-\alpha_t))^2}\left(\frac{1}{h'(\alpha_t)}\right)^2\right\} z_{tO}z_{tR}\right)$$
$$= -\sum_{t=1}^{n}\frac{1}{\alpha_t(1-\alpha_t)}\left(\frac{1}{h'(\alpha_t)}\right)^2 z_{tO}z_{tR}.$$

Para $r = 1,\ldots,m$ e $o = 1,\ldots,m$,

$$U_{ro} = \frac{\partial^2\ell_2(\beta,\lambda)}{\partial\beta_r\beta_o} = \sum_{t:y_t\in(0,1)}\left\{\left(\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\mu_t^2}\right)\left(\frac{d\mu_t}{d\eta_t}\right)^2 + \left(\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\mu_t}\left(\frac{\partial}{\partial\mu_t}\frac{d\mu_t}{d\eta_t}\right)\right.\right.$$
$$\left.\left.\times\frac{d\mu_t}{d\eta_t}\right)\right\}x_{to}x_{tr}$$
$$= \sum_{t:y_t\in(0,1)} -\left\{\phi_t^2 v_t^*\left(\frac{1}{g'(\mu_t)}\right)^2 + \left((y_t^*-\mu_t^*)\phi_t\left(\frac{g''(\mu_t)}{(g'(\mu_t))^2}\right)\frac{1}{g'(\mu_t)}\right)\right\}x_{to}x_{tr}.$$

Para cumulantes que envolvem os vetores de parâmetros β e λ é necessário usar o fato de que se $\tau : (0,1) \to \mathbb{R}$ é uma função contínua, então (Ospina, 2008)

$$\mathbb{E}\left(\sum_{t:y_t\in(0,1)}\tau(y_t)\right) = \sum_{t=1}^{n}(1-\alpha_t)\mathbb{E}\left(\tau(y_t)|y_t^c = 0\right).$$

Como $\mathbb{E}(y_t^*|y_t^c = 0) = \mu_t^*$, temos que

$$\mathbb{E}(U_{ro}) = \mathbb{E}\left(\sum_{t:y_t\in(0,1)} -\left\{\phi_t^2 v_t^*\left(\frac{1}{g'(\mu_t)}\right)^2\right\}x_{to}x_{tr}\right)$$
$$= -\sum_{t=1}^{n}(1-\alpha_t)\left[\phi_t^2 v_t^*\left(\frac{1}{g'(\mu_t)}\right)^2\right]x_{to}x_{tr}.$$

Adicionalmente, para $p = 1,\ldots,q$ e $l = 1,\ldots,q$,

$$U_{pl} = \frac{\partial^2\ell_2(\beta,\lambda)}{\partial\lambda_p\lambda_l} = \sum_{t:y_t\in(0,1)}\left\{\left(\frac{\partial^2\ell_t(\mu_t,\phi_t)}{\partial\phi_t^2}\left(\frac{d\phi_t}{d\kappa_t}\right)^2\right) + \left(\frac{\partial\ell_t(\mu_t,\phi_t)}{\partial\phi_t}\left(\frac{\partial}{\partial\phi_t}\frac{d\phi_t}{d\kappa_t}\right)\right.\right.$$
$$\left.\left.\times\frac{d\phi_t}{d\kappa_t}\right)\right\}s_{tl}s_{tp}$$
$$= \sum_{t:y_t\in(0,1)} -\left\{\left(\mu_t^2 v_t^* + 2\mu_t c_t + v_t^\dagger\right)\left(\frac{1}{b'(\phi_t)}\right)^2 + \left((y_t^*-\mu_t^*)\mu_t + (y_t^\dagger-\mu_t^\dagger)\right)\right.$$
$$\left.\times\left(\frac{b''(\phi_t)}{(b'(\phi_t))^2}\frac{1}{b'(\phi_t)}\right)\right\}s_{tl}s_{tp}.$$

Como $\mathbb{E}(y_t^\dagger | y_t^c = 0) = \mu_t^\dagger$, segue que

$$\begin{aligned}\mathbb{E}(U_{pl}) &= \mathbb{E}\left(\sum_{t:y_t\in(0,1)} -\left\{ \left(\mu_t^2 v_t^* + 2\mu_t c_t + v_t^\dagger\right)\left(\frac{1}{b'(\phi_t)}\right)^2 \right\} s_{tl} s_{tp} \right) \\ &= -\sum_{t=1}^{n} (1-\alpha_t)\left[\left(\mu_t^2 v_t^* + 2\mu_t c_t + v_t^\dagger\right)\left(\frac{1}{b'(\phi_t)}\right)^2 \right] s_{tl} s_{tp}.\end{aligned}$$

Por fim, para $r = 1,\ldots,m$ e $p = 1,\ldots,q$,

$$\begin{aligned}U_{rp} &= \frac{\partial^2 \ell_2(\beta,\lambda)}{\partial\beta_r \lambda_p} = \sum_{t:y_t\in(0,1)} \frac{\partial}{\partial\phi_t}\left(\frac{\partial \ell_2(\beta,\lambda)}{\partial\beta_r}\right)\frac{d\phi_t}{d\kappa_t}\frac{\partial\kappa_t}{\partial\lambda_p} \\ &= \sum_{t:y_t\in(0,1)} \left\{ (y_t^* - \mu_t^*) - \phi_t\left[\mu_t v_t^* + c_t\right] \right\} \frac{1}{g'(\mu_t)}\frac{1}{b'(\phi_t)} x_{tr} s_{tp}.\end{aligned}$$

Dessa forma,

$$\begin{aligned}\mathbb{E}(U_{rp}) &= \mathbb{E}\left(\sum_{t:y_t\in(0,1)} -\left\{ \phi_t\left[\mu_t v_t^* + c_t\right] \right\} \frac{1}{g'(\mu_t)}\frac{1}{b'(\phi_t)} x_{tr} s_{tp} \right) \\ &= -\sum_{t=1}^{n} (1-\alpha_t)\left[\phi_t\left(\mu_t v_t^* + c_t\right) \right] \frac{1}{g'(\mu_t)}\frac{1}{b'(\phi_t)} x_{tr} s_{tp}.\end{aligned}$$

Dada a separabilidade dos vetores de parâmetros γ e $(\beta^\top, \lambda^\top)$, temos que $U_{Rp} = U_{rR} = 0$.

A matriz de informação observada de Fisher tem a forma

$$J(\theta) = \begin{pmatrix} J_{\gamma\gamma} & 0 & 0 \\ 0 & J_{\beta\beta} & J_{\beta\lambda} \\ 0 & J_{\lambda\beta} & J_{\lambda\lambda} \end{pmatrix},$$

em que

$$\begin{aligned}J_{\gamma\gamma} &= Z^\top \left\{ P\left[Y^c(\mathcal{I} - 2\mathcal{M}^c) + \mathcal{M}^{c2} \right] PG + P(Y^c - \mathcal{M}^c) W G^2 \right\} GZ, \\ J_{\beta\beta} &= X^\top H \left\{ \Phi V^* T + DT^2(Y^* - \mathcal{M}^*) \right\} \Phi T X, \\ J_{\beta\lambda} &= J_{\lambda\beta}{}^\top = -X^\top H \left\{ (Y^* - \mathcal{M}^*) - \Phi(\mathcal{M}V^* + C) \right\} TVS, \\ J_{\lambda\lambda} &= S^\top H \Big\{ V(\mathcal{M}^2 V^* + 2\mathcal{M}C + V^\dagger) + QV^2[(Y^* - V\mathcal{M}^*)\mathcal{M} \\ &\quad + (Y^\dagger - \mathcal{M}^\dagger)] \Big\} VS,\end{aligned}$$

com $Y^c = \text{diag}\{y_1^c,\ldots,y_n^c\}$, $Y^* = \text{diag}\{y_1^*,\ldots,y_n^*\}$, $Y^\dagger = \text{diag}\{y_1^\dagger,\ldots,y_n^\dagger\}$, $\mathcal{M}^c = \text{diag}\{\mu_1^c,\ldots,\mu_n^c\}$, $\mathcal{M}^* = \text{diag}\{\mu_1^*,\ldots,\mu_n^*\}$, $\mathcal{M}^\dagger = \text{diag}\{\mu_1^\dagger,\ldots,\mu_n^\dagger\}$, $V^* = \text{diag}\{v_1^*,\ldots,v_n^*\}$, $V^\dagger = \text{diag}\{v_1^\dagger,\ldots,v_n^\dagger\}$ e $C = \text{diag}\{c_1,\ldots,c_n\}$. Adicionalmente, $W = \text{diag}\{h''(\alpha_1),\ldots,h''(\alpha_n)\}$, $D = \text{diag}\{g''(\mu_1),\ldots,g''(\mu_n)\}$ e $Q = \text{diag}\{b''(\phi_1),\ldots,b''(\phi_n)\}$.

