



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PÓS-GRADUAÇÃO EM ESTATÍSTICA

DANIELE DE BRITO TRINDADE

MODELO DE REGRESSÃO BETA NÃO LINEAR COM ERROS
NAS VARIÁVEIS

Recife
2018

DANIELE DE BRITO TRINDADE

**MODELO DE REGRESSÃO BETA NÃO LINEAR COM ERROS
NAS VARIÁVEIS**

Tese apresentada ao programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para obtenção do grau de Doutor em Estatística.

Orientadora: Prof^a Dr^a Patrícia Leone Espinheira Ospina

Recife
2018

Catálogo na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

T833m Trindade, Daniele de Brito
Modelo de regressão beta não linear com erros nas variáveis / Daniele de Brito Trindade. – 2018.
257 f.: il., fig., tab.

Orientadora: Patrícia Leone Espinheira Ospina.
Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN, Estatística, Recife, 2018.
Inclui referências e apêndices.

1. Estatística. 2. Regressão beta. I. Ospina, Patrícia Leone Espinheira (orientadora). II. Título.

310

CDD (23. ed.)

UFPE- MEI 2018-039

DANIELE DE BRITO TRINDADE

MODELO DE REGRESSÃO BETA NÃO LINEAR COM ERROS NAS VARIÁVEIS

Tese apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Estatística.

Aprovada em: 21 de fevereiro de 2018.

BANCA EXAMINADORA

Prof.^a Patrícia Leone Espinheira Ospina
UFPE

Prof. Francisco Cribari Neto
UFPE

Prof. Aldo William Medina Garay
UFPE

Prof.^a Hildete Prisco Pinheiro
Unicamp

Prof. Jalmar Manuel Farfan Carrasco
UFBA

Dedico esta tese a minha mãe, Maria Terezinha, e a meu pai, Luiz Carlos (in memoriam), pelo amor incondicional e apoio em todos os momentos da minha vida.

Agradecimentos

Primeiramente, agradeço a Deus pela constante presença na minha vida, dando-me tranquilidade e equilíbrio para vencer todos os obstáculos que encontrei durante esta longa e árdua caminhada.

Agradeço também aos meus guias e aos meus Orixás por mostrarem nos mínimos detalhes a forte presença nos momentos mais difíceis da minha vida, *Olorum Modupé!*

Agradeço ao meu alicerce, minha mãe, Maria Terezinha Vieira de Brito, pelo apoio, pela força, pela superação da distância e principalmente pelas palavras de conforto.

A minha orientadora, amiga, mãe, confidente, Patrícia Espinheira Ospina pela orientação, apoio, paciência, compreensão e confiança depositada em mim durante a elaboração deste trabalho. Muito obrigada por acreditar em mim e dividir o seu conhecimento comigo.

Aos meus colegas e amigos de pós-graduação que me apoiaram e me ajudaram durante as dificuldades. Em especial, a Jéssica Rivas que esteve junto comigo desde a graduação superando todas as dificuldades e a Enai por me ajudar e apoiar no início desta caminhada. Agradeço também a Alisson, a Luana, a Eveline, por dividirem seus conhecimentos comigo e ajudar no desenvolvimento deste trabalho. A Cris, a Priscila e a Raabe que mesmo distantes estão presentes no meu coração.

Agradeço também ao meu colega e grande amigo Fábio que me deu muito apoio nos momentos que eu me senti sem direção. Obrigada por sentar, olhar, gastar o seu precioso tempo comigo e com o fruto deste trabalho. Você, sem dúvidas, foi fundamental para o desenvolvimento de muitas simulações e muitos ajustes de conjuntos de dados, além disso, foi um porto seguro.

Aos meus amigos Ana Clara Paixão, Danila Silva, Cleber Xavier, Lucas Santos, Nadja, Jaqueline Braga, Rosemeire Fiaccone, Nívea Bispo e Paula Xavier por me ajudarem nas horas mais nebulosas, me dando apoio e palavras de conforto e incentivo. Agradeço por muitas vezes acreditarem mais em mim do que eu mesma. Serei eternamente grata pelo apoio e amor que vocês têm por mim.

Agradeço também as minhas amigas pernambucanas Priscila Alburquerque, Xênia Caroline, Elizabeth da Silva e Luzia Vaz que me ajudaram a superar a distância da minha terra natal e me acolherem com a alma.

Aos meus professores da graduação e pós-graduação, por me ajudarem a superar as dificuldades e por dividir os seus conhecimentos comigo. Em especial agradeço à professora

Leila Denise Alves Ferreira Amorim, por toda confiança depositada em mim desde a graduação, por ter sido minha coorientadora no mestrado, por desenvolver trabalhos de pesquisas sólidos e importantes na área de saúde e por ser meu parâmetro profissional. Também agradeço aos professores Raydonal Ospina, Abraão Nascimento e Jalmar Carrasco por depositarem a sua confiança e dividir os seus conhecimentos comigo.

Agradeço a professora Rita Ribeiro, juntamente com a escola de Nutrição da UFBA, por disponibilizar a sua pesquisa e fornecer as bases de dados utilizadas nas aplicações desta tese. Obrigada por todo o carinho e disponibilidade. Sua ajuda foi fundamental para o desenvolvimento da minha tese.

A Valéria Bittencourt, por todo carinho, dedicação e gentileza. Muito obrigada pelas palavras de conforto e todo amor que teve por mim durante os sete anos de muita luta e dedicação vividos na UFPE.

Agradeço aos meus colegas do Instituto Federal Baiano - *campus* Guanambi, por todo apoio durante a finalização deste trabalho.

A Banca examinadora, pelo tempo disponível e sugestões para o melhoramento da tese.

A Universidade Federal da Bahia pelo espaço físico e materiais disponíveis para a execução do meu trabalho. Em especial à Dona Zezé, Cláudio e Artur.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo suporte financeiro concedido por meio da bolsa de doutorado.

Ao Instituto Federal Baiano - *campus* Guanambi pelo afastamento para o desenvolvimento e dedicação ao doutorado.

Sem sonhos, a vida não tem brilho. Sem metas, os sonhos não têm alicerces. Sem prioridades, os sonhos não se tornam reais. Sonhe, trace metas, estabeleça prioridades e corra riscos para executar seus sonhos. Melhor é errar por tentar do que errar por se omitir!

Augusto Cury

Resumo

Nesta tese generalizamos o modelo de regressão beta com erros de medidas (Carrasco, 2014) propondo um modelo com formas funcionais não lineares para os preditores. Foram avaliados três métodos de estimação denotados por máxima verossimilhança aproximada, máxima pseudo-verossimilhança aproximada e calibração da regressão. Para avaliação das metodologias de estimação realizamos um estudo de simulação de Monte Carlo sob diferentes cenários. Foi possível evidenciar que os métodos de máxima verossimilhança aproximada e máxima pseudo-verossimilhança aproximada tiveram melhor desempenho, em geral, quando comparados aos métodos de calibração da regressão e *naive*. Vale salientar também que os desempenhos dos estimadores associados à não linearidade e dispersão variável são afetados negativamente quando o coeficiente de confiabilidade diminui. Intervalos de confiança foram construídos com o objetivo de avaliar, através dos estudos de simulação de Monte Carlo, taxas de cobertura. Novos resíduos foram propostos para o modelo de regressão beta não linear com erros de medidas com o objetivo de verificar as suposições assumidas ao modelo e detectar pontos aberrantes. Também foram construídas medidas de predição e de qualidade de ajuste considerando os três métodos de estimação e tais medidas foram avaliadas através de estudos de simulação de Monte Carlo. Três aplicações considerando dados reais são analisadas. Utilizamos as linguagens de programação Ox (Doornik, 2011) e R como suportes computacionais.

Palavras-chave: Regressão beta. Erros de medida. Máxima verossimilhança aproximada. Máxima pseudo-verossimilhança aproximada. Simulação de Monte Carlo. Predição.

Abstract

In this thesis we generalize the beta regression model with measurement errors (Carrasco, 2014) proposing a model with nonlinear functional forms. Three estimation methods denoted by approximate maximum likelihood, approximate maximum pseudo likelihood, and regression calibration were evaluated. To evaluate the estimation methodologies, we performed a Monte Carlo simulation study on different scenarios. It was possible to show that the approximate maximum likelihood and approximate maximum pseudo-likelihood methods performed better, in general, when compared to the calibration regression and the naive methods. Besides that the performance of estimators associated with nonlinearity and variable dispersion are negatively affected when degree of reliability decreases. Confidence intervals were constructed through the Monte Carlo simulation studies with the objective of evaluating the coverage rates. New residuals were proposed for the nonlinear beta regression model with measurement errors in order to verify the assumptions over the model and to detect aberrant points. Prediction and quality of fit measures were also constructed considering the three estimation methods. Also, the methods were evaluated through Monte Carlo simulation studies. Three applications considering real data are analyzed. We use the programming languages Ox (Doornik, 2011) and R as computational supports.

Keywords: Beta regression. Measurement errors. Approximate maximum likelihood. Approximate maximum pseudo-likelihood. Monte Carlo simulation. Prediction.

Lista de Figuras

2.1	Vieses estimados dos estimadores de α_0 , α_1 e β_1 considerando os coeficientes de confiabilidade $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 1: Modelo com dispersão constante.	56
2.2	Vieses estimados do estimador de γ_0 considerando os coeficientes de confiabilidade $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Verdadeiros valores dos parâmetros: $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 1: Modelo com dispersão constante.	57
2.3	Vieses estimados dos estimadores de α_0 , α_1 e β_1 para $\phi = 16.44$, $\phi = 54.60$ e $\phi = 298.87$ com o coeficiente de confiabilidade $k_x = 0.50$ (erro de medida alto). Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 2: Modelo com dispersão constante.	61
2.4	Vieses estimados dos estimadores de γ_0 para $\phi = 16.44$, $\phi = 54.60$ e $\phi = 298.87$ com o coeficiente de confiabilidade $k_x = 0.50$ (erro de medida alto). Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 2: Modelo com dispersão constante.	62
2.5	Vieses estimados dos estimadores de α_0 , α_1 e β_1 para $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Valores verdadeiros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 3: Modelo com precisão variável.	65
2.6	Vieses estimados dos estimadores de γ_0 e λ_1 para $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Valores verdadeiros: $\gamma_0 = 2.5$, $\lambda_1 = 0.9$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 3: Modelo com precisão variável.	66
2.7	Vieses estimados dos estimadores de α_0 , α_1 e β_1 para diferentes pontos de quadratura $Q = 12$, $Q = 30$ e $Q = 80$ com o coeficiente de confiabilidade $k_x = 0,50$ (erro de medida alto). Valores verdadeiros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 4: Modelo com dispersão variável. . . .	70

2.8	Vieses estimados dos estimadores de γ_0 e λ_1 para diferentes pontos de quadratura $Q = 12$, $Q = 30$ e $Q = 80$ com o coeficiente de confiabilidade $k_x = 0,50$ (erro de medida alto). Valores verdadeiros: $\gamma_0 = 2,5$, $\lambda_1 = 0,9$, $\mu_x = 0,0$ e $\sigma_x^2 = 1,0$. Cenário 4: Modelo com dispersão variável.	71
2.9	Vieses estimados dos estimadores de α_0 , α_1 e β_1 para os coeficientes de confiabilidade $k_x = 0,95$ (erro de medida baixo), $k_x = 0,75$ (erro de medida moderado) e $k_x = 0,50$ (erro de medida alto). Valores verdadeiros: $\alpha_0 = 0,7$, $\alpha_1 = 2,0$, $\beta_1 = -1,5$, $\mu_x = 1,5$ e $\sigma_x^2 = 0,5$. Cenário 5: Modelo com não linearidade na dispersão.	75
2.10	Vieses estimados dos estimadores de γ_0 , γ_1 e λ_1 para os coeficientes de confiabilidade $k_x = 0,95$ (erro de medida baixo), $k_x = 0,75$ (erro de medida moderado) e $k_x = 0,50$ (erro de medida alto). Valores verdadeiros: $\gamma_0 = 1,5$, $\gamma_1 = 2,0$, $\lambda_1 = 1,3$, $\mu_x = 1,5$ e $\sigma_x^2 = 0,5$. Cenário 5: Modelo com não linearidade na dispersão.	76
3.1	Taxas de cobertura dos intervalos de 95% de confiança para os parâmetros α_0 , α_1 e β_1 do modelo de regressão beta não linear com erros de medida com dispersão contante. Coeficiente de confiabilidade: $k_x = 0,95$ (erro de medida baixo), $k_x = 0,75$ (erro de medida moderado) e $k_x = 0,50$ (erro de medida alto). Cenário 1.	84
3.2	Taxas de cobertura dos intervalos de 95% de confiança para o parâmetro γ_0 do modelo de regressão beta não linear com erros de medida com dispersão contante. Coeficiente de confiabilidade: $k_x = 0,95$ (erro de medida baixo), $k_x = 0,75$ (erro de medida moderado) e $k_x = 0,50$ (erro de medida baixo). Cenário 1.	85
3.3	Taxas de cobertura dos intervalos de 95% de confiança para os parâmetros γ_0 e λ_1 do modelo de regressão beta não linear com erros de medida com dispersão variável. Coeficiente de confiabilidade: $k_x = 0,95$ (erro de medida baixo), $k_x = 0,75$ (erro de medida moderado) e $k_x = 0,50$ (erro de medida baixo). Cenário 2.	90
3.4	Taxas de cobertura dos intervalos de 95% de confiança para os parâmetros α_0 , α_1 e β_1 do modelo de regressão beta não linear com erros de medida com dispersão variável. Coeficiente de confiabilidade: $k_x = 0,95$ (erro de medida baixo), $k_x = 0,75$ (erro de medida moderado) e $k_x = 0,50$ (erro de medida baixo). Cenário 2.	91

3.5	Taxas de cobertura dos intervalos de 95% de confiança para os parâmetros α_0 , α_1 e β_1 do modelo de regressão beta não linear com erros de medida com não linearidade na dispersão. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida baixo). Cenário 3.	95
3.6	Taxas de cobertura dos intervalos de 95% de confiança para os parâmetros γ_0 , γ_1 e λ_1 do modelo de regressão beta não linear com erros de medida com não linearidade na dispersão. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida baixo). Cenário 3.	96
4.1	Gráficos normais de probabilidade para os resíduos ponderados obtidos utilizando o método de máxima verossimilhança aproximada para estimar os parâmetros. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.	110
4.2	Gráficos normais de probabilidade para os resíduos ponderados padronizados obtidos utilizando o método de máxima verossimilhança aproximada para estimar os parâmetros. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.	111
4.3	Gráficos normais de probabilidade para os resíduos combinados obtidos utilizando o método de máxima verossimilhança aproximada para estimar os parâmetros. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.	113
4.4	Gráficos normais de probabilidade para os resíduos ponderados obtidos utilizando o método de máxima pseudo-verossimilhança aproximada para estimar os parâmetros. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.	116
4.5	Gráficos normais de probabilidade para os resíduos ponderados padronizados obtidos utilizando o método de máxima pseudo-verossimilhança aproximada para estimar os parâmetros. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.	117

4.6	Gráficos normais de probabilidade para os resíduos combinados obtidos utilizando o método de máxima pseudo-verossimilhança aproximada para estimar os parâmetros. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.	118
4.7	Gráficos normais de probabilidade para os resíduos ponderados obtidos utilizando o método da calibração da regressão para estimar os parâmetros. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.	121
4.8	Gráficos normais de probabilidade para os resíduos ponderados padronizados obtidos utilizando o método da calibração da regressão para estimar os parâmetros. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.	122
4.9	Gráficos normais de probabilidade para os resíduos combinados obtidos utilizando o método da calibração da regressão para estimar os parâmetros. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.	123
4.10	Gráficos normais de probabilidade para os resíduos ponderados obtidos utilizando o método de máxima verossimilhança aproximada para estimar os parâmetros com coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado). Modelo com dispersão variável não linear.	125
4.11	Gráficos normais de probabilidade para os resíduos ponderados padronizados obtidos utilizando o método de máxima verossimilhança aproximada para estimar os parâmetros com coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado). Modelo com dispersão variável não linear.	125
4.12	Gráficos normais de probabilidade para os resíduos combinados obtidos utilizando o método de máxima verossimilhança aproximada para estimar os parâmetros com coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado). Modelo com dispersão variável não linear.	125
4.13	Gráficos normais de probabilidade para os resíduos ponderados obtidos utilizando o método de máxima pseudo-verossimilhança aproximada para estimar os parâmetros com coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado). Modelo com dispersão variável não linear.	126

4.14	Gráficos normais de probabilidade para os resíduos ponderados padronizados obtidos utilizando o método de máxima pseudo-verossimilhança aproximada para estimar os parâmetros com coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado). Modelo com dispersão variável não linear.	126
4.15	Gráficos normais de probabilidade para os resíduos combinados obtidos utilizando o método de máxima pseudo-verossimilhança aproximada para estimar os parâmetros com coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado). Modelo com dispersão variável não linear.	127
4.16	Gráficos normais de probabilidade para os resíduos ponderados obtidos utilizando o método da calibração da regressão para estimar os parâmetros com coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado). Modelo com dispersão variável não linear.	127
4.17	Gráficos normais de probabilidade para os resíduos ponderados padronizados obtidos utilizando o método da calibração da regressão para estimar os parâmetros com coeficiente de confiabilidade $k_x = 0,75$ (erro de medida moderado). Modelo com dispersão variável não linear.	128
4.18	Gráficos normais de probabilidade para os resíduos combinados obtidos utilizando o método da calibração da regressão para estimar os parâmetros com coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado). Modelo com dispersão variável não linear.	128
5.1	<i>Boxplots</i> das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$) e R_{LR}^2 corrigida (R_{LRc}^2) obtidas pelo método da máxima verossimilhança aproximada . Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.	135
5.2	<i>Boxplots</i> das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$) e R_{LR}^2 corrigida (R_{LRc}^2) obtidas pelo método da máxima pseudo-verossimilhança aproximada . Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.	137
5.3	<i>Boxplots</i> das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$) e R_{LR}^2 corrigida (R_{LRc}^2) corrigida obtidas pelo método da calibração da regressão . Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.	138

5.4	<i>Boxplots</i> das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$) e R_{LR}^2 corrigida (R_{LRc}^2) corrigida considerando o coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado) e tamanho amostral $n = 160$. Modelo com erro de medida no preditor da dispersão.	140
5.5	<i>Boxplots</i> das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$) e R_{LR}^2 corrigida (R_{LRc}^2) obtidas pelos métodos avaliados considerando os coeficientes de confiabilidade $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão variável não linear. Tamanho amostral $n = 160$	142
5.6	Função densidade de probabilidade da distribuição <i>t</i> -Student para diferentes graus de liberdade (gl).	148
5.7	Função densidade de probabilidade da distribuição Log-Normal para diferentes parâmetros μ e σ^2	149
5.8	Função densidade de probabilidade para diferentes parâmetros k e λ	150
5.9	Função densidade de probabilidade da distribuição beta para diferentes valores dos parâmetros α e β	151
5.10	Histogramas considerando 100 ocorrências geradas da variável aleatória X que segue diferentes distribuições: (a) $X \sim t(3)$; (b) $X \sim LN(2.5, 2.7)$; (c) $X \sim Weibull(1.5, 1)$; (d) $X \sim Beta(0.6, 0.6)$	153
5.11	<i>Boxplots</i> das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$), R_c^2 e R_{LR}^2 corrigida (R_{LRc}^2) obtidas pelos métodos: (a) máxima verossimilhança aproximada, (b) pseudo-verossimilhança aproximada e (c) calibração da regressão. Coeficiente de confiabilidade $k_x = 0.75$. Modelo com dispersão constante. $X \sim t(3)$	156
5.12	<i>Boxplots</i> das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$), R_c^2 e R_{LR}^2 corrigida (R_{LRc}^2) obtidas pelos métodos: (a) máxima verossimilhança aproximada, (b) pseudo-verossimilhança aproximada e (c) calibração da regressão. Coeficiente de confiabilidade $k_x = 0.75$. Modelo com dispersão constante. $X \sim LN(2.5, 2.7)$	157
5.13	<i>Boxplots</i> das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$), R_c^2 e R_{LR}^2 corrigida (R_{LRc}^2) obtidas pelos métodos: (a) máxima verossimilhança aproximada, (b) pseudo-verossimilhança aproximada e (c) calibração da regressão. Coeficiente de confiabilidade $k_x = 0.75$. Modelo com dispersão constante. $X \sim Weibull(1.5, 1.0)$	157

5.14	<i>Boxplots</i> das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$), R_c^2 e R_{LR}^2 corrigida (R_{LRc}^2) obtidas pelos métodos: (a) máxima verossimilhança aproximada, (b) pseudo-verossimilhança aproximada e (c) calibração da regressão. Coeficiente de confiabilidade $k_x = 0.75$. Modelo com dispersão constante. $X \sim Beta(0.6, 0.6)$	158
5.15	<i>Boxplots</i> das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$), R_c^2 e R_{LR}^2 corrigida (R_{LRc}^2) obtidas pelos métodos: (a) máxima verossimilhança aproximada, (b) pseudo-verossimilhança aproximada e (c) calibração da regressão. Coeficiente de confiabilidade $k_x = 0.75$. Modelo com dispersão constante. $X \sim N(0.0, 1.0)$	158
5.16	<i>Boxplots</i> das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$), R_c^2 e R_{LR}^2 corrigida (R_{LRc}^2) obtidas pelos métodos: (a) máxima verossimilhança aproximada, (b) pseudo-verossimilhança aproximada e (c) calibração da regressão. Coeficiente de confiabilidade $k_x = 0.75$. Avaliação da homoscedasticidade.	159
6.1	<i>Boxplot</i> da variável de interesse (y_t) razão entre a circunferência da cintura (cm) e a altura (cm).	161
6.2	<i>Boxplots</i> com as medidas resumo das covariáveis idade (a) e peso (d) e gráficos de colunas com as porcentagens das covariáveis estado nutricional (b) e gênero (c) das 166 crianças.	162
6.3	Resíduos e envelopes simulados considerando os resíduos ponderados (a) e (d), padronizados (b) e (e) e combinados (c) e (f) utilizando o método da calibração da regressão para o ajuste do modelo de regressão beta com erros de medidas. Modelo (6.1).	166
6.4	Resíduos e envelopes simulados considerando os resíduos ponderados (a) e (d), padronizados (b) e (e) e combinados (c) e (f) utilizando o método da máxima verossimilhança aproximada para o ajuste do modelo de regressão beta com erros de medidas. Modelo (6.1).	167
6.5	Resíduos e envelopes simulados considerando os resíduos ponderados (a) e (d), padronizados (b) e (e) e combinados (c) e (f) utilizando o método da máxima pseudo-verossimilhança aproximada para o ajuste do modelo de regressão beta com erros de medidas. Modelo (6.1).	168
6.6	Gráfico de dispersão da concentração de mmol vanádio/mmol alumínio dos dois métodos.	171

6.7	<i>Boxplots</i> com as medidas descritivas da porcentagem de cristalinidade (a), do vapor d'água (b) e da concentração de vanádio (c). Dados de craqueamento catalítico fluido (FCC).	172
6.8	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método <i>naive</i> para o ajuste do modelo de regressão beta linear com erros de medidas e dispersão constante, Equação (6.4). 175	175
6.9	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da calibração da regressão (RC) para o ajuste do modelo de regressão beta linear com erros de medidas e dispersão constante, Equação (6.4)	175
6.10	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da máxima verossimilhança aproximada (MVa) para o ajuste do modelo de regressão beta com erros de medidas e dispersão constante, Equação (6.4).	176
6.11	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da máxima pseudo-verossimilhança aproximada (MPVa) para o ajuste do modelo de regressão beta linear com erros de medidas e dispersão constante, Equação (6.4).	177
6.12	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método <i>naive</i> para o ajuste do modelo de regressão beta não linear com erros de medidas. Modelo 2.	179
6.13	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da calibração da regressão para o ajuste do modelo de regressão beta não linear com erros de medidas. Modelo 2. 179	179
6.14	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da máxima verossimilhança aproximada para o ajuste do modelo de regressão beta não linear com erros de medidas. Modelo 2.	180
6.15	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da máxima pseudo-verossimilhança aproximada para o ajuste do modelo de regressão beta não linear com erros de medidas. Modelo 2.	180
6.16	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método <i>naive</i> para o ajuste do modelo de regressão beta não linear com erros de medidas. Modelo 3.	182
6.17	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da calibração da regressão para o ajuste do modelo de regressão beta não linear com erros de medidas. Modelo 3. 183	183

6.18	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da máxima verossimilhança aproximada para o ajuste do modelo de regressão beta não linear com erros de medidas. Modelo 3.	183
6.19	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da máxima pseudo-verossimilhança aproximada para o ajuste do modelo de regressão beta não linear com erros de medidas. Modelo 3.	184
6.20	Identificação dos pontos aberrantes considerando os resíduos combinados utilizando os métodos naive (a), calibração da regressão (b), máxima verossimilhança aproximada (c) e máxima pseudo-verossimilhança aproximada (d) para o ajuste do modelo de regressão beta não linear com erros de medidas. Modelo 3. . . .	185
6.21	<i>Boxplot</i> da variável de interesse é a razão entre a circunferência da cintura (cm) e a altura (cm).	193
6.22	<i>Boxplots</i> com as medidas resumo das covariáveis (a),(b),(d)-(e) e gráfico de colunas do gênero dos adolescentes (c).	193
6.23	Gráfico de dispersão do colesterol total (w_t) e colesterol “ruim” LDL (x_t) ambas medições em mg/dl.	194
6.24	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método padrão para o ajuste do modelo de regressão beta tradicional. Modelo 2.	198
6.25	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método naive para o ajuste do modelo de regressão beta com erros de medidas. Modelo 2.	199
6.26	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da calibração da regressão para o ajuste do modelo de regressão beta com erros de medidas. Modelo 2.	199
6.27	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da máxima verossimilhança aproximada para o ajuste do modelo de regressão beta com erros de medidas. Modelo 2.	200
6.28	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da máxima pseudo-verossimilhança aproximada para o ajuste do modelo de regressão beta com erros de medidas. Modelo 2.	200

6.29	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método <i>naive</i> para o ajuste do modelo de regressão beta com erros de medidas.	203
6.30	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da calibração da regressão para o ajuste do modelo de regressão beta com erros de medidas.	203
6.31	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da máxima verossimilhança aproximada para o ajuste do modelo de regressão beta com erros de medidas.	204
6.32	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da máxima pseudo-verossimilhança aproximada para o ajuste do modelo de regressão beta com erros de medidas.	204
6.33	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da calibração da regressão para o ajuste do modelo de regressão beta com erros de medidas.	206
6.34	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da máxima verossimilhança aproximada para o ajuste do modelo de regressão beta com erros de medidas.	207
6.35	Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da máxima pseudo-verossimilhança aproximada para o ajuste do modelo de regressão beta com erros de medidas.	207
1	Vieses estimados dos estimadores de α_0 , α_1 e β_1 para $\phi = 16.44$, $\phi = 54.60$ e $\phi = 298.87$ com o coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado). Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 2: Modelo com dispersão constante. . .	236
2	Vieses estimados dos estimadores de γ_0 para $\phi = 16.44$, $\phi = 54.60$ e $\phi = 298.87$ com o coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado). Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 2: Modelo com dispersão constante.	237
3	Vieses estimados dos estimadores de α_0 , α_1 e β_1 para $\phi = 16.44$, $\phi = 54.60$ e $\phi = 298.87$ com o coeficiente de confiabilidade $k_x = 0.95$ (erro de medida baixo). Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 2: Modelo com dispersão constante. . .	238

- 4 Vieses estimados dos estimadores de γ_0 para $\phi = 16.44$, $\phi = 54.60$ e $\phi = 298.87$ com o coeficiente de confiabilidade $k_x = 0.95$ (erro de medida baixo). Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 2: Modelo com dispersão constante. 239
- 5 Vieses estimados dos estimadores de α_0 , α_1 e β_1 para $Q = 12$ (coluna esquerda), $Q = 30$ (coluna central) e $Q = 80$ (coluna direita) com o coeficiente de confiabilidade $k_x = 0.75$. Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 4: Modelo com dispersão variável. 245
- 6 Vieses estimados dos estimadores de γ_0 e λ_1 para $Q = 12$ (coluna esquerda), $Q = 30$ (coluna central) e $Q = 80$ (coluna direita) com o coeficiente de confiabilidade $k_x = 0.75$. Verdadeiros valores dos parâmetros: $\gamma_0 = 2.5$, $\lambda_1 = 0.9$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 4: Modelo com dispersão variável. . . . 246
- 7 Vieses estimados dos estimadores de α_0 , α_1 e β_1 para $Q = 12$ (coluna esquerda), $Q = 30$ (coluna central) e $Q = 80$ (coluna direita) com o coeficiente de confiabilidade $k_x = 0.95$. Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 4: Modelo com dispersão variável. 247
- 8 Vieses estimados dos estimadores de γ_0 e λ_1 para $Q = 12$ (coluna esquerda), $Q = 30$ (coluna central) e $Q = 80$ (coluna direita) com o coeficiente de confiabilidade $k_x = 0.95$. Verdadeiros valores dos parâmetros: $\gamma_0 = 2.5$, $\lambda_1 = 0.9$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 4: Modelo com dispersão variável. . . . 248

Lista de Tabelas

- 2.1 Limites inferiores médios (LI) e superiores médios (LS) e suas respectivas medianas (Md) para a média da variável resposta (μ_t) considerando o modelo de regressão beta não linear com erros de medida com dispersão constante para diferentes coeficientes de confiabilidade (k_x) e tamanhos amostrais (n). 54
- 2.2 Raízes dos erros quadráticos médios (REQM) para os estimadores de α_0 , α_1 , β_1 e γ_0 considerando o modelo com dispersão constante. Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 1. 55
- 2.3 Limites inferiores médios (LI) e superiores médios (LS) e suas respectivas medianas (Md) para a média da variável resposta (μ_t) considerando o modelo de regressão beta não linear com erros de medida com dispersão constante para diferentes valores de ϕ e o coeficiente de confiabilidade $k_x = 0.50$ 58
- 2.4 Raízes dos erros quadráticos médios (REQM) para os estimadores de α_0 , α_1 , β_1 e γ_0 considerando o modelo com dispersão constante com o coeficiente de confiabilidade $k_x = 0.50$. Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 2. 59
- 2.5 Limites inferiores médios (LI) e superiores médios (LS) e suas respectivas medianas (Md_{μ_t}) para a média da variável resposta (μ_t) e médias dos mínimos (Mín), máximos (Máx) e medianas (Md_{ϕ_t}) de ϕ_t e seus respectivos graus de heterogeneidade (δ) para o modelo de regressão beta não linear com erros de medida e dispersão variável. 63
- 2.6 Raízes dos erros quadráticos médios (REQM) para os estimadores de α_0 , α_1 , β_1 , γ_0 e λ_1 considerando o modelo de regressão beta não linear com dispersão variável. Valores verdadeiros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 2.5$, $\lambda_1 = 0.9$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 3. 64
- 2.7 Limites inferiores médios (LI) e superiores médios (LS) e suas respectivas medianas (Md_{μ_t}) para a média da variável resposta (μ_t) e médias dos mínimos (Mín), máximos (Máx) e medianas (Md_{ϕ_t}) de ϕ_t e seus respectivos graus de heterogeneidade (δ) para o modelo de regressão beta não linear com erros de medida com dispersão variável. Coeficiente de confiabilidade $k_x = 0.50$ 68

2.8	Raízes dos erros quadráticos médios (REQM) para os estimadores de α_0 , α_1 , β_1 , γ_0 e λ_1 considerando o modelo de regressão beta não linear com dispersão variável. Coeficiente de confiabilidade $k_x = 0.50$. Valores verdadeiros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 2.5$, $\lambda_1 = 0.9$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 4.	69
2.9	Limites inferiores médios (LI) e superiores médios (LS) e suas respectivas medianas (Md_{μ_t}) para a média da variável resposta (μ_t) e médias dos mínimos (Mín), máximos (Máx) e medianas (Md_{ϕ_t}) de ϕ_t e seus respectivos graus de heterogeneidade (δ) para o modelo de regressão beta não linear com erros de medidas e não linearidade na dispersão.	73
2.10	Raízes dos erros quadráticos médios (REQM) para os estimadores de α_0 , α_1 , β_1 , γ_0 , γ_1 e λ_1 considerando o modelo de regressão beta não linear com erros de medidas e não linearidade na dispersão. Valores verdadeiros: $\alpha_0 = 0,7$, $\alpha_1 = 2.0$, $\beta_1 = -1.5$, $\gamma_0 = 1.5$, $\gamma_1 = 2.0$, $\lambda_1 = 1.3$, $\mu_x = 1.5$ e $\sigma_x^2 = 0.5$. Cenário 5.	74
3.1	Taxas de cobertura e comprimentos médios dos intervalos de 90% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão constante, utilizando os estimadores obtidos pelos métodos <i>naive</i> (<i>Naive</i>), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.95$. Cenário 1.	86
3.2	Taxas de cobertura e comprimentos médios dos intervalos de 90% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão constante, utilizando os estimadores obtidos pelos métodos <i>naive</i> (<i>Naive</i>), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.75$. Cenário 1.	87
3.3	Taxas de cobertura e comprimentos médios dos intervalos de 90% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão constante, utilizando os estimadores obtidos pelos métodos <i>naive</i> (<i>Naive</i>), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.50$. Cenário 1.	88

3.4	Taxas de cobertura e comprimentos médios dos intervalos de 90% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão variável, utilizando os estimadores obtidos pelos métodos <i>naive</i> (<i>Naive</i>), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.95$. Cenário 2.	92
3.5	Taxas de cobertura e comprimentos médios dos intervalos de 90% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão variável, utilizando os estimadores obtidos pelos métodos <i>naive</i> (<i>Naive</i>), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.75$. Cenário 2.	93
3.6	Taxas de cobertura e comprimentos médios dos intervalos de 90% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão variável, utilizando os estimadores obtidos pelos métodos <i>naive</i> (<i>Naive</i>), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.50$. Cenário 2.	94
3.7	Taxas de cobertura e comprimentos médios dos intervalos de 90% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com não linearidade na dispersão, utilizando os estimadores obtidos pelos métodos <i>naive</i> (<i>Naive</i>), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.95$. Cenário 3.	97
3.8	Taxas de cobertura e comprimentos médios dos intervalos de 90% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com não linearidade na dispersão, utilizando os estimadores obtidos pelos métodos <i>naive</i> (<i>Naive</i>), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.75$. Cenário 3.	98
3.9	Taxas de cobertura e comprimentos médios dos intervalos de 90% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com não linearidade na dispersão, utilizando os estimadores obtidos pelos métodos <i>naive</i> (<i>Naive</i>), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.50$. Cenário 3.	99

4.1	Médias, erros-padrão, assimetrias e curtoses empíricas dos resíduos ponderado (r^α), ponderado padronizado (r_p^α) e combinado ($r^{\alpha\gamma}$). Modelo: $\log(\mu_t/[1-\mu_t]) = \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}$, $t = 1, \dots, 20$, $\phi \approx 150$. Coeficiente de Confiabilidade $k_x = 0.50$. Método: MVa.	108
4.2	Médias, erros-padrão, assimetrias e curtoses empíricas dos resíduos ponderado (r^α), ponderado padronizado (r_p^α) e combinado ($r^{\alpha\gamma}$). Modelo: $\log(\mu_t/[1-\mu_t]) = \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}$, $t = 1, \dots, 20$, $\phi \approx 150$. Coeficiente de Confiabilidade $k_x = 0.75$. Método: MVa.	109
4.3	Médias, erros-padrão, assimetrias e curtoses empíricas dos resíduos ponderado (r^α), ponderado padronizado (r_p^α) e combinado ($r^{\alpha\gamma}$). Modelo: $\log(\mu_t/[1-\mu_t]) = \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}$, $t = 1, \dots, 20$, $\phi \approx 150$. Coeficiente de Confiabilidade $k_x = 0.95$. Método: MVa.	109
4.4	Médias, erros-padrão, assimetrias e curtoses empíricas dos resíduos ponderado (r^α), ponderado padronizado (r_p^α) e combinado ($r^{\alpha\gamma}$). Modelo: $\log(\mu_t/[1-\mu_t]) = \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}$, $t = 1, \dots, 20$, $\phi \approx 150$. Coeficiente de Confiabilidade $k_x = 0.50$. Método: MPVa.	114
4.5	Médias, erros-padrão, assimetrias e curtoses empíricas dos resíduos ponderado (r^α), ponderado padronizado (r_p^α) e combinado ($r^{\alpha\gamma}$). Modelo: $\log(\mu_t/[1-\mu_t]) = \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}$, $t = 1, \dots, 20$, $\phi \approx 150$. Coeficiente de Confiabilidade $k_x = 0.75$. Método: MPVa.	114
4.6	Médias, erros-padrão, assimetrias e curtoses empíricas dos resíduos ponderado (r^α), ponderado padronizado (r_p^α) e combinado ($r^{\alpha\gamma}$). Modelo: $\log(\mu_t/[1-\mu_t]) = \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}$, $t = 1, \dots, 20$, $\phi \approx 150$. Coeficiente de Confiabilidade $k_x = 0.95$. Método: MPVa.	115
4.7	Médias, erros-padrão, assimetrias e curtoses empíricas dos resíduos ponderado (r^α), ponderado padronizado (r_p^α) e combinado ($r^{\alpha\gamma}$). Modelo: $\log(\mu_t/[1-\mu_t]) = \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}$, $t = 1, \dots, 20$, $\phi \approx 150$. Coeficiente de Confiabilidade $k_x = 0.50$. Método: RC.	119
4.8	Médias, erros-padrão, assimetrias e curtoses empíricas dos resíduos ponderado (r^α), ponderado padronizado (r_p^α) e combinado ($r^{\alpha\gamma}$). Modelo: $\log(\mu_t/[1-\mu_t]) = \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}$, $t = 1, \dots, 20$, $\phi \approx 150$. Coeficiente de Confiabilidade $k_x = 0.75$. Método: RC.	119
4.9	Médias, erros-padrão, assimetrias e curtoses empíricas dos resíduos ponderado (r^α), ponderado padronizado (r_p^α) e combinado ($r^{\alpha\gamma}$). Modelo: $\log(\mu_t/[1-\mu_t]) = \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}$, $t = 1, \dots, 20$, $\phi \approx 150$. Coeficiente de Confiabilidade $k_x = 0.95$. Método: RC.	120

5.1	Valores das estatísticas considerando os coeficientes de confiabilidade $k_x = 0.50$, $k_x = 0.75$ e $k_x = 0.95$ e os diferentes métodos de estimação. Modelo com dispersão constante $\phi \approx 150$	134
5.2	Valores das estatísticas considerando os coeficientes de confiabilidade $k_x = 0.75$ e $k_x = 0.95$ e os diferentes métodos de estimação. Modelo sem erro de medida no preditor da dispersão.	139
5.3	Valores das estatísticas considerando os coeficientes de confiabilidade $k_x = 0.75$ e $k_x = 0.95$ e os diferentes métodos de estimação. Modelo com não linearidade na dispersão.	141
5.4	Valores das estatísticas. Modelo mal especificado: omissão de covariáveis. O modelo verdadeiro é $g(\mu_t) = \alpha_0 + \alpha_1 z_{t1} + \alpha_2 z_{t2} + \alpha_3 z_{t3} + \beta_1 x_{t1}$, $t = 1, \dots, n$ e ϕ constante. Método: Máxima verossimilhança aproximada. Coeficiente de confiabilidade: $k_x = 0.75$	144
5.5	Valores das estatísticas. Modelo mal especificado: omissão de covariáveis. O modelo verdadeiro é $g(\mu_t) = \alpha_0 + \alpha_1 z_{t1} + \alpha_2 z_{t2} + \alpha_3 z_{t3} + \beta_1 x_{t1}$, $t = 1, \dots, n$ e ϕ constante. Método: Máxima pseudo-verossimilhança aproximada. Coeficiente de confiabilidade: $k_x = 0.75$	145
5.6	Valores das estatísticas. Modelo mal especificado: omissão de covariáveis. O modelo verdadeiro é $g(\mu_t) = \alpha_0 + \alpha_1 z_{t1} + \alpha_2 z_{t2} + \alpha_3 z_{t3} + \beta_1 x_{t1}$, $t = 1, \dots, n$ e ϕ constante. Método: Calibração da regressão. Coeficiente de confiabilidade: $k_x = 0.75$	146
5.7	Vieses, raízes do erros quadráticos médios (REQM) e taxas de cobertura com 95% de confiança para os estimadores de α_0 , α_1 , β_1 e γ_0 considerando o modelo com dispersão constante. Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$ e $\gamma_0 = 5.0$	154
5.8	Vieses, raízes do erros quadráticos médios (REQM) e taxas de cobertura com 95% de confiança para os estimadores de α_0 , α_1 , β_1 e γ_0 considerando o modelo com dispersão constante. Verdadeiros valores de parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$ e $\gamma_0 = 5.0$	156
5.9	Vieses, raízes do erros quadráticos médios (REQM) e taxas de cobertura com 95% de confiança para os estimadores de α_0 , α_1 , β_1 e γ_0 considerando o modelo com dispersão constante. Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$ e $\gamma_0 = 5.0$	159

6.1	Estimativas, erros-padrão e p -valores obtidos pelos métodos <i>Naive</i> , calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa). Considerando o modelo de regressão beta linear com erros de medida com todas as covariáveis.	164
6.2	Estimativas, erros-padrão e p -valores obtidos pelos métodos <i>Naive</i> , calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa). Modelo (6.1).	165
6.3	Estatísticas de predição e medidas de qualidade do ajuste dos modelos ajustado considerando os métodos calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa). . .	167
6.4	Curva de calibração. Método 1.	170
6.5	Concentração de vanádio, absorvância, peso do zeolito sem água e relação entre vanádio e alumínio (mmol) das 5 amostras estudadas. Método 2.	170
6.6	Conversão da concentração ppm vanádio ± 0.2022 para mmol vanádio/mmol alumínio.	170
6.7	Estimativas, erros-padrão e p -valores obtidos pelos métodos <i>Naive</i> , calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa). Modelo beta linear com erros de medida e dispersão variável.	173
6.8	Estimativas, erros-padrão e p -valores obtidos pelos métodos <i>Naive</i> , calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa). Modelo de regressão beta linear com erros de medidas com dispersão constante (Modelo 1).	174
6.9	Estimativas, erros-padrão e p -valores obtidos pelos métodos <i>naive</i> , calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa). Modelo de regressão beta não linear com erros de medidas com dispersão constante, considerando a função de ligação logito associado à media (Modelo 2).	178
6.10	Estimativas, erros-padrão e p -valores obtidos pelos métodos <i>naive</i> , calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa). Modelos de regressão beta não linear com erros de medida considerando a função de ligação complemento log-log associada à média. Modelo 3.	181
6.11	Estatísticas de predição e medidas para avaliação do ajuste do modelo considerando o modelo de regressão linear beta com erros de medida (Modelo 1).	184

6.12	Estatísticas de predição e medidas para avaliação do ajuste do modelo considerando os modelos de regressão linear beta não linear com erros de medida com função de ligação logito (Modelo 2) e função de ligação complemento log-log (Modelo 3).	185
6.13	Estimativas, erros-padrão e p -valores obtidos pelo método da calibração da regressão (RC).	186
6.14	Estimativas, erros-padrão e p -valores obtidos pelo método da máxima verossimilhança aproximada (MVa).	187
6.15	Estimativas, erros-padrão e p -valores obtidos pelo método da máxima pseudo-verossimilhança aproximada (MPVa).	188
6.16	Limites inferiores (LI), limites superiores (LS) e comprimento (Comp) dos intervalos de confiança considerando os níveis 90%, 95% e 99% de confiança, para os métodos <i>naive</i> , calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa). Modelo de regressão beta linear com erros de medida (Modelo Linear).	189
6.17	Limites inferiores (LI), limites superiores (LS) e comprimento (Comp) dos intervalos de confiança considerando os níveis 90%, 95% e 99% de confiança, para os métodos <i>naive</i> , calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa). Modelo de regressão beta não linear com erros de medida considerando a função de ligação logito (Modelo 2).	190
6.18	Limites inferiores (LI), limites superiores (LS) e comprimento (Comp) dos intervalos de confiança considerando os níveis 90%, 95% e 99% de confiança, para os métodos <i>naive</i> , calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa). Modelo de regressão não linear com erros de medida considerando a função de ligação complemento log-log (Modelo 3).	191
6.19	Estimativas, erros-padrão e p -valores obtidos pelos métodos <i>naive</i> , calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa).	195
6.20	Estimativas, erros-padrão e p -valores obtidos pelo modelo de regressão beta tradicional.	196
6.21	Estimativas, erros-padrão e p -valores obtidos pelo modelo de regressão beta com covariáveis significativas.	197
6.22	Estimativas, erros-padrão e p -valores obtidos pelos métodos <i>naive</i> , calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa).	197

6.23	Estimativas, erros-padrão e p -valores obtidos pelos métodos <i>naive</i> , calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa). Modelo 6.8.	201
6.24	Estimativas, erros-padrão e p -valores obtidos pelo modelo de regressão beta. Modelo 6.8.	202
6.25	Estimativas, erros-padrão e p -valores obtidos pelos métodos <i>naive</i> , calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa).	205
6.26	Estatísticas de predição e medidas de qualidade do ajuste dos modelos ajustado considerando os métodos <i>naive</i> , calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa)	209
6.27	Limites inferiores (LI), limites superiores (LS) e comprimento (Comp) dos intervalos de confiança considerando os níveis 90%, 95% e 99% de confiança, para os métodos <i>naive</i> , calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa). Modelo 3.	210
1	Limites inferiores médios (LI) e superiores médios (LS) e suas respectivas medianas (Md) para a média da variável resposta (μ_t) considerando o modelo de regressão beta não linear com erros de medida com dispersão constante. Coeficiente de confiabilidade $k_x = 0.75$	235
2	Limites inferiores médios (LI) e superiores médios (LS) e suas respectivas medianas (Md) para a média da variável resposta (μ_t) considerando o modelo de regressão beta não linear com erros de medida com dispersão constante. Coeficiente de confiabilidade $k_x = 0.95$	235
3	Raízes dos erros quadráticos médios (REQM) para os estimadores de α_0 , α_1 , β_1 e γ_0 considerando o modelo com dispersão constante com o coeficiente de confiabilidade $k_x = 0.75$. Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 2.	240
4	Raízes dos erros quadráticos médios (REQM) para os estimadores de α_0 , α_1 , β_1 e γ_0 considerando o modelo com dispersão constante com o coeficiente de confiabilidade $k_x = 0.95$. Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 2.	241

5	Limites inferiores médios (LI) e superiores médios (LS) e suas respectivas medianas (Md_{μ_t}) para a média da variável resposta (μ_t) e médias dos mínimos (Mín), máximos (Máx) e medianas (Md_{ϕ_t}) de ϕ_t e seus respectivos graus de heterogeneidade (δ) para o modelo de regressão beta não linear com erros de medida com dispersão variável. Coeficiente de confiabilidade $k_x = 0.75$	242
6	Limites inferiores médios (LI) e superiores médios (LS) e suas respectivas medianas (Md_{μ_t}) para a média da variável resposta (μ_t) e médias dos mínimos (Mín), máximos (Máx) e medianas (Md_{ϕ_t}) de ϕ_t e seus respectivos graus de heterogeneidade (δ) para o modelo de regressão beta não linear com erros de medida com dispersão variável. Coeficiente de confiabilidade $k_x = 0,95$	242
7	Raiz do erro quadrático médio (REQM) para os estimadores de $\alpha_0, \alpha_1, \beta_1, \gamma_0$ e λ_1 considerando o modelo de regressão beta não linear com dispersão variável. Coeficiente de confiabilidade $k_x = 0.75$. Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6, \alpha_1 = 2.4, \beta_1 = 0.8, \gamma_0 = 2.5, \lambda_1 = 0.9, \mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 4.	243
8	Raízes dos erros quadráticos médios (REQM) para os estimadores de $\alpha_0, \alpha_1, \beta_1, \gamma_0$ e λ_1 considerando o modelo de regressão beta não linear com dispersão variável. Coeficiente de confiabilidade $k_x = 0.95$. Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6, \alpha_1 = 2.4, \beta_1 = 0.8, \gamma_0 = 2.5, \lambda_1 = 0.9, \mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 4.	244
9	Taxas de cobertura e comprimentos médios dos intervalos de 99% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão constante, utilizando os estimadores obtidos pelos métodos <i>naive</i> (<i>Naive</i>), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.95$. Cenário 1.	249
10	Taxas de cobertura e comprimentos médios dos intervalos de 99% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão constante, utilizando os estimadores obtidos pelos métodos <i>naive</i> (<i>Naive</i>), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.75$. Cenário 1.	250

11	Taxas de cobertura e comprimentos médios dos intervalos de 99% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão constante, utilizando os estimadores obtidos pelos métodos <i>naive</i> (<i>Naive</i>), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.95$. Cenário 1.	251
12	Taxas de cobertura e comprimentos médios dos intervalos de 99% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão variável, utilizando os estimadores obtidos pelos métodos <i>naive</i> (<i>Naive</i>), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0,95$. Cenário 2.	252
13	Taxas de cobertura e comprimentos médios dos intervalos de 99% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão variável, utilizando os estimadores obtidos pelos métodos <i>naive</i> (<i>Naive</i>), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0,75$. Cenário 2.	253
14	Taxas de cobertura e comprimentos médios dos intervalos de 99% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão variável, utilizando os estimadores obtidos pelos métodos <i>naive</i> (<i>Naive</i>), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0,50$. Cenário 2.	254
15	Taxas de cobertura e comprimentos médios dos intervalos de 99% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com não linearidade na dispersão, utilizando os estimadores obtidos pelos métodos <i>naive</i> (<i>Naive</i>), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.95$. Cenário 3.	255
16	Taxas de cobertura e comprimentos médios dos intervalos de 99% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com não linearidade na dispersão, utilizando os estimadores obtidos pelos métodos <i>naive</i> (<i>Naive</i>), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.75$. Cenário 3.	256

17	Taxas de cobertura e comprimentos médios dos intervalos de 99% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com não linearidade na dispersão, utilizando os estimadores obtidos pelos métodos <i>naive</i> (<i>Naive</i>), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.50$. Cenário 3.	257
----	---	-----

Sumário

1	INTRODUÇÃO	35
1.1	Objetivos	39
1.2	Organização da Tese	39
1.3	Suporte Computacional	40
2	MODELO DE REGRESSÃO BETA NÃO LINEAR COM ERROS DE MEDIDA	41
2.1	O Modelo	41
2.2	Estimação	43
2.2.1	Estimação por máxima verossimilhança aproximada	46
2.2.2	Estimação por pseudo-verossimilhança aproximada	48
2.2.3	Estimação por calibração da regressão	50
2.2.4	Valores iniciais para o algoritmo de otimização	51
2.3	Avaliação numérica	53
2.3.1	Cenário 1: Dispersão constante	54
2.3.2	Cenário 2: Avaliação do parâmetro de precisão ϕ	58
2.3.3	Cenário 3: Dispersão variável	63
2.3.4	Cenário 4: Avaliação dos pontos de quadratura “Q”	67
2.3.5	Cenário 5: Não linearidade na dispersão	72
2.4	Conclusões	76
3	DISTRIBUIÇÕES ASSINTÓTICAS	78
3.1	Distribuições assintóticas dos estimadores	78
3.1.1	Máxima verossimilhança aproximada	78
3.1.2	Máxima pseudo-verossimilhança aproximada	80
3.1.3	Calibração da regressão	82
3.2	Avaliação numérica	83
3.2.1	Cenário 1: Dispersão constante	83
3.2.2	Cenário 2: Dispersão variável	89
3.2.3	Cenário 3: Não linearidade na dispersão	95
3.3	Conclusões	100
4	ANÁLISE DE RESÍDUOS	102
4.1	Introdução	102

4.2	Resíduos para o modelo de regressão beta não linear	103
4.2.1	Resíduo ponderado	104
4.2.2	Resíduo ponderado padronizado	105
4.2.3	Resíduo combinado	105
4.3	Resíduos para o modelo de regressão beta não linear com erros nas variáveis	107
4.4	Avaliação numérica	107
4.4.1	Cenário 1: Dispersão constante	108
4.4.2	Cenário 2: Não linearidade na dispersão	124
4.5	Novos limites para detecção de pontos aberrantes	128
4.6	Conclusões	129
5	ESTATÍSTICAS PRESS, P^2 E MEDIDAS DE QUALIDADE DE AJUSTE PARA O MODELO DE REGRESSÃO BETA NÃO LINEAR COM ERROS DE MEDIDA	130
5.1	Estatísticas PRESS e P^2	131
5.2	Avaliação numérica - Modelo de regressão beta não linear com erros de medida	133
5.2.1	Cenário 1: Dispersão constante	134
5.2.2	Cenário 2: Dispersão variável	136
5.2.3	Cenário 3: Não linearidade na dispersão	141
5.2.4	Cenário 4 - Modelo Linear: Má especificação do modelo	143
5.3	Avaliação da má especificação da distribuição mensurada com erro de medida	147
5.3.1	Distribuição <i>t-Student</i>	147
5.3.2	Distribuição Log-Normal	148
5.3.3	Distribuição Weibull	149
5.3.4	Distribuição Beta	150
5.4	Avaliação numérica - Má especificação da distribuição	151
5.5	Conclusões	160
6	APLICAÇÕES	161
6.1	Aplicação 1 - Modelo Linear: Sobre peso em crianças considerando o peso mensurado duas vezes.	161
6.1.1	Modelos ajustados	163
6.2	Aplicação 2 - Modelo Não Linear: Dados de craqueamento catalítico fluido (FCC)	169
6.2.1	Modelos ajustados	172

6.2.2	Intervalos de confiança	188
6.3	Aplicação 3 - Modelo Não Linear: sobrepeso em adolescentes	192
6.3.1	Modelos ajustados	194
7	CONSIDERAÇÕES FINAIS	212
7.1	Trabalhos futuros	214
	REFERÊNCIAS	215
	APÊNDICE A - FUNÇÃO ESCORE E MATRIZ DE INFORMAÇÃO OBSERVADA	220
	Apêndice A.1 - Função escore	220
	Apêndice A.1.1 - Máxima verossimilhança aproximada	220
	Apêndice A.1.2 - Máxima pseudo-verossimilhança aproximada	224
	Apêndice A.1.3 - Calibração da regressão	225
	Apêndice A.2 - Matriz de Informação Observada - MVa	228
	APÊNDICE B - RESULTADOS DOS ESTUDOS DE SIMULAÇÃO DOS CAPÍTULOS 2 E 3	235
	Apêndice B.1 - Cenário 2: Avaliação do parâmetro de precisão ϕ	235
	Apêndice B.2 - Cenário 4: Avaliação dos pontos de quadratura “Q”	242

1 INTRODUÇÃO

Os modelos matemáticos que relacionam e explicam o comportamento de uma variável resposta através de uma função das covariáveis (variáveis explicativas), conhecidos como modelos de regressão, vêm sendo comumente utilizados em diferentes áreas do conhecimento, tais como saúde, economia, ciências biológicas, entre outras. Os modelos de regressão podem ser classificados como: modelos lineares ou não lineares, modelos homoscedásticos ou heteroscedásticos, modelos com ou sem erros nas variáveis, por exemplo. Dentro de cada classe de modelos existem suposições que devem ser respeitadas para que o processo inferencial seja válido.

Quando as covariáveis não são observadas diretamente ou estão sujeitas a erros de medição surge uma classe de modelos de regressão denominados modelos com erros de medida ou modelos com erros nas variáveis. É importante ressaltar que, se esta suposição não é respeitada, serão obtidos resultados inferenciais não confiáveis (Stefanski, 1985; Stefanski e Carroll, 1985; Rosner e Spiegelman, 1989). Nestas circunstâncias, os modelos de regressão com erros de medida usualmente são definidos e estruturados de modo que a resposta y seja uma função de preditores ou covariáveis x cujas medições são imprecisas. Assim, no lugar do verdadeiro valor da variável preditora, x , considera-se o valor de uma outra variável preditora, w , que está associada a um erro de mensuração.

Existem três modelagens para solucionar a problemática gerada pela mensuração de covariáveis com erro. A primeira, denotada por modelo estrutural, considera uma distribuição de probabilidade para as covariáveis não observadas. Na segunda abordagem, chamada de modelo funcional, os valores desconhecidos de x são vistos como parâmetros. A última abordagem, conhecida como modelo ultraestrutural, é uma generalização dos modelos estrutural e funcional. Neste trabalho, focaremos apenas no modelo estrutural. Preditores com erros de medida podem estar associados a diversos tipos de resposta. Aqui nosso interesse são respostas que se distribuem de forma contínua no intervalo $(0, 1)$.

A literatura apresenta diversas metodologias para analisar variáveis dependentes y limitadas ao intervalo $(0,1)$, como proporções. O modelo de regressão beta, proposto por Ferrari e Cribari-Neto (2004), vem sendo bastante utilizado nos últimos anos e se tornou uma abordagem apropriada para modelar variáveis contínuas restritas ao intervalo unitário, isto é, quando $0 < y < 1$, que podem ser explicadas por covariáveis através de uma estrutura de regressão para a média. Neste modelo assume-se que o parâmetro de dispersão é constante para todas as observações. Além disso, a estrutura de regressão para a média da variável resposta é linear.

