



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE BACHARELADO EM ESTATÍSTICA



Kleber Henrique dos Santos

Teste de especificação correta em modelos de regressão Beta Prime

João Pessoa

2021

Kleber Henrique dos Santos

Teste de especificação correta em modelos de regressão Beta Prime

Monografia apresentada ao Curso de Bacharelado em Estatística da Universidade Federal da Paraíba, como requisito parcial para obtenção do título de Bacharel em Estatística.

Universidade Federal da Paraíba – UFPB

Centro de Ciências Exatas e da Natureza

Departamento de Estatística

Curso de Bacharelado em Estatística

Orientador: Profa. Dra. Tarciana Liberal Pereira

João Pessoa

2021

Catálogo na publicação
Seção de Catalogação e Classificação

S237t Santos, Kleber Henrique dos.

Teste de especificação correta em modelos de regressão Beta Prime / Kleber Henrique dos Santos. - João Pessoa, 2021.

71 f. : il.

Orientação: Tarciana Liberal Pereira.

TCC (Graduação/Bacharelado em Estatística) - UFPB/CCEN.

1. Análise de regressão. 2. Teste de erro de especificação. 3. Modelo de regressão Beta Prime. 4. Função de ligação incorreta. 5. Simulação de Monte Carlo. 6. Omissão de variável importante. I. Pereira, Tarciana Liberal Pereira. II. Título.

UFPB/CCEN

CDU 519.233.5(043.2)



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
COORDENAÇÃO DO CURSO DE ESTATÍSTICA



ATA DE DEFESA DO TRABALHO DE CONCLUSÃO DE CURSO

“Teste de Especificação Correta em Modelos de Regressão Beta Prime”

Kleber Henrique dos Santos

No nono dia do mês de Julho de 2021 às 14:00 horas via Internet, a Banca Examinadora do Trabalho de Conclusão de Curso do(a) discente Kleber Henrique dos Santos, matrícula 2016096042, foi composta pelos professores: Dra. Tarciana Liberal Pereira, Presidente/Orientadora (Departamento de Estatística - UFPB), Dra. Ana Hermínia Andrade e Silva, Examinadora (Departamento de Estatística - UFPB), Dra. Tatiene Correia de Souza, Examinadora (Departamento de Estatística - UFPB) e Dr. Luiz Medeiros de Araújo Lima Filho, Examinador Suplente (Departamento de Estatística - UFPB). Iniciando-se os trabalhos, a presidente da Banca Examinadora cumprimentou os presentes, comunicou aos mesmos a finalidade da reunião e passou a palavra ao candidato para que se fizesse, oralmente, a exposição do Trabalho de Conclusão de Curso intitulado “Teste de Especificação Correta em Modelos de Regressão Beta Prime”. Concluída a apresentação, o(a) candidato(a) foi arguido(a) pela Banca Examinadora que sugeriu que o(a) discente fizesse algumas alterações até o dia 25 de Julho de 2021. Uma vez entregue a versão final do Trabalho de Conclusão de Curso à Coordenação do Bacharelado em Estatística com as alterações solicitadas pela banca examinadora dentro do prazo que o(a) discente recebeu. O(a) discente será aprovado(a) com a nota 10,0 (dez), resultante da média aritmética das notas atribuídas pelos membros da Banca Examinadora.

Dra. Tarciana Liberal Pereira
(Professora Orientadora)

Dra. Ana Hermínia Andrade e Silva
(Professora Examinadora)

Dra. Tatiene Correia de Souza
(Professora Examinadora)

Kleber Henrique dos Santos
(Discente)

João Pessoa, 09 de Julho de 2021.

Este trabalho é dedicado a minha família e amigos,
que, certamente, em muito me agregaram ao longo
deste curso.

Agradecimentos

Agradeço primeiramente a Deus pelo dom da vida e por me permitir dar mais esse passo. E também a Nossa Senhora por toda proteção.

Aos meus pais, Edgley e Tiane, muito obrigado por todo incentivo e motivação, pois sem vocês eu não iria conseguir chegar até aqui. Obrigado também por todos os conselhos, paciência, compreensão e principalmente todo amor. Minha eterna gratidão.

Aos meus irmãos, Keyla e Junior, por terem me suportado em meus piores momentos e por me motivarem sempre.

A toda a minha família, que não irei citar nomes para não cometer nenhuma injustiça, muito obrigado por toda torcida e por sempre acreditarem em mim, mesmo quando nem eu mesmo acreditava.

Ao meu GO que é o meu lugar de descanso, onde minha alma encontra paz. Obrigado por todo suporte que vocês me proporcionam e por compreenderem as minhas ausências.

Aos meus amigos da turma de 2016, por todo companherismo, amizade, brincadeiras e aprendizados, em especial a Franciele, Milleny, Marina, Natan, Ullysses e Gleizi. Vocês são muito especiais e são para a vida.

A turma de 2017, ao qual fui agregado posteriormente, muito obrigado por compartilharem comigo essa reta final de curso, pela troca de experiência, pela amizade, pelo elo profissional que foi construído. Em Especial Wanusa, Leticia, Nicollas, Cecilia, Carol e Bianca.

A todos os colegas e amigos que a estatística me deu. Em especial Fernanda, Kelfanio, Luis Eduardo, Juliana, João Paulo, Natanael, Matheus P., Matheus A., Giovanni, Manoel, Natasha e Dani.

Aos professores do DE, meu muitíssimo obrigado por tudo que pude aprender com vocês ao longo do curso, com toda certeza irei levar para a vida todos os conselhos que obtive de vocês, e que muito contribuíram para o meu crescimento profissional. Em especial gostaria de agradecer aos professores Ana Hermínia, Ana Flávia, Isabel, Maria Lídia, Cláudio, Rodrigo, Marcelo, Hemílio, Luiz, Pedro, Eufrásio, Ulisses, Ronei e Neir. De modo especial quero agradecer a professora Everlane por ter me dado a oportunidade de ser seu aluno PIBIC, por todo aprendizado, parceria e amizade.

A professora Tatiene, nunca consegui dizer isso pessoalmente mas não poderia deixar de agradecer-lá nesse momento. Acredito que mesmo sem saber, ela tem uma grande contribuição para que eu conseguisse chegar até o fim desse curso. Em um dos momen-

tos, em que quase abandonei a graduação ela teve esse papel de devolver a paixão pela estatística, que por vários motivos eu estava perdendo.

Um agradecimento mais que especial a minha orientadora, professora Tarciana. Obrigado pela orientação ao longo desses quase dois anos, entre PIBIC e TCC. Minha gratidão por tudo que pude aprender com essa grande profissional no decorrer desse tempo. Muito obrigado por todos os ensinamentos, conselhos, pela confiança e motivação.

As professoras Tatiene e Ana Hermínia pela solicitude em participar da banca que contribuiu no enriquecimento desse trabalho.

Ao CNPQ pelo apoio financeiro durante a graduação.

“Todas as vitórias ocultam uma abdicação.”
(Simone de Beauvoir)

Resumo

O modelo de regressão beta prime (BOURGUIGNON; SANTOS-NETO; CASTRO, 2018) é uma alternativa aos modelos lineares generalizados bastante útil para modelar dados positivos assimétricos. Neste trabalho, nós propomos dois testes de especificação correta, baseados no teste RESET (RAMSEY, 1969), para o modelo de regressão beta prime. No primeiro teste a variável de teste é adicionada ao submodelo da média, ao passo que no segundo teste, as variáveis de teste são acrescentadas aos submodelos da média e da precisão. Através de simulações de Monte Carlo, em amostras finitas, avaliamos o desempenho dos testes por meio das propriedades de tamanho e poder, obtendo assim informações sobre a melhor combinação de estatística e variável de teste para realização dos testes propostos. Duas aplicações a dados reais são apresentadas para demonstrar a aplicabilidade e importância dos testes propostos.

Palavras-chave: Teste de erro de especificação, Modelo de regressão Beta Prime, Função de ligação incorreta, Simulação de Monte Carlo, Omissão de variável importante.

Abstract

The beta prime regression model proposed by BOURGUIGNON; SANTOS-NETO; CASTRO (2018) is a very useful alternative to generalized linear model for modeling positive asymmetric data. In this paper, we propose two general misspecification tests based on the RESET test RAMSEY (1969) for beta prime regression models with varying precision. In the first test, we add the testing variable in the mean submodel, whereas the second test focuses on adding the testing variables in all submodels. We conduct an extensive Monte Carlo simulation study to evaluate the performance of the proposed tests in finite sample size in terms of their sizes and powers, thus obtaining information about the best combination of test statistics and testing variables to perform the proposed tests. We also present and discuss two empirical applications to show the applicability and importance of the proposed tests

Keywords: Misspecification test, Beta prime regression model, Incorrect link function, Monte Carlo simulation, Omitted variables.

Lista de ilustrações

- Figura 1 – Envelope simulado dos resíduos quantílicos para os dados do milho. . . 40
- Figura 2 – Envelope simulado dos resíduos quantílicos para os dados de cerejeiras. 41

Lista de tabelas

Tabela 1 – Cenários considerados na simulação.	24
Tabela 2 – Taxa de rejeição da hipótese nula (%) para a o cenário I.	27
Tabela 3 – Taxa de rejeição da hipótese nula (%) para a o cenário II.	28
Tabela 4 – Taxa de rejeição da hipótese nula (%) para a o cenário III.	29
Tabela 5 – Taxa de rejeição da hipótese nula (%) para a o cenário IV.	30
Tabela 6 – Taxa de rejeição sob a hipótese alternativa (função de ligação incorreta), cenário I.	31
Tabela 7 – Taxa de rejeição sob a hipótese alternativa (função de ligação incorreta), cenário II.	32
Tabela 8 – Taxa de rejeição sob a hipótese alternativa (função de ligação incorreta), cenário III.	33
Tabela 9 – Taxa de rejeição sob a hipótese alternativa (função de ligação incorreta), cenário IV.	34
Tabela 10 – Taxa de rejeição sob a hipótese alternativa (omissão de variável importante), cenário I.	35
Tabela 11 – Taxa de rejeição sob a hipótese alternativa (omissão de variável importante), cenário II.	36
Tabela 12 – Taxa de rejeição sob a hipótese alternativa (omissão de variável importante), cenário III.	37
Tabela 13 – Taxa de rejeição sob a hipótese alternativa (omissão de variável importante), cenário IV.	38
Tabela 14 – Estimativa dos parâmetros para os dados do milho.	39
Tabela 15 – Estimativa dos parâmetros dos dados de cerejeiras.	41

Lista de abreviaturas e siglas

BP	Beta Prime
MLGs	Modelos lineares generalizados
FDP	Função densidade de probabilidade
EMV	Estimadores de máxima verossimilhança
RV	Razão de verossimilhança
S	Escore
W	Wald
G	Gradiente

Sumário

	Lista de ilustrações	8
1	INTRODUÇÃO	12
1.1	Introdução	12
1.2	Organização do trabalho	13
2	DISTRIBUIÇÃO BETA PRIME	15
3	MODELO DE REGRESSÃO BETA PRIME	18
4	TESTE DE ERRO DE ESPECIFICAÇÃO PARA O MODELO DE REGRESSÃO BETA PRIME	20
5	ESTUDO DE SIMULAÇÃO	23
5.1	Simulação de Tamanho	23
5.2	Simulação de Poder	25
5.2.1	Função de ligação incorreta	25
5.2.2	Omissão de variável importante	25
6	APLICAÇÃO	39
6.1	Dados do Milho	39
6.2	Dados de Cerejeiras	40
7	CONSIDERAÇÕES FINAIS	42
7.1	Conclusão	42
7.2	Publicação	42
	REFERÊNCIAS	43
A	APÊNDICE	46
A.1	Vetor Escore	46
A.2	Matriz de Informação de Fisher	47
B	IMPLEMENTAÇÃO COMPUTACIONAL - R	49

1 Introdução

1.1 Introdução

Modelos de regressão são bastantes utilizados quando se deseja observar o comportamento de uma variável resposta em um conjunto de variáveis explicativas. Os modelos lineares generalizados (MLGs), introduzidos na literatura por NELDER; WEDDERBURN (1972), são utilizados, em geral, quando o modelo normal linear não é adequado. Nos MLGs a distribuição da variável resposta pertence a família exponencial e a média é modelada por meio de uma estrutura de regressão que envolve parâmetros desconhecidos, covariáveis e uma função de ligação. A distribuição beta prime (*BP*) foi inicialmente proposta por KEEPING (1962) e MCDONALD (1984) em termos de parâmetros de forma e o suporte são os reais positivos ($y > 0$). BOURGUIGNON; SANTOS-NETO; CASTRO (2018) propuseram uma nova parametrização para a distribuição *BP* em termos da média, μ , e do parâmetro de precisão, ϕ . Os autores também propuseram o modelo de regressão *BP* baseado na nova parametrização da distribuição. Uma vantagem do modelo proposto, em relação a outros modelos de regressão para dados positivos, como por exemplo o modelo de regressão Gamma e o modelo Birnbaum-Saunders reparametrizado, é a sua grande flexibilidade em lidar com dados positivos assimétricos.

Uma etapa importante, quando está se buscando um modelo de regressão, é a verificação de que o modelo retrata adequadamente a realidade do fenômeno em estudo. Na modelagem estatística, ao utilizarmos um modelo de regressão não sabemos se o modelo irá ajustar bem os dados do referente problema, isso pode ocorrer por vários motivos entre eles a especificação incorreta do modelo, que pode acarretar em inferências imprecisas na estimação dos parâmetros, testes de hipóteses e intervalos de confiança. RAMSEY (1969) propôs o teste RESET (regression specification error test) que é bastante utilizado para averiguar se a suposição de especificação correta é válida no modelo de regressão linear clássico. O impacto causado ao usar uma especificação incorreta no modelo pode ser considerado por uma função analítica (desconhecida) do preditor linear, aproximada por um polinômio. Dessa forma, o teste se reduz a incluir uma forma não-linear por meio de variáveis de teste e testar a exclusão das variáveis adicionadas através de valores críticos da distribuição *F* de Snedecor. O propósito do teste é verificar se as variáveis de teste têm poder em explicar a variável dependente. Para mais detalhes ver GOODFREY (1988) .

No que tange as propriedades do teste RESET, RAMSEY; GILBERT (1972) recomendaram a segunda, terceira e quarta potência dos valores ajustados como variáveis de teste. THURSBY; SCHMIDT (1977) recomendaram a utilização da segunda, terceira e quarta potência das variáveis independentes. GODFREY; ORME (1994) e HOROWITZ

(1994) relataram boas propriedades a respeito do tamanho e do poder sob a especificação de forma funcional. PETERS (2000) estudou o desempenho do teste RESET para os modelos de regressão linear, Tobit e Weibull e evidenciou que o teste é válido para qualquer modelo econométrico. SHUKUR; EDGERTON (2002) generalizaram o teste RESET por meio de um sistema de equações simultâneas. SAPRA et al. (2005) estenderam o teste RESET para MLGs. MANTALOS; SHUKUR (2007) investigaram a robustez do teste RESET para erros não-normais. LIMA (2007) desenvolveu um teste de erro de especificação para o modelo de regressão beta, com dispersão fixa. OLIVEIRA (2013) utilizou o teste de erro de especificação para o modelo de regressão beta com dispersão fixa e variável. PEREIRA; CRIBARI-NETO (2014) propuseram o teste para o modelo de regressão beta inflacionado. CANTERLE; BAYER (2015) fizeram o uso do teste RESET para modelos lineares generalizados para dados binários. SANTOS; CRIBARI-NETO (2017) introduziram um teste de especificação correta para o modelo de regressão Log-Birnbaum-Saunders.

Este trabalho tem como objetivo propor dois testes de erro de especificação para o modelo de regressão *BP* com dispersão variável baseados no teste proposto por RAMSEY (1969). Os testes consistem em verificar se a especificação do modelo de regressão beta prime está correta, ou seja, se o modelo estimado está de acordo com o processo de geração dos dados. O procedimento do teste equivale a incluir no modelo variáveis de teste e testar a exclusão dessas variáveis, por meio de um teste de hipótese tal como os testes da razão de verossimilhança, score, Wald e gradiente. Aproximações das distribuições nulas exatas dessas estatísticas de teste pela distribuição qui-quadrado podem ser pobres em pequenas amostras (PEREIRA; CRIBARI-NETO, 2014). Dessa forma, torna-se primordial avaliar numericamente o desempenho desses testes de hipóteses em amostras de tamanho finito. Adicionalmente, consideramos diferentes variáveis de teste como as potências do valor predito e do preditor linear ajustado. Nos centralizamos em investigar dois tipos de erro de especificação que foram a omissão de variável importante e função de ligação incorreta. Por meio de simulações de Monte Carlo avaliamos o desempenho dos testes através das taxas de rejeição da hipótese nula (tamanho) e das taxas de rejeição da hipótese alternativa (poder) e os resultados indicam que os testes propostos podem ser bastantes úteis para identificar erros de especificação no modelo.

1.2 Organização do trabalho

Além deste capítulo de introdução, este trabalho é composto por mais seis capítulos, como segue:

- No Capítulo 2 apresentamos a distribuição *BP*;
- No Capítulo 3 abordamos o modelo de regressão *BP* com dispersão variável;

-
- No Capítulo 4 apresentamos os testes de erro de especificação propostos;
 - No Capítulo 5 avaliamos os testes propostos, com diferentes implementações, em amostras finitas por meio de simulações de Monte Carlo;
 - No Capítulo 6 apresentamos aplicações a dados reais;
 - Por fim, no Capítulo 7 são apresentadas as considerações finais.