A matriz de informação (esperada) de Fisher pode ser escrita como

$$I(\theta) = \begin{pmatrix} I_{\gamma\gamma} & 0 & 0 \\ 0 & I_{\beta\beta} & I_{\beta\lambda} \\ 0 & I_{\lambda\beta} & I_{\lambda\lambda} \end{pmatrix},$$

em que

$$\begin{aligned} I_{\gamma\gamma} &= Z^\top GPGZ, \\ I_{\beta\beta} &= X^\top \Delta\Phi T V^* T\Phi X, \\ I_{\beta\lambda} &= I_{\lambda\beta}{}^\top = X^\top \Delta\Phi T(\mathcal{M}V^* + C)VS, \\ I_{\lambda\lambda} &= S^\top \Delta V(\mathcal{M}^2 V^* + 2\mathcal{M}C + V^\dagger)VS, \end{aligned}$$

com $\Delta = \text{diag}\{1-\alpha_1,\ldots,1-\alpha_n\}$.

4.3 Testes da razão de verossimilhanças modificados

Sejam $y_1,\ldots,y_n$ variáveis aleatórias independentes, em que cada y_t, $t = 1,\ldots,n$, segue a função de densidade (4.2.1). Seja $\theta = (\gamma^\top,\beta^\top,\lambda^\top)^\top$ o vetor de parâmetros desconhecidos que indexam o modelo de regressão beta inflacionado em 0 ou 1. Considere que o vetor de parâmetros θ é escrito como $\theta = (\nu^\top,\tau^\top)^\top$, em que $\nu = (\nu_1,\ldots,\nu_r)^\top$ representa o vetor de parâmetros de interesse e $\tau = (\tau_1,\ldots,\tau_s)^\top$ representa o vetor de parâmetros de perturbação, com $M+m+q = r+s$. O interesse reside em testar a hipótese nula $\mathcal{H}_0 : \nu = \nu_0$, em que ν_0 é um vetor de constantes especificado de dimensão r.

A estatística da razão de verossimilhanças para testar $\mathcal{H}_0$ pode ser escrita como

$$LR = 2\{\ell(\hat{\theta}) - \ell(\tilde{\theta})\}, \tag{4.3.1}$$

em que $\hat{\theta} = (\hat{\nu}^\top,\hat{\tau}^\top)^\top$ é o estimador de máxima verossimilhança irrestrito de θ e $\tilde{\theta} = (\nu_0^\top,\tilde{\tau}^\top)^\top$ é o estimador de máxima verossimilhança de θ obtido pela imposição da hipótese nula, ou seja, o estimador de máxima verossimilhança restrito. Sob condições usuais de regularidade e sob $\mathcal{H}_0$ a distribuição da estatística da razão de verossimilhanças pode ser aproximada pela distribuição χ^2_r com erro de ordem n^{-1}. Contudo, este teste pode ter tamanho distorcido se o tamanho da amostra não for suficientemente grande para garantir uma boa aproximação da distribuição nula da estatística da razão de verossimilhanças pela distribuição χ^2. Uma outra alternativa para testar $\mathcal{H}_0$, no caso do parâmetro de interesse ser escalar, é a estatística da razão de verossimilhanças sinalizada dada por

$$R = \text{sign}(\hat{\nu} - \nu_0)\sqrt{LR}. \tag{4.3.2}$$

Sob a hipótese nula, a distribuição da estatística da razão de verossimilhanças sinalizada pode ser aproximada pela distribuição $\mathcal{N}(0,1)$ com erro de ordem $n^{-1/2}$. A vantagem desse teste é que a hipótese alternativa pode ser unilateral ou bilateral.

Para melhorar a aproximação da distribuição da estatística pela distribuição normal padrão, Barndorff-Nielsen (1986, 1991) propôs a estatística da razão de verossimilhanças sinalizada modificada dada por

$$R1 = R - \frac{1}{R}\log\varepsilon. \tag{4.3.3}$$

A expressão geral para a quantidade ε é

$$\varepsilon = |\hat{J}|^{1/2}|\tilde{U}'|^{-1}|\tilde{J}_{\tau\tau}|^{1/2}\frac{R}{[(\hat{\ell}' - \tilde{\ell}')^\top(\tilde{U}')^{-1}]_\nu}, \tag{4.3.4}$$

em que J é a informação observada de Fisher e U é a função escore total. Aqui, 'chapéu' denota avaliação no estimador de máxima verossimilhança irrestrito e 'til' denota avaliação no estimador de máxima verossimilhança restrito. Adicionalmente, $J_{\tau\tau}$ denota a matriz de informação observada $s \times s$ correspondente ao vetor τ e $[(\hat{\ell}' - \tilde{\ell}')^\top(\tilde{U}')^{-1}]_\nu$ é o elemento do vetor $(\hat{\ell}' - \tilde{\ell}')^\top(\tilde{U}')^{-1}$ correspondente ao parâmetro ν, que nesse caso é escalar. Sob a hipótese nula, $R1$ segue distribuição normal padrão com um erro de ordem $n^{-3/2}$. A expressão (4.3.4) envolve derivadas com respeito ao espaço amostral em $\tilde{U}'$ e $\hat{\ell}' - \tilde{\ell}'$. Uma vez que derivadas relativas ao espaço amostral podem ser difíceis de serem obtidas, Skovgaard (1996) forneceu aproximações explícitas para essas quantidades

$$(\hat{\ell}' - \tilde{\ell}')^\top \approx \hat{q}\hat{I}^{-1}\hat{J} \quad \text{e} \quad \tilde{U}' \approx \hat{\Upsilon}\hat{I}^{-1}\hat{J},$$

em que I é a informação (esperada) de Fisher. As quantidades $\hat{q}$ e $\hat{\Upsilon}$ são obtidas, respectivamente, de

$$q = \mathbb{E}_{\theta_1}[U(\theta_1)(\ell(\theta_1) - \ell(\theta))] \quad \text{e} \quad \Upsilon = \mathbb{E}_{\theta_1}[U(\theta_1)U^\top(\theta)] \tag{4.3.5}$$

substituindo $\hat{\theta}$ por θ_1 e $\tilde{\theta}$ por θ após os valores esperados serem computados. Dessa forma, uma aproximação geral para a quantidade ε dada em (4.3.4) é

$$\bar{\varepsilon} = |\hat{J}|^{-1/2}|\hat{I}||\hat{\Upsilon}|^{-1}|\tilde{J}_{\tau\tau}|^{1/2}\frac{R}{[\hat{\Upsilon}^{-1}\hat{q}]_\nu}. \tag{4.3.6}$$

Assim, a estatística da razão de verossimilhanças sinalizada modificada dada em (4.3.3) pode ser escrita como

$$R1 = R - \frac{1}{R}\log\bar{\varepsilon}. \tag{4.3.7}$$

Estudos numéricos realizados por Bellio e Brazzale (1999) apontaram para uma melhor aproximação da distribuição nula da estatística $R1$ pela distribuição normal relativamente à estatística R.

A estatística $R1$ foi desenvolvida para testar hipóteses em que o parâmetro de interesse é unidimensional. Frequentemente é desejável testar hipóteses envolvendo vários parâmetros. Skovgaard (2001) propôs uma generalização da estatística de Barndorff-Nielsen para o caso em que o parâmetro de interesse é multidimensional. Considerando a hipótese nula $\mathcal{H}_0 : \nu = \nu_0$,

em que ν e ν_0 são vetores, a estatística da razão de verossimilhanças modificada é dada por (Skovgaard, 2001)

$$LR1 = LR\left(1 - \frac{1}{LR}\log\bar{\xi}\right)^2, \tag{4.3.8}$$

em que

$$\bar{\xi} = |\tilde{I}|^{1/2}|\hat{I}|^{1/2}|\hat{\Upsilon}|^{-1}|\tilde{J}_{\tau\tau}|^{1/2}|[\tilde{I}\hat{\Upsilon}^{-1}\hat{J}\hat{I}^{-1}\hat{\Upsilon}]_{\tau\tau}|^{-1/2}\frac{\{\tilde{U}^\top\hat{\Upsilon}^{-1}\hat{I}\hat{J}^{-1}\hat{\Upsilon}\tilde{I}^{-1}\tilde{U}\}^{r/2}}{LR^{r/2-1}\,\tilde{U}^\top\hat{\Upsilon}^{-1}\hat{q}}.$$

Uma estatística assintoticamente equivalente a $LR1$ é dada por

$$LR2 = LR - 2\log\bar{\xi}. \tag{4.3.9}$$

Sob $\mathcal{H}_0$, as estatísticas $LR1$ e $LR2$ são aproximadamente distribuídas como χ^2_r com alto grau de precisão. A versão $LR1$ tem a vantagem de ser sempre não-negativa e de reduzir-se ao quadrado da estatística de Barndorff-Nielsen ($R1$) quando $r = 1$. Skovgaard (2001) mostrou, através de estudos de simulação, que testes baseados nas estatísticas $LR1$ e $LR2$ têm desempenhos superiores ao teste baseado na estatística LR e que a estatística $LR2$ possui desempenho um pouco superior ao de $LR1$ em alguns casos.