No entanto, sabe-se que em muitas aplicações tais suposições não são satisfeitas. Desta forma, Simas *et al.* (2010) estenderam o modelo de Ferrari e Cribari-Neto (2004) que de maneira semelhante à estrutura da média de y_t , propõem uma estrutura de regressão para o parâmetro de precisão. Os autores ainda generalizaram a proposta ao considerarem que essas estruturas de regressão sejam também não lineares.

Considere y_1, \dots, y_n variáveis aleatórias independentes tais que, para cada $t = 1, \dots, n$, y_t segue distribuição beta com função de densidade dada por

$$f(y_t; \mu_t, \phi_t) = \frac{\Gamma(\phi_t)}{\Gamma(\mu_t \phi_t) \Gamma((1 - \mu_t) \phi_t)} y_t^{\mu_t \phi_t - 1} (1 - y_t)^{(1 - \mu_t) \phi_t - 1}, \quad 0 < y_t < 1, \quad (1.1)$$

em que $0 < \mu_t < 1$, $\phi_t > 0$. Aqui,

$$\mathbb{E}(y_t) = \mu_t \quad \text{e} \quad \text{Var}(y_t) = \frac{V(\mu_t)}{1 + \phi_t},$$

em que, $V(\mu_t) = \mu_t(1 - \mu_t)$ é a função de variância. Uma vez fixada a média da variável resposta quanto maior o valor de ϕ_t menor será a variância de y_t . Assim, ϕ_t pode ser interpretado como um parâmetro de precisão do modelo e ϕ_t^{-1} como o parâmetro de dispersão.

O modelo de regressão beta com dispersão variável é dado pela função de densidade (1.1), sendo impostas as componentes sistemáticas

$$g(\mu_t) = \mathbf{z}_t^\top \boldsymbol{\alpha}, \quad (1.2)$$

$$h(\phi_t) = \mathbf{v}_t^\top \boldsymbol{\gamma}, \quad (1.3)$$

em que $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^\top \in \mathbb{R}^p$, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_{\check{q}})^\top \in \mathbb{R}^{\check{q}}$ são vetores de parâmetros desconhecidos, a serem estimados; $\mathbf{z}_t^\top = (z_{t1}, \dots, z_{tp})$ e $\mathbf{v}_t^\top = (v_{t1}, \dots, v_{t\check{q}})$ são vetores de observações fixas e conhecidas de p e \check{q} covariáveis ($p + \check{q} < n$). A função de ligação $g : (0, 1) \rightarrow \mathbb{R}$ é, por suposição, estritamente monótona, contínua e duas vezes diferenciável e a função de ligação $h : (0, \infty) \rightarrow \mathbb{R}$ é estritamente monótona, contínua e duas vezes diferenciável.

Há inúmeras funções de ligação para a média que podem ser utilizadas no modelo de regressão beta, entre elas, a função *logit* $g(\mu) = \log(\mu/1 - \mu)$, *probit* $g(\mu) = \Phi^{-1}(\mu)$, em que $\Phi(\cdot)$ representa a função acumulada da distribuição normal, complemento log-log $g(\mu) = \log(-\log(1 - \mu))$, entre outras. Particularmente, quando se utiliza a função de ligação *logit* é possível interpretar os parâmetros do modelo de regressão em termos da razão de chances. Existem diversas funções de ligação para ϕ_t , por exemplo, pode-se utilizar a especificação logarítmica, $h(\phi_t) = \log(\phi_t)$ ou raiz quadrada $h(\phi_t) = \sqrt{\phi_t}$.

Na prática é razoável pensar que algumas variáveis explicativas dos vetores \mathbf{z}_t e \mathbf{v}_t em (1.2) e (1.3) não sejam observadas diretamente, mas sim adquiridas com possíveis erros de

mensuração. Carrasco *et al.* (2014) propuseram um modelo de regressão beta linear com erros de medida e dispersão variável, tal que

$$\begin{aligned} g(\mu_t) &= \mathbf{z}_t^\top \boldsymbol{\alpha} + \mathbf{x}_t^\top \boldsymbol{\beta}, \\ h(\phi_t) &= \mathbf{v}_t^\top \boldsymbol{\gamma} + \mathbf{m}_t^\top \boldsymbol{\lambda}, \end{aligned}$$

em que $\boldsymbol{\beta} = (\beta_1, \dots, \beta_r)^\top \in \mathbb{R}^r$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_s)^\top \in \mathbb{R}^s$ são parâmetros desconhecidos, $(p+r) + (\check{q}+s) < n$, $\mathbf{x}_t^\top = (x_{t1}, \dots, x_{tr})$ e $\mathbf{m}_t^\top = (m_{t1}, \dots, m_{ts})$ são vetores de covariáveis não observadas diretamente ou associadas a erros de medição que no enfoque estrutural são consideradas como variáveis latentes.

Carrasco *et al.* (2014) apresentaram os métodos de máxima verossimilhança aproximada, pseudo-verossimilhança aproximada e calibração da regressão para estimar os parâmetros e avaliaram numericamente, através de estudos de Monte Carlo, os comportamentos dos estimadores dos parâmetros do modelo calculando as suas médias, vieses e os erros quadráticos médios. Os resultados mostraram melhores desempenhos para os métodos de máxima verossimilhança aproximada e pseudo-verossimilhança aproximada. Também foram avaliadas as taxas de cobertura para os intervalos de 90%, 95% e 99% de confiança dos parâmetros do modelo, considerando todos os métodos. Os autores concluíram que os métodos de máxima verossimilhança aproximada e pseudo-verossimilhança aproximada apresentaram melhores resultados.

Modelos não lineares com erros de medida estão sendo recentemente estudados, como mostra a literatura (Chesher, 1991; Huwang e Huang, 2000; Wang, 2003; Wang, 2004; Schennach, 2012). Chesher (1991) obteve uma fórmula aproximada para caracterizar o viés do estimador para modelo de regressão não linear geral, em que o regressor é medido com erro e a distribuição do erro não necessariamente é normal. Huwang e Huang (2000) estudaram um modelo polinomial univariado em que $g(x; \beta)$ é um polinômio em x de ordem conhecida e mostraram que o estimador de mínimos quadrados baseados nos dois primeiros momentos condicionais de y dado w são consistentes. Wang (2003) considerou modelos não lineares univariados em que todos os erros aleatórios, denotados por ϵ , são distribuídos normalmente e mostrou que o estimador da menor distância baseado nos dois primeiros momentos condicionais de y dados w é consistente e assintoticamente normal. Wang (2004) propôs um método de estimação para modelos não lineares com erros de medida tipo Berkson. Este estimador é baseado nos primeiro e segundo momentos condicionais da variável resposta e considera as variáveis predictoras do modelo. Schennach (2012) apresentou uma visão geral da recente literatura sobre os modelos não lineares com erros de medida, aplicados principalmente na área de econometria. A autora explicou as origens do conceito erro de medição, a identificação e estimativa de modelos não lineares com erro de medição. Também descreveu abordagens de estimação simples que dependem

de conhecimento da distribuição em relação à variável com erro de medição. Em seguida, descreveu métodos que garantiam a identificação do erro de medida através de variáveis auxiliares mais facilmente disponíveis (como medidas repetidas, variáveis instrumentais e dados de painel).

Uma etapa importante no ajuste de um modelo de regressão é a verificação de possíveis afastamentos das suposições feitas ao modelo, especialmente para o componente aleatório e para a parte sistemática do modelo, bem como a avaliação do efeito de observações discrepantes nos resultados do ajuste. Tal etapa, conhecida como análise de diagnóstico, inclui, entre outros, a análise de resíduos que tem como objetivo detectar a presença de possíveis pontos aberrantes e avaliar a adequação da distribuição proposta para a variável resposta. As técnicas utilizadas para análise de resíduos e diagnóstico para os MLG's são semelhantes às utilizadas no modelo clássico de regressão beta. A literatura apresenta algumas definições de resíduos para os modelos de regressão beta. Ferrari e Cribari-Neto (2004) e Espinheira *et al.* (2008) propuseram alguns resíduos para o modelo de regressão beta com dispersão constante. Ferrari *et al.* (2011) estenderam um dos resíduos propostos por Espinheira *et al.* (2008) para o modelo que considera a dispersão variável. Rocha e Simas (2011) definiram um novo resíduo para o modelo de regressão beta não linear e generalizaram alguns dos resíduos definidos por Espinheira *et al.* (2008). Espinheira *et al.* (2017) propuseram um resíduo, denotado por resíduo combinado, para o modelo de regressão beta não linear. Carrasco (2012) utilizou um resíduo proposto por Espinheira *et al.* (2008) para o modelo de regressão beta linear com erros de medidas.

Outro aspecto importante na modelagem de regressão é a seleção de covariáveis para o preditor linear ou não linear, bem como a avaliação da qualidade de ajuste do modelo proposto aos dados. Diversos critérios para seleção de modelos são apresentados na literatura com o objetivo de escolher um modelo que apresente o número mínimo de parâmetros possíveis a serem estimados e que explique bem o comportamento da variável resposta. Entretanto, muitas medidas não fornecem nenhuma visão da qualidade dos valores preditos. Desta forma, Allen (1971) propôs a estatística soma de quadrados dos valores preditos para a resposta (PRESS), que é utilizada na identificação do modelo com o melhor poder preditivo (Liu *et al.*, 1999). Baseada na estatística PRESS é possível calcular a estatística P^2 , que é semelhante ao coeficiente de determinação (R^2). Vale salientar que esta estatística mensura a qualidade de predição enquanto que o R^2 mede a qualidade do ajuste. Silva (2015) propôs a estatística PRESS e a P^2 para o modelo de regressão beta linear e não linear, baseando-se nos resíduos ponderado e combinado.

1.1 Objetivos

A presente tese tem como objetivo estudar os modelos de regressão beta não linear com erros de medida sob o enfoque estrutural. Serão estudados os seguintes métodos de estimação: máxima verossimilhança aproximada, máxima pseudo-verossimilhança aproximada e calibração da regressão. Com o intuito de avaliar o comportamento dos estimadores, serão realizados estudos de simulação via Monte Carlo.

Objetivamos também, encontrar aproximações para a distribuição assintótica dos estimadores dos parâmetros associados à média e à dispersão de forma análoga aos trabalhos de Gong e Samaniego (1981), Carroll *et al.* (2006) e Guolo (2011), além de obter intervalos de confiança para os parâmetros.

Para possibilitar diagnóstico da qualidade de ajuste do modelo, resíduos para o modelo de regressão beta não linear com erros de medidas serão propostos e avaliados via simulações e aplicações à dados reais.

Além disso, coeficientes de predição PRESS, P^2 e medidas de qualidade de ajuste para o modelo de regressão beta não linear com erros de medidas serão propostos e avaliados através de estudos de simulação e aplicação a dados reais.

1.2 Organização da Tese

Esta tese encontra-se dividida em sete capítulos. No Capítulo 2 apresentamos o modelo de regressão beta não linear na presença de erros de medida sob o enfoque estrutural. Desenvolvemos inferência baseada nos métodos de máxima verossimilhança aproximada, máxima pseudo-verossimilhança aproximada e calibração da regressão. Além disso, expressões para o vetor escore são obtidas e realizamos um estudo de simulação para avaliar o comportamento assintótico dos estimadores associados ao modelo. No Capítulo 3 as distribuições assintóticas dos estimadores dos parâmetros do modelo de regressão beta não linear com erros de medida são definidas. Estudos de simulação para avaliar a taxa de cobertura dos intervalos de confiança, considerando os métodos de estimação abordados, foram realizados. O Capítulo 4 apresenta a análise de resíduos, considerando os resíduos ponderados, resíduos ponderados padronizados e os resíduos combinados, dos modelos avaliados para os três métodos de estimação estudados. No Capítulo 5 propomos os coeficientes de predição PRESS, P^2 e medidas de qualidade do ajuste para o modelo de regressão beta não linear com erros de medidas. Aplicações a dados reais são apresentadas no Capítulo 6. Por fim, o Capítulo 7 apresenta as principais considerações finais da tese.

1.3 Suporte Computacional

A linguagem de programação Ox (versão 7.2) constitui a plataforma computacional utilizada no desenvolvimento da tese. Ela permite a implementação de técnicas estatísticas com facilidade e atende a requisitos como precisão e eficiência, o que tem contribuído para sua ampla utilização no campo da computação científica. Detalhes sobre a linguagem de programação Ox podem ser encontrados em Doornik (2006, 2007, 2009, 2011).

Utilizamos também a linguagem de programação R (versão 3.2.2) para a produção de gráficos. R é um sistema integrado para computação estatística e geração de gráficos (Cribari-Neto e Zarkos, 1999; Venables e Ripley, 2002; Dalgaard, 2002; Murrell, 2006; e Venables *et al.*, 2009).

A presente tese foi digitada com auxílio do sistema tipográfico L^AT_EX. Consiste em uma série de macros ou rotinas do sistema T_EX que facilitam o desenvolvimento da edição de textos (Knuth, 1986). Mais detalhes podem ser encontrados em L^Amp^ort (1994) ou através do site <http://www.tex.ac.uk/CTAN/latex>.

2 MODELO DE REGRESSÃO BETA NÃO LINEAR COM ERROS DE MEDIDA

Em muitas pesquisas científicas existe o interesse na relação não linear

$$y = g(x; \beta) + \epsilon, \quad (2.1)$$

em que $y \in \mathbb{R}$ é a variável resposta, $x \in \mathbb{R}^{r_1}$ é a variável preditora, $\beta \in \mathbb{R}^r$ é o parâmetro de regressão desconhecido, sendo $r_1 < r$, e ϵ é o erro aleatório. Em muitos experimentos é impossível mensurar o preditor x de forma exata. Assim, uma aproximação de x pode ser feita através da mensuração de w (Wang, 2004).

Por exemplo, um epidemiologista estuda a gravidade de uma doença pulmonar, y , entre os residentes de uma cidade em relação à quantidade de determinados poluentes atmosféricos, x . Assume-se que os poluentes são medidos em uma certa estação de observação na cidade. A exposição atual dos residentes aos poluentes x , contudo, pode variar de forma aleatória a partir dos valores w mensurados naquelas estações.

Outros exemplos incluem estudos agrícolas em que as relações entre a produção de uma colheita, y , e a quantidade de fertilizante usada, x , são estudadas. A absorção do fertilizante pelo solo pode variar aleatoriamente por causa das condições do local. Ou ainda, estudos médicos em que as relações entre a eficácia de uma droga, y , e a quantidade de droga usada, x , são analisadas. A absorção da droga no sangue do paciente pode variar de forma aleatória por conta das condições biológicas de cada indivíduo (Wang, 2004).

O objetivo principal deste capítulo é introduzir uma nova classe de modelos denotados por modelos de regressão beta não lineares em que uma ou mais variáveis explicativas são mensuradas com erros. As técnicas de estimação de máxima verossimilhança aproximada, pseudo-verossimilhança aproximada e calibração da regressão são apresentadas e discutidas para este modelo. Para avaliar o comportamento dos estimadores dos parâmetros encontrados pelos métodos de estimação citados acima é realizado um estudo de simulação de Monte Carlo.

2.1 O Modelo

Sejam y_1, \dots, y_n variáveis aleatórias independentes tais que, para cada $t = 1, \dots, n$, y_t segue distribuição beta com função de densidade dada por

$$f(y_t; \mu_t, \phi_t) = \frac{\Gamma(\phi_t)}{\Gamma(\mu_t \phi_t) \Gamma((1 - \mu_t) \phi_t)} y_t^{\mu_t \phi_t - 1} (1 - y_t)^{(1 - \mu_t) \phi_t - 1}, \quad 0 < y_t < 1, \quad (2.2)$$

em que $0 < \mu_t < 1$, $\phi_t > 0$. Desta forma, $y_t \sim \text{Beta}(\mu_t, \phi_t)$.

O modelo de regressão beta não linear com erros de medida assume que a média de y_t e o parâmetro de precisão se relacionam através das funções $g(\cdot)$ e $h(\cdot)$ com os preditores não lineares η_{1t} e η_{2t} , envolvendo covariáveis, com e sem erros de medida, e parâmetros desconhecidos. Assim, o modelo de regressão beta não linear com erros de medida baseado em (2.2) assume que a média e o parâmetro de precisão de y_t satisfazem as seguintes relações funcionais

$$g(\mu_t) = \eta_{1t} = f_1(\mathbf{z}_t^\top, \mathbf{x}_t^\top; \boldsymbol{\alpha}, \boldsymbol{\beta}) \quad (2.3)$$

$$\text{e} \quad h(\phi_t) = \eta_{2t} = f_2(\mathbf{v}_t^\top, \mathbf{m}_t^\top; \boldsymbol{\gamma}, \boldsymbol{\lambda}), \quad (2.4)$$

em que $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^\top \in \mathbb{R}^p$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_r)^\top \in \mathbb{R}^r$, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_{\check{q}})^\top \in \mathbb{R}^{\check{q}}$ e $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_s)^\top \in \mathbb{R}^s$ são parâmetros desconhecidos, $\mathbf{z}_t^\top = (z_{t1}, \dots, z_{tp_1})$ e $\mathbf{v}_t^\top = (v_{t1}, \dots, v_{t\check{q}_1})$ são vetores de observações fixas e conhecidas de p_1 e \check{q}_1 covariáveis, $\mathbf{x}_t^\top = (x_{t1}, \dots, x_{tr_1})$ e $\mathbf{m}_t^\top = (m_{t1}, \dots, m_{ts_1})$ são vetores de covariáveis não observadas diretamente ou associadas a erros de medição, com $p_1 \leq p$, $\check{q}_1 \leq \check{q}$, $r_1 \leq r$ e $s_1 \leq s$. As funções de ligação $g(\cdot) : (0, 1) \rightarrow \mathbb{R}$ e $h(\cdot) : (0, \infty) \rightarrow \mathbb{R}$ são, por condição, estritamente monótonas, contínuas e duas vezes diferenciáveis. As funções $f_1(\cdot)$ e $f_2(\cdot)$ são funções contínuas e diferenciáveis, tais que as matrizes de derivadas $\mathbf{F}_1 = \partial\eta_1/\partial\boldsymbol{\alpha}$, $\mathbf{F}_2 = \partial\eta_1/\partial\boldsymbol{\beta}$, $\mathbf{F}_3 = \partial\eta_2/\partial\boldsymbol{\gamma}$ e $\mathbf{F}_4 = \partial\eta_2/\partial\boldsymbol{\lambda}$ têm postos p, r, \check{q} e s , respectivamente.

Como no modelo de regressão beta com erros de medida proposto por Carrasco *et al.* (2014), os vetores de variáveis explicativas mensuradas sem erro, \mathbf{z}_t e \mathbf{v}_t , podem não conter variáveis em comum, ocorrendo o mesmo com as variáveis medidas com erros, \mathbf{x}_t e \mathbf{m}_t . Com o objetivo de facilitar a descrição das metodologias que serão abordadas é considerado que \mathbf{x}_t e \mathbf{m}_t , coincidam, isto é, utilizamos as mesmas variáveis mensuradas com erros na especificação da média e da precisão da variável resposta.

No modelo de regressão beta não linear com erros de medida, o vetor de medidas aleatórias $\mathbf{x}_t = (x_{t1}, \dots, x_{tr_1})^\top$ não é observado diretamente, isto é, observa-se $\mathbf{w}_t = (w_{t1}, \dots, w_{tr})^\top$, que se relaciona com \mathbf{x}_t da seguinte forma:

$$\mathbf{w}_t = \mathfrak{R}_1(\mathbf{x}_t; \boldsymbol{\tau}) + \mathbf{e}_t, \quad (2.5)$$

em que $\mathfrak{R}_1(\cdot; \boldsymbol{\tau})$ denota uma função linear ou não linear do vetor de parâmetros $\boldsymbol{\tau} \in \mathbb{R}^{r_2}$, $r_2 \leq r$ e $\mathbf{e}_t = (e_{t1}, \dots, e_{tr})^\top$ é um vetor de erros aleatórios. Uma particularidade de (2.5) é dado por:

$$\mathbf{w}_t = \boldsymbol{\tau}_0 + (\boldsymbol{\tau}_1 \circ \mathbf{x}_t) + \mathbf{e}_t,$$

em que $\boldsymbol{\tau}_0 = (\tau_{01}, \dots, \tau_{0r})^\top \in \mathbb{R}^r$ e $\boldsymbol{\tau}_1 = (\tau_{11}, \dots, \tau_{1r})^\top \in \mathbb{R}^r$ são vetores de parâmetros desconhecidos e o operador “ \circ ” representa o produto de Hadamard (produto elemento

a elemento). É possível considerar que $\boldsymbol{\tau}_0$ é um vetor de zeros e $\boldsymbol{\tau}_1$ é um vetor de uns. Então, a equação anterior se reduz ao modelo clássico (Carrasco *et al.*, 2014)

$$\mathbf{w}_t = \mathbf{x}_t + \mathbf{e}_t. \quad (2.6)$$

Considerando o enfoque estrutural, admite-se que a distribuição do vetor da variável não observada é normal, ou seja, $x_{ti} \sim N(\mu_{xi}, \sigma_{xi}^2)$, com $\mu_{xi} \in \mathbb{R}$ e $\sigma_{xi}^2 > 0$ denotando, respectivamente, a média e a variância da i -ésima variável explicativa não observável, independentes, $i = 1, \dots, r$. Além disso, supõe-se que os erros aleatórios são independentes tais que $e_{ti} \sim N(0, \sigma_{ei}^2)$. Finalmente, a independência entre x_{ti} e $e_{t'i'}$, com $t, t' = 1, \dots, n$ e $i, i' = 1, \dots, r$, é assumida.

Como nos modelos lineares com erros de medida, os modelos não lineares com erros de medida sob enfoque estrutural também devem ser tratados levando-se em conta o problema de identificabilidade (Fuller, 1987). Para contornar tal problema, usualmente assume-se conhecida alguma quantidade, por exemplo, as variâncias dos erros de medida, ou uma das razões $\sigma_{xi}^2/\sigma_{ei}^2$ ou $\sigma_{xi}^2/(\sigma_{xi}^2 + \sigma_{ei}^2)$, sendo esta última denotada como coeficiente de confiabilidade (Cheng e Van Ness, 1999; Fuller, 1987).

É possível estimar as variâncias dos erros de medida sempre que existem n_{ti} réplicas para w_{ti} , $i = 1, \dots, r$ (Carroll *et al.*, 2006). Assim, sob esta suposição, o modelo com erros de medida (2.6) tem a seguinte estrutura:

$$w_{tij} = x_{ti} + e_{ti}, \quad t = 1, \dots, n, \quad i = 1, \dots, r, \quad j = 1, \dots, n_{ti},$$

em que w_{tij} é a j -ésima observação da i -ésima covariável medida com erros para o indivíduo t . Na presença de réplicas, o desvio das observações em torno da média representa a variabilidade intrapessoal para cada t -ésimo indivíduo. Sendo assim, as variâncias dos erros de medida podem ser estimadas por

$$\hat{\sigma}_{ei}^2 = \frac{1}{\sum_{t=1}^n (n_{ti} - 1)} \sum_{t=1}^n \sum_{j=1}^{n_{ti}} (w_{tij} - \bar{w}_{ti.})^2, \quad (2.7)$$

em que

$$\bar{w}_{ti.} = \frac{1}{n_{ti}} \sum_{j=1}^{n_{ti}} w_{tij}.$$

Uma propriedade importante deste estimador é que $\mathbb{E}[\hat{\sigma}_{ei}^2] = \sigma_{ei}^2$, ou seja, o estimador de σ_{ei}^2 é não viesado (Carrasco, 2012).

2.2 Estimação

Os parâmetros do modelo de regressão beta não linear com erros de medida podem ser estimados através de três métodos, da mesma forma que o modelo proposto por Carrasco

et al. (2014). No método de máxima verossimilhança, as estimativas dos parâmetros são encontradas a partir da maximização do logaritmo da função de verossimilhança. Já no método da máxima pseudo-verossimilhança os parâmetros de perturbação são estimados anteriormente através de alguma metodologia usual, como a máxima verossimilhança, para depois serem substituídos no logaritmo da função de verossimilhança original. Por fim, no método de calibração da regressão os estimadores são definidos a partir da troca do vetor de covariáveis não observadas \mathbf{x}_t por um vetor de estimativas do valor esperado condicional de \mathbf{x}_t dado \mathbf{w}_t , ou seja, $\mathbf{x}_t|\mathbf{w}_t$.

A estrutura do modelo proposto é similar à do modelo linear de regressão beta com erros de medida proposto por Carrasco *et al.* (2014). Desta forma, seja um conjunto finito de pares $(y_1, \mathbf{w}_1), (y_2, \mathbf{w}_2), \dots, (y_n, \mathbf{w}_n)$ de variáveis observadas. A função de densidade conjunta de (y_t, \mathbf{w}_t) , que é o par de observações para o t -ésimo indivíduo, $t = 1, \dots, n$ é obtida integrando a função de densidade conjunta dos dados completos $(y_t, \mathbf{w}_t, \mathbf{x}_t)$, definida por $f(y_t, \mathbf{w}_t, \mathbf{x}_t; \Psi)$, com respeito a \mathbf{x}_t . Aqui, $\Psi = (\boldsymbol{\theta}^\top, \boldsymbol{\xi}^\top, \boldsymbol{\sigma}_e^{2\top})^\top$ representa o vetor de parâmetros desconhecidos, em que $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\lambda}^\top)^\top$ denota o vetor de parâmetros de interesse e $(\boldsymbol{\xi}^\top, \boldsymbol{\sigma}_e^{2\top})^\top$ o vetor de parâmetros de perturbação, sendo $\boldsymbol{\xi}^\top = (\boldsymbol{\mu}_x^\top, \boldsymbol{\sigma}_x^{2\top})^\top$. Os vetores de médias, de variâncias de \mathbf{x}_t e o vetor de variâncias do erro são definidos por $\boldsymbol{\mu}_x = (\mu_{x1}, \dots, \mu_{xr})^\top$, $\boldsymbol{\sigma}_x^2 = (\sigma_{x1}^2, \dots, \sigma_{xr}^2)^\top$ e $\boldsymbol{\sigma}_e^2 = (\sigma_{e1}^2, \dots, \sigma_{er}^2)^\top$, respectivamente. Desta forma, a função de densidade conjunta de (y_t, \mathbf{w}_t) é dada por

$$\begin{aligned} f(y_t, \mathbf{w}_t; \Psi) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_t, \mathbf{w}_t, \mathbf{x}_t; \boldsymbol{\theta}) d\mathbf{x}_t, \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_t|\mathbf{w}_t, \mathbf{x}_t; \boldsymbol{\theta}) f(\mathbf{w}_t|\mathbf{x}_t; \boldsymbol{\sigma}_e^2) f(\mathbf{x}_t; \boldsymbol{\xi}) d\mathbf{x}_t, \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_t|\mathbf{x}_t; \boldsymbol{\theta}) f(\mathbf{w}_t|\mathbf{x}_t; \boldsymbol{\sigma}_e^2) f(\mathbf{x}_t; \boldsymbol{\xi}) d\mathbf{x}_t, \end{aligned} \quad (2.8)$$

em que $f(y_t|\mathbf{x}_t; \boldsymbol{\theta})$ é a função de densidade da distribuição beta dada em (2.2), $f(\mathbf{x}_t; \boldsymbol{\xi})$ é a função de densidade da variável explicativa \mathbf{x}_t , cuja distribuição é $N_r(\boldsymbol{\mu}_x, \text{diag}(\sigma_{x1}^2, \dots, \sigma_{xr}^2))$ e $f(\mathbf{w}_t|\mathbf{x}_t; \boldsymbol{\sigma}_e^2)$ é a função de densidade condicional de \mathbf{w}_t dado \mathbf{x}_t com distribuição $N_r(\mathbf{x}_t, \text{diag}(\sigma_{e1}^2, \dots, \sigma_{er}^2))$. Aqui, $\text{diag}(a_1, \dots, a_r)$ é uma matriz diagonal com elementos a_1, \dots, a_r . Observe que, para encontrar (2.8), deve-se assumir que o erro de medida é não diferenciável, ou seja, que a distribuição de y_t dado $(\mathbf{w}_t, \mathbf{x}_t)$ depende apenas de \mathbf{x}_t (Carroll *et al.*, 2006).

Desta forma, o logaritmo da função de verossimilhança das n observações independentes correspondente à função de densidade conjunta (2.8) da amostra observada é dado por

$$\begin{aligned}
\ell(\Psi) &= \sum_{t=1}^n \log \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_t | \mathbf{x}_t; \boldsymbol{\theta}) f(\mathbf{x}_t | \mathbf{w}_t; \boldsymbol{\xi}, \sigma_e^2) f(\mathbf{w}_t; \boldsymbol{\xi}, \sigma_e^2) d\mathbf{x}_t \\
&= \sum_{t=1}^n \log f(\mathbf{w}_t; \boldsymbol{\xi}, \sigma_e^2) + \sum_{t=1}^n \log \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_t | \mathbf{x}_t; \boldsymbol{\theta}) f(\mathbf{x}_t | \mathbf{w}_t; \boldsymbol{\xi}, \sigma_e^2) d\mathbf{x}_t, \quad (2.9)
\end{aligned}$$

em que, $f(\mathbf{x}_t | \mathbf{w}_t; \boldsymbol{\xi}, \sigma_e^2)$ denota a função de densidade condicional de \mathbf{x}_t dado \mathbf{w}_t com distribuição $N_r(\boldsymbol{\mu}_{x_t | w_t}, \boldsymbol{\sigma}_{x_t | w_t}^2)$, com os elementos de $\boldsymbol{\mu}_{x_t | w_t}$ sendo $\mu_{xi} + [\sigma_{xi}^2 / (\sigma_{xi}^2 + \sigma_{ei}^2)](w_{ti} - \mu_{xi})$ e $\boldsymbol{\sigma}_{x_t | w_t}^2$ sendo uma matriz diagonal cujo i -ésimo elemento da diagonal é $\sigma_{xi}^2 \sigma_{ei}^2 / (\sigma_{xi}^2 + \sigma_{ei}^2)$, para todo $i = 1, 2, \dots, r$.

Para exemplificar os métodos de estimação abordados nas subseções seguintes, uma única covariável mensurada com erro é considerada. Entretanto, vale salientar que esses métodos podem ser facilmente generalizados para r covariáveis com erros de medida. Desta forma,

$$y_t | x_t, w_t \sim \text{Beta}(\mu_t, \phi_t), \quad (2.10)$$

$$g(\mu_t) = f_1(\mathbf{z}_t^\top, x_t; \boldsymbol{\alpha}, \beta), \quad h(\phi_t) = f_2(\mathbf{v}_t^\top, x_t; \boldsymbol{\gamma}, \lambda), \quad (2.11)$$

$$w_t = \tau_0 + \tau_1 x_t + e_t, \quad x_t \stackrel{\text{ind}}{\sim} N(\mu_x, \sigma_x^2), \quad e_t \stackrel{\text{ind}}{\sim} N(0, \sigma_e^2). \quad (2.12)$$

Da Equação (2.12) tem-se

$$w_t \stackrel{\text{ind}}{\sim} N(\tau_0 + \tau_1 \mu_x, \tau_1^2 \sigma_x^2 + \sigma_e^2), \quad x_t | w_t \stackrel{\text{ind}}{\sim} N(\mu_{x_t | w_t}, \sigma_{x_t | w_t}^2), \quad (2.13)$$

em que

$$\mu_{x_t | w_t} = \mu_x + k_x [w_t - (\tau_0 + \tau_1 \mu_x)], \quad \sigma_{x_t | w_t}^2 = \frac{\sigma_e^2 k_x}{\tau_1^2}, \quad (2.14)$$

com $k_x = \tau_1^2 \sigma_x^2 / (\tau_1^2 \sigma_x^2 + \sigma_e^2)$ denotada como coeficiente de confiabilidade (Carrasco *et al.*, 2014).

Desta forma, a Equação (2.9) pode ser reescrita como

$$\ell(\Psi) = \sum_{t=1}^n \ell_{1t}(\boldsymbol{\xi}, \sigma_e^2) + \sum_{t=1}^n \ell_{2t}(\boldsymbol{\theta}, \boldsymbol{\xi}, \sigma_e^2), \quad (2.15)$$

em que

$$\ell_{1t}(\boldsymbol{\xi}, \sigma_e^2) = -\frac{1}{2} \log[2\pi(\tau_1^2 \sigma_x^2 + \sigma_e^2)] - \frac{1}{2(\tau_1^2 \sigma_x^2 + \sigma_e^2)} [w_t - (\tau_0 + \tau_1 \mu_x)]^2 \quad (2.16)$$

e

$$\ell_{2t}(\boldsymbol{\theta}, \boldsymbol{\xi}, \sigma_e^2) = \log \int_{-\infty}^{\infty} f(y_t | x_t; \boldsymbol{\theta}, \boldsymbol{\xi}, \sigma_e^2) \frac{1}{\sqrt{2\pi\sigma_{x_t | w_t}^2}} \exp\left[-\frac{(x_t - \mu_{x_t | w_t})^2}{2\sigma_{x_t | w_t}^2}\right] dx_t. \quad (2.17)$$

2.2.1 Estimação por máxima verossimilhança aproximada

Quando o interesse é obter uma integral de uma função regular $f(x) : \mathbb{R} \rightarrow \mathbb{R}$ os métodos de integração numérica baseiam-se na expressão de $f(x)$ como o produto de uma função de pesos $\nu(x) : \mathbb{R} \rightarrow \mathbb{R}^+$ e outra função $g(x) : \mathbb{R} \rightarrow \mathbb{R}$, isto é,

$$f(x) = \nu(x)g(x).$$

Muitos desses métodos aproximam uma integral por uma soma ponderada do tipo

$$\int_{\mathbb{R}} f(x)dx \approx \sum_{q=1}^Q \nu_q g(x_q), \quad (2.18)$$

em que os pontos x_q denotam os nós e os ν_q denotam os pesos, e juntos constituem uma regra de integração (Fahrmeir e Tutz, 2001). Uma proposta de integral baseada em (2.18) é a quadratura de Gauss-Hermite (QGH) que tem sido amplamente utilizada, como mostram, por exemplo, Golub e Welsch (1967), Liu e Pierce (1994), Rabe-Hesketh *et al.* (2005), Villegas *et al.* (2011) e Manco (2013).

O método da quadratura de Gauss-Hermite aproxima integrais da forma

$$\int_{-\infty}^{\infty} e^{-x^2} f(x)dx,$$

por uma soma ponderada, isto é,

$$\int_{-\infty}^{\infty} e^{-x^2} f(x)dx \approx \sum_{q=1}^Q \nu_q f(s_q), \quad (2.19)$$

em que s_q e ν_q representa o q -ésimo zero e o peso, respectivamente, do polinômio ortogonal de Hermite de ordem Q . Aqui, Q denota o número de pontos de quadratura (Abramowitz and Stegun, 1964).

O segundo termo do logaritmo da função de verossimilhança $\ell(\Psi)$ na Equação (2.15), que depende de uma integral de inviável aproximação analítica, como visto na Equação (2.17), pode ser aproximado utilizando o método da quadratura de Gauss-Hermite.

Considerando a transformação $u_t = (x_t - \mu_{x_t|w_t})/\sqrt{2\sigma_{x_t|w_t}^2}$ na Equação (2.17), tem-se que $x_t = \mu_{x_t|w_t} + \sqrt{2\sigma_{x_t|w_t}^2} u_t$ e $dx_t = \sqrt{2\sigma_{x_t|w_t}^2} du_t$. Portanto, para $(y_1, \mathbf{w}_1), \dots, (y_n, \mathbf{w}_n)$, os n pares de variáveis aleatórias observadas, como no modelo proposto por Carrasco *et al.* (2014), o logaritmo da função de verossimilhança em (2.8), admitindo que uma única covariável é medida com erro, é dado por

$$\ell_a(\Psi) = \sum_{t=1}^n \ell_{1t}(\boldsymbol{\xi}, \sigma_e^2) + \sum_{t=1}^n \ell_{2t}(\boldsymbol{\theta}, \boldsymbol{\xi}, \sigma_e^2), \quad (2.20)$$

em que

$$\begin{aligned}\ell_{1t}(\boldsymbol{\xi}, \sigma_e^2) &= \log f(w_t; \boldsymbol{\xi}, \sigma_e^2) = -\frac{1}{2} \log[2\pi(\tau_1^2 \sigma_x^2 + \sigma_e^2)] - \frac{1}{2(\tau_1^2 \sigma_x^2 + \sigma_e^2)} [w_t - (\tau_0 + \tau_1 \mu_x)]^2 \quad \text{e} \\ \ell_{2t}(\boldsymbol{\theta}, \boldsymbol{\xi}, \sigma_e^2) &\simeq \log \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp[\ell_{tq}(\mu_{tq}, \phi_{tq})],\end{aligned}$$

sendo

$$\begin{aligned}\ell_{tq}(\mu_{tq}; \phi_{tq}) &= \log \Gamma(\phi_{tq}) - \log \Gamma(\mu_{tq} \phi_{tq}) - \log \Gamma[(1 - \mu_{tq}) \phi_{tq}] + (\mu_{tq} \phi_{tq} - 1) \log y_t + \\ &\quad [(1 - \mu_{tq}) \phi_{tq} - 1] \log(1 - y_t), \\ g(\mu_{tq}) &= f_1(\mathbf{z}_t^\top, x_t^*; \boldsymbol{\alpha}, \beta), \\ h(\phi_{tq}) &= f_2(\mathbf{v}_t^\top, x_t^*; \boldsymbol{\gamma}, \lambda), \\ x_t^* &= \mu_{x_t|w_t} + \sqrt{2\sigma_{x_t|w_t}^2} s_q, \\ \mu_{x_t|w_t} &= \mu_x + k_x [w_t - (\tau_0 + \tau_1 \mu_x)], \\ \sigma_{x_t|w_t}^2 &= \frac{\sigma_e^2 k_x}{\tau_1^2}, \\ k_x &= \frac{\tau_1^2 \sigma_x^2}{(\tau_1^2 \sigma_x^2 + \sigma_e^2)}, \quad \text{em que } k_x \text{ pode ser conhecido.}\end{aligned}$$

É importante lembrar que as funções $f_1(\cdot)$ e $f_2(\cdot)$ são funções não lineares, contínuas e diferenciáveis, tal que $\mathbf{F}_1 = \partial g(\mu_{tq})/\partial \boldsymbol{\alpha}$, $F_2 = \partial g(\mu_{tq})/\partial \beta$, $\mathbf{F}_3 = \partial h(\phi_{tq})/\partial \boldsymbol{\gamma}$ e $F_4 = \partial h(\phi_{tq})/\partial \lambda$ têm postos p , 1, \checkmark e 1, respectivamente. Para evitar problemas de identificabilidade nos modelos com erros de medida é usual assumir que a variância do erro de medida σ_e^2 é conhecida (Carrasco, 2012). O mesmo se aplica aos modelos não lineares com erros de medida.

Calculando as derivadas parciais do logaritmo da função de verossimilhança aproximada com respeito aos vetores de parâmetros $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \beta, \boldsymbol{\gamma}^\top, \lambda)^\top$ e de perturbação $\boldsymbol{\xi}^\top = (\mu_x, \sigma_x^2)^\top$, obtêm-se os componentes do vetor escore, denotado por

$$U_{\boldsymbol{\theta}}(\boldsymbol{\theta}^\top, \boldsymbol{\xi}^\top) = (U_{\boldsymbol{\alpha}}^\top, U_\beta, U_{\boldsymbol{\gamma}}^\top, U_\lambda, U_{\mu_x}, U_{\sigma_x^2})^\top,$$

com

$$\begin{aligned}U_{\boldsymbol{\alpha}}(\boldsymbol{\theta}) &= \mathbf{F}_1^\top \left[(\boldsymbol{\Phi} \circ \mathbf{G})(\mathbf{P} \circ \mathbf{V})^\top \circ \mathbf{B} \right] \mathbf{L} \\ U_\beta(\boldsymbol{\theta}) &= \mathbf{F}_2^\top \left[(\boldsymbol{\Phi} \circ \mathbf{G})(\mathbf{P} \circ \mathbf{V})^\top \circ \mathbf{B} \right] \mathbf{L} \\ U_{\boldsymbol{\gamma}}(\boldsymbol{\theta}) &= \mathbf{F}_3^\top \left[\mathbf{H} \mathbf{a}^{*\top} \circ \mathbf{B} \right] \mathbf{L}, \\ U_\lambda(\boldsymbol{\theta}) &= \mathbf{F}_4^\top \left[\mathbf{H} \mathbf{a}^{*\top} \circ \mathbf{B} \right] \mathbf{L}, \\ U_{\mu_x}(\boldsymbol{\xi}) &= \mathbf{W}_{\mu_x} + \mathbf{B}^\top \mathbf{A}_{\mu_x} \quad \text{e} \\ U_{\sigma_x^2}(\boldsymbol{\xi}) &= \mathbf{W}_{\sigma_x^2} + \mathbf{B}^\top \mathbf{A}_{\sigma_x^2},\end{aligned}$$

em que “ \circ ” denota o produto de Hadamard (elemento a elemento), \mathbf{L} é um vetor $n \times 1$ de uns. $\mathbf{F}_1 = \partial\eta_1/\partial\boldsymbol{\alpha}$ e $\mathbf{F}_3 = \partial\eta_2/\partial\boldsymbol{\gamma}$ são matrizes de derivadas de ordens $(n \times p)$ e $(n \times \check{q})$, respectivamente. $\mathbf{F}_2 = \partial\eta_1/\partial\beta$ e $\mathbf{F}_4 = \partial\eta_2/\partial\lambda$ são vetores de derivadas r -dimensional e s -dimensional, respectivamente, em que $r = s = 1$. $\boldsymbol{\Phi}$, \mathbf{G} , \mathbf{P} , \mathbf{H} são matrizes de ordem $(n \times q)$ e \mathbf{a}^* e o vetor n -dimensional \mathbf{B} são definidas no Apêndice 7.1.

O estimador de máxima verossimilhança de $\boldsymbol{\Psi}$, $\hat{\boldsymbol{\Psi}}$, é solução do sistema de equações $\mathbf{U}\boldsymbol{\alpha} = 0$, $U_\beta = 0$, $\mathbf{U}\boldsymbol{\gamma} = 0$, $U_\lambda = 0$, $U_{\mu_x} = 0$ e $U_{\sigma_x^2} = 0$, em que a utilização de métodos iterativos é necessário para sua solução. Algumas propriedades assintóticas dos estimadores de máxima verossimilhança aproximada serão apresentadas no Capítulo 3.

2.2.2 Estimação por pseudo-verossimilhança aproximada

Um dos métodos mais conhecidos e comumente utilizados na estimação dos parâmetros de um modelo estatístico é o método de máxima verossimilhança. Segundo Cramér (1999), as equações de verossimilhança possuem raízes fortemente consistentes, assintoticamente eficientes e distribuídas normalmente, válidas condições usuais de regularidade (Cox and Hinkley, 1974).

Na presença de parâmetros de perturbação, diversos métodos de estimação alternativos vêm sendo propostos, tais como o método de máxima verossimilhança parcial (Cox, 1975), máxima verossimilhança condicional (Godambe, 1991) e máxima pseudo-verossimilhança (Gong e Samaniego, 1981). De maneira geral, os processos de estimação na presença de parâmetros de perturbação estão focados em eliminá-los.

Aqui utiliza-se o método de máxima pseudo-verossimilhança com o objetivo de contornar a presença dos parâmetros de perturbação referentes às distribuições da variável medida com erros. A ideia principal deste método é substituir o vetor de parâmetros de perturbação por uma estimativa consistente na função de máxima verossimilhança original (2.15), gerando assim uma função de pseudo-verossimilhança que depende somente dos parâmetros de interesse. As estimativas de máxima pseudo-verossimilhança poderão ser calculadas por algum método confiável (Carrasco *et al.*, 2014). Nosso intuito é explorar este método no modelo de regressão beta não linear com erros nas variáveis.

Considere $(\boldsymbol{\xi}^\top, \sigma_e^2)^\top$, em que $\boldsymbol{\xi}^\top = (\mu_x, \sigma_x^2)^\top$ é o vetor de parâmetros de perturbação e $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \beta, \boldsymbol{\gamma}^\top, \lambda)^\top$ é o vetor de parâmetros de interesse. Como não é possível eliminar o vetor $\boldsymbol{\xi}$ em (2.9) através de integração, utiliza-se a abordagem de pseudo-verossimilhança para estimar os parâmetros (Gong e Samaniego, 1981; Guolo, 2011). Desta forma, o logaritmo da função verossimilhança (2.15) é maximizado em duas etapas. Na primeira etapa, estima-se o vetor de parâmetros de perturbação $\boldsymbol{\xi}$ a partir da maximização do logaritmo da função de verossimilhança reduzida, como sugerem Buonaccorsi e Tosteson

(1993) e Guolo (2011). Ou seja,

$$\ell_r(\boldsymbol{\xi}, w) = \sum_{t=1}^n \log f(w; \boldsymbol{\xi}, \sigma_e^2) \quad (2.21)$$

denota o logaritmo da função de verossimilhança reduzida, em que a variância do erro de medição, dada por σ_e^2 , é assumida conhecida ou é estimada quando existem réplicas para w_t . Na segunda etapa, a estimativa de $\boldsymbol{\xi}$, denotada por $\hat{\boldsymbol{\xi}}$, obtida da maximização de (2.21) é substituída no logaritmo da função de verossimilhança original (2.15), o que resulta no logaritmo da pseudo-verossimilhança

$$\ell_p(\boldsymbol{\theta}, \hat{\boldsymbol{\xi}}, \hat{\sigma}_e^2) = \sum_{t=1}^n \ell_{1t}(\hat{\boldsymbol{\xi}}, \hat{\sigma}_e^2) + \sum_{t=1}^n \ell_{2t}(\boldsymbol{\theta}, \hat{\boldsymbol{\xi}}, \hat{\sigma}_e^2), \quad (2.22)$$

sendo $\hat{\sigma}_e^2$ o valor conhecido ou estimado de σ_e^2 .

Vale salientar que, assim como $\ell_a(\boldsymbol{\Psi})$, a função $\ell_p(\boldsymbol{\theta}, \hat{\boldsymbol{\xi}}, \hat{\sigma}_e^2)$ é analiticamente intratável. Contudo, ao contrário do que ocorre com $\ell_a(\boldsymbol{\Psi})$, em $\ell_p(\boldsymbol{\theta}, \hat{\boldsymbol{\xi}}, \hat{\sigma}_e^2)$ a parcela que envolve uma integral depende apenas dos parâmetros de interesse. A maximização de $\ell_p(\boldsymbol{\theta}, \hat{\boldsymbol{\xi}}, \hat{\sigma}_e^2)$ também será solucionado pelo método de quadratura de Gauss-Hermite. Desta forma, o logaritmo da pseudo-verossimilhança aproximada para o modelo de regressão beta não linear com erros de medida unidimensional é dado por (2.22), sendo

$$\ell_{1t}(\hat{\boldsymbol{\xi}}, \hat{\sigma}_e^2) = -\frac{1}{2} \log[2\pi(\tau_1^2 \hat{\sigma}_x^2 + \hat{\sigma}_e^2)] - \frac{1}{2(\tau_1^2 \hat{\sigma}_x^2 + \hat{\sigma}_e^2)} [w_t - (\tau_0 + \tau_1 \hat{\mu}_x)]^2$$

e

$$\ell_{2t}(\boldsymbol{\theta}, \hat{\boldsymbol{\xi}}, \hat{\sigma}_e^2) \simeq \sum_{t=1}^n \log \left(\sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp[\ell_{tq}(\boldsymbol{\theta}; \hat{\boldsymbol{\xi}}, \hat{\sigma}_e^2)] \right),$$

em que

$$\begin{aligned} \ell_{tq}(\boldsymbol{\theta}; \hat{\boldsymbol{\xi}}, \hat{\sigma}_e^2) &= \log \Gamma(\phi_{tq}) - \log \Gamma(\mu_{tq} \phi_{tq}) - \log \Gamma[(1 - \mu_{tq}) \phi_{tq}] + (\mu_{tq} \phi_{tq} - 1) \log y_t + \\ &\quad [(1 - \mu_{tq}) \phi_{tq} - 1] \log(1 - y_t), \\ g(\mu_{tq}) &= f_1(\mathbf{z}_t^\top, x_t^*; \boldsymbol{\alpha}, \beta), \\ h(\phi_{tq}) &= f_2(\mathbf{v}_t^\top, x_t^*; \boldsymbol{\gamma}, \lambda), \\ x_t^* &= \hat{\mu}_{x_t|w_t} + \sqrt{2\hat{\sigma}_{x_t|w_t}^2} s_q, \\ \hat{\mu}_{x_t|w_t} &= \hat{\mu}_x + \hat{k}_x [w_t - (\tau_0 + \tau_1 \hat{\mu}_x)], \\ \hat{\sigma}_{x_t|w_t}^2 &= \frac{\hat{\sigma}_e^2 \hat{k}_x}{\tau_1^2}, \\ \hat{k}_x &= \frac{\tau_1^2 \hat{\sigma}_x^2}{(\tau_1^2 \hat{\sigma}_x^2 + \hat{\sigma}_e^2)}. \end{aligned}$$

Aqui, s_q e ν_q representam o q -ésimo zero e peso, respectivamente, do polinômio ortogonal de Hermite de ordem Q . As Equações (2.21) e (2.22) são idênticas às obtidas no modelo

proposto por Carrasco *et al.* (2014), a diferença está em $g(\mu_t)$ e $h(\phi_t)$ que são formadas por equações não lineares.

Calculando as derivadas parciais de $\ell_p(\boldsymbol{\theta}; \hat{\boldsymbol{\xi}}, \hat{\sigma}_e^2)$ com respeito ao vetor de parâmetros de interesse $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \beta, \boldsymbol{\gamma}^\top, \lambda)^\top$, obtêm-se os componentes do vetor pseudo-escore representado por $U_p(\boldsymbol{\theta}) = (U_{p\boldsymbol{\alpha}}^\top, U_{p\beta}, U_{p\boldsymbol{\gamma}}^\top, U_{p\lambda})^\top$, com

$$\begin{aligned} U_{p\boldsymbol{\alpha}}(\boldsymbol{\theta}) &= \mathbf{F}_1^\top \left[(\boldsymbol{\Phi} \circ \mathbf{G}) (\mathbf{P} \circ \mathbf{V})^\top \circ \mathbf{B} \right] \mathbf{L} \\ U_{p\beta}(\boldsymbol{\theta}) &= \mathbf{F}_2^\top \left[(\boldsymbol{\Phi} \circ \mathbf{G}) (\mathbf{P} \circ \mathbf{V})^\top \circ \mathbf{B} \right] \mathbf{L} \\ U_{p\boldsymbol{\gamma}}(\boldsymbol{\theta}) &= \mathbf{F}_3^\top \left[\mathbf{H}\mathbf{a}^{*\top} \circ \mathbf{B} \right] L, \\ U_{p\lambda}(\boldsymbol{\theta}) &= \mathbf{F}_4^\top \left[\mathbf{H}\mathbf{a}^{*\top} \circ \mathbf{B} \right] L. \end{aligned}$$

As matrizes são obtidas da mesma forma que as apresentadas para a máxima verossimilhança aproximada, considerando \hat{x}_t^* . O estimador de máxima pseudo-verossimilhança de $\boldsymbol{\theta}$, $\tilde{\boldsymbol{\theta}}$, é solução do sistema de equações não linear $\mathbf{U}_p\boldsymbol{\alpha} = 0$, $U_{p\beta} = 0$, $U_{p\boldsymbol{\gamma}} = 0$ e $U_{p\lambda} = 0$, que requer também a utilização de métodos iterativos para a sua solução.

2.2.3 Estimação por calibração da regressão

O método de estimação por calibração da regressão vem sendo muito utilizado na última década por sua facilidade na estimação de parâmetros quando se evidencia erros de mensuração nas covariáveis, ver Carroll *et al.* (2006), Freedman *et al.* (2008), Thurston *et al.* (2005) e Guolo (2011) para maiores detalhes. A principal ideia deste método é substituir a variável não observável, x_t por uma estimativa da esperança condicional de x_t dado w_t , $\mathbb{E}(x_t|w_t)$, denotada por *função de calibração*, que depende de um vetor de parâmetros $\boldsymbol{\kappa}$. A troca da variável não observada x_t por $r(w_t, \hat{\boldsymbol{\kappa}}) = \mathbb{E}(x_t|w_t)$ estabelece um modelo modificado para os dados. Aqui, $\hat{\boldsymbol{\kappa}}$ é uma estimativa de $\boldsymbol{\kappa}$.

Assim como no modelo de regressão beta com erros de medida sob o enfoque estrutural, o modelo de regressão beta não linear com erros de medida dado pelas Equações (2.2), (2.3), (2.4) e (2.6), com uma única variável medida com erro, apresenta que a função de calibração dada por

$$r(w_t, \psi) = \mu_x + k_x [w_t - (\tau_0 + \tau_1 \mu_x)],$$

em que $\boldsymbol{\kappa} = (\mu_x, \sigma_x^2, \sigma_e^2)^\top$ e $k_x = \tau_1^2 \sigma_x^2 / (\tau_1^2 \sigma_x^2 + \sigma_e^2)$ é conhecido como coeficiente de confiabilidade. Sob a suposição de que $w_t \sim N(\tau_0 + \tau_1 \mu_x, \tau_1^2 \sigma_x^2 + \sigma_e^2)$, segue que $\bar{w} = \sum_{t=1}^n w_t / n$ e $s_w^2 = \sum_{t=1}^n (w_t - \bar{w})^2 / (n - 1)$ são estimadores ótimos de $\tau_0 + \tau_1 \mu_x$ e $\tau_1^2 \sigma_x^2 + \sigma_e^2$, respectivamente. Entretanto, não é possível estimar σ_x^2 através dos dados observados w_1, \dots, w_n . Assume-se, então, que k_x ou σ_e^2 é conhecido, ou, alternativamente, que σ_e^2 pode ser estimado desde que seja possível observar réplicas de w_t .

Ao substituir $r(w_t, \psi)$ estimada em $f(y_t|x_t; \boldsymbol{\alpha}, \beta, \boldsymbol{\gamma}, \lambda)$ dado em (2.2), ou seja, a função de calibração estimada na função densidade de probabilidade de y_t dado x_t , obtém-se o logaritmo da função de verossimilhança modificada, que é

$$\ell_{rc}(\boldsymbol{\theta}) = \sum_{t=1}^n \ell(\mu_t, \phi_t),$$

sendo

$$\begin{aligned} \ell(\mu_t, \phi_t) &= \log \Gamma(\phi_t) - \log \Gamma(\mu_t \phi_t) - \log \Gamma[(1 - \mu_t)\phi_t] + (\mu_t \phi_t - 1) \log y_t + \\ &\quad [(1 - \mu_t)\phi_t - 1] \log(1 - y_t), \\ g(\mu_t) &= f_1(\mathbf{z}_t^\top, x_t^*; \boldsymbol{\alpha}, \beta), \\ h(\phi_t) &= f_2(\mathbf{v}_t^\top, x_t^*; \boldsymbol{\gamma}, \lambda), \\ x_t^* &= \bar{w} + \hat{k}_x[w_t - (\tau_0 + \tau_1 \bar{w})], \end{aligned}$$

em que \hat{k}_x é o estimador de k_x .

Observe que o logaritmo da função de verossimilhança modificada depende apenas do parâmetro de interesse $\boldsymbol{\theta}$ e coincide com o do modelo de regressão beta usual, uma vez que x_t^* atua como uma variável explicativa, medida sem erro. Novamente, os resultados obtidos são idênticos ao do modelo proposto por Carrasco *et al.* (2014), exceto para $g(\mu_t)$ e $h(\phi_t)$.

Derivando o $\ell_{rc}(\hat{\boldsymbol{\theta}})$ com respeito ao vetor de parâmetros de interesse $\boldsymbol{\theta}$, obtemos os componentes do vetor escore baseado na calibração da regressão representado por $U_{rc}(\boldsymbol{\theta}) = (U_{rc\boldsymbol{\alpha}}^\top, U_{rc\beta}, U_{rc\boldsymbol{\gamma}}, U_{rc\lambda})^\top$, definidos por

$$\begin{aligned} U_{rc\boldsymbol{\alpha}}(\boldsymbol{\theta}) &= \mathbf{F}_1^\top \boldsymbol{\Phi} \mathbf{T} (\mathbf{y}^* - \boldsymbol{\mu}^*) \\ U_{rc\beta}(\boldsymbol{\theta}) &= \mathbf{F}_2^\top \boldsymbol{\Phi} \mathbf{T} (\mathbf{y}^* - \boldsymbol{\mu}^*) \\ U_{rc\boldsymbol{\gamma}}(\boldsymbol{\theta}) &= \mathbf{F}_3^\top \mathbf{H} \mathbf{a} \\ U_{rc\lambda}(\boldsymbol{\theta}) &= \mathbf{F}_4^\top \mathbf{H} \mathbf{a}, \end{aligned}$$

em que $\mathbf{F}_1^\top = \partial \eta_1 / \partial \boldsymbol{\alpha}$ é uma matriz de ordem $n \times p$, $\mathbf{F}_2^\top = \partial \eta_1 / \partial \beta$ é uma vetor r -dimensional, $\mathbf{F}_3^\top = \partial \eta_2 / \partial \boldsymbol{\gamma}$ é uma matriz de ordem $n \times \check{q}$ e $\mathbf{F}_4^\top = \partial \eta_2 / \partial \lambda$ é um vetor s -dimensional, em que $r = s = 1$. $\boldsymbol{\Phi} = \text{diag}\{\phi_t, \dots, \phi_n\}$, $\mathbf{T} = \text{diag}\{1/g'(\mu_1), \dots, 1/g'(\mu_n)\}$, $\mathbf{H} = \text{diag}\{1/h'(\phi_1), \dots, 1/h'(\phi_n)\}$ e $\mathbf{a} = \text{diag}(a_1, \dots, a_n)^\top$.