2 Distribuição Beta Prime

A distribuição *BP*, foi introduzida inicialmente por KEEPING (1962) e MCDONALD (1984), e também é conhecida como distribuição beta inversa ou beta do tipo dois. Trata-se de uma distribuição de probabilidade contínua com suporte nos reais positivos, $y > 0$. Entretanto, há poucos estudos envolvendo a distribuição *BP*. Em MCDONALD (1987) foram discutidas propriedades e os estimadores de máxima verossimilhança foram obtidos para os parâmetros do modelo.

Seja Y uma variável aleatória que segue distribuição *BP*, $Y \sim BP(\alpha, \beta)$ cuja função densidade de probabilidade (fdp) é dada por

$$f(y; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} y^{\alpha-1} (1+y)^{-(\alpha+\beta)}, \quad y > 0 \quad (2.1)$$

em que $\alpha > 0$, $\beta > 0$ são parâmetros de forma e $B(\alpha, \beta)$ é a função beta que é dada por

$$\int_0^\infty t^{\alpha-1} (1+t)^{-(\alpha+\beta)} dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)},$$

sendo $\Gamma(\cdot)$ a função gama que é definida por

$$\Gamma(a) = \int_0^\infty e^{-u} u^{a-1} du.$$

A função de distribuição acumulada da distribuição *BP* é da forma

$$F(y; \alpha, \beta) = I_{\frac{y}{1+y}}(\alpha, \beta),$$

em que

$$I_y(\alpha, \beta) = \frac{B(y; \alpha, \beta)}{B(\alpha, \beta)}$$

é a função beta incompleta regularizada. A função densidade de probabilidade da distribuição *BP* (2.1) é decrescente com $f(y; \alpha, \beta) \rightarrow \infty$ de forma que $y \rightarrow 0$, se $0 < \alpha < 1$; $f(y; \alpha, \beta)$ é decrescente com moda em $y = 0$ no caso de $\alpha = 1$, $f(y; \alpha, \beta)$ cresce e depois decresce com moda igual a $(\alpha - 1)/(\beta + 1)$. Além do mais, para $0 < \alpha < 1$, $f(y; \alpha, \beta)$ é côncava; se $1 < \alpha \leq 2$, $f(y; \alpha, \beta)$ é convexa e, em seguida, para cima com ponto de inflexão x_2 ; e caso $\alpha > 2$, $f(y; \alpha, \beta)$ é côncava e então para baixo e volta a ser para cima com pontos de inflexão em x_1 e x_2 , em que

$$x_1 = \frac{(\alpha - 1)(\alpha + 2) - \sqrt{(\alpha - 1)(\beta + 2)(\alpha + \beta)}}{(\beta + 2)(\beta + 1)}$$

e

$$x_2 = \frac{(\alpha - 1)(\alpha + 2) + \sqrt{(\alpha - 1)(\beta + 2)(\alpha + \beta)}}{(\beta + 2)(\beta + 1)}.$$

O r -ésimo momento de uma variável aleatória Y que segue distribuição BP pode ser descrito da forma

$$E[Y^r] = \frac{B(\alpha + r, \beta - r)}{B(\alpha, \beta)}, \quad -\alpha < r < \beta.$$

Para $r \in \mathbf{N}$ e $r < \beta$ o r -ésimo momento é dado por

$$E[Y^r] = \prod_{i=1}^r \frac{\alpha + i - 1}{\beta - i}.$$

A média e a variância são expressas da seguinte forma

$$E[Y] = \frac{\alpha}{\beta - 1}, \quad \beta > 1, \quad \text{e} \quad \text{Var}(Y) = \frac{\alpha(\alpha + \beta - 1)}{(\beta - 2)(\beta - 1)^2}, \quad \beta > 2,$$

respectivamente.

A distribuição BP apresenta relação com algumas outras distribuições, que são:

- Se $Y \sim F(2\alpha, 2\beta)$, então $\frac{\alpha}{\beta}Y \sim BP(\alpha, \beta)$;
- Se $Z \sim B(\alpha, \beta)$, então $\frac{Z}{1-Z} \sim BP(\alpha, \beta)$;
- Se $V \sim \text{Gama}(\alpha, 1)$ e $W \sim \text{Gama}(\beta, 1)$, então $\frac{V}{W} \sim BP(\alpha, \beta)$;
- Se Y tem distribuição Lomax (Pareto tipo II), com parâmetros de escala α e de forma λ , então $\frac{Y}{\lambda} \sim BP(1, \alpha)$;
- Se Y tem distribuição Pareto com mínimo y_1 e parâmetro de forma α , então $Y - y_1 \sim BP(1, \alpha)$;
- Se Y tem distribuição Pareto padrão tipo IV com parâmetro de forma α e parâmetro de desigualdade γ , então $Y^{\frac{1}{\gamma}} \sim BP(1, \alpha)$.

BOURGUIGNON; SANTOS-NETO; CASTRO (2018) propuseram uma parametrização da distribuição BP em termo dos parâmetros da média e da precisão, considerando $\mu = \frac{\alpha}{(\beta-1)}$ e $\phi = \beta - 2$, sendo assim $\alpha = \mu(1 + \phi)$ e $\beta = 2 + \phi$.

Portanto, considerando essa nova parametrização a média e variância são expressas da seguinte forma, respectivamente

$$E(Y) = \mu \quad \text{e} \quad \text{Var}(Y) = \frac{\mu(1 + \mu)}{\phi}.$$

Em termos de notação a variável aleatória Y segue distribuição **BP** com média μ e parâmetro de precisão ϕ , $Y \sim \mathbf{BP}(\mu, \phi)$. A função de variância, $V(\mu) = \mu(1 + \mu)$, é similar a da distribuição gama, em que a variância tem relação quadrática com a média. A função densidade de probabilidade, considerando a nova parametrização, pode ser escrita da seguinte forma

$$f(y; \mu, \phi) = \frac{y^{\mu(\phi+1)-1} (1+y)^{-[\mu(\phi+1)+\phi+2]}}{B(\mu(1+\phi), \phi+2)}, \quad y > 0, \quad (2.2)$$

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

3 Modelo de Regressão Beta Prime

No contexto de modelos de regressão, para variáveis aleatórias positivas, a distribuição **BP** foi utilizada por MCDONALD; BUTLER (1990) e TULUPYEV et al. (2013) considerando a parametrização usual (2.1). BOURGUIGNON; SANTOS-NETO; CASTRO (2018) a distribuição **BP** também foi utilizada em modelos de regressão considerando a nova parametrização (2.2).

Sejam Y_1, \dots, Y_n variáveis aleatórias independentes, no qual cada Y_i , $i = 1, \dots, n$ tem fdp expressa em (2.2) com média μ_i e parâmetro de precisão ϕ_i . Considerando que a média e o parâmetro de precisão para y_i satisfazem as seguintes relações

$$g_1(\mu_i) = \eta_{1i} = x_i^\top \boldsymbol{\beta} \quad \text{e} \quad g_2(\phi_i) = \eta_{2i} = z_i^\top \mathbf{v}, \quad i = 1, \dots, n. \quad (3.1)$$

em que $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)^\top$ e $\mathbf{v} = (v_1, \dots, v_M)^\top$ são vetores de parâmetros de regressão desconhecidos, $\boldsymbol{\beta} \in \mathbb{R}^m$ e $\mathbf{v} \in \mathbb{R}^M$ com $m + M < n$, η_{1i} e η_{2i} são os preditores lineares, $x_i = (x_{i1}, \dots, x_{im})^\top$ e $z_i = (z_{i1}, \dots, z_{iM})^\top$ são observações de m e M regressores conhecidos, para $i = 1, \dots, n$. Assumindo que as matrizes de covariáveis $\mathbf{X} = (x_1, \dots, x_n)^\top$ e $\mathbf{Z} = (z_1, \dots, z_n)^\top$ têm posto m e M , respectivamente, as funções de ligação $g_1 : \mathbb{R} \rightarrow \mathbb{R}^+$ e $g_2 : \mathbb{R} \rightarrow \mathbb{R}^+$ em (3.1) devem ser estritamente monótonas, positivas e ao menos duas vezes diferenciáveis, de forma que $\mu_i = g_1^{-1}(x_i^\top \boldsymbol{\beta})$ e $\phi_i = g_2^{-1}(z_i^\top \mathbf{v})$. As funções de ligação que podem ser empregadas nesse modelo são a raiz quadrada e logarítmica para $\boldsymbol{\mu}$ e $\boldsymbol{\phi}$.

A função log-verossimilhança é da forma

$$\ell(\boldsymbol{\beta}, \mathbf{v}) = \sum_{i=1}^n \ell(\mu_i, \phi_i), \quad i = 1, \dots, n, \quad (3.2)$$

em que

$$\begin{aligned} \ell(\mu_i, \phi_i) = & [\mu_i(1 + \phi_i) - 1] \log(y_i) - [\mu_i(1 + \phi_i) + \phi_i + 2] \log(1 + y_i) - \log[\Gamma(\mu_i(1 + \phi_i))] \\ & - \log[\Gamma(\phi_i + 2)] + \log[\Gamma(\mu_i(1 + \phi_i) + \phi_i + 2)]. \end{aligned}$$

O vetor escore, obtido diferenciando a função log-verossimilhança (3.2), pode ser escrito na forma matricial para $\boldsymbol{\beta}$ e \mathbf{v} , respectivamente, como

$$\begin{aligned} U_{\boldsymbol{\beta}} &= \mathbf{X}^\top \boldsymbol{\Phi} D_1(\mathbf{y}^* - \boldsymbol{\mu}^*), \\ U_{\mathbf{v}} &= \mathbf{Z}^\top D_2(\mathbf{y}^* - \boldsymbol{\mu}^*) \end{aligned}$$

em que $\mathbf{D}_1 = \text{diag}[\frac{\partial \mu_1}{\partial \eta_{11}}, \dots, \frac{\partial \mu_n}{\partial \eta_{1n}}]$, $\mathbf{D}_2 = \text{diag}[\frac{\partial \phi_1}{\partial \eta_{21}}, \dots, \frac{\partial \phi_n}{\partial \eta_{2n}}]$, $\Phi = \text{diag}[(1 + \phi_1), \dots, (1 + \phi_n)]$, $\mathbf{y}^* = (y_1^*, \dots, y_n^*)^\top$ em que $y_i^* = \log(\frac{y_i}{1+y_i})$, $\mathbf{y}^* = (y_1^*, \dots, y_n^*)^\top$ em que $y_i^* = \log(\frac{y_i^{\mu_i}}{(1+y_i)^{1+\mu_i}})$, $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_n^*)^\top$ em que $\mu_i^* = [\Psi^{(0)}(\mu_i(1 + \phi_i)) - \Psi^{(0)}(\mu_i(1 + \phi_i) + \phi_i + 2)]$ e $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_n^*)^\top$ em que $\mu_i^* = \mu_i(\mu_i^* - \frac{\gamma_i}{\mu_i})$ com $\gamma_i = \Psi^{(0)}(\mu_i(1 + \phi_i) + \phi_i + 2) - \Psi^{(0)}(\phi_i + 2)$. Para mais detalhes, ver (A.1).

A matriz de informação de Fisher pode ser expressa da seguinte forma

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{\beta\beta} & \mathbf{K}_{\beta\nu} \\ \mathbf{K}_{\nu\beta} & \mathbf{K}_{\nu\nu} \end{bmatrix}, \quad (3.3)$$

em que $\mathbf{K}_{\beta\beta} = \mathbf{X}^\top E_3 \mathbf{X}$, $\mathbf{K}_{\beta\nu} = (\mathbf{K}_{\nu\beta})^\top = \mathbf{Z}^\top E_5 \mathbf{X}$ e $\mathbf{K}_{\nu\nu} = \mathbf{Z}^\top E_4 \mathbf{Z}$. O i -ésimo elemento de E_3 , E_4 e E_5 são dados, respectivamente, por $e_{3i} = (1 + \phi_i)^2 [\Psi^{(1)}(\mu_i(1 + \phi_i)) - \Psi^{(1)}(\mu_i(1 + \phi_i) + \phi_i + 2)] (\frac{\partial \mu_i}{\partial \eta_{1i}})^2$, $e_{4i} = [\mu_i^2 \Psi^{(1)}(\mu_i(1 + \phi_i)) - (1 + \mu_i)^2 \Psi^{(1)}(\mu_i(1 + \phi_i) + \phi_i + 2) + \Psi^{(1)}(\phi_i + 2)] (\frac{\partial \phi_i}{\partial \eta_{2i}})^2$ e $e_{5i} = -(1 + \phi_i) \{ \Psi^{(1)}(\mu_i(1 + \phi_i) + \phi_i + 2) + \mu_i [\Psi^{(1)}(\mu_i(1 + \phi_i) + \phi_i + 2) - \Psi^{(1)}(\mu_i(1 + \phi_i))] \} (\frac{\partial \mu_i}{\partial \eta_{1i}}) (\frac{\partial \phi_i}{\partial \eta_{2i}})$. Para mais detalhes, ver (A.2).

Os estimadores de máxima verossimilhança (EMV) não apresentam forma fechada e são obtidos através de métodos iterativos de otimização não-linear. O modelo de regressão **BP** é da classe de modelos GAMLSS (Modelos Aditivos Generalizados para Localização Escala e Forma); ver STASINOPOULOS; ENEA; RIGBY (2017). As estimativas dos parâmetros do modelo de regressão **BP** (3.1) podem ser obtidas por meio do pacote `gamlss`, contido no R (R Core Team, 2021).

4 Teste de erro de especificação para o modelo de regressão Beta Prime

Os testes propostos aqui são extensões do teste RESET de especificação correta proposto por RAMSEY (1969), para verificar a forma funcional incorreta e omissão de variável no contexto de modelos de regressão linear. O teste é realizado em duas etapas. Na primeira etapa, variáveis de teste são incluídas e em seguida, na segunda etapa, testa-se a exclusão dessas variáveis adicionadas através de valores críticos da distribuição F de snedecor. O modelo está corretamente especificado se as variáveis de teste não explica a variável dependente. No primeiro teste proposto neste artigo as variáveis de teste são acrescentadas ao submodelo da média, e no segundo teste as variáveis de teste são acrescentadas aos submodelos da média e da precisão. Considere o modelo de regressão BP

$$g_1(\mu_i) = X\beta, \quad i = 1, \dots, n.$$

$$g_2(\phi_i) = Zv, \quad i = 1, \dots, n.$$

em que os vetores β e v são $m \times 1$ e $M \times 1$, respectivamente, X e Z são matrizes de regressores $n \times m$ e $n \times M$, respectivamente, μ e ϕ são vetores $n \times 1$ e $i = 1, \dots, n$.

Considerando o primeiro teste, iremos, na primeira etapa, incluir as variáveis de teste como covariáveis, no submodelo da média. Assim,

$$g_1(\mu_i) = X\beta + A_1\tau_1,$$

em que A_1 é uma matriz $n \times s$ de variáveis de teste e τ_1 é um vetor de parâmetros $s \times 1$. A segunda etapa consiste em testar as hipóteses

$$H_0 : \tau_1 = 0 \quad (\text{corretamente especificado}),$$

$$H_1 : \tau_1 \neq 0 \quad (\text{incorretamente especificado}).$$

As variáveis de teste utilizadas são potências de $\hat{\eta}_1$ e de $\hat{\mu}$. Para testar a exclusão das variáveis de teste, considerando o modelo de regressão beta prime, precisamos recorrer a testes assintóticos, a saber: razão de verossimilhanças (RV) (NEYMAN; PEARSON, 1928), escore (S) (RAO, 1948), wald (W) (WALD, 1943) e gradiente (G) (TERRELL, 2002).

Os testes RV , S e W , apresentam estatísticas de testes equivalentes até a primeira ordem e segue distribuição assintótica χ_k^2 com erro de ordem n^{-1} , k é o número de restrições impostas pela hipótese nula e n o tamanho da amostra. O teste G tem uma estrutura mais simplificada pois, não depende da matriz da informação de fisher. A estatística de teste G segue distribuição assintótica χ_k^2 a menos de termos de ordem n^{-1} . As estatísticas de teste são dadas, respectivamente, por

$$\begin{aligned} RV &= 2[\ell(\hat{\zeta}) - \ell(\tilde{\zeta})], \\ S &= \tilde{U}_{1\beta}^\top \tilde{K}_{11}^{\beta\beta} \tilde{U}_{1\beta}, \\ W &= \hat{\tau}_1^\top \hat{K}_{11}^{\beta\beta} \hat{\tau}_1, \\ G &= \tilde{U}_{1\beta}^\top \hat{\tau}_1, \end{aligned}$$

em que $\hat{\zeta} = (\hat{\tau}_1^\top, \hat{\beta}^\top, \hat{\nu}^\top)^\top$ é o estimador de máxima verossimilhança de ζ e $\tilde{\zeta} = (0^\top, \tilde{\beta}^\top, \tilde{\nu}^\top)^\top$ é o estimador de máxima verossimilhança restrito obtido sob a imposição da hipótese nula. $\tilde{U}_{1\beta}$ é o vetor de dimensão s que contém os primeiros s elementos do vetor escore $U_\beta(\beta, \nu)$ e $\tilde{K}_{11}^{\beta\beta}$ é a matriz $s \times s$ formada pelas s primeiras linhas e s primeiras colunas da inversa da matriz $\mathbf{K}_{\beta\beta}$, ambas estão sendo avaliadas no estimador de máxima verossimilhança restrito e $\hat{K}_{11}^{\beta\beta}$ corresponde a $K_{11}^{\beta\beta}$ avaliado no estimador de máxima verossimilhança irrestrito.