Nós derivamos, para a classe de modelos de regressão beta inflacionados, estatísticas ajustadas da razão de verossimilhanças e da razão de verossimilhanças sinalizada de forma a obter aproximações mais precisas das distribuições nulas das estatísticas pela distribuição χ^2, mesmo em amostras de tamanho pequeno ou moderado. As quantidades não usuais q e Υ relativas ao vetor escore e ao logaritmo da função de verossimilhanças são apresentadas a seguir. Uma derivação detalhada dos resultados é apresentada no Apêndice A. Da primeira equação em (4.3.5), temos que

$$q = \begin{bmatrix} \mathbb{E}_{\theta_1}[U_\gamma(\theta_1)\ell(\theta_1)] - \mathbb{E}_{\theta_1}[U_\gamma(\theta_1)\ell(\theta)] \\ \mathbb{E}_{\theta_1}[U_\beta(\theta_1)\ell(\theta_1)] - \mathbb{E}_{\theta_1}[U_\beta(\theta_1)\ell(\theta)] \\ \mathbb{E}_{\theta_1}[U_\lambda(\theta_1)\ell(\theta_1)] - \mathbb{E}_{\theta_1}[U_\lambda(\theta_1)\ell(\theta)] \end{bmatrix}.$$

Nós obtemos

$$\mathbb{E}_\theta[U_\gamma(\theta)\ell(\theta)] = Z^\top PGV^c\alpha^*\iota.$$

Uma vez que $\mathbb{E}_{\theta_1}[(y_t^* - \mu_t^{*(1)})(y_t^* - \mu_t^*)] = v_t^{*(1)}$, sobrescrito (1) indicando avaliação em θ_1, temos que

$$\mathbb{E}_{\theta_1}[U_\gamma(\theta_1)\ell(\theta)] = Z^\top P^{(1)}G^{(1)}V^{c(1)}\alpha^*\iota$$

e, assim,

$$\mathbb{E}_{\theta_1}[U_\gamma(\theta_1)\ell(\theta_1)] - \mathbb{E}_{\theta_1}[U_\gamma(\theta_1)\ell(\theta)] = Z^\top P^{(1)}G^{(1)}V^{c(1)}\{\alpha^{*(1)} - \alpha^*\}\iota.$$

Adicionalmente,

$$\mathbb{E}_\theta[U_\beta(\theta)\ell(\theta)] = X^\top\Phi\Delta T\{V^*(\Phi\mathcal{M} - \mathcal{I}) + C(\Phi - 2\mathcal{I})\}\iota.$$

Portanto,

$$\begin{aligned}\mathbb{E}_{\theta_1}[U_\beta(\theta_1)\ell(\theta_1)] - \mathbb{E}_{\theta_1}[U_\beta(\theta_1)\ell(\theta)] = X^\top\Phi^{(1)}\Delta^{(1)}T^{(1)}\{V^{*(1)}(\Phi^{(1)}\mathcal{M}^{(1)} - \Phi\mathcal{M}) \\ + C^{(1)}(\Phi^{(1)} - \Phi)\}\iota.\end{aligned}$$

Por fim,

$$\mathbb{E}_\theta[U_\lambda(\theta)\ell(\theta)] = S^\top V\Delta\{(V^*\mathcal{M}+C)(\Phi\mathcal{M}-\mathcal{I})+(V^\dagger+C\mathcal{M})(\Phi-2\mathcal{I})\}\iota.$$

Segue então que

$$\begin{aligned}\mathbb{E}_\theta[U_\lambda(\theta_1)\ell(\theta_1)] - \mathbb{E}_\theta[U_\lambda(\theta_1)\ell(\theta)] = S^\top V^{(1)}\Delta^{(1)}\{(V^{*(1)}\mathcal{M}^{(1)}+C^{(1)})(\Phi^{(1)}\mathcal{M}^{(1)} \\ -\Phi\mathcal{M})+(V^{\dagger(1)}+C^{(1)}\mathcal{M}^{(1)})(\Phi^{(1)}-\Phi)\}\iota.\end{aligned}$$

Da segunda equação em (4.3.5), temos

$$\Upsilon = \begin{bmatrix} \mathbb{E}_{\theta_1}[U_\gamma(\theta_1)U_\gamma^\top(\theta)] & \mathbb{E}_{\theta_1}[U_\gamma(\theta_1)U_\beta^\top(\theta)] & \mathbb{E}_{\theta_1}[U_\gamma(\theta_1)U_\lambda^\top(\theta)] \\ \mathbb{E}_{\theta_1}[U_\beta(\theta_1)U_\gamma^\top(\theta)] & \mathbb{E}_{\theta_1}[U_\beta(\theta_1)U_\beta^\top(\theta)] & \mathbb{E}_{\theta_1}[U_\beta(\theta_1)U_\lambda^\top(\theta)] \\ \mathbb{E}_{\theta_1}[U_\lambda(\theta_1)U_\gamma^\top(\theta)] & \mathbb{E}_{\theta_1}[U_\lambda(\theta_1)U_\beta^\top(\theta)] & \mathbb{E}_{\theta_1}[U_\lambda(\theta_1)U_\lambda^\top(\theta)] \end{bmatrix}.$$

Assim, nós obtemos

$$\bar{q} = \begin{bmatrix} Z^\top\hat{P}\hat{G}\hat{V}^c\{\hat{\alpha}^*-\tilde{\alpha}^*\}\iota \\ X^\top\hat{\Phi}\hat{\Delta}\hat{T}\{\hat{V}^*(\hat{\Phi}\hat{\mathcal{M}}-\tilde{\Phi}\tilde{\mathcal{M}})+\hat{C}(\hat{\Phi}-\tilde{\Phi})\}\iota \\ S^\top\hat{V}\hat{\Delta}\{(\hat{V}^*\hat{\mathcal{M}}+\hat{C})(\hat{\Phi}\hat{\mathcal{M}}-\tilde{\Phi}\tilde{\mathcal{M}})+(\hat{V}^\dagger+\hat{C}\hat{\mathcal{M}})(\hat{\Phi}-\tilde{\Phi})\}\iota \end{bmatrix}$$

e

$$\bar{\Upsilon} = \begin{bmatrix} Z^\top\hat{P}\hat{G}\hat{V}^c\tilde{G}\tilde{P}Z & 0 & 0 \\ 0 & X^\top\hat{\Phi}\hat{T}\hat{V}^*\hat{\Delta}\tilde{T}\tilde{\Phi}X & X^\top\hat{\Phi}\hat{T}\hat{\Delta}(\hat{V}^*\tilde{\mathcal{M}}+\hat{C})\tilde{V}S \\ 0 & S^\top\hat{V}\hat{\Delta}(\hat{V}^*\hat{\mathcal{M}}+\hat{C})\tilde{T}\tilde{\Phi}X & S^\top\hat{V}\hat{\Delta}\{\hat{V}^*\hat{\mathcal{M}}\tilde{\mathcal{M}}+\hat{C}(\hat{\mathcal{M}}+\tilde{\mathcal{M}})+\hat{V}^\dagger\}\tilde{V}S \end{bmatrix}$$

(matrizes necessárias para a obtenção das estatísticas $LR1$, $LR2$ e $R1$ na classe de modelos de regressão beta inflacionados).

4.4 Avaliação numérica

Nosso objetivo nesta seção é comparar, através de simulações de Monte Carlo, os desempenhos de testes baseados na estatística da razão de verossimilhanças usual e sinalizada em modelos de regressão beta inflacionados. Para a estatística da razão de verossimilhanças (LR) e suas versões corrigidas ($LR1$ e $LR2$) os testes realizados foram bilaterais. No caso da estatística da razão de verossimilhanças sinalizada (R) e sua versão corrigida ($R1$) os testes realizados foram unilaterais. Primeiramente, testes foram realizados sobre os parâmetros do submodelo de regressão da média.