O estimador de máxima verossimilhança usando calibração da regressão de $\boldsymbol{\theta}$, $\check{\boldsymbol{\theta}}$, é solução simultânea do sistema de equações $U_{rc\boldsymbol{\alpha}} = 0$, $U_{rc\beta} = 0$, $U_{rc\boldsymbol{\gamma}} = 0$ e $U_{rc\lambda} = 0$, que requer também a utilização de métodos iterativos para calcular a solução.

2.2.4 Valores iniciais para o algoritmo de otimização

Como dito anteriormente, os estimadores aqui propostos não possuem forma fechada. Assim, é necessário utilizar algoritmos de otimização não linear tais como, Newton-Raphson,

Escore de Fisher, BHHH ou BFGS, como métodos iterativos de maximização.

Esses processos de maximização necessitam da especificação de valores iniciais para os parâmetros a serem estimados. Aqui, utilizou-se o valor inicial proposto por Espinheira *et al.* (2017), que será sucintamente apresentado.

Quando $p_1 = p$ e $\check{q}_1 = \check{q}$, uma proposta de valor inicial utiliza a linearização da função não linear avaliada com base na expansão em série de Taylor. Seja $f(z_t, \alpha)$ a função não linear. Sua expansão até a primeira ordem ao redor do ponto $\alpha^{(0)}$ é

$$f(z_t, \alpha) \approx f(z_t, \alpha^{(0)}) + \sum_{i=1}^p \left[\frac{\partial f(z_t, \alpha)}{\partial \alpha_i} \right]_{\alpha=\alpha^{(0)}} (\alpha_i - \alpha_i^{(0)}),$$

em que $\alpha_i^{(0)} = (\alpha_1^{(0)}, \dots, \alpha_p^{(0)})$ é um valor inicial. Assim,

$$f(z_t, \alpha) = f(z_t, \alpha^{(0)}) + \sum_{i=1}^p j_{ti}^{(0)} (\alpha_i - \alpha_i^{(0)}). \quad (2.23)$$

Analogamente ao que é feito no modelo normal não linear em que $y_t = f(z_t, \alpha)$, aqui consideraremos $g(y_t) = f(z_t, \alpha)$. Adicionalmente, supondo $\theta_i^{(0)} = (\alpha_i - \alpha_i^{(0)})$, segue que

$$g(y_t) - f(z_t, \alpha^{(0)}) = \sum_{i=1}^p j_{ti}^{(0)} \theta_i^{(0)},$$

que pode ser visto como um modelo linear e a aproximação por mínimos quadrados de $\theta^{(0)}$ será dado por

$$\hat{\theta}^{(0)} = (J^{(0)\top} J^{(0)})^{-1} J^{(0)\top} (g(y) - f(z, \alpha^{(0)})), \quad (2.24)$$

em que $J^{(0)} = [\partial \eta_1 / \partial \alpha]_{\alpha=\alpha^{(0)}}$ e $\hat{\theta}_i^0 = (\hat{\alpha}_i - \alpha_i^{(0)})$, então $\alpha_i^{(0)} = \hat{\theta}_i^0 + \hat{\alpha}_i$. Essa relação pode ser generalizada para construir um processo iterativo de passo m tal que $\alpha_i^m = \theta_i^{(m-1)} + \alpha_i^{(m-1)}$. Generalizando (2.24) temos que

$$\alpha^{(m)} = (J^{(m-1)\top} J^{(m-1)})^{-1} J^{(m-1)\top} (g(y) - f(z, \alpha^{(m-1)})) + \alpha^{(m-1)}.$$

A proposta de Espinheira *et al.* (2017) é considerar

$$\alpha_{NL}^{(0)} = (J_1^{(0)\top} J_1^{(0)})^{-1} J_1^{(0)\top} (g(y) - f(z, \alpha_L^{(0)})) \quad \text{e} \quad \alpha_L^{(0)} = (Z^\top Z)^{-1} Z^\top g(y).$$

Já para o modelo da precisão o mesmo raciocínio será utilizado, mas agora o parâmetro será γ e na equação equivalente a (4) considera-se $h(\phi_t) = f(v_t, \gamma)$, o que implica

$$\gamma_{NL}^{(0)} = (J_2^{(0)\top} J_2^{(0)})^{-1} J_2^{(0)\top} \left(h(\phi_{NL}^{(0)}) - f(v, \gamma_L^{(0)}) \right) \quad \text{e} \quad \alpha_L^{(0)} = (V^\top V)^{-1} V^\top h(\phi_L^{(0)}). \quad (2.25)$$

Aqui, $J_2^{(0)} = [\partial \eta_2 / \partial \gamma]_{\gamma=\gamma^{(0)}}$,

$$\phi_{Lt}^{(0)} = \frac{\check{\mu}_{NLt}(1 - \check{\mu}_{NLt})}{\hat{\sigma}_{NLt}^2} - 1,$$

em que $\mu_{NLt} = g^{-1}(f_1(z_t^\top; \alpha_L))$, $\hat{\sigma}_{NLt}^2 = \check{e}_{NL}^\top \check{e}_{NL} / [(n-p)g'(\check{\mu}_{NLt})]$ com $\check{e}_{NL} = g(y) - \check{\mu}_{NL}$.

Para obter os valores iniciais para os modelos de regressão beta não lineares com erros de medida foi utilizado a proposta de Espinheira *et al.* (2017) descrita anteriormente.

2.3 Avaliação numérica

Realizou-se um estudo de simulação de Monte Carlo com $R = 10000$ réplicas cujo objetivo é investigar o comportamento do viés e do erro quadrático médio dos estimadores dos parâmetros do modelo obtidos pelos métodos de estimação descritos nas seções anteriores. Para $\Psi = (\Psi_1, \Psi_2, \Psi_3, \Psi_4, \Psi_5, \Psi_6)^\top = (\alpha, \beta, \gamma, \lambda, \mu_x, \sigma_x^2)^\top$, tem-se que

$$\bar{\Psi}_j = \frac{1}{R} \sum_{i=1}^R \hat{\Psi}_j^{(i)}, \quad \widehat{\text{Viés}}(\hat{\Psi}_j) = \bar{\Psi}_j - \Psi_j \quad \text{e} \quad \text{REQM}(\hat{\Psi}_j) = \sqrt{\frac{1}{R} \sum_{i=1}^R (\hat{\Psi}_j^{(i)} - \Psi_j)^2},$$

em que $j = 1, \dots, 6$. $\widehat{\text{Viés}}(\hat{\Psi}_j)$ e $\text{REQM}(\hat{\Psi}_j)$ são o viés estimado e a raiz do erro quadrático médio dos estimadores dos parâmetros. Aqui, $\hat{\Psi}_j^{(i)}$ denota a estimativa de Ψ_j obtida na i -ésima amostra simulada.

Cinco cenários foram analisados para o modelo de regressão beta não linear com erros de medida sob o enfoque estrutural. Geraram-se diferentes amostras aleatórias de tamanho n igual a 40, 80 e 160 para os quatro primeiros cenários e para o quinto cenário as amostras aleatórias possuem tamanho n igual a 60, 80 e 160. No último cenário avaliamos um modelo mais complexo, pois ele apresenta não linearidade nos submodelos da média e da dispersão. Desta forma, foi necessário aumentar o tamanho amostral para diminuir o problema de convergência.

Para essas amostras foram considerados valores verdadeiros de x , que segue distribuição normal com média μ_x e variância σ_x^2 . As ligações logito e logarítmica foram escolhidas, respectivamente, para os parâmetros μ_t e ϕ_t obtidos a partir das Equações (2.3) e (2.4) as quais envolvem parâmetros desconhecidos, para os quais também foram atribuídos valores. As ocorrências da variável dependente y_t foram geradas a partir da distribuição beta com média μ_t e precisão ϕ_t . Depois, geraram-se ocorrências da covariável w_t a partir de sua distribuição condicional dado x_t , ou seja, com uma distribuição normal com média x_t e variância σ_e^2 . O coeficiente de confiabilidade do modelo foi considerado conhecido e os seguintes valores foram atribuídos $k_x = 0.50$ (erro de medida alto), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.95$ (erro de medida baixo), que representam a $\sigma_e^2 = \sigma_x^2$, $\sigma_e^2 = \sigma_x^2/3$ e $\sigma_e^2 = \sigma_x^2/19$, respectivamente. A variável w foi utilizada nos estimadores dos parâmetros de interesse descritos anteriormente e respectivos intervalos de confiança que serão apresentados no Capítulo 3.

Para cada amostra simulada, os parâmetros do modelo foram estimados sob dois enfoques: primeiro, assumimos (erroneamente) que não há erros de medida na covariável x_t , ou seja, utilizamos o método *naive* (*Naive*); segundo, reconhecemos que x_t é medida com erros e, nesse caso, utilizamos os três métodos descritos nas seções anteriores, a saber: máxima verossimilhança aproximada (MVa), máxima pseudo-verossimilhança aproximada

(MPVa) e calibração da regressão (RC).

2.3.1 Cenário 1: Dispersão constante

Neste cenário admite-se que o modelo de regressão beta não linear com erros de medida com dispersão constante, sob o enfoque estrutural, apresenta a seguinte estrutura:

$$\begin{aligned}g(\mu_t) &= \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}, \\h(\phi) &= \gamma_0, \\w_t &= x_t + e_t,\end{aligned}$$

em que $t = 1, 2, \dots, n$. Para cada réplica de Monte Carlo, geraram-se ocorrências da variável sem erro de medida z_{t1} com distribuição uniforme (0.2, 1.2). Os seguintes valores para os parâmetros que permanecem fixos para todas as simulações foram escolhidos: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Neste cenário, considerou-se $Q = 50$ pontos de quadratura para aproximar os métodos de estimação. Como para cada réplica de Monte Carlo são geradas diferentes amostras aleatórias de tamanhos $n = 40, 80$ e 160, não é possível obter um único intervalo para a média da variável resposta.

Tabela 2.1: Limites inferiores médios (LI) e superiores médios (LS) e suas respectivas medianas (Md) para a média da variável resposta (μ_t) considerando o modelo de regressão beta não linear com erros de medida com dispersão constante para diferentes coeficientes de confiabilidade (k_x) e tamanhos amostrais (n).

k_x	n	μ_t		
		LI	LS	Md
0.95	40	0.1354	0.8832	0.5135
	80	0.1089	0.8970	0.4968
	160	0.0875	0.9186	0.4942
0.75	40	0.1355	0.8836	0.5136
	80	0.1085	0.8974	0.4967
	160	0.0879	0.9195	0.4944
0.50	40	0.1350	0.8836	0.5135
	80	0.1085	0.8972	0.4969
	160	0.0879	0.9192	0.4941

A Tabela 2.1 apresenta os limites inferiores e superiores médios e suas respectivas medianas da média da variável resposta (μ_t) para os diferentes tamanhos amostrais. Note que μ_t não é variável aleatória, esse procedimento é realizado para verificar em que subintervalo do intervalo unitário a média da variável resposta se encontra.

Considerando o tamanho amostral fixo, pode-se verificar que há uma pequena diferença entre as medidas avaliadas quando se aumenta o coeficiente de confiabilidade (k_x). Além disso, vale salientar que os valores dos parâmetros considerados cobrem

praticamente todo o espaço paramétrico de μ_t e estão bem distribuídos dentro do intervalo (0,1), como mostra a mediana. O valor do parâmetro γ_0 conduziu a $\phi \approx 150$ (148.41), constante. Além disso, neste cenário o grau de heterogeneidade, definido como $\delta = \max(\phi_1, \dots, \phi_n) / \min(\phi_1, \dots, \phi_n)$, é igual a 1.

Tabela 2.2: Raízes dos erros quadráticos médios (REQM) para os estimadores de α_0 , α_1 , β_1 e γ_0 considerando o modelo com dispersão constante. Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 1.

k_x	n	Método	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_0$
0.95	40	<i>Naive</i>	0.04294	0.33423	0.06304	0.64440
		RC	0.04368	0.33423	0.04649	0.64440
		MVa	0.04375	0.33635	0.04738	0.68518
		MPVa	0.04375	0.33635	0.04738	0.68501
	80	<i>Naive</i>	0.03364	0.27754	0.05507	0.67067
		RC	0.03401	0.27754	0.03276	0.68547
		MVa	0.03386	0.27827	0.03298	0.41468
		MPVa	0.03386	0.27827	0.03298	0.41465
	160	<i>Naive</i>	0.02236	0.18087	0.05087	0.68547
		RC	0.02260	0.18087	0.02340	0.68547
		MVa	0.02252	0.17968	0.02303	0.25614
		MPVa	0.02252	0.17968	0.02303	0.25613
0.75	40	<i>Naive</i>	0.07456	0.56166	0.23366	1.72891
		RC	0.08727	0.56166	0.12134	1.72891
		MVa	0.08943	0.58798	0.11511	1.72876
		MPVa	0.08910	0.58655	0.11044	1.65586
	80	<i>Naive</i>	0.05883	0.47678	0.22946	1.77145
		RC	0.06503	0.47678	0.07799	1.77145
		MVa	0.06625	0.49010	0.07596	1.32015
		MPVa	0.06607	0.48920	0.07409	1.28590
	160	<i>Naive</i>	0.03919	0.31980	0.22779	1.78955
		RC	0.04319	0.31980	0.05821	1.78955
		MVa	0.04358	0.31383	0.05260	0.98699
		MPVa	0.04351	0.31322	0.05203	0.97005
0.50	40	<i>Naive</i>	0.10013	0.74308	0.43476	2.31854
		RC	0.81127	0.74308	3.64234	2.31854
		MVa	0.15825	0.80544	0.29861	1.69729
		MPVa	0.16722	0.80079	0.38526	1.55503
	80	<i>Naive</i>	0.07957	0.63399	0.43181	2.36738
		RC	0.11146	0.63399	0.47713	2.36738
		MVa	0.10889	0.67340	0.15704	1.36716
		MPVa	0.10911	0.66527	0.17644	1.26691
	160	<i>Naive</i>	0.05360	0.45177	0.43224	2.38938
		RC	0.07009	0.45177	0.12758	2.38938
		MVa	0.07266	0.44610	0.09885	1.11365
		MPVa	0.07246	0.44038	0.10124	1.05377

A Tabela 2.2 apresenta os resultados da raiz do erro quadrático médio (REQM) dos

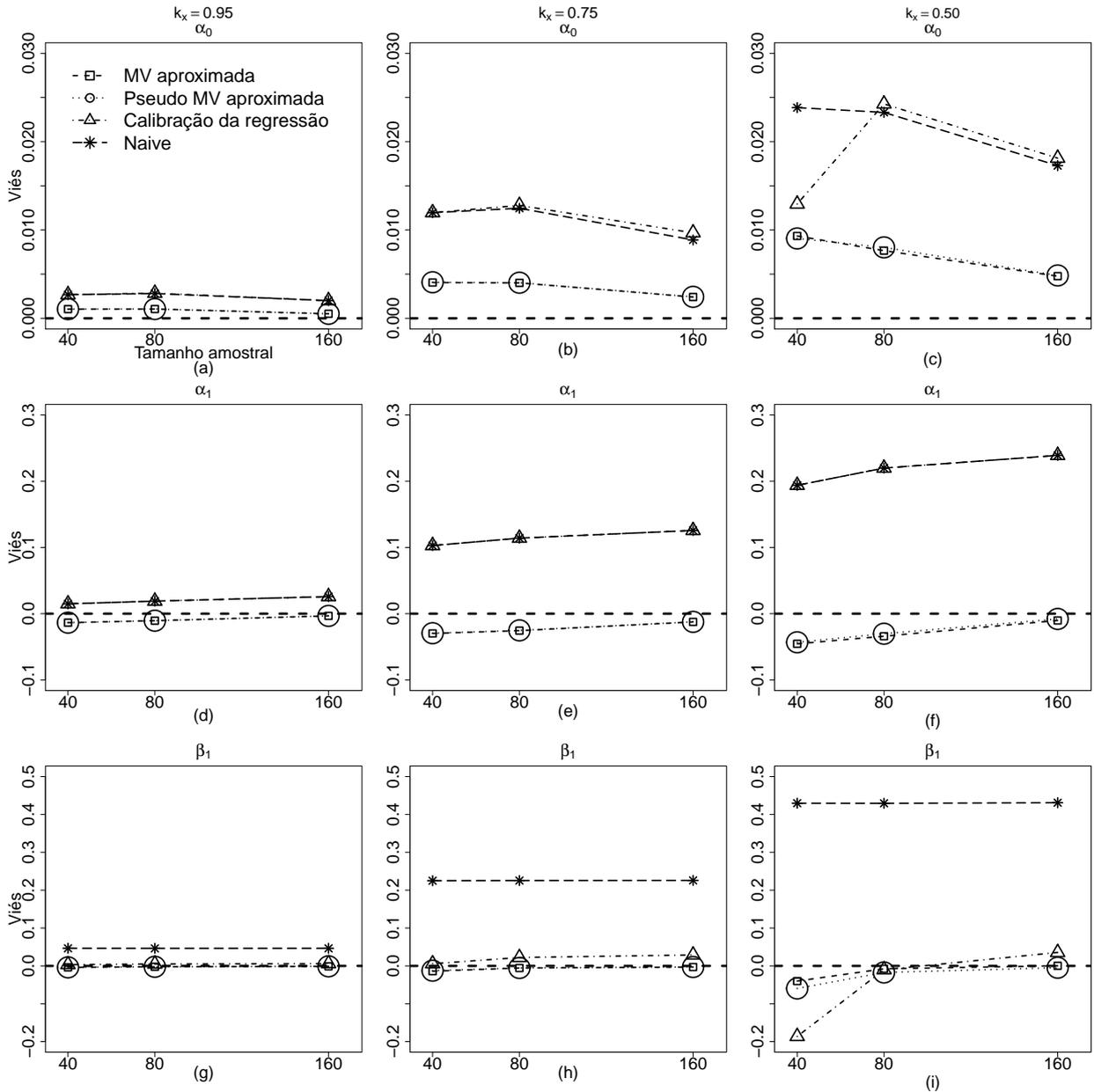


Figura 2.1: Vieses estimados dos estimadores de α_0 , α_1 e β_1 considerando os coeficientes de confiabilidade $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 1: Modelo com dispersão constante.

estimadores considerando o modelo com dispersão constante. Note que a raiz do erro quadrático médio diminui com o aumento do tamanho amostral para todos os coeficientes de confiabilidade considerados, exceto para o parâmetro de dispersão, em que a raiz do erro quadrático médio aumenta para os métodos *Naive* e calibração da regressão (RC). Por exemplo, considerando o método RC, quando $k_x = 0.50$, temos que para $n = (40, 80, 160)$, $\text{REQM}(\gamma_0) = (2.31854, 2.36738, 2.38938)$, respectivamente.

Considerando o parâmetro associado à variável com erro de medida, β_1 , tem-se que os métodos de máxima verossimilhança aproximada (MVA) e máxima pseudo-verossimilhança

aproximada (MPVa) apresentaram os melhores desempenhos. Vale salientar que para $n = 40$, o método RC apresentou piores resultados para os parâmetros α_0 e β_1 , $REQM(\alpha_0) = 0.81127$ e $REQM(\beta_1) = 3.64234$, respectivamente (Tabela 2.2).

É interessante notar a dificuldade em estimar o parâmetro associado à não linearidade, α_1 , em especial quando o coeficiente de confiabilidade é pequeno, $k_x = 0.50$. Está dificuldade na estimação é verificada através dos resultados da REQM dos estimadores deste parâmetro que são muito superiores aos dos estimadores dos demais parâmetros, mesmo para os métodos MVa e MPVa. Note também que os valores da REQM considerando os métodos MVa e MPVa são melhores em relação ao *naive* apenas quando o tamanho da amostra é grande, ou seja, para $n = 160$ (Tabela 2.2).

As Figuras 2.1 e 2.2 apresentam os vieses estimados dos parâmetros associados aos submodelos da média e da dispersão, respectivamente, considerando o modelo de regressão beta não linear com erros de medida e dispersão constante para os diferentes métodos de estimação estudados. Verifica-se que os métodos da máxima verossimilhança aproximada (MV aproximada) e pseudo-verossimilhança aproximada (Pseudo MV aproximada) apresentam melhores desempenhos relativamente aos métodos *Naive* (Naive) e calibração da regressão (Calibração da regressão).

Além disso, quando se considera os métodos *Naive* e *RC* observa-se que com o aumento do erro de medida os vieses estimados se afastam de zero, ou seja, os estimadores obtidos através destes métodos são mais viesados quando comparados com os obtidos pelos métodos da MVa e MPVa. Vale salientar que o método *Naive* apresentou o pior desempenho para todos os parâmetros, exceto para o erro pequeno ($k_x = 0.95$), cujo comportamento é semelhante aos métodos que consideram erros de medida. Isto é esperado, pois para $k_x = 0.95$, tem-se que $\sigma_e^2 = 0.053$, ou seja, um erro de medida relativamente pequeno.

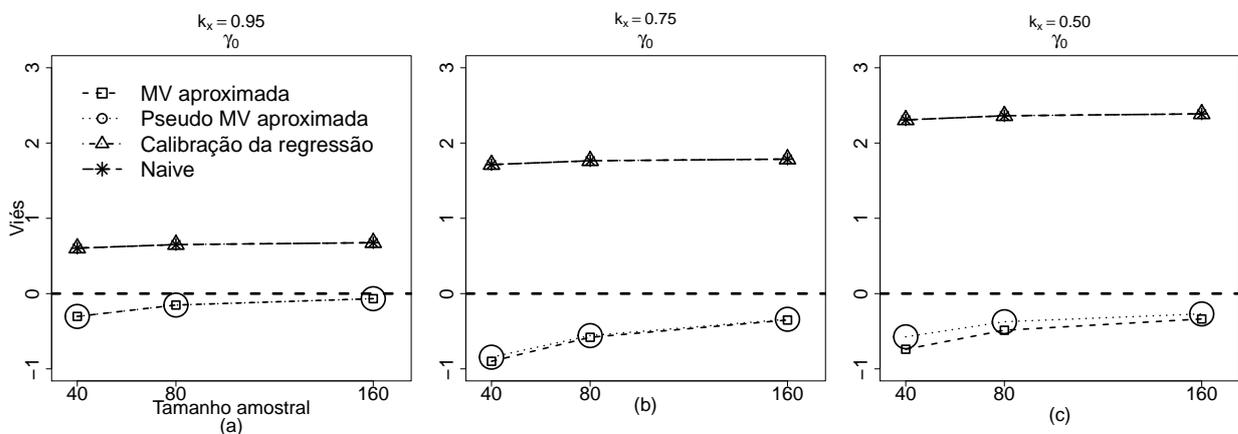


Figura 2.2: Vieses estimados do estimador de γ_0 considerando os coeficientes de confiabilidade $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Verdadeiros valores dos parâmetros: $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 1: Modelo com dispersão constante.

Com relação ao parâmetro associado a não linearidade, α_1 , os melhores e bons desempenhos estão relacionados aos métodos da máxima verossimilhança aproximada e máxima pseudo-verossimilhança aproximada. Outro fator importante está associado ao parâmetro da dispersão, γ_0 . Note que apenas nos métodos da máxima verossimilhança aproximada (MV aproximada) e máxima pseudo-verossimilhança aproximada (Pseudo MV aproximada) o viés diminui com o aumento do tamanho da amostra, como mostra a Figura 2.2 (a)-(c).

2.3.2 Cenário 2: Avaliação do parâmetro de precisão ϕ

Neste cenário é considerado o modelo de regressão beta não linear com erros de medida sob o enfoque estrutural apresentado no Cenário 1. Ou seja,

$$\begin{aligned} g(\mu_t) &= \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}, \\ h(\phi) &= \gamma_0, \\ w_t &= x_t + e_t, \end{aligned}$$

em que $t = 1, 2, \dots, n$, $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$.

Tabela 2.3: Limites inferiores médios (LI) e superiores médios (LS) e suas respectivas medianas (Md) para a média da variável resposta (μ_t) considerando o modelo de regressão beta não linear com erros de medida com dispersão constante para diferentes valores de ϕ e o coeficiente de confiabilidade $k_x = 0.50$.

ϕ	n	μ_t		
		LI	LS	Md
16.44	40	0.1362	0.8826	0.5136
	80	0.1089	0.8965	0.4971
	160	0.0877	0.9183	0.4943
54.60	40	0.1347	0.8831	0.5143
	80	0.1076	0.8971	0.4971
	160	0.0880	0.9190	0.4942
298.87	40	0.1354	0.8830	0.5137
	80	0.1081	0.8971	0.4968
	160	0.0879	0.9191	0.4942

Aqui, o objetivo é avaliar os desempenhos dos estimadores em relação à variação da precisão do modelo, medida por ϕ . Para tanto, utilizou-se valores de $\gamma_0 = (2.8, 4.0, 5.7)$ que conduziu a valores de $\phi = (16.44, 54.60, 298.87)$, respectivamente. Estes valores foram escolhidos de forma que fossem obtidos valores baixo, intermediário e alto para a precisão do modelo. Adicionalmente, o número de pontos de quadratura considerado é $Q = 50$.

As simulações foram realizadas para diferentes coeficientes de confiabilidade. Aqui, são apresentados os resultados considerando $k_x = 0.50$. Os resultados considerando os coeficientes de confiabilidade $k_x = 0.75$ e $k_x = 0.95$ encontram-se disponíveis nas Tabelas 1 e 2, respectivamente, no Apêndice 7.1, Seção 7.1.

Tabela 2.4: Raízes dos erros quadráticos médios (REQM) para os estimadores de α_0 , α_1 , β_1 e γ_0 considerando o modelo com dispersão constante com o coeficiente de confiabilidade $k_x = 0.50$. Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 2.

ϕ	n	Método	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_0$
16.44	40	<i>Naive</i>	0.13124	0.92746	0.43574	0.73316
		RC	0.28865	0.92746	2.92557	0.73316
		MVa	0.20267	1.02012	0.44365	1.83898
		MPVa	0.20790	1.01637	0.52486	1.69142
	80	<i>Naive</i>	0.10380	0.79152	0.43184	0.76609
		RC	0.12915	0.79152	0.25299	0.76609
		MVa	0.13284	0.86020	0.24983	1.22622
		MPVa	0.13170	0.85491	0.24805	1.11016
	160	<i>Naive</i>	0.06922	0.54617	0.43059	0.77962
		RC	0.08316	0.54617	0.13786	0.77962
		MVa	0.08590	0.55044	0.16759	0.77837
		MPVa	0.08544	0.54844	0.16057	0.71491
54.60	40	<i>Naive</i>	0.10932	0.78669	0.43246	1.47890
		RC	0.43258	0.78669	3.24796	1.47890
		MVa	0.17045	0.85063	0.31898	2.00728
		MPVa	0.17656	0.84264	0.42223	1.83764
	80	<i>Naive</i>	0.08585	0.67615	0.43208	1.53036
		RC	0.21884	0.67615	1.37494	1.53036
		MVa	0.11387	0.71302	0.17912	1.54887
		MPVa	0.11606	0.70669	0.23510	1.43277
	160	<i>Naive</i>	0.05756	0.48215	0.43175	1.54569
		RC	0.07381	0.48215	0.12817	1.54569
		MVa	0.07590	0.47519	0.12037	1.22764
		MPVa	0.07567	0.47136	0.11927	1.16501
298.87	40	<i>Naive</i>	0.09702	0.71864	0.43399	2.96359
		RC	0.48103	0.71864	2.15074	2.96359
		MVa	0.15793	0.78323	0.30087	1.48117
		MPVa	0.16988	0.77725	0.38529	1.38905
	80	<i>Naive</i>	0.07791	0.61805	0.43237	3.01863
		RC	0.14474	0.61805	2.01565	3.01863
		MVa	0.10753	0.66570	0.14718	1.22445
		MPVa	0.10804	0.65616	0.19992	1.17087
	160	<i>Naive</i>	0.05264	0.45004	0.43210	3.03524
		RC	0.06958	0.45004	0.12600	3.03524
		MVa	0.07181	0.44705	0.09041	0.98278
		MPVa	0.07175	0.44113	0.09368	0.95650

A Tabela 2.3 apresenta os limites inferiores e superiores médios e suas respectivas medianas para a média da variável resposta (μ_t), considerando os diferentes valores dos parâmetros de precisão ϕ . Pode-se verificar que há uma pequena diferença entre as medidas avalidas em relação ao valor de ϕ , para um tamanho amostral fixo. Por exemplo,

considerando $n = 80$, tem-se $LI = (0.1089, 0.1076, 0.1081)$ para $\phi = (16.44, 54.60, 298.87)$, respectivamente.

Além disso, os valores dos parâmetros considerados cobrem praticamente todo o espaço paramétrico de μ_t e estão bem distribuídos dentro do intervalo $(0,1)$, como mostra a mediana. Considerando por exemplo $\phi = 16.44$, tem-se que $Md = (0.5136, 0.4971, 0.4943)$ para $n = (40, 80, 160)$, respectivamente. Resultados semelhantes são obtidos considerando os coeficientes de confiabilidade $k_x = 0,75$ e $k_x = 0,95$.

A Tabela 2.4 apresenta as raízes dos erros quadráticos médios (REQM) dos estimadores para cada valor de ϕ , considerando o modelo de regressão beta não linear com erros de medida e dispersão constante com coeficiente de confiabilidade $k_x = 0.50$. Note que o valor da raiz do erro quadrático médio diminui com o aumento do tamanho amostral para todos os valores de ϕ avaliados, exceto para o parâmetro de dispersão, em que a raiz do erro quadrático médio aumenta para os métodos *Naive* e RC. Por exemplo, considerando o método *Naive*, quando $\phi = 298.87$, para $n = (40, 80, 160)$ tem-se $REQM(\gamma_0) = (2.96359; 3.01863; 3.03524)$, respectivamente.

Mais uma vez fica evidente a dificuldade de estimação do parâmetro associado à não linearidade, α_1 . Mesmo com o aumento da precisão do modelo os valores da REQM do estimador associado à não linearidade considerando todos os métodos de estimação continuam altos e substancialmente superiores aos dos estimadores dos demais parâmetros do modelo. Novamente os valores da REQM considerando os métodos MVa e MPVa são melhores em relação ao *naive* apenas para $n = 160$, exceto para $\phi = 16.44$.

Considerando o parâmetro associado à variável com erro de medida, β_1 , tem-se que os métodos da máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa) apresentaram os melhores desempenhos, exceto para $\phi = 16.44$ e $n = 40$. Vale salientar que para $n = 40$, o método da calibração da regressão (RC) apresentou piores resultados para os parâmetros α_0 e β_1 , para todos os ϕ avaliados (Tabela 2.2).

As Tabelas 3 e 4 do Apêndice 7.1 apresentam as raízes dos erros quadráticos médios (REQM) dos estimadores para cada valor de ϕ , considerando o modelo de regressão beta não linear com dispersão constante com $k_x = 0.75$ e $k_x = 0.95$, respectivamente. É possível notar que, para um valor fixo de ϕ , com o aumento do coeficiente de confiabilidade há uma diminuição na REQM, como verificado no Cenário 1.

As Figuras 2.3 e 2.4 apresentam os vieses estimados considerando o modelo de regressão beta não linear com erros de medida e dispersão constante para os diferentes métodos de estimação estudados, em relação ao parâmetro de precisão ϕ com $k_x = 0.50$. Verificou-se que os métodos da máxima verossimilhança aproximada (MV aproximada) e pseudo-verossimilhança aproximada (Pseudo MV aproximada) apresentam melhores desempenhos

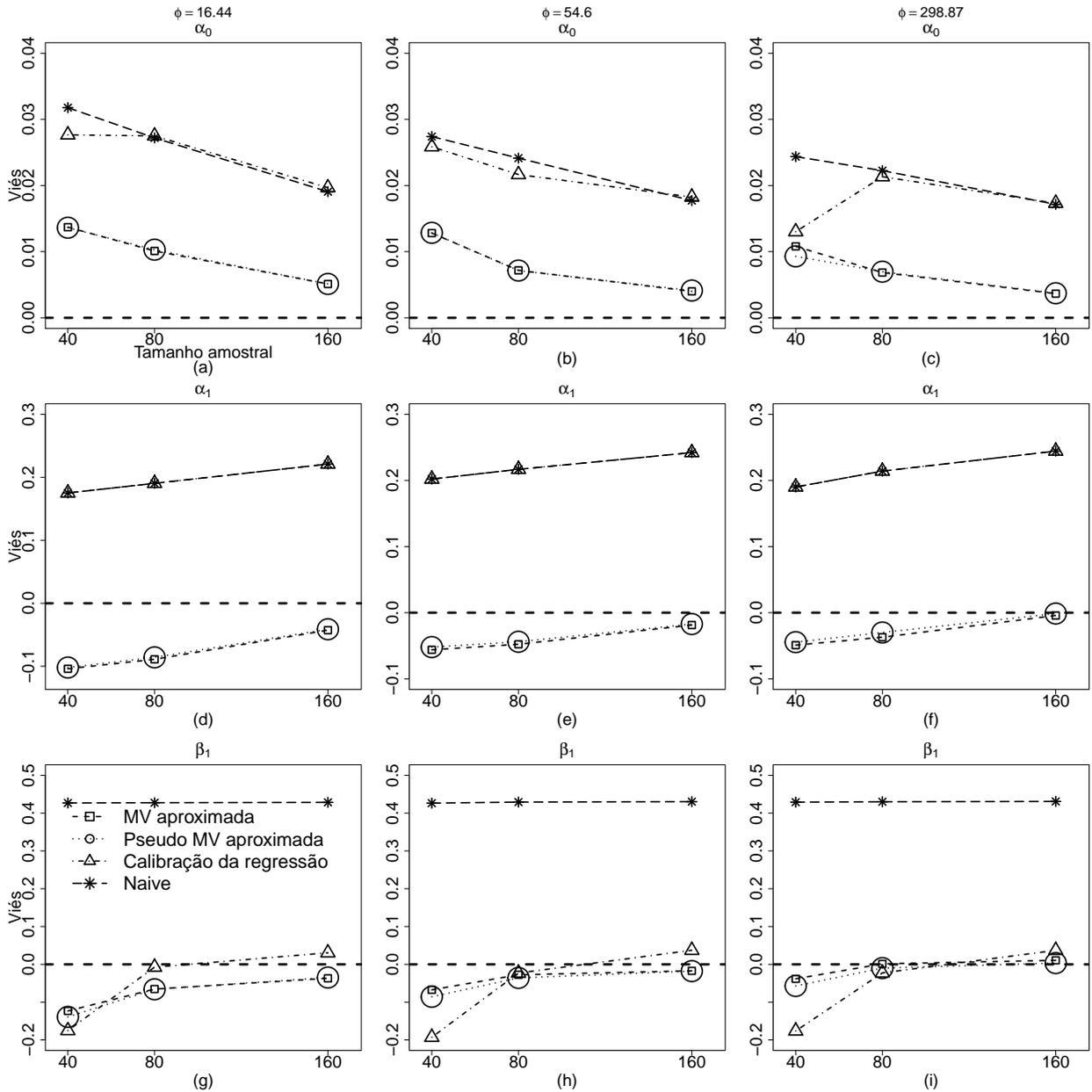


Figura 2.3: Vieses estimados dos estimadores de α_0 , α_1 e β_1 para $\phi = 16.44$, $\phi = 54.60$ e $\phi = 298.87$ com o coeficiente de confiabilidade $k_x = 0.50$ (erro de medida alto). Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 2: Modelo com dispersão constante.

quando comparados com os métodos *naive* e calibração da regressão.

Observa-se que com o aumento do tamanho amostral os vieses estimados dos estimadores se aproximam de zero para todos os valores de ϕ avaliados, considerando os métodos da máxima verossimilhança aproximada e pseudo-verossimilhança aproximada. Em relação ao parâmetro associado à variável com erro de medida (β_1), observou-se que o modelo *Naive* apresentou o pior desempenho (Figuras 2.3 (g)-(i)).

Com relação ao parâmetro α_1 , associado à não linearidade, nota-se que os valores de ϕ também interferem no desempenho de seu estimador, tal que quanto menor o valor de ϕ

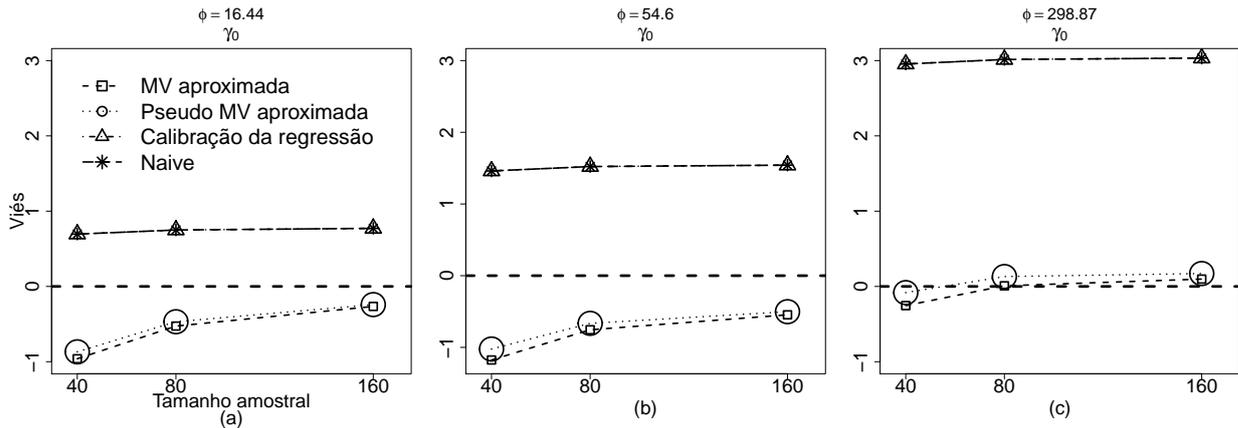


Figura 2.4: Vieses estimados dos estimadores de γ_0 para $\phi = 16.44$, $\phi = 54.60$ e $\phi = 298.87$ com o coeficiente de confiabilidade $k_x = 0.50$ (erro de medida alto). Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 2: Modelo com dispersão constante.

mais distante de zero encontra-se o viés, ou seja $\hat{\alpha}_1$ se torna mais viesado, em especial para os métodos *naive* e RC. No entanto, para os métodos da máxima verossimilhança aproximada e máxima pseudo-verossimilhança aproximada o desempenho do estimador deste parâmetro, quanto ao viés, fica próximo do desempenho dos estimadores dos demais parâmetros, ou seja, quanto maior o tamanho amostral menor o viés do estimador (Figuras 2.3 (d)-(f)).

Outro fator importante está associado ao parâmetro da dispersão, γ_0 . Note que quanto maior o valor de ϕ , menor o viés, para os métodos da máxima verossimilhança aproximada (MV aproximada) e máxima pseudo-verossimilhança aproximada (Pseudo MV aproximada). Neste caso, os métodos *naive* (Naive) e calibração da regressão (Calibração da regressão) são semelhantes e fornecem vieses muito altos, especialmente com o aumento de ϕ (Figuras 2.4 (a)-(c)).

No Apêndice 7.1 encontram-se as Figuras 1 e 2 com os vieses estimados considerando o modelo de regressão beta não linear com dispersão constante para os diferentes métodos de estimação estudados, em relação ao parâmetro de precisão ϕ para $k_x = 0.75$. Também são apresentadas as Figuras 3 e 4 com os vieses estimados considerando o modelo de regressão beta não linear com dispersão constante para os diferentes métodos de estimação estudados, em relação ao parâmetro de precisão ϕ para $k_x = 0.95$.

2.3.3 Cenário 3: Dispersão variável

Neste cenário admite-se que o modelo de regressão beta não linear com erros de medida e dispersão variável, sob o enfoque estrutural, apresenta a seguinte estrutura:

$$\begin{aligned} g(\mu_t) &= \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}, \\ h(\phi_t) &= \gamma_0 + \lambda_1 x_{t1}, \\ w_t &= x_t + e_t, \end{aligned}$$

em que $t = 1, 2, \dots, n$. Para cada réplica de Monte Carlo, geraram-se ocorrências da variável sem erro de medida z_{t1} com distribuição uniforme (0.2, 1.2). Os seguintes valores para os parâmetros que permanecem fixos para todas as simulações foram escolhidos: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 2.5$, $\lambda_1 = 0.9$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Como para cada réplica de Monte Carlo são geradas diferentes amostras aleatórias de tamanhos $n = 40, 80$ e 160 , não é possível obter um único intervalo para a média da variável resposta, bem como um único grau de heterogeneidade da dispersão. Os resultados são obtidos considerando $Q = 50$ pontos de quadratura.

Tabela 2.5: Limites inferiores médios (LI) e superiores médios (LS) e suas respectivas medianas (Md_{μ_t}) para a média da variável resposta (μ_t) e médias dos mínimos (Mín), máximos (Máx) e medianas (Md_{ϕ_t}) de ϕ_t e seus respectivos graus de heterogeneidade (δ) para o modelo de regressão beta não linear com erros de medida e dispersão variável.

k_x	n	μ_t			ϕ_t			δ
		LI	LS	Md_{μ_t}	Mín	Máx	Md_{ϕ_t}	
0.95	40	0.1499	0.8829	0.5156	2.1219	94.4678	12.5317	50.6266
	80	0.1245	0.8969	0.4989	1.7209	117.8152	12.4209	76.1024
	160	0.1042	0.9188	0.4962	1.4157	145.5675	12.3707	112.9515
0.75	40	0.1496	0.8829	0.5157	2.1106	93.2445	12.5002	50.4809
	80	0.1245	0.8970	0.4989	1.7202	117.6784	12.4217	76.0508
	160	0.1041	0.9188	0.4962	1.4161	145.6058	12.3718	112.9530
0.50	40	0.1498	0.8823	0.5160	2.1182	92.8067	12.5075	50.0786
	80	0.1245	0.8969	0.4988	1.7201	117.2303	12.4187	75.7692
	160	0.1041	0.9187	0.4962	1.4160	145.4027	12.3708	112.7931

A Tabela 2.5 apresenta os limites inferiores e superiores médios e suas respectivas medianas para a média da variável resposta (μ_t), considerando os diferentes tamanhos amostrais e as médias dos mínimos (Mín) e máximos (Máx) de ϕ_t e seus respectivos graus de heterogeneidade (δ) para o modelo de regressão beta não linear com erros de medida e dispersão variável.

Como no Cenário 1, é possível notar que os coeficientes de confiabilidade, k_x , não influenciam muito nos limites dos intervalos de valores para a média da variável resposta. O mesmo é verificado para as médias do mínimo e máximo do parâmetro de precisão. Em

Tabela 2.6: Raízes dos erros quadráticos médios (REQM) para os estimadores de α_0 , α_1 , β_1 , γ_0 e λ_1 considerando o modelo de regressão beta não linear com dispersão variável. Valores verdadeiros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 2.5$, $\lambda_1 = 0.9$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 3.

k_x	n	Método	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_0$	$\hat{\lambda}_1$
0.95	40	<i>Naive</i>	0.11328	0.76821	0.11633	0.24269	0.29035
		RC	0.11375	0.76821	0.11234	0.24333	0.29554
		MVa	0.11389	0.77171	0.11355	0.33487	0.31417
		MPVa	0.11389	0.77171	0.11355	0.33487	0.31418
	80	<i>Naive</i>	0.08580	0.63583	0.08674	0.16871	0.21624
		RC	0.08612	0.63583	0.07424	0.16918	0.20155
		MVa	0.08582	0.63778	0.07463	0.20427	0.19798
		MPVa	0.08582	0.63778	0.07463	0.20427	0.19798
	160	<i>Naive</i>	0.05816	0.41016	0.07233	0.13687	0.18577
		RC	0.05836	0.41016	0.05285	0.13749	0.15929
		MVa	0.05857	0.41226	0.05306	0.13433	0.13814
		MPVa	0.05857	0.41226	0.05306	0.13433	0.13815
0.75	40	<i>Naive</i>	0.13400	0.88915	0.25536	0.42745	0.47632
		RC	0.14303	0.88933	0.18033	0.43529	0.42326
		MVa	0.14097	0.90139	0.18363	0.71166	0.53361
		MPVa	0.14049	0.90064	0.17997	0.68113	0.50719
	80	<i>Naive</i>	0.10389	0.74483	0.23976	0.45471	0.45473
		RC	0.10947	0.74483	0.11542	0.45940	0.33535
		MVa	0.10353	0.73817	0.11736	0.38013	0.32022
		MPVa	0.10350	0.73803	0.11668	0.37455	0.31390
	160	<i>Naive</i>	0.07055	0.50008	0.23473	0.46615	0.44803
		RC	0.07497	0.50008	0.08023	0.47031	0.30520
		MVa	0.07029	0.48403	0.07970	0.21708	0.20224
		MPVa	0.07031	0.48403	0.07985	0.21743	0.20150
0.50	40	<i>Naive</i>	0.15949	1.00409	0.44757	0.69569	0.66567
		RC	1.10224	1.00409	4.01695	2.20079	7.54922
		MVa	0.21241	1.04200	0.41636	1.34326	0.99749
		MPVa	0.22590	1.03179	0.51507	1.22537	0.98922
	80	<i>Naive</i>	0.12761	0.85402	0.44108	0.74753	0.66151
		RC	0.15578	0.85402	0.27771	0.75825	0.49109
		MVa	0.13898	0.84712	0.22592	0.78858	0.55717
		MPVa	0.13913	0.84435	0.23373	0.74641	0.51627
	160	<i>Naive</i>	0.08914	0.60799	0.43886	0.76607	0.65986
		RC	0.10658	0.60799	0.15211	0.77372	0.43132
		MVa	0.09137	0.55663	0.14563	0.49363	0.34391
		MPVa	0.09152	0.55579	0.14703	0.47856	0.32879

relação ao grau de heterogeneidade, vale destacar que o mesmo se modifica com o aumento do tamanho da amostra. Além disso, os coeficientes de confiabilidade não influenciam os resultados.

A Tabela 2.6 apresenta as raízes dos erros quadráticos médios (REQM) dos estimadores considerando o modelo de regressão beta não linear com dispersão variável. Similarmente

aos resultados verificados nos outros cenários, a raiz do erro quadrático médio dos estimadores diminui com o aumento do tamanho amostral, para todos os coeficientes de confiabilidade considerados, exceto quanto ao parâmetro de dispersão, γ_0 , em que a raiz do erro quadrático médio aumenta para os métodos *Naive* e calibração da regressão (RC) quando se considera os coeficientes de confiabilidade $k_x = 0.75$ e $k_x = 0.50$. Por exemplo, considerando o método *Naive*, quando $k_x = 0.50$, temos que para $n = (40, 80, 160)$ tem-se $\text{REQM}(\gamma_0) = (0.69569, 0.74753, 0.76607)$, respectivamente.

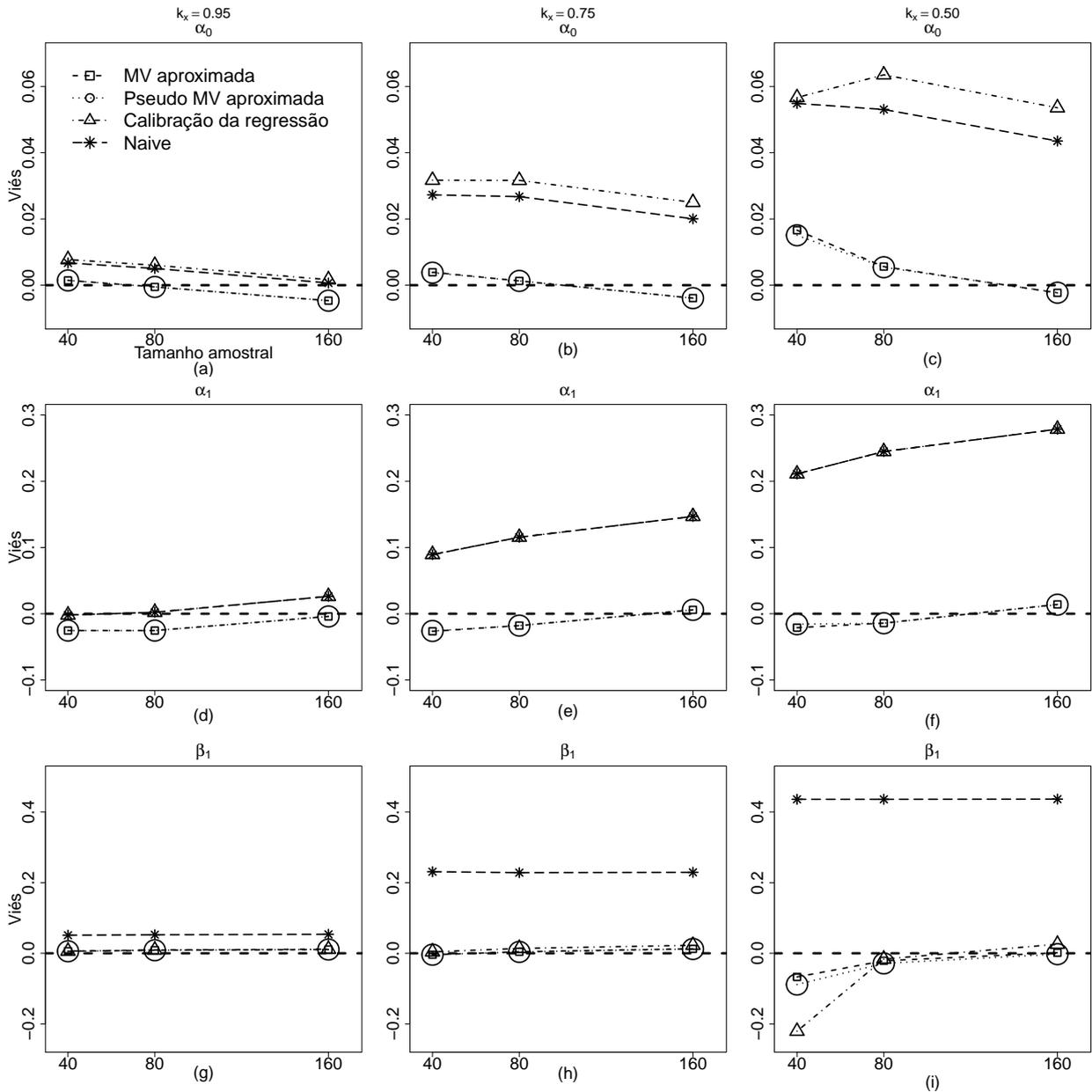


Figura 2.5: Vieses estimados dos estimadores de α_0 , α_1 e β_1 para $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Valores verdadeiros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 3: Modelo com precisão variável.

Nota-se neste cenário uma dificuldade ainda maior em estimar o parâmetro associado

à não linearidade. Os valores da REQM são consideravelmente superiores aos dos demais cenários. Considerando o parâmetro associado à variável com erro de medida, β_1 , tem-se que os métodos de máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa) apresentaram os melhores desempenhos. Para $k_x = 0.50$, o método da calibração de regressão (RC) apresenta pior desempenho para $n = 40$, por exemplo, considerando o parâmetro λ_1 , $REQM(\lambda_1) = 7.54922$.

No entanto, nota-se que a mediana do ϕ encontra-se próximo de 13 (Tabela 2.5) que é um valor consideravelmente pequeno para a precisão dos dados. Desta forma o cenário da Tabela 2.6 pode ser crítico do ponto de vista de estimação e variância dos estimadores afetando consequentemente os respectivos valores da REQM. Em especial os estimadores de α_1 , parâmetro que determina a não linearidade do modelo, são os mais afetados. No entanto, quanto ao viés, como veremos mais adiante na Figura 2.5 mesmo neste cenário crítico os estimadores de α_1 obtidos pelos métodos da máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa) mostram um bom desempenho.

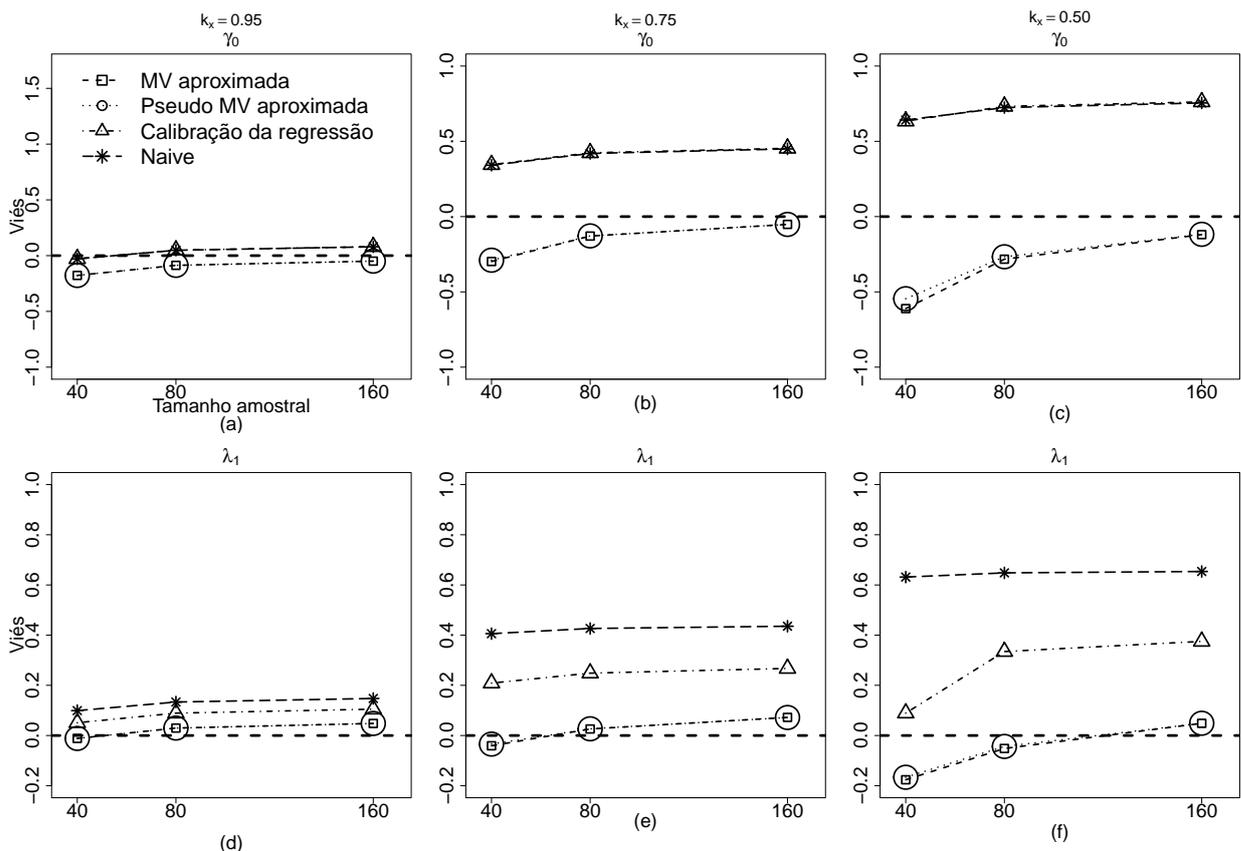


Figura 2.6: Vieses estimados dos estimadores de γ_0 e λ_1 para $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Valores verdadeiros: $\gamma_0 = 2.5$, $\lambda_1 = 0.9$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 3: Modelo com precisão variável.

As Figuras 2.5 e 2.6 apresentam os vieses estimados dos estimadores dos parâmetros associados aos modelos da média e da dispersão, respectivamente, considerando o modelo de

regressão beta não linear com erros de medida e dispersão variável. Novamente, os métodos da máxima verossimilhança aproximada (MV aproximada) e pseudo-verossimilhança aproximada (Pseudo MV aproximada) apresentam os melhores resultados do que os métodos *naive* (Naive) e calibração da regressão (Calibração da regressão).

Além disso, fixando o tamanho amostral, observou-se que com o aumento do erro de medida, ou seja, com a diminuição do coeficiente de confiabilidade os vieses estimados se afastam de zero, considerando todos os métodos de estimação, em especial os métodos *naive* e calibração da regressão. Por exemplo, para o método da máxima verossimilhança aproximada quando $n = 40$ tem-se que para $k_x = (0.95, 0.75, 0.50)$ os vieses são $(-0.0116, -0.2984, -0.6111)$ considerando o estimador de γ_0 . O método *naive* apresentou o pior desempenho quanto ao viés para todos os parâmetros, exceto quando o erro era pequeno ($k_x = 0.95$). Outro fator importante está associado ao modelo da dispersão, note que apenas os métodos da máxima verossimilhança aproximada (MV aproximada) e máxima pseudo-verossimilhança aproximada (Pseudo MV aproximada) diminuem os vieses dos estimadores de γ_0 (Figuras 2.6 (a)-(c)) e λ_1 (Figuras 2.6 (d)-(f)) com o aumento do tamanho da amostra.