Sob condições usuais de regularidade e sob a hipótese nula, as estatística de teste têm distribuição assintótica χ_s^2 , sendo s o número de restrições impostas por H_0 (VARGAS; FERRARI; LEMONTE, 2014). Neste caso considerando o teste RESET para $s = 1$, se o valor da estatística de teste for menor que o quantil usual da χ_s^2 , não rejeita-se a hipótese nula e assumimos que o modelo está corretamente especificado, caso contrário rejeita-se a hipótese nula.

O segundo teste de erro de especificação proposto é obtido acrescentando variáveis de teste aos dois submodelos, da média e do parâmetro de precisão. Considerando os seguintes modelos aumentados

$$\begin{aligned} g_1(\mu_i) &= X\beta + A_1 \tau_1, \\ g_2(\phi_i) &= Z\nu + A_2 \tau_2, \end{aligned}$$

em que A_2 é uma matriz $n \times r$ de variáveis de teste e τ_2 é um vetor de parâmetros $r \times 1$. A hipótese a ser testada será

$$\begin{aligned} H_0 &: \tau_1 = \tau_2 = 0 \quad (\text{corretamente especificado}), \\ H_1 &: \tau_i \neq 0 \quad (\text{incorretamente especificado}), \end{aligned}$$

ao menos um τ_i diferente de zero.

A estatística de teste RV é da mesma forma que foi vista anteriormente, já as estatísticas S, W e G, são expressas como a soma de duas formas quadráticas:

$$S = \tilde{U}_{1\beta}^\top \tilde{K}_{11}^{\beta\beta} \tilde{U}_{1\beta} + \tilde{U}_{1v}^\top \tilde{K}_{11}^{vv} \tilde{U}_{1v},$$

$$W = \hat{\tau}_1^\top \hat{K}_{11}^{\beta\beta} \hat{\tau}_1 + \hat{\tau}_2^\top \hat{K}_{11}^{vv} \hat{\tau}_2,$$

$$G = \tilde{U}_{1\beta}^\top \hat{\tau}_1 + \tilde{U}_{1v}^\top \hat{\tau}_2,$$

em que \tilde{U}_{1v} é o vetor de dimensão r que contém os primeiros r elementos do vetor escore $U_v(\beta, v)$ e \tilde{K}_{11}^{vv} é a matriz $r \times r$ formada pelas r primeiras linhas e r primeiras colunas da inversa da matriz \mathbf{K}_{vv} , ambas estão sendo avaliadas no estimador de máxima verossimilhança restrito e \hat{K}_{11}^{vv} corresponde a K_{11}^{vv} avaliado no estimador de máxima verossimilhança irrestrito. Sob condições usuais de regularidade e sob a hipótese nula, as estatísticas de teste têm distribuição assintótica χ_{s+r}^2 , sendo s e r o número de restrições impostas por H_0 .

5 Estudo de Simulação

Neste capítulo, através de simulação de Monte Carlo avaliamos o desempenho dos testes de erro de especificação propostos para o modelo de regressão *BP*. 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).

Consideramos o modelo de regressão *BP*

$$g_1(\boldsymbol{\mu}_i) = \beta_0 + \beta_1 x_i \quad \text{e} \quad g_2(\boldsymbol{\phi}_i) = v_0 + v_1 z_i, \quad (5.1)$$

em que $g_1(\cdot)$ e $g_2(\cdot)$ são funções de ligação e x_i e z_i são covariáveis. Para cada covariável foram geradas 50 observações $\mathcal{U}(0, 1)$ e replicadas 2 e 4 vezes obtendo tamanhos de amostra 100 e 200. Em cada réplica geramos uma amostra aleatória da variável dependente $Y = (Y_1, \dots, Y_n)^\top$ com $Y_i \sim BP(\boldsymbol{\mu}_i, \boldsymbol{\phi}_i)$, em que $\boldsymbol{\mu}_i = g_1^{-1}(\boldsymbol{\eta}_{1i})$ e $\boldsymbol{\phi}_i = g_2^{-1}(\boldsymbol{\eta}_{2i})$. Todas as simulações realizadas foram utilizando a linguagem de programação R (R Core Team, 2021) e 10000 réplicas de Monte Carlo.

Os testes foram realizados considerando como variáveis de teste o quadrado e o cubo dos preditores lineares assim como o quadrado e cubo dos valores preditos. Todos os testes foram realizados considerando três diferentes níveis nominais: 1%, 5% e 10%. De modo que a comparação do poder ocorra entre testes com mesmo tamanho, os valores críticos usados nas simulações de poder são extraídos das simulações de tamanho. O teste em que as variáveis de teste foram acrescentadas na estrutura do submodelo da média é denominado de R_μ e quando as variáveis de teste foram acrescentadas nos submodelos da média e do parâmetro de precisão é denominado R_θ , com $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\phi})^\top$. Na simulação de poder os erros de especificação utilizados foram referentes a função de ligação incorreta e omissão de variável importante.

Considerando o modelo de regressão *BP*, dado em (5.1), foram considerados quatro cenários com algumas características importantes, a saber: média baixa e precisão alta (cenário I), média baixa e precisão baixa (cenário II), média alta e precisão alta (cenário III) e média alta e precisão baixa (cenário IV).

5.1 Simulação de Tamanho

As simulações de tamanho são realizadas sob a hipótese nula e a variável resposta é gerada sem erro de especificação, considerando os parâmetros obtidos na Tabela 1, com

Tabela 1 – Cenários considerados na simulação.

Cenários	Valores iniciais				Intervalo	
	β_0	β_1	v_0	v_1	μ	ϕ
I	1,9	-1,5	3,5	1	(1,5; 6,62)	(33,51; 89,96)
II	1,9	-1,5	1,8	-0,4	(1,5; 6,62)	(4,08; 5,99)
III	4,0	-3,5	3,5	1	(1,68; 53,39)	(33,51; 89,96)
IV	4,0	-3,5	1,8	-0,4	(1,68; 53,39)	(4,08; 5,99)

função de ligação logarítmica para os submodelos da média, μ , e da precisão, ϕ . Para o teste R_μ consideramos as estatística de teste RV, S, W e G e para o teste R_θ apenas as estatística de teste RV e G, pois os testes baseados nas estatísticas S e W não apresentaram bons resultados.

As Tabelas 2 a 5 apresentam os percentuais de rejeição da hipótese nula de que o modelo está corretamente especificado. À medida que o tamanho da amostra aumenta, os percentuais de rejeição convergem para os níveis nominais especificados. De forma geral, em todos os cenários o teste considerando as estatísticas de teste S e G apresentaram as taxas de rejeição mais próximas dos níveis nominais para o teste R_μ e o teste baseado na estatística G apresentou melhor desempenho pra o teste R_θ . Por exemplo, na Tabela 3, para $n = 100$ e $\alpha = 5\%$ o teste baseado na variável $\hat{\mu}^3$ obteve tamanhos estimados de 5,60%, 4,92%, 6,24% e 5,32%, respectivamente, para as estatísticas RV, S, W e G, para o teste R_μ .

Na Tabela 4 para $n = 200$, $\alpha = 1\%$ incluindo a variável de teste $\hat{\eta}^2$ os tamanhos dos teste R_μ , considerando as estatística RV, S, W e G são, respectivamente, 0,98%, 0,83%, 1,10% e 0,90%. Ou seja, o teste baseado na estatística RV também apresenta bons desempenhos, principalmente para amostras de tamanho maiores, e isso ocorre também no teste R_θ .

Em geral, o teste considerando a estatística W foi a que apresentou o pior desempenho para o teste R_μ . Por exemplo, na Tabela 5 para $n = 50$ e $\alpha = 10\%$ na inclusão das variável de teste $\hat{\eta}^2$, $\hat{\eta}^3$, $\hat{\mu}^2$ e $\hat{\mu}^3$ o teste baseado na estatística W apresentou tamanho de 14,50%, 14,59%, 14,11% e 14,07%, respectivamente.

No teste R_θ , para os níveis de significância de 5% e 10% o teste considerando as estatísticas apresentaram, em geral, tamanhos maiores que os níveis nominais, ou seja, os testes tendem a rejeitar a hipótese de que o modelo está corretamente especificado mais que o esperado. Há indícios suficientes de que o teste R_μ é mais eficiente que o teste R_θ , pois apresenta menor distorção de tamanho, principalmente quando a variável de teste é $\hat{\eta}^2$.

5.2 Simulação de Poder

As simulações de poder foram realizadas considerando erro de especificação no submodelo da média, teste R_μ , e nos dois submodelos da média e da precisão, teste R_θ .

5.2.1 Função de ligação incorreta

No que se refere à identificação de erro na especificação da função de ligação, os dados são gerados com função de ligação raiz quadrada para μ e logarítmica para ϕ , no teste R_μ . Adicionalmente, considerando o teste R_θ , os dados são gerados com função de ligação raiz quadrada para μ e ϕ . Em ambos os testes o modelo foi estimado considerando a função de ligação logarítmica. Os valores iniciais dos parâmetros, estão apresentados na Tabela 1.

Os resultados de simulação de poder dos testes no que tange à identificação do erro na especificação da função de ligação em uso, encontram-se exibidos nas Tabelas 6 a 9. Analisando os resultados, o teste R_μ apresenta, em geral, poderes mais elevados do que o teste R_θ . Em particular, no que tange ao teste R_μ , o teste baseado na estatística S resultou em poderes do teste levemente maiores do que os poderes do teste considerando a estatística G. No que tange o teste R_θ , o teste baseado na estatística G apresentou poderes levemente superiores ao teste baseado na estatística RV.

É importante observar que, em geral, quando o tamanho da amostra aumenta, ambos os testes apresentaram bons resultados de poder. Por exemplo, na Tabela 8 para o teste R_μ , com $n = 100$ e $n = 200$, em todos os níveis nominais e na inclusão das variáveis de teste $\hat{\eta}^2$, $\hat{\eta}^3$, $\hat{\mu}^2$ e $\hat{\mu}^3$, resultou em uma taxa de rejeição da hipótese alternativa de 100%. Assim como para $n = 200$ no teste R_θ .

Em geral, o teste baseado na estatística W apresentou o pior resultado para o teste R_μ , em todos os cenários. Por exemplo, na Tabela 7 para $n = 50$ e $\alpha = 1\%$, na inclusão das variáveis de teste $\hat{\eta}^2$, $\hat{\eta}^3$, $\hat{\mu}^2$ e $\hat{\mu}^3$ o teste considerando a estatística W apresentou taxas de rejeição sob a hipótese alternativa de 13,14%, 4,68%, 9,90% e 7,86%, respectivamente.

Podemos concluir que os testes apresentaram melhores resultados nos cenários em que foram considerados a precisão alta (cenários I e III). Dentre esses dois cenários, o cenário III, considerando valores altos para a média, foi o que apresentou melhores resultados.

5.2.2 Omissão de variável importante

Um importante erro de especificação é a não inclusão de uma variável importante no modelo. Considere o seguinte modelo:

$$g_1(\mu_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2,$$

em que β_0 e β_1 são apresentados na Tabela 1, para os cenários I e II, $\beta_2 = 3,0$ e para os cenários III e IV, $\beta_2 = 2,8$. Quando o erro de especificação ocorre nos dois submodelos, da média e do parâmetro de precisão, temos o seguinte modelo para estrutura de geração da precisão

$$g_2(\phi_i) = v_0 + v_1 z_1 + v_2 z_2,$$

em que os valores iniciais de v_0 e v_1 estão apresentados na Tabela 1, $v_2 = 1,5$ para os cenários I e III e $v_2 = 1,2$ para os cenários II e IV. A variável omitida na estimação do modelo foi x_2 , para o teste R_μ e x_2 e z_2 para o teste R_θ .

Conforme pode ser visto nas tabelas 10 a 13 os resultados da simulação de poder com base na omissão de variável importante são apresentados. Nos cenários com precisão alta (I e III) e em todos os níveis nominais considerados, tanto o teste R_μ quanto o teste R_θ apresentam taxas de rejeição sob a hipótese alternativa superior a 99%.

Nos cenários com precisão baixa (II e IV) e $n < 200$, o teste R_μ , baseado na estatística G, é mais poderoso do que o teste baseado na estatística S. Por exemplo, considerando o cenário II (Tabela 11) e nível nominal de 1%, os poderes do teste R_μ baseado na estatística G foram 81,37 e 99,53 para $n = 50$ e $n = 100$, respectivamente. Por outro lado, considerando a teste baseado na estatística S, os poderes foram 69,38 e 98,54, respectivamente. No teste R_θ , o teste baseado na estatística RV apresenta taxas de rejeição sob a hipótese alternativa superior a estatística G.

Um ponto importante que é possível observar, é de que nos cenários com precisão baixa (II e IV), o teste baseado na estatística W apresentou as maiores taxas de rejeição sob a hipótese alternativa, no que tange o teste R_θ , conforme pode ser visto nas tabelas 11 e 13.

Tabela 2 – Taxa de rejeição da hipótese nula (%) para a o cenário I.

Teste R_μ										
Regressores adicionadas	Testes	$n = 50$			$n = 100$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$\hat{\eta}^2$	<i>RV</i>	1,57	6,46	12,05	1,25	5,76	11,45	1,24	5,23	10,88
	<i>S</i>	1,07	5,16	10,59	1,01	5,16	10,80	1,16	4,97	10,53
	<i>W</i>	2,36	7,93	13,65	1,42	6,39	12,20	1,36	5,46	11,23
	<i>G</i>	1,18	5,80	11,59	1,07	5,45	11,22	1,19	5,07	10,71
$\hat{\eta}^3$	<i>RV</i>	1,65	6,47	12,30	1,25	5,76	11,24	1,29	5,43	10,94
	<i>S</i>	1,06	5,16	10,80	1,05	5,22	10,54	1,18	5,14	10,55
	<i>W</i>	2,42	7,81	13,74	1,57	6,40	11,94	1,40	5,59	11,39
	<i>G</i>	1,31	5,98	11,65	1,10	5,51	10,96	1,22	5,26	10,81
$\hat{\mu}^2$	<i>RV</i>	1,67	6,73	12,12	1,25	5,84	10,91	1,28	5,56	10,87
	<i>S</i>	0,99	5,40	10,85	1,07	5,32	10,42	1,19	5,31	10,54
	<i>W</i>	2,41	8,08	13,54	1,53	6,40	11,63	1,41	5,80	11,21
	<i>G</i>	1,20	6,04	11,53	1,12	5,58	10,67	1,24	5,43	10,75
$\hat{\mu}^3$	<i>RV</i>	1,65	6,94	12,21	1,17	5,78	11,13	1,24	5,64	10,90
	<i>S</i>	0,90	5,48	10,79	0,99	5,12	10,32	1,09	5,41	10,58
	<i>W</i>	2,50	8,11	13,67	1,52	6,36	11,89	1,37	5,91	11,33
	<i>G</i>	1,15	6,21	11,57	1,03	5,46	10,68	1,14	5,55	10,76
Teste R_θ										
Regressores adicionadas	Testes	$n = 50$			$n = 100$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$(\hat{\eta}^2, \hat{\kappa}^2)$	<i>RV</i>	1,99	8,14	14,27	1,48	6,15	11,64	1,22	5,48	10,62
	<i>G</i>	1,52	6,68	12,89	1,22	5,73	11,11	1,06	5,24	10,39
$(\hat{\eta}^2, \hat{\eta}^2)$	<i>RV</i>	1,57	6,93	12,85	1,30	5,50	11,14	1,14	5,50	10,47
	<i>G</i>	1,03	5,96	11,84	1,03	5,17	10,57	1,01	5,38	10,31
$(\hat{\mu}^2, \hat{\phi}^2)$	<i>RV</i>	2,26	8,54	15,08	1,52	6,26	11,82	1,37	5,51	10,89
	<i>G</i>	1,46	6,81	13,44	1,19	5,69	11,30	1,19	5,31	10,65
$(\hat{\eta}^2, \hat{\mu}^2)$	<i>RV</i>	1,59	6,95	13,17	1,27	5,80	11,23	1,15	5,43	10,58
	<i>G</i>	1,01	5,98	12,00	1,01	5,36	10,69	1,03	5,30	10,36
$(\hat{\eta}^2, \hat{\eta}^3)$	<i>RV</i>	1,54	6,94	12,91	1,28	5,73	11,19	1,15	5,46	10,52
	<i>G</i>	0,99	6,12	11,86	1,03	5,29	10,53	1,09	5,33	10,25

Tabela 3 – Taxa de rejeição da hipótese nula (%) para a o cenário II.