O modelo beta inflacionado em um usado é

$$\begin{aligned} h(\alpha_t) &= \gamma_0+\gamma_1 z_{t1}, \\ g(\mu_t) &= \beta_0+\beta_1 x_{t1}+\beta_2 x_{t2}, \\ b(\phi_t) &= \lambda_0+\lambda_1 s_{t1}. \end{aligned} \tag{4.4.1}$$

Tabela 4.1 Tamanhos dos testes (%), submodelo de μ.

Teste bilateral										
r	Estatística	$n = 30$			$n = 40$			$n = 50$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
1	LR	2.50	9.24	15.82	2.60	8.98	15.44	1.76	7.28	13.76
	$LR1$	1.40	5.78	11.22	1.30	5.68	11.42	1.28	5.30	10.90
	$LR2$	1.16	5.18	10.54	0.96	5.06	10.80	1.04	4.98	10.42
2	LR	2.86	9.50	16.74	2.48	8.94	15.58	2.28	8.16	14.04
	$LR1$	1.12	5.30	10.48	0.90	5.52	10.24	1.22	5.56	10.60
	$LR2$	1.04	4.96	10.24	0.84	5.30	10.06	1.22	5.48	10.46
Teste unilateral										
r	Estatística	$n = 30$			$n = 40$			$n = 50$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
1	R	2.12	8.08	13.74	2.18	7.70	12.40	1.74	6.50	11.46
	$R1$	1.22	5.66	10.78	1.28	5.60	10.16	1.26	5.30	10.10

Os valores das covariáveis foram selecionados aleatoriamente da distribuição $\mathcal{U}(0,1)$. Todas as simulações foram baseadas em 5,000 réplicas e nos tamanhos amostrais 30, 40 e 50. Em cada réplica foi gerada uma amostra aleatória da variável dependente $y = (y_1, \ldots, y_n)^\top$, com y_t seguindo distribuição beta inflacionada em um, $t = 1, \ldots, n$. Nós consideramos duas hipóteses nulas diferentes: $\mathcal{H}_0 : \beta_2 = 0$ ($r = 1$) e $\mathcal{H}_0 : \beta_1 = \beta_2 = 0$ ($r = 2$). No primeiro caso ($r = 1$), os verdadeiros valores dos parâmetros são $\gamma_0 = -2.0$, $\gamma_1 = 1.5$, $\beta_0 = -1.0$, $\beta_1 = 3.5$, $\beta_2 = 0$, $\lambda_0 = 5.1$ e $\lambda_1 = -2.8$. Para o segundo caso ($r = 2$), $\beta_0 = 2.0$ e $\beta_1 = \beta_2 = 0$. Os resultados foram obtidos usando a linguagem de programação R (ver `http://www.R-project.org`).

A Tabela 4.1 contém as taxas de rejeição (%) da hipótese nula dos diferentes testes quando essa hipótese é verdadeira. O teste da razão de verossimilhanças original é notavelmente liberal ao passo que os testes corrigidos apresentam bom desempenho. Por exemplo, para $n = 30$, $r = 1$ e ao nível de significância de 5% as taxas de rejeição da hipótese nula dos testes baseados em LR, $LR1$ e $LR2$ são, respectivamente, 9.24%, 5.78% e 5.18%. Ou seja, a taxa de rejeição da hipótese nula do teste LR é quase o dobro do nível nominal selecionado. O impacto do número de parâmetros de interesse é maior sobre o desempenho do teste da razão de verossimilhanças do que sobre os de suas versões corrigidas. O teste baseado em $LR2$ apresentou os tamanhos estimados mais próximos dos níveis nominais. Considerando os testes unilaterais, a versão corrigida do teste da razão de verossimilhanças sinalizada ($R1$) apresentou melhor desempenho no que se refere ao tamanho estimado. Por exemplo, para $n = 30$ e ao nível de significância de 5% a taxa de rejeição da hipótese nula é 8.08% para o teste baseados em R e 5.66% para o teste baseado em $R1$.

A Tabela 4.2 apresenta as taxas de rejeição dos testes sob as hipóteses alternativas $\mathcal{H}_1 : \beta_2 = \delta$ ($r = 1$) e $\mathcal{H}_1 : \beta_1 = \beta_2 = \delta$ ($r = 2$). Os resultados foram obtidos para $n = 30, 50$, $\delta = -1.0, -0.5, 0.5, 1.0$ e as estatísticas $LR1$ e $LR2$. Por ser consideravelmente liberal, a estatística LR não foi incluída na análise. É possível notar que ambos os testes corrigidos apresentam

Tabela 4.2 Poder dos testes da razão de verossimilhanças ajustados (%), submodelo de μ.

Teste bilateral						
r	δ	Estatística	$n = 30$		$n = 50$	
			5%	10%	5%	10%
1	−1.0	$LR1$	98.56	99.44	99.92	99.96
		$LR2$	98.56	99.44	99.84	99.96
	−0.5	$LR1$	61.74	72.60	74.98	84.12
		$LR2$	62.12	73.16	76.00	84.62
	0.5	$LR1$	55.08	67.38	66.72	77.10
		$LR2$	55.74	68.08	68.12	77.50
	1.0	$LR1$	96.88	98.64	99.08	99.66
		$LR2$	96.94	98.68	99.12	99.68
2	−1.0	$LR1$	99.76	99.94	100.00	100.00
		$LR2$	99.74	99.94	100.00	100.00
	−0.5	$LR1$	66.90	78.40	93.14	96.64
		$LR2$	66.82	78.24	93.08	96.62
	0.5	$LR1$	36.20	49.20	66.94	78.00
		$LR2$	36.22	49.30	66.20	78.06
	1.0	$LR1$	78.60	87.24	98.10	99.18
		$LR2$	78.56	87.32	98.10	99.18

desempenhos similares no que tange ao poder, o teste baseado em $LR2$ sendo levemente mais poderoso.

Considerando o teste unilateral a Tabela 4.3 apresenta as taxas de rejeição sob a hipótese alternativa $\mathcal{H}_1 : \beta_2 > \delta$. Nós não incluímos a estatística R nas comparações de poder uma vez que ela é consideravelmente liberal e não deve, portanto, ser recomendada. Notamos que o teste da razão de verossimilhanças sinalizado ajustado apresenta bom desempenho no que diz respeito ao poder. Por exemplo, para $n = 30$, $\delta = 1.0$ e ao nível nominal de 5% o poder do teste é 98.60%.

Nós consideramos agora testes sobre os parâmetros do submodelo de regressão de ϕ. O

Tabela 4.3 Poder do teste da razão de verossimilhanças sinalizado ajustado (%), submodelo de μ.

Teste unilateral					
δ	Estatística	$n = 30$		$n = 50$	
		5%	10%	5%	10%
3.0	$R1$	100.00	100.00	100.00	100.00
2.0		100.00	100.00	100.00	100.00
1.0		98.60	99.56	99.66	99.92
0.5		67.18	80.62	77.18	88.04

Tabela 4.4 Tamanhos dos testes (%), submodelo de ϕ.

Teste bilateral										
r	Estatística	$n = 30$			$n = 40$			$n = 50$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
1	LR	2.24	8.22	14.78	1.82	6.32	12.62	1.74	6.66	12.82
	$LR1$	2.92	7.94	13.56	1.80	5.86	11.48	1.48	5.46	11.22
	$LR2$	1.56	6.30	11.62	1.34	5.30	10.68	1.16	5.04	10.62
2	LR	2.92	10.10	16.50	2.40	8.34	14.54	1.88	7.42	14.28
	$LR1$	1.28	5.76	10.82	1.34	5.52	10.42	1.10	5.04	10.50
	$LR2$	1.22	5.58	10.18	1.30	5.38	10.26	1.04	4.94	10.30
Teste unilateral										
r	Estatística	$n = 30$			$n = 40$			$n = 50$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
1	R	2.20	7.92	13.66	1.82	7.28	13.10	1.54	6.36	12.26
	$R1$	2.06	6.76	12.10	1.28	5.26	10.14	1.12	5.04	10.52

modelo beta inflacionado em zero ou um usado é

$$\begin{aligned} h(\alpha_t) &= \gamma_0 + \gamma_1 z_{t1}, \\ g(\mu_t) &= \beta_0 + \beta_1 x_{t1}, \\ b(\phi_t) &= \lambda_0 + \lambda_1 s_{t1} + \lambda_2 s_{t2}. \end{aligned} \tag{4.4.2}$$

Para $r = 1$, os verdadeiros valores dos parâmetros são $\gamma_0 = -2.0$, $\gamma_1 = 1.5$, $\beta_0 = -1.0$, $\beta_1 = 3.5$, $\lambda_0 = 5.1$, $\lambda_1 = -2.8$ e $\lambda_2 = 0$. Para o segundo caso ($r = 2$), fixamos $\lambda_0 = 5.1$, $\lambda_1 = \lambda_2 = 0$. Note que quando $r = 2$ a hipótese nula em teste é a de dispersão constante.