2.3.4 Cenário 4: Avaliação dos pontos de quadratura “Q”

Neste cenário é considerado o modelo de regressão beta não linear com erros de medida e dispersão variável sob o enfoque estrutural apresentado no Cenário 3. Ou seja,

$$\begin{aligned} g(\mu_t) &= \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}, \\ h(\phi_t) &= \gamma_0 + \lambda_1 x_{t1}, \\ w_t &= x_t + e_t, \end{aligned}$$

em que $t = 1, 2, \dots, n$, $z_{t1} \sim U(0.2, 1.2)$, $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 2.5$, $\lambda_1 = 0.9$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Aqui, o objetivo é avaliar o desempenho dos estimadores em relação número de pontos de quadratura Q . Para tanto, consideramos $Q = (12, 30, 80)$.

As simulações foram realizadas para diferentes coeficientes de confiabilidade. Aqui são apresentados os resultados considerando $k_x = 0.50$. Os resultados para os coeficientes de confiabilidade $k_x = 0.75$ e $k_x = 0.95$ encontram-se disponíveis no Apêndice 7.1 Seção 7.1.

Como para cada réplica de Monte Carlo são geradas diferentes amostras aleatórias de tamanhos $n = 40, 80$ e 160 , não é possível obter um único intervalo para a média da variável resposta e para o grau de hereogeneidade δ .

A Tabela 2.7 apresenta os limites inferiores e superiores médios e suas respectivas medianas para a média da variável resposta (μ_t), considerando os diferentes tamanhos amostrais e as médias dos mínimos (Mín) e máximos (Máx) de ϕ_t e seus respectivos

Tabela 2.7: Limites inferiores médios (LI) e superiores médios (LS) e suas respectivas medianas (Md_{μ_t}) para a média da variável resposta (μ_t) e médias dos mínimos (Mín), máximos (Máx) e medianas (Md_{ϕ_t}) de ϕ_t e seus respectivos graus de heterogeneidade (δ) para o modelo de regressão beta não linear com erros de medida com dispersão variável. Coeficiente de confiabilidade $k_x = 0.50$.

Q	n	μ_t			ϕ_t			δ
		LI	LS	Md_{μ_t}	Mín	Máx	Md_{ϕ_t}	
12	40	0.1500	0.8826	0.5161	2.1160	94.0082	12.5286	50.7622
	80	0.1245	0.8970	0.4989	1.7206	117.8071	12.4212	76.1043
	160	0.1041	0.9188	0.4962	1.4162	145.5935	12.3717	112.9302
30	40	0.1498	0.8823	0.5158	2.1157	92.8782	12.5175	12.5175
	80	0.1245	0.8970	0.4989	1.7207	117.6527	12.4184	76.0025
	160	0.1041	0.9187	0.4962	1.4161	145.4276	12.3703	112.7615
80	40	0.1498	0.8823	0.5157	2.1171	92.7216	12.4978	50.0801
	80	0.1244	0.8968	0.4989	1.7198	117.1709	12.4223	75.7012
	160	0.1041	0.9188	0.4962	1.4156	145.4827	12.3710	112.8931

graus de heterogeneidade (δ) para o modelo de regressão beta não linear com erros de medida com dispersão variável. Como no Cenário 3, é possível notar que os coeficientes de confiabilidade não influenciam muito nos limites dos intervalos de variação de valores para a média da variável resposta. O mesmo é verificado para as médias do mínimo e máximo do parâmetro de precisão. Em relação ao grau de heterogeneidade, vale destacar que o mesmo modifica com o aumento do tamanho da amostra. Resultados semelhantes são obtidos considerando os coeficientes de confiabilidade $k_x = 0.75$ e $k_x = 0.95$ e os resultados encontram-se, respectivamente, nas Tabelas 5 e 6, no Apêndice 7.1.

A Tabela 2.8 mostra as raízes dos erros quadráticos médios (REQM) dos estimadores dos parâmetros considerando o modelo de regressão beta não linear com dispersão variável com coeficiente de confiabilidade $k_x = 0.50$ e variando o número de pontos de quadratura. Fixando o ponto de quadratura, os valores das raízes dos erros quadráticos médios (REQM) diminuem com o aumento do tamanho amostral em todos os métodos de estimação, exceto para o estimador do parâmetro de dispersão, $\hat{\gamma}_0$, em que os valores da REQM aumentam para os métodos *Naive* e RC. Por exemplo, considerando o método *Naive*, quando $Q = 30$, temos que para $n = (40, 80, 160)$ tem-se $\text{REQM}(\hat{\gamma}_0) = (0.69376, 0.74754, 0.76610)$, respectivamente.

Considerando o parâmetro associado à variável com erro de medida, β_1 , tem-se que os métodos de máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa) apresentaram os melhores resultados. Note que, para todos os valores de Q avaliados, o método da calibração de regressão (RC) apresenta pior desempenho para $n = 40$. Por exemplo, considerando $Q = 12$ e o parâmetro λ_1 , $\text{REQM}(\hat{\lambda}_1) = 7.63364$. Ou seja, para o método da calibração, o valor da REQM é praticamente sete vezes superior

Tabela 2.8: Raízes dos erros quadráticos médios (REQM) para os estimadores de α_0 , α_1 , β_1 , γ_0 e λ_1 considerando o modelo de regressão beta não linear com dispersão variável. Coeficiente de confiabilidade $k_x = 0.50$. Valores verdadeiros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 2.5$, $\lambda_1 = 0.9$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 4.

Q	n	Método	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_0$	$\hat{\lambda}_1$
12	40	<i>Naive</i>	0.15947	1.00488	0.44486	0.69187	0.66528
		RC	1.16270	1.00488	4.53059	2.20590	7.63364
		MVa	0.23545	1.07661	0.52990	1.09976	1.03705
		MPVa	0.23198	1.06026	0.54575	1.00861	1.04088
	80	<i>Naive</i>	0.12740	0.85444	0.44047	0.74710	0.66112
		RC	0.15582	0.85444	0.28305	0.75791	0.49083
		MVa	0.14122	0.86250	0.23905	0.65878	0.50336
		MPVa	0.14045	0.85685	0.23016	0.62357	0.46018
	160	<i>Naive</i>	0.08910	0.60824	0.43878	0.76590	0.65977
		RC	0.10660	0.60824	0.15210	0.77362	0.43120
		MVa	0.09189	0.56325	0.14965	0.41717	0.30804
		MPVa	0.09171	0.56184	0.14583	0.40229	0.29347
30	40	<i>Naive</i>	0.15968	1.00507	0.44638	0.69376	0.66568
		RC	1.16394	1.00507	4.54550	2.20960	7.64550
		MVa	0.21622	1.04948	0.42746	1.28354	1.05662
		MPVa	0.22854	1.04020	0.53551	1.17073	1.01945
	80	<i>Naive</i>	0.12757	0.85436	0.44065	0.74754	0.66140
		RC	0.15593	0.85436	0.28270	0.75828	0.49099
		MVa	0.13935	0.85237	0.23224	0.78553	0.57157
		MPVa	0.13941	0.84671	0.23647	0.73610	0.52188
	160	<i>Naive</i>	0.08918	0.60834	0.43880	0.76610	0.65979
		RC	0.10663	0.60834	0.15217	0.77377	0.43124
		MVa	0.09176	0.55674	0.15102	0.47894	0.34051
		MPVa	0.09187	0.55576	0.14990	0.46294	0.32330
80	40	<i>Naive</i>	0.15978	1.00418	0.44860	0.69788	0.66662
		RC	1.09549	1.00418	3.99032	2.19737	7.53546
		MVa	0.20806	1.02387	0.41641	1.32431	0.92197
		MPVa	0.22081	1.01922	0.50380	1.23298	0.93713
	80	<i>Naive</i>	0.12762	0.85435	0.44120	0.74820	0.66177
		RC	0.15532	0.85435	0.27694	0.75889	0.49129
		MVa	0.13746	0.83709	0.22059	0.78667	0.54738
		MPVa	0.13824	0.83447	0.23271	0.74262	0.51329
	160	<i>Naive</i>	0.08912	0.60840	0.43886	0.76635	0.65974
		RC	0.10658	0.60840	0.15210	0.77402	0.43106
		MVa	0.09134	0.55442	0.14290	0.48530	0.33709
		MPVa	0.09166	0.55350	0.14571	0.47228	0.32631

aos valores da REQM em comparação aos outros três métodos investigados.

Verificou-se que os pontos de quadratura não influenciam expressivamente os resultados, uma vez que os mesmos estão similares. Considere, por exemplo, o parâmetro α_1 (não linearidade), o tamanho amostral $n = 80$ e o método de máxima verossimilhança

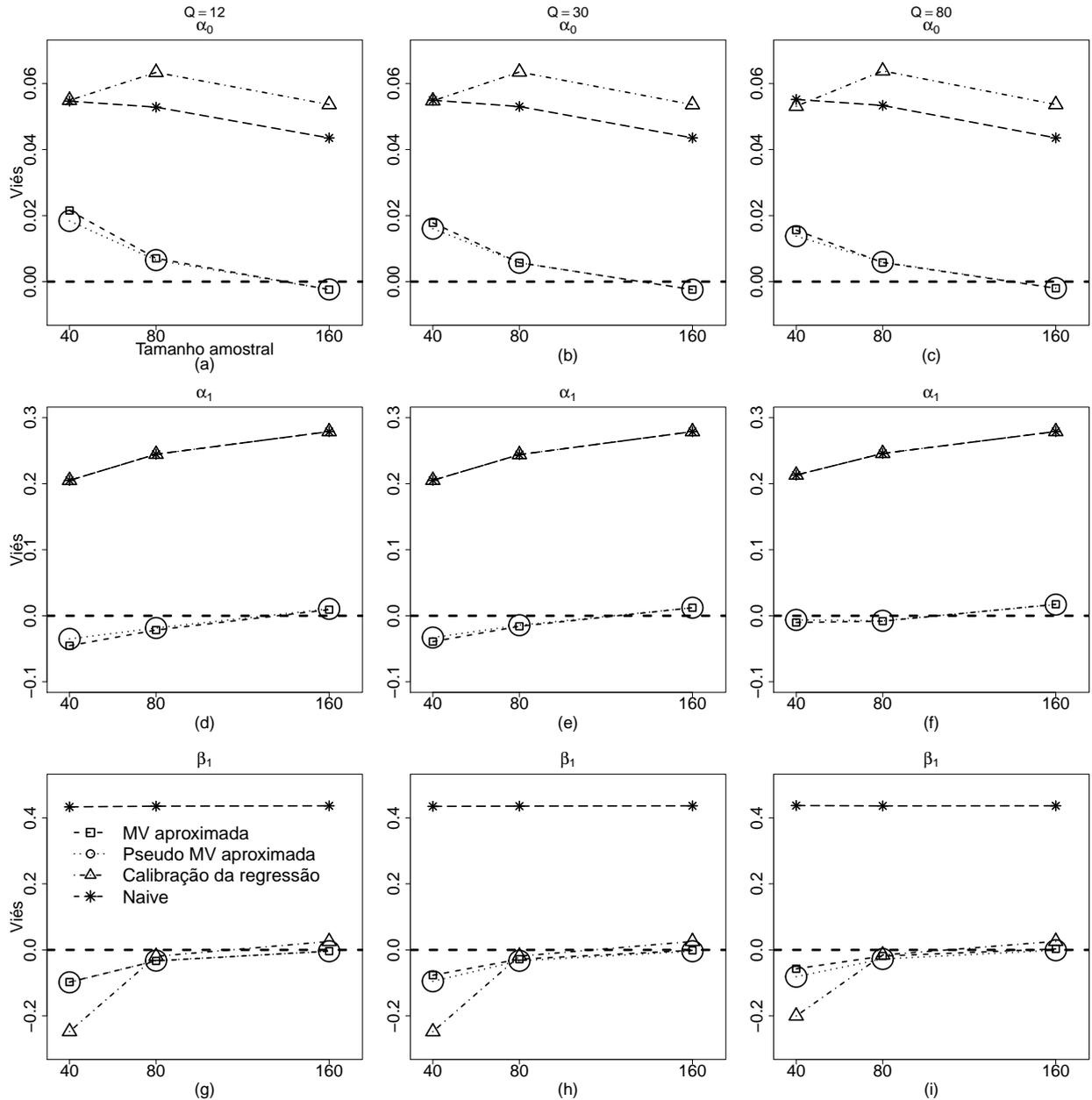


Figura 2.7: Vieses estimados dos estimadores de α_0 , α_1 e β_1 para diferentes pontos de quadratura $Q = 12$, $Q = 30$ e $Q = 80$ com o coeficiente de confiabilidade $k_x = 0,50$ (erro de medida alto). Valores verdadeiros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 4: Modelo com dispersão variável.

aproximada (MVa), então para $Q = (12, 30, 80)$ tem-se $REQM = (0.8625; 0.8524; 0.8371)$, respectivamente.

As Tabelas 7 e 8 apresentam, respectivamente, as raízes dos erros quadráticos médios dos estimadores de α_0 , α_1 , β_1 , γ_0 e λ_1 considerando o modelo de regressão beta não linear com dispersão variável para os coeficientes de confiabilidade $k_x = 0.75$ e $k_x = 0.95$.

As Figuras 2.7 e 2.8 apresentam os vieses estimados dos estimadores associados aos submodelos da média e da dispersão, respectivamente, do modelo de regressão beta não linear com erros de medida e dispersão variável, considerando diferentes pontos de

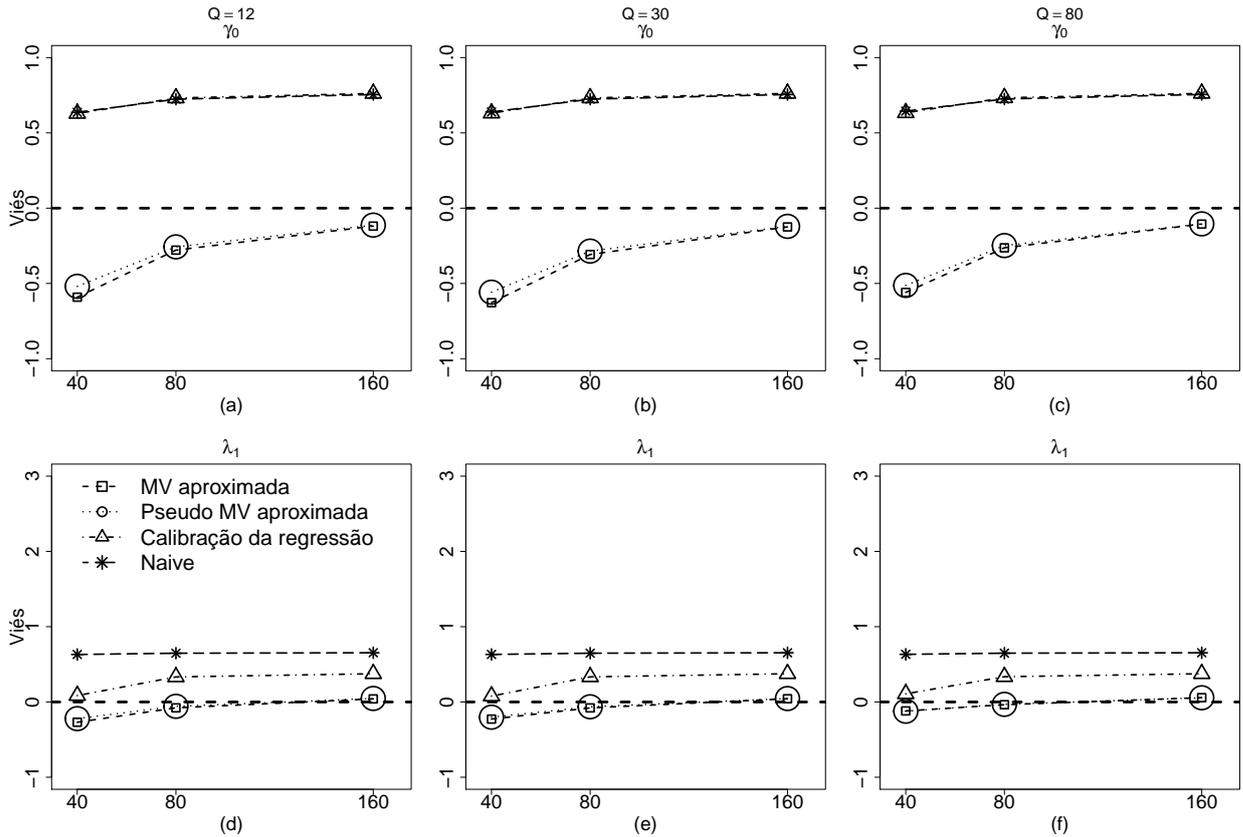


Figura 2.8: Vieses estimados dos estimadores de γ_0 e λ_1 para diferentes pontos de quadratura $Q = 12$, $Q = 30$ e $Q = 80$ com o coeficiente de confiabilidade $k_x = 0,50$ (erro de medida alto). Valores verdadeiros: $\gamma_0 = 2,5$, $\lambda_1 = 0,9$, $\mu_x = 0,0$ e $\sigma_x^2 = 1,0$. Cenário 4: Modelo com dispersão variável.

quadratura, com coeficiente de confiabilidade de $k_x = 0,50$. Novamente, os métodos da máxima verossimilhança aproximada (MV aproximada) e máxima pseudo-verossimilhança aproximada (Pseudo MV aproximada) apresentam melhores desempenhos do que os métodos *naive* (Naive) e calibração da regressão (Calibração da regressão).

Como esperado o método *naive* apresentou o pior desempenho quanto ao viés para todos os parâmetros, especialmente para o parâmetro associado à variável com erro de medida, β_1 (Figuras 2.7 (g)-(i)). Vale ressaltar também que o número de pontos de quadratura não influenciou de forma expressiva o desempenho dos estimadores com relação ao viés.

Outro fator importante está associado aos parâmetros do modelo da dispersão. Note que, para o parâmetro γ_0 , apenas os métodos da máxima verossimilhança aproximada (MV aproximada) e pseudo-verossimilhança aproximada (Pseudo MV aproximada) diminuem o viés com o aumento do tamanho da amostra (Figuras 2.8 (a)-(c)). Porém, considerando o parâmetro associado à variável com erro de medida, λ_1 , os métodos da MVa e MPVa apresentam vieses próximos de zero mesmo para tamanhos de amostras mais reduzidos (Figuras 2.8 (d)-(f)).

As Figuras 5 e 6 apresentam, respectivamente, os vieses estimados dos estimadores associados aos modelos da média e da dispersão do modelo de regressão beta não linear com erros de medida e dispersão variável, considerando diferentes pontos de quadratura para o coeficiente de confiabilidade $k_x = 0.75$ e encontram-se disponíveis no Apêndice 7.1.

As Figuras 7 e 8 apresentam, respectivamente, os vieses para os estimadores associados aos modelos da média e da dispersão do modelo de regressão beta não linear com erros de medida e dispersão variável, considerando diferentes pontos de quadratura para o coeficiente de confiabilidade $k_x = 0.95$ e encontram-se disponíveis no Apêndice 7.1.

2.3.5 Cenário 5: Não linearidade na dispersão

Neste cenário admite-se que o modelo de regressão beta não linear com erros de medida e não linearidade na variável, sob o enfoque estrutural, apresenta as seguintes componentes sistemáticas

$$\begin{aligned} g(\mu_t) &= \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}, \\ h(\phi_t) &= \gamma_0 + v_{t1}^{\gamma_1} + \lambda_1 x_{t1}, \\ w_t &= x_t + e_t, \end{aligned}$$

em que, $t = 1, 2, \dots, n$. Para cada réplica de Monte Carlo, geraram-se ocorrências das variáveis sem erro de medida z_{t1} e v_{t1} , ambas com distribuição uniforme com $(0.2, 1.2)$. Os seguintes valores para os parâmetros que permanecem fixos para todas as simulações foram escolhidos: $\alpha_0 = 0.7$, $\alpha_1 = 2.0$, $\beta_1 = -1.5$, $\gamma_0 = 1.5$, $\gamma_1 = 2.0$, $\lambda_1 = 1.3$, $\mu_x = 1.5$ e $\sigma_x^2 = 0.5$. Como para cada réplica de Monte Carlo são geradas diferentes amostras aleatórias de tamanhos $n = 60, 80$ e 160 , não é possível obter um único intervalo para a média da variável resposta. Os pontos de quadratura inicialmente estão fixados em $Q = 50$.

A Tabela 2.9 apresenta os limites inferiores e superiores médios e suas respectivas medianas para a média da variável resposta (μ_t), considerando os diferentes tamanhos amostrais e as médias dos mínimos (Mín) e máximos (Máx) de ϕ_t e seus respectivos graus de heterogeneidade (δ) para o modelo de regressão beta não linear com erros de medidas e não linearidade na dispersão. É possível notar que os coeficientes de confiabilidade não influenciam muito os limites dos intervalos de variação dos valores para a média da variável resposta. O mesmo é verificado para as médias do mínimo e máximo do parâmetro de precisão. Em relação ao grau de heterogeneidade, vale destacar que o mesmo se modifica com o aumento do tamanho da amostra. Apesar de ter médias relativamente próximas ao extremo inferior, ocorrem valores de ϕ_t altos. Ou seja, alta precisão, o que pode amenizar problemas de estimação do modelo de regressão beta típicos quando os dados encontram-se próximos aos extremos do intervalo $(0, 1)$.

Tabela 2.9: Limites inferiores médios (LI) e superiores médios (LS) e suas respectivas medianas (Md_{μ_t}) para a média da variável resposta (μ_t) e médias dos mínimos (Mín), máximos (Máx) e medianas (Md_{ϕ_t}) de ϕ_t e seus respectivos graus de heterogeneidade (δ) para o modelo de regressão beta não linear com erros de medidas e não linearidade na dispersão.

k_x	n	μ_t			ϕ_t			δ
		LI	LS	Md_{μ_t}	Mín	Máx	Md_{ϕ_t}	
0.95	60	0.0291	0.8240	0.2620	8.7692	1345.9608	95.9047	186.5761
	80	0.0255	0.8389	0.2600	7.8235	1518.0251	96.3248	232.2263
	160	0.0198	0.8773	0.2687	6.0253	1762.6010	86.2667	336.6258
0.75	60	0.0290	0.8207	0.2620	8.9305	1324.0551	95.9861	178.3026
	80	0.0255	0.8359	0.2602	7.9402	1481.7885	96.2731	223.2474
	160	0.0198	0.8757	0.2686	6.0559	1716.1938	86.2943	327.1333
0.50	60	0.0284	0.8211	0.2623	8.9631	1353.5230	96.1855	182.6783
	80	0.0251	0.8351	0.2603	7.9171	1511.6661	96.3918	229.0370
	160	0.0196	0.8758	0.2687	6.0751	1746.4286	86.2813	331.9241

A Tabela 2.10 mostra as raízes dos erros quadráticos médios (REQM) dos estimadores considerando o modelo de regressão beta não linear com erros de medidas e não linearidade na dispersão, considerando diferentes coeficientes de confiabilidade. É possível verificar que a raiz do erro quadrático médio diminui com o aumento do tamanho amostral, para todos os coeficientes de confiabilidade considerados. Pode-se notar também que o método *naive* apresenta o pior desempenho, especialmente quando o coeficiente de confiabilidade é igual a $k = 0.50$. Neste caso, os métodos de máxima verossimilhança aproximada (MVa) e pseudo-verossimilhança aproximada (MPVa) apresentam melhores resultados, quando considera-se os estimadores associados à média.

É notável que os valores da REQM dos estimadores de todos os parâmetros são expressivamente reduzidos neste cenário em que os valores de ϕ são consideravelmente superiores aos valores de ϕ no Cenário 4, mesmo considerando a estimação do parâmetro adicional ligado a não linearidade do modelo de dispersão e o fato dos valores da média estarem próximos do limite inferior do intervalo unitário.

As Figuras 2.9 e 2.10 apresentam os vieses estimados dos estimadores associados aos modelos da média e dispersão, respectivamente, do modelo de regressão beta não linear com erros de medida e não linearidade na dispersão, considerando diferentes coeficientes de confiabilidade. Os métodos da máxima verossimilhança aproximada (MV aproximada) e pseudo-verossimilhança aproximada (Pseudo MV aproximada) apresentam melhores resultados do que os métodos *naive* (Naive) e calibração da regressão (Calibração da regressão). Como esperado o método *naive* apresentou o pior desempenho, especialmente para os parâmetros associado à variável com erro de medida, β_1 (Figuras 2.9 (g)-(i)) e λ_1 (Figuras 2.10 (g)-(i)).

Outro fator importante está associado aos parâmetros do modelo da dispersão. Note

Tabela 2.10: Raízes dos erros quadráticos médios (REQM) para os estimadores de α_0 , α_1 , β_1 , γ_0 , γ_1 e λ_1 considerando o modelo de regressão beta não linear com erros de medidas e não linearidade na dispersão. Valores verdadeiros: $\alpha_0 = 0,7$, $\alpha_1 = 2,0$, $\beta_1 = -1,5$, $\gamma_0 = 1,5$, $\gamma_1 = 2,0$, $\lambda_1 = 1,3$, $\mu_x = 1,5$ e $\sigma_x^2 = 0,5$. Cenário 5.

k_x	n		α_0	α_1	β_1	γ_0	γ_1	λ_1
0.95	60	Naive	0.20428	0.40902	0.12478	0.49266	1.25042	0.44051
		RC	0.14835	0.40902	0.08259	0.53427	1.25042	0.41499
		MVa	0.15697	0.42310	0.08785	0.88640	1.31510	0.50744
		MPVa	0.15685	0.42315	0.08777	0.88563	1.31407	0.50696
	80	Naive	0.18479	0.35107	0.11605	0.41797	1.20129	0.40392
		RC	0.12505	0.35107	0.07057	0.46201	1.20129	0.37302
		MVa	0.12612	0.36464	0.07013	0.65221	1.17428	0.41134
		MPVa	0.12606	0.36461	0.07009	0.65185	1.17361	0.41115
	160	Naive	0.15678	0.24570	0.10300	0.29966	1.07865	0.36320
		RC	0.08790	0.24570	0.05058	0.34893	1.07865	0.32253
		MVa	0.08779	0.26119	0.04889	0.39655	0.95334	0.26493
		MPVa	0.08778	0.26117	0.04888	0.39655	0.95290	0.26493
0.75	60	Naive	0.68453	0.69175	0.46401	0.64930	1.65628	0.79470
		RC	0.31434	0.69175	0.19643	0.97705	1.65628	0.66835
		MVa	0.32192	0.85809	0.19386	2.12353	2.20139	1.06102
		MPVa	0.31137	0.85625	0.18554	2.05383	2.15598	1.02825
	80	Naive	0.66686	0.60702	0.45637	0.61357	1.67535	0.77804
		RC	0.26802	0.60702	0.16992	0.92863	1.67535	0.63399
		MVa	0.27579	0.81356	0.16549	1.71962	2.12388	0.89203
		MPVa	0.26732	0.81263	0.15932	1.67585	2.08388	0.86703
	160	Naive	0.64120	0.44555	0.44731	0.59714	1.62987	0.74931
		RC	0.18786	0.44555	0.12835	0.89719	1.62987	0.57940
		MVa	0.18440	0.67889	0.11076	0.94733	1.80180	0.57941
		MPVa	0.18133	0.67845	0.10853	0.92657	1.78092	0.56796
0.50	60	Naive	1.21768	0.88650	0.84122	0.80027	1.77735	1.00054
		RC	0.86322	0.88655	0.56083	1.60099	1.77727	0.84716
		MVa	0.59950	1.05943	0.37695	2.05466	2.38974	1.12708
		MPVa	0.57699	1.05200	0.36270	1.93120	2.30627	1.07053
	80	Naive	1.20496	0.79073	0.83553	0.79238	1.81518	0.98872
		RC	0.67878	0.79073	0.44125	1.49155	1.81518	0.76113
		MVa	0.49298	0.98860	0.30881	1.75572	2.37612	0.98145
		MPVa	0.48611	0.98378	0.30322	1.66581	2.28477	0.93848
	160	Naive	1.18223	0.61159	0.83129	0.80253	1.77508	0.97930
		RC	0.37259	0.61159	0.25619	1.37536	1.77508	0.68587
		MVa	0.31677	0.80122	0.20083	1.18807	2.02659	0.68066
		MPVa	0.30853	0.80046	0.19538	1.14019	1.96493	0.65577

que, para os parâmetros γ_0 (Figura 2.10 (a)-(c)), γ_1 (Figura 2.10 (d)-(f)) e λ_1 ((Figura 2.10 (g)-(i)), apenas os métodos da máxima verossimilhança aproximada (MV aproximada) e pseudo-verossimilhança (Pseudo MV aproximada) diminuem o viés com o aumento do tamanho da amostra.

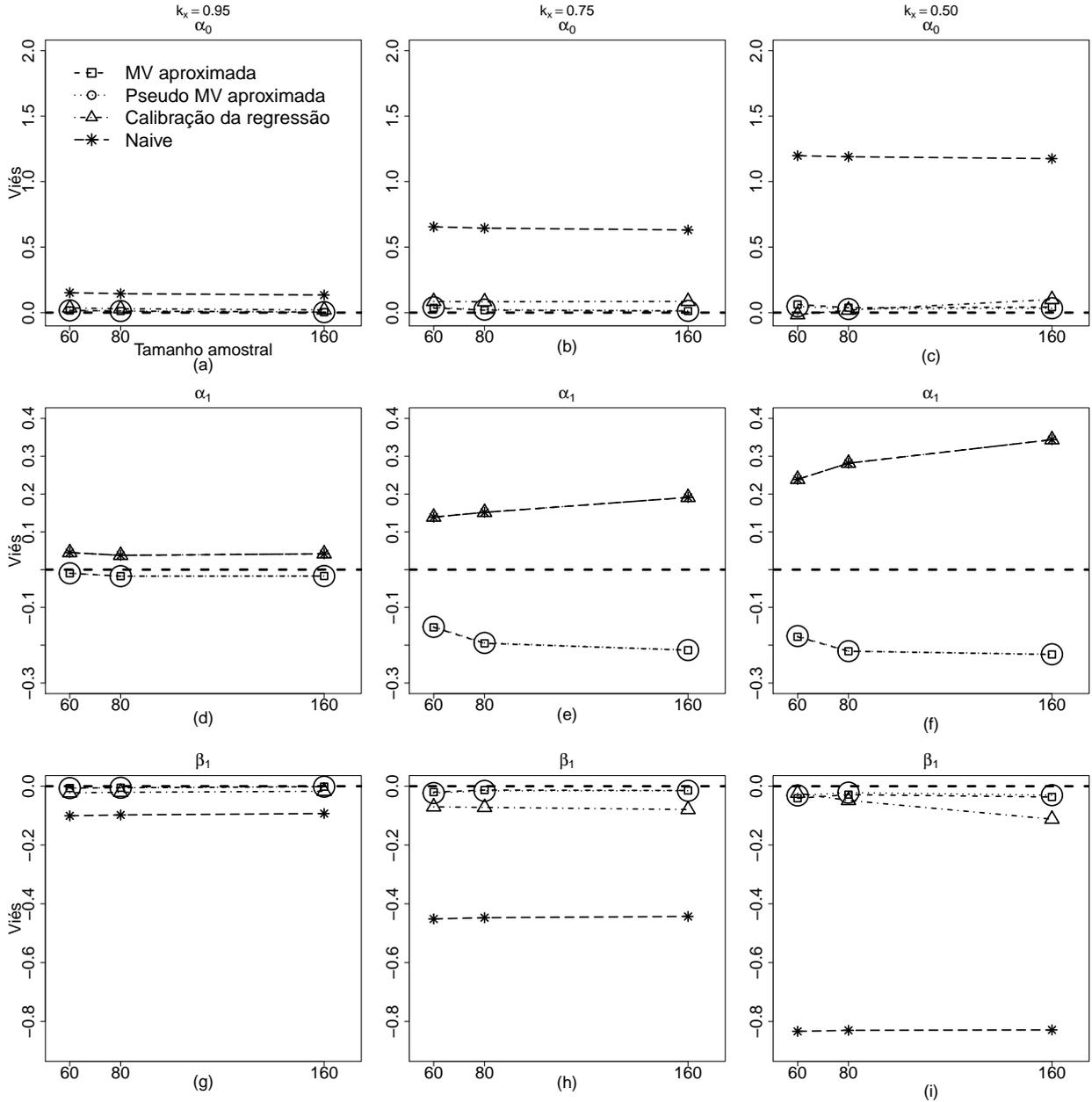


Figura 2.9: Vieses estimados dos estimadores de α_0 , α_1 e β_1 para os coeficientes de confiabilidade $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Valores verdadeiros: $\alpha_0 = 0.7$, $\alpha_1 = 2.0$, $\beta_1 = -1.5$, $\mu_x = 1.5$ e $\sigma_x^2 = 0.5$. Cenário 5: Modelo com não linearidade na dispersão.

Tanto os valores da REQM quanto o comportamento do vieses estimados dos estimadores apontam para a grande dificuldade em estimar o modelo conjuntamente não linear para a média e para a dispersão. A maior dificuldade encontra-se na estimação dos parâmetros associados a não linearidade, neste caso α_1 (Figura 2.9 (d)-(f)) e γ_1 (Figura 2.10 (d)-(f)). Mas os estimadores do parâmetro associado ao erro de medida no modelo da dispersão também apresentam um pobre desempenho tanto quanto a viés quanto a REQM.

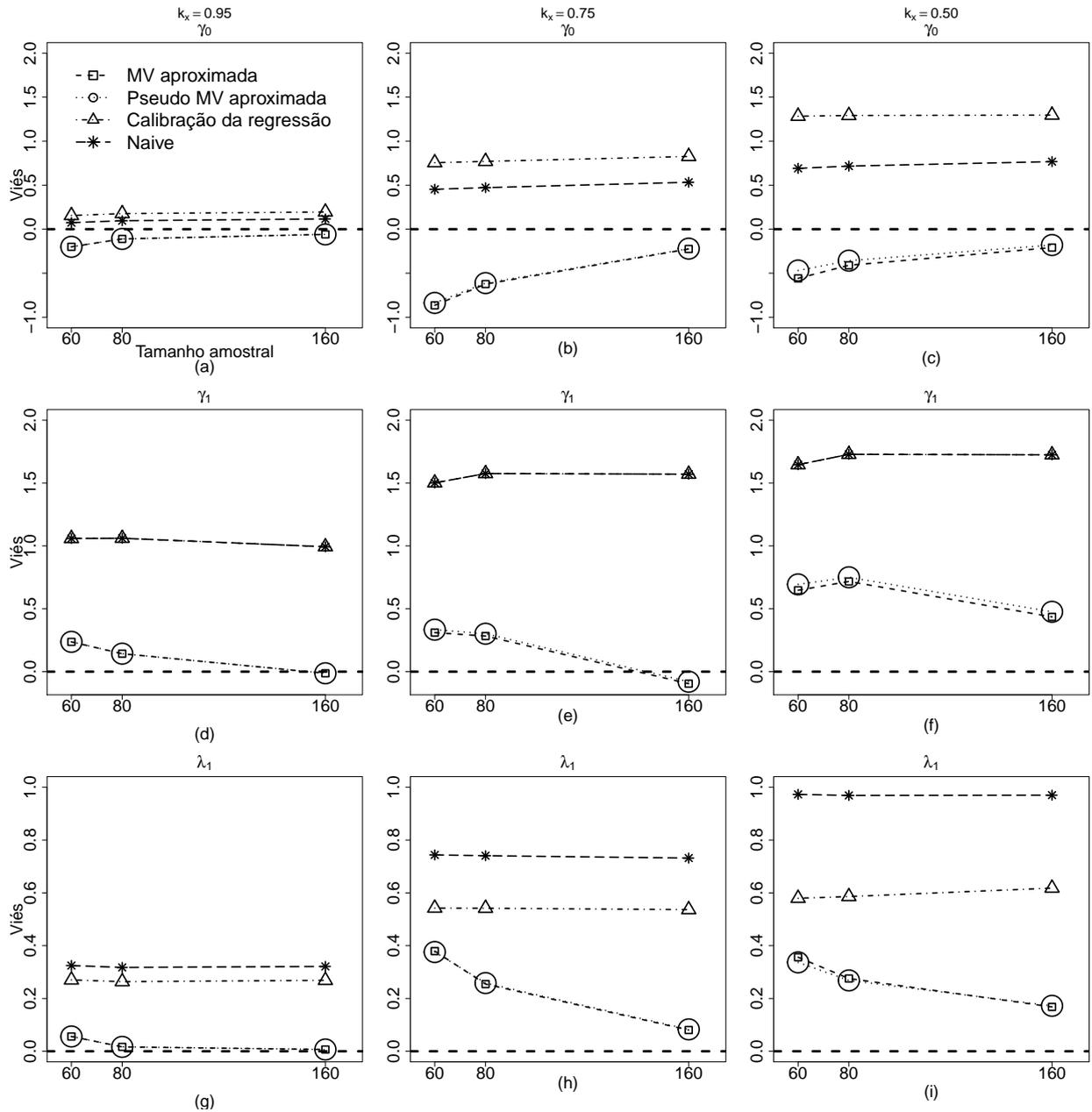


Figura 2.10: Vieses estimados dos estimadores de γ_0 , γ_1 e λ_1 para os coeficientes de confiabilidade $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Valores verdadeiros: $\gamma_0 = 1.5$, $\gamma_1 = 2.0$, $\lambda_1 = 1.3$, $\mu_x = 1.5$ e $\sigma_x^2 = 0.5$. Cenário 5: Modelo com não linearidade na dispersão.

2.4 Conclusões

Neste capítulo propomos o modelo de regressão beta não linear com erros de medidas. Utilizamos três métodos de estimação: máxima verossimilhança aproximada (MVa), máxima pseudo-verossimilhança aproximada (MPVa) e calibração da regressão (RC). Também avaliamos o método *naive*, em que ajusta-se o modelo considerando erroneamente que a variável não apresenta erro de medida.

Através das simulações de Monte Carlo avaliamos o desempenho dos estimadores

estudados através da média, do viés e da raiz do erro quadrático médio. Concluímos que, de modo geral, os métodos da MVa e MPVa apresentaram os melhores desempenhos. Além disso, verificamos que o desempenho dos estimadores associados à não linearidade e dispersão variável são afetados negativamente quando o coeficiente de confiabilidade diminui.

3 DISTRIBUIÇÕES ASSINTÓTICAS

As distribuições assintóticas dos estimadores obtidos pelos métodos estudados no Capítulo 2 serão discutidas aqui. As propriedades assintóticas dos estimadores são fundamentais para a construção dos intervalos de confiança e testes de hipóteses. Neste capítulo serão apresentados intervalos de confiança associados aos diversos métodos de estimação e os mesmos são avaliados por estudos de simulação.

3.1 Distribuições assintóticas dos estimadores

3.1.1 Máxima verossimilhança aproximada

Um dos métodos comumente utilizados na literatura na estimação paramétrica é o método de máxima verossimilhança, cujos estimadores obtidos sob certas condições apresentam tipicamente boas propriedades como consistência, eficiência e normalidade assintótica (Cramér, 1999). Considere o vetor de parâmetros desconhecidos $\Psi = (\boldsymbol{\theta}^\top, \boldsymbol{\xi}^\top, \boldsymbol{\sigma}_e^{2\top})^\top$, em que $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\lambda}^\top)^\top$ denota o vetor de parâmetros de interesse e $(\boldsymbol{\xi}^\top, \boldsymbol{\sigma}_e^{2\top})^\top$ o vetor de parâmetros de perturbação, sendo $\boldsymbol{\xi}^\top = (\mu_x^\top, \sigma_x^{2\top})^\top$. Para grandes amostras e sob as condições usuais de regularidade, a distribuição aproximada do estimador de máxima verossimilhança do vetor de parâmetros Ψ , denotado por $\hat{\Psi}$, é normal multivariada de média Ψ e matriz de variâncias e covariâncias $I(\hat{\Psi})^{-1}$ em que $I(\Psi) = \mathbb{E}[-\partial^2 \ell(\Psi) / \partial \Psi \partial \Psi^\top]$ é a matriz de informação esperada.

O conceito de informação está associado à curvatura média do logaritmo da função de verossimilhança, no sentido de que quanto maior a curvatura mais precisa é a informação contida com base em diversas amostras, ou equivalentemente. Podemos considerar $I(\Psi)$ como uma medida de informação global, mas em muitas situações não é possível encontrar os valores esperados ou a informação baseada em várias amostras, por exemplo, em modelos associados a tempos de vida na presença de censura. Assim, recomenda-se o uso da matriz de informação observada, denotada por $J(\Psi) = -\partial^2 \ell(\Psi) / \partial \Psi \partial \Psi^\top$. Em modelos não lineares com erros de medida, utilizar a matriz de informação observada no lugar da matriz de informação esperada se torna útil (Buonaccorsi e Tosteson, 1993).

Os modelos mistos, modelos com efeitos aleatórios e modelos com erros de medida, possuem o logaritmo da função de verossimilhança que depende de integrais difíceis de serem resolvidas. Para solucionar esta problemática técnicas numéricas de integração são tipicamente utilizadas. Aqui, fizemos uso do método de quadratura de Gauss-Hermite.

A partir do logaritmo da função de verossimilhança aproximada é obtido o vetor

escore aproximado, $U_a = \partial \ell_a(\Psi)/\partial(\Psi)$, e a matriz de informação observada aproximada, $J_a(\Psi) = -\partial^2 \ell_a(\Psi)/\partial(\Psi)\partial(\Psi)^\top$, em que $\Psi = (\alpha^\top, \beta^\top, \gamma^\top, \lambda^\top, \mu_x, \sigma_x^2)^\top$. O vetor escore aproximado para o modelo de regressão beta com erros de medida foi apresentado no Capítulo 2. A matriz de informação observada aproximada definida por

$$\mathbf{J}_a(\Psi) = \begin{bmatrix} \frac{\partial^2 \ell_a(\Psi)}{\partial \alpha \partial \alpha^\top} & \frac{\partial^2 \ell_a(\Psi)}{\partial \alpha d\beta} & \frac{\partial^2 \ell_a(\Psi)}{\partial \alpha \partial \gamma^\top} & \frac{\partial^2 \ell_a(\Psi)}{\partial \alpha d\lambda} & \frac{\partial^2 \ell_a(\Psi)}{\partial \alpha d\mu_x} & \frac{\partial^2 \ell_a(\Psi)}{\partial \alpha d\sigma_x^2} \\ & \frac{\partial^2 \ell_a(\Psi)}{d\beta d\beta} & \frac{\partial^2 \ell_a(\Psi)}{d\beta \partial \gamma^\top} & \frac{\partial^2 \ell_a(\Psi)}{d\beta d\lambda} & \frac{\partial^2 \ell_a(\Psi)}{d\beta d\mu_x} & \frac{\partial^2 \ell_a(\Psi)}{d\beta d\sigma_x^2} \\ & & \frac{\partial^2 \ell_a(\Psi)}{\partial \gamma \partial \gamma^\top} & \frac{\partial^2 \ell_a(\Psi)}{\partial \gamma d\lambda} & \frac{\partial^2 \ell_a(\Psi)}{\partial \gamma d\mu_x} & \frac{\partial^2 \ell_a(\Psi)}{\partial \gamma d\sigma_x^2} \\ & & & \frac{\partial^2 \ell_a(\Psi)}{d\lambda d\lambda} & \frac{\partial^2 \ell_a(\Psi)}{d\lambda d\mu_x} & \frac{\partial^2 \ell_a(\Psi)}{\partial \lambda \partial \sigma_x^2} \\ & & & & \frac{\partial^2 \ell_a(\Psi)}{d\mu_x d\mu_x} & \frac{\partial^2 \ell_a(\Psi)}{d\mu_x d\sigma_x^2} \\ & & & & & \frac{\partial^2 \ell_a(\Psi)}{d\sigma_x^2 d\sigma_x^2} \end{bmatrix}_{n(\Psi) \times n(\Psi)}, \quad (3.1)$$

em que $n(\Psi) = (p + 1) + (\check{q} + 1) + 2 < n$, encontra-se detalhada no Apêndice 7.1.

A distribuição assintótica do estimador MVA, ou seja, considerando o método de máxima verossimilhança aproximada, é normal com média Ψ e matriz de variâncias e covariâncias $J_a^{-1}(\Psi)$, ou seja, para n e Q suficientemente grande,

$$\hat{\Psi} \sim N(\Psi, J_a^{-1}(\Psi)), \quad (3.2)$$

aproximadamente (Guolo, 2011).

Considere a seguinte matriz de informação observada contendo apenas os parâmetros de interesse:

$$\mathbf{J}_a(\alpha, \beta, \gamma, \lambda) = \begin{bmatrix} \mathbf{J}_a(\alpha\alpha) & \mathbf{J}_a(\alpha\beta) & \mathbf{J}_a(\alpha\gamma) & \mathbf{J}_a(\alpha\lambda) \\ & \mathbf{J}_a(\beta\beta) & \mathbf{J}_a(\beta\gamma) & \mathbf{J}_a(\beta\lambda) \\ & & \mathbf{J}_a(\gamma\gamma) & \mathbf{J}_a(\gamma\lambda) \\ & & & \mathbf{J}_a(\lambda\lambda) \end{bmatrix}_{n(\alpha, \beta, \gamma, \lambda) \times n(\alpha, \beta, \gamma, \lambda)}, \quad (3.3)$$

em que $n(\Psi) = (p + 1) + (\check{q} + 1) < n$ e os elementos são detalhados no Apêndice 7.1. Com base na Equação (3.2) e utilizando o método delta podemos construir intervalos de confiança assintóticos para os parâmetros do modelo definido em (2.3) e (2.4) com nível de 95% de confiança dado por

$$(g^{-1}(\hat{\eta}_{1t} - 1.96\text{e.p.}(\hat{\eta}_{1t})), g^{-1}(\hat{\eta}_{1t} + 1.96\text{e.p.}(\hat{\eta}_{1t})))$$

e

$$(g^{-1}(\hat{\eta}_{2t} - 1.96\text{e.p.}(\hat{\eta}_{2t})), g^{-1}(\hat{\eta}_{2t} + 1.96\text{e.p.}(\hat{\eta}_{2t}))),$$

em que $\hat{\eta}_{1t} = f_1(\mathbf{z}_t^\top, x_t^*; \hat{\boldsymbol{\alpha}}, \hat{\beta})$ e $\hat{\eta}_{2t} = f_2(\mathbf{v}_t^\top, x_t^*; \hat{\boldsymbol{\gamma}}, \hat{\lambda})$.

Considerando, por exemplo β , parâmetro associado ao modelo da média, temos que o erro-padrão é dado por e.p. $(\hat{\eta}_{1t(\beta)}) = \sqrt{\left(\frac{\partial \eta_{1t}}{\partial \beta} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}^\top \mathbf{J}_{a(\beta\beta)}^* \frac{\partial \eta_{1t}}{\partial \beta} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right)}$, em que $\mathbf{J}_{a(\beta\beta)}^*$ é a inversa da matriz de informação observada $\mathbf{J}_{a(\beta\beta)}$.

3.1.2 Máxima pseudo-verossimilhança aproximada

A máxima pseudo-verossimilhança é uma técnica comumente utilizada para a estimação dos parâmetros quando o modelo avaliado tem parâmetros de interesse e perturbação. A teoria de equações de estimação é usada para determinar a distribuição assintótica dos estimadores de máxima pseudo-verossimilhança aproximada (ver Carroll *et al.*, 2006 e Guolo, 2011). Carroll e Ruppert (1988) utilizaram a teoria de equações de estimação para analisar vários métodos de transformação e ponderação em modelos de regressão. Outra boa referência é Godambe (1991). Anteriormente denotamos os vetores de parâmetros de interesse e perturbação por $\boldsymbol{\theta}$ e $(\boldsymbol{\xi}^\top, \boldsymbol{\sigma}_e^2)^\top$, respectivamente. Como a variância do erro $\boldsymbol{\sigma}_e^2$ é conhecida, o parâmetro $\boldsymbol{\sigma}_e^2$ é omitido da notação. Considere (y_t, \mathbf{w}_t) , para $t = 1, \dots, n$, os n pares de variáveis aleatórias observadas, em que a cada t -ésima unidade amostral associa-se uma função de estimação ψ_t . A função de estimação para uma amostra é dada pela seguinte expressão:

$$\boldsymbol{\Psi}_n(\mathbf{y}, \mathbf{w}; \boldsymbol{\theta}, \boldsymbol{\xi}) = \sum_{t=1}^n \psi_t(y_t, w_t; \boldsymbol{\theta}, \boldsymbol{\xi}), \quad (3.4)$$

em que $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$. Assim, a Equação (3.4) é restringida de tal forma que $\boldsymbol{\Psi}_n(\mathbf{y}, \mathbf{w}; \boldsymbol{\theta}, \boldsymbol{\xi}) = 0$ denota a equação de estimação.

Como discutido anteriormente, a estimação por máxima pseudo-verossimilhança estima o vetor de parâmetros de perturbação, $\boldsymbol{\xi}$, a partir de uma equação de estimação que é definida por

$$\sum_{t=1}^n \psi_{1t}(w_t; \boldsymbol{\xi}) = 0. \quad (3.5)$$

Note que a função de estimação (3.5) depende apenas de w_t e $\boldsymbol{\xi}$. Após a estimação de $\boldsymbol{\xi}$, estima-se o vetor de parâmetros de interesse a partir de outra equação de estimação dada por

$$\sum_{t=1}^n \psi_{2t}(y_t, w_t; \boldsymbol{\theta}, \hat{\boldsymbol{\xi}}) = 0, \quad (3.6)$$

em que $\hat{\boldsymbol{\xi}}$ é fixo e obtido em (3.5). Como o par $(\boldsymbol{\theta}, \hat{\boldsymbol{\xi}})$ é solução simultânea de (3.5) e (3.6), a distribuição assintótica de $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\xi}})$ é determinada a partir de uma equação de estimação

única que é dada por

$$n^{-1} \sum_{t=1}^n \begin{pmatrix} \psi_{1t}(w_t; \boldsymbol{\xi}) \\ \psi_{2t}(y_t, w_t; \boldsymbol{\theta}, \hat{\boldsymbol{\xi}}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix};$$

para mais detalhes, ver Carroll *et al.* (2006). O objetivo aqui é obter a distribuição assintótica do estimador do vetor de parâmetros de interesse $\boldsymbol{\theta}$. Sob as condições de regularidade dadas em Gong e Samaniego (1981), a distribuição assintótica de $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ é normal com média 0 e matriz de variâncias e covariâncias $\boldsymbol{\Sigma}$ (Carroll *et al.*, 2006; Guolo, 2011), em que

$$\begin{aligned} \boldsymbol{\Sigma} &= I_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1} I_p I_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}, \\ I_p &= \boldsymbol{\Sigma}_{\boldsymbol{\theta}\boldsymbol{\theta}} - I_{\boldsymbol{\theta}\boldsymbol{\xi}} I_{\boldsymbol{\xi}\boldsymbol{\xi}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\theta}} - \boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\theta}}^\top I_{\boldsymbol{\xi}\boldsymbol{\xi}}^{-1} I_{\boldsymbol{\xi}\boldsymbol{\theta}}^\top + I_{\boldsymbol{\theta}\boldsymbol{\xi}} I_{\boldsymbol{\xi}\boldsymbol{\xi}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\xi}} I_{\boldsymbol{\xi}\boldsymbol{\xi}}^{-1} I_{\boldsymbol{\theta}\boldsymbol{\xi}}^\top. \end{aligned} \quad (3.7)$$

Os componentes da matriz $\boldsymbol{\Sigma}$ são dados por

$$\begin{aligned} I_{\boldsymbol{\theta}\boldsymbol{\theta}} &= - \sum_{t=1}^n \mathbb{E} \left\{ \frac{\partial l_{pt}(\boldsymbol{\theta}, \boldsymbol{\xi})}{\partial \boldsymbol{\theta} \boldsymbol{\theta}^\top} \right\} \Big|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}}, \\ I_{\boldsymbol{\xi}\boldsymbol{\xi}} &= - \sum_{t=1}^n \mathbb{E} \left\{ \frac{\partial l_{rt}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi} \boldsymbol{\xi}^\top} \right\}, \\ I_{\boldsymbol{\theta}\boldsymbol{\xi}} &= - \sum_{t=1}^n \mathbb{E} \left\{ \frac{\partial l_{pt}(\boldsymbol{\theta}, \boldsymbol{\xi})}{\partial \boldsymbol{\theta} \boldsymbol{\xi}^\top} \right\} \Big|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}}, \\ \boldsymbol{\Sigma}_{\boldsymbol{\theta}\boldsymbol{\theta}} &= \sum_{t=1}^n \frac{\partial l_{pt}(\boldsymbol{\theta}, \boldsymbol{\xi})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}} \left\{ \frac{\partial l_{pt}(\boldsymbol{\theta}, \boldsymbol{\xi})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}} \right\}^\top, \\ \boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\theta}} &= \sum_{t=1}^n \frac{\partial l_{pt}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} \left\{ \frac{\partial l_{pt}(\boldsymbol{\theta}, \boldsymbol{\xi})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}} \right\}^\top, \\ \boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\xi}} &= \sum_{t=1}^n \frac{\partial l_{rt}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} \left\{ \frac{\partial l_{pt}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} \right\}^\top, \end{aligned}$$

em que $l_{rt}(\boldsymbol{\xi})$ e $l_{pt}(\boldsymbol{\theta}, \boldsymbol{\xi})$ são as t -ésimas parcelas do logaritmo da função de verossimilhança de $l_r(\boldsymbol{\xi})$ e $l_p(\boldsymbol{\theta}, \boldsymbol{\xi})$, respectivamente (Guolo, 2011), $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\lambda}^\top)^\top$ denota o vetor de parâmetros de interesse e $\boldsymbol{\xi} = (\mu_x, \sigma_x^2)^\top$ denota o vetor de parâmetros de perturbação. Como no modelo proposto por Carrasco *et al.* (2014), para o modelo de regressão beta não linear com erros de medida, as matrizes $I_{\boldsymbol{\theta}\boldsymbol{\theta}}$, $I_{\boldsymbol{\xi}\boldsymbol{\xi}}$ e $I_{\boldsymbol{\theta}\boldsymbol{\xi}}$ não possuem formas explícitas. Uma vez que a integral é aproximada numericamente, é utilizado simplesmente as segundas derivadas de $l_{rt}(\boldsymbol{\xi})$ e $l_{pt}(\boldsymbol{\theta}, \boldsymbol{\xi})$, desta forma a matriz de informação esperada é substituída pela matriz de informação observada (Buonaccorsi e Tosteson, 1993).

Os intervalos de confiança assintóticos de nível 95% de confiança para η_{1t} e η_{2t} são dados, respectivamente, por

$$(g^{-1}(\hat{\eta}_{1t} - 1.96\text{e.p.}(\hat{\eta}_{1t})), g^{-1}(\hat{\eta}_{1t} + 1.96\text{e.p.}(\hat{\eta}_{1t})))$$

e

$$(g^{-1}(\hat{\eta}_{2t} - 1.96\text{e.p.}(\hat{\eta}_{2t})), g^{-1}(\hat{\eta}_{2t} + 1.96\text{e.p.}(\hat{\eta}_{2t}))),$$

em que $\hat{\eta}_{1t} = f_1(\mathbf{z}_t^\top, x_t^*; \hat{\boldsymbol{\alpha}}, \hat{\beta})$ e $\hat{\eta}_{2t} = f_2(\mathbf{v}_t^\top, x_t^*; \hat{\boldsymbol{\gamma}}, \hat{\lambda})$. Considerando, por exemplo, o parâmetro associado ao modelo da média, β , temos que o erro-padrão é dado por $\text{e.p.}(\hat{\eta}_{1t(\beta)}) = \sqrt{\left(\frac{\partial \eta_{1t}}{\partial \beta} \Big|_{\beta=\hat{\beta}}^\top \hat{\Sigma}_{\beta\beta} \frac{\partial \eta_{1t}}{\partial \beta} \Big|_{\beta=\hat{\beta}} \right)}$.

3.1.3 Calibração da regressão

O método de calibração da regressão é bastante usado em modelos com erros de medida por sua simplicidade. A ideia principal deste método é substituir a variável não observada x_t por alguma estimativa da esperança condicional de x_t dado w_t , $E(x_t|w_t)$.

Os intervalos de confiança assintóticos de nível 95% de confiança para η_{1t} e η_{2t} são dados, respectivamente, por

$$(g^{-1}(\hat{\eta}_{1t} - 1.96\text{e.p.}(\hat{\eta}_{1t})), g^{-1}(\hat{\eta}_{1t} + 1.96\text{e.p.}(\hat{\eta}_{1t})))$$

e

$$(g^{-1}(\hat{\eta}_{2t} - 1.96\text{e.p.}(\hat{\eta}_{2t})), g^{-1}(\hat{\eta}_{2t} + 1.96\text{e.p.}(\hat{\eta}_{2t}))),$$

em que $\hat{\eta}_{1t} = f_1(\mathbf{z}_t^\top, E(x|w_t); \hat{\boldsymbol{\alpha}}, \hat{\beta})$ e $\hat{\eta}_{2t} = f_2(\mathbf{v}_t^\top, E(x|w_t); \hat{\boldsymbol{\gamma}}, \hat{\lambda})$.