Teste R_μ										
Regressores adicionadas	Testes	$n = 50$			$n = 100$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$\hat{\eta}^2$	<i>RV</i>	1,40	6,06	11,57	1,39	5,76	10,93	1,13	5,30	10,26
	<i>S</i>	0,80	4,64	9,91	1,13	5,31	10,27	1,05	5,08	10,07
	<i>W</i>	2,11	7,51	13,39	1,79	6,25	11,63	1,25	5,59	10,66
	<i>G</i>	1,03	5,36	11,00	1,22	5,52	10,48	1,03	5,18	10,25
$\hat{\eta}^3$	<i>RV</i>	1,42	6,06	11,58	1,34	5,85	10,61	1,07	5,28	10,47
	<i>S</i>	0,84	4,72	9,95	1,11	5,31	10,02	1,00	5,08	10,10
	<i>W</i>	2,04	7,36	13,39	1,78	6,39	11,19	1,21	5,61	10,72
	<i>G</i>	1,07	5,45	11,01	1,17	5,56	10,47	1,05	5,16	10,38
$\hat{\mu}^2$	<i>RV</i>	1,42	6,02	11,71	1,34	5,68	10,77	1,14	5,29	10,50
	<i>S</i>	0,88	4,79	10,06	1,11	5,18	10,15	0,99	5,24	10,14
	<i>W</i>	2,09	7,30	13,48	1,70	6,30	11,37	1,23	5,63	10,83
	<i>G</i>	1,06	5,45	11,09	1,18	5,39	10,55	1,06	5,30	10,34
$\hat{\mu}^3$	<i>RV</i>	1,36	6,01	11,71	1,40	5,60	10,86	1,10	5,32	10,56
	<i>S</i>	0,86	4,92	10,09	1,08	4,92	10,25	0,95	5,15	10,17
	<i>W</i>	2,08	7,21	13,35	1,58	6,24	11,33	1,20	5,65	10,94
	<i>G</i>	1,02	5,50	11,20	1,17	5,32	10,67	1,03	5,27	10,46
Teste R_θ										
Regressores adicionadas	Testes	$n = 50$			$n = 100$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$(\hat{\eta}^2, \hat{\kappa}^2)$	<i>RV</i>	2,10	7,48	14,32	1,59	6,58	12,21	1,18	5,48	10,69
	<i>G</i>	1,63	6,58	12,81	1,40	6,15	11,64	1,18	5,37	10,52
$(\hat{\eta}^2, \hat{\eta}^2)$	<i>RV</i>	1,51	7,25	13,20	1,47	6,60	11,86	1,19	5,71	11,00
	<i>G</i>	1,11	6,21	12,23	1,19	6,02	11,21	1,09	5,47	10,74
$(\hat{\mu}^2, \hat{\phi}^2)$	<i>RV</i>	2,00	7,68	14,34	1,46	6,42	11,91	1,15	5,47	10,83
	<i>G</i>	1,66	6,82	12,90	1,25	5,92	11,28	1,13	5,21	10,69
$(\hat{\eta}^2, \hat{\mu}^2)$	<i>RV</i>	1,60	7,28	13,09	1,45	6,41	12,06	1,20	5,87	11,05
	<i>G</i>	1,25	6,21	11,89	1,14	5,84	11,26	1,08	5,60	10,79
$(\hat{\eta}^2, \hat{\eta}^3)$	<i>RV</i>	1,58	7,27	13,15	1,48	6,47	12,02	1,12	5,88	11,07
	<i>G</i>	1,20	6,17	11,93	1,17	5,90	11,18	1,08	5,50	10,73

Tabela 4 – Taxa de rejeição da hipótese nula (%) para a o cenário III.

Teste R_μ										
Regressores adicionadas	Testes	$n = 50$			$n = 100$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$\hat{\eta}^2$	<i>RV</i>	1,34	5,99	11,69	1,20	5,42	10,63	0,98	5,13	10,20
	<i>S</i>	0,85	5,05	10,33	1,00	5,02	9,92	0,83	4,90	9,80
	<i>W</i>	1,99	7,43	13,18	1,53	5,95	11,41	1,10	5,35	10,57
	<i>G</i>	0,98	5,51	11,16	1,09	5,25	10,42	0,90	5,05	10,09
$\hat{\eta}^3$	<i>RV</i>	1,34	6,21	11,36	1,28	5,65	10,66	1,01	5,13	10,18
	<i>S</i>	0,89	5,18	10,06	1,05	5,21	10,06	0,84	4,89	9,88
	<i>W</i>	2,18	7,34	12,94	1,52	6,10	11,50	1,08	5,35	10,51
	<i>G</i>	1,00	5,63	10,81	1,09	5,40	10,40	0,94	5,01	10,06
$\hat{\mu}^2$	<i>RV</i>	1,62	6,49	11,88	1,31	5,92	11,22	1,23	5,14	10,59
	<i>S</i>	1,02	5,21	10,44	1,02	5,35	10,62	1,13	4,86	10,17
	<i>W</i>	2,47	7,95	13,28	1,56	6,54	11,99	1,36	5,44	10,90
	<i>G</i>	1,17	5,68	11,12	1,10	5,60	10,94	1,17	5,00	10,36
$\hat{\mu}^3$	<i>RV</i>	1,61	6,72	11,90	1,35	6,07	11,41	1,27	5,41	10,30
	<i>S</i>	0,96	5,43	10,64	1,06	5,39	10,67	1,14	5,17	9,86
	<i>W</i>	2,46	7,90	13,60	1,61	6,83	12,02	1,43	5,67	10,61
	<i>G</i>	1,07	5,97	11,20	1,11	5,66	11,03	1,18	5,24	10,11
Teste R_θ										
Regressores adicionadas	Testes	$n = 50$			$n = 100$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$(\hat{\eta}^2, \hat{\kappa}^2)$	<i>RV</i>	2,13	8,68	14,75	1,36	6,34	11,93	1,00	5,58	10,90
	<i>G</i>	1,73	7,46	13,45	1,13	5,84	11,47	0,91	5,32	10,66
$(\hat{\eta}^2, \hat{\eta}^2)$	<i>RV</i>	1,50	6,31	12,18	1,03	5,51	10,91	1,02	5,48	10,57
	<i>G</i>	1,14	5,39	11,16	0,88	5,08	10,33	0,95	5,30	10,33
$(\hat{\mu}^2, \hat{\phi}^2)$	<i>RV</i>	2,51	9,38	15,56	1,47	6,52	12,29	1,25	5,71	10,86
	<i>G</i>	1,56	7,56	13,94	1,15	6,86	11,47	1,08	5,29	10,55
$(\hat{\eta}^2, \hat{\mu}^2)$	<i>RV</i>	1,69	6,94	12,65	1,17	5,89	11,25	1,17	5,96	11,22
	<i>G</i>	1,05	5,62	11,29	0,91	5,29	10,75	0,97	5,59	10,76
$(\hat{\eta}^2, \hat{\eta}^3)$	<i>RV</i>	1,58	6,41	11,99	1,12	5,53	11,07	1,03	5,48	10,81
	<i>G</i>	1,13	5,38	11,05	0,87	5,17	10,44	0,91	5,19	10,56

Tabela 5 – Taxa de rejeição da hipótese nula (%) para a o cenário IV.

Teste R_μ										
Regressores adicionadas	Testes	$n = 50$			$n = 100$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$\hat{\eta}^2$	<i>RV</i>	1,65	6,90	12,85	1,30	6,09	11,59	1,22	5,17	10,24
	<i>S</i>	1,04	5,54	11,16	1,02	5,31	11,03	1,01	4,91	9,85
	<i>W</i>	2,45	8,51	14,50	1,70	6,92	12,45	1,34	5,50	10,63
	<i>G</i>	1,20	6,19	12,20	1,15	5,70	11,33	1,11	5,11	10,11
$\hat{\eta}^3$	<i>RV</i>	1,58	6,85	12,96	1,26	6,14	11,64	1,17	5,23	10,26
	<i>S</i>	1,06	5,48	11,32	1,13	5,21	11,00	1,05	4,97	9,86
	<i>W</i>	2,41	8,50	14,59	1,64	6,71	12,41	1,31	5,55	10,69
	<i>G</i>	1,20	6,35	12,35	1,13	5,72	11,15	1,10	5,10	10,21
$\hat{\mu}^2$	<i>RV</i>	1,50	6,86	12,75	1,21	5,95	11,21	1,17	5,31	10,45
	<i>S</i>	1,06	5,69	11,02	0,91	5,30	10,57	1,10	5,08	10,07
	<i>W</i>	2,31	8,41	14,11	1,57	6,50	11,87	1,27	5,66	10,84
	<i>G</i>	1,09	6,41	11,97	1,05	5,59	10,90	1,11	5,17	10,29
$\hat{\mu}^3$	<i>RV</i>	1,50	6,99	12,55	1,30	5,82	10,97	1,20	5,39	10,35
	<i>S</i>	0,99	5,38	10,97	0,93	5,42	10,19	1,11	5,12	10,14
	<i>W</i>	2,33	8,44	14,07	1,67	6,42	11,69	1,36	5,80	10,81
	<i>G</i>	1,10	6,21	11,96	1,10	5,44	10,65	1,09	5,26	10,22
Teste R_θ										
Regressores adicionadas	Testes	$n = 50$			$n = 100$			$n = 200$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$(\hat{\eta}^2, \hat{\kappa}^2)$	<i>RV</i>	2,14	8,02	14,46	1,41	6,35	12,02	1,24	5,41	10,69
	<i>G</i>	1,71	6,93	13,19	1,18	5,85	11,46	1,24	5,28	10,39
$(\hat{\eta}^2, \hat{\eta}^2)$	<i>RV</i>	2,04	7,85	13,93	1,41	6,08	11,83	0,92	5,33	10,75
	<i>G</i>	1,37	6,83	12,87	1,20	5,44	11,25	0,83	5,11	10,32
$(\hat{\mu}^2, \hat{\phi}^2)$	<i>RV</i>	2,29	8,48	14,78	1,35	6,10	11,96	1,33	5,54	10,66
	<i>G</i>	1,72	7,36	13,60	1,10	5,63	11,41	1,23	5,38	10,52
$(\hat{\eta}^2, \hat{\mu}^2)$	<i>RV</i>	2,02	8,01	14,45	1,52	6,31	12,22	1,04	5,44	11,23
	<i>G</i>	1,69	6,76	12,83	1,13	5,48	11,35	0,83	5,11	10,66
$(\hat{\eta}^2, \hat{\eta}^3)$	<i>RV</i>	2,03	7,82	13,90	1,41	6,15	11,67	0,95	5,41	10,84
	<i>G</i>	1,44	6,81	12,67	1,17	5,50	11,02	0,85	5,22	10,55

Tabela 6 – Taxa de rejeição sob a hipótese alternativa (função de ligação incorreta), cenário I.

Erro de especificação no submodelo da média										
Teste R_μ										
Regressores adicionadas	Testes	n = 50			n = 100			n = 200		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$\hat{\eta}^2$	<i>RV</i>	94,08	98,72	99,48	99,99	100,00	100,00	100,00	100,00	100,00
	<i>S</i>	94,77	98,92	99,62	99,99	100,00	100,00	100,00	100,00	100,00
	<i>W</i>	91,69	98,44	99,41	99,99	100,00	100,00	100,00	100,00	100,00
	<i>G</i>	94,68	98,79	99,55	99,99	100,00	100,00	100,00	100,00	100,00
$\hat{\eta}^3$	<i>RV</i>	51,90	78,26	87,07	90,62	97,82	99,13	99,90	100,00	100,00
	<i>S</i>	53,01	79,91	87,90	91,16	98,02	99,21	99,91	100,00	100,00
	<i>W</i>	48,75	75,72	85,81	89,95	97,60	99,11	99,90	100,00	100,00
	<i>G</i>	50,94	78,39	87,08	90,61	97,82	99,13	99,90	100,00	100,00
$\hat{\mu}^2$	<i>RV</i>	80,74	93,97	97,15	99,33	99,95	99,98	100,00	100,00	100,00
	<i>S</i>	83,94	94,72	97,45	99,47	99,95	99,98	100,00	100,00	100,00
	<i>W</i>	75,42	92,59	96,68	99,22	99,93	99,97	100,00	100,00	100,00
	<i>G</i>	82,12	94,33	97,26	99,43	99,95	99,98	100,00	100,00	100,00
$\hat{\mu}^3$	<i>RV</i>	69,63	88,42	94,05	97,87	99,65	99,90	100,00	100,00	100,00
	<i>S</i>	74,02	89,72	94,66	98,06	99,68	99,92	100,00	100,00	100,00
	<i>W</i>	63,78	86,90	93,36	97,38	99,63	99,88	100,00	100,00	100,00
	<i>G</i>	71,58	89,13	94,27	97,98	99,68	99,91	100,00	100,00	100,00
Erro de especificação nos submodelos da média e da precisão										
Teste R_θ										
Regressores adicionadas	Testes	n = 50			n = 100			n = 200		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$(\hat{\eta}^2, \hat{\kappa}^2)$	<i>RV</i>	32,86	58,31	70,78	74,51	90,28	94,71	98,73	99,66	99,80
	<i>G</i>	33,18	59,95	72,00	76,40	90,78	94,94	98,85	99,71	99,80
$(\hat{\eta}^2, \hat{\eta}^2)$	<i>RV</i>	36,68	59,90	71,95	76,15	91,37	95,25	98,88	99,76	99,92
	<i>G</i>	37,47	60,95	72,18	77,30	91,61	95,30	99,00	99,79	99,93
$(\hat{\mu}^2, \hat{\phi}^2)$	<i>RV</i>	19,25	42,00	56,12	54,18	77,99	86,58	92,44	98,29	99,16
	<i>G</i>	21,16	43,90	57,15	56,61	78,88	87,02	93,02	98,37	99,20
$(\hat{\eta}^2, \hat{\mu}^2)$	<i>RV</i>	35,33	59,82	71,73	76,92	91,01	95,00	98,86	99,76	99,94
	<i>G</i>	37,89	61,15	72,33	77,54	91,45	95,16	98,90	99,79	99,94
$(\hat{\eta}^2, \hat{\eta}^3)$	<i>RV</i>	36,75	60,08	71,73	76,82	91,05	95,14	98,77	99,77	99,95
	<i>G</i>	38,97	61,36	72,74	77,70	91,48	95,24	98,93	99,79	99,95

Tabela 7 – Taxa de rejeição sob a hipótese alternativa (função de ligação incorreta), cenário II.

Erro de especificação no submodelo da média										
Teste R_μ										
Regressores adicionadas	Testes	n = 50			n = 100			n = 200		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$\hat{\eta}^2$	<i>RV</i>	19,00	40,58	53,00	44,91	68,30	79,39	83,22	94,36	97,26
	<i>S</i>	26,19	46,22	56,65	49,60	71,56	81,32	85,50	95,08	97,47
	<i>W</i>	13,14	34,34	48,40	37,17	65,12	77,11	79,39	93,33	96,83
	<i>G</i>	22,65	43,11	54,85	47,58	69,72	80,09	84,61	94,72	97,29
$\hat{\eta}^3$	<i>RV</i>	5,53	16,87	26,37	12,73	28,99	42,22	30,54	53,85	65,60
	<i>S</i>	6,59	18,22	27,59	13,06	30,07	43,23	32,06	54,89	65,44
	<i>W</i>	4,68	15,61	25,27	11,68	28,07	41,48	29,94	53,06	65,10
	<i>G</i>	5,56	17,18	26,32	12,69	28,91	42,29	30,58	54,13	65,72
$\hat{\mu}^2$	<i>RV</i>	13,14	30,81	42,48	30,01	55,06	67,18	66,33	84,63	91,31
	<i>S</i>	16,54	34,59	45,98	35,05	58,70	69,62	69,83	86,17	92,31
	<i>W</i>	9,90	27,54	39,76	25,38	52,20	65,01	62,48	83,35	90,50
	<i>G</i>	14,20	31,92	43,58	31,33	56,12	67,81	67,06	84,95	91,55
$\hat{\mu}^3$	<i>RV</i>	10,34	26,34	37,29	24,21	87,42	60,18	55,84	77,03	85,31
	<i>S</i>	13,54	29,03	40,64	27,80	50,92	62,31	59,80	78,84	86,63
	<i>W</i>	7,86	23,60	34,85	20,08	45,23	58,27	52,30	75,77	84,29
	<i>G</i>	11,37	27,07	38,12	25,21	48,48	60,59	56,69	77,43	85,60
Erro de especificação nos submodelos da média e da precisão										
Teste R_θ										
Regressores adicionadas	Testes	n = 50			n = 100			n = 200		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$(\hat{\eta}^2, \hat{\kappa}^2)$	<i>RV</i>	7,91	22,63	34,05	22,43	43,30	56,75	56,11	76,88	84,63
	<i>G</i>	9,44	25,36	35,90	23,19	44,97	58,07	57,42	78,17	85,23
$(\hat{\eta}^2, \hat{\eta}^2)$	<i>RV</i>	7,20	18,11	26,80	22,57	44,05	56,56	56,16	76,25	84,70
	<i>G</i>	8,19	18,98	27,08	24,59	45,20	57,82	58,14	77,22	85,26
$(\hat{\mu}^2, \hat{\phi}^2)$	<i>RV</i>	0,28	0,88	1,38	14,56	32,98	46,08	39,45	63,40	74,05
	<i>G</i>	0,33	0,99	1,45	15,50	34,45	47,59	40,46	64,49	74,36
$(\hat{\eta}^2, \hat{\mu}^2)$	<i>RV</i>	9,67	23,27	34,97	23,14	44,92	57,48	55,45	75,99	84,83
	<i>G</i>	10,38	25,15	36,51	25,36	46,23	58,17	57,28	76,95	85,22
$(\hat{\eta}^2, \hat{\eta}^3)$	<i>RV</i>	10,19	24,31	35,70	23,41	45,37	57,53	56,06	76,96	85,01
	<i>G</i>	10,87	25,89	37,22	25,71	46,79	58,40	57,93	77,59	85,64

Tabela 8 – Taxa de rejeição sob a hipótese alternativa (função de ligação incorreta), cenário III.