Os resultados apresentados na Tabela 4.4 mostram que o teste $LR2$ é o que apresenta o melhor desempenho. Mesmo para $n = 50$ o teste da razão de verossimilhanças original se mostrou liberal. Por exemplo, para $n = 50$, $r = 2$ e ao nível de significância de 10%, as taxas de rejeição da hipótese nula dos testes baseados em LR, $LR1$ e $LR2$ são, respectivamente, 14.28%, 10.50% e 10.30%. No que se refere aos testes unilaterais é possível observar que o desempendo do teste basedo na estatística da razão de verossimilhanças sinalizada modificada é melhor do que o de sua versão original. As taxas de rejeição dos testes se aproximam dos níveis nominais à medida que o tamanho da amostra aumenta, como esperado. Por exemplo, ao nível de 5% de significância os tamanhos estimados dos testes baseados em R e em $R1$ são, respectivamente, 7.92% e 6.76% para $n = 30$ e 6.36% e 5.04% para $n = 50$.

A Tabela 4.5 contém as taxas de rejeição sob as hipóteses alternativas $\mathcal{H}_1 : \lambda_2 = \delta$ ($r = 1$) e $\mathcal{H}_1 : \lambda_1 = \lambda_2 = \delta$ ($r = 2$). Os resultados foram obtidos para $n = 30, 50$, $\delta = -4.0, -3.0, 3.0, 4.0$ e as estatísticas $LR1$ e $LR2$. Comparado ao teste baseado em $LR1$, para $r = 1$, o teste baseado em $LR2$ apresentou poderes um pouco superiores. Por exemplo, para $n = 30$, $r = 1$, $\delta = -3$ e ao nível de significância de 5% os poderes dos testes $LR1$ e $LR2$ foram, respectivamente, 66.82% e 73.72%. Todavia, quando $r = 2$, os resultados da tabela sugerem que ambos os testes são igualmente poderosos.

Tabela 4.5 Poder dos testes da razão de verossimilhanças ajustados (%), submodelo de ϕ.

Teste bilateral						
r	δ	Estatística	$n = 30$		$n = 50$	
			5%	10%	5%	10%
1	−4.0	$LR1$	88.00	94.38	98.32	99.26
		$LR2$	91.12	95.08	98.44	99.28
	−3.0	$LR1$	66.82	80.58	89.04	93.58
		$LR2$	73.72	82.34	89.82	93.82
	3.0	$LR1$	52.54	69.68	79.40	86.96
		$LR2$	60.00	71.42	80.34	87.32
	4.0	$LR1$	77.16	88.28	95.10	97.52
		$LR2$	82.10	89.04	95.38	97.52
2	−4.0	$LR1$	97.36	98.54	99.98	99.98
		$LR2$	97.32	98.56	99.98	99.98
	−3.0	$LR1$	85.60	91.34	98.68	99.32
		$LR2$	85.18	90.98	98.70	99.32
	3.0	$LR1$	73.42	83.28	95.74	97.50
		$LR2$	72.76	82.42	95.66	94.42
	4.0	$LR1$	90.44	94.48	99.64	99.84
		$LR2$	89.62	93.78	99.64	99.84

Considerando o teste unilateral a Tabela 4.6 apresenta as taxas de rejeição sob a hipótese alternativa $\mathcal{H}_1 : \lambda_2 > \delta$. O teste da razão de verossimilhanças sinalizado ajustado apresenta bons poderes. Note ainda que os poderes do teste crescem quando aumentamos o tamanho da amostra e o valor de δ, como esperado. Por exemplo, ao nível de significância de 5%, os poderes do teste $R1$ para $n = 30$ foram, respectivamente, 42.16% para $\delta = 2$ e 88.42% para $\delta = 4$.

Testes sobre os parâmetros do submodelo de regressão de α também foram avaliados, mas não apresentaram bom desempenho em pequenas amostras. Por isso, esses resultados não serão apresentados.

Tabela 4.6 Poder do teste da razão de verossimilhanças sinalizado ajustado (%), submodelo de ϕ.

Teste unilateral					
δ	Estatística	$n = 30$		$n = 50$	
		5%	10%	5%	10%
4.0	$R1$	88.42	93.66	97.70	98.96
3.0		69.74	81.92	87.96	93.48
2.0		42.16	57.86	62.22	73.64
1.0		17.48	30.22	25.96	37.74

4.5 Conclusões

Este capítulo focou na realização de inferências via testes de hipóteses no modelo de regressão beta inflacionado. Em particular, nós focamos no teste da razão de verossimilhanças que usualmente apresenta tamanhos distorcidos em amostras de tamanho pequeno a moderado. Nós obtivemos ajustes para as estatísticas da razão de verossimilhanças e da razão de verossimilhanças sinalizada usando o resultado de Skovgaard (2001). As estatísticas ajustadas podem ser usadas para testar hipóteses nulas que envolvem os parâmetros que definem a média e/ou o parâmetro de precisão. Resultados de simulação foram obtidos para o modelo de regressão beta inflacionado e conduziram a conclusões importantes relativas ao comportamento dos testes em amostras de tamanho pequeno e moderado. O teste da razão de verossimilhanças usual e sua versão sinalizada não são recomendados, uma vez que eles podem ser consideravelmente liberais em amostras pequenas. Os testes ajustados se comportaram de forma mais precisa. Nós recomendamos o uso do teste da razão de verossimilhanças sinalizada corrigido, $R1$, bem como os testes da razão de verossimilhanças ajustados, em particular o teste baseado na estatística $LR2$.

CAPÍTULO 5

Considerações Finais

5.1 Conclusões

Ao longo deste trabalho desenvolvemos alguns aspectos de inferência e ilustramos a aplicabilidade prática da regressão beta inflacionada na modelagem de dados de proporções, taxas ou frações na presença de zeros ou uns. As principais contribuições e conclusões deste trabalho encontram-se resumidas a seguir:

1. No Capítulo 2, avaliamos a eficiência do gasto público municipal no estado de São Paulo utilizando modelos de regressão para lidar com dados em forma de frações, taxas ou proporções correlacionados espacialmente e medidos continuamente no intervalo aberto $(0,1)$, mas que podem conter zeros e/ou uns. A análise foi focada na comparação dos modelos de regressão beta e beta inflacionado. Adicionalmente, uma comparação com os modelos de regressão linear e com inferência realizada via quasi-verossimilhança foi apresentada. As avaliações foram baseadas nos gráficos dos resíduos, pseudo R^2, medidas descritivas e construção de mapas coropléticos. A análise conduziu a conclusões importantes. O modelo de regressão beta inflacionado foi o mais adequado para explicar os escores de eficiência média dos municípios. A variável que mais influencia a eficiência administrativa de municipalidades paulistas é a que indica se os municípios são gerenciados pelo partido político PT. A variável que introduz a dependência espacial e a taxa de urbanização também exercem efeito positivo sobre a eficiência, ao passo que o rendimento médio dos trabalhadores exerce efeito negativo. Uma breve avaliação do gasto público municipal no Brasil e nos demais estados do Brasil excluindo o estado de São Paulo foi apresentada. Para o Brasil, as capitais dos estados tendem a ser mais eficientes e os municípios novos tendem a ser menos eficientes. Ao contrário do que se esperava, municípios que recebem *royalties* tendem a ser menos eficientes, tanto para o Brasil como para os demais estados. É importante citar que esta receita adicional não deveria gerar gastos indevidos e despesas ineficientes. O recebimento de *royalties* deveria ser canalizado em prol de uma melhor qualidade de vida medida pelo acesso aos serviços públicos.