As estimativas dos erros-padrão dos estimadores que são usados para construção dos intervalos de confiança podem ser aproximadas utilizando o método de *bootstrap* não paramétrico (Carroll *et al.*, 2006). Para determinar os erros-padrão dos estimadores, utilizou-se o método *bootstrap* que segue o seguinte algoritmo:

1. Selecionou-se B amostras independentes, $\mathbf{x}^{*1}, \dots, \mathbf{x}^{*B}$, cada uma consistindo de n valores selecionados com reposição de \mathbf{x} . Tomou-se $B = 30$.
2. Calculou-se a réplica *bootstrap* para cada amostra *bootstrap*:

$$\hat{\theta}^*(b) = s(\mathbf{x}^{*b}), \quad b = 1, \dots, B.$$

3. Estimou-se o erro-padrão $\text{e.p.}_{F(\hat{\theta})}$ pelo desvio padrão amostral das B réplicas:

$$\widehat{\text{e.p.}}_B = \left[\frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}^*(b) - \hat{\theta}^*(\cdot))^2 \right]^{\frac{1}{2}},$$

em que

$$\hat{\theta}^*(\cdot) = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^*(b).$$

3.2 Avaliação numérica

Os desempenhos dos estimadores intervalares dos parâmetros do modelo de regressão beta não linear com erros de medida foram avaliados através do estudo de simulações de Monte Carlo com $R = 10000$ réplicas.

A seguir são apresentadas as taxas de cobertura (TC) dos intervalos de confiança com coeficientes de confiança nominal 90%, 95% e 99%. Esta medida foi estimada da seguinte forma

$$TC = \frac{\#\{\Psi_j \in IC[\Psi_j; 1 - \alpha]\}}{R},$$

em que Ψ_j é o j -ésimo elemento de $\Psi = (\Psi_1, \Psi_2, \Psi_3, \Psi_4)^\top = (\alpha, \beta, \gamma, \lambda)^\top$, R é o número de réplicas de Monte Carlo e $IC[\Psi_j; 1 - \alpha]$ é o intervalo de confiança para Ψ_j com coeficiente nominal $1 - \alpha$.

3.2.1 Cenário 1: Dispersão constante

Considere o modelo de regressão beta não linear com erros de medida sob o enfoque estrutural com a seguinte estrutura:

$$\begin{aligned} g(\mu_t) &= \alpha_0 + z_1^{\alpha_1} + \beta_1 x_1, \\ h(\phi) &= \gamma_0, \\ w_t &= x_t + e_t, \end{aligned}$$

em que, $t = 1, 2, \dots, n$. Para cada réplica de Monte Carlo, geraram-se ocorrências da variável sem erro de medida z_{t1} com distribuição uniforme (0.2, 1.2). Os seguintes valores para os parâmetros que permanecem fixos para todas as simulações foram escolhidos: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Como para cada réplica de Monte Carlo são geradas diferentes amostras aleatórias de tamanhos $n = 40, 80$ e 160 , não é possível obter um único intervalo para a média da variável resposta. Vale salientar também que o parâmetro γ_0 conduziu ao valor de $\phi \approx 150$ (148.41), constante. Inicialmente consideramos $Q = 50$ pontos de quadratura.

As Figuras 3.1 e 3.2 apresentam as taxas de cobertura dos intervalos com 95% de confiança para os estimadores do modelo do Cenário 1, considerando diferentes valores para o coeficiente de confiabilidade. É possível observar que os métodos aproximados de máxima verossimilhança (MV aproximada) e pseudo-verossimilhança (Pseudo MV aproximada) têm comportamentos semelhantes e com o aumento do número de observações as taxas de cobertura tendem para os valores desejados. Já os métodos *naive* (Naive) e calibração de regressão (Calibração da regressão) apresentam os piores desempenhos para os parâmetros β_1 e γ_0 , havendo uma diminuição da taxa de cobertura com o aumento do número de

observações. Com o aumento do erro de medida, ou seja, a diminuição do coeficiente de confiabilidade, nota-se que é necessário um número maior de observações para que a taxa de cobertura definida seja alcançada, especialmente para o parâmetro associado à dispersão.

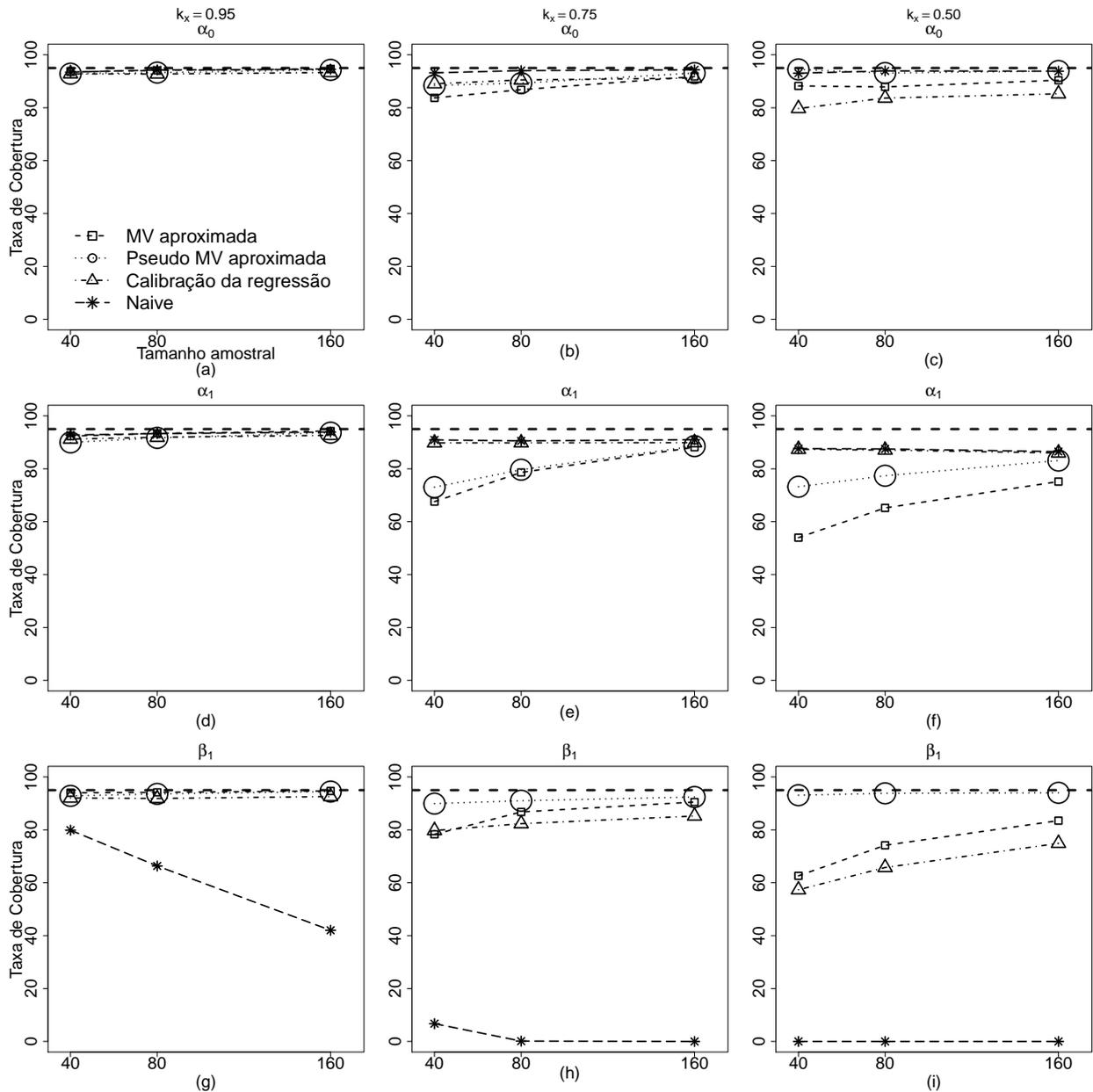


Figura 3.1: Taxas de cobertura dos intervalos de 95% de confiança para os parâmetros α_0 , α_1 e β_1 do modelo de regressão beta não linear com erros de medida com dispersão contante. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Cenário 1.

Para um erro pequeno, ou seja, $k_x = 0.95$, é possível verificar que, considerando os parâmetros α_0 e α_1 , todos os estimadores apresentam resultados semelhantes. Para o parâmetro β_1 nota-se que o método *Naive* apresenta pior resultado, o que é esperado, pois o mesmo não considera o erro de medida, como mostra as Figuras 3.1 (g)-(i). É notável como

o aumento do erro de medida prejudica os desempenhos dos intervalos de confiança para α_1 , parâmetro associado a não linearidade, no que diz respeito aos métodos de estimação da máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa), como mostra as Figuras 3.1 (d)-(f).

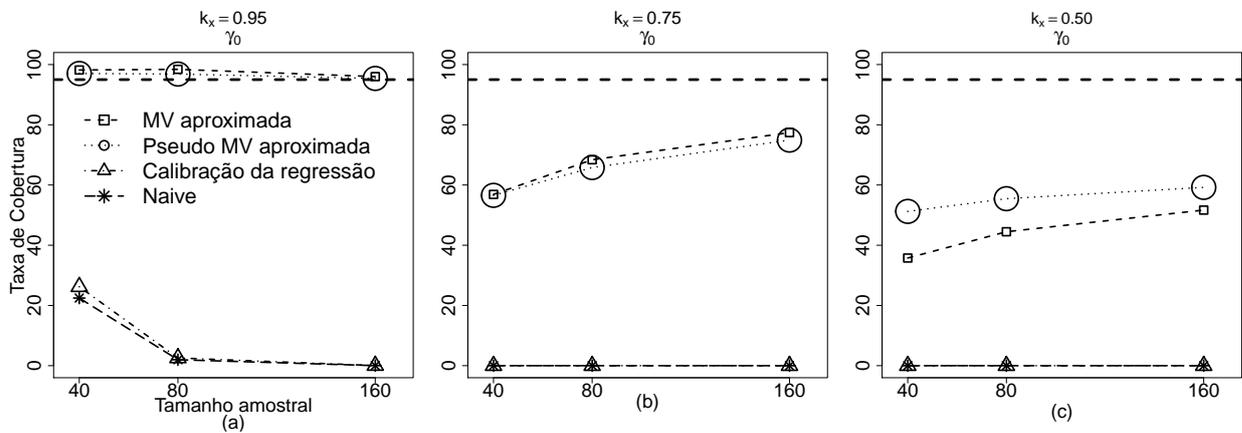


Figura 3.2: Taxas de cobertura dos intervalos de 95% de confiança para o parâmetro γ_0 do modelo de regressão beta não linear com erros de medida com dispersão contante. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida baixo). Cenário 1.

As Tabelas 3.1, 3.2 e 3.3 apresentam as taxas de cobertura (certo), os comprimentos médios (comp) e as taxas de não cobertura (inf e sup) dos intervalos de 90% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão constante, considerando os coeficientes de confiabilidade iguais a 0.95, 0.75 e 0.50, respectivamente. Nota-se que as taxas de cobertura para o parâmetro de não linearidade α_1 decaem com o aumento do erro de medida, ou seja, quando o coeficiente de confiabilidade diminui. De fato, para $k_x = 0.50$ as taxas de cobertura são muito baixas para os métodos MVa e MPVa os quais apresentam desempenho tipicamente bom com relação aos demais parâmetros do modelo. Ressalta-se que em todos os casos $\phi = 148.41$ e que uma investigação para valores maiores de ϕ se faz necessária.

De modo geral, para todos os coeficientes de confiabilidade, verificou-se que a taxa de cobertura aumenta com o aumento do número de observações. Para o estimador associado à variável com erros de medidas, β_1 , o método *Naive* apresenta o pior desempenho, ou seja, o método superestima o parâmetro β_1 . Além disso, considerando o parâmetro associado à dispersão, γ_0 , notou-se que os métodos *Naive* e RC apresentam péssimos desempenhos quanto a cobertura e que os métodos MVa e MPVa possuem intervalos cujos comprimentos se reduzem com o aumento do tamanho da amostra. Vale salientar que, com o aumento do erro de medida, os métodos *Naive* e RC pioram muito seus resultados em relação ao parâmetro γ_0 .

Tabela 3.1: Taxas de cobertura e comprimentos médios dos intervalos de 90% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão constante, utilizando os estimadores obtidos pelos métodos *naive* (*Naive*), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.95$. Cenário 1.

k_x	$\hat{\theta}$	Intervalo	$n = 40$						$n = 80$						$n = 160$					
			$Md_{\mu_t} = 0, 5135$			$Md_{\mu_t} = 0, 4968$			$Md_{\mu_t} = 0, 4968$			$Md_{\mu_t} = 0, 4942$								
			inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp		
0.95	α_0	<i>Naive</i>	5.52	87.66	6.82	0.13	5.00	88.88	6.12	0.11	4.58	89.27	6.15	0.07						
		RC	5.62	87.40	6.98	0.14	5.56	87.64	6.80	0.11	5.12	87.95	6.93	0.07						
		MVa	5.84	87.85	6.31	0.14	5.55	88.94	5.51	0.11	5.14	89.55	5.31	0.07						
		MPVa	6.31	87.19	6.50	0.14	6.09	87.89	6.02	0.11	5.34	89.13	5.53	0.07						
	α_1	<i>Naive</i>	5.00	87.39	7.61	1.05	4.36	87.90	7.74	0.88	3.79	88.13	8.08	0.58						
		RC	5.83	86.12	8.05	1.08	5.26	86.46	8.28	0.89	4.38	86.64	8.98	0.58						
		MVa	6.54	86.91	6.55	1.05	5.47	88.28	6.25	0.88	5.07	89.04	5.89	0.58						
		MPVa	8.27	83.99	7.74	1.01	6.78	86.03	7.19	0.86	5.71	88.17	6.12	0.57						
	β_1	<i>Naive</i>	0.29	70.06	29.65	0.14	0.06	54.66	45.28	0.10	0.01	29.79	70.20	0.07						
		RC	5.29	86.79	7.92	0.14	4.45	86.54	9.01	0.10	3.43	86.73	9.84	0.07						
		MVa	6.11	88.90	4.99	0.15	5.88	89.00	5.12	0.11	5.61	89.86	4.53	0.08						
		MVPa	6.69	87.70	5.61	0.15	6.16	88.44	5.40	0.11	5.89	89.34	4.77	0.07						
γ_0	<i>Naive</i>	0.00	14.58	85.42	0.73	0.00	0.91	99.09	0.52	0.00	0.01	99.99	0.36							
	RC	0.00	17.31	82.69	0.76	0.00	1.11	98.89	0.52	0.00	0.01	99.99	0.37							
	MVa	0.63	96.74	2.63	1.97	2.08	94.60	3.32	1.21	5.19	90.68	4.13	0.79							
	MVpa	2.58	93.92	3.50	1.88	4.26	91.99	3.75	1.18	5.62	89.97	4.41	0.77							

Tabela 3.2: Taxas de cobertura e comprimentos médios dos intervalos de 90% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão constante, utilizando os estimadores obtidos pelos métodos *naive* (*Naive*), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.75$. Cenário 1.

k_x	$\hat{\theta}$	Intervalo	$n = 40$						$n = 80$						$n = 160$												
			$Md_{\mu_t} = 0.5136$			$Md_{\mu_t} = 0.4967$			$Md_{\mu_t} = 0.4944$			$Md_{\mu_t} = 0.4967$			$Md_{\mu_t} = 0.4944$			$Md_{\mu_t} = 0.4944$									
			inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp					
0.75	α_0	<i>Naive</i>	4.50	87.86	7.64	0.23	3.97	88.44	7.59	0.19	3.42	88.96	7.62	0.12	4.50	87.86	7.64	0.23	3.97	88.44	7.59	0.19	3.42	88.96	7.62	0.12	
		RC	6.97	82.74	10.29	0.24	5.82	84.27	9.91	0.19	4.74	85.02	10.24	0.13	6.97	82.74	10.29	0.24	5.82	84.27	9.91	0.19	4.74	85.02	10.24	0.13	
		MVa	10.59	77.75	11.66	0.23	8.95	81.46	9.59	0.19	6.67	86.23	7.10	0.14	10.59	77.75	11.66	0.23	8.95	81.46	9.59	0.19	6.67	86.23	7.10	0.14	
		MPVa	8.11	83.09	8.80	0.38	7.72	83.93	8.35	0.22	5.91	87.70	6.39	0.14	8.11	83.09	8.80	0.38	7.72	83.93	8.35	0.22	5.91	87.70	6.39	0.14	
	α_1	<i>Naive</i>	3.77	85.45	10.78	1.74	3.06	85.23	11.71	1.47	1.60	85.19	13.21	0.97	3.77	85.45	10.78	1.74	3.06	85.23	11.71	1.47	1.60	85.19	13.21	0.97	
		RC	4.44	84.48	11.08	1.79	3.71	83.81	12.48	1.49	2.15	83.81	14.04	0.96	4.44	84.48	11.08	1.79	3.71	83.81	12.48	1.49	2.15	83.81	14.04	0.96	
		MVa	18.47	63.50	18.03	1.29	13.08	73.46	13.46	1.26	8.31	82.81	8.88	0.92	18.47	63.50	18.03	1.29	13.08	73.46	13.46	1.26	8.31	82.81	8.88	0.92	
		MPVa	15.87	68.37	15.76	1.87	12.70	74.56	12.74	1.41	8.08	83.30	8.62	0.97	15.87	68.37	15.76	1.87	12.70	74.56	12.74	1.41	8.08	83.30	8.62	0.97	
	β_1	<i>Naive</i>	0.01	3.75	96.24	0.21	0.00	0.06	99.94	0.15	0.00	0.00	100.00	0.10	0.01	3.75	96.24	0.21	0.00	0.06	99.94	0.15	0.00	0.00	0.00	100.00	0.10
		RC	10.29	71.76	17.95	0.25	4.77	74.93	20.30	0.19	1.70	77.49	20.81	0.15	10.29	71.76	17.95	0.25	4.77	74.93	20.30	0.19	1.70	77.49	20.81	0.15	
		MVa	17.95	73.13	8.92	0.28	11.66	81.64	6.70	0.21	8.85	85.53	5.62	0.16	17.95	73.13	8.92	0.28	11.66	81.64	6.70	0.21	8.85	85.53	5.62	0.16	
		MPVa	7.88	84.40	7.72	0.51	7.80	85.96	6.24	0.27	7.10	87.27	5.63	0.17	7.88	84.40	7.72	0.51	7.80	85.96	6.24	0.27	7.10	87.27	5.63	0.17	
γ_0	<i>Naive</i>	0.00	0.00	100.00	0.72	0.00	0.00	100.00	0.51	0.00	0.00	100.00	0.36	0.00	0.00	100.00	0.72	0.00	0.00	100.00	0.51	0.00	0.00	100.00	0.36		
	RC	0.00	0.00	100.00	0.75	0.00	0.00	100.00	0.52	0.00	0.00	100.00	0.36	0.00	0.00	100.00	0.75	0.00	0.00	100.00	0.52	0.00	0.00	100.00	0.36		
	MVa	38.27	54.79	6.94	3.03	27.30	65.59	7.11	2.71	18.39	73.89	7.72	2.27	38.27	54.79	6.94	3.03	27.30	65.59	7.11	2.71	18.39	73.89	7.72	2.27		
	MPVa	38.94	53.18	7.88	5.56	30.44	61.85	7.71	3.40	21.02	71.04	7.94	2.49	38.94	53.18	7.88	5.56	30.44	61.85	7.71	3.40	21.02	71.04	7.94	2.49		

Tabela 3.3: Taxas de cobertura e comprimentos médios dos intervalos de 90% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão constante, utilizando os estimadores obtidos pelos métodos *naive* (*Naive*), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.50$. Cenário 1.

k_x	$\hat{\theta}$	Intervalo	$n = 40$						$n = 80$						$n = 160$					
			$Md_{\mu_t} = 0.5135$			$Md_{\mu_t} = 0.4969$			$Md_{\mu_t} = 0.4941$			$Md_{\mu_t} = 0.4941$			$Md_{\mu_t} = 0.4941$					
			inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp		
0.50	α_0	<i>Naive</i>	3.94	87.67	8.39	0.31	3.11	88.61	8.28	0.25	2.70	88.36	8.94	0.17						
		RC	11.06	72.27	16.67	0.34	8.31	76.29	15.40	0.26	6.89	77.48	15.63	0.16						
		MVa	8.00	82.14	9.86	0.42	8.33	81.95	9.72	0.30	7.06	84.37	8.57	0.21						
		MPVa	3.85	90.63	5.52	5.69	5.06	88.36	6.58	0.76	4.64	89.16	6.20	0.33						
	α_1	<i>Naive</i>	3.20	82.19	14.61	2.21	2.25	81.64	16.11	1.87	0.85	80.15	19.00	1.24						
		RC	3.82	81.88	14.30	2.28	2.84	80.89	16.27	1.91	1.35	79.55	19.10	1.24						
		MVa	25.35	49.62	25.03	1.33	20.12	60.02	19.86	1.35	14.95	69.56	15.49	1.04						
		MPVa	15.47	68.95	15.58	7.54	13.69	72.17	14.14	3.44	10.51	77.90	11.59	1.58						
β_1	<i>Naive</i>	0.00	0.01	99.99	0.21	0.00	0.00	100.00	0.15	0.00	0.00	100.00	0.11							
	RC	21.50	50.05	28.45	0.44	13.11	57.45	29.44	0.31	5.40	66.58	28.02	0.26							
	MVa	26.80	56.63	16.57	0.41	18.98	68.33	12.69	0.33	12.62	77.53	9.85	0.26							
	MPVa	0.70	89.78	9.52	21.79	1.97	89.98	8.05	1.25	3.78	89.12	7.10	0.48							
γ_0	<i>Naive</i>	0.00	0.00	100.00	0.71	0.00	0.00	100.00	0.50	0.00	0.00	100.00	0.36							
	RC	0.00	0.00	100.00	0.74	0.00	0.00	100.00	0.51	0.00	0.00	100.00	0.36							
	MVa	55.13	33.51	11.36	2.61	47.01	41.79	11.20	2.47	40.27	48.63	11.10	2.19							
	MPVa	40.77	46.76	12.47	36.61	36.83	51.48	11.69	9.58	33.99	54.51	11.50	4.53							

Para os intervalos com 99% de confiança, resultados similares foram encontrados e os mesmos estão disponíveis nas Tabelas 9, 10 e 11 no Apêndice 7.1, considerando os coeficientes de confiabilidade $k_x = 0.50$, $k_x = 0.75$ e $k_x = 0.95$, respectivamente. Novamente, para todos os coeficientes de confiabilidade, verificou-se que a taxa de cobertura aumenta com o aumento do número de observações. Resultados similares são verificados quando consideramos o nível de 99% de confiança.

3.2.2 Cenário 2: Dispersão variável

Neste cenário admite-se que o modelo de regressão beta não linear com erros de medida com dispersão variável, sob o enfoque estrutural, apresenta a estrutura

$$\begin{aligned} g(\mu_t) &= \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}, \\ h(\phi_t) &= \gamma_0 + \lambda_1 x_{t1} \quad e \\ w_t &= x_t + e_t, \quad t = 1, 2, \dots, n. \end{aligned}$$

Aqui, foram geradas ocorrências da covariável $z_{t1} \sim U(0.2, 1.2)$ e os verdadeiros valores dos parâmetros são $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 2.5$, $\lambda_1 = 0.9$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Considerou-se diferentes tamanhos de amostras $n = (40, 80, 160)$ e o ponto de quadratura foi $Q = 50$.

As Figuras 3.3 e 3.4 apresentam as taxas de cobertura dos intervalos com 95% de confiança para os estimadores associados aos modelos da dispersão e da média, respectivamente, para o modelo do Cenário 2, considerando diferentes valores para o coeficiente de confiabilidade.

Note que os métodos de MVa e MPVa apresentam melhores desempenhos para todos os coeficientes de confiabilidade considerados. Considerando o parâmetro associado à variável com erro de medida, β_1 , é possível notar que o método *Naive* apresenta o pior resultado, especialmente com o aumento do coeficiente de confiabilidade (k_x) e tamanho amostral (n). Em relação aos parâmetros associados à dispersão do modelo, mais uma vez os estimadores MVa e MPVa apresentam os melhores resultados, tendo para a taxa de cobertura com o aumento do tamanho amostral.

Neste cenário em que a dispersão é modelada é notável a melhora do desempenho dos intervalos de confiança para α_1 , parâmetro que define a não linearidade do modelo, no que diz respeito aos métodos MVa e MPVa (Figura 3.4 (d)-(f)).

As Tabelas 3.4, 3.5 e 3.6 apresentam as taxas de cobertura (certo), os comprimentos médios (comp) e as taxas de não cobertura (inf e sup) dos intervalos de 90% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão variável, para os coeficientes de confiabilidade de $k_x = 0.95$, $k_x = 0.75$ e $k_x = 0.50$, respectivamente.

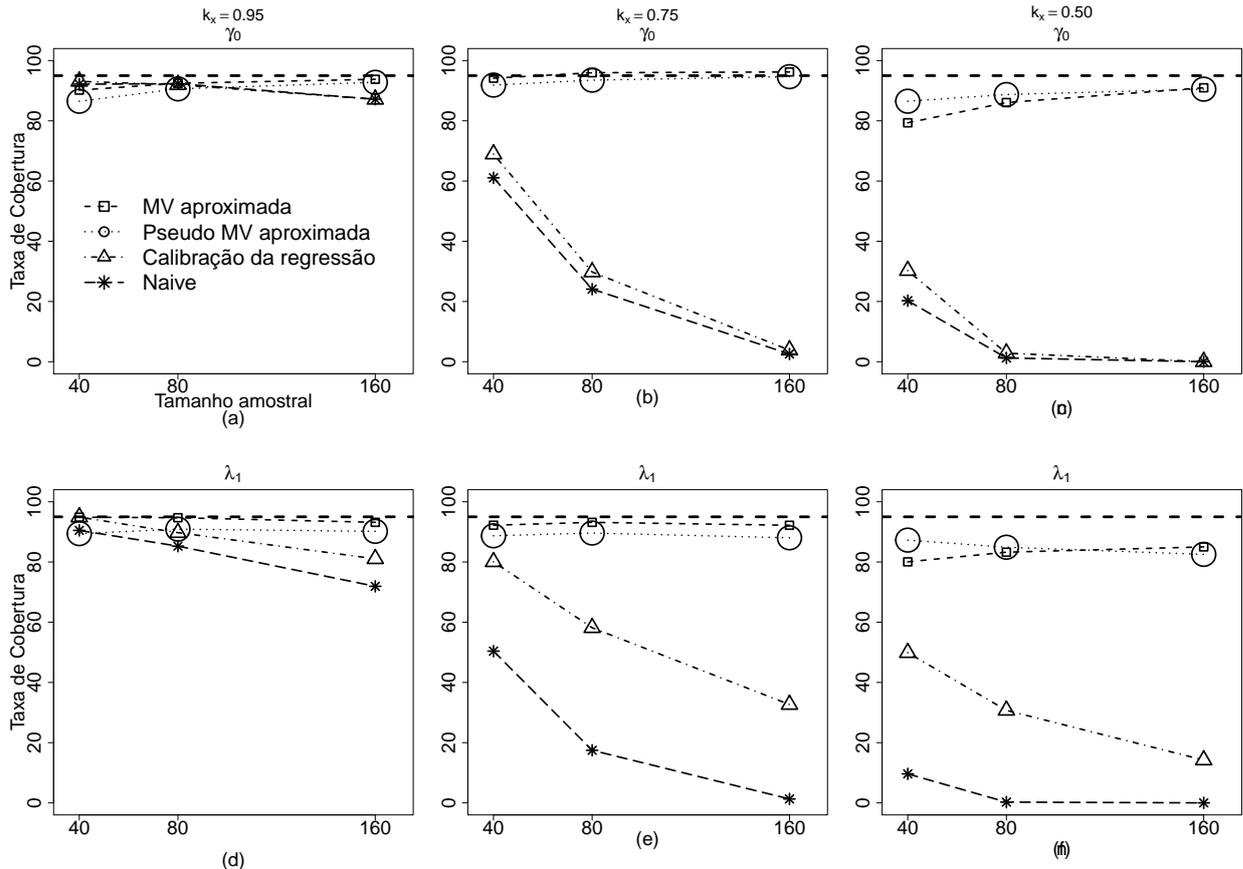


Figura 3.3: Taxas de cobertura dos intervalos de 95% de confiança para os parâmetros γ_0 e λ_1 do modelo de regressão beta não linear com erros de medida com dispersão variável. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Cenário 2.

Considerando o coeficiente de confiabilidade, $k_x = 0.95$, verificou-se que a taxa de cobertura diminui com o aumento do número de medidas, em relação ao parâmetro α_0 , apenas para o método RC.

Em relação ao parâmetro, α_1 (não linearidade) verificou-se que a taxa de cobertura aumenta com o aumento do número de medidas, em todos os métodos avaliados. Para o estimador associado à variável com erros de medidas, β_1 , o método MVa apresenta o melhor desempenho. Além disso, o método *Naive* superestima o parâmetro β_1 . Considerando o parâmetro associado à dispersão, γ_0 , notou-se que os métodos *Naive* e RC apresentam os piores desempenhos e para o parâmetro relacionado à variável com erro de medida, λ_1 , os métodos MVa e MPVa possuem os melhores resultados (Tabela 3.4).

O mesmo pode ser verificado quando se consideram os coeficientes de confiabilidade iguais a $k_x = 0.75$ e $k_x = 0.50$ (Tabela 3.5 e Tabela 3.6, respectivamente), exceto em relação ao parâmetro α_0 , em que a taxa de cobertura aumenta com o aumento do tamanho da amostra para todos os métodos avaliados.

Vale salientar que, com o aumento do erro de medida, os métodos *Naive* e RC pioraram muito seus resultados quanto à estimação intervalar em relação aos parâmetros γ_0 e λ_1

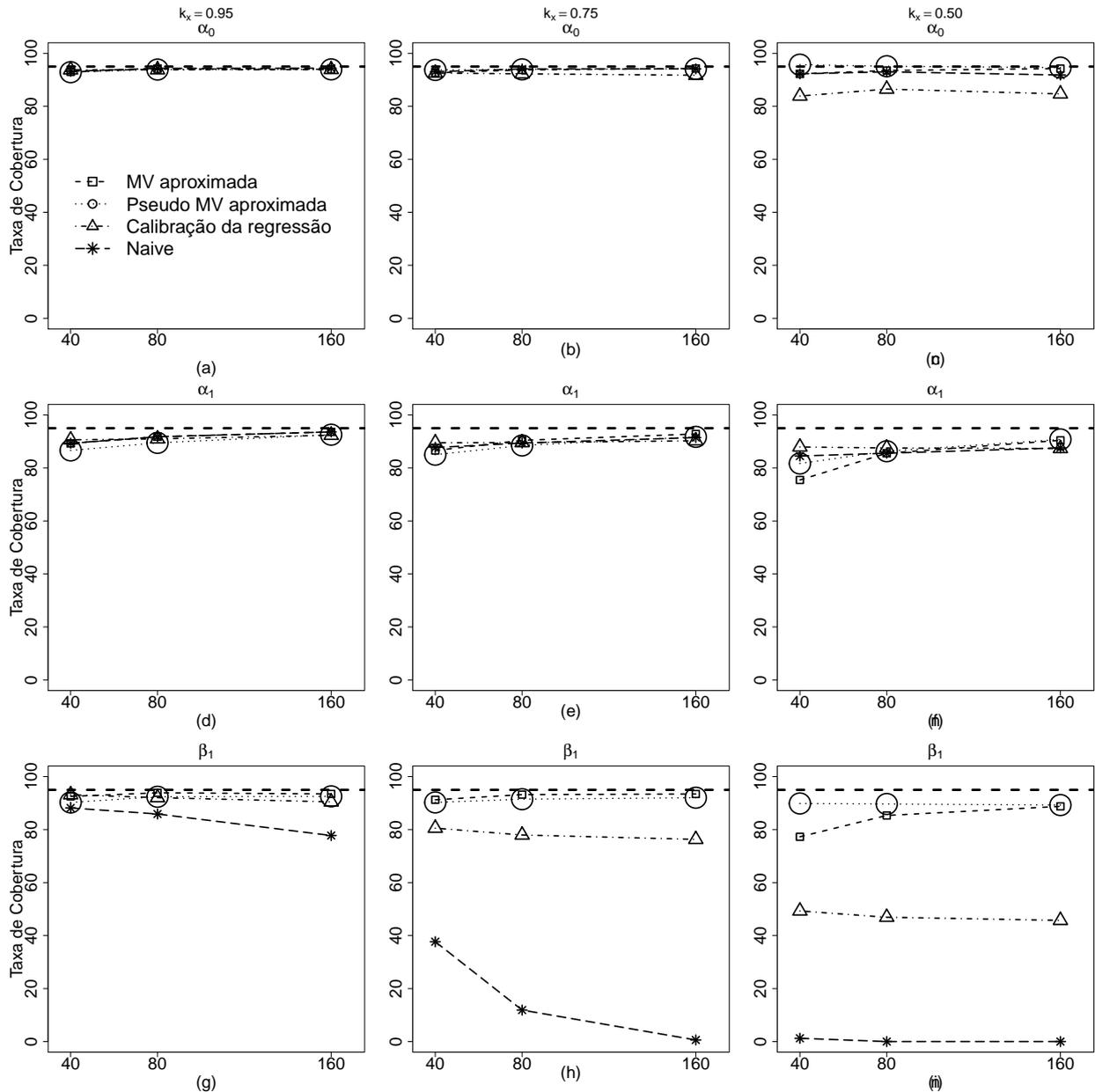


Figura 3.4: Taxas de cobertura dos intervalos de 95% de confiança para os parâmetros α_0 , α_1 e β_1 do modelo de regressão beta não linear com erros de medida com dispersão variável. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida baixo). Cenário 2.

(associado ao erro de medida da covariável do submodelo da dispersão). Além disso, há um aumento considerável no comprimento dos intervalos obtido pelo método MPVa. Nota-se uma melhoria considerável nas taxas de cobertura relativas ao intervalo de confiança dos parâmetros de não linearidade da média quantos aos métodos MVa e MPVa.

Para os intervalos com 99% de confiança, resultados similares foram encontrados e os mesmos estão disponíveis nas Tabelas 12, 13 e 14, Apêndice 7.1, considerando os coeficientes de confiabilidade $k_x = 0.50$, $k_x = 0.75$ e $k_x = 0.95$, respectivamente.

Tabela 3.4: Taxas de cobertura e comprimentos médios dos intervalos de 90% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão variável, utilizando os estimadores obtidos pelos métodos *naive* (*Naive*), calibração da regressão (RC), máxima verossimilhança aproximada (MPVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.95$. Cenário 2.

k_x	$\hat{\theta}$	Intervalo	$n = 40$				$n = 80$				$n = 160$			
			$Md_{\mu_t} = 0.5156$				$Md_{\mu_t} = 0.4989$				$Md_{\mu_t} = 0.4962$			
			$\phi_t \in [2.1219; 94.4678]$				$\phi_t \in [1.7209; 117.8152]$				$\phi_t \in [1.4157; 145.5675]$			
	inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp		
0.95	α_0	<i>Naive</i>	6.00	87.38	6.62	0.35	5.26	89.14	5.60	0.27	5.76	89.08	5.16	0.19
		RC	4.97	89.18	5.85	0.38	5.17	89.05	5.78	0.28	6.08	88.43	5.49	0.19
		MVa	6.58	87.32	6.10	0.35	6.27	88.68	5.05	0.27	7.15	88.47	4.38	0.19
		MPVa	6.66	87.03	6.31	0.36	6.43	88.17	5.40	0.27	7.09	88.62	4.29	0.19
	α_1	<i>Naive</i>	5.26	84.15	10.59	2.34	3.97	86.97	9.06	2.01	3.62	88.71	7.67	1.33
		RC	5.17	85.88	8.95	2.56	4.99	86.11	8.90	2.08	4.78	87.12	8.10	1.33
		MVa	5.83	84.13	10.04	2.33	4.77	86.58	8.65	2.00	4.70	88.36	6.94	1.32
		MPVa	7.62	81.13	11.25	2.27	6.55	83.85	9.60	1.96	5.42	87.28	7.30	1.31
	β_1	<i>Naive</i>	2.15	81.70	16.15	0.32	0.93	77.99	21.08	0.22	0.33	67.36	32.31	0.15
		RC	5.27	87.97	6.76	0.37	5.03	86.50	8.47	0.23	4.47	84.58	10.95	0.16
		MVa	5.13	86.96	7.91	0.35	3.65	88.83	7.52	0.24	3.08	88.13	8.79	0.17
		MVPa	6.36	84.28	9.36	0.33	4.51	86.99	8.50	0.23	3.76	86.56	9.68	0.16
γ_0	<i>Naive</i>	8.91	86.25	4.84	0.72	3.61	85.88	10.51	0.50	1.21	79.11	19.68	0.35	
	RC	8.05	87.70	4.25	0.80	3.92	85.93	10.15	0.52	1.37	79.28	19.35	0.36	
	MVa	15.72	83.08	1.20	0.86	11.62	86.86	1.52	0.59	10.37	87.86	1.77	0.41	
	MVpa	19.18	78.90	1.92	0.81	13.39	84.50	2.11	0.57	10.82	87.07	2.11	0.40	
λ_1	<i>Naive</i>	3.04	83.70	13.26	0.81	1.01	76.40	22.59	0.53	0.22	60.65	39.13	0.36	
	RC	2.96	91.24	5.80	1.07	2.38	83.19	14.43	0.58	1.02	72.76	26.22	0.37	
	MVa	4.59	89.51	5.90	0.98	3.04	89.40	7.56	0.64	1.88	87.59	10.53	0.43	
	MPVa	7.80	83.30	8.90	0.86	4.75	84.98	10.27	0.57	2.74	84.11	13.15	0.40	

Tabela 3.5: Taxas de cobertura e comprimentos médios dos intervalos de 90% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão variável, utilizando os estimadores obtidos pelos métodos *naive* (*Naive*), calibração da regressão (RC), máxima verossimilhança aproximada (MPVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.75$. Cenário 2.

k_x	$\hat{\theta}$	Intervalo	$n = 40$				$n = 80$				$n = 160$			
			$Md_{\mu_t} = 0.5157$				$Md_{\mu_t} = 0.4989$				$Md_{\mu_t} = 0.4962$			
			$\phi_t \in [2.1106; 93.2445]$				$\phi_t \in [1.7202; 117.6784]$				$\phi_t \in [1.4161; 145.6058]$			
	inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp		
α_0	<i>Naive</i>	4.67	87.63	7.70	0.41	3.59	88.89	7.52	0.32	3.25	88.46	8.29	0.22	
	RC	4.65	87.23	8.12	0.44	3.82	86.80	9.38	0.33	3.78	85.27	10.95	0.22	
	MVa	6.88	87.19	5.93	0.43	6.32	88.65	5.03	0.33	7.12	88.53	4.35	0.22	
	MPVa	6.33	88.63	5.04	0.67	6.29	88.71	5.00	0.34	7.17	88.53	4.30	0.22	
α_1	<i>Naive</i>	4.05	82.67	13.28	2.73	2.68	84.42	12.90	2.36	1.86	86.38	11.76	1.57	
	RC	4.79	84.16	11.05	2.88	4.21	84.12	11.67	2.36	3.10	84.81	12.09	1.54	
	MVa	7.14	81.34	11.52	2.65	5.11	85.38	9.51	2.31	4.50	87.80	7.70	1.54	
	MPVa	8.24	79.08	12.68	3.39	6.54	82.95	10.51	2.28	5.43	86.48	8.09	1.52	
0.75	<i>Naive</i>	0.03	27.84	72.13	0.34	0.00	7.04	92.96	0.23	0.00	0.26	99.74	0.16	
	RC	12.10	73.20	14.70	0.38	11.86	69.86	18.28	0.24	8.70	68.25	23.05	0.16	
	MVa	4.89	86.54	8.57	0.52	3.84	87.98	8.18	0.36	2.68	88.21	9.11	0.25	
	MPVa	5.07	84.86	10.07	1.36	4.90	85.78	9.32	0.34	5.43	86.27	10.34	0.24	
γ_0	<i>Naive</i>	0.53	50.60	48.87	0.71	0.02	16.18	83.80	0.50	0.00	1.43	98.57	0.35	
	RC	0.57	58.23	41.20	0.83	0.03	19.84	80.13	0.55	0.00	1.91	98.09	0.39	
	MVa	6.90	90.88	2.22	1.47	5.39	91.42	3.19	0.96	5.04	91.21	3.75	0.63	
	MPVa	10.49	86.51	3.00	3.78	8.19	88.01	3.80	0.95	6.44	89.40	4.16	0.62	
λ_1	<i>Naive</i>	0.16	39.31	60.53	0.71	0.00	11.66	88.34	0.47	0.00	0.67	99.33	0.32	
	RC	3.71	71.85	24.44	0.96	1.45	48.96	49.59	0.53	0.32	24.81	74.87	0.34	
	MVa	4.22	88.54	7.24	1.43	2.04	88.68	9.28	0.92	0.72	86.32	12.96	0.60	
	MPVa	5.82	83.80	10.38	2.46	4.07	83.42	12.51	0.82	1.83	81.01	17.16	0.54	

Tabela 3.6: Taxas de cobertura e comprimentos médios dos intervalos de 90% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão variável, utilizando os estimadores obtidos pelos métodos *naive* (*Naive*), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.50$. Cenário 2.

k_x	$\hat{\theta}$	Intervalo	$n = 40$						$n = 80$						$n = 160$					
			$Md_{\mu_t} = 0.5160$						$Md_{\mu_t} = 0.4988$						$Md_{\mu_t} = 0.4962$					
			$\phi_t \in [2.1182; 92.8067]$						$\phi_t \in [1.7201; 117.2303]$						$\phi_t \in [1.4160; 145.4027]$					
			inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp		
0.50	α_0	<i>Naive</i>	3.95	85.65	10.40	0.46	2.57	86.93	10.50	0.36	1.89	85.47	12.64	0.25						
		RC	7.15	77.20	15.65	0.50	4.40	79.20	16.40	0.38	3.47	76.79	19.74	0.25						
		MVa	6.21	86.91	6.88	0.60	6.11	88.24	5.65	0.43	6.51	89.12	4.37	0.29						
		MPVa	4.37	91.53	4.10	10.94	5.28	90.45	4.27	0.90	6.10	89.81	4.09	0.33						
	α_1	<i>Naive</i>	3.01	79.65	17.34	3.00	1.96	80.71	17.33	2.60	1.14	81.72	17.14	1.75						
		RC	4.31	81.63	14.06	3.13	3.55	81.52	14.93	2.61	2.22	81.02	16.76	1.73						
		MVa	12.56	70.73	16.71	2.60	7.00	80.58	12.42	2.45	5.32	85.55	9.13	1.72						
		MPVa	8.79	76.76	14.45	22.49	6.71	81.21	12.08	5.10	5.60	85.33	9.07	1.89						
	β_1	<i>Naive</i>	0.00	0.66	99.34	0.31	0.00	0.00	100.00	0.22	0.00	0.00	100.00	0.15						
		RC	28.86	42.49	28.65	0.35	25.80	40.17	34.03	0.22	20.43	39.12	40.45	0.15						
		MVa	15.42	73.57	11.01	0.79	9.05	80.96	9.99	0.58	5.68	84.00	10.32	0.41						
		MPVa	2.78	85.19	12.03	51.58	4.56	84.28	11.16	2.02	5.23	83.44	11.33	0.48						
γ_0	<i>Naive</i>	0.06	14.25	85.69	0.71	0.00	0.69	99.31	0.48	0.00	0.00	100.00	0.34							
	RC	0.54	21.39	78.07	0.85	0.01	1.53	98.46	0.57	0.00	0.00	100.00	0.41							
	MVa	18.96	76.59	4.45	2.05	11.75	83.09	5.16	1.47	6.29	87.58	6.13	0.99							
	MPVa	12.19	82.56	5.25	73.83	9.62	84.56	5.82	7.58	7.27	85.90	6.83	1.42							
λ_1	<i>Naive</i>	0.00	6.21	93.79	0.57	0.00	0.11	99.89	0.37	0.00	0.00	100.00	0.26							
	RC	11.97	42.13	45.90	0.79	5.73	25.24	69.03	0.44	1.52	11.44	87.04	0.28							
	MVa	13.65	75.94	10.41	1.79	9.92	78.76	11.32	1.19	5.13	79.60	15.27	0.79							
	MPVa	5.80	82.49	11.71	100.07	6.34	79.21	14.45	4.36	4.39	75.71	19.90	0.98							

3.2.3 Cenário 3: Não linearidade na dispersão

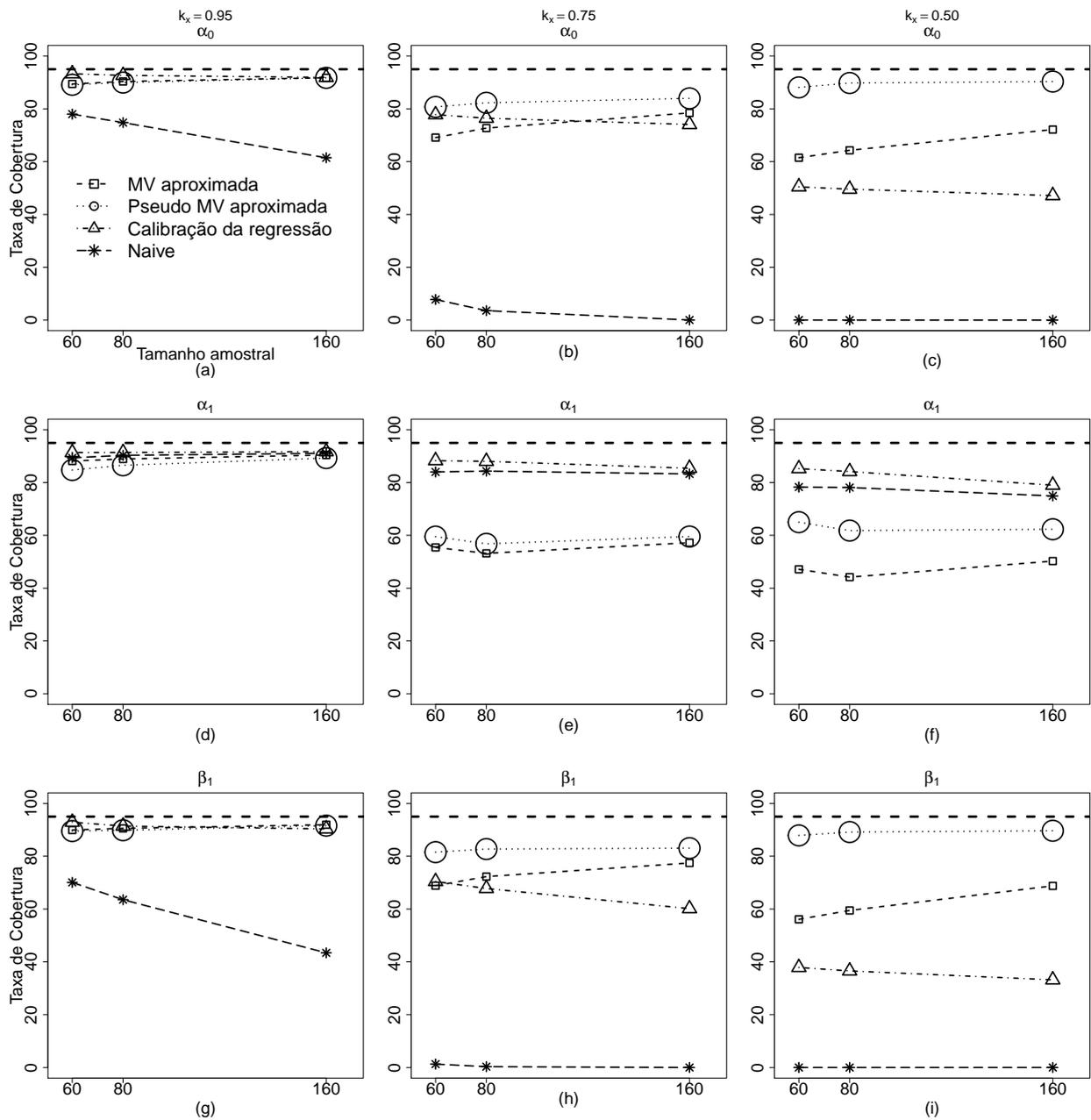


Figura 3.5: Taxas de cobertura dos intervalos de 95% de confiança para os parâmetros α_0 , α_1 e β_1 do modelo de regressão beta não linear com erros de medida com não linearidade na dispersão. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida baixo). Cenário 3.

Neste cenário admite-se que o modelo de regressão beta não linear com erros de medida com não linearidade na dispersão, sob o enfoque estrutural, apresenta a estrutura

$$\begin{aligned}
 g(\mu_t) &= \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}, \\
 h(\phi_t) &= \gamma_0 + v_{t1}^{\gamma_1} + \lambda_1 x_{t1} \quad e \\
 w_t &= x_t + e_t,
 \end{aligned}$$

em que $t = 1, 2, \dots, n$. Para cada réplica de Monte Carlo, geraram-se ocorrências das variáveis sem erro de medida z_{t1} e v_{t1} , ambas com distribuição uniforme $(0.2, 1.2)$. Os seguintes valores para os parâmetros que permanecem fixos para todas as simulações foram escolhidos: $\alpha_0 = 0.7$, $\alpha_1 = 2.0$, $\beta_1 = -1.5$, $\gamma_0 = 1.5$, $\gamma_1 = 2.0$, $\lambda_1 = 1.3$, $\mu_x = 1.5$ e $\sigma_x^2 = 0.5$. Diferentes tamanhos de amostras aleatórias $n = 60, 80$ e 160 foram considerados e os pontos de quadratura foram $Q = 50$.

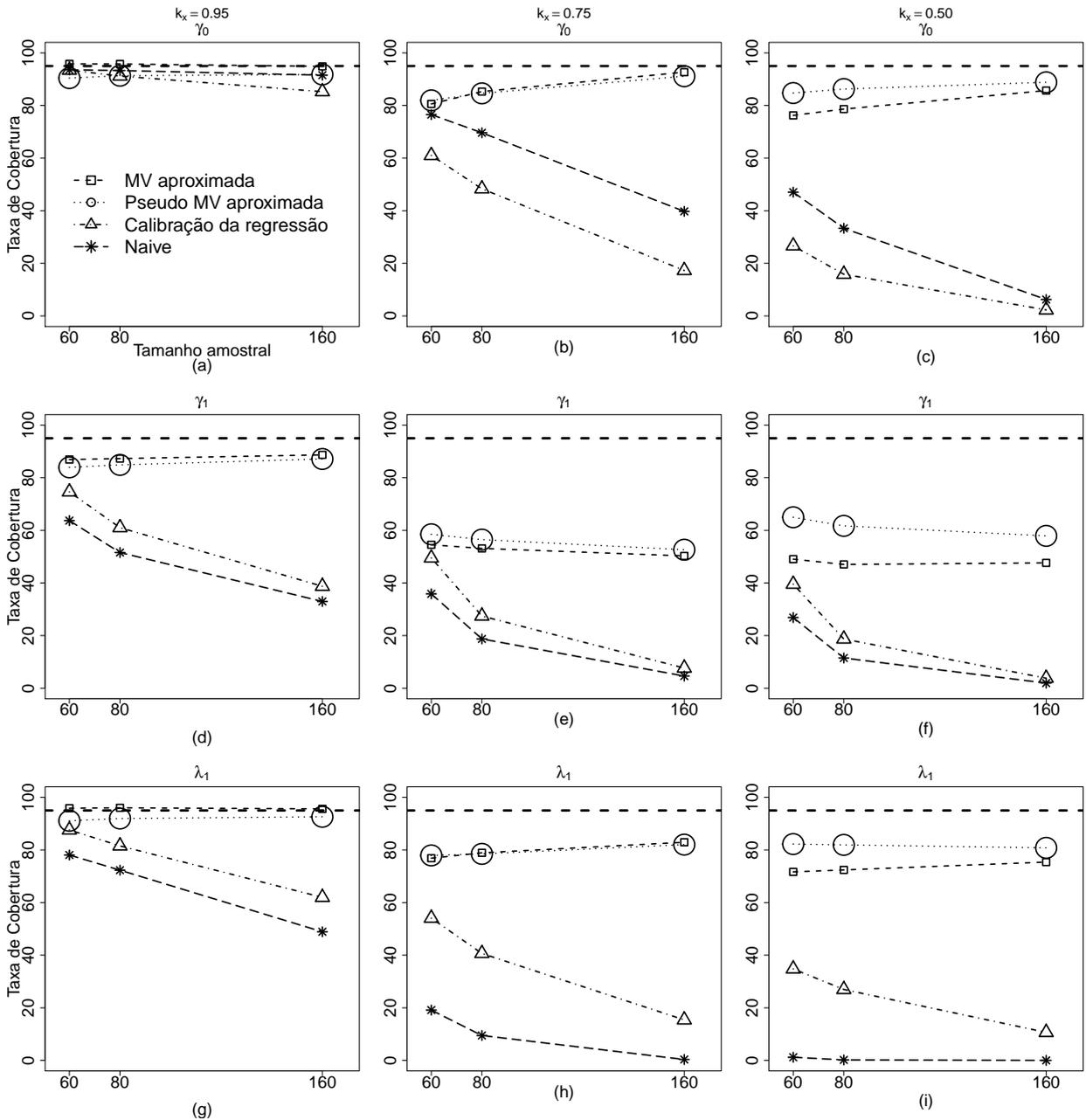


Figura 3.6: Taxas de cobertura dos intervalos de 95% de confiança para os parâmetros γ_0 , γ_1 e λ_1 do modelo de regressão beta não linear com erros de medida com não linearidade na dispersão. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Cenário 3.