Erro de especificação no submodelo da média										
Teste R_μ										
Regressores adicionadas	Testes	n = 50			n = 100			n = 200		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$\hat{\eta}^2$	<i>RV</i>	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	<i>S</i>	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	<i>W</i>	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	<i>G</i>	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
$\hat{\eta}^3$	<i>RV</i>	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	<i>S</i>	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	<i>W</i>	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	<i>G</i>	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
$\hat{\mu}^2$	<i>RV</i>	99,90	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	<i>S</i>	99,94	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	<i>W</i>	99,75	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	<i>G</i>	99,94	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
$\hat{\mu}^3$	<i>RV</i>	98,87	99,90	99,99	100,00	100,00	100,00	100,00	100,00	100,00
	<i>S</i>	99,25	99,94	99,99	100,00	100,00	100,00	100,00	100,00	100,00
	<i>W</i>	98,17	99,84	99,98	100,00	100,00	100,00	100,00	100,00	100,00
	<i>G</i>	99,32	99,94	99,99	100,00	100,00	100,00	100,00	100,00	100,00
Erro de especificação nos submodelos da média e da precisão										
Teste R_θ										
Regressores adicionadas	Testes	n = 50			n = 100			n = 200		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$(\hat{\eta}^2, \hat{\kappa}^2)$	<i>RV</i>	94,63	98,74	99,30	99,96	99,98	99,98	100,00	100,00	100,00
	<i>G</i>	94,68	98,81	99,30	99,95	99,97	99,97	100,00	100,00	100,00
$(\hat{\eta}^2, \hat{\eta}^2)$	<i>RV</i>	96,02	99,27	99,68	99,98	100,00	100,00	100,00	100,00	100,00
	<i>G</i>	96,29	99,29	99,71	99,98	100,00	100,00	100,00	100,00	100,00
$(\hat{\mu}^2, \hat{\phi}^2)$	<i>RV</i>	68,33	89,03	94,29	99,13	99,90	99,96	100,00	100,00	100,00
	<i>G</i>	74,14	90,61	94,94	99,32	99,91	99,96	100,00	100,00	100,00
$(\hat{\eta}^2, \hat{\mu}^2)$	<i>RV</i>	95,45	99,13	99,65	99,98	100,00	100,00	100,00	100,00	100,00
	<i>G</i>	96,73	99,28	99,73	99,98	100,00	100,00	100,00	100,00	100,00
$(\hat{\eta}^2, \hat{\eta}^3)$	<i>RV</i>	96,11	99,25	99,71	99,98	100,00	100,00	100,00	100,00	100,00
	<i>G</i>	96,56	99,29	99,75	99,98	100,00	100,00	100,00	100,00	100,00

Tabela 9 – Taxa de rejeição sob a hipótese alternativa (função de ligação incorreta), cenário IV.

Erro de especificação no submodelo da média										
Teste R_μ										
Regressores adicionadas	Testes	n = 50			n = 100			n = 200		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$\hat{\eta}^2$	<i>RV</i>	74,14	88,72	93,50	98,16	99,67	99,89	100,00	100,00	100,00
	<i>S</i>	79,76	91,30	94,91	98,83	99,81	99,92	100,00	100,00	100,00
	<i>W</i>	63,25	85,28	92,25	97,24	99,55	99,87	100,00	100,00	100,00
	<i>G</i>	76,49	89,84	94,10	98,38	99,73	99,89	100,00	100,00	100,00
$\hat{\eta}^3$	<i>RV</i>	66,12	84,15	90,71	96,41	99,24	99,70	99,99	100,00	100,00
	<i>S</i>	72,35	87,23	92,30	97,36	99,48	99,79	100,00	100,00	100,00
	<i>W</i>	57,05	80,79	88,83	94,83	98,87	99,64	99,98	100,00	100,00
	<i>G</i>	68,52	85,22	91,24	96,55	99,29	99,73	99,99	100,00	100,00
$\hat{\mu}^2$	<i>RV</i>	46,93	69,62	80,19	86,05	95,47	97,99	99,58	99,96	100,00
	<i>S</i>	53,38	73,94	82,63	89,63	96,33	98,24	99,64	99,96	100,00
	<i>W</i>	39,48	64,97	77,56	82,28	94,80	97,67	99,52	99,95	99,98
	<i>G</i>	48,68	70,77	80,79	86,98	95,66	98,05	99,60	99,96	100,00
$\hat{\mu}^3$	<i>RV</i>	34,40	58,64	70,15	73,71	89,77	94,61	97,94	99,57	99,79
	<i>S</i>	41,85	63,90	73,19	78,74	91,03	95,28	98,41	99,64	99,82
	<i>W</i>	28,83	53,68	67,05	69,61	88,41	94,12	97,32	99,30	99,76
	<i>G</i>	36,83	59,94	71,12	74,87	90,30	94,80	98,10	99,58	99,80
Erro de especificação nos submodelos da média e da precisão										
Teste R_θ										
Regressores adicionadas	Testes	n = 50			n = 100			n = 200		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$(\hat{\eta}^2, \hat{\kappa}^2)$	<i>RV</i>	30,29	49,87	58,13	84,61	94,34	96,98	99,48	99,90	99,93
	<i>G</i>	31,99	52,11	59,78	86,04	94,92	97,37	99,49	99,90	99,92
$(\hat{\eta}^2, \hat{\eta}^2)$	<i>RV</i>	40,97	64,73	75,95	84,15	94,77	97,01	99,74	99,95	99,99
	<i>G</i>	43,17	65,86	76,77	85,58	95,19	97,20	99,77	99,95	99,99
$(\hat{\mu}^2, \hat{\phi}^2)$	<i>RV</i>	21,69	43,49	57,01	60,72	80,79	88,34	93,77	98,54	99,34
	<i>G</i>	23,26	45,90	58,60	61,61	81,68	88,72	94,05	98,49	99,28
$(\hat{\eta}^2, \hat{\mu}^2)$	<i>RV</i>	39,44	65,04	75,53	83,70	94,64	97,15	99,71	99,95	99,98
	<i>G</i>	38,42	66,64	77,63	86,25	95,46	97,48	99,75	99,95	99,98
$(\hat{\eta}^2, \hat{\eta}^3)$	<i>RV</i>	39,82	64,98	75,74	84,06	94,82	97,06	99,75	99,95	99,99
	<i>G</i>	42,24	66,88	77,32	85,74	95,16	97,34	99,77	99,95	99,99

Tabela 11 – Taxa de rejeição sob a hipótese alternativa (omissão de variável importante), cenário II.

Erro de especificação no submodelo da média										
Teste R_μ										
Regressores adicionadas	Testes	n = 50			n = 100			n = 200		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$\hat{\eta}^2$	<i>RV</i>	85,77	96,30	98,36	99,68	99,98	99,99	100,00	100,00	100,00
	<i>S</i>	69,38	90,98	95,46	98,54	99,85	99,98	100,00	100,00	100,00
	<i>W</i>	87,06	96,62	98,53	99,73	99,99	99,99	100,00	100,00	100,00
	<i>G</i>	81,37	94,78	97,49	99,53	99,98	99,99	100,00	100,00	100,00
$\hat{\eta}^3$	<i>RV</i>	86,58	96,40	98,52	99,71	99,98	99,99	100,00	100,00	100,00
	<i>S</i>	77,76	94,05	97,32	99,33	99,98	99,99	100,00	100,00	100,00
	<i>W</i>	87,56	96,60	98,58	99,73	99,99	99,99	100,00	100,00	100,00
	<i>G</i>	83,96	95,86	98,22	99,65	99,98	99,99	100,00	100,00	100,00
$\hat{\mu}^2$	<i>RV</i>	85,58	95,99	98,14	99,65	99,98	99,99	100,00	100,00	100,00
	<i>S</i>	76,27	94,11	97,57	99,29	99,96	99,99	100,00	100,00	100,00
	<i>W</i>	86,04	96,10	98,30	99,67	99,98	99,99	100,00	100,00	100,00
	<i>G</i>	83,08	95,47	97,98	99,56	99,97	99,99	100,00	100,00	100,00
$\hat{\mu}^3$	<i>RV</i>	83,23	95,18	97,78	99,51	99,96	99,99	100,00	100,00	100,00
	<i>S</i>	74,34	93,41	97,13	99,12	99,95	99,99	100,00	100,00	100,00
	<i>W</i>	83,64	95,31	97,89	99,58	99,96	99,99	100,00	100,00	100,00
	<i>G</i>	80,91	94,81	97,65	99,43	99,95	99,99	100,00	100,00	100,00
Erro de especificação nos submodelos da média e da precisão										
Teste R_θ										
Regressores adicionadas	Testes	n = 50			n = 100			n = 200		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$(\hat{\eta}^2, \hat{\kappa}^2)$	<i>RV</i>	27,92	55,11	68,98	70,28	87,66	93,52	98,10	99,62	99,83
	<i>G</i>	21,36	51,22	65,99	64,52	86,18	92,87	97,74	99,58	99,82
$(\hat{\eta}^2, \hat{\eta}^2)$	<i>RV</i>	32,55	56,47	69,32	69,82	88,13	93,62	98,22	99,69	99,88
	<i>G</i>	27,45	53,66	67,88	67,84	87,27	93,31	98,08	99,69	99,85
$(\hat{\mu}^2, \hat{\phi}^2)$	<i>RV</i>	16,77	40,17	55,88	50,30	74,08	84,02	90,05	97,44	98,84
	<i>G</i>	14,68	39,12	54,85	48,56	73,50	83,88	89,34	97,41	98,80
$(\hat{\eta}^2, \hat{\mu}^2)$	<i>RV</i>	31,41	55,43	68,93	70,15	88,42	93,91	98,13	99,69	99,89
	<i>G</i>	25,15	53,45	67,95	68,04	87,67	93,46	97,94	99,65	99,85
$(\hat{\eta}^2, \hat{\eta}^3)$	<i>RV</i>	31,69	55,83	68,88	69,96	88,44	93,65	98,12	99,71	99,89
	<i>G</i>	25,55	53,48	67,70	67,89	87,62	93,25	97,91	99,66	99,84

Tabela 13 – Taxa de rejeição sob a hipótese alternativa (omissão de variável importante), cenário IV.

Erro de especificação no submodelo da média										
Teste R_μ										
Regressores adicionadas	Testes	n = 50			n = 100			n = 200		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$\hat{\eta}^2$	<i>RV</i>	81,98	94,45	97,55	99,43	99,96	99,98	100,00	100,00	100,00
	<i>S</i>	44,79	68,84	78,91	88,19	96,92	98,53	99,99	100,00	100,00
	<i>W</i>	82,19	94,88	97,94	99,59	99,96	99,99	100,00	100,00	100,00
	<i>G</i>	64,87	83,81	90,19	96,99	99,48	99,75	100,00	100,00	100,00
$\hat{\eta}^3$	<i>RV</i>	81,77	94,31	97,54	99,45	99,96	99,98	100,00	100,00	100,00
	<i>S</i>	62,81	86,73	93,17	97,47	99,68	99,89	100,00	100,00	100,00
	<i>W</i>	83,24	95,09	97,89	99,62	99,96	99,98	100,00	100,00	100,00
	<i>G</i>	76,41	91,80	96,28	99,08	99,91	99,96	100,00	100,00	100,00
$\hat{\mu}^2$	<i>RV</i>	79,34	92,99	96,65	99,14	99,86	99,97	100,00	100,00	100,00
	<i>S</i>	69,31	89,68	95,16	98,48	99,77	99,89	100,00	100,00	100,00
	<i>W</i>	81,35	93,98	97,26	99,37	99,93	99,98	100,00	100,00	100,00
	<i>G</i>	78,00	92,41	96,50	99,08	99,86	99,96	100,00	100,00	100,00
$\hat{\mu}^3$	<i>RV</i>	76,40	91,56	95,64	98,78	99,78	99,92	100,00	100,00	100,00
	<i>S</i>	68,07	88,95	94,13	97,84	99,62	99,87	100,00	100,00	100,00
	<i>W</i>	79,39	92,50	96,26	99,04	99,84	99,95	100,00	100,00	100,00
	<i>G</i>	75,89	91,26	95,58	98,62	99,78	99,92	100,00	100,00	100,00
Erro de especificação nos submodelos da média e da precisão										
Teste R_θ										
Regressores adicionadas	Testes	n = 50			n = 100			n = 200		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
$(\hat{\eta}^2, \hat{\kappa}^2)$	<i>RV</i>	83,05	95,41	97,95	99,86	99,97	99,98	99,99	99,99	99,99
	<i>G</i>	66,29	89,81	95,01	99,25	99,93	99,97	99,99	99,99	99,99
$(\hat{\eta}^2, \hat{\eta}^2)$	<i>RV</i>	83,56	94,93	97,79	99,82	99,98	100,00	100,00	100,00	100,00
	<i>G</i>	70,17	88,85	94,56	99,14	99,93	99,97	100,00	100,00	100,00
$(\hat{\mu}^2, \hat{\phi}^2)$	<i>RV</i>	80,31	93,63	97,15	99,65	99,96	99,98	100,00	100,00	100,00
	<i>G</i>	76,22	92,91	96,86	99,59	99,95	99,97	100,00	100,00	100,00
$(\hat{\eta}^2, \hat{\mu}^2)$	<i>RV</i>	81,98	94,81	97,57	99,77	99,98	100,00	100,00	100,00	100,00
	<i>G</i>	64,05	89,42	94,67	99,16	99,93	99,97	100,00	100,00	100,00
$(\hat{\eta}^2, \hat{\eta}^3)$	<i>RV</i>	82,53	94,96	97,74	99,81	99,98	100,00	100,00	100,00	100,00
	<i>G</i>	69,25	89,55	94,64	99,14	99,93	99,97	100,00	100,00	100,00

6 Aplicação

Neste capítulo apresentamos duas aplicações a dados reais utilizando os testes de especificação propostos neste estudo.

6.1 Dados do Milho

A primeira aplicação é referente a um experimento, descrito por GRIFFITHS; HILL; JUDGE (1993) e apresentado por BOURGUIGNON; SANTOS-NETO; CASTRO (2018). A variável dependente é a produção de milho e as variáveis independentes são as combinações de nitrogênio (x_{1i}) e fosfato (x_{2i}), $i = 1, \dots, n$.

A Tabela 14 apresenta as estimativas de máxima verossimilhança dos parâmetros do modelo de regressão BP escolhido utilizando a função de ligação logarítmica para os submodelos da média μ e do parâmetro de precisão ϕ . O ajuste considerado para o modelo é o mesmo apresentado por BOURGUIGNON; SANTOS-NETO; CASTRO (2018), a saber

$$\log(\mu_i) = \beta_0 + \beta_1 \log(x_{1i}) + \beta_2 \log(x_{2i}) \quad \text{e} \quad \log(\phi_i) = v_0 + v_1 x_{2i}, \quad i = 1, \dots, 30.$$

Tabela 14 – Estimativa dos parâmetros para os dados do milho.

Parâmetro	Estimativa	Erro Padrão	P-Valor
Submodelo para μ			
β_0	0,5207	0,2788	0,0739
β_1	0,3506	0,0330	9,3e-11
β_2	0,3990	0,0423	1,0e-09
Submodelo para ϕ			
v_0	2,7027	0,6650	0,0004
v_1	0,0072	0,0034	0,0432

É possível observar que os coeficientes são estatisticamente significativos. Notamos uma relação positiva entre a variável resposta com o nitrogênio, assim como com o fosfato. O nosso propósito consiste em testar se o modelo utilizado está corretamente especificado. Para isso, são utilizados os testes propostos nesse estudo. Foram considerados o teste baseado nas estatísticas S e G para o teste R_μ baseado em $\hat{\eta}^2$ e o teste considerando a estatística G para o teste R_θ baseado em $(\hat{\eta}^2, \hat{\eta}^2)$. O teste considerando as estatísticas S e G para R_μ são, respectivamente, 0,00001 e 0,0270 e o teste baseado na estatística G para o teste R_θ é 1,3474. Em ambos os testes a hipótese de que o modelo está corretamente especificado não é rejeitada ao nível de significância de 5%.

A Figura 1 apresenta o envelope simulado para o resíduo quantílico proposto por DUNN; SMYTH (1996), que se baseiam no teorema da inversa da função de distribuição acumulada. É possível verificar que os resíduos, em geral, permanecem dentro das bandas de confiança dos envelopes simulados e dessa forma não há indícios de afastamento da suposição de que o modelo de regressão beta prime é adequado para os dados.

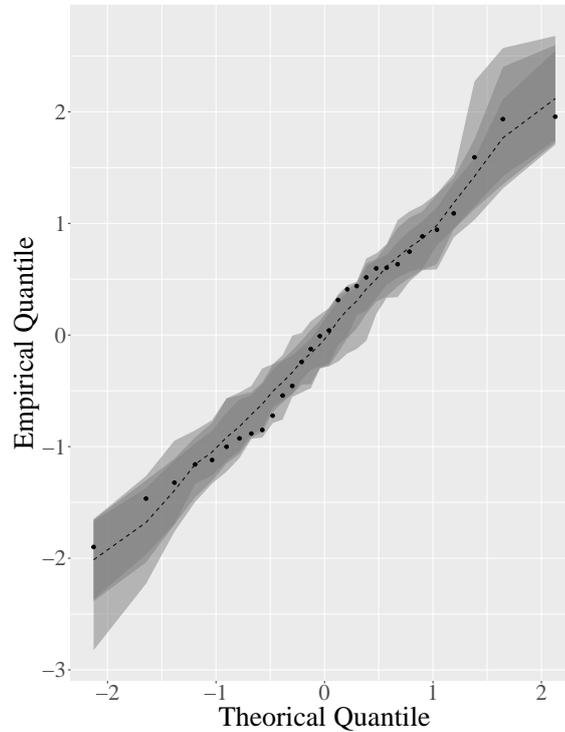


Figura 1 – Envelope simulado dos resíduos quantílicos para os dados do milho.