2. No Capítulo 3, nós desenvolvemos, baseado no teste RESET (Ramsey, 1969), um teste de erro de especificação para modelos de regressão beta inflacionados tanto sob dispersão fixa quanto sob dispersão variável. Em particular, nós propomos duas variantes do teste. Na primeira variante, nós adicionamos variáveis de teste apenas ao submodelo da média condicional. A segunda variante segue da adição de variáveis de teste aos submodelos da

média condicional, da probabilidade de que a variável é igual a zero ou um e do parâmetro de precisão. Nós realizamos simulações de Monte Carlo para avaliar os desempenhos em amostras finitas dos testes (tamanho e poder) e também para decidir quais variáveis devem ser usadas como variáveis de teste e qual procedimento de teste assintótico conduz às inferências mais confiáveis. Nós consideramos vários tipos de erros de especificação: não-linearidade negligenciada, função de ligação incorreta, variáveis omitidas, correlação espacial negligenciada e dispersão variável não modelada. Por fim, uma aplicação empírica que emprega dados reais foi apresentada e discutida. Os resultados mostraram que o teste proposto apresenta bons poderes, exceto quando a correlação espacial é negligenciada, neste caso tamanhos amostrais maiores são necessários para obter menores probabilidades do erro tipo II. Evidências numéricas mostraram que o mais simples dos dois testes, isto é, quando apenas o submodelo da média é aumentado, apresenta tipicamente o melhor desempenho. Este teste é realizado usando o quadrado do preditor linear estimado como variável de teste.

3. No Capítulo 4, baseados na proposta de Skovgaard (2001), derivamos ajustes para as estatísticas da razão de verossimilhanças usual e sinalizada em modelos de regressão beta inflacionados. As estatísticas ajustadas podem ser usadas para testar hipóteses nulas que envolvem os parâmetros que definem a média e/ou o parâmetro de precisão. Uma avaliação dos testes corrigidos foi realizada via simulação de Monte Carlo. Resultados numéricos conduziram a conclusões importantes relativas ao comportamento dos testes em amostras de tamanho pequeno e moderado. O teste da razão de verossimilhanças usual e sua versão sinalizada não são recomendados, uma vez que podem ser consideravelmente liberais em amostras pequenas. Os testes ajustados se comportaram de forma mais precisa. Assim, comparadas com as estatísticas da razão de verossimilhanças usual e sinalizada, inferências baseadas nas estatísticas ajustadas que nós propomos são muito mais confiáveis e efetivas.

É possível verificar que as conclusões obtidas são bem gerais. Primeiro, com base nos resultados obtidos a partir do ajuste do modelo de regressão beta inflacionado aos dados de eficiência administrativa é possível fornecer uma visão geral de como os municípios têm-se comportado com relação ao gerenciamento de recursos públicos, avaliar os desempenhos de administrações municipais e fornecer instrumentos que podem ser usados para avaliar os governos locais. Segundo, o teste de erro de especificação proposto pode ser bastante útil em situações práticas, uma vez que é capaz de detectar várias formas de erro de especificação em modelos de regressão beta inflacionados. Terceiro, as correções propostas para os testes baseados na função de verossimilhança podem ser utilizadas para desenvolver testes mais confiáveis na classe de modelos de regressão beta inflacionados, no que diz respeito à aproximação distribucional assintótica mesmo em pequenas amostras.

5.2 Trabalhos Futuros

Algumas linhas de pesquisa devem ser desenvolvidas. Por exemplo, serão foco de nossas pesquisas futuras:

1. Desenvolver testes de especificação para os modelos de regressão beta inflacionados baseado no teste do arco-íris (‘rainbow test’) e comparar o desempenho do teste com o desempenho do teste de especificação proposto neste trabalho.

2. Desenvolver modelos beta para dados geoestatísticos a fim de lidar com dados distribuídos continuamente no intervalo $(0, 1)$ e que são correlacionados espacialmente. A geoestatística é um ramo da estatística espacial no qual os dados consistem de uma amostra finita de valores medidos relacionados a um fenômeno contínuo espacial subjacente. Usualmente, esse tipo de dado é resultante de levantamento de recursos naturais e incluem mapas geológicos, topográfico, ecológicos, fitogeográficos e pedológicos.

Apêndice A

Obtenção das quantidades propostas por Skovgaard para a classe de modelos de regressão beta inflacionados

Neste apêndice, apresentamos os detalhes necessários para a obtenção das quantidades $\hat{q}$ e $\hat{\Upsilon}$. Das equações (4.2.2) e (4.2.3), é possível obter

$$\begin{aligned}
\mathbb{E}_\theta[U_\gamma(\theta)\ell(\theta)] &= \mathbb{E}_\theta[Z^\top PG(y^c-\mu^c)\{(y^c-\mu^c)^\top\alpha^* + a^\top + [(y^*-\mu^*)^\top \\
&\quad \times(\Phi\mathcal{M}-\mathcal{I}) + (y^\dagger-\mu^\dagger)^\top(\Phi-2\mathcal{I}) + b^\top]H\}\iota] \\
&= Z^\top PG\{\mathbb{E}_\theta[(y^c-\mu^c)(y^c-\mu^c)^\top]\alpha^* + \mathbb{E}_\theta[(y^c-\mu^c)]a^\top \\
&\quad + \mathbb{E}_\theta[(y^c-\mu^c)(y^*-\mu^*)^\top H](\Phi\mathcal{M}-\mathcal{I}) \\
&\quad + \mathbb{E}_\theta[(y^c-\mu^c)(y^\dagger-\mu^\dagger)^\top H](\Phi-2\mathcal{I}) \\
&\quad + \mathbb{E}_\theta[H(y^c-\mu^c)]b^\top\}\iota.
\end{aligned}$$

Como $\mathbb{E}_\theta[(y_t^*-\mu_t^*)|y_t^c=0] = \mathbb{E}_\theta[(y_t^\dagger-\mu_t^\dagger)|y_t^c=0] = 0$ e $\mathbb{E}_\theta[(y_t^c-\mu_t^c)(y_t^c-\mu_t^c)] = v_t^c$, segue que

$$\begin{aligned}
\mathbb{E}_\theta[(y_t^c-\mu_t^c)(y_t^*-\mu_t^*)(1-y_t^c)] &= \mathbb{E}_\theta[(y_t^c-\mu_t^c)(y_t^*-\mu_t^*)(1-y_t^c)|y_t^c=0] \\
&\quad \times(1-\alpha_t) + \mathbb{E}_\theta[(y_t^c-\mu_t^c)(y_t^*-\mu_t^*) \\
&\quad \times(1-y_t^c)|y_t^c=1]\alpha_t = 0,
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}_\theta[(y_t^c-\mu_t^c)(y_t^\dagger-\mu_t^\dagger)(1-y_t^c)] &= \mathbb{E}_\theta[(y_t^c-\mu_t^c)(y_t^\dagger-\mu_t^\dagger)(1-y_t^c)|y_t^c=0] \\
&\quad \times(1-\alpha_t) + \mathbb{E}_\theta[(y_t^c-\mu_t^c)(y_t^\dagger-\mu_t^\dagger) \\
&\quad \times(1-y_t^c)|y_t^c=1]\alpha_t = 0
\end{aligned}$$

e

$$\begin{aligned}
\mathbb{E}_\theta[(y_t^c-\mu_t^c)(1-y_t^c)] &= \mathbb{E}_\theta[(y_t^c-\mu_t^c)(1-y_t^c)|y_t^c=0](1-\alpha_t) \\
&\quad + \mathbb{E}_\theta[(y_t^c-\mu_t^c)(1-y_t^c)|y_t^c=1]\alpha_t = 0.
\end{aligned}$$

Assim,

$$\mathbb{E}_\theta[U_\gamma(\theta)\ell(\theta)] = Z^\top PGV^c\alpha^*\iota.$$

Adicionalmente,

$$\begin{aligned}\mathbb{E}_\theta[U_\beta(\theta)\ell(\theta)] = X^\top \Phi T\{&\mathbb{E}_\theta[H(y^*-\mu^*)(y^c-\mu^c)^\top]\alpha^* + \mathbb{E}_\theta[H(y^*-\mu^*)]a^\top \\ &+ \mathbb{E}_\theta[H(y^*-\mu^*)(y^*-\mu^*)^\top H](\Phi\mathcal{M}-\mathcal{I}) \\ &+ \mathbb{E}_\theta[H(y^*-\mu^*)(y^\dagger-\mu^\dagger)^\top H](\Phi-2\mathcal{I}) \\ &+ \mathbb{E}_\theta[H(y^*-\mu^*)b^\top\}H]\iota.\end{aligned}$$