Tabela 3.7: Taxas de cobertura e comprimentos médios dos intervalos de 90% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com não linearidade na dispersão, utilizando os estimadores obtidos pelos métodos *naive* (*Naive*), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.95$. Cenário 3.

k_x	$\hat{\theta}$	Intervalo	$n = 60$						$n = 80$						$n = 160$											
			$Md_{\mu_t} = 0.2620$						$Md_{\mu_t} = 0.2600$						$Md_{\mu_t} = 0.2687$											
			$\phi_t \in [8.7692; 1345.9608]$		sup		comp		$\phi_t \in [7.8235; 1518.0251]$		sup		comp		$\phi_t \in [6.0253; 1762.6010]$		sup		comp							
0.95	α_0	<i>Naive</i>	0.29	68.13	31.58	0.44	0.32	64.11	35.57	0.38	0.06	49.51	50.43	0.27	0.29	68.13	31.58	0.44	0.32	64.11	35.57	0.38	0.06	49.51	50.43	0.27
		RC	2.98	88.69	8.33	0.49	3.68	87.10	9.22	0.40	3.61	87.14	9.25	0.27	2.98	88.69	8.33	0.49	3.68	87.10	9.22	0.40	3.61	87.14	9.25	0.27
		MVa	12.69	76.52	10.79	0.41	11.11	79.30	9.59	0.36	7.74	85.38	6.88	0.27	12.69	76.52	10.79	0.41	11.11	79.30	9.59	0.36	7.74	85.38	6.88	0.27
		MPVa	12.65	76.44	10.91	0.41	11.01	79.36	9.63	0.36	7.68	85.20	7.12	0.27	12.65	76.44	10.91	0.41	11.01	79.36	9.63	0.36	7.68	85.20	7.12	0.27
	α_1	<i>Naive</i>	5.62	83.69	10.69	1.19	5.46	83.90	10.64	1.03	4.30	84.91	10.79	0.73	5.62	83.69	10.69	1.19	5.46	83.90	10.64	1.03	4.30	84.91	10.79	0.73
		RC	3.88	87.14	8.98	1.42	4.16	86.89	8.95	1.19	3.57	86.36	10.07	0.79	3.88	87.14	8.98	1.42	4.16	86.89	8.95	1.19	3.57	86.36	10.07	0.79
		MVa	13.47	75.40	11.13	1.12	11.17	78.89	9.94	1.00	8.69	83.19	8.12	0.73	13.47	75.40	11.13	1.12	11.17	78.89	9.94	1.00	8.69	83.19	8.12	0.73
		MPVa	15.15	72.59	12.26	1.08	12.86	75.81	11.33	0.97	9.39	81.96	8.65	0.73	15.15	72.59	12.26	1.08	12.86	75.81	11.33	0.97	9.39	81.96	8.65	0.73
	β_1	<i>Naive</i>	41.55	58.32	0.13	0.24	47.81	52.12	0.07	0.20	67.95	32.04	0.01	0.14	41.55	58.32	0.13	0.24	47.81	52.12	0.07	0.20	67.95	32.04	0.01	0.14
		RC	9.85	87.33	2.82	0.27	11.14	85.57	3.29	0.22	12.29	84.69	3.02	0.15	9.85	87.33	2.82	0.27	11.14	85.57	3.29	0.22	12.29	84.69	3.02	0.15
		MVa	10.38	76.74	12.88	0.23	9.29	79.46	11.25	0.20	6.81	85.07	8.12	0.15	10.38	76.74	12.88	0.23	9.29	79.46	11.25	0.20	6.81	85.07	8.12	0.15
		MVPa	10.64	76.73	12.63	0.23	9.44	79.52	11.04	0.20	6.90	85.17	7.93	0.15	10.64	76.73	12.63	0.23	9.44	79.52	11.04	0.20	6.90	85.17	7.93	0.15
γ_0	<i>Naive</i>	4.75	87.36	7.89	1.51	4.09	87.35	8.56	1.27	2.19	84.82	12.99	0.87	4.75	87.36	7.89	1.51	4.09	87.35	8.56	1.27	2.19	84.82	12.99	0.87	
	RC	2.95	87.66	9.39	1.76	3.20	84.86	11.94	1.39	1.89	77.68	20.43	0.89	2.95	87.66	9.39	1.76	3.20	84.86	11.94	1.39	1.89	77.68	20.43	0.89	
	MVa	8.10	89.48	2.42	2.36	7.18	89.87	2.95	1.97	6.15	90.07	3.78	1.26	8.10	89.48	2.42	2.36	7.18	89.87	2.95	1.97	6.15	90.07	3.78	1.26	
	MVPa	11.89	82.32	5.79	2.15	9.95	84.26	5.79	1.81	8.22	86.35	5.43	1.17	11.89	82.32	5.79	2.15	9.95	84.26	5.79	1.81	8.22	86.35	5.43	1.17	
γ_1	<i>Naive</i>	0.27	51.41	48.32	2.12	0.14	42.49	57.37	1.79	0.02	22.69	77.29	1.35	0.27	51.41	48.32	2.12	0.14	42.49	57.37	1.79	0.02	22.69	77.29	1.35	
	RC	0.12	62.82	37.06	2.62	0.09	50.96	48.95	2.06	0.02	27.69	72.29	1.46	0.12	62.82	37.06	2.62	0.09	50.96	48.95	2.06	0.02	27.69	72.29	1.46	
	MVa	20.97	69.61	9.42	3.47	17.83	74.37	7.80	3.10	11.29	81.61	7.10	2.57	20.97	69.61	9.42	3.47	17.83	74.37	7.80	3.10	11.29	81.61	7.10	2.57	
	MVPa	21.53	66.28	12.19	3.24	18.51	71.64	9.85	2.95	11.98	79.71	8.31	2.51	21.53	66.28	12.19	3.24	18.51	71.64	9.85	2.95	11.98	79.71	8.31	2.51	
λ_1	<i>Naive</i>	0.36	66.16	33.48	0.93	0.14	59.14	40.72	0.78	0.05	36.55	63.40	0.53	0.36	66.16	33.48	0.93	0.14	59.14	40.72	0.78	0.05	36.55	63.40	0.53	
	RC	0.64	78.88	20.48	1.12	0.41	71.10	28.49	0.87	0.24	51.19	48.57	0.55	0.64	78.88	20.48	1.12	0.41	71.10	28.49	0.87	0.24	51.19	48.57	0.55	
	MVa	1.60	89.79	8.61	1.60	2.16	90.87	6.97	1.35	3.17	91.05	5.78	0.88	1.60	89.79	8.61	1.60	2.16	90.87	6.97	1.35	3.17	91.05	5.78	0.88	
	MPVa	4.29	83.98	11.73	1.48	4.39	85.94	9.67	1.25	4.96	87.02	8.02	0.82	4.29	83.98	11.73	1.48	4.39	85.94	9.67	1.25	4.96	87.02	8.02	0.82	

Tabela 3.8: Taxas de cobertura e comprimentos médios dos intervalos de 90% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com não linearidade na dispersão, utilizando os estimadores obtidos pelos métodos *naive* (Naive), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.75$. Cenário 3.

k_x	$\hat{\theta}$	Intervalo	$n = 60$						$n = 80$						$n = 160$							
			$Md_{\mu_t} = 0.2620$						$Md_{\mu_t} = 0.2602$						$Md_{\mu_t} = 0.2686$							
			$\phi_t \in [8.9305; 1324.0551]$		sup		comp		$\phi_t \in [7.9402; 1481.7885]$		sup		comp		$\phi_t \in [6.0559; 1716.1938]$		sup		comp			
0.75	α_0	Naive	0.00	4.68	95.32	0.61	0.00	1.72	98.28	0.53	0.06	49.51	50.43	0.27	0.00	1.72	98.28	0.53	0.06	49.51	50.43	0.27
		RC	8.76	69.98	21.26	0.67	8.13	68.98	22.89	0.56	3.61	87.14	9.25	0.27	3.61	87.14	9.25	0.56	3.61	87.14	9.25	0.27
		MVa	20.74	56.21	23.05	0.56	20.97	57.84	21.19	0.49	7.74	85.38	6.88	0.27	7.74	85.38	6.88	0.49	7.74	85.38	6.88	0.27
		MPVa	11.51	70.62	17.87	1.41	12.67	69.92	17.41	0.86	7.68	85.20	7.12	0.27	7.68	85.20	7.12	0.86	7.68	85.20	7.12	0.27
	α_1	Naive	4.87	76.58	18.55	1.85	3.83	77.98	18.19	1.64	4.30	84.91	10.79	0.73	4.30	84.91	10.79	1.64	4.30	84.91	10.79	0.73
		RC	2.94	82.79	14.27	2.29	2.65	82.85	14.50	1.96	3.57	86.36	10.07	0.79	3.57	86.36	10.07	1.96	3.57	86.36	10.07	0.79
		MVa	39.55	41.35	19.10	1.07	44.00	36.91	19.09	0.87	8.69	83.19	8.12	0.73	8.69	83.19	8.12	0.87	8.69	83.19	8.12	0.73
		MPVa	36.11	46.16	17.73	2.14	41.40	40.45	18.15	1.32	9.39	81.96	8.65	0.73	9.39	81.96	8.65	1.32	9.39	81.96	8.65	0.73
	β_1	Naive	99.22	0.78	0.00	0.33	99.90	0.10	0.00	0.28	67.95	32.04	0.01	0.14	67.95	32.04	0.01	0.28	67.95	32.04	0.01	0.14
		RC	28.72	62.37	8.91	0.36	32.14	60.11	7.75	0.30	12.29	84.69	3.02	0.15	12.29	84.69	3.02	0.30	12.29	84.69	3.02	0.15
		MVa	24.47	55.62	19.91	0.34	23.34	57.37	19.29	0.29	6.81	85.07	8.12	0.15	6.81	85.07	8.12	0.29	6.81	85.07	8.12	0.15
		MVPa	18.97	71.03	10.00	0.89	18.40	70.72	10.88	0.52	6.90	85.17	7.93	0.15	6.90	85.17	7.93	0.52	6.90	85.17	7.93	0.15
γ_0	Naive	0.95	67.86	31.19	1.37	0.50	59.24	40.26	1.16	2.19	84.82	12.99	0.87	2.19	84.82	12.99	1.16	2.19	84.82	12.99	0.87	
	RC	0.91	51.91	47.18	1.63	0.35	38.78	60.87	1.29	1.89	77.68	20.43	0.89	1.89	77.68	20.43	1.29	1.89	77.68	20.43	0.89	
	MVa	21.72	74.32	3.96	3.45	18.00	77.60	4.40	3.06	6.15	90.07	3.78	1.26	6.15	90.07	3.78	3.06	6.15	90.07	3.78	1.26	
	MVPa	18.31	76.16	5.53	8.54	15.40	78.80	5.80	4.87	8.22	86.35	5.43	1.17	8.22	86.35	5.43	4.87	8.22	86.35	5.43	1.17	
γ_1	Naive	0.12	25.68	74.20	2.07	0.01	13.32	86.67	1.71	0.02	22.69	77.29	1.35	0.02	22.69	77.29	1.71	0.02	22.69	77.29	1.35	
	RC	0.01	37.78	62.21	2.57	0.01	19.65	80.34	1.99	0.02	27.69	72.29	1.46	0.02	27.69	72.29	1.99	0.02	27.69	72.29	1.46	
	MVa	44.81	40.67	14.52	3.19	49.21	36.82	13.97	2.76	11.29	81.61	7.10	2.57	11.29	81.61	7.10	2.76	11.29	81.61	7.10	2.57	
	MVPa	39.74	46.00	14.26	7.27	44.94	41.62	13.44	4.62	11.98	79.71	8.31	2.51	11.98	79.71	8.31	4.62	11.98	79.71	8.31	2.51	
λ_1	Naive	0.00	12.88	87.12	0.84	0.00	5.95	94.05	0.71	0.05	36.55	63.40	0.53	0.05	36.55	63.40	0.71	0.05	36.55	63.40	0.53	
	RC	0.62	44.16	55.22	1.03	0.41	32.55	67.04	0.81	0.24	51.19	48.57	0.55	0.24	51.19	48.57	0.81	0.24	51.19	48.57	0.55	
	MVa	7.47	67.63	24.90	2.06	9.61	69.35	21.04	1.81	3.17	91.05	5.78	0.88	3.17	91.05	5.78	1.81	3.17	91.05	5.78	0.88	
	MPVa	6.24	70.99	22.77	5.41	8.39	71.36	20.25	3.10	4.96	87.02	8.02	0.82	4.96	87.02	8.02	3.10	4.96	87.02	8.02	0.82	

Tabela 3.9: Taxas de cobertura e comprimentos médios dos intervalos de 90% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com não linearidade na dispersão, utilizando os estimadores obtidos pelos métodos *naive* (*Naive*), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.50$. Cenário 3.

k_x	$\hat{\theta}$	Intervalo	$n = 60$						$n = 80$						$n = 160$							
			$Md_{\mu_t} = 0.2623$						$Md_{\mu_t} = 0.2603$						$Md_{\mu_t} = 0.2687$							
			$\phi_t \in [8.9631; 1353.5230]$		sup		comp		$\phi_t \in [7.9631; 1511.6661]$		sup		comp		$\phi_t \in [6.0751; 1746.4286]$		sup		comp			
inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp			
	α_0	<i>Naive</i>	0.00	0.01	99.99	0.65	0.00	0.00	100.00	0.57	0.00	0.00	100.00	0.40	0.00	0.00	100.00	0.40	0.00	0.00	100.00	0.40
		RC	22.64	42.57	34.79	0.73	21.29	42.37	36.34	0.61	16.60	40.11	43.29	0.41	16.60	40.11	43.29	0.41	16.60	40.11	43.29	0.41
		MVa	19.96	53.33	26.71	0.93	19.43	54.74	25.83	0.80	14.34	62.63	23.03	0.61	14.34	62.63	23.03	0.61	14.34	62.63	23.03	0.61
		MPVa	1.30	83.60	15.10	8.40	2.02	83.91	14.07	6.27	2.97	83.95	13.08	1.78	2.97	83.95	13.08	1.78	2.97	83.95	13.08	1.78
	α_1	<i>Naive</i>	4.74	71.42	23.84	2.24	3.03	72.21	24.76	2.00	1.93	67.97	30.10	1.41	1.93	67.97	30.10	1.41	1.93	67.97	30.10	1.41
		RC	2.75	79.69	17.56	2.79	1.87	78.67	19.46	2.43	1.37	72.65	25.98	1.62	1.37	72.65	25.98	1.62	1.37	72.65	25.98	1.62
		MVa	34.25	40.90	24.85	1.29	39.48	37.63	22.89	1.07	42.64	40.03	17.33	0.85	42.64	40.03	17.33	0.85	42.64	40.03	17.33	0.85
		MPVa	22.78	59.29	17.93	7.87	27.91	55.04	17.05	5.11	34.68	52.56	12.76	2.07	34.68	52.56	12.76	2.07	34.68	52.56	12.76	2.07
0.50	β_1	<i>Naive</i>	100.00	0.00	0.00	0.33	100.00	0.00	0.00	0.28	100.00	0.00	0.00	0.20	100.00	0.00	0.00	0.20	100.00	0.00	0.00	0.20
		RC	44.83	32.00	23.17	0.37	47.49	31.01	21.50	0.30	57.71	27.92	14.37	0.21	57.71	27.92	14.37	0.21	57.71	27.92	14.37	0.21
		MVa	29.47	48.44	22.09	0.54	29.03	50.66	20.31	0.47	27.63	58.64	13.73	0.36	27.63	58.64	13.73	0.36	27.63	58.64	13.73	0.36
		MVPa	15.72	83.30	0.98	5.59	15.06	83.77	1.17	4.16	15.78	82.26	1.96	1.15	15.78	82.26	1.96	1.15	15.78	82.26	1.96	1.15
	γ_0	<i>Naive</i>	0.11	36.12	63.77	1.13	0.05	23.92	76.03	0.96	0.00	3.53	96.47	0.66	0.00	3.53	96.47	0.66	0.00	3.53	96.47	0.66
		RC	0.96	20.72	78.32	1.35	0.50	11.89	87.61	1.08	0.00	1.57	98.43	0.72	0.00	1.57	98.43	0.72	0.00	1.57	98.43	0.72
		MVa	20.50	70.16	9.34	3.61	17.66	72.61	9.73	3.26	10.93	79.85	9.22	2.43	10.93	79.85	9.22	2.43	10.93	79.85	9.22	2.43
		MVpa	10.75	80.72	8.53	31.65	9.68	81.84	8.48	16.03	7.07	84.23	8.70	5.28	7.07	84.23	8.70	5.28	7.07	84.23	8.70	5.28
	γ_1	<i>Naive</i>	0.03	18.32	81.65	2.02	0.00	7.06	92.94	1.64	0.00	0.96	99.04	1.23	0.00	0.96	99.04	1.23	0.00	0.96	99.04	1.23
		RC	0.02	28.20	71.78	2.48	0.00	12.01	87.99	1.91	0.00	1.86	98.14	1.34	0.00	1.86	98.14	1.34	0.00	1.86	98.14	1.34
		MVa	31.48	41.79	26.73	3.28	32.08	39.40	28.52	2.88	38.63	36.98	24.39	2.54	38.63	36.98	24.39	2.54	38.63	36.98	24.39	2.54
		MVpa	21.09	58.47	20.44	29.58	22.33	54.75	22.92	17.83	31.04	48.15	20.81	6.95	31.04	48.15	20.81	6.95	31.04	48.15	20.81	6.95
	λ_1	<i>Naive</i>	0.00	0.68	99.32	0.67	0.00	0.10	99.90	0.56	0.00	0.00	100.00	0.38	0.00	0.00	100.00	0.38	0.00	0.00	100.00	0.38
		RC	5.02	28.54	66.44	0.83	3.38	22.24	74.38	0.65	0.84	7.79	91.37	0.42	0.84	7.79	91.37	0.42	0.84	7.79	91.37	0.42
		MVa	9.77	64.37	25.86	2.06	10.56	64.49	24.95	1.83	10.90	65.99	23.11	1.38	10.90	65.99	23.11	1.38	10.90	65.99	23.11	1.38
		MPVa	3.97	77.35	18.68	14.94	4.75	76.86	18.39	10.06	6.95	73.93	19.12	3.46	6.95	73.93	19.12	3.46	6.95	73.93	19.12	3.46

As Figuras 3.5 e 3.6 apresentam as taxas de cobertura dos intervalos com 95% de confiança para os estimadores associados aos modelos da média e da dispersão, respectivamente, considerando diferentes valores para o coeficiente de confiabilidade.

Considerando os parâmetros associados ao modelo da média é possível verificar que os métodos MVa e MPVa não estimam bem o parâmetro de não linearidade, α_1 , principalmente com o aumento do erros de medidas (Figuras 3.5 (d)-(f)). Vale salientar também que para os outros parâmetros o mesmo não é observado, ou seja, os métodos MVa e MPVa apresentam melhores resultados quando comparados com os métodos *Naive* e calibração da regressão.

Em relação aos parâmetros associados à dispersão do modelo, os estimadores obtidos pelos métodos MVa e MPVa apresentam os melhores resultados, tendendo para a taxa de cobertura fixada com o aumento do tamanho amostral. Note que, para o parâmetro associado à não linearidade, γ_1 , a taxa de cobertura não alcança o limite fixado de 95% de confiança (Figuras 3.6 (d)-(f)).

As Tabelas 3.7, 3.8 e 3.9 apresentam as taxas de cobertura (certo), os comprimentos médios (comp) e as taxas de não cobertura (inf e sup) dos intervalos de 90% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com não linearidade na dispersão, para os coeficientes de confiabilidade de $k_x = 0.95$, $k_x = 0.75$ e $k_x = 0.50$, respectivamente.

De modo geral, os métodos MVa e MPVa apresentam melhores desempenhos quanto ao intervalo de confiança do parâmetros. Vale salientar que com o aumento da amostra as taxas de cobertura vão aumentando. Novamente, é possível observar a diminuição da taxa de cobertura com o aumento do erro de medida.

Para os intervalos com 99% de confiança, resultados similares foram encontrados e os mesmos estão disponíveis nas Tabelas 15, 16 e 17, Apêndice 7.1, considerando os coeficientes de confiabilidade, $k_x = 0.50$, $k_x = 0.75$ e $k_x = 0.95$, respectivamente. Resultados similares aos dos intervalos com 90% de confiança foram obtidos considerando o nível de 99% de confiança.

3.3 Conclusões

Neste capítulo as distribuições assintóticas dos estimadores obtidos pelos métodos de máxima verossimilhança aproximada (MVa), pseudo-verossimilhança aproximada (MPVa) e calibração da regressão (RC) e os respectivos intervalos de confiança são apresentados.

Com o objetivo de avaliar a taxa de cobertura, bem como as taxas de não cobertura, realizamos simulações de Monte Carlo e concluímos que os métodos da máxima verossimilhança aproximada e pseudo-verossimilhança aproximada apresentam melhores

desempenhos quando comparados com os métodos *naive* e calibração da regressão.

Verificamos também que quanto menor o coeficiente de confiabilidade mais necessidade de um tamanho amostral maior que as taxas de cobertura fixadas sejam alcançadas, em especial para os parâmetros associados à dispersão do modelo. Concluímos também que o método da pseudo-verossimilhança aproximada apresentam os melhores resultados, porém o comprimento do intervalo é muito superior ao obtido pelo método da máxima verossimilhança aproximada. Através da estimação intervalar foi possível notar que para o erro de medida maior os métodos de estimação apresentam dificuldades para estimar parâmetros associados à não linearidade.

4 ANÁLISE DE RESÍDUOS

4.1 Introdução

A análise de resíduos é uma das etapas utilizadas para a escolha do modelo que melhor se adequa aos dados. Através desta etapa é possível detectar a presença de pontos aberrantes. Tais pontos devem ser investigados uma vez que os mesmos podem ser influentes ou não. Outro objetivo da análise de resíduos é verificar se existe adequação na distribuição proposta para a variável resposta e na função de ligação. Adicionalmente, com base em gráficos de resíduos é possível identificar tendências de não linearidade e dispersão erroneamente modeladas. Considere o resíduo para a t -ésima observação definido como

$$R_t = r_t(y_t, \hat{\mu}_t), \quad (4.1)$$

que mede a discrepância entre o valor observado (y_t) e o valor ajustado ($\hat{\mu}_t$) da t -ésima observação (Cox e Snell, 1968). A função $r_t(\cdot)$ é determinada de forma a estabilizar a variância e/ou induzir a simetria da distribuição amostral de R_t (Cordeiro, 1986). As escolhas mais comuns são

$$R_t = \frac{y_t - \hat{\mu}_t}{\sqrt{\widehat{Var}(y_t)}} \quad \text{e} \quad R_t = \frac{y_t - \hat{\mu}_t}{\sqrt{\widehat{Var}(y_t - \hat{\mu}_t)}}.$$

É interessante que a definição de resíduos através da Equação (4.1) satisfaça algumas propriedades, tais como $\mathbb{E}(R_t) = 0$, variância constante e $Cov(R_t, R_k) = 0, t \neq k$, pois em muitos casos estas condições facilitam a determinação aproximada da distribuição de R_t , o que torna a interpretação dos gráficos dos resíduos, com determinação dos limites de confiança, por exemplo, mais fácil e até mesmo mais confiável (Espinheira *et al.*, 2017). A análise gráfica é crucial na realização da análise dos resíduos. Alguns gráficos comumente utilizados são: resíduos versus índices das observações, resíduos versus valores ajustados, entre outros.

Considerando os modelos de regressão beta, a literatura apresenta uma gama de definições de resíduos. Ferrari e Cribari-Neto (2004) e Espinheira *et al.* (2008) definiram alguns resíduos para o modelo de regressão beta com dispersão constante. Ferrari *et al.* (2011) estenderam um dos resíduos propostos por Espinheira *et al.* (2008) para o modelo que considera a dispersão variável. Rocha e Simas (2011) definiram um novo resíduo para o modelo de regressão beta não linear e generalizaram alguns dos resíduos definidos por Espinheira *et al.* (2008). Espinheira *et al.* (2017) propuseram um novo resíduo para a classe de modelos de regressão beta denotado por resíduo combinado.

O objetivo deste capítulo é apresentar, sucintamente, os resíduos ponderado, ponderado padronizado e combinado para o modelo de regressão beta não linear e estender para o modelo de regressão beta não linear com erros de medida. Aqui, serão avaliados os resíduos obtidos pelos três métodos de estimação, denotados por máxima verossimilhança aproximada (MVa), máxima pseudo-verossimilhança aproximada (MPVa) e calibração da regressão (RC). Como o método *naive* apresentou, em geral, os piores desempenhos tanto em relação aos vieses e erros quadráticos médios quanto as taxas de cobertura, o mesmo é omitido neste capítulo.

4.2 Resíduos para o modelo de regressão beta não linear

O resíduo ponderado padronizado 1, denotado aqui por resíduo ponderado, foi proposto por Espinheira *et al.* (2008) para o modelo de regressão beta com dispersão constante. Uma possível padronização do resíduo ponderado, também proposto por Espinheira *et al.* (2008), é denotado aqui por resíduo ponderado padronizado. Ferrari *et al.* (2011) propuseram uma extensão do resíduo ponderado, adaptando-o para os casos em que a dispersão é variável.

Para o modelo de regressão beta não linear, os resíduos ponderado e ponderado padronizado foram generalizados por Rocha e Simas (2011). Espinheira *et al.* (2017) propuseram o resíduo combinado tanto para o modelo beta linear quanto para o modelo beta não linear.

Sejam y_1, \dots, y_n variáveis aleatórias independentes tais que, para cada $t = 1, \dots, n$, y_t segue distribuição beta com função de densidade dada por

$$f(y_t; \mu_t, \phi_t) = \frac{\Gamma(\phi_t)}{\Gamma(\mu_t \phi_t) \Gamma((1 - \mu_t) \phi_t)} y_t^{\mu_t \phi_t - 1} (1 - y_t)^{(1 - \mu_t) \phi_t - 1}, \quad 0 < y_t < 1. \quad (4.2)$$

No modelo de regressão beta não linear a média e o parâmetro de precisão as seguintes relações funcionais são satisfeitas:

$$g(\mu_t) = \eta_{1t} = f_1(\mathbf{z}_t^\top; \boldsymbol{\alpha})$$

e

$$h(\phi_t) = \eta_{2t} = f_2(\mathbf{v}_t^\top; \boldsymbol{\gamma}), \quad (4.3)$$

em que $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^\top \in \mathbb{R}^p$ e $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_{\check{q}}) \in \mathbb{R}^{\check{q}}$ são vetores de parâmetros desconhecidos, com $p + \check{q} < n$, η_{1t} e η_{2t} são os preditores e $\mathbf{z}_t^\top = (z_{t1}, \dots, z_{tp_1})$ e $\mathbf{v}_t^\top = (v_{t1}, \dots, v_{t\check{q}_1})$ são vetores de observações fixas e conhecidas com $t = 1, \dots, n$, $p_1 \leq p$ e $\check{q}_1 < \check{q}$. As funções de ligação $g(\cdot) : (0, 1) \rightarrow \mathbb{R}$ e $h(\cdot) : (0, \infty) \rightarrow \mathbb{R}$ são estritamente monótonas e duas vezes diferenciáveis.

4.2.1 Resíduo ponderado

Com base no vetor escore e na matriz de informação de Fisher (Santos, 2015) obtidos para o modelo de regressão beta não linear, segue que o proceso iterativo Scoring de Fisher para estimar α é dado por

$$\alpha^{(m+1)} = \alpha^{(m)} + (J_1^\top \Phi^{(m)} W^m J_1)^{-1} J_1^\top \Phi^{(m)} T^{(m)} (y^* - \mu^{*(m)}), \quad (4.4)$$

em que m representa os passos necessários até a convergência do processo, $J_1^\top = \partial \eta_1 / \partial \alpha$ é uma matriz de ordem $n \times p$, $\Phi = \text{diag}\{\phi_1, \dots, \phi_n\}$, $T = \text{diag}\{1/g'(\mu_1), \dots, 1/g'(\mu_n)\}$, $W = \text{diag}\{\omega_1, \dots, \omega_n\}$, com $\omega_t = \phi_t v_t \{1/[g'(\mu)]^2\}$, com

$$v_t = \text{Var}(y_t^*) = \psi'((1 - \mu_t)\phi_t) + \psi'(\mu_t\phi_t), \quad (4.5)$$

para todo $t = 1, \dots, n$.

Na forma de um processo iterativo de mínimos quadrados ponderados o processo é dado por

$$\alpha^{(m+1)} = (J_1^\top \Phi^{(m)} W^m J_1)^{-1} J_1^\top \Phi^{(m)} W^{(m)} u_1^{(m)},$$

em que $u_1^{(m)} = J_1 \alpha_1^{(m)} + W^{-1(m)} T^{(m)} (y^* - \mu^{*(m)})$, $y_t^* = \log\left(\frac{y_t}{1-y_t}\right)$ e $\mu_t^* = \psi(\mu_t\phi_t) - \psi((1 - \mu_t)\phi_t)$.

Após convergência,

$$\hat{\alpha} = (J_1^\top \hat{\Phi} \hat{W} J_1)^{-1} J_1^\top \hat{\Phi} \hat{W} u_1, \quad (4.6)$$

com $u_1 = J_1 \hat{\alpha} + \hat{W}^{-1} \hat{T} (y^* - \hat{\mu}^*)$, que pode ser visto como um estimador de mínimos quadrados considerando a regressão linear de $\hat{\Phi}^{1/2} \hat{W}^{1/2} u_1$ em $\hat{\Phi}^{1/2} \hat{W}^{1/2} J_1$. O resíduo dessa regressão é dado por

$$\begin{aligned} r^\alpha &= \hat{\Phi}^{1/2} \hat{W}^{1/2} u_1 - \hat{\Phi}^{1/2} \hat{W}^{1/2} J_1 \hat{\alpha} \\ &= \hat{\Phi}^{1/2} \hat{W}^{1/2} (u_1 - J_1 \hat{\alpha}) \\ &= \hat{\Phi}^{1/2} \hat{W}^{1/2} \hat{T} (y^* - \hat{\mu}^*), \end{aligned} \quad (4.7)$$

ou seja,

$$r_t^\alpha = \frac{y_t^* - \hat{\mu}_t^*}{\sqrt{\hat{v}_t}} \quad (4.8)$$

é o resíduo ponderado, em que $\text{Var}(y_t^*) = v_t$ é definido em (4.5).

4.2.2 Resíduo ponderado padronizado

Para a padronização do resíduo (4.8) baseada na variância de u_1 , consideramos que $Cov(\hat{\alpha}) \approx (J_1^\top \Phi W J_1)^{-1}$ e reescrevemos a Equação (4.6) como

$$(J_1^\top \Phi W J_1) \hat{\alpha} = J_1^\top \Phi W u_1. \quad (4.9)$$

Considerando $\widehat{W} \approx W$ e $\widehat{\Phi} \approx \Phi$ em (4.9), segue que

$$\begin{aligned} (J_1^\top \Phi W J_1) Cov(\hat{\alpha}) (J_1^\top \Phi W J_1)^\top &\approx (J_1^\top \Phi W) Cov(u_1) (J_1^\top \Phi W)^\top \\ (J_1^\top \Phi W J_1) (J_1^\top \Phi W J_1)^{-1} (J_1^\top \Phi W J_1) &\approx J_1^\top \Phi W Cov(u_1) (W \Phi J_1) \\ Cov(u_1) &\approx W^{-1} \Phi^{-1}. \end{aligned}$$

Assim, $\widehat{Cov}(u_1) \approx \widehat{W}^{-1} \widehat{\Phi}^{-1}$.

Baseando-se em (4.6) o resíduo em (4.7) também pode ser expresso como

$$\begin{aligned} r^\alpha &= \widehat{\Phi}^{1/2} \widehat{W}^{1/2} (u_1 - J_1 (J_1^\top \widehat{\Phi} \widehat{W} J_1)^{-1} J_1^\top \widehat{\Phi} \widehat{W} u_1) \\ &= (\widehat{\Phi}^{1/2} \widehat{W}^{1/2} - \widehat{\Phi}^{1/2} \widehat{W}^{1/2} J_1 (J_1^\top \widehat{\Phi} \widehat{W} J_1)^{-1} J_1^\top \widehat{\Phi} \widehat{W}) u_1. \end{aligned}$$

Então, considerando que $\widehat{Cov}(u_1) \approx \widehat{W}^{-1} \widehat{\Phi}^{-1}$, $\widehat{W} \approx W$ e $\widehat{\Phi} \approx \Phi$, segue que

$$\begin{aligned} Cov(r^\alpha) &\approx (\Phi^{1/2} W^{1/2} - \widehat{\Phi}^{1/2} W^{1/2} J_1 (J_1^\top \Phi W J_1)^{-1} J_1^\top \Phi W) Cov(u_1) \\ &\times (\Phi^{1/2} W^{1/2} - \widehat{\Phi}^{1/2} W^{1/2} J_1 (J_1^\top \Phi W J_1)^{-1} J_1^\top \Phi W)^\top \\ &\approx (I - \Phi^{1/2} W^{1/2} J_1 (J_1^\top \Phi W J_1)^{-1} J_1^\top \Phi^{1/2} W^{1/2}) \\ &\approx (I - G). \end{aligned}$$

Logo, $\widehat{Cov}(r^\beta) \approx (I - \widehat{G})$, com $G = \widehat{\Phi}^{1/2} \widehat{W}^{1/2} J_1 (J_1^\top \widehat{\Phi} \widehat{W} J_1)^{-1} J_1^\top \widehat{\Phi}^{1/2} \widehat{W}^{1/2}$, em que G é a matriz de projeção de $\Phi^{1/2} W^{1/2} u_1$ em $\Phi^{1/2} W^{1/2} J_1$ e é uma matriz simétrica e idempotente.

Assim, o resíduo ponderado padronizado é dado por

$$r_{pt}^\alpha = \frac{r_t^\alpha}{\sqrt{\widehat{Cov}(r_t^\alpha)}} = \frac{y_t^* - \widehat{\mu}_t^*}{\sqrt{\widehat{v}_t(1 - \widehat{g}_{tt})}}, \quad (4.10)$$

em que g_{tt} é o t -ésimo elemento da matriz diagonal G e v_t é definido em (4.5).

4.2.3 Resíduo combinado

O processo iterativo *Scoring* de Fisher para estimar γ é dado por

$$\gamma^{(m+1)} = \gamma^{(m)} + (J_2^\top D^{*(m)} J_2)^{-1} J_2^\top H^{(m)}(a^{(m)}),$$

em que $J_2^\top = \partial \eta_2 / \partial \gamma$ é uma matriz de ordem $n \times q$, $H = \text{diag}\{1/h'(\phi_1), \dots, 1/h'(\phi_n)\}$ e $a = (a_1, \dots, a_n)^\top$ com

$$a_t = \mu_t(y_t^* - \mu_t^*) + \log(1 - y_t) - \psi((1 - \mu_t)\phi_t) + \psi(\phi_t). \quad (4.11)$$

O processo iterativo de mínimos quadrados reponderados é expresso por

$$\gamma^{(m+1)} = (J_2^\top D^{*(m)} J_2)^{-1} J_2^\top D^{*(m)} u_2^{(m)},$$

em que $u_2^{(m)} = J_2 \gamma^{(m)} + D^{*-1(m)} H^{(m)}(a^{(m)})$. Após convergência do processo, temos que

$$\hat{\gamma} = (J_2^\top \hat{D}^* J_2)^{-1} J_2^\top \hat{D}^* \hat{u}_2,$$

com $\hat{u}_2 = J_2 \hat{\gamma} + \hat{D}^{*-1} \hat{H} \hat{a}$, que pode ser visto como um estimador de mínimos quadrados considerando a regressão linear de $D^{1/2} J_2$ em $D^{1/2} u_2$. O resíduo dessa regressão é dado por

$$\begin{aligned} r^\gamma &= \hat{D}^{*1/2} u_2 - \hat{D}^{*1/2} J_2 \hat{\gamma} \\ &= \hat{D}^{*1/2} (\hat{u}_2 - J_2 \hat{\gamma}) \\ &= \hat{D}^{*1/2} \hat{H} \hat{a}. \end{aligned}$$

Assim, o resíduo da t -ésima observação é

$$r_t^\gamma = \frac{\hat{a}_t}{\sqrt{\hat{\xi}_t}},$$

em que

$$\xi_t = \mu_t^2 \psi'(\mu_t \phi_t) + (1 - \mu_t)^2 \psi'((1 - \mu_t) \phi_t) - \psi'(\phi_t) = \text{Var}(a_t)$$

e a_t está definido em (4.11).

O resíduo combinado é construído com componentes dos resíduos obtidos através dos processos iterativos *Scoring* de Fisher de β e γ , sendo definido por

$$r_t^{\alpha\gamma} = (y_t^* - \hat{\mu}_t^*) + \hat{a}_t. \quad (4.12)$$

A variância de (4.12) pode ser obtida através de

$$\text{Var}(y_t^* + \mu_t^* y_t^* + \log(1 - y_t)) = (1 + \mu_t)^2 \phi'(\mu_t \phi_t) + \mu_t^2 \phi'((1 - \mu_t) \phi_t) - \psi'(\phi_t) = \zeta_t.$$

Desta forma, o resíduo combinado padronizado é definido por

$$r_t^{\alpha\gamma} = \frac{(y_t^* - \hat{\mu}_t^*) + \hat{a}_t}{\sqrt{\hat{\zeta}_t}}, \quad (4.13)$$

em que $\hat{\zeta}_t$ é dado por ζ_t avaliado em $\hat{\mu}_t$ e $\hat{\phi}_t$.

4.3 Resíduos para o modelo de regressão beta não linear com erros nas variáveis

Para os modelos de regressão que apresentam erros de medida é sugerido avaliar o t -ésimo resíduo baseado na estimativa corrigida, ou seja, na estimativa obtida considerando a presença de erros de medida na variável e nos valores observados (Fuller, 1987; Carroll e Spiegelman, 1922; Buonaccorsi, 2010)

Assim, Carrasco (2012), propôs o uso dos resíduos ponderados padronizados, expresso na Equação (4.8), com as estimativas obtidas pelos três métodos avaliados: máxima verossimilhança aproximada (MVa), máxima pseudo-verossimilhança aproximada (MPVa) e calibração da regressão (RC), com o objetivo de avaliar os resíduos para o modelo de regressão beta com erros de medida.

O objetivo desse capítulo é aplicar esta mesma técnica para os modelos de regressão beta não linear com erros de medida para os três resíduos apresentados anteriormente, a saber: resíduo ponderado (4.8), resíduo ponderado padronizado (4.11) e resíduo combinado (4.13), com as estimativas obtidas pelos métodos da máxima verossimilhança aproximada, máxima pseudo-verossimilhança aproximada e calibração da regressão. Realizamos estudos de simulação considerando dois cenários distintos que serão apresentados a seguir.

4.4 Avaliação numérica

Realizou-se um estudo de simulação de Monte Carlo com $R = 10000$ réplicas cujo objetivo era investigar os comportamentos das distribuições dos resíduos apresentados na Seção 4.3 considerando os métodos de estimação apresentados nesta tese.

Com o objetivo de verificar o comportamento dos resíduos estudados em relação às medidas de posição, dispersão, assimetria e curtose, considerou-se apenas para o primeiro cenário uma amostra de tamanho 20 e os resultados foram apresentados em tabelas para cada observação.

Para a construção dos gráficos normais de probabilidade foram geradas diferentes amostras aleatórias de tamanho n igual a 40, 80 e 160 para o primeiro cenário e para o segundo cenário, que apresenta não linearidade nos submodelos da média e da dispersão, foram geradas amostras aleatórias de tamanho n igual a 60, 80 e 160, com o objetivo de minimizar os problemas de convergência. Essas amostras foram tomadas com valores verdadeiros de x , que segue distribuição normal com média μ_x e variância σ_x^2 . As ocorrências da variável dependente y_t foram geradas a partir da distribuição beta com média μ_t e precisão ϕ_t . Depois, geraram-se ocorrências da covariável w_t a partir de sua distribuição condicional dado x_t , ou seja, com uma distribuição normal com média x_t e variância σ_e^2 .

O coeficiente de confiabilidade do modelo foi considerado conhecido e os seguintes valores foram atribuídos $k_x = 0.50$ (erro de medida alto), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.95$ (erro de medida baixo), que representam a $\sigma_e^2 = \sigma_x^2$, $\sigma_e^2 = \sigma_x^2/3$ e $\sigma_e^2 = \sigma_x^2/19$, respectivamente. A variável w_t foi utilizada na construção dos estimadores dos parâmetros de interesse descritos anteriormente.

4.4.1 Cenário 1: Dispersão constante

Considerou-se o modelo de regressão beta não linear com erros de medida sob o enfoque estrutural com a seguinte estrutura:

$$\begin{aligned} g(\mu_t) &= \alpha_0 + z_1^{\alpha_1} + \beta_1 x_1, \\ h(\phi) &= \gamma_0, \\ w_t &= x_t + e_t, \end{aligned}$$

em que $t = 1, 2, \dots, n$. Para cada réplica de Monte Carlo, geraram-se ocorrências da variável sem erro de medida z_{t1} com distribuição uniforme (0.2, 1.2). Os seguintes valores para os parâmetros que permanecem fixos para todas as simulações foram escolhidos: $\alpha_0 = -0.6$, $\alpha_1 = 1.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Como para cada réplica de Monte Carlo são geradas diferentes amostras aleatórias de tamanhos $n = 40, 80$ e 160 , não é possível obter um único intervalo para a média da variável resposta. Neste cenário, considerou-se $Q = 50$ pontos de quadratura para aproximar os métodos de estimação.

Tabela 4.1: Médias, erros-padrão, assimetrias e curtoses empíricas dos resíduos ponderado (r^α), ponderado padronizado (r_p^α) e combinado ($r^{\alpha\gamma}$). Modelo: $\log(\mu_t/[1 - \mu_t]) = \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}$, $t = 1, \dots, 20$, $\phi \approx 150$. Coeficiente de Confiabilidade $k_x = 0.50$. Método: MVa.

t	Média			Erro-padrão			Assimetria			Curtose		
	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$
1	0.149	0.165	0.020	4.975	5.346	4.975	0.023	0.041	-0.293	4.994	4.998	5.118
2	-0.511	-0.572	-0.645	5.021	5.392	5.134	-0.484	-0.557	-0.838	5.591	5.854	6.340
3	0.267	0.298	0.114	5.453	6.034	5.364	0.419	0.399	0.115	4.868	4.968	4.651
4	0.684	0.743	0.547	5.142	5.551	5.052	0.337	0.352	-0.021	5.716	5.840	5.734
5	0.463	0.514	0.269	6.054	7.412	5.853	0.716	0.563	0.395	4.823	4.751	4.267
6	-0.257	-0.286	-0.373	4.825	5.230	4.873	-0.100	-0.030	-0.422	5.832	7.199	5.849
7	0.218	0.239	0.085	5.086	5.468	5.082	-0.037	-0.045	-0.336	4.833	4.883	4.995
8	0.205	0.183	0.017	5.954	7.128	5.859	0.345	0.198	0.028	4.483	4.502	4.343
9	0.358	0.394	0.228	5.039	5.408	5.020	0.123	0.130	-0.167	4.740	4.716	4.848
10	-0.401	-0.449	-0.530	4.997	5.352	5.070	-0.239	-0.269	-0.595	5.854	5.944	6.050
11	0.248	0.285	0.076	5.677	6.436	5.551	0.507	0.461	0.173	4.924	4.964	4.672
12	-0.414	-0.473	-0.548	5.101	5.454	5.226	-0.604	-0.670	-0.956	5.811	6.016	6.512
13	0.476	0.517	0.351	4.942	5.310	4.925	0.022	0.060	-0.319	5.563	5.696	5.739
14	0.327	0.347	0.199	5.065	5.406	5.056	-0.097	-0.148	-0.440	5.995	6.112	6.382
15	0.430	0.478	0.285	5.297	5.742	5.206	0.404	0.418	0.077	5.193	5.220	4.849
16	0.439	0.479	0.311	4.973	5.335	4.940	0.075	0.083	-0.232	5.031	5.099	5.069
17	0.192	0.209	0.061	5.094	5.502	5.073	0.188	0.201	-0.112	5.111	5.243	5.094
18	0.180	0.198	0.042	5.183	5.607	5.194	0.065	0.156	-0.305	5.985	6.679	5.954
19	0.310	0.344	0.177	5.044	5.421	5.040	-0.059	-0.037	-0.363	5.037	5.118	5.147
20	-0.319	-0.366	-0.447	5.017	5.400	5.105	-0.343	-0.373	-0.654	5.276	5.440	5.497

A seguir serão apresentados os resultados dos resíduos ponderado (r^α), ponderado padronizado (r_p^α) e combinado ($r^{\alpha\gamma}$) considerando os métodos de estimação estudados nessa

tese, a saber: máxima verossimilhança aproximada (MVa), máxima pseudo-verossimilhança aproximada (MPVa) e calibração da regressão (RC).

Método da máxima verossimilhança aproximada - MVa

Tabela 4.2: Médias, erros-padrão, assimetrias e curtoses empíricas dos resíduos ponderado (r^α), ponderado padronizado (r_p^α) e combinado ($r^{\alpha\gamma}$). Modelo: $\log(\mu_t/[1 - \mu_t]) = \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}$, $t = 1, \dots, 20$, $\phi \approx 150$. Coeficiente de Confiabilidade $k_x = 0.75$. Método: MVa.

t	Média			Erro-padrão			Assimetria			Curtose		
	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$
1	0.218	0.236	0.123	4.873	5.235	4.859	0.200	0.217	-0.074	5.487	5.486	5.594
2	-0.173	-0.204	-0.259	4.668	5.000	4.709	-0.215	-0.257	-0.457	5.199	5.315	5.389
3	0.179	0.223	0.085	4.877	5.393	4.847	0.254	0.298	0.022	5.077	5.122	5.004
4	0.450	0.489	0.366	4.620	4.974	4.597	0.063	0.044	-0.176	5.238	5.412	5.382
5	0.258	0.297	0.133	5.588	6.918	5.484	0.608	0.488	0.328	5.581	5.578	5.211
6	-0.342	-0.381	-0.429	4.729	5.084	4.786	-0.333	-0.387	-0.629	6.358	6.470	6.654
7	0.006	0.013	-0.087	4.826	5.184	4.834	-0.040	-0.062	-0.292	5.308	5.369	5.504
8	0.266	0.304	0.148	5.441	6.478	5.362	0.589	0.506	0.291	5.885	5.748	5.558
9	0.159	0.173	0.067	4.893	5.287	4.890	-0.067	-0.118	-0.413	7.511	7.990	7.985
10	-0.257	-0.291	-0.344	4.706	5.038	4.742	-0.195	-0.247	-0.444	5.619	5.761	5.721
11	0.265	0.318	0.148	5.357	6.094	5.242	0.786	0.783	0.514	5.691	5.820	5.247
12	-0.341	-0.388	-0.427	4.655	4.938	4.732	-0.582	-0.645	-0.830	5.734	5.799	6.303
13	0.284	0.308	0.193	4.804	5.172	4.787	0.167	0.178	-0.089	5.643	5.705	5.641
14	0.265	0.281	0.180	4.685	4.986	4.666	0.128	0.124	-0.116	5.477	5.488	5.559
15	0.310	0.368	0.210	5.091	5.515	5.040	0.455	0.538	0.164	6.279	6.372	6.056
16	0.343	0.376	0.254	4.688	5.021	4.653	0.321	0.335	0.064	5.429	5.451	5.479
17	0.231	0.249	0.145	4.695	5.058	4.676	0.220	0.223	-0.019	5.383	5.402	5.398
18	0.194	0.211	0.100	4.967	5.367	4.968	-0.071	-0.092	-0.322	5.839	5.984	5.965
19	0.313	0.348	0.216	4.975	5.343	4.967	0.074	0.096	-0.193	5.584	5.587	5.695
20	-0.275	-0.318	-0.364	4.748	5.090	4.812	-0.448	-0.508	-0.687	5.433	5.555	5.696

Tabela 4.3: Médias, erros-padrão, assimetrias e curtoses empíricas dos resíduos ponderado (r^α), ponderado padronizado (r_p^α) e combinado ($r^{\alpha\gamma}$). Modelo: $\log(\mu_t/[1 - \mu_t]) = \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}$, $t = 1, \dots, 20$, $\phi \approx 150$. Coeficiente de Confiabilidade $k_x = 0.95$. Método: MVa.

t	Média			Erro-padrão			Assimetria			Curtose		
	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$
1	0.008	0.007	-0.004	1.810	1.958	1.817	-1.662	-1.857	-1.983	26.525	29.720	29.070
2	-0.013	-0.016	-0.026	1.732	1.838	1.730	0.424	0.359	0.225	11.509	11.308	11.414
3	0.035	0.042	0.020	1.827	2.010	1.827	0.020	0.103	-0.191	12.282	12.142	12.381
4	0.086	0.088	0.075	1.754	1.875	1.746	1.077	1.058	0.893	13.189	12.931	12.792
5	0.134	0.163	0.122	2.026	2.497	1.990	3.600	3.495	3.197	42.346	40.635	38.283
6	-0.167	-0.174	-0.178	1.826	1.933	1.855	-3.048	-2.932	-3.401	33.518	32.485	37.041
7	-0.075	-0.081	-0.087	1.696	1.812	1.699	0.092	0.074	-0.155	14.843	14.333	14.776
8	0.010	0.015	-0.002	1.952	2.318	1.955	-0.189	-0.068	-0.567	27.107	26.642	27.984
9	0.019	0.024	0.008	1.812	1.961	1.812	0.191	0.444	-0.135	23.424	24.963	23.199
10	-0.082	-0.090	-0.093	1.718	1.832	1.725	-0.259	-0.320	-0.444	10.664	10.650	10.865
11	0.054	0.064	0.041	1.765	2.005	1.738	2.425	2.437	1.957	30.981	30.939	25.344
12	-0.126	-0.138	-0.140	1.840	1.950	1.858	-1.001	-1.139	-1.236	14.119	14.635	14.695
13	0.124	0.136	0.111	2.034	2.196	2.025	0.901	1.182	0.408	37.130	38.332	37.757
14	0.086	0.090	0.074	1.723	1.825	1.719	0.451	0.428	0.269	10.799	10.513	10.701
15	-0.004	-0.001	-0.015	1.846	1.988	1.848	-0.027	-0.057	-0.279	17.963	17.611	17.670
16	0.072	0.079	0.060	1.790	1.907	1.779	1.111	1.148	0.838	20.293	20.116	19.087
17	0.034	0.036	0.024	1.680	1.805	1.685	-0.814	-0.803	-0.998	12.607	12.523	13.145
18	-0.005	-0.005	-0.016	1.742	1.880	1.743	-0.180	-0.092	-0.404	13.294	13.686	13.188
19	0.078	0.085	0.066	1.880	2.013	1.863	2.415	2.479	2.039	30.803	31.561	27.849
20	-0.108	-0.118	-0.118	1.562	1.662	1.568	-0.146	-0.172	-0.293	7.582	7.566	7.574

As Tabelas 4.1, 4.2 e 4.3 apresentam as médias (medida de posição), erros-padrão (medida de dispersão), assimetrias (medida de assimetria) e curtoses (medida de curtose) empíricas dos resíduos ponderado (r^α), ponderado padronizado (r_p^α) e combinado ($r^{\alpha\gamma}$)

considerando os coeficientes de confiabilidade $k_x = 0.50$ (erro de medida alto), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.95$ (erro de medida baixo), respectivamente, para 20 observações.

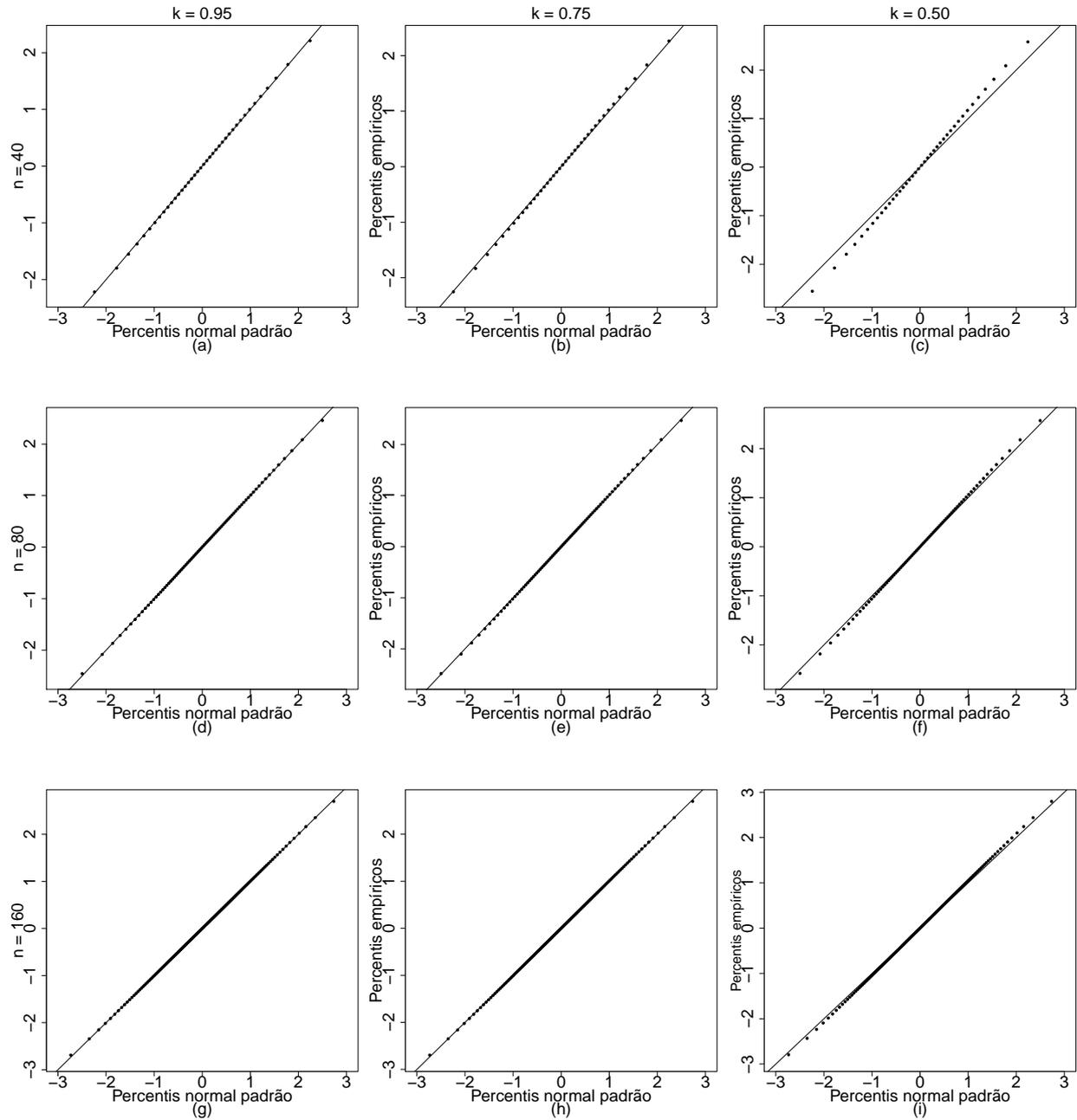


Figura 4.1: Gráficos normais de probabilidade para os **resíduos ponderados** obtidos utilizando o método de máxima verossimilhança aproximada para estimar os parâmetros. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.

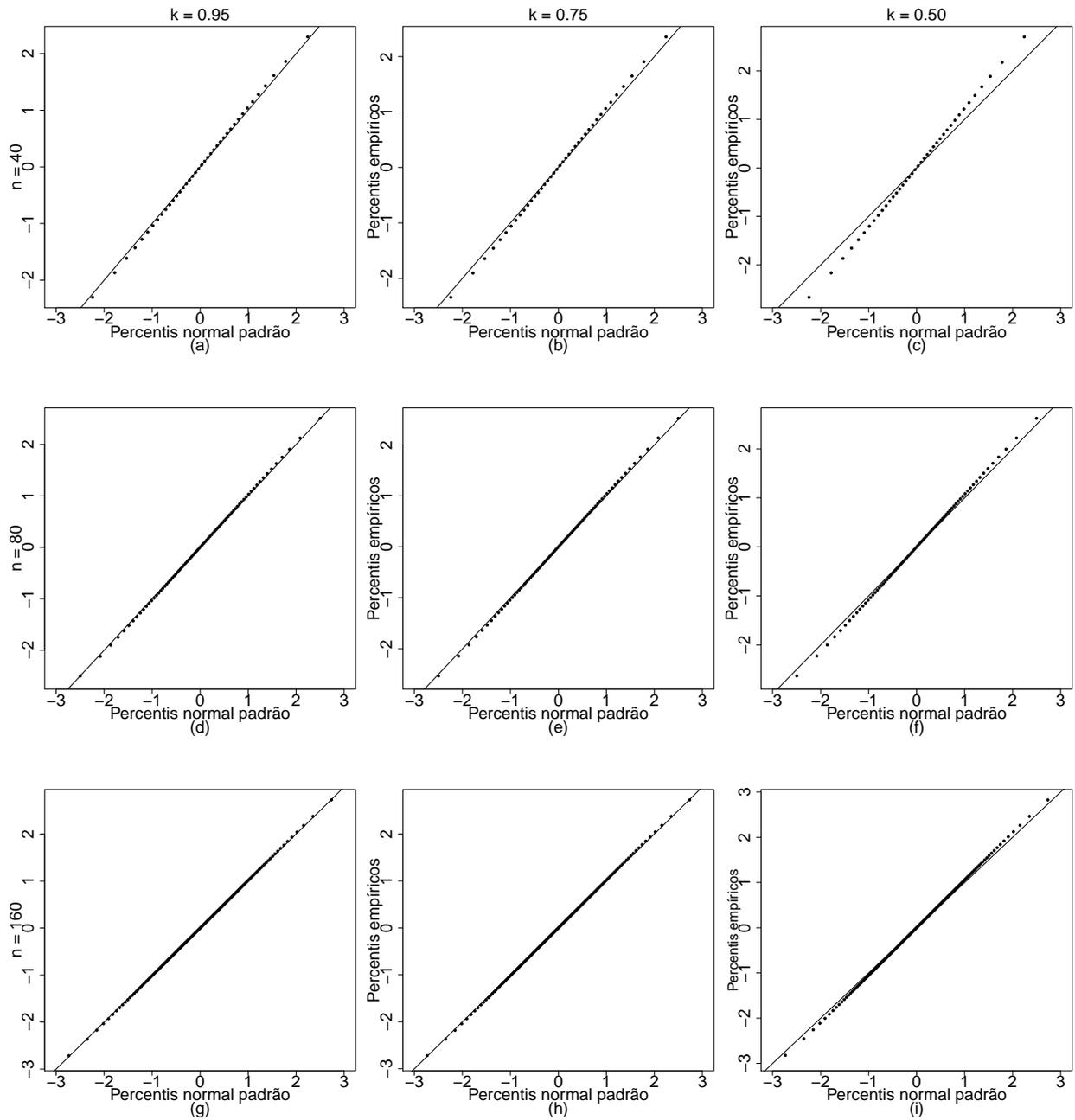


Figura 4.2: Gráficos normais de probabilidade para os resíduos ponderados padronizados obtidos utilizando o método de máxima verossimilhança aproximada para estimar os parâmetros. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.

Para o coeficiente de confiabilidade $k_x = 0.50$, verifica-se que as médias não estão próximas de zero, porém o resíduo combinado, $r^{\alpha\gamma}$, fornece os menores valores médios. Em relação aos erros-padrão, verifica-se uma similaridade entre os mesmos e os resultados encontram-se nos intervalos (4.845, 6.054), (5.230, 7.412) e (4.873, 5.859) para os resíduos ponderado, ponderado padronizado e combinado, respectivamente. Considerando as medidas de assimetria é possível notar que, em geral, o resíduo combinado possui assimetria

negativa enquanto os outros resíduos têm assimetria, na sua maioria, positiva. A curtose está variando nos intervalos (4.483, 5.995), (4.502, 7.199) e (4.343, 6.512) para os resíduos ponderado, ponderado padronizado e combinado, respectivamente, e apresenta, na sua maioria, uma similaridade entre os resíduos (Tabela 4.1).

Para o coeficiente de confiabilidade $k_x = 0.75$, é possível notar novamente que as médias não estão próximas de zero, por exemplo, para a observação 4 tem-se (0.450, 0.489, 0.366) para os resíduos ponderado, ponderado padronizado e combinado, respectivamente. Novamente o residuo combinado fornece os menores valores médios. Aqui, os resíduos avaliados também apresentam erros-padrão similares, sendo que o resíduo ponderado padronizado apresenta os maiores resultados. Considerando, por exemplo, a observação 6, tem-se que $r_p^\alpha = 6.918$. Para as medidas de assimetria observou-se que, novamente, o resíduo combinado possui assimetria negativa enquanto os outros resíduos não apresentam um padrão positivo ou negativo. Os valores mínimos da curtose para os resíduos ponderado, ponderado padronizado e combinado, respectivamente, foram da observação 3 (5.077, 5.122, 5.004) e os valores máximos foram da observação 9 (7.511, 7.990, 7.985). (Tabela 4.2).