6.2 Dados de Cerejeiras

Os dados da segunda aplicação referem-se a uma amostra de dados de 31 cerejeiras em uma floresta do estado da Pensilvânia, EUA (RYAN; JOINER, 1994). As variáveis independentes dizem respeito a altura (h) e diâmetro (d). A variável resposta é o volume. O modelo ajustado para os dados é dado por

$$\log(\mu_i) = \beta_0 + \beta_1 \log(h) + \beta_2 \log(d) \quad \text{e} \quad \log(\phi_i) = \nu_0, \quad i = 1, \dots, n.$$

A Tabela 15 exibe as estimativas de máxima verossimilhança dos parâmetros do modelo de regressão *BP* ajustado, utilizando a função de ligação logarítmica no submodelo da média μ e no submodelo da precisão ϕ .

É possível notar que os coeficientes são estatisticamente significativas e que há uma relação positiva entre a variável resposta e a altura, como também há uma relação positiva da variável resposta e o diâmetro. Utilizando os testes propostos nesse estudo e fazendo o uso do teste considerando as estatísticas S e G para o teste R_μ , baseado em

Tabela 15 – Estimativa dos parâmetros dos dados de cerejeiras.

Parâmetro	Estimativa	Erro Padrão	P-Valor
Submodelo para μ			
β_0	-6,5705	0,7656	3,4e-09
β_1	1,9858	0,0719	0.2e-16
β_2	1,1019	0,1953	5,5e-06
Submodelo para ϕ			
ν_0	5,1361	0,2567	0.2e-16

$\hat{\eta}^2$, obtivemos os valores 0,4808 e 0,4806. Adicionalmente, considerando a estatística G para o teste R_θ , baseado em $(\hat{\eta}^2, \hat{\eta}^2)$, obtivemos o valor 0,9544. Portanto, a especificação correta do modelo não é rejeitada ao nível de 5% de significância em ambos os testes, R_μ e R_θ .

Na Figura 2, apresentamos o envelope simulado dos resíduos quantílicos. Podemos verificar que a especificação do modelo está correta, pois os resíduos em sua grande parte se encontram dentro do envelope, ou seja não há violação grave das premissas distributivas.

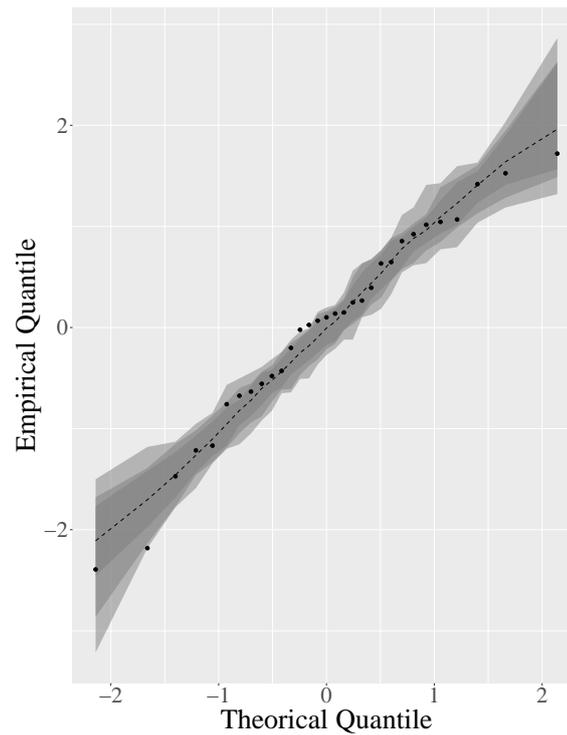


Figura 2 – Envelope simulado dos resíduos quantílicos para os dados de cerejeiras.

7 Considerações Finais

7.1 Conclusão

Neste trabalho, foram propostos dois testes de erro de especificação para o modelo de regressão *BP* com dispersão variável. No primeiro teste, a variável de teste é adicionada apenas ao submodelo da média e no segundo teste a variável é adicionada aos dois submodelos. Os testes foram implementados considerando as estatísticas, a saber: razão de verossimilhança, escore, wald e gradiente. Os resultados apresentados foram obtidos por meio de simulação de Monte Carlo em amostras finitas. Foram considerados como erro de especificação a função de ligação incorreta e a omissão de variável importante.

Os resultados apresentados sugerem que o tamanho empírico dos testes está muito próximo dos níveis de significância indicados (1%, 5% e 10%) a medida que o tamanho da amostra aumenta. Evidências numéricas mostram que os testes são capazes de detectar a especificação incorreta do modelo, principalmente o teste R_μ utilizando como variável de teste o quadrado do preditor linear estimado. Indica-se o uso das estatísticas escore e gradiente para testar a correta especificação do modelo.

Em termos práticos, os testes propostos podem ser bastante úteis. Dessa forma, recomendamos os testes propostos a todos usuários que desejam avaliar se o modelo de regressão *BP* está ou não bem especificado.

7.2 Publicação

Este trabalho gerou um artigo intitulado "A misspecification test for beta prime regression models" que foi submetido a revista *Communications in Statistics - Simulation and Computation*, e encontra-se sob revisão.

Referências

- BOURGUIGNON, M.; SANTOS-NETO, M.; CASTRO, M. de. A new regression model for positive data. arXiv preprint arXiv:1804.07734, 2018. Citado 6 vezes nas páginas 6, 7, 12, 16, 18 e 39.
- CANTERLE, D. R.; BAYER, F. M. Testes de especificação para a função de ligação em modelos lineares generalizados para dados binários. *Ciência e Natura*, Universidade Federal de Santa Maria, v. 37, n. 1, p. 1–11, 2015. Citado na página 13.
- DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, Taylor & Francis, v. 5, n. 3, p. 236–244, 1996. Citado na página 40.
- GODFREY, L. G.; ORME, C. D. The sensitivity of some general checks to omitted variables in the linear model. *International Economic Review*, JSTOR, p. 489–506, 1994. Citado na página 12.
- GOODFREY, L. Misspecification tests in econometrics: the lagrange multiplier principle and other approaches. *Econometric Society Monographs*, n. 16, 1988. Citado na página 12.
- GRIFFITHS, W. E.; HILL, R. C.; JUDGE, G. G. *Learning and practicing econometrics*. John Wiley and Sons, Inc. New York, USA, 1993. Citado na página 39.
- HOROWITZ, J. L. Bootstrap-based critical values for the information matrix test. *Journal of Econometrics*, Elsevier, v. 61, n. 2, p. 395–411, 1994. Citado 2 vezes nas páginas 12 e 13.
- KEEPING, E. *Introduction to statistical inference*, vol. 26. Dover Publications. com, 1962. Citado 2 vezes nas páginas 12 e 15.
- LIMA, L. B. d. Um teste de especificação correta em modelos de regressão beta. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2007. Citado na página 13.
- MANTALOS, P.; SHUKUR, G. The robustness of the reset test to non-normal error terms. *Computational Economics*, Springer, v. 30, n. 4, p. 393–408, 2007. Citado na página 13.
- MCDONALD, J. B. Some generalized functions for the size distribution of income. *Econometrica*, v. 52, p. 647–663, 1984. Citado 2 vezes nas páginas 12 e 15.
- MCDONALD, J. B. Model selection: some generalized distributions. *Communications in Statistics-Theory and Methods*, Taylor & Francis, v. 16, n. 4, p. 1049–1074, 1987. Citado na página 15.
- MCDONALD, J. B.; BUTLER, R. J. Regression models for positive random variables. *Journal of Econometrics*, Elsevier, v. 43, n. 1-2, p. 227–251, 1990. Citado na página 18.

- NELDER, J. A.; WEDDERBURN, R. W. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, Wiley Online Library, v. 135, n. 3, p. 370–384, 1972. Citado na página 12.
- NEYMAN, J.; PEARSON, E. S. On the use and interpretation of certain test criteria for purposes of statistical inference: Part i. *Biometrika*, JSTOR, p. 175–240, 1928. Citado na página 20.
- OLIVEIRA, J. S. C. d. Detectando má especificação em regressão beta. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2013. Citado na página 13.
- PEREIRA, T. L.; CRIBARI-NETO, F. Detecting model misspecification in inflated beta regressions. *Communications in Statistics-Simulation and Computation*, Taylor & Francis, v. 43, n. 3, p. 631–656, 2014. Citado na página 13.
- PETERS, S. On the use of the reset test in microeconomic models. *Applied Economics Letters*, Taylor & Francis, v. 7, n. 6, p. 361–365, 2000. Citado na página 13.
- R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria, 2021. Disponível em: <<https://www.R-project.org/>>. Citado 2 vezes nas páginas 19 e 23.
- RAMSEY, J.; GILBERT, R. A monte carlo study of some small sample properties of tests for specification error. *Journal of the American Statistical Association*, Taylor & Francis Group, v. 67, n. 337, p. 180–186, 1972. Citado na página 12.
- RAMSEY, J. B. Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, v. 31, n. 2, p. 350–371, 1969. Citado 5 vezes nas páginas 6, 7, 12, 13 e 20.
- RAO, C. R. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. In: CAMBRIDGE UNIVERSITY PRESS. *Mathematical Proceedings of the Cambridge Philosophical Society*. [S.l.], 1948. v. 44, n. 1, p. 50–57. Citado na página 20.
- RYAN, B. F.; JOINER, B. L. *Minitab handbook*. [S.l.]: Duxbury Resource Center, 1994. Citado na página 40.
- SANTOS, J.; CRIBARI-NETO, F. Hypothesis testing in log-birnbaum-saunders regressions. *Communications in Statistics-Simulation and Computation*, Taylor & Francis, v. 46, n. 5, p. 3990–4003, 2017. Citado na página 13.
- SAPRA, S. et al. A regression error specification test (reset) for generalized linear models. *Economics Bulletin*, AccessEcon, v. 3, n. 1, p. 1–6, 2005. Citado na página 13.
- SHUKUR, G.; EDGERTON, D. The small sample properties of the reset test as applied to systems of equations. *Journal of Statistical Computation and Simulation*, Taylor & Francis, v. 72, n. 12, p. 909–924, 2002. Citado na página 13.
- STASINOPOULOS, M.; ENEA, M.; RIGBY, R. A. Zero adjusted distributions on the positive real line. 2017. Citado na página 19.

TERRELL, G. R. The gradient statistic. *Computing Science and Statistics*, v. 34, n. 34, p. 206–215, 2002. Citado na página 20.

THURSBY, J. G.; SCHMIDT, P. Some properties of tests for specification error in a linear regression model. *Journal of the American Statistical Association*, Taylor & Francis, v. 72, n. 359, p. 635–641, 1977. Citado na página 12.

TULUPYEV, A. et al. Beta prime regression with application to risky behavior frequency screening. *Statistics in medicine*, Wiley Online Library, v. 32, n. 23, p. 4044–4056, 2013. Citado na página 18.

VARGAS, T. M.; FERRARI, S. L.; LEMONTE, A. J. Improved likelihood inference in generalized linear models. *Computational Statistics & Data Analysis*, Elsevier, v. 74, p. 110–124, 2014. Citado na página 21.

WALD, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical society*, JSTOR, v. 54, n. 3, p. 426–482, 1943. Citado na página 20.

A Apêndice

A.1 Vetor Escore

Primeiramente mostraremos como foram obtidos os elementos do vetor escore para essa classe de modelos. A função escore é obtida pela diferenciação da função de log-verossimilhança (3.2) com respeito ao β_j e \mathbf{v}_k e são dados por

$$\begin{aligned}
 U_j &= \frac{\partial \ell(\beta, \mathbf{v})}{\partial \beta_j} = \sum_{t=1}^n \frac{\partial \ell(\mu_t, \phi_t)}{\partial \mu_t} \frac{d\mu_t}{d\eta_{1t}} \frac{\partial \eta_{1t}}{\partial \beta_j} \\
 &= \sum_{t=1}^n [(1 + \phi_t) \log y_t - (1 + \phi_t) \log(1 + y_t) - (1 + \phi_t) \psi(\mu_t(1 + \phi_t)) \\
 &\quad + (1 + \phi_t) \psi(\mu_t(1 + \phi_t) + \phi_t + 2)] \frac{d\mu_t}{d\eta_{1t}} x_{tj} \\
 &= \sum_{t=1}^n (1 + \phi_t) [\log(\frac{y_t}{(1 + y_t)}) - \psi(\mu_t(1 + \phi_t) + \phi_t + 2)] \frac{d\mu_t}{d\eta_{1t}} x_{tj} \\
 &= \sum_{t=1}^n (1 + \phi_t) (y_t^* - \mu_t^*) \frac{d\mu_t}{d\eta_{1t}} x_{tj},
 \end{aligned}$$

e

$$\begin{aligned}
 U_k &= \frac{\partial \ell(\beta, \mathbf{v})}{\partial \mathbf{v}_k} = \sum_{t=1}^n \frac{\partial \ell(\mu_t, \phi_t)}{\partial \phi_t} \frac{d\phi_t}{d\eta_{2t}} \frac{\partial \eta_{2t}}{\partial \mathbf{v}_k} \\
 &= \sum_{t=1}^n [\mu_t \log y_t - (\mu_t + 1) \log(1 + y_t) - \mu_t \psi(\mu_t(1 + \phi_t)) - \psi(\phi_t + 2) \\
 &\quad + (\mu_t + 1) \psi(\mu_t(1 + \phi_t) + \phi_t + 2)] \frac{d\phi_t}{d\eta_{2t}} z_{tk} \\
 &= \sum_{t=1}^n [\log(\frac{y_t^{\mu_t}}{(1 + y_t)^{(\mu_t + 1)}}) - \mu_t (\psi(\mu_t(1 + \phi_t)) - \psi(\mu_t(1 + \phi_t) + \phi_t + 2)) - \psi(\phi_t + 2) \\
 &\quad + \psi(\mu_t(1 + \phi_t) + \phi_t + 2)] \frac{d\phi_t}{d\eta_{2t}} z_{tk} \\
 &= \sum_{t=1}^n [y_t^* - \mu_t (\mu_t^* - \frac{\gamma}{\mu_t})] \frac{d\phi_t}{d\eta_{2t}} z_{tk} \\
 &= \sum_{t=1}^n (y_t^* - \mu_t^*) \frac{d\phi_t}{d\eta_{2t}} z_{tk},
 \end{aligned}$$

em que, $\psi(\cdot) = \frac{d \log \Gamma(\cdot)}{d \cdot}$ é a função digamma, $y_t^* = \log(\frac{y_t}{(1 + y_t)})$, $\mu_t^* = \psi(\mu_t(1 + \phi_t)) - \psi(\mu_t(1 +$

$\phi_t) + \phi_t + 2)$, $y_t^* = \log\left(\frac{y_t^{\mu_t}}{(1+y_t)^{(\mu_t+1)}}\right)$ e $\mu_t^* = \mu_t(\mu_t^* - \frac{\gamma}{\mu_t})$ onde $\gamma = \psi(\phi_t + 2) - \psi(\mu_t(1 + \phi_t) + \phi_t + 2)$.

A função escore pode também ser escrita na forma matricial

$$U_{\beta} = X^{\top} \Phi D_1 (y^* - \mu^*),$$

$$U_{\nu} = Z^{\top} D_2 (y^* - \mu^*),$$

em que $\mathbf{D}_1 = \text{diag}[\frac{\partial \mu_1}{\partial \eta_{11}}, \dots, \frac{\partial \mu_n}{\partial \eta_{1n}}]$, $\mathbf{D}_2 = \text{diag}[\frac{\partial \phi_1}{\partial \eta_{21}}, \dots, \frac{\partial \phi_n}{\partial \eta_{2n}}]$, $\Phi = \text{diag}[(1 + \phi_1), \dots, (1 + \phi_n)]$, $\mathbf{y}^* = (y_1^*, \dots, y_n^*)^{\top}$, $\mathbf{y}^* = (y_1^*, \dots, y_n^*)^{\top}$, $\mu^* = (\mu_1^*, \dots, \mu_n^*)^{\top}$ e $\mu^* = (\mu_1^*, \dots, \mu_n^*)^{\top}$.