Lembrando que $\mathbb{E}_\theta[(y_t^*-\mu_t^*)(y_t^*-\mu_t^*)|y_t^c=0] = v_t^*$ e $\mathbb{E}_\theta[(y_t^*-\mu_t^*)(y_t^\dagger-\mu_t^\dagger)|y_t^c=0] = c_t$, temos que

$$\begin{aligned}\mathbb{E}_\theta[(1-y_t^c)(y_t^*-\mu_t^*)] = &\mathbb{E}_\theta[(1-y_t^c)(y_t^*-\mu_t^*)|y_t^c=0](1-\alpha_t) \\ &+ \mathbb{E}_\theta[(1-y_t^c)(y_t^*-\mu_t^*)|y_t^c=1]\alpha_t = 0,\end{aligned}$$

$$\begin{aligned}\mathbb{E}_\theta[(1-y_t^c)(y_t^*-\mu_t^*)(y_t^*-\mu_t^*)(1-y_t^c)] = &\mathbb{E}_\theta[(1-y_t^c)(y_t^*-\mu_t^*)(y_t^*-\mu_t^*) \\ &\times(1-y_t^c)|y_t^c=0](1-\alpha_t) \\ &+ \mathbb{E}_\theta[(1-y_t^c)(y_t^*-\mu_t^*)(y_t^*-\mu_t^*) \\ &\times(1-y_t^c)|y_t^c=1]\alpha_t \\ &= v_t^*(1-\alpha_t),\end{aligned}$$

$$\begin{aligned}\mathbb{E}_\theta[(1-y_t^c)(y_t^*-\mu_t^*)(y_t^\dagger-\mu_t^\dagger)(1-y_t^c)] = &\mathbb{E}_\theta[(1-y_t^c)(y_t^*-\mu_t^*)(y_t^\dagger-\mu_t^\dagger) \\ &\times(1-y_t^c)|y_t^c=0](1-\alpha_t) \\ &+ \mathbb{E}_\theta[(1-y_t^c)(y_t^*-\mu_t^*)(y_t^\dagger-\mu_t^\dagger) \\ &\times(1-y_t^c)|y_t^c=1]\alpha_t \\ &= c_t(1-\alpha_t)\end{aligned}$$

e

$$\begin{aligned}\mathbb{E}_\theta[(1-y_t^c)(y_t^*-\mu_t^*)(1-y_t^c)] = &\mathbb{E}_\theta[(1-y_t^c)(y_t^*-\mu_t^*)(1-y_t^c)|y_t^c=0] \\ &\times(1-\alpha_t) + \mathbb{E}_\theta[(1-y_t^c)(y_t^*-\mu_t^*) \\ &\times(1-y_t^c)|y_t^c=1]\alpha_t = 0.\end{aligned}$$

Assim,

$$\mathbb{E}_\theta[U_\beta(\theta)\ell(\theta)] = X^\top\Phi\Delta T\{V^*(\Phi\mathcal{M}-\mathcal{I}) + C(\Phi-2\mathcal{I})\}\iota.$$

Por fim,

$$\begin{aligned}\mathbb{E}_\theta[U_\lambda(\theta)\ell(\theta)] = S^\top V\{&\mathbb{E}_\theta[H(y^*-\mu^*)(y^c-\mu^c)^\top]\mathcal{M}\alpha^* + \mathbb{E}_\theta[H(y^*-\mu^*)]a^\top\mathcal{M} \\ &+ \mathbb{E}_\theta[H(y^*-\mu^*)(y^*-\mu^*)^\top H]\mathcal{M}(\Phi\mathcal{M}-\mathcal{I}) \\ &+ \mathbb{E}_\theta[H(y^*-\mu^*)(y^\dagger-\mu^\dagger)^\top H]\mathcal{M}(\Phi-2\mathcal{I}) + \mathbb{E}_\theta[H(y^*-\mu^*)b^\top \\ &\times H]\mathcal{M} + \mathbb{E}_\theta[H(y^\dagger-\mu^\dagger)(y^c-\mu^c)^\top]\alpha^* + \mathbb{E}_\theta[H(y^\dagger-\mu^\dagger)]a^\top \\ &+ \mathbb{E}_\theta[H(y^\dagger-\mu^\dagger)(y^*-\mu^*)^\top H](\Phi\mathcal{M}-\mathcal{I}) \\ &+ \mathbb{E}_\theta[H(y^\dagger-\mu^\dagger)(y^\dagger-\mu^\dagger)^\top H](\Phi-2\mathcal{I}) \\ &+ \mathbb{E}_\theta[H(y^\dagger-\mu^\dagger)b^\top H]\}\iota.\end{aligned}$$

Lembrando que $\mathbb{E}_\theta[(y_t^\dagger - \mu_t^\dagger)(y_t^\dagger - \mu_t^\dagger)|y_t^c = 0] = v_t^\dagger$, obtemos

$$\mathbb{E}_\theta[(1-y_t^c)(y_t^\dagger - \mu_t^\dagger)(y_t^\dagger - \mu_t^\dagger)(1-y_t^c)] = v_t^\dagger(1-\alpha_t).$$

Assim,

$$\mathbb{E}_\theta[U_\lambda(\theta)\ell(\theta)] = S^\top V\Delta\{(V^*\mathcal{M}+C)(\Phi\mathcal{M}-\mathcal{I})+(V^\dagger+C\mathcal{M})(\Phi-2\mathcal{I})\}\iota.$$

Para obter $\bar{\Upsilon}$ é necessário calcular as seguintes esperanças:

$$\begin{aligned}\mathbb{E}_{\theta_1}[U_\gamma(\theta_1)U_\gamma^\top(\theta)] &= \mathbb{E}_{\theta_1}[Z^\top P^{(1)}G^{(1)}(y^c-\mu^{c(1)})(y^c-\mu^c)^\top GPZ]\\ &= Z^\top P^{(1)}G^{(1)}V^{c(1)}GPZ,\end{aligned}$$

$$\mathbb{E}_{\theta_1}[U_\gamma(\theta_1)U_\beta^\top(\theta)] = \mathbb{E}_{\theta_1}[U_\beta(\theta_1)U_\gamma^\top(\theta)] = 0,$$

$$\mathbb{E}_{\theta_1}[U_\gamma(\theta_1)U_\lambda^\top(\theta)] = \mathbb{E}_{\theta_1}[U_\lambda(\theta_1)U_\gamma^\top(\theta)] = 0,$$

$$\begin{aligned}\mathbb{E}_{\theta_1}[U_\beta(\theta_1)U_\beta^\top(\theta)] &= \mathbb{E}_{\theta_1}[X^\top\Phi^{(1)}HT^{(1)}(y^*-\mu^{*(1)})(y^*-\mu^*)^\top T\Phi HX]\\ &= X^\top\Phi^{(1)}T^{(1)}V^{*(1)}\Delta^{(1)}T\Phi X,\end{aligned}$$

$$\begin{aligned}\mathbb{E}_{\theta_1}[U_\beta(\theta_1)U_\lambda^\top(\theta)] &= \mathbb{E}_{\theta_1}[X^\top\Phi^{(1)}HT^{(1)}(y^*-\mu^{*(1)})\{(y^*-\mu^*)^\top\mathcal{M}\\ &\quad +(y^\dagger-\mu^\dagger)^\top\}VHS]\\ &= X^\top\Phi^{(1)}T^{(1)}\Delta^{(1)}(V^{*(1)}\mathcal{M}+C^{(1)})VS,\end{aligned}$$

$$\begin{aligned}\mathbb{E}_{\theta_1}[U_\lambda(\theta_1)U_\beta^\top(\theta)] &= \mathbb{E}_{\theta_1}[S^\top HV^{(1)}\{\mathcal{M}^{(1)}(y^*-\mu^{*(1)})+(y^\dagger-\mu^{\dagger(1)})\}\\ &\quad\times(y^*-\mu^*)^\top T\Phi HX]\\ &= S^\top V^{(1)}\Delta^{(1)}(V^{*(1)}\mathcal{M}^{(1)}+C^{(1)})T\Phi X\end{aligned}$$

e

$$\begin{aligned}\mathbb{E}_{\theta_1}[U_\lambda(\theta_1)U_\lambda^\top(\theta)] &= \mathbb{E}_{\theta_1}[S^\top HV^{(1)}\{\mathcal{M}^{(1)}(y^*-\mu^{*(1)})+(y^\dagger-\mu^{\dagger(1)})\}\\ &\quad\times\{(y^*-\mu^*)^\top\mathcal{M}+(y^\dagger-\mu^\dagger)^\top\}VHS]\\ &= S^\top V^{(1)}\Delta^{(1)}\{V^{*(1)}\mathcal{M}^{(1)}\mathcal{M}+C^{(1)}(\mathcal{M}^{(1)}+\mathcal{M})\\ &\quad+V^{\dagger(1)}\}VS.\end{aligned}$$

Referências Bibliográficas

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

[2] Aitchison, J. *The Lognormal Distribution*. Cambridge: Cambridge University Press, 1969.