Avaliando a Tabela 4.3, que apresenta os resultados para o coeficiente de confiabilidade $k_x = 0.95$, encontramos as médias de todos os resíduos próximas de zero, tendo como melhores resultados o residuo combinado, por exemplo, para a observação 1 $r^{\alpha\gamma} = -0.004$. Para os erros-padrão observou-se resultados bem menores que os apresentados para os coeficientes de confiabilidade menores. Para as medidas de assimetria há uma similaridade entre os resultados sem padrão em relação tipo de assimetria. A curtose apresentou resultados bastante diferentes dos verificados anteriores variando entre (7.582, 42.346), (7.566, 40.635) e (7.574, 38.283), para os resíduos ponderado, padronizado e combinado, respectivamente.

As Figuras 4.1, 4.2 e 4.3 contêm os gráficos normais de probabilidade para os resíduos ponderados, resíduos ponderados padronizados e resíduos combinados, respectivamente, obtidos utilizando o método de máxima verossimilhança aproximada para estimar os parâmetros. Nota-se que com a diminuição do coeficiente de confiabilidade é possível verificar um leve afastamento da reta normal. Vale salientar que o tamanho da amostra está influenciando levemente nos resultados. Além disso, deve-se focar nos resultados do resíduo combinado que, parece ter uma distribuição melhor aproximada pela normal padrão 4.3.

Máxima pseudo-verossimilhança aproximada - MPVa

Para verificar o comportamento dos resíduos obtidos utilizando o método de estimação da máxima pseudo-verossimilhança aproximada foram calculados as médias, os erros-padrão, as assimetrias e as curtoses dos resíduos ponderado (r^α), ponderado padronizado

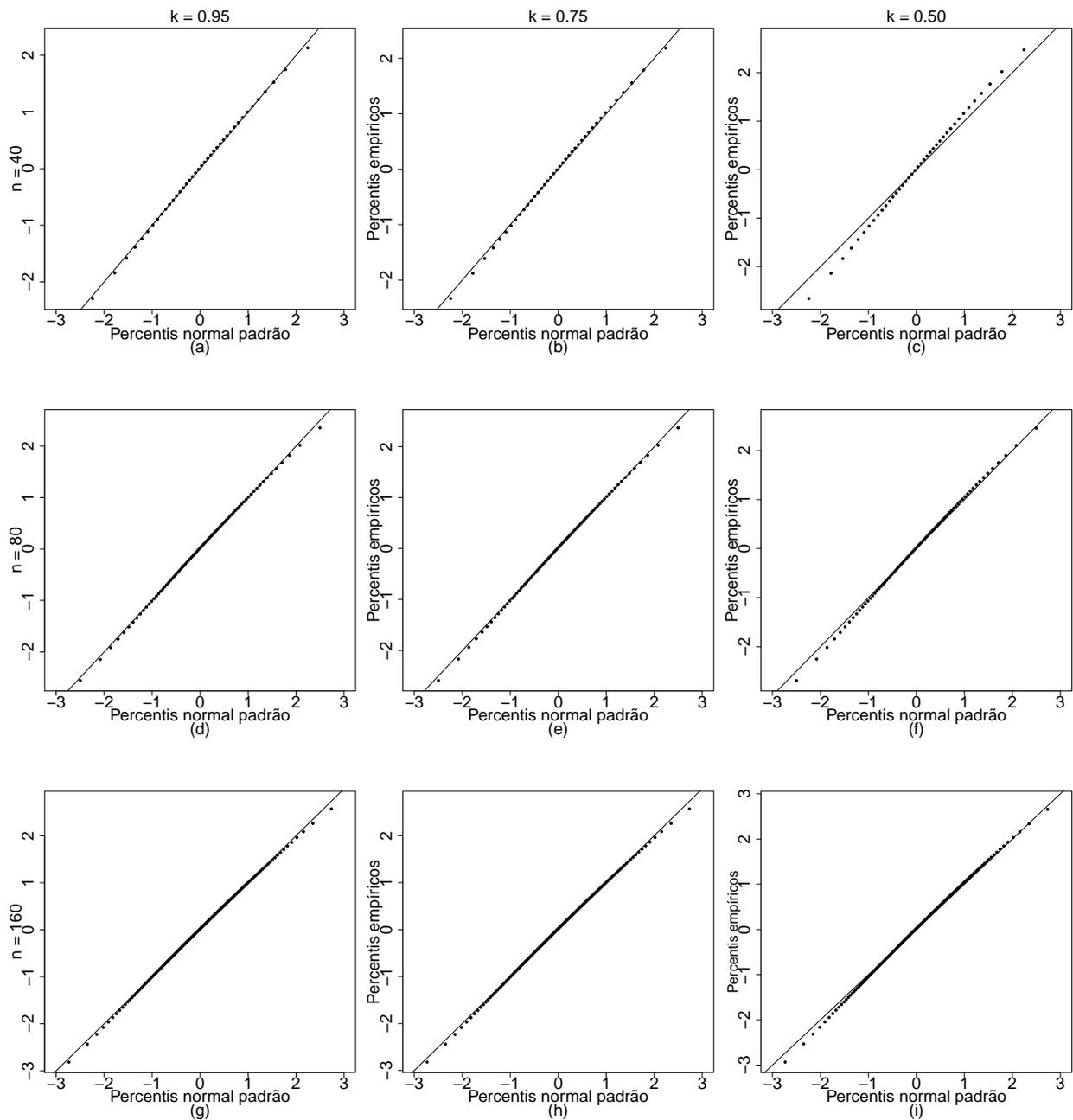


Figura 4.3: Gráficos normais de probabilidade para os **resíduos combinados** obtidos utilizando o método de máxima verossimilhança aproximada para estimar os parâmetros. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.

(r_p^α) e combinado ($r^{\alpha\gamma}$) em relação os coeficientes de confiabilidade $k_x = 0.50$ (erro de medida alto), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.95$ (erro de medida baixo), estando os resultados disponíveis nas Tabelas 4.4, 4.5 e 4.6, respectivamente.

Como observado no método de máxima verossimilhança aproximada, para o método da máxima pseudo-verossimilhança aproximada, considerando o coeficiente de confiabilidade $k_x = 0.50$, verifica-se que as médias não estão próximas de zero, porém o resíduo combinado

Tabela 4.4: Médias, erros-padrão, assimetrias e curtoses empíricas dos resíduos ponderado (r^α), ponderado padronizado (r_p^α) e combinado ($r^{\alpha\gamma}$). Modelo: $\log(\mu_t/[1 - \mu_t]) = \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}$, $t = 1, \dots, 20$, $\phi \approx 150$. Coeficiente de Confiabilidade $k_x = 0.50$. Método: MPVa.

t	Média			Erro-padrão			Assimetria			Curtose		
	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$
1	-0.044	-0.045	-0.197	5.017	5.398	5.067	-0.101	-0.118	-0.595	6.256	6.191	7.078
2	-0.518	-0.588	-0.673	5.023	5.397	5.171	-0.496	-0.375	-0.822	6.725	6.600	6.994
3	0.368	0.426	0.197	5.316	5.852	5.164	0.758	0.723	0.363	5.532	5.464	4.788
4	0.733	0.803	0.579	5.036	5.434	4.909	0.536	0.556	0.213	5.494	5.247	4.948
5	0.583	0.727	0.364	5.896	7.402	5.604	1.128	0.946	0.528	7.672	5.545	4.638
6	-0.425	-0.462	-0.575	4.961	5.437	5.093	-0.316	-0.441	-0.837	6.773	5.712	6.271
7	0.172	0.192	0.021	5.020	5.394	5.008	0.140	0.146	-0.214	5.402	5.436	5.362
8	0.437	0.491	0.226	5.794	6.930	5.595	0.581	0.667	0.266	4.996	5.086	4.428
9	0.222	0.236	0.069	5.070	5.446	5.079	0.052	0.087	-0.264	4.936	4.882	5.073
10	-0.455	-0.512	-0.609	4.955	5.321	5.082	-0.406	-0.366	-0.742	5.311	5.194	5.659
11	0.294	0.354	0.093	5.600	6.358	5.403	0.743	0.727	0.322	5.547	5.408	4.584
12	-0.408	-0.474	-0.568	5.113	5.483	5.341	-1.152	-1.056	-1.596	8.450	8.043	10.831
13	0.462	0.501	0.306	5.036	5.410	4.993	0.263	0.261	-0.146	5.374	5.388	5.363
14	0.364	0.394	0.214	4.957	5.287	4.885	0.308	0.299	-0.047	4.974	4.870	4.825
15	0.417	0.483	0.257	5.108	5.562	4.984	0.597	0.516	0.133	5.845	5.486	5.028
16	0.210	0.228	0.060	4.942	5.298	4.917	0.170	0.181	-0.166	4.814	4.794	4.780
17	0.096	0.104	-0.057	5.113	5.500	5.099	0.285	0.288	-0.114	5.925	5.949	5.916
18	0.106	0.116	-0.053	5.167	5.572	5.205	-0.035	-0.058	-0.453	5.716	5.599	5.974
19	0.272	0.299	0.123	4.948	5.301	4.964	-0.178	-0.185	-0.539	5.029	5.016	5.561
20	-0.601	-0.686	-0.772	5.265	5.687	5.450	-0.656	-0.580	-0.956	5.540	5.368	6.055

Tabela 4.5: Médias, erros-padrão, assimetrias e curtoses empíricas dos resíduos ponderado (r^α), ponderado padronizado (r_p^α) e combinado ($r^{\alpha\gamma}$). Modelo: $\log(\mu_t/[1 - \mu_t]) = \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}$, $t = 1, \dots, 20$, $\phi \approx 150$. Coeficiente de Confiabilidade $k_x = 0.75$. Método: MPVa.

t	Média			Erro-padrão			Assimetria			Curtose		
	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$
1	0.100	0.108	0.009	4.628	4.979	4.614	0.121	0.137	-0.124	5.118	5.117	5.103
2	-0.130	-0.157	-0.214	4.490	4.805	4.515	-0.086	-0.119	-0.332	5.279	5.366	5.401
3	0.223	0.268	0.125	4.838	5.342	4.780	0.570	0.608	0.290	6.219	6.291	5.713
4	0.537	0.581	0.453	4.408	4.749	4.350	0.334	0.372	0.098	4.981	5.165	4.835
5	0.377	0.445	0.254	5.357	6.596	5.226	0.812	0.772	0.534	5.815	5.837	5.237
6	-0.394	-0.435	-0.474	4.363	4.683	4.431	-0.451	-0.494	-0.770	6.732	6.743	7.332
7	-0.031	-0.029	-0.117	4.485	4.811	4.486	0.054	0.034	-0.183	5.110	5.112	5.120
8	0.356	0.411	0.234	5.283	6.324	5.183	0.669	0.584	0.384	5.596	5.441	5.167
9	0.148	0.163	0.057	4.668	5.038	4.657	0.151	0.144	-0.162	6.457	6.563	6.353
10	-0.274	-0.312	-0.354	4.359	4.684	4.401	-0.263	-0.333	-0.504	5.544	5.774	5.690
11	0.384	0.455	0.268	5.140	5.873	5.016	0.811	0.833	0.506	6.056	6.497	5.544
12	-0.274	-0.318	-0.356	4.386	4.672	4.457	-0.544	-0.651	-0.792	5.625	5.947	6.092
13	0.268	0.291	0.184	4.450	4.789	4.437	0.160	0.176	-0.086	5.486	5.644	5.553
14	0.235	0.249	0.153	4.400	4.678	4.379	0.174	0.173	-0.086	5.742	5.718	5.685
15	0.310	0.357	0.217	4.787	5.180	4.726	0.635	0.702	0.318	7.534	7.653	6.938
16	0.271	0.296	0.188	4.423	4.731	4.383	0.364	0.360	0.101	5.900	5.888	5.734
17	0.106	0.113	0.025	4.378	4.704	4.372	0.100	0.090	-0.134	5.202	5.196	5.265
18	0.159	0.173	0.071	4.621	4.985	4.620	0.064	0.051	-0.194	5.680	5.655	5.731
19	0.193	0.213	0.107	4.499	4.826	4.496	0.102	0.111	-0.146	5.398	5.390	5.416
20	-0.265	-0.306	-0.353	4.528	4.861	4.603	-0.562	-0.640	-0.834	5.907	6.139	6.388

fornece, em geral, os menores valores médios. Por exemplo, para a observação 6, tem-se que os resíduos são $(-0.425, -0.462, -0.575)$ para os resíduos ponderado, ponderado padronizado e combinado, respectivamente (Tabela 4.4). Em relação aos erros-padrão, verifica-se uma similaridade entre os mesmos e os resultados encontram-se nos intervalos $(4.942, 5.896)$, $(5.287, 7.402)$ e $(4.885, 5.604)$ para os resíduos ponderado, ponderado padronizado e combinado, respectivamente. Considerando as medidas de assimetria é possível notar que não existe um padrão em relação ao tipo de assimetria. A observação 16 apresenta os menores valores da curtose, ou seja, $(4.814, 4.794, 4.780)$ para os resíduos

Tabela 4.6: Médias, erros-padrão, assimetrias e curtoses empíricas dos resíduos ponderado (r^α), ponderado padronizado (r_p^α) e combinado ($r^{\alpha\gamma}$). Modelo: $\log(\mu_t/[1 - \mu_t]) = \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}$, $t = 1, \dots, 20$, $\phi \approx 150$. Coeficiente de Confiabilidade $k_x = 0.95$. Método: MPVa.

t	Média			Erro-padrão			Assimetria			Curtose		
	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$
1	0.013	0.012	0.001	1.858	2.009	1.861	-0.770	-0.960	-1.141	32.099	34.734	33.483
2	-0.019	-0.021	-0.031	1.739	1.844	1.737	0.173	0.126	-0.016	10.898	10.772	10.931
3	0.017	0.023	0.002	1.847	2.027	1.855	-1.148	-1.021	-1.383	19.615	18.454	20.919
4	0.071	0.073	0.060	1.718	1.838	1.712	0.766	0.747	0.585	12.296	12.007	11.950
5	0.109	0.131	0.097	2.001	2.471	1.969	3.269	3.124	2.832	42.624	40.762	38.215
6	-0.167	-0.174	-0.178	1.810	1.916	1.838	-2.828	-2.720	-3.175	30.830	30.031	34.184
7	-0.074	-0.080	-0.086	1.694	1.810	1.696	0.311	0.285	0.049	16.670	16.078	16.371
8	0.017	0.024	0.004	2.028	2.410	2.025	0.342	0.499	-0.062	28.817	28.705	29.066
9	0.013	0.017	0.001	1.843	1.996	1.843	0.279	0.530	-0.050	23.845	25.428	23.406
10	-0.096	-0.104	-0.107	1.781	1.897	1.792	-0.934	-0.956	-1.149	14.963	14.612	15.717
11	0.005	0.009	-0.007	1.767	2.009	1.765	-0.707	-0.699	-0.945	16.054	16.678	17.019
12	-0.130	-0.143	-0.144	1.892	2.017	1.913	-1.602	-1.925	-1.872	20.592	24.128	22.022
13	0.128	0.140	0.115	2.035	2.196	2.026	0.844	1.101	0.373	34.620	35.631	35.348
14	0.070	0.072	0.058	1.764	1.872	1.763	-0.011	-0.062	-0.207	12.786	12.667	12.972
15	-0.013	-0.010	-0.025	1.876	2.019	1.878	-0.109	-0.143	-0.374	19.062	18.668	18.786
16	0.064	0.070	0.052	1.746	1.858	1.738	0.624	0.620	0.383	16.082	15.455	15.282
17	0.045	0.049	0.036	1.692	1.818	1.695	-0.421	-0.398	-0.610	13.267	13.074	13.615
18	-0.002	-0.002	-0.014	1.720	1.854	1.720	-0.147	-0.046	-0.375	13.302	13.735	13.184
19	0.067	0.073	0.055	1.883	2.017	1.867	2.330	2.387	1.949	30.855	31.600	27.870
20	-0.115	-0.125	-0.125	1.630	1.730	1.639	-1.048	-1.006	-1.233	17.130	16.389	17.908

ponderado, ponderado padronizado e combinado, respectivamente. Já os maiores valores da curtose são encontrados na observação 12 dados por (8.450, 8.043, 10.831). Vale salientar que a curtose apresenta uma amplitude maior do que a verificada para o método de máxima verossimilhança aproximada.

A Tabela 4.5 apresenta as médias, erros-padrão, assimetrias e curtoses dos três resíduos avaliados considerando o coeficiente de confiabilidade $k_x = 0.75$. É possível notar que as médias continuam afastadas do valor zero, por exemplo, os resíduos ponderado, ponderado padronizado e combinado, são respectivamente, (0.148, 0.163, 0.057) para a observação 9. Mais uma vez, o resíduo combinado fornece os menores valores médios. Aqui, os resíduos avaliados também apresentam similares para os erros-padrão. Note que, os valores dos erros-padrão diminuem em comparação com os apresentados para o coeficiente de confiabilidade $k_x = 0.50$ (A Tabela 4.4). As medidas de assimetria não apresenta um padrão. Entretanto, é possível verificar que para os resíduos combinados são, na sua maioria, os valores da assimetria são negativos. A curtose encontra-se nos intervalos (4.981, 7.534), (5.112, 7.653) e (4.835, 7.332) para os resíduos ponderado, ponderado padronizado e combinado, respectivamente, e apresenta, na sua maioria, uma similaridade entre os resíduos.

Em relação ao coeficiente de confiabilidade $k_x = 0.95$ tem-se que as médias de todos os resíduos estão próximas de zero, tendo como resultados mais próximos o resíduo combinado, por exemplo, para a observação 4, os resíduos ponderado, ponderado padronizado e combinado, são (0.071, 0.073, 0.060), respectivamente. É possível notar a influência do erro de medida nos resultados dos erros-padrão, uma vez que, para o erro de medida baixo,

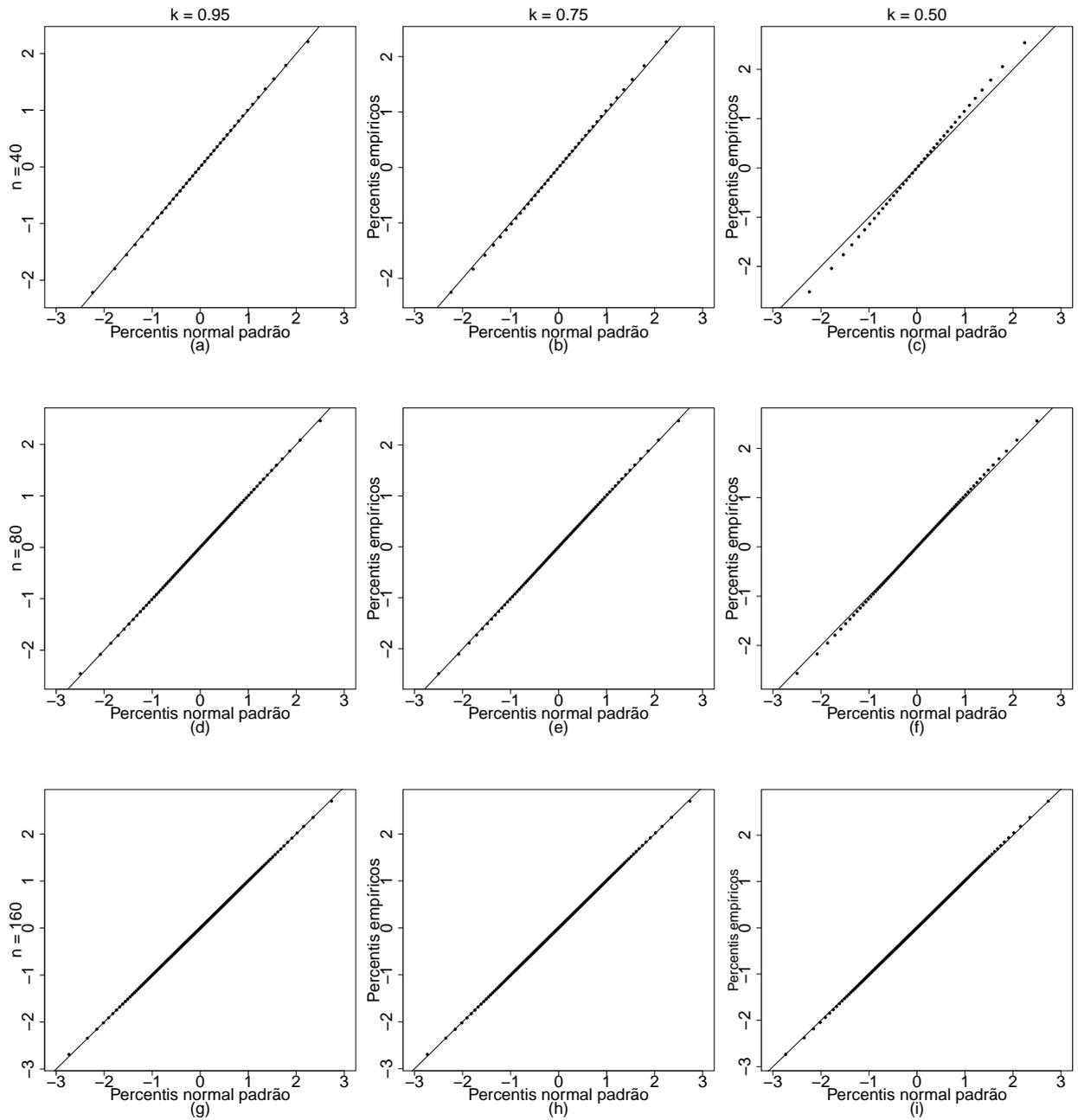


Figura 4.4: Gráficos normais de probabilidade para os **resíduos ponderados** obtidos utilizando o método de máxima pseudo-verossimilhança aproximada para estimar os parâmetros. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.

os valores são bem menores do que os valores apresentados nas Tabelas 4.4 e 4.5. Por exemplo, para a observação 4, os erros-padrão são (1.718, 1.838, 1.712) para os resíduos ponderado, ponderado padronizado e combinado, respectivamente. Já as medidas de assimetria apresentam uma similaridade entre os resultados, em que o resíduo combinado, em geral, apresenta assimetria negativa. A curtose encontra-se nos intervalos (10.898, 42.624), (10.772, 40.762) e (10.931, 38.215) para os resíduos ponderado, padronizado e

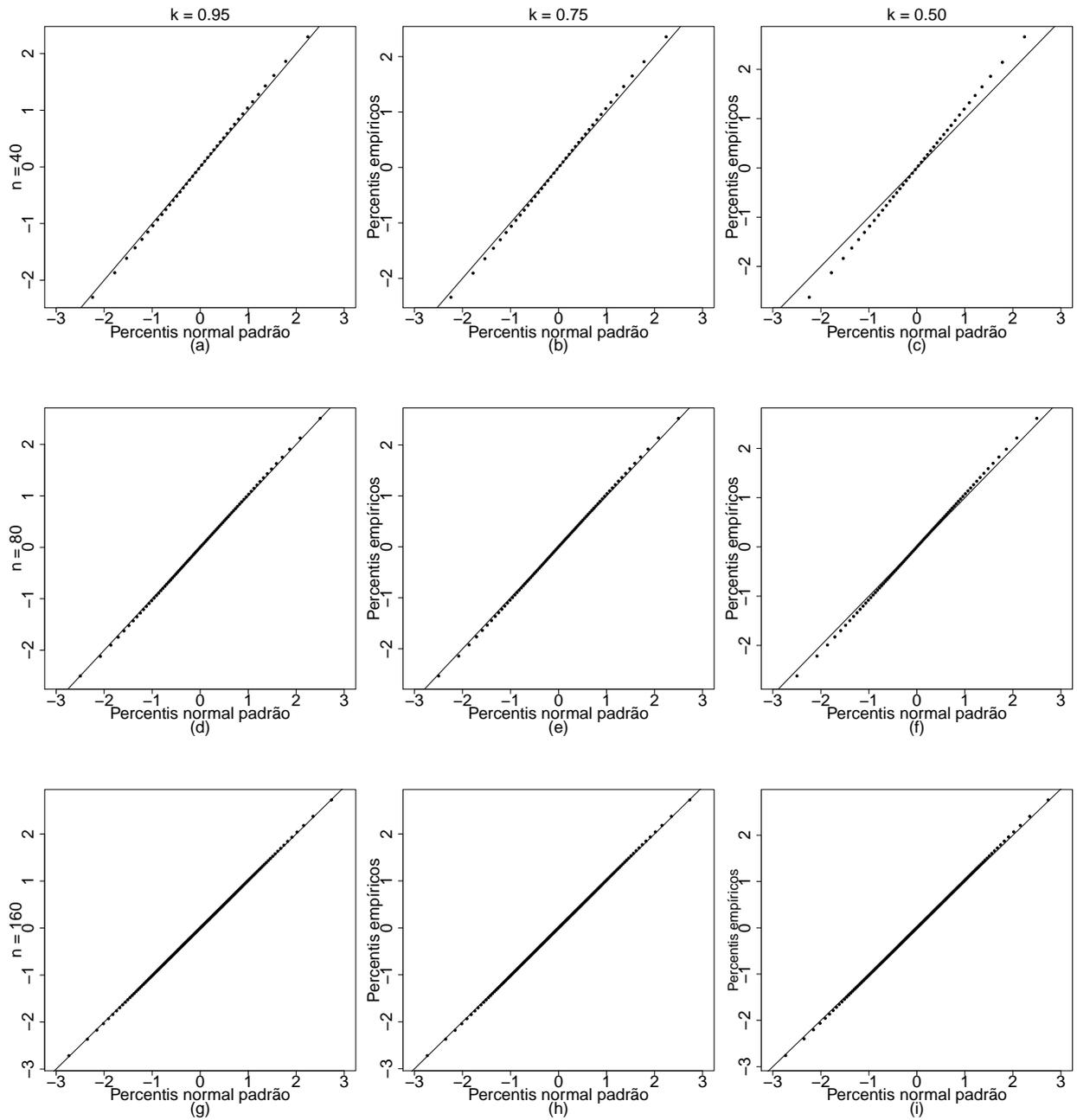


Figura 4.5: Gráficos normais de probabilidade para os **resíduos ponderados padronizados** obtidos utilizando o método de máxima pseudo-verossimilhança aproximada para estimar os parâmetros. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.

combinado, respectivamente (Tabela 4.6).

As Figuras 4.4, 4.5 e 4.6 contêm gráficos normais de probabilidade para os resíduos ponderados, resíduos ponderados padronizados e resíduos combinados, respectivamente. Como nos resultados obtidos pelo método da máxima verossimilhança, é possível notar que com o aumento do erro de medida, ou seja, com a diminuição do coeficiente de confiabilidade há um leve afastamento da reta e um pequeno aumento do intervalo de

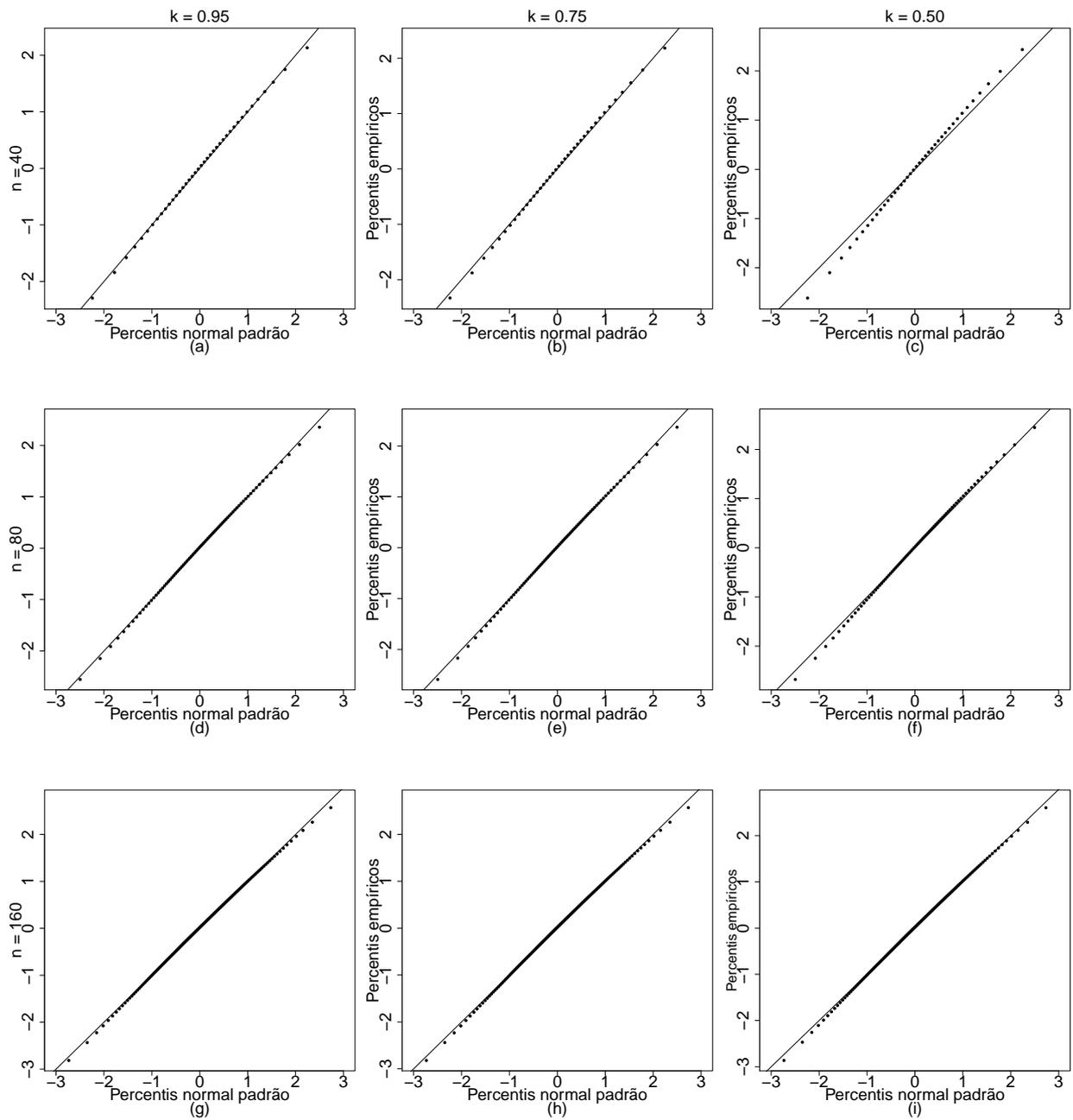


Figura 4.6: Gráficos normais de probabilidade para os **resíduos combinados** obtidos utilizando o método de máxima pseudo-verossimilhança aproximada para estimar os parâmetros. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.

amplitude dos três resíduos estudados. Também observa-se que o tamanho amostral está influenciando levemente nos resultados, ou seja, quanto maior o tamanho da amostra mais próximo os resíduos encontram-se da reta. Além disso, deve-se focar nos resultados do resíduo combinado que parece ter sua distribuição melhor aproximada pela normal padrão.

Calibração da Regressão - RC

As Tabelas 4.7, 4.8 e 4.9 apresentam as médias, erros-padrão, assimetrias e curtoses empíricas dos três resíduos estudados considerando os coeficientes de confiabilidade $k_x = 0.50$ (erro de medida alto), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.95$ (erro de medida baixo), respectivamente.

Tabela 4.7: Médias, erros-padrão, assimetrias e curtoses empíricas dos resíduos ponderado (r^α), ponderado padronizado (r_p^α) e combinado ($r^{\alpha\gamma}$). Modelo: $\log(\mu_t/[1 - \mu_t]) = \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}$, $t = 1, \dots, 20$, $\phi \approx 150$. Coeficiente de Confiabilidade $k_x = 0.50$. Método: RC.

t	Média			Erro-padrão			Assimetria			Curtose		
	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$
1	-0.076	-0.084	-0.058	0.814	0.875	0.836	-0.362	-0.368	-0.760	3.947	3.918	4.730
2	-0.043	-0.049	-0.026	0.822	0.881	0.846	-0.399	-0.406	-0.887	4.646	4.493	6.167
3	0.065	0.076	0.077	0.861	0.948	0.831	0.516	0.555	0.128	3.777	3.815	3.351
4	0.112	0.120	0.131	0.787	0.848	0.765	0.324	0.350	-0.001	3.483	3.505	3.270
5	0.090	0.105	0.094	0.894	1.089	0.849	0.724	0.712	0.367	3.649	3.555	2.972
6	-0.044	-0.052	-0.029	0.833	0.896	0.867	-0.655	-0.648	-1.136	4.837	4.800	6.194
7	0.003	0.003	0.019	0.843	0.907	0.846	-0.054	-0.046	-0.447	3.744	3.740	4.238
8	0.075	0.089	0.080	0.901	1.079	0.858	0.715	0.766	0.348	3.723	3.953	2.969
9	-0.014	-0.014	0.003	0.829	0.890	0.844	-0.296	-0.290	-0.720	4.059	3.981	4.876
10	-0.061	-0.068	-0.046	0.837	0.899	0.866	-0.450	-0.466	-0.859	4.044	4.008	4.829
11	0.053	0.065	0.061	0.896	1.010	0.861	0.632	0.662	0.245	3.882	3.930	3.328
12	0.006	-0.000	0.021	0.842	0.903	0.874	-0.694	-0.696	-1.141	4.463	4.469	5.780
13	0.018	0.020	0.038	0.796	0.856	0.800	-0.137	-0.133	-0.573	4.217	4.159	4.962
14	0.029	0.030	0.049	0.805	0.857	0.794	0.252	0.255	-0.127	3.830	3.819	3.678
15	0.041	0.048	0.056	0.828	0.899	0.810	0.338	0.381	-0.013	3.499	3.531	3.291
16	-0.001	0.000	0.017	0.820	0.877	0.823	-0.019	-0.005	-0.423	3.966	3.924	4.345
17	-0.036	-0.040	-0.018	0.819	0.882	0.838	-0.390	-0.386	-0.841	4.479	4.412	5.692
18	-0.017	-0.018	0.002	0.818	0.881	0.827	-0.112	-0.117	-0.502	3.848	3.819	4.314
19	0.015	0.017	0.034	0.806	0.869	0.809	-0.103	-0.096	-0.439	3.394	3.355	3.702
20	-0.062	-0.071	-0.046	0.832	0.894	0.870	-0.721	-0.746	-1.126	4.358	4.342	5.446

Tabela 4.8: Médias, erros-padrão, assimetrias e curtoses empíricas dos resíduos ponderado (r^α), ponderado padronizado (r_p^α) e combinado ($r^{\alpha\gamma}$). Modelo: $\log(\mu_t/[1 - \mu_t]) = \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}$, $t = 1, \dots, 20$, $\phi \approx 150$. Coeficiente de Confiabilidade $k_x = 0.75$. Método: RC.

t	Média			Erro-padrão			Assimetria			Curtose		
	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$
1	-0.030	-0.034	-0.007	0.681	0.734	0.686	-0.254	-0.280	-0.503	4.056	4.155	4.634
2	-0.016	-0.020	0.007	0.693	0.744	0.700	-0.363	-0.412	-0.712	5.443	5.672	6.909
3	0.046	0.055	0.066	0.727	0.808	0.717	0.353	0.441	0.111	3.940	4.219	3.793
4	0.087	0.095	0.111	0.651	0.703	0.641	0.297	0.324	0.061	4.024	4.196	3.623
5	0.055	0.067	0.069	0.770	0.946	0.749	0.699	0.739	0.414	4.461	4.665	3.717
6	-0.039	-0.044	-0.016	0.671	0.724	0.686	-0.668	-0.673	-1.095	6.518	6.648	8.590
7	-0.023	-0.023	0.002	0.662	0.713	0.663	-0.021	-0.018	-0.215	3.349	3.403	3.383
8	0.044	0.053	0.059	0.768	0.919	0.749	0.612	0.695	0.242	5.269	5.565	4.222
9	0.002	0.003	0.025	0.688	0.744	0.688	-0.051	-0.055	-0.325	4.323	4.473	4.413
10	-0.024	-0.029	-0.001	0.676	0.729	0.682	-0.243	-0.278	-0.506	4.286	4.437	4.608
11	0.037	0.046	0.054	0.763	0.867	0.747	0.586	0.652	0.309	4.514	4.843	4.086
12	0.014	0.011	0.038	0.656	0.702	0.662	-0.393	-0.440	-0.646	4.098	4.341	4.690
13	0.015	0.018	0.040	0.668	0.720	0.670	-0.159	-0.154	-0.439	4.325	4.380	4.937
14	0.017	0.018	0.042	0.644	0.686	0.640	0.132	0.128	-0.104	3.876	3.900	3.804
15	0.018	0.024	0.041	0.682	0.741	0.677	0.146	0.207	-0.089	3.970	4.024	3.981
16	0.012	0.014	0.036	0.677	0.726	0.676	-0.044	-0.037	-0.325	4.370	4.357	4.776
17	-0.014	-0.015	0.010	0.677	0.733	0.681	-0.229	-0.235	-0.515	4.623	4.700	5.449
18	-0.015	-0.016	0.009	0.672	0.726	0.677	-0.261	-0.276	-0.482	3.730	3.784	4.045
19	-0.005	-0.004	0.019	0.686	0.740	0.691	-0.224	-0.222	-0.472	3.904	3.902	4.251
20	-0.015	-0.020	0.008	0.678	0.729	0.684	-0.228	-0.301	-0.513	4.592	4.778	4.817

Para o coeficiente de confiabilidade $k_x = 0.50$ é possível verificar que as médias estão bem próximas de zero. Por exemplo, considerando a observação 19, tem-se (0.015, 0.017,

Tabela 4.9: Médias, erros-padrão, assimetrias e curtoses empíricas dos resíduos ponderado (r^α), ponderado padronizado (r_p^α) e combinado ($r^{\alpha\gamma}$). Modelo: $\log(\mu_t/[1 - \mu_t]) = \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}$, $t = 1, \dots, 20$, $\phi \approx 150$. Coeficiente de Confiabilidade $k_x = 0.95$. Método: RC.

t	Média			Erro-padrão			Assimetria			Curtose		
	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$	r^α	r_p^α	$r^{\alpha\gamma}$
1	0.005	0.005	0.016	0.771	0.829	0.769	0.084	0.077	-0.040	3.216	3.238	3.250
2	-0.006	-0.007	0.004	0.797	0.847	0.797	0.008	0.011	-0.113	3.171	3.185	3.182
3	0.003	0.005	0.012	0.816	0.897	0.817	-0.036	-0.027	-0.143	2.833	2.870	2.855
4	0.030	0.030	0.041	0.760	0.813	0.759	0.012	-0.003	-0.091	2.937	2.946	2.954
5	0.037	0.045	0.047	0.767	0.946	0.763	0.204	0.214	0.085	3.445	3.438	3.403
6	-0.048	-0.050	-0.037	0.759	0.807	0.761	-0.098	-0.076	-0.226	3.392	3.379	3.552
7	-0.028	-0.031	-0.018	0.766	0.821	0.766	0.010	0.007	-0.100	2.894	2.870	2.883
8	0.003	0.004	0.013	0.767	0.915	0.767	0.060	0.045	-0.052	3.084	3.057	3.051
9	0.003	0.004	0.014	0.758	0.818	0.758	-0.058	-0.051	-0.165	2.936	2.971	2.971
10	-0.026	-0.027	-0.015	0.773	0.825	0.773	0.007	0.020	-0.096	2.888	2.925	2.892
11	-0.004	-0.004	0.005	0.785	0.888	0.783	0.150	0.159	0.036	3.115	3.179	3.068
12	-0.017	-0.020	-0.008	0.794	0.839	0.796	-0.175	-0.198	-0.293	3.064	3.115	3.213
13	0.035	0.038	0.046	0.761	0.819	0.760	-0.014	-0.013	-0.130	3.001	3.009	3.053
14	0.014	0.014	0.024	0.779	0.827	0.778	0.029	0.018	-0.080	2.898	2.908	2.869
15	-0.002	-0.000	0.009	0.767	0.829	0.767	0.069	0.087	-0.053	3.264	3.293	3.280
16	0.021	0.024	0.032	0.776	0.828	0.774	0.035	0.032	-0.075	2.969	2.971	2.947
17	0.016	0.017	0.028	0.742	0.799	0.742	-0.133	-0.140	-0.236	2.904	2.922	2.941
18	-0.009	-0.010	0.002	0.768	0.828	0.769	-0.059	-0.054	-0.176	3.137	3.247	3.158
19	0.002	0.003	0.012	0.766	0.825	0.768	-0.110	-0.087	-0.227	3.129	3.203	3.212
20	-0.024	-0.027	-0.014	0.761	0.809	0.763	-0.099	-0.117	-0.210	3.055	3.100	3.147

0.034) para os resíduos ponderado, ponderado padronizado e combinado, respectivamente. Em relação aos erros-padrão, verifica-se uma similaridade entre os mesmos e os resultados encontram-se próximos de 1 (um), variando de 0.787 a 0.901 para o resíduo ponderado, de 0.848 a 1.089 para o resíduo ponderado padronizado e de 0.765 a 0.874 para o resíduo combinado. Considerando as medidas de assimetria verificamos que, em geral, os resíduos possuem assimetria negativa. A curtose encontra-se nos intervalos (3.394, 4.837), (3.355, 4.800) e (2.969, 6.194) para os resíduos ponderado, ponderado padronizado e combinado, respectivamente (Tabela 4.7).

Em relação ao coeficiente de confiabilidade $k_x = 0.75$ as médias também estão bem próximas de zero. Por exemplo, considerando a observação 8, tem-se (0.044, 0.053, 0.059) para os resíduos ponderado, ponderado padronizado e combinado, respectivamente. Os erros-padrão considerando o método da calibração da regressão são bem menores quando comparados com os obtidos pelos métodos da máxima verossimilhança aproximada e máxima pseudo-verossimilhança aproximada. Aqui, os valores dos erros-padrão encontram-se próximos de 1 (um). Considerando as medidas de assimetria verificamos que, em geral, os resíduos possuem assimetria negativa, principalmente o resíduo combinado. A curtose encontra-se nos intervalos (3.349, 6.518), (3.403, 6.648) e (3.383, 8.590) para os resíduos ponderado, ponderado padronizado e combinado, respectivamente (Tabela 4.8).

A Tabela 4.7 apresenta os resultados considerando o coeficiente de confiabilidade $k_x = 0.95$, ou seja, um erro de medida baixo. Novamente, as médias estão bem próximas de zero e os erros-padrão encontram-se próximos de 1 (um). Resultados similares são vistos para a assimetria e a curtose.

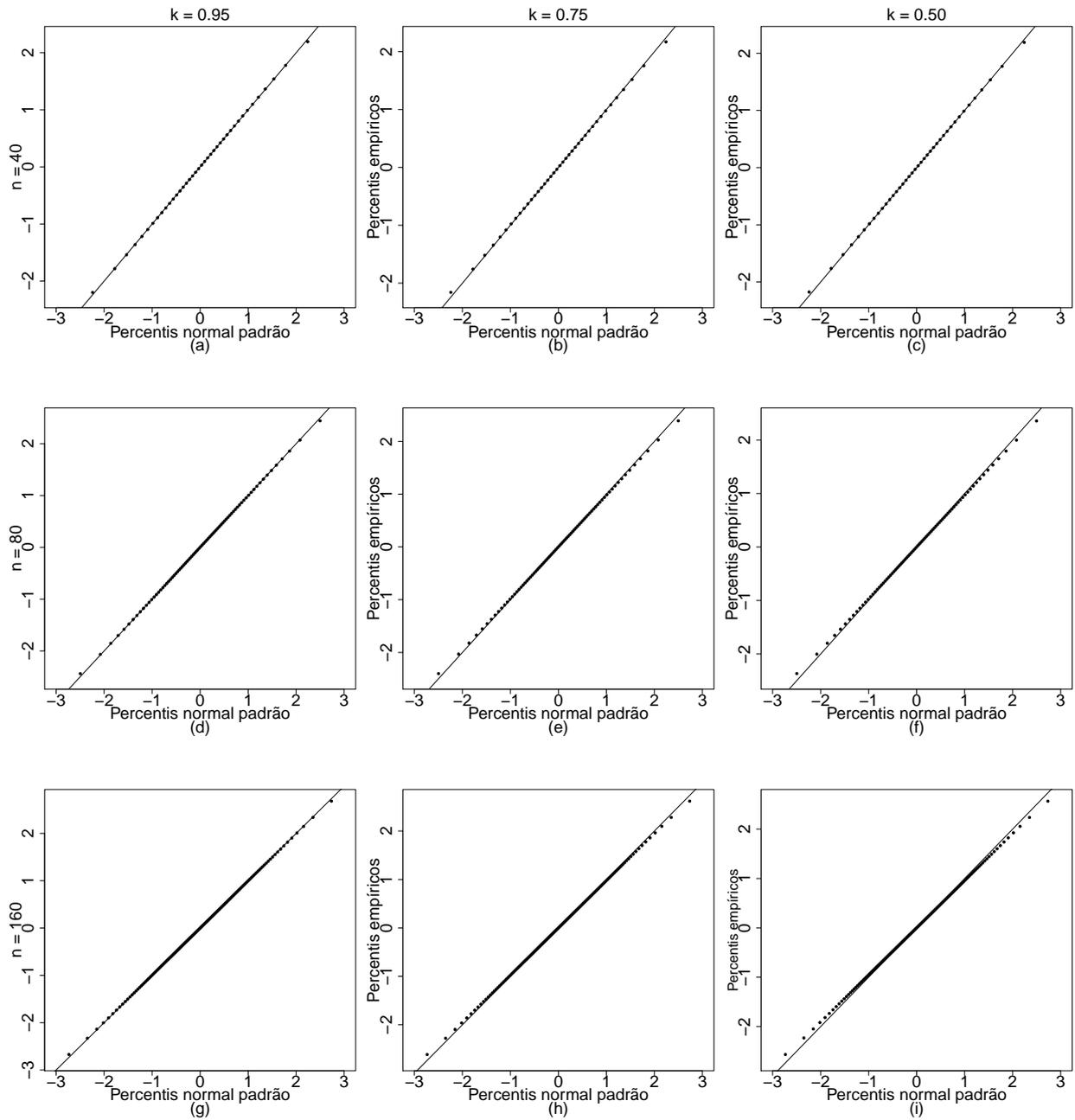


Figura 4.7: Gráficos normais de probabilidade para os **resíduos ponderados** obtidos utilizando o método da calibração da regressão para estimar os parâmetros. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.

As Figuras 4.7, 4.8 e 4.9 apresentam os gráficos normais de probabilidade para os resíduos ponderados, resíduos ponderados padronizados e resíduos combinados, respectivamente, obtidos utilizando o método de calibração da regressão para estimar os parâmetros. Nota-se que com a diminuição do coeficiente de confiabilidade é possível verificar o afastamento da reta. Em relação à amplitude dos três resíduos estudados, é possível verificar que os resíduos estão variando de -2.0 a 2.0 . Vale salientar que o tamanho da amostra

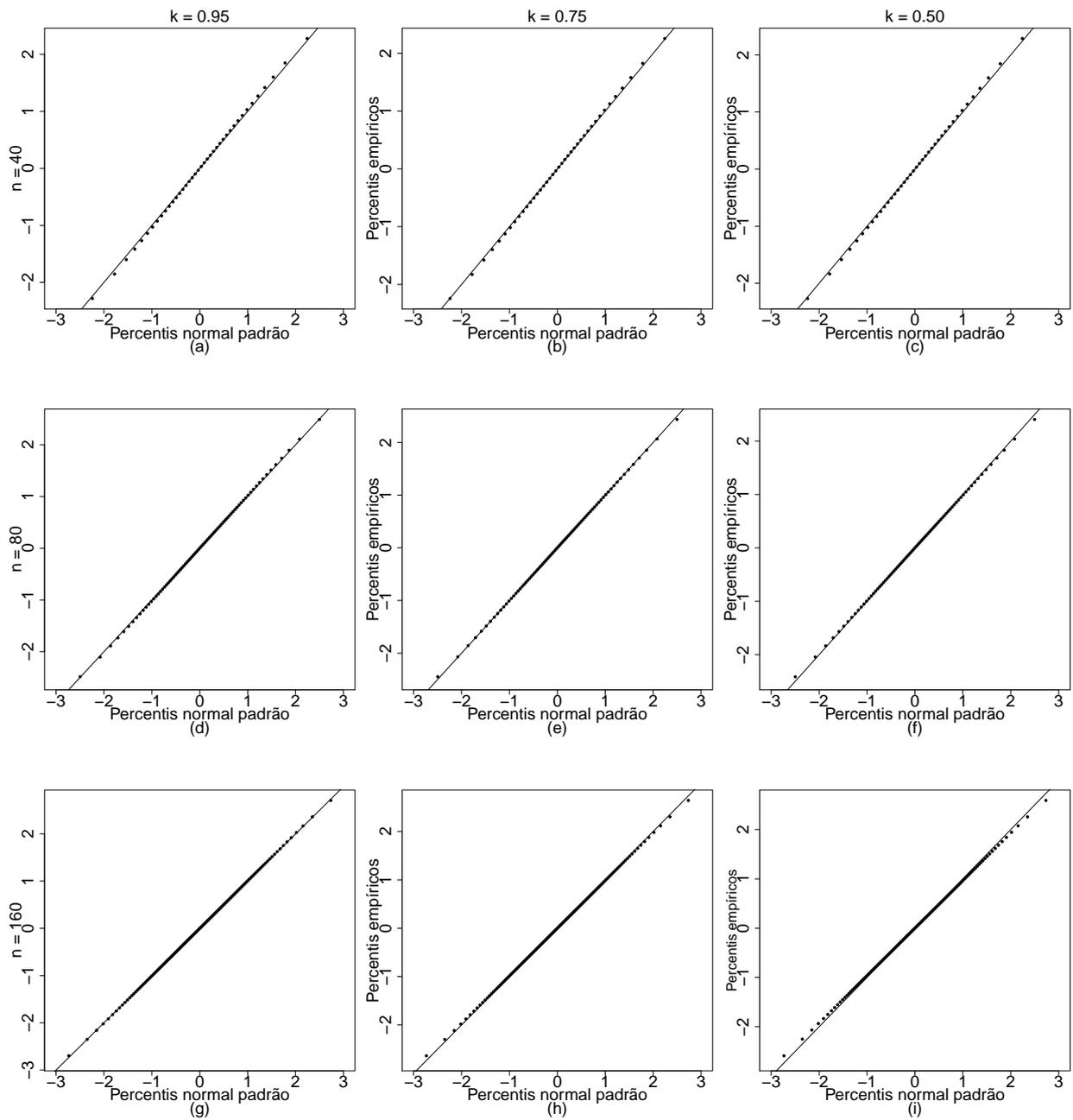


Figura 4.8: Gráficos normais de probabilidade para os **resíduos ponderados padronizados** obtidos utilizando o método da calibração da regressão para estimar os parâmetros. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.

está influenciando fortemente nos resultados. Além disso, deve-se focar nos resultados do resíduo combinado.

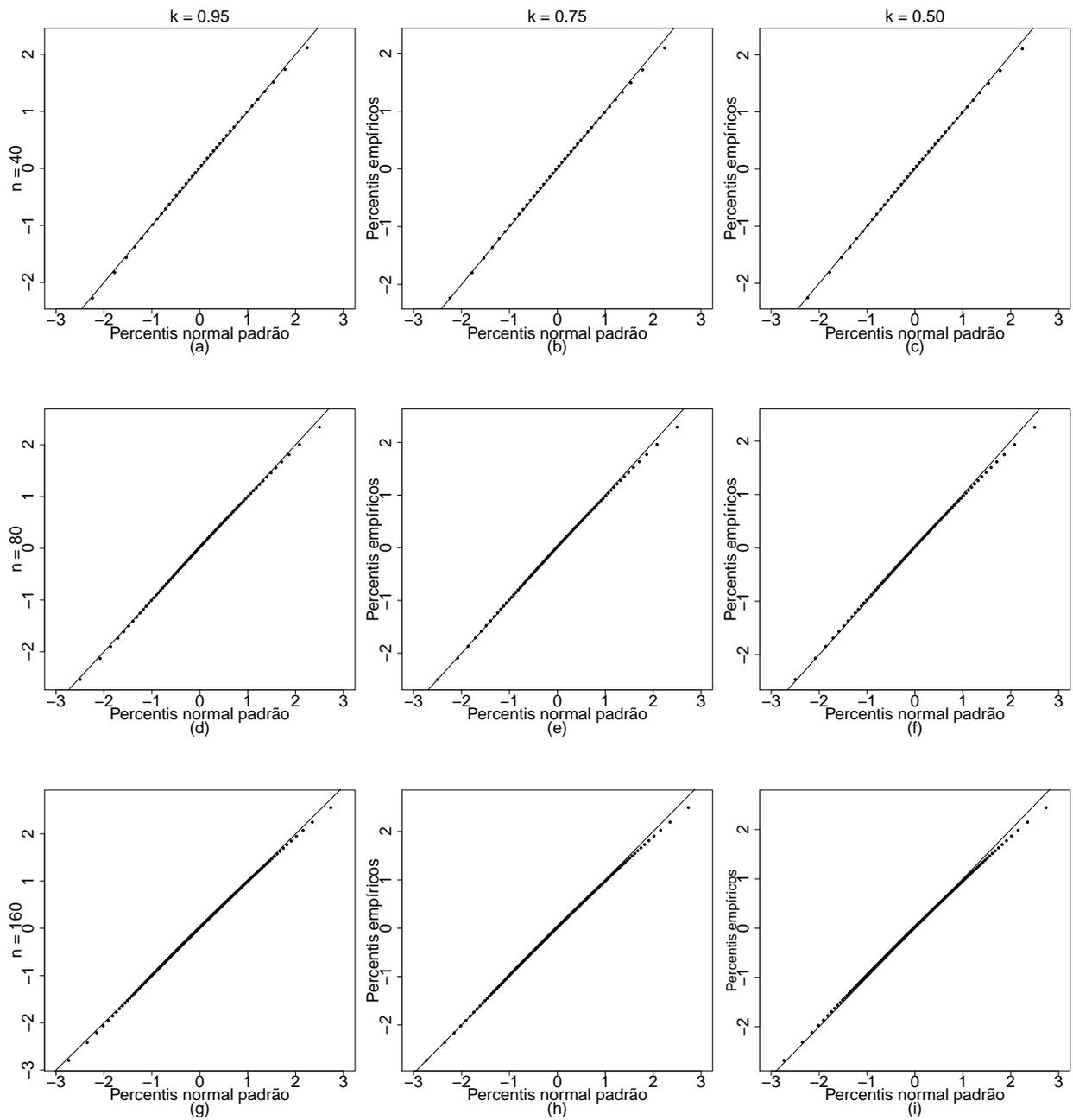


Figura 4.9: Gráficos normais de probabilidade para os **resíduos combinados** obtidos utilizando o método da calibração da regressão para estimar os parâmetros. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.

4.4.2 Cenário 2: Não linearidade na dispersão

Neste cenário admite-se que o modelo de regressão beta não linear com erros de medida com dispersão variável, sob o enfoque estrutural, apresenta a estrutura

$$\begin{aligned}g(\mu_t) &= \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}, \\h(\phi_t) &= \gamma_0 + v_{t1}^{\gamma_1} + \lambda_1 x_{t1}, \\w_t &= x_t + e_t,\end{aligned}$$

em que $t = 1, 2, \dots, n$. Para cada réplica de Monte Carlo, geraram-se ocorrências das variáveis sem erro de medida z_{t1} e v_{t1} , ambas com distribuição uniforme (0.2, 1.2). Os seguintes valores para os parâmetros que permanecem fixos para todas as simulações foram escolhidos: $\alpha_0 = 0.7$, $\alpha_1 = 2.0$, $\beta_1 = -1.5$, $\gamma_0 = 1.5$, $\gamma_1 = 2.0$, $\lambda_1 = 1.3$, $\mu_x = 1.5$ e $\sigma_x^2 = 0.5$. Foram considerados $Q = 50$ pontos de quadratura. A seguir são apresentados os Q-Q *plots* dos resíduos avaliados para os três métodos de estimação abordados nesta tese, considerando o coeficiente de confiabilidade $k_x = 0.75$.

Máxima Verossimilhança Aproximada - MVa

As Figuras 4.10, 4.11 e 4.12 apresentam os gráficos normais de probabilidade para os resíduos ponderados, ponderados padronizados e combinado, respectivamente, considerando o métodos de estimação da máxima verossimilhança aproximada.

Pelos gráficos Q-Q *plots* verificamos que para o modelo de regressão beta não linear com erros de medida e não linearidade na dispersão a distribuição dos resíduos não é bem aproximada pela distribuição normal padrão. Diante disto, os limites usuais $[-2.0, 2.0]$ utilizados para detectar pontos aberrantes não são adequados. Para resolver esse problema Espinheira *et al.* (2017) propõem novos limites baseados na distribuição empírica dos resíduos obtida com base nos envelopes simulados.

Vale salientar que com a diminuição do coeficiente de confiabilidade mais evidente fica que os resíduos não seguem a distribuição normal padrão para o modelo de regressão beta não linear com erros de medida e dispersão não linear.