A.2 Matriz de Informação de Fisher

A matriz de informação de Fisher é obtida a partir da segunda derivada da função log-verossimilhança (3.2), que são expressos da forma a seguir:

1. para β_j e β_l , com $j, l = 1, \dots, m$, temos

$$U_{jl} = \frac{\partial^2 \ell(\beta, \nu)}{\partial \beta_j \partial \beta_l} = \sum_{t=1}^n \left\{ \frac{\partial^2 \ell(\mu_t, \phi_t)}{\partial \mu_t^2} \left(\frac{d\mu_t}{d\eta_{1t}} \right)^2 + \frac{\partial \ell(\mu_t, \phi_t)}{\partial \mu_t} \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_{1t}} \right) \frac{d\mu_t}{d\eta_{1t}} \right\}_{x_{tj} x_{tl}}$$

$$= \sum_{t=1}^n \left\{ -(1 + \phi_t)^2 [\psi^{(1)}(\mu_t(1 + \phi_t)) - \psi^{(1)}(\mu_t(1 + \phi_t) + \phi_t + 2)] \left(\frac{d\mu_t}{d\eta_{1t}} \right)^2 \right.$$

$$\left. + \frac{\partial \ell(\mu_t, \phi_t)}{\partial \mu_t} \left(\frac{\partial}{\partial \mu_t} \frac{d\mu_t}{d\eta_{1t}} \right) \frac{d\mu_t}{d\eta_{1t}} \right\}_{x_{tk} x_{tl}}.$$

Sob condições de regularidade, obtemos $E\left(\frac{\partial \ell(\mu_t, \phi_t)}{\partial \mu_t}\right) = 0$. Assim, temos que

$$K_{\beta\beta} = -E(U_{jl})$$

$$= \sum_{t=1}^n (1 + \phi_t)^2 [\psi^{(1)}(\mu_t(1 + \phi_t)) - \psi^{(1)}(\mu_t(1 + \phi_t) + \phi_t + 2)] \left(\frac{d\mu_t}{d\eta_{1t}} \right)^2_{x_{tj} x_{tl}}.$$

2. Para β_j e ν_k , com $j = 1, \dots, m$ e $k = 1, \dots, M$, temos

$$U_{jk} = \frac{\partial^2 \ell(\beta, \nu)}{\partial \beta_j \partial \nu_k} = \sum_{t=1}^n \frac{\partial^2 \ell(\mu_t, \phi_t)}{\partial \mu_t \partial \phi_t} \frac{d\mu_t}{d\eta_{1t}} \frac{d\phi_t}{d\eta_{2t}}_{x_{tj} z_{tk}}$$

$$= \sum_{t=1}^n \left\{ (y_t^* - \mu_t^*) + (1 + \phi_t) [\psi^{(1)}(\mu_t(1 + \phi_t) + \phi_t + 2) + \mu_t [\psi^{(1)}(\mu_t(1 + \phi_t) + \phi_t + 2) \right.$$

$$\left. - \psi^{(1)}(\mu_t(1 + \phi_t))] \right\} \frac{d\mu_t}{d\eta_{1t}} \frac{d\phi_t}{d\eta_{2t}}_{x_{tj} z_{tk}}.$$

Tomando o valor esperado sobre a função, temos que

$$\begin{aligned} K_{\beta v} &= -E(U_{jk}) \\ &= -\sum_{t=1}^n (1 + \phi_t) [\psi^{(1)}(\mu_t(1 + \phi_t) + \phi_t + 2) + \mu_t [\psi^{(1)}(\mu_t(1 + \phi_t) + \phi_t + 2) - \psi^{(1)}(\mu_t(1 + \phi_t))] \\ &\quad \frac{d\mu_t}{d\eta_{1t}} \frac{d\phi_t}{d\eta_{2t}} x_{tj} z_{tk}. \end{aligned}$$

3. Para v_k e v_p , com $k, p = 1, \dots, m$, temos

$$\begin{aligned} U_{kp} &= \frac{\partial^2 \ell(\beta, v)}{\partial v_k \partial v_p} = \sum_{t=1}^n \left\{ \frac{\partial^2 \ell(\mu_t, \phi_t)}{\partial \phi_t^2} \left(\frac{d\phi_t}{d\eta_{2t}} \right)^2 + \frac{\partial \ell(\mu_t, \phi_t)}{\partial \phi_t} \left(\frac{\partial}{\partial \phi_t} \frac{d\phi_t}{d\eta_{2t}} \right) \frac{d\phi_t}{d\eta_{2t}} \right\} z_{tk} z_{tp} \\ &= \sum_{t=1}^n \left\{ -[\mu_t^2 \psi^{(1)}(\mu_t(1 + \phi_t)) - (\mu_t + 1)^2 \psi^{(1)}(\mu_t(1 + \phi_t) + \phi_t + 2) + \psi^{(1)}(\phi_t + 2)] \left(\frac{d\phi_t}{d\eta_{2t}} \right)^2 \right. \\ &\quad \left. + \frac{\partial \ell(\mu_t, \phi_t)}{\partial \phi_t} \left(\frac{\partial}{\partial \phi_t} \frac{d\phi_t}{d\eta_{2t}} \right) \frac{d\phi_t}{d\eta_{2t}} \right\} z_{tk} z_{tp}. \end{aligned}$$

Sob condições de regularidade, obtemos $E\left(\frac{\partial \ell(\mu_t, \phi_t)}{\partial \phi_t}\right) = 0$. Assim, temos que

$$\begin{aligned} K_{vv} &= -E(U_{kp}) \\ &= \sum_{t=1}^n \left(\mu_t^2 \psi^{(1)}(\mu_t(1 + \phi_t)) - (\mu_t + 1)^2 \psi^{(1)}(\mu_t(1 + \phi_t) + \phi_t + 2) + \psi^{(1)}(\phi_t + 2) \right) \\ &\quad \left(\frac{d\phi_t}{d\eta_{2t}} \right)^2 z_{tk} z_{tp}. \end{aligned}$$

A matriz de informação de Fisher pode ser escrita na forma matricial

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{\beta\beta} & \mathbf{K}_{\beta v} \\ \mathbf{K}_{v\beta} & \mathbf{K}_{vv} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top \mathbf{E}_3 \mathbf{X} & \mathbf{X}^\top \mathbf{E}_5 \mathbf{Z} \\ \mathbf{Z}^\top \mathbf{E}_5 \mathbf{X} & \mathbf{Z}^\top \mathbf{E}_4 \mathbf{Z} \end{bmatrix},$$

em que cada elemento de E_3 , E_4 e E_5 são, respectivamente, dado por

$$e_{3i} = (1 + \phi_i)^2 [\Psi^{(1)}(\mu_i(1 + \phi_i)) - \Psi^{(1)}(\mu_i(1 + \phi_i) + \phi_i + 2)] \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right)^2,$$

$$e_{4i} = [\mu_i^2 \Psi^{(1)}(\mu_i(1 + \phi_i)) - (1 + \mu_i)^2 \Psi^{(1)}(\mu_i(1 + \phi_i) + \phi_i + 2) + \Psi^{(1)}(\phi_i + 2)] \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right)^2,$$

e

$$e_{5i} = -(1 + \phi_i) \{ \Psi^{(1)}(\mu_i(1 + \phi_i) + \phi_i + 2) + \mu_i [\Psi^{(1)}(\mu_i(1 + \phi_i) + \phi_i + 2) - \Psi^{(1)}(\mu_i(1 + \phi_i))] \} \left(\frac{\partial \mu_i}{\partial \eta_{1i}} \right) \left(\frac{\partial \phi_i}{\partial \eta_{2i}} \right).$$

B Implementação Computacional - R

```
rm(list = ls())

setwd("C:/Users/klebe/Dropbox/PIBIC/Script")
source("gamlss_BP.r")
source("glmBP.r")
source("Estimation.r")
source("simu_BP_v11.r")

# Pacotes -----

library(gamlss)
library(gamlss.dist)
library(gamlss.data)
library(splines)

# Simulação do Tamanho -----

set.seed(144)

n=50          # tamanho amostra
j=1          # indice para usar tamanhos maiores
NR=10000     # Numero de replicas de MC
pw=2        # potencia variavel de teste
n1=n*j      # Quando n for 100, n1=n*j=50*2

beta=c(1.9,-1.5) #valores reais vetor beta
x0=rep(1,n)
x01=runif(n)
x02=runif(n)

x0=rep(x0,j)
x01=rep(x01,j)
x02=rep(x02,j)
```

```
X=cbind(x0,x01) #matriz de regressores de mu
eta1=X%*%beta #preditor linear
mu = exp(eta1)
summary(mu)

lambda=c(1.8,-0.4)
z0=rep(1,n)
z01=runif(n)
z02=runif(n)

z0=rep(z0,j)
z01=rep(z01,j)
z02=rep(z02,j)
Z=cbind(z0,z01)
kappa=Z%*%lambda #preditor linear
phi = exp(kappa)
summary(phi)

Vmu=mu*(1+mu)
vary= Vmu/phi
summary(vary)

#Inicializando vetores para guardar estatisticas de teste
Etest1=rep(0,NR)
Etest2=rep(0,NR)
Etest3=rep(0,NR)
Etest4=rep(0,NR)

contC=0
contC1=0

# PRIMEIRO LAÇO DE MONTE CARLO

for(i in 1:NR) {

  y=rBP(n1,mu,phi)
  ystar=log(y/(1+y))
```

```
fit <- gamlss(y~(X-1),sigma.formula=~(Z-1),
             family = BP(mu.link = "log",sigma.link = "log"),
             trace=FALSE)

if(fit$conv==FALSE){

  i=i-1
  contC=contC+1
  next

}

# ADIÇÃO DAS NOVAS VARIÁVEIS DE TESTE

etatest=fit$mu.lp           #Preditor linear
mutest=fitted.values(fit) #Valor ajustado
x1=mutest^pw
x1=as.vector(x1)
Xnew=cbind(X,x1)           #Nova matriz de regressores de mu

fitnew <- gamlss(y~(Xnew-1),sigma.formula=~(Z-1),
                family = BP(mu.link = "log",sigma.link = "log"),
                trace=FALSE)

if (fitnew$conv==FALSE){

  i=i-1
  contC1=contC1+1
  next

}

# FUNÇÃO ESCORE VEROSSIMILHANÇA RESTRITA E INVERSA DE FISHER RESTRITA

mx1=2
mx2=1
mx3=3
```

```

zerox=rep(0,mx2)
betatil=c(fit$mu.coef, zerox)
etatil1=Xnew%%betatil
mutil=exp(etatil1)

ms1=2
ms2=1
lambdatil=c(fit$sigma.coef)
etatil2=Z%%lambdatil
phitil=exp(etatil2)

Phitil=diag(as.vector(1+phitil))
D1 <- diag(as.vector(exp(etatil1)))
ystartil=ystar
mustartil=digamma(mutil*(1+phitil))-digamma(mutil*(1+phitil)+phitil+2)
escorebetatil=t(Xnew)%%Phitil%%D1%%(ystartil-mustartil)

# INVERSA DE FISHER RESTRITA

deltatil=(1+phitil)
Deltatil=as.vector(deltatil)

psi1til=trigamma(mutil*(1+phitil))
psi2til=trigamma(mutil*(1+phitil)+phitil+2)
wtil=psi1til-psi2til
Wtil=as.vector(wtil)

psi3til=trigamma(phitil+2)
ctil=(mutil^2)*psi1til-((1+mutil)^2)*psi2til+psi3til
Ctil=as.vector(ctil)

dtil=psi2til+mutil*(psi2til-psi1til)
Dtil=as.vector(dtil)

e3til=(Deltatil^2)*Wtil*as.vector(exp(etatil1)^2)
e4til=Ctil*as.vector(exp(etatil2))^2
e5til=(-Deltatil)*Dtil*as.vector(exp(etatil1))*as.vector(exp(etatil2))

K.bbtil=t(Xnew)%%diag(as.vector(e3til))%%Xnew

```

```

K.bltil=t(Xnew)%*%diag(as.vector(e5til))%*%Z
K.lbttil=t(Z)%*%diag(as.vector(e5til))%*%Xnew
K.llttil=t(Z)%*%diag(as.vector(e4til))%*%Z

bloc.suptil=cbind(K.bbttil,K.bltil)
bloc.inftil=cbind(K.lbttil,K.llttil)

Fisher.bltil=rbind(bloc.suptil,bloc.inftil)

Inver.Fisher.bltil <- try(solve(Fisher.bltil))

# INVERSA DE FISHER IRRESTRITA

muhat=fitnew$mu.fv
phihat=fitnew$sigma.fv

eta1hat=fitnew$mu.lp
eta2hat=fitnew$sigma.lp

betahat=fitnew$mu.coefficients
lambdahat=fitnew$sigma.coefficients

deltahat=1+phihat
Delta=as.vector(deltahat)

psi1=trigamma(muhat*(1+phihat))
psi2=trigamma(muhat*(1+phihat)+phihat+2)
w=psi1-psi2
W=as.vector(w)

psi3=trigamma(phihat+2)
c=(muhat^2)*psi1-((1-muhat)^2)*psi2+psi3
C=as.vector(c)

d=psi2+muhat*(psi2-psi1)
D=as.vector(d)

e3= (Delta^2)*W*(as.vector(exp(eta1hat))^2)
e4= C*(as.vector(exp(eta2hat))^2)

```

```
e5= -Delta*D*as.vector(exp(eta1hat))*as.vector(exp(eta2hat))

K.bb=t(Xnew)%%diag(as.vector(e3))%*%Xnew
K.bl=t(Xnew)%%diag(as.vector(e5))%*%Z
K.lb=t(Z)%%diag(as.vector(e5))%*%Xnew
K.ll=t(Z)%%diag(as.vector(e4))%*%Z

bloc.sup=cbind(K.bb,K.bl)
bloc.inf=cbind(K.lb,K.ll)

Fisher.bl=rbind(bloc.sup,bloc.inf)

Inver.Fisher.bl <- try(solve(Fisher.bl))

# ESTATISTICAS DOS TESTES

# RAZAO DE VEROSSIMILHANÇA

lhat<-logLik(fitnew)
ltil<-logLik(fit)

Etest1[i]=2*(lhat[1]-ltil[1])

# ESCORE

Etest2[i]= t(escorebetatil[mx3])%*((Inver.Fisher.bltil[mx3,mx3])%*%
escorebetatil[mx3])

# WALD

Etest3[i]= t(betahat[mx3])%*((Inver.Fisher.bl[mx3,mx3])^(-1))%*%betahat[mx3]

# Gradiente

Etest4[i]= t(escorebetatil[mx3])%*%betahat[mx3]
```

```
qchisq90=qchisq(0.90,1)
qchisq95=qchisq(0.95,1)
qchisq99=qchisq(0.99,1)

# FIM DO LAÇO DE MONTE CARLO
}

print(contC)
print(contC1)
print(max(phi)/min(phi))
print(max(vary)/min(vary))

cont90RV=ifelse(Etest1 > qchisq90,1,0)
result_tam90RV=((sum(cont90RV))/NR)*100

cont90E=ifelse(Etest2 > qchisq90,1,0)
result_tam90E=((sum(cont90E))/NR)*100

cont90W=ifelse(Etest3 > qchisq90,1,0)
result_tam90W=((sum(cont90W))/NR)*100

cont90G=ifelse(Etest4 > qchisq90,1,0)
result_tam90G=((sum(cont90G))/NR)*100

cont95RV=ifelse(Etest1 > qchisq95,1,0)
result_tam95RV=((sum(cont95RV))/NR)*100

cont95E=ifelse(Etest2 > qchisq95,1,0)
result_tam95E=((sum(cont95E))/NR)*100

cont95W=ifelse(Etest3 > qchisq95,1,0)
result_tam95W=((sum(cont95W))/NR)*100

cont95G=ifelse(Etest4 > qchisq95,1,0)
result_tam95G=((sum(cont95G))/NR)*100
```

```
cont99RV=ifelse(Etest1 > qchisq99,1,0)
result_tam99RV=((sum(cont99RV))/NR)*100

cont99E=ifelse(Etest2 > qchisq99,1,0)
result_tam99E=((sum(cont99E))/NR)*100

cont99W=ifelse(Etest3 > qchisq99,1,0)
result_tam99W=((sum(cont99W))/NR)*100

cont99G=ifelse(Etest4 > qchisq99,1,0)
result_tam99G=((sum(cont99G))/NR)*100

print("resultRV")
print(result_tam90RV)
print(result_tam95RV)
print(result_tam99RV)

print("resultE")
print(result_tam90E)
print(result_tam95E)
print(result_tam99E)

print("resultW")
print(result_tam90W)
print(result_tam95W)
print(result_tam99W)

print("resultG")
print(result_tam90G)
print(result_tam95G)
print(result_tam99G)

# QUANTIS ESTIMADOS

quantilRV=quantile(Etest1, probs=c(.90,.95,.99))
quantilE=quantile(Etest2, probs=c(.90,.95,.99))
quantilW=quantile(Etest3, probs=c(.90,.95,.99))
```

```
quantilG=quantile(Etest4, probs=c(.90,.95,.99))

print("quantilRV")
print(quantilRV)

print("quantileE")
print(quantileE)

print("quantilW")
print(quantilW)

print("quantilG")
print(quantilG)

d<-date()
print(d)

resultRV=cbind(result_tam90RV,result_tam95RV,result_tam99RV)
resultE=cbind(result_tam90E,result_tam95E,result_tam99E)
resultW=cbind(result_tam90W,result_tam95W,result_tam99W)
resultG=cbind(result_tam90G,result_tam95G,result_tam99G)

result100=rbind(resultRV,resultE,resultW,resultG)
result100

# Simulação do Poder (FLI) -----

set.seed (144)

Etest1=rep(0,NR)
Etest2=rep(0,NR)
Etest3=rep(0,NR)
Etest4=rep(0,NR)

contC=0
contC1=0
```

```
beta=c(1.9,-1.5) #valores reais vetor beta
x0=rep(1,n)
x01=runif(n)
x02=runif(n)

x0=rep(x0,j)
x01=rep(x01,j)
x02=rep(x02,j)
X=cbind(x0,x01) #matriz de regressores de mu
eta1=X%*%beta
mu = (eta1)^2 # função de ligação raiz

for(i in 1:NR) {

  y=rBP(n1,mu,phi)
  ystar=log(y/(1+y))

  fit <- gamlss(y~(X-1),sigma.formula=~(Z-1),
               family = BP(mu.link = "log",sigma.link = "log"),
               trace=FALSE)

  if(fit$conv==FALSE){

    i=i-1
    contC=contC+1
    next

  }

  # ADIÇÃO DAS NOVAS VARIÁVEIS DE TESTE

  etatest=fit$mu.lp           #Preditor linear
  mutest=fitted.values(fit) #Valor ajustado
  x1=mutest^pw
  x1=as.vector(x1)
  Xnew=cbind(X,x1)          #Nova matriz de regressores de mu
```

```

fitnew <- gamlss(y~(Xnew-1),sigma.formula=~(Z-1),
               family = BP(mu.link = "log",sigma.link = "log"),
               trace=FALSE)

if (fitnew$conv==FALSE){

  i=i-1
  contC1=contC1+1
  next

}

# FUNÇÃO ESCORE VEROSSIMILHANÇA RESTRITA E INVERSA DE FISHER RESTRITA

mx1=2
mx2=1
mx3=3
zerox=rep(0,mx2)
betatil=c(fit$mu.coef, zerox)
etatil1=Xnew%%betatil
mutil=exp(etatil1)

ms1=2
ms2=1
lambdatil=c(fit$sigma.coef)
etatil2=Z%%lambdatil
phital=exp(etatil2)

Phital=diag(as.vector(1+phital))
D1 <- diag(as.vector(exp(etatil1)))
ystartil=ystar
mustartil=digamma(mutil*(1+phital))-digamma(mutil*(1+phital)+phital+2)
escorebetatil=t(Xnew)%%Phital%%D1%%(ystartil-mustartil)

# INVERSA DE FISHER RESTRITA

deltatil=(1+phital)
Deltatil=as.vector(deltatil)