[3] Assunção, R. M. *Estatística Espacial com Aplicações em Epidemiologia, Economia e Sociologia*. São Carlos: 7^a Escola de Modelos de Regressão, 2001.

[4] Banker, R.; Charnes, A.; Cooper, W. W. Some models for estimating technical and scale efficiencies in data envelopment analisys. *Management Science 30* (1984), 1078–1092.

[5] Barndorff-Nielsen, O. E. Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika 73* (1986), 307–322.

[6] Barndorff-Nielsen, O. E. Modified signed log likelihood ratio. *Biometrika 78* (1991), 557–563.

[7] Bartlett, M. S. Properties of sufficiency and statistical tests. *Proceedings of the Royal Society A 160* (1937), 268–282.

[8] Bartlett, M. S. Further aspects of the theory of multiple regression. *Mathematical Proceedings of the Cambridge Philosophical Society 34* (1938), 33–40.

[9] Bartlett, M. S. Multivariate analysis (with discussion). *Supplement to the Journal of the Royal Statistical Society 9* (1947), 176–197.

[10] Bartlett, M. S. A note on the multiplying factors for various χ^2 approximations. *Journal of the Royal Statistical Society B 16* (1954), 296–298.

[11] Bellio, R.; Brazzale, A. R. Higher-order likelihood-based inference in nonlinear regression. In *Proceedings of the14th international workshop on statistical modelling* (1999), Technical university, Graz, pp. 440–443.

[12] Box, G. E. P. A general distribution theory for a class of likelihood criteria. *Biometrika 36* (1949), 317–346.

[13] Charnes, A.; Cooper, W. W.; Rhodes, E. Measuring efficiency of decision making units. *European Journal of Operational Research 1* (1978), 429–444.

[14] Cook, D. O.; Kieschnick, R.; McCullough, B. D. On the hetergeneity of corporate capital structures and its implications. Disponível em: http://ssrn.com/abstract=671061, 2006.

[15] Cox, D. R.; Reid, N. Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society B 49* (1987), 1–39.

[16] Cressie, N. A. C. *Statistics for Spatial Data*. New York:Wiley, 1993.

[17] Cribari-Neto; Pereira, T. L. Uma avaliação da eficiência de administrações municipais no estado de são paulo. Texto para discussão, 2010.

[18] Cribari-Neto, F.; Lima, L. B. A misspecification test for beta regressions. Texto para discussão, 2007.

[19] Cysneiros, A. H. M. A.; Ferrari, S. L. P. An improved likelihood ratio test for varying dispersion in exponential family nonlinear models. *Statistics and Probability Letters 76* (2006), 255–265.

[20] Espinheira, P. L.; Ferrari, S. L. P.; Cribari-Neto, F. Influence diagnostics in beta regression. *Computational Statistics and Data Analysis 52* (2008a), 4417–4431.

[21] Espinheira, P. L.; Ferrari, S. L. P.; Cribari-Neto, F. On beta regression residuals. *Journal of Applied Statistics 35* (2008b), 407–419.

[22] Farrel, M. J. The measurement of productive efficiency. *Journal of the Royal Statistical Society Series A 120* (1957), 253–281.

[23] Ferrari, S. L. P.; Cribari-Neto, F. Corrected modified profile likelihood heteroskedasticity tests. *Statistics and Probability Letters 57* (2002), 353–361.

[24] Ferrari, S. L. P.; Cribari-Neto, F. Beta regression for modelling rates and proportions. *Journal of Applied Statistics 31* (2004), 799–815.

[25] Ferrari, S. L. P.; Cysneiros, A. H. M. A. Skovgaard´s adjustment to likelihood ratio tests in exponential family nonlinear models. *Statistics and Probability Letters 78* (2008), 3047–3055.

[26] Ferrari, S. L. P.; Pinheiro, E. C. Improved likelihood inference in beta regression. Artigo aceito para publicação no Journal of Statistical Computation and Simulation, 2010.

[27] Feuerverger, A. On some methods of analysis for weather experiments. *Biometrika 66* (1979), 665–668.

[28] Hausman, J. A. Specification tests in econometrics. *Econometrica 46* (1978), 1251–1271.

[29] Heller, G.; Stasinopoulos, M.; Rigby, B. The zero-adjusted inverse gaussian distribution as a model for insurance claims. In *Proceedings of the 21th International Workshop on Statistical Modelling* (2006), J. Hinde; J. Einbeck; ; J. Newell, Eds., Galway: Irland, pp. 226–233.

[30] Hoof, A. Second stage dea: comparison of approaches for modelling the dea escore. *European Journal of Operational Research 181* (2007), 425–435.

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

[32] Mantalos, P.; Shukur, G. The robustness of the reset test to non-normal error terms. *Computational Economics 30* (2007), 393–408.

[33] McCullagh, P.; Nelder, J. A. *Generalized Linear Models*. London: Chapman and Hall, 1989.

[34] McFadden, D. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics 1* (1974), 105–142. In P. Zarembka (ed.), Academic Press: New York.

[35] Moran, P. A. P. Notes on continuous stochastic phenomena. *Biometrika 37* (1950), 17–23.

[36] Ospina, R. *Modelos de regressão beta inflacionados*. Tese de doutorado, Instituto de Matemática e Estatística da Universidade de São Paulo, 2008.

[37] Ospina, R.; Cribari-Neto; Vasconcellos, K. L. P. Improved point and interval estimation for a beta regression model. *Computational Statistics and Data Analysis 51* (2006), 960–981.

[38] Ospina, R.; Ferrari, S. L. P. Inflated beta distributions. *Statistical Papers 51* (2010), 111–126.

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

[40] Pereira, T. L.; Cribari-Neto, F. Testing model specification in inflated beta regressions. Artigo submetido para publicação, 2010.

[41] Ramsey, J. B. Tests for specification erros in classical linear least squares regression analysis. *Journal of the Royal Statistical Society B 31* (1969), 350–371.

[42] Ramsey, J. B.; Gilbert, R. A monte carlo study of some small sample properties of tests for specification error. *Journal of the American Statistical Association 67* (1972), 180–186.

[43] Ramsey, J. B.; Schmidt, P. Some further results on the use of ols and blus residuals in specification error tests. *Journal of the American Statistical Association 71* (1976), 389–390.

[44] Sampaio de Souza, M. C.; Cribari-Neto, F.; Stosic, B. D. Explaining dea technical efficiency scores in an outlier corrected environment: the case of public services in brazilian municipalities. *Brazilian Review of Econometrics 25* (2005), 289–315.

[45] Santos, F. C. B.; Cribari-Neto, F.; Sampaio de Souza, M. C. Uma avaliação da eficiência do gasto público municipal no brasil. *Revista Brasileira de Estatística 68* (2007), 7–55.

[46] Shukur, G.; Edgerton, D. L. The small sample properties of the reset test as applied to systems of equations. *Journal of Statistical Computation and Simulation 72* (2002), 909–924.

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

[48] Simonoff, J. S.; Tsai, C. H. Use of modified profile likelihood for improved tests of constancy of variance in regression. *Applied Statistics 43* (1994), 357–370.

[49] Skovgaard, I. M. An explicit large-deviation approximation to one-parameter tests. *Bernoulli 2* (1996), 145–165.

[50] Skovgaard, I. M. Likelihood asymptotics. *Scandinavian Journal of Statistics 28* (2001), 3–32.

[51] Thursby, J. G.; Schmidt, P. Some properties of tests for specification error in a linear regression model. *Journal of the American Statistical Association 72* (1977), 635–641.

[52] Tobin, J. Estimation of relationships for limited dependent variables. *Econometrika 26* (1958), 24–36.

[53] Vasconcellos, K. L. P.; Cribari-Neto, F. Improved maximum likelihood estimation in a new class of beta regression models. *Brazilian Journal of Probability and Statistics 19* (2005), 13–31.

[54] Wedderburn, R. W. M. Quasi-likelihood functions, generalized linear models, and the gauss-newton method. *Biometrika 61* (1974), 439–447.

[55] Wei, B. C.; Shi, J. Q.; Fung, W. K.; Hu, Y. Q. Testing for varying dispersion in exponential family nonlinear models. *Annals of the Institute of Statistical Mathematics 50* (1998), 277–294.

[56] Whittle, P. On stationary processes in the plane. *Biometrika 41* (1954), 434–449.

[57] Yoo, S. A note on an approximation of the mobile communications expeditures distribution function using a mixture model. *Applied Statistics 31* (2004), 747–752.