```

```

psi1til=trigamma(mutil*(1+phitil))
psi2til=trigamma(mutil*(1+phitil)+phitil+2)
wtil=psi1til-psi2til
Wtil=as.vector(wtil)

psi3til=trigamma(phitil+2)
ctil=(mutil^2)*psi1til-((1+mutil)^2)*psi2til+psi3til
Ctil=as.vector(ctil)

dtil=psi2til+mutil*(psi2til-psi1til)
Dtil=as.vector(dtil)

e3til=(Deltatil^2)*Wtil*as.vector(exp(etatil1))^2
e4til=Ctil*as.vector(exp(etatil2))^2
e5til=(-Deltatil)*Dtil*as.vector(exp(etatil1))*as.vector(exp(etatil2))

K.bbttil=t(Xnew)%*%diag(as.vector(e3til))%*%Xnew
K.bltil=t(Xnew)%*%diag(as.vector(e5til))%*%Z
K.lbttil=t(Z)%*%diag(as.vector(e5til))%*%Xnew
K.lltil=t(Z)%*%diag(as.vector(e4til))%*%Z

bloc.suptil=cbind(K.bbttil,K.bltil)
bloc.inftil=cbind(K.lbttil,K.lltil)

Fisher.bltil=rbind(bloc.suptil,bloc.inftil)

Inver.Fisher.bltil <- try(solve(Fisher.bltil))

# INVERSA DE FISHER IRRESTRITA

muhat=fitnew$mu.fv
phihat=fitnew$sigma.fv

eta1hat=fitnew$mu.lp
eta2hat=fitnew$sigma.lp

betahat=fitnew$mu.coefficients

```

```

lambdahat=fitnew$sigma.coefficients

deltahat=1+phihat
Delta=as.vector(deltahat)

psi1=trigamma(muhat*(1+phihat))
psi2=trigamma(muhat*(1+phihat)+phihat+2)
w=psi1-psi2
W=as.vector(w)

psi3=trigamma(phihat+2)
c=(muhat^2)*psi1-((1-muhat)^2)*psi2+psi3
C=as.vector(c)

d=psi2+muhat*(psi2-psi1)
D=as.vector(d)

e3= (Delta^2)*W*(as.vector(exp(eta1hat)))^2
e4= C*(as.vector(exp(eta2hat))^2)
e5= -Delta*D*as.vector(exp(eta1hat))*as.vector(exp(eta2hat))

K.bb=t(Xnew)%*%diag(as.vector(e3))%*%Xnew
K.bl=t(Xnew)%*%diag(as.vector(e5))%*%Z
K.lb=t(Z)%*%diag(as.vector(e5))%*%Xnew
K.ll=t(Z)%*%diag(as.vector(e4))%*%Z

bloc.sup=cbind(K.bb,K.bl)
bloc.inf=cbind(K.lb,K.ll)

Fisher.bl=rbind(bloc.sup,bloc.inf)

Inver.Fisher.bl <- try(solve(Fisher.bl))

# ESTATISTICAS DOS TESTES

# RAZAO DE VEROSSIMILHANÇA

```

```
lhat<-logLik(fitnew)
ltil<-logLik(fit)

Etest1[i]=2*(lhat[1]-ltil[1])

# ESCORE

Etest2[i]= t(escorebetatil[mx3])%*%(Inver.Fisher.bltil[mx3,mx3])%*%
escorebetatil[mx3]

# WALD

Etest3[i]= t(betahat[mx3])%*%((Inver.Fisher.bl[mx3,mx3])^(-1))%*%betahat[mx3]

# Gradiente

Etest4[i]= t(escorebetatil[mx3])%*%betahat[mx3]

# FIM DO LAÇO DE MONTE CARLO
}

qchisqRV= quantilRV
qchisqE=quantile
qchisqW=quantilW
qchisqG=quantilG

print(contC)
print(contC1)
print(max(phi)/min(phi))
print(max(vary)/min(vary))
length(Etest1 < qchisqRV[1])

cont90RV=ifelse(Etest1 <= qchisqRV[1],1,0)
result_poder90RV=(1-(sum(cont90RV))/NR)*100

cont90E=ifelse(Etest2 <= qchisqE[1],1,0)
```

```
result_poder90E=(1-(sum(cont90E))/NR)*100

cont90W=ifelse(Etest3 <= qchisqW[1],1,0)
result_poder90W=(1-(sum(cont90W))/NR)*100

cont90G=ifelse(Etest4 <= qchisqG[1],1,0)
result_poder90G=(1-(sum(cont90G))/NR)*100

cont95RV=ifelse(Etest1 <= qchisqRV[2],1,0)
result_poder95RV=(1-(sum(cont95RV))/NR)*100

cont95E=ifelse(Etest2 <= qchisqE[2],1,0)
result_poder95E=(1-(sum(cont95E))/NR)*100

cont95W=ifelse(Etest3 <= qchisqW[2],1,0)
result_poder95W=(1-(sum(cont95W))/NR)*100

cont95G=ifelse(Etest4 <= qchisqG[2],1,0)
result_poder95G=(1-(sum(cont95G))/NR)*100

cont99RV=ifelse(Etest1 <= qchisqRV[3],1,0)
result_poder99RV=(1-(sum(cont99RV))/NR)*100

cont99E=ifelse(Etest2 <= qchisqE[3],1,0)
result_poder99E=(1-(sum(cont99E))/NR)*100

cont99W=ifelse(Etest3 <= qchisqW[3],1,0)
result_poder99W=(1-(sum(cont99W))/NR)*100

cont99G=ifelse(Etest4 <= qchisqG[3],1,0)
result_poder99G=(1-(sum(cont99G))/NR)*100

print("resultRV")
print(result_poder90RV)
print(result_poder95RV)
print(result_poder99RV)
```

```
print("resultE")
print(result_poder90E)
print(result_poder95E)
print(result_poder99E)

print("resultW")
print(result_poder90W)
print(result_poder95W)
print(result_poder99W)

print("resultG")
print(result_poder90G)
print(result_poder95G)
print(result_poder99G)

d<-Sys.time()
print(d)

resultPoderRV=cbind(result_poder90RV,result_poder95RV,result_poder99RV)
resultPoderE=cbind(result_poder90E,result_poder95E,result_poder99E)
resultPoderW=cbind(result_poder90W,result_poder95W,result_poder99W)
resultPoderG=cbind(result_poder90G,result_poder95G,result_poder99G)

resultPoder50=rbind(resultPoderRV,resultPoderE,resultPoderW,resultPoderG)
resultPoder50

# Simulação do Poder (OVI) -----

set.seed (144)

Etest1=rep(0,NR)
Etest2=rep(0,NR)
Etest3=rep(0,NR)
Etest4=rep(0,NR)

contC=0
contC1=0
```

```
betaovi=c(4,-3.5, 2.8) #valores reais vetor beta
x02=(x01)^2
Xovi=cbind(x0,x01,x02)#matriz de regressores de mu
etaovi=Xovi%%betaovi #preditor linear
muovi = exp(etaovi)
summary(muovi)

for(i in 1:NR) {

  y=rBP(n1,muovi,phi)
  ystar=log(y/(1+y))

  fit <- gamlss(y~(X-1),sigma.formula=~(Z-1),
               family = BP(mu.link = "log",sigma.link = "log"),
               trace=FALSE)

  if(fit$conv==FALSE){

    i=i-1
    contC=contC+1
    next

  }

  # ADIÇÃO DAS NOVAS VARIÁVEIS DE TESTE

  etatest=fit$mu.lp          #Preditor linear
  mutest=fitted.values(fit) #Valor ajustado
  x1=etatest^pw
  x1=as.vector(x1)
  Xnew=cbind(X,x1)          #Nova matriz de regressores de mu

  fitnew <- gamlss(y~(Xnew-1),sigma.formula=~(Z-1),
                  family = BP(mu.link = "log",sigma.link = "log"),
                  trace=FALSE)

  if (fitnew$conv==FALSE){
```

```

    i=i-1
    contC1=contC1+1
    next
}

# FUNÇÃO ESCORE VEROSSIMILHANÇA RESTRITA E INVERSA DE FISHER RESTRITA

mx1=2
mx2=1
mx3=3
zerox=rep(0,mx2)
betatil=c(fit$mu.coef, zerox)
etatil1=Xnew%%betatil
mutil=exp(etatil1)

ms1=2
ms2=1
lambdatil=c(fit$sigma.coef)
etatil2=Z%%lambdatil
phital=exp(etatil2)

Phital=diag(as.vector(1+phital))
D1 <- diag(as.vector(exp(etatil1)))
ystartil=ystar
mustartil=digamma(mutil*(1+phital))-digamma(mutil*(1+phital)+phital+2)
escorebetatil=t(Xnew)%%Phital%%D1%%(ystartil-mustartil)

# INVERSA DE FISHER RESTRITA

deltatil=1+as.vector(phital)
Deltatil=as.vector(deltatil)

psi1til=trigamma(mutil*(1+phital))
psi2til=trigamma(mutil*(1+phital)+phital+2)
wttil=psi1til-psi2til
Wtil=as.vector(wttil)

```

```

psi3til=trigamma(phitil+2)
ctil=(mutil^2)*psi1til-((1+mutil)^2)*psi2til+psi3til
Ctil=as.vector(ctil)

dtil=psi2til+mutil*(psi2til-psi1til)
Dtil=as.vector(dtil)

e3til=(Deltatil^2)*Wtil*as.vector(exp(etatil1))^2
e4til=Ctil*as.vector(exp(etatil2))^2
e5til=(-Deltatil)*Dtil*as.vector(exp(etatil1))*as.vector(exp(etatil2))

K.bbtil=t(Xnew)%*%diag(as.vector(e3til))%*%Xnew
K.bltil=t(Xnew)%*%diag(as.vector(e5til))%*%Z
K.lbtil=t(Z)%*%diag(as.vector(e5til))%*%Xnew
K.lltil=t(Z)%*%diag(as.vector(e4til))%*%Z

bloc.suptil=cbind(K.bbtil,K.bltil)
bloc.inftil=cbind(K.lbtil,K.lltil)

Fisher.bltil=rbind(bloc.suptil,bloc.inftil)

Inver.Fisher.bltil <- chol2inv(chol(Fisher.bltil))

# INVERSA DE FISHER IRRESTRITA

muhat=fitnew$mu.fv
phihat=fitnew$sigma.fv

eta1hat=fitnew$mu.lp
eta2hat=fitnew$sigma.lp

betahat=fitnew$mu.coefficients
lambdahat=fitnew$sigma.coefficients

deltahat=1+phihat
Delta=as.vector(deltahat)

psi1=trigamma(muhat*(1+phihat))
psi2=trigamma(muhat*(1+phihat)+phihat+2)

```

```

w=psi1-psi2
W=as.vector(w)

psi3=trigamma(phihat+2)
c=(muhat^2)*psi1-((1+muhat)^2)*psi2+psi3
C=as.vector(c)

d=psi2+muhat*(psi2-psi1)
D=as.vector(d)

e3= (Delta^2)*W*(as.vector(exp(eta1hat))^2)
e4= C*(as.vector(exp(eta2hat))^2)
e5= -Delta*D*as.vector(exp(eta1hat))*as.vector(exp(eta2hat))

K.bb=t(Xnew)%%diag(as.vector(e3))%*%Xnew
K.bl=t(Xnew)%%diag(as.vector(e5))%*%Z
K.lb=t(Z)%%diag(as.vector(e5))%*%Xnew
K.ll=t(Z)%%diag(as.vector(e4))%*%Z

bloc.sup=cbind(K.bb,K.bl)
bloc.inf=cbind(K.lb,K.ll)

Fisher.bl=rbind(bloc.sup,bloc.inf)

Inver.Fisher.bl <- chol2inv(chol(Fisher.bl))

# ESTATISTICAS DOS TESTES

# RAZAO DE VEROSSIMILHANÇA

lhat<-logLik(fitnew)
ltil<-logLik(fit)

Etest1[i]=2*(lhat[1]-ltil[1])

# ESCORE

```

```
Etest2[i]= t(escorebetatil[mx3])%*%(Inver.Fisher.bltil[mx3,mx3])%*%
  escorebetatil[mx3]

# WALD

Etest3[i]= t(betahat[mx3])%*%((Inver.Fisher.bl[mx3,mx3])^(-1))%*%betahat[mx3]

# Gradiente

Etest4[i]= t(escorebetatil[mx3])%*%betahat[mx3]

# FIM DO LAÇO DE MONTE CARLO
}

qchisqRV= quantilRV
qchisqE=quantile
qchisqW=quantilW
qchisqG=quantilG

print(contC)
print(contC1)
print(max(phi)/min(phi))
print(max(vary)/min(vary))

cont90RV=ifelse(Etest1 <= qchisqRV[1],1,0)
result_poder90RV=(1-(sum(cont90RV))/NR)*100

cont90E=ifelse(Etest2 <= qchisqE[1],1,0)
result_poder90E=(1-(sum(cont90E))/NR)*100

cont90W=ifelse(Etest3 <= qchisqW[1],1,0)
result_poder90W=(1-(sum(cont90W))/NR)*100

cont90G=ifelse(Etest4 <= qchisqG[1],1,0)
result_poder90G=(1-(sum(cont90G))/NR)*100
```

```
cont95RV=ifelse(Etest1 <= qchisqRV[2],1,0)
result_poder95RV=(1-(sum(cont95RV))/NR)*100
```

```
cont95E=ifelse(Etest2 <= qchisqE[2],1,0)
result_poder95E=(1-(sum(cont95E))/NR)*100
```

```
cont95W=ifelse(Etest3 <= qchisqW[2],1,0)
result_poder95W=(1-(sum(cont95W))/NR)*100
```

```
cont95G=ifelse(Etest4 <= qchisqG[2],1,0)
result_poder95G=(1-(sum(cont95G))/NR)*100
```

```
cont99RV=ifelse(Etest1 <= qchisqRV[3],1,0)
result_poder99RV=(1-(sum(cont99RV))/NR)*100
```

```
cont99E=ifelse(Etest2 <= qchisqE[3],1,0)
result_poder99E=(1-(sum(cont99E))/NR)*100
```

```
cont99W=ifelse(Etest3 <= qchisqW[3],1,0)
result_poder99W=(1-(sum(cont99W))/NR)*100
```

```
cont99G=ifelse(Etest4 <= qchisqG[3],1,0)
result_poder99G=(1-(sum(cont99G))/NR)*100
```

```
print("resultRV")
print(result_poder90RV)
print(result_poder95RV)
print(result_poder99RV)
```

```
print("resultE")
print(result_poder90E)
print(result_poder95E)
print(result_poder99E)
```

```
print("resultW")
print(result_poder90W)
print(result_poder95W)
```

```
print(result_poder99W)
```

```
print("resultG")
```

```
print(result_poder90G)
```

```
print(result_poder95G)
```

```
print(result_poder99G)
```

```
d<-Sys.time()
```

```
print(d)
```

```
resultPoderRV=cbind(result_poder90RV,result_poder95RV,result_poder99RV)
```

```
resultPoderE=cbind(result_poder90E,result_poder95E,result_poder99E)
```

```
resultPoderW=cbind(result_poder90W,result_poder95W,result_poder99W)
```

```
resultPoderG=cbind(result_poder90G,result_poder95G,result_poder99G)
```

```
resultPoder50=rbind(resultPoderRV,resultPoderE,resultPoderW,resultPoderG)
```

```
resultPoder50
```