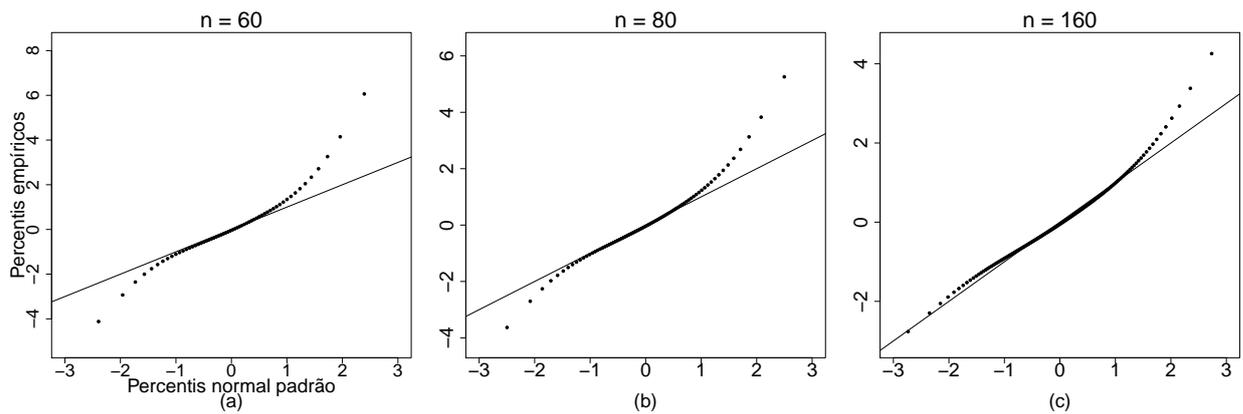


Figura 4.10: Gráficos normais de probabilidade para os **resíduos ponderados** obtidos utilizando o método de máxima verossimilhança aproximada para estimar os parâmetros com coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado). Modelo com dispersão variável não linear.

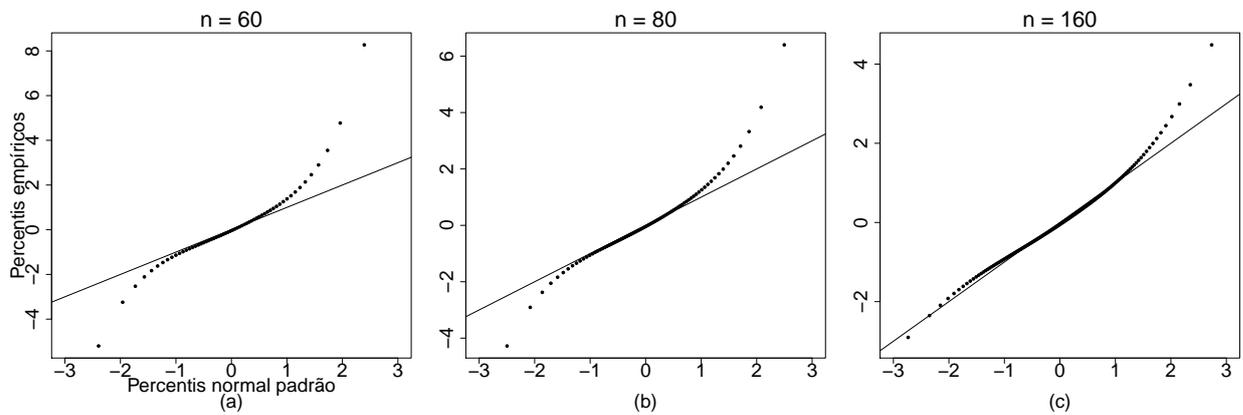


Figura 4.11: Gráficos normais de probabilidade para os **resíduos ponderados padronizados** obtidos utilizando o método de máxima verossimilhança aproximada para estimar os parâmetros com coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado). Modelo com dispersão variável não linear.

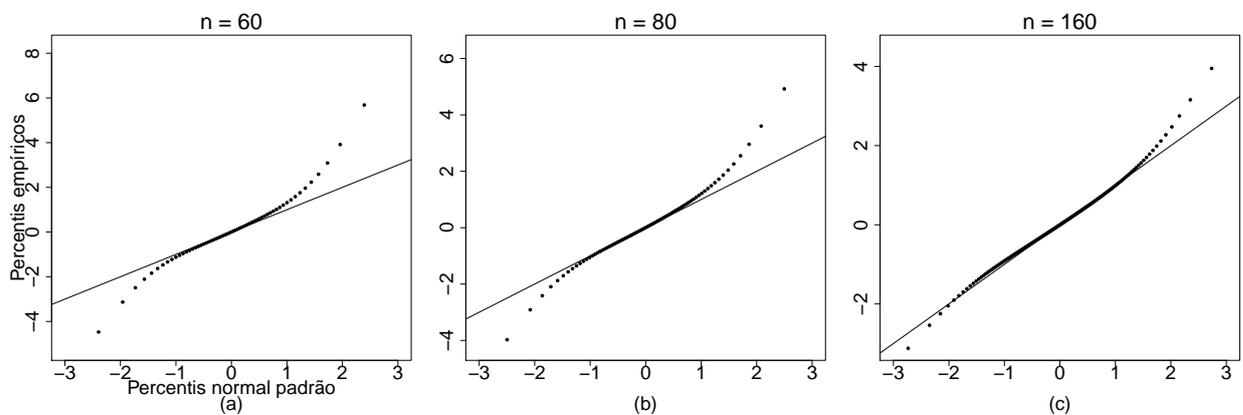


Figura 4.12: Gráficos normais de probabilidade para os **resíduos combinados** obtidos utilizando o método de máxima verossimilhança aproximada para estimar os parâmetros com coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado). Modelo com dispersão variável não linear.

Máxima pseudo-verossimilhança aproximada - MPVa

As Figuras 4.13, 4.14 e 4.15 apresentam os gráficos normais de probabilidade para os resíduos ponderados, ponderados padronizados e combinado, respectivamente, considerando o método de estimação da máxima pseudo-verossimilhança aproximada.

Note que os resultados são similares aos obtidos utilizando o método de máxima verossimilhança aproximada. Ou seja, quanto menor o coeficiente de confiabilidade mais distante a distribuição dos resíduos em relação à distribuição normal padrão.

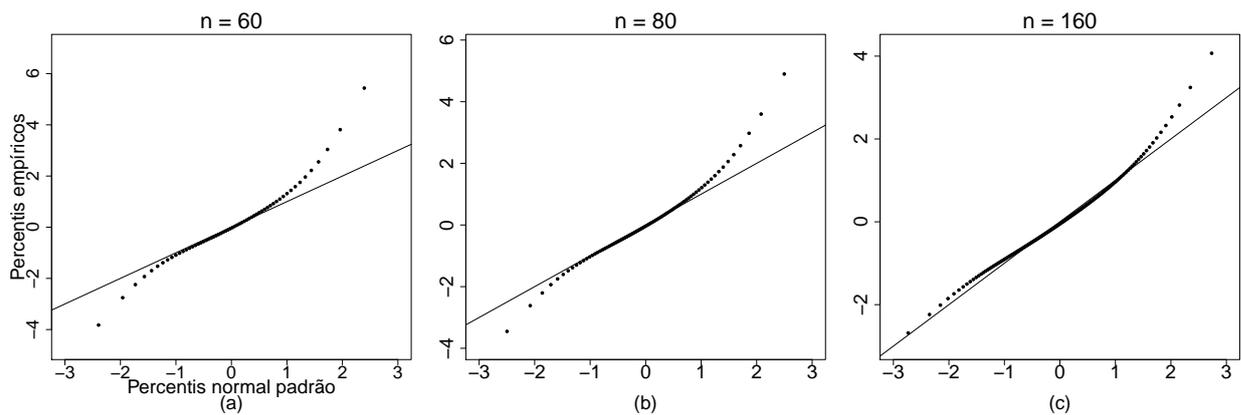


Figura 4.13: Gráficos normais de probabilidade para os **resíduos ponderados** obtidos utilizando o método de máxima pseudo-verossimilhança aproximada para estimar os parâmetros com coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado). Modelo com dispersão variável não linear.

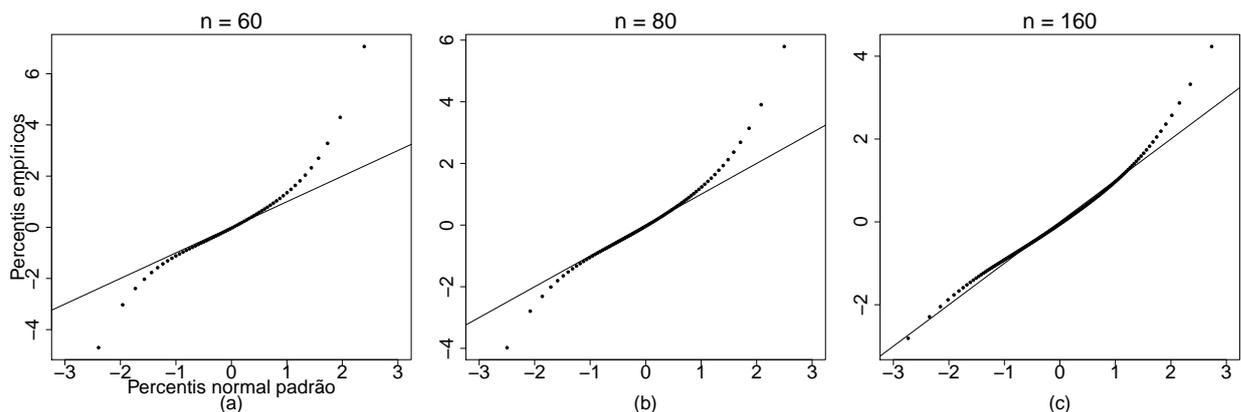


Figura 4.14: Gráficos normais de probabilidade para os **resíduos ponderados padronizados** obtidos utilizando o método de máxima pseudo-verossimilhança aproximada para estimar os parâmetros com coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado). Modelo com dispersão variável não linear.

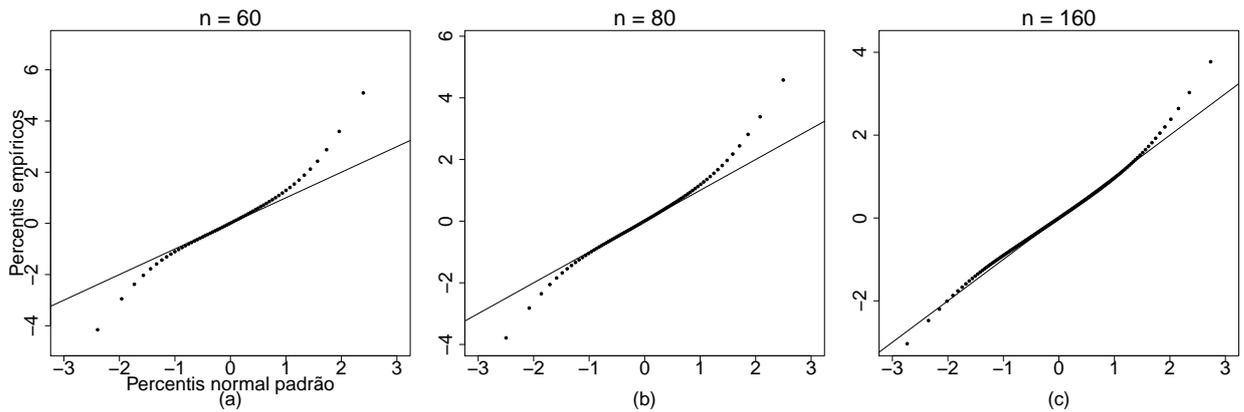


Figura 4.15: Gráficos normais de probabilidade para os **resíduos combinados** obtidos utilizando o método de máxima pseudo-verossimilhança aproximada para estimar os parâmetros com coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado). Modelo com dispersão variável não linear.

Calibração da Regressão - RC

As Figuras 4.16, 4.17 e 4.18 apresentam os gráficos normais de probabilidade para os resíduos ponderados, ponderados padronizados e combinado, respectivamente, considerando os métodos de estimação da calibração da regressão.

Aqui, os resultados são similares aos obtidos utilizando os métodos anteriores. Novamente, é possível concluir que distribuições dos resíduos não é bem aproximada pela normal padrão. Vale salientar também que especialmente para o método da calibração da regressão a estimação do modelo conjuntamente não linear apresenta problemas de estimação.

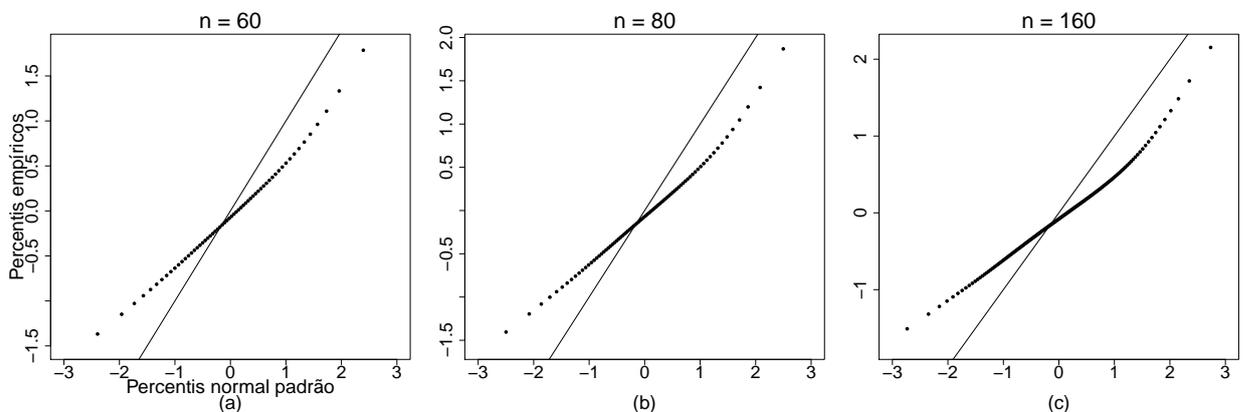


Figura 4.16: Gráficos normais de probabilidade para os **resíduos ponderados** obtidos utilizando o método da calibração da regressão para estimar os parâmetros com coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado). Modelo com dispersão variável não linear.

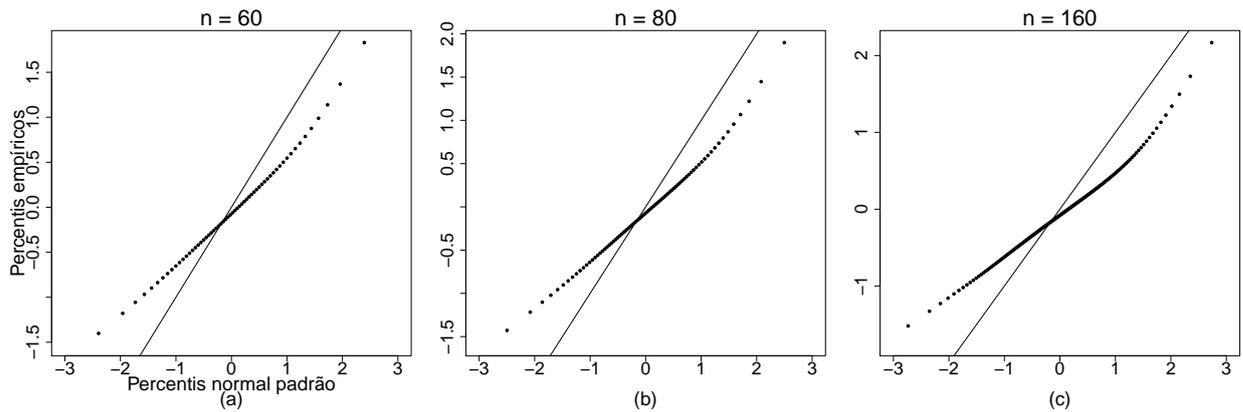


Figura 4.17: Gráficos normais de probabilidade para os **resíduos ponderados padronizados** obtidos utilizando o método da calibração da regressão para estimar os parâmetros com coeficiente de confiabilidade $k_x = 0,75$ (erro de medida moderado). Modelo com dispersão variável não linear.

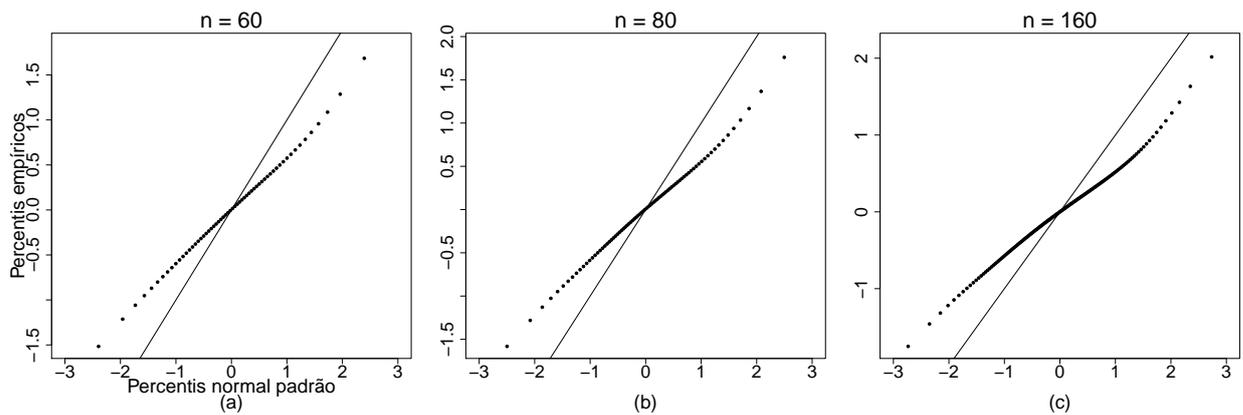


Figura 4.18: Gráficos normais de probabilidade para os **resíduos combinados** obtidos utilizando o método da calibração da regressão para estimar os parâmetros com coeficiente de confiabilidade $k_x = 0,75$ (erro de medida moderado). Modelo com dispersão variável não linear.

4.5 Novos limites para detecção de pontos aberrantes

Considerando os resultados obtidos na seção anterior verificamos que as distribuições dos resíduos podem não ser bem aproximadas pela distribuição normal padrão. E mesmo quando há fortes indícios que os resíduos investigados são aproximadamente normais, há de se ter cautela.

Por exemplo, se para detectar observações aberrantes nos gráficos dos resíduos utilizamos os limites convencionais da distribuição normal padrão $[-2.0, 2.0]$, poderíamos estar cometendo um erro, pois a distribuição empírica dos resíduos tem valores da assimetria e da curtose afastados dos valores da distribuição normal padrão, como visto nos estudos de simulação. Assim, propomos que os limites para a detecção de pontos aberrantes sejam

obtidos a partir dos envelopes simulados. Logo, para a construção dos gráficos de resíduos versus índices das observações estabelecemos como bandas de corte $[\omega_{0.025}; \omega_{0.975}]$, que são os quantis 2.5% e 97.5% da distribuição empírica dos resíduos gerados para a obtenção do envelope. Vale salientar que esta etapa não é computacionalmente custosa, uma vez que os resíduos já foram simulados.

4.6 Conclusões

Neste capítulo utilizamos os resíduos ponderado, ponderado padronizado, propostos por Espinheira *et al.* (2008), e combinado (Espinheira *et al.*, 2017), para realizar uma análise de resíduos dos modelos de regressão beta não linear com erros de medida.

Através de simulações de Monte Carlo os desempenhos dos resíduos propostos na presença de erros de medida foram avaliados e concluímos que as distribuições dos resíduos não podem ser aproximadas pela distribuição normal padrão, portanto não se deve considerar os intervalos $[-2.0, 2.0]$ para identificar pontos aberrantes.

Além disso, ocorreram problemas de convergência associados aos métodos da máxima pseudo-verossimilhança aproximada (MPVa) e calibração da regressão (RC) quando se tem um coeficiente de confiabilidade e tamanho amostral pequenos.

5 ESTATÍSTICAS PRESS, P^2 E MEDIDAS DE QUALIDADE DE AJUSTE PARA O MODELO DE REGRESSÃO BETA NÃO LINEAR COM ERROS DE MEDIDA

Segundo Bozdangan (1987) definir um modelo que apresente o menor número possível de parâmetros possíveis a serem estimados e que explique bem o comportamento da variável resposta, ou seja, buscar um modelo mais parcimonioso é um tópico extremamente importante na análise de dados. Nesta linha, diversos critérios para seleção de modelos são apresentados na literatura dentre os principais é possível citar a seleção *best subset* (Garside, 1965), critério de informação de Akaike (AIC) (Akaike, 1973), C_p de Mallows (Mallows, 1973), validação cruzada (Stone, 1974), critério bayesiano de Schwarz (BIC) (Schwarz, 1978), seleção *stepwise, forward* ou eliminação *backward* (Draper e Smith, 1981), soma de quadrados de resíduos (SQR), e várias funções do SQR tais como R^2 e o R^2 ajustado (Bayer e Cribari-Neto, 2017). Outras referências interessantes são Wolfinger (1993) e Littell *et al.* (2002).

Tais medidas, entretanto, não fornecem visão da qualidade dos valores preditos. Assim, Allen (1971) propôs a soma de quadrados preditos (PRESS), que é utilizada na identificação do modelo com a melhor habilidade para predição (Liu *et al.*, 1999). O cálculo da PRESS consiste em ajustar o modelo, repetidamente, excluindo uma observação de cada vez. Para cada repetição, o modelo é usado para prever a observação que ficou de fora e o modelo mais adequado para predição apresentará o valor da PRESS relativamente pequeno.

Também é possível calcular a estatística P^2 , que é semelhante ao coeficiente de determinação (R^2), através da PRESS. A estatística P^2 é definida como o complementar do quociente entre a PRESS e a soma de quadrados totais sem a t -ésima observação. Esta estatística mensura a qualidade de predição enquanto o R^2 mede a qualidade do ajuste. Silva (2015) propõe a PRESS e a P^2 para o modelo de regressão beta linear e não linear, baseando-se nos resíduos ponderado e combinado.

Neste capítulo propomos o uso das estatísticas PRESS e P^2 (Silva, 2015) e das medidas de qualidade de ajuste (Bayer e Cribari-Neto, 2017) para o modelo de regressão beta não linear com erros de medidas. Para tal, apresentamos o método de construção da estatística PRESS e P^2 para os modelos beta não lineares em geral (Silva, 2015) e as medidas de qualidade de ajuste para o modelo beta.

5.1 Estatísticas PRESS e P^2

Considere y_1, \dots, y_n variáveis aleatórias independentes tais que, para cada $t = 1, \dots, n$, y_t segue distribuição beta com função de densidade dada por

$$f(y_t; \mu_t, \phi_t) = \frac{\Gamma(\phi_t)}{\Gamma(\mu_t \phi_t) \Gamma((1 - \mu_t) \phi_t)} y_t^{\mu_t \phi_t - 1} (1 - y_t)^{(1 - \mu_t) \phi_t - 1}, \quad 0 < y_t < 1. \quad (5.1)$$

No modelo de regressão beta não linear a média e o parâmetro de precisão as seguintes relações funcionais são satisfeitas:

$$\begin{aligned} g(\mu_t) &= \eta_{1t} = f_1(\mathbf{z}_t^\top; \boldsymbol{\alpha}) \quad \text{e} \\ h(\phi_t) &= \eta_{2t} = f_2(\mathbf{v}_t^\top; \boldsymbol{\gamma}), \end{aligned}$$

em que $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^\top \in \mathbb{R}^p$ e $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_{\check{q}})^\top \in \mathbb{R}^{\check{q}}$ são vetores de parâmetros desconhecidos, com $p + \check{q} < n$, η_{1t} e η_{2t} são os preditores não lineares e $\mathbf{z}_t^\top = (z_{t1}, \dots, z_{tp_1})$ e $\mathbf{v}_t^\top = (v_{t1}, \dots, v_{tq_1})$ são vetores de observações fixas e conhecidas com $t = 1, \dots, n$, $p_1 \leq p$ e $q_1 < \check{q}$. As funções de ligação $g(\cdot) : (0, 1) \rightarrow \mathbb{R}$ e $h(\cdot) : (0, \infty) \rightarrow \mathbb{R}$ são estritamente monótonas e duas vezes diferenciáveis.

No modelo de regressão beta não linear, $\hat{\alpha}$ dado na Equação (4.4) pode ser visto como a solução de mínimos quadrados ordinários da regressão

$$\check{y} = \hat{\Phi}^{1/2} \hat{W}^{1/2} \quad \text{sobre} \quad \check{J}_1 = \hat{\Phi}^{1/2} \hat{W}^{1/2} J_1, \quad (5.2)$$

em que $\Phi = \text{diag}(\phi_1, \dots, \phi_n)$, $J_1 = \partial \eta_1 / \partial \boldsymbol{\alpha}$, $T = \text{diag}\{1/g'(\mu_1), \dots, 1/g'(\mu_n)\}$, $u_1 = J_1 \hat{\alpha} + \hat{W}^{-1} \hat{T} (y^* - \hat{\mu}^*)$ e $W = \text{diag}\{w_1, \dots, w_n\}$, com $w_t = \phi_t v_t \{1/[g'(\mu)]^2\}$, sendo

$$v_t = \text{Var}(y_t^*) = \psi'((1 - \mu_t) \phi_t) + \psi'(\mu_t \phi_t). \quad (5.3)$$

Portanto, o erro de predição é $\check{y}_t - \hat{y}_{(t)} = \hat{\phi}_t^{1/2} \hat{w}_t^{1/2} u_{1,t} - \hat{\phi}_t^{1/2} \hat{w}_t^{1/2} j_{1t}^\top \hat{\alpha}_{(t)}$. Usando as ideias propostas por Pregibon (1981) tem-se que $\hat{\alpha}_{(t)} = \hat{\alpha} - \{(J_1^\top \hat{\Phi} \hat{W} J_1)^\top j_{1t} \hat{\phi}_t^{1/2} \hat{w}_t^{1/2} r_t^\alpha\} / (1 - g_{tt})$, em que r_t^α é o resíduo ponderado definido em (4.8) e g_{tt} é o t -ésimo elemento da matriz diagonal $G = \hat{\Phi}^{1/2} \hat{W}^{1/2} J_1 (J_1^\top \hat{\Phi} \hat{W} J_1)^{-1} J_1 \hat{\Phi}^{1/2} \hat{W}^{1/2}$. Assim, segue que $\check{y}_t - \hat{y}_{(t)} = r_t^\alpha / (1 - g_{tt})$.

Para o modelo de regressão beta não linear, a estatística PRESS é definida por

$$PRESS = \sum_{t=1}^n \left(\frac{r_t^\alpha}{1 - g_{tt}} \right)^2. \quad (5.4)$$

Considerando a mesma abordagem para a determinação do coeficiente R^2 , é possível pensar em um coeficiente de predição baseado na PRESS, denotado por estatística P^2 e definido por

$$P^2 = 1 - \frac{PRESS}{SST_{(t)}}, \quad (5.5)$$

em que $SST_{(t)} = \sum_{t=1}^n (y_t - \bar{y}_{(t)})^2$, sendo $\bar{y}_{(t)}$ a média aritmética de y sem a t -ésima observação. Pode-se mostrar que $SST_{(t)} = [n/(n-k)]^2 SST$, sendo k o número de parâmetros estimados do modelo. Para o modelo de regressão beta com dispersão variável temos que $SST = \sum_{t=1}^n (\check{y}_t - \bar{\check{y}})^2$, $\bar{\check{y}}$ é a média aritmética da variável $\check{y}_t = \hat{\phi}_t^{1/2} \hat{w}_t^{1/2} u_{1,t}$, $t = 1, \dots, n$ e $k = p + \check{q}$.

Como sugerido por Cook e Weisberg (1982), versões alternativas da estatística PRESS podem ser definidas, baseadas em diferentes resíduos. Assim, as estatísticas PRESS e P^2 associadas, considerando o resíduo combinado, foram propostas por Silva (2015) e são definidas por

$$PRESS_{\alpha\gamma} = \sum_{t=1}^n \left(\frac{r_t^{\alpha\gamma}}{1 - g_{tt}} \right)^2 \quad \text{e} \quad P_{\alpha\gamma}^2 = 1 - \frac{PRESS_{\alpha\gamma}}{SST_{(t)}}, \quad (5.6)$$

respectivamente. Em (5.6), $r_t^{\alpha\gamma}$ é o resíduo combinado proposto por Espinheira *et al.* (2017) definido em (4.13).

As medidas R^2 e P^2 são distintas, uma vez que o R^2 mede a qualidade do ajuste do modelo e a P^2 e $P_{\alpha\gamma}^2$ medem o poder de predição. Adicionalmente, P^2 e $P_{\alpha\gamma}^2$ não são medidas positivas. De fato, $PRESS/SST_{(t)}$ é uma quantidade positiva, assim a P^2 e a $P_{\alpha\gamma}^2$ assumem valores no intervalo $(-\infty; 1]$. Quanto mais próximo de um, melhor poder preditivo do modelo.

Silva (2015) propôs as versões penalizadas das estatísticas P^2 e $P_{\alpha\gamma}^2$, respectivamente, dadas por

$$P_c^2 = 1 - \frac{(1 - P^2)(n - 1)}{(n - p)} \quad (5.7)$$

e

$$P_{\alpha\gamma c}^2 = 1 - \frac{(1 - P_{\alpha\gamma}^2)(n - 1)}{(n - p)}. \quad (5.8)$$

Para avaliar a bondade do ajuste do modelo consideramos a medida de adequação R^2 baseada na razão de verossimilhanças aqui denotada como pseudo- R_{LR}^2 e dada por

$$R_{LR}^2 = 1 - \left(\frac{L_{null}}{L_{fit}} \right)^{2/n}, \quad (5.9)$$

em que L_{null} é a função de verossimilhança maximizada considerando o modelo sem regressores e L_{fit} é a função de verossimilhança maximizada do modelo ajustado.

Bayer e Cribari-Neto (2017) propuseram o critério de seleção baseado na medida dada por

$$R_{LRc}^2 = 1 - (1 - R_{LR}^2) \frac{n - 1}{n - p}$$

para modelos com dispersão constante e

$$R_{LRc}^2 = 1 - (1 - R_{LR}^2) \left(\frac{n-1}{n - (1+\varpi)kk_1 - (1-\varpi)kk_2} \right)^\delta \quad (5.10)$$

para modelos com dispersão variável, em que $0 \leq \varpi \leq 1$ e $\delta > 0$. kk_1 e kk_2 denotam o número de parâmetros dos submodelos da média e da dispersão, respectivamente.

Consideramos ainda a medida global de ajuste definida por Ferrari e Cribari-Neto (2004) como o quadrado do coeficiente de correlação amostral entre $g(y)$ e $\hat{\eta}_1$ e denotada por $\text{ps-}R^2$. Baseado nela Bayer e Cribari-Neto (2017) propuseram o critério de seleção $\text{ps-}R^2$ corrigido dada por

$$\text{ps-}R_c^2 = 1 - (1 - \text{ps-}R^2) \frac{n-1}{n-p}. \quad (5.11)$$

5.2 Avaliação numérica - Modelo de regressão beta não linear com erros de medida

Realizou-se um estudo de simulação de Monte Carlo com $R = 10000$ réplicas com o objetivo de investigar os comportamentos das estatísticas para o modelo de regressão beta não linear com erros de medidas, considerando os métodos de estimação apresentados nesta tese.

Foram geradas diferentes amostras aleatórias de tamanho n igual a 40, 80 e 160 para os dois primeiros cenários e para o terceiro cenário geraram-se amostras aleatórias de tamanho n igual a 60, 80 e 160, para evitar problemas de convergência. Essas amostras foram tomadas com valores verdadeiros de x_t , que segue distribuição normal com média μ_x e variância σ_x^2 . Ocorrências da variável dependente y_t foram geradas a partir da distribuição beta com média μ_t e precisão ϕ_t . Depois, geraram-se ocorrências da covariável w_t a partir de sua distribuição condicional dado x_t , ou seja, com uma distribuição normal com média x_t e variância σ_e^2 . O coeficiente de confiabilidade do modelo foi considerado conhecido e os seguintes valores foram atribuídos $k_x = 0.50$ (erro de medida alto), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.95$ (erro de medida baixo), que representam a $\sigma_e^2 = \sigma_x^2$, $\sigma_e^2 = \sigma_x^2/3$ e $\sigma_e^2 = \sigma_x^2/19$, respectivamente. A variável w_t foi utilizada para gerar os estimadores dos parâmetros de interesse descritos anteriormente. Em todos os cenários foram considerados $Q = 50$ pontos de quadratura.

5.2.1 Cenário 1: Dispersão constante

Considerou-se o modelo de regressão beta não linear com erros de medida sob o enfoque estrutural com a seguinte estrutura:

$$\begin{aligned} g(\mu_t) &= \alpha_0 + z_1^{\alpha_1} + \beta_1 x_1, \\ h(\phi) &= \gamma_0, \\ w_t &= x_t + e_t, \end{aligned}$$

em que $t = 1, 2, \dots, n$. Para cada réplica de Monte Carlo, geraram-se ocorrências da variável sem erro de medida z_{t1} com distribuição uniforme em $(0.2, 1.2)$. Os seguintes valores para os parâmetros que permanecem fixos para todas as simulações foram escolhidos: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$.

Tabela 5.1: Valores das estatísticas considerando os coeficientes de confiabilidade $k_x = 0.50$, $k_x = 0.75$ e $k_x = 0.95$ e os diferentes métodos de estimação. Modelo com dispersão constante $\phi \approx 150$.

		Cenário 1 : $g(\mu_t) = \alpha_0 + z_1^{\alpha_1} + \beta_1 x_1$ e $h(\phi_t) = \gamma_0$								
		$k_x = 0.95$			$k_x = 0.75$			$k_x = 0.50$		
n		MVa	MPVa	RC	MVa	MPVa	RC	MVa	MPVa	RC
40	P^2	0.8331	0.8331	0.8331	0.8292	0.8292	0.8292	0.8101	0.8108	0.8091
	P_c^2	0.8191	0.8191	0.8191	0.8150	0.81500	0.8150	0.7942	0.7950	0.7931
	$P_{\alpha\gamma}^2$	0.8331	0.8331	0.8331	0.8293	0.8294	0.8294	0.8103	0.8111	0.8095
	$P_{\alpha\gamma c}^2$	0.8192	0.8192	0.8192	0.8151	0.8151	0.8152	0.7945	0.7953	0.7936
	ps- R^2	0.8306	0.8306	0.8306	0.8279	0.8279	0.8279	0.8178	0.8180	0.8180
	ps- R_c^2	0.8165	0.8165	0.8165	0.8136	0.8136	0.8136	0.8026	0.8028	0.8028
	R_{LR}^2	0.8371	0.8371	0.9902	0.8292	0.8292	0.9912	0.8188	0.8187	0.9944
	R_{LRc}^2	0.8235	0.8235	0.9894	0.8149	0.8149	0.9912	0.8037	0.8036	0.9939
80	P^2	0.8104	0.8104	0.8104	0.8083	0.8083	0.8083	0.8013	0.8016	0.8016
	P_c^2	0.8029	0.8029	0.8029	0.8008	0.8008	0.8008	0.7934	0.7938	0.7938
	$P_{\alpha\gamma}^2$	0.8105	0.8105	0.8105	0.8084	0.8084	0.8084	0.8013	0.8016	0.8017
	$P_{\alpha\gamma c}^2$	0.8030	0.8030	0.8030	0.8008	0.8008	0.8008	0.7934	0.7938	0.7938
	ps- R^2	0.8112	0.8112	0.8112	0.8095	0.8095	0.8095	0.8049	0.8049	0.8049
	ps- R_c^2	0.8037	0.8037	0.8037	0.8020	0.8020	0.8020	0.7972	0.7972	0.7972
	R_{LR}^2	0.8140	0.8140	0.9893	0.8049	0.8049	0.9911	0.7941	0.7941	0.9938
	R_{LRc}^2	0.8067	0.8067	0.9888	0.7972	0.7974	0.9908	0.7859	0.7859	0.9935
160	P^2	0.8229	0.8229	0.8229	0.8220	0.8220	0.8220	0.8175	0.8199	0.8199
	P_c^2	0.8195	0.8195	0.8195	0.8186	0.8186	0.8186	0.8140	0.8164	0.8164
	$P_{\alpha\gamma}^2$	0.8229	0.8229	0.8229	0.8220	0.8220	0.8221	0.8177	0.8199	0.8200
	$P_{\alpha\gamma c}^2$	0.8195	0.8195	0.8195	0.8186	0.8186	0.8186	0.8142	0.8165	0.8165
	ps- R^2	0.8251	0.8251	0.8251	0.8244	0.8244	0.8244	0.8222	0.8228	0.8228
	ps- R_c^2	0.8217	0.8217	0.8217	0.8211	0.8211	0.8211	0.8188	0.8194	0.8194
	R_{LR}^2	0.8261	0.8261	0.9901	0.8178	0.8178	0.9919	0.8075	0.8076	0.9943
	R_{LRc}^2	0.8228	0.8228	0.9899	0.8143	0.8143	0.9917	0.8038	0.8039	0.9942

A Tabela 5.1 apresenta os valores médios das estatísticas apresentadas na seção anterior considerando os coeficientes de confiabilidade $k_x = 0.50$, $k_x = 0.75$ e $k_x = 0.95$ e os métodos da máxima verossimilhança aproximada (MVa), máxima pseudo-verossimilhança aproximada (MPVa) e calibração da regressão (RC) para estimação dos parâmetros

o modelo com dispersão constante. Através dos resultados obtidos verifica-se que as estatísticas estão bem semelhantes para todos os métodos considerados. Também é notável uma sucinta diminuição nas estatísticas P^2 e suas versões corrigidas, com a diminuição do coeficiente de confiabilidade, indicando que à medida que o erro de medida aumenta fica mais difícil fazer predição no modelo. Vale salientar que as medidas R_{LR}^2 e R_{LRc}^2 são bem próximas de um, indicando que os métodos estão ajustando bem os modelos. Além disso, devemos salientar que as duas medidas de qualidade de ajuste ps- R^2 e R_{LR}^2 apresentam comportamentos equivalentes.

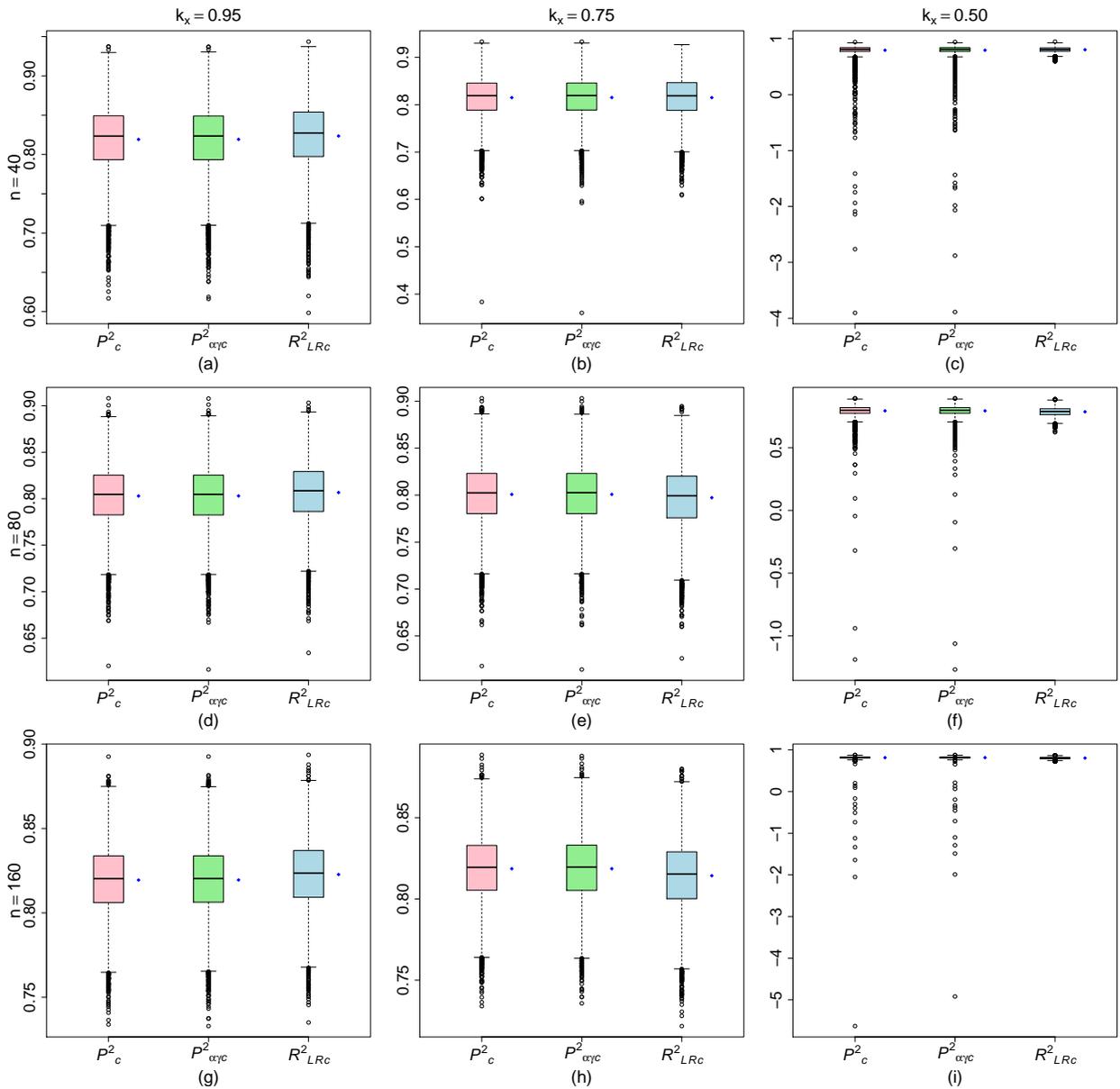


Figura 5.1: *Boxplots* das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$) e R_{LR}^2 corrigida (R_{LRc}^2) obtidas pelo método da **máxima verossimilhança aproximada**. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.

Com o objetivo de fazer uma análise mais detalhada do comportamento das estatísticas de medidas de predição e de qualidade de ajuste propostas para o modelo de regressão beta não linear com erros de medida foram construídos *boxplots* considerando os métodos da máxima verossimilhança aproximada (MVa), pseudo-verossimilhança aproximada (MPVa) e calibração da regressão (RC) e os mesmos são mostrados nas Figuras 5.1, 5.2 e 5.3, respectivamente. Foram avaliadas as medidas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando resíduo combinado ($P_{\alpha\gamma c}^2$) e R_{LR}^2 corrigida (R_{LRc}^2).

As medidas P^2 e $P_{\alpha\gamma}^2$ indicam que à medida que o coeficiente de confiabilidade diminui a qualidade de predição do modelo é substancialmente afetada negativamente mesmo quando o modelo está bem especificado. Este resultado corrobora com o desempenho dos estimadores do parâmetro associado a não linearidade o qual é prejudicado tanto quanto ao viés quanto às taxas de coberturas dos intervalos de confiança.

Note que o mesmo não ocorre com a qualidade do ajuste do modelo avaliada com base na R_{LRc}^2 . À medida que o erro de medida aumenta maior o sentido em considerar um modelo que considere este fato. Assim, à medida que k_x (coeficiente de confiabilidade) diminui, mais os valores da estatística R_{LRc}^2 tendem a se concentrar próximo de um.

Comparando os três métodos nota-se que o método da máxima pseudo-verossimilhança aproximada é o que apresenta o melhor desempenho quanto à predição visto que os *boxplots* das medidas P_c^2 e $P_{\alpha\gamma c}^2$ revelam que os valores das estatísticas tendem a se concentrar próximas de um à medida que o tamanho da amostra aumenta mesmo quando $k_x = 0.50$ (Figura 5.2). Do ponto de vista de qualidade de ajuste é o método de calibração da regressão que apresenta o melhor desempenho, visto que independente do tamanho da amostra os valores da estatística R_{LRc}^2 estão muito concentradas no valor máximo, igual a um (Figura 5.3). O método da calibração da regressão apresentou melhores resultados quando comparados aos outros métodos. Este fato pode estar associado à questão deste método apresentar um bom desempenho quanto ao processo de estimação do parâmetro de não linearidade.

5.2.2 Cenário 2: Dispersão variável

Para avaliação das medidas de predição e qualidade de ajuste em modelos com dispersão variável consideramos dois casos: (i) modelo sem erro de medida no preditor da dispersão e (ii) modelo com erro de medida no preditor da dispersão.

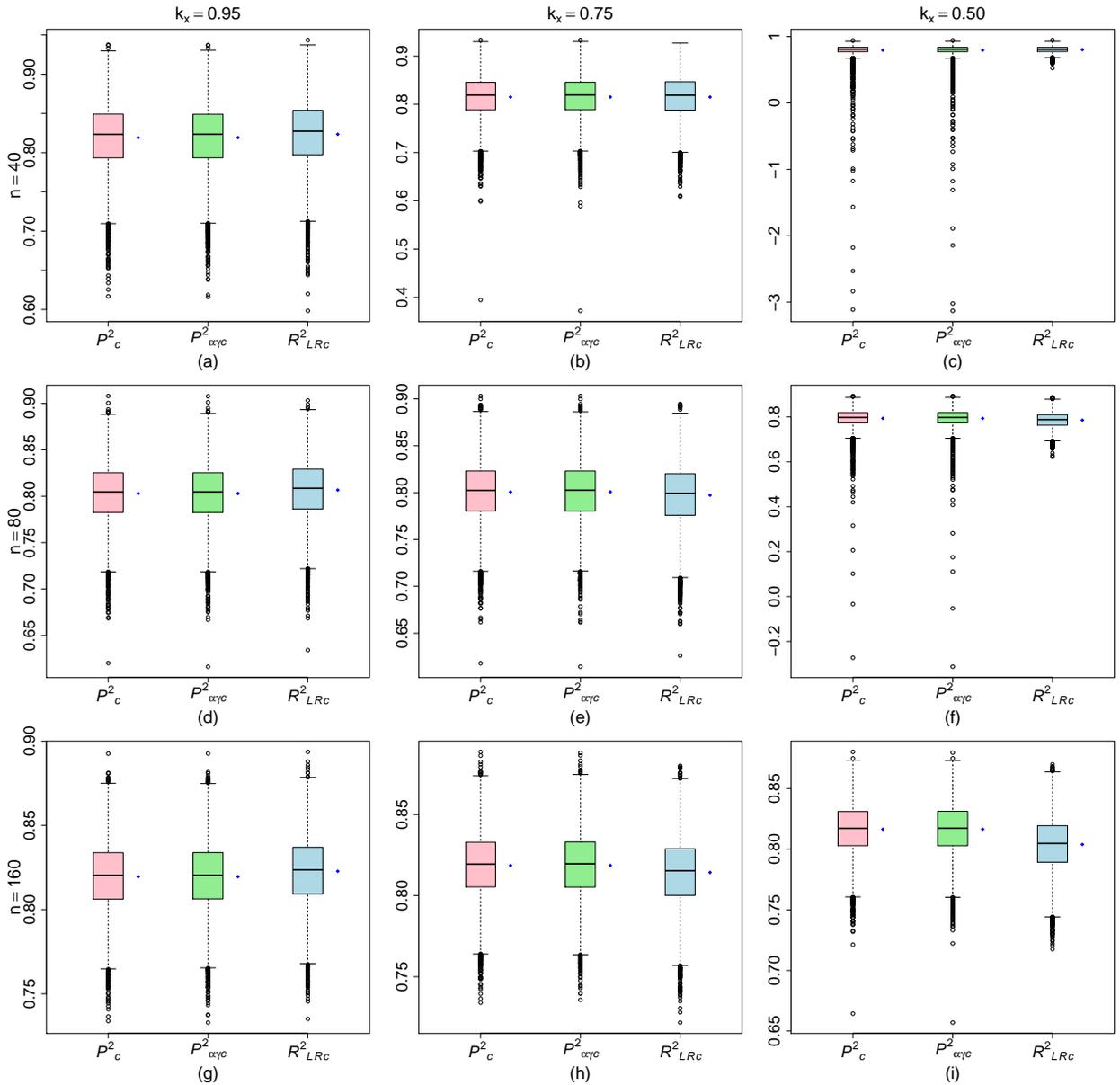


Figura 5.2: *Boxplots* das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$) e R_{LRc}^2 corrigida (R_{LRc}^2) obtidas pelo método da **máxima pseudo-verossimilhança aproximada**. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.

Modelo sem erro de medida no preditor da dispersão

Neste cenário admite-se que o modelo de regressão beta não linear com erros de medida e dispersão variável, sob o enfoque estrutural, apresenta a seguinte estrutura:

$$g(\mu_t) = \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1},$$

$$h(\phi_t) = \gamma_0 + \gamma_1 z_{t1},$$

$$w_t = x_t + e_t,$$

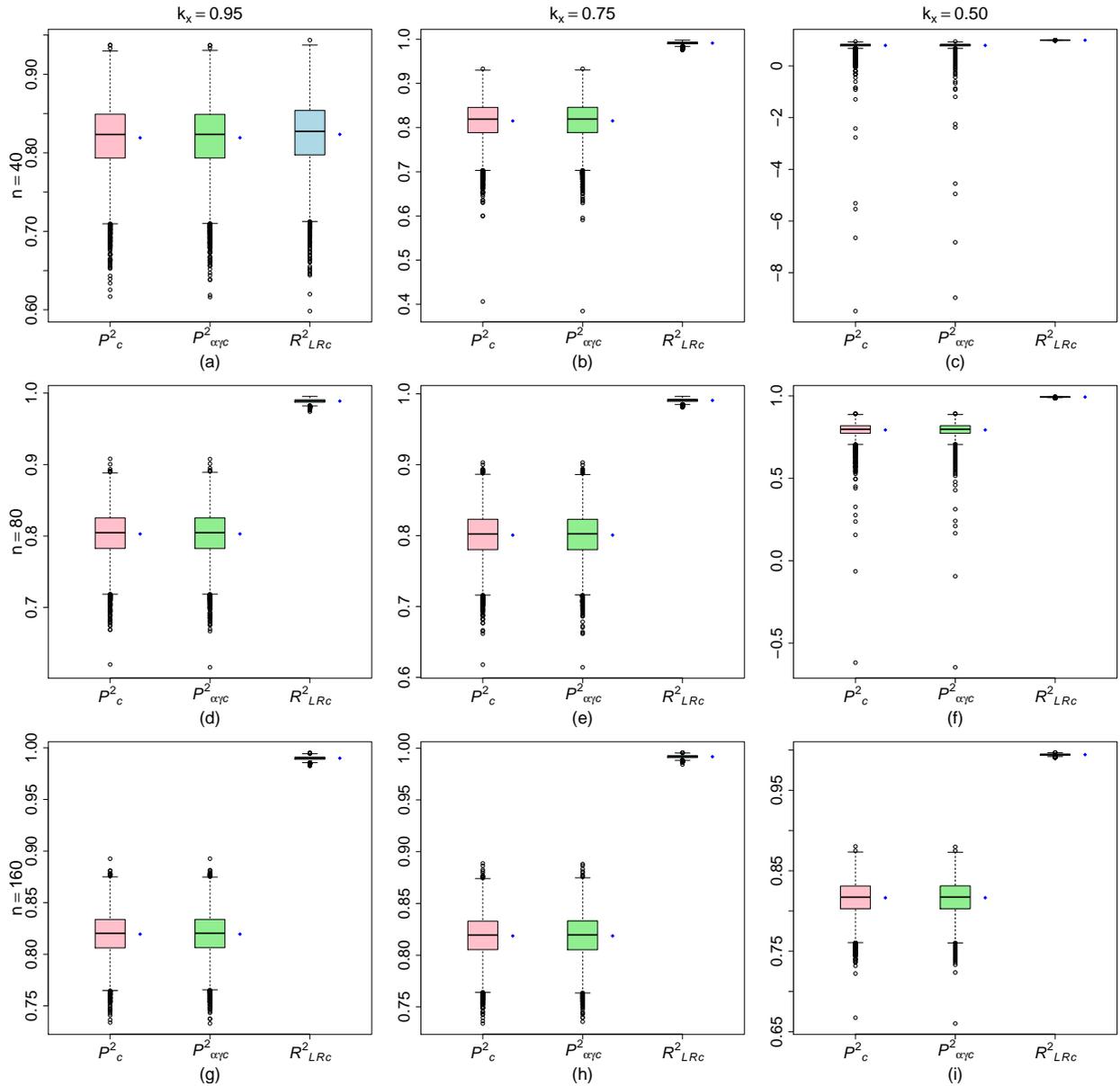


Figura 5.3: *Boxplots* das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$) e R_{LR}^2 corrigida (R_{LRc}^2) corrigida obtidas pelo método da **calibração da regressão**. Coeficiente de confiabilidade: $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão constante.

em que $t = 1, 2, \dots, n$. Para cada réplica de Monte Carlo, geraram-se ocorrências da variável sem erro de medida z_{t1} com distribuição uniforme (0.2, 1.2). Os seguintes valores para os parâmetros que permanecem fixos para todas as simulações foram escolhidos: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 2.5$, $\gamma_1 = 0.9$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$.

A Tabela 5.2 apresenta os valores das estatísticas considerando os coeficientes de confiabilidade $k_x = 0.75$ e $k_x = 0.95$ e os métodos da máxima verossimilhança aproximada (MVa), máxima pseudo-verossimilhança aproximada (MPVa) e calibração da regressão (RC) para estimação dos parâmetros do modelo com dispersão variável sem erro de medida.

Tabela 5.2: Valores das estatísticas considerando os coeficientes de confiabilidade $k_x = 0.75$ e $k_x = 0.95$ e os diferentes métodos de estimação. Modelo sem erro de medida no preditor da dispersão.

n		Cenário 2: $g(\mu_t) = \alpha_0 + z_{1t}^{\alpha_1} + \beta_1 x_{1t}$ e $h(\phi_t) = \gamma_0 + \gamma_1 z_{1t}$					
		$k_x = 0.95$			$k_x = 0.75$		
		MVa	MPVa	RC	MVa	MPVa	RC
40	P^2	0.79417	0.79417	0.79402	0.72783	0.73510	0.76461
	P_c^2	0.77065	0.77065	0.77048	0.69673	0.70483	0.73771
	$P_{\alpha\gamma}^2$	0.79423	0.79423	0.79423	0.72887	0.73574	0.76513
	$P_{\alpha\gamma c}^2$	0.77072	0.77071	0.77071	0.69788	0.70554	0.73829
	ps- R^2	0.78277	0.78277	0.78273	0.77706	0.77721	0.77732
	ps- R_c^2	0.75794	0.75794	0.75790	0.75158	0.75175	0.75187
	R_{LR}^2	0.78793	0.78793	0.98695	0.67540	0.67411	0.98406
	R_{LRc}^2	0.74630	0.74630	0.98439	0.61167	0.61013	0.98093
80	P^2	0.77175	0.77175	0.77147	0.75305	0.75408	0.75998
	P_c^2	0.75958	0.75958	0.75929	0.73988	0.74096	0.74718
	$P_{\alpha\gamma}^2$	0.77180	0.77180	0.77169	0.75317	0.75413	0.76046
	$P_{\alpha\gamma c}^2$	0.75963	0.75963	0.75951	0.74000	0.74102	0.74768
	ps- R^2	0.77659	0.77659	0.77658	0.77397	0.77402	0.77411
	ps- R_c^2	0.76467	0.76467	0.76466	0.76191	0.76197	0.76207
	R_{LR}^2	0.77285	0.77285	0.98667	0.64884	0.64865	0.98373
	R_{LRc}^2	0.75282	0.75282	0.98550	0.61788	0.61768	0.98229
160	P^2	0.76872	0.76872	0.76868	0.76188	0.76189	0.76494
	P_c^2	0.76275	0.76275	0.76271	0.75573	0.75574	0.75888
	$P_{\alpha\gamma}^2$	0.76866	0.76866	0.76879	0.76176	0.76176	0.76535
	$P_{\alpha\gamma c}^2$	0.76269	0.76269	0.76283	0.75561	0.75562	0.75930
	ps- R^2	0.78172	0.78172	0.78172	0.78054	0.78054	0.78057
	ps- R_c^2	0.77609	0.77609	0.77609	0.77488	0.77488	0.77490
	R_{LR}^2	0.77416	0.77416	0.98711	0.64719	0.64719	0.98410
	R_{LRc}^2	0.76469	0.76469	0.98657	0.63240	0.63239	0.98343

Podemos observar que com o aumento do erro de medida, ou seja, quando o coeficiente de confiabilidade diminui, há uma diminuição nas medidas de predição, evidenciando a dificuldade na qualidade da predição quando o erro de medida vai aumentando, mesmo com o modelo especificado corretamente. Já em relação às medidas de qualidade de ajuste existe um aumento nos resultados com o aumento do tamanho amostral, por exemplo, considerando o método MVa e o R_{LRc}^2 tem-se que $R_{LRc}^2 = (0.74630, 0.75282, 0.76469)$, para $n = (40, 60, 160)$.

Vale salientar que considerando o erro de medida baixo, ou seja $k_x = 0.95$, os valores das estatísticas dos métodos de máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa) são idênticos (Tabela 5.2). Já em relação ao erro de medida ponderado, ou seja, $k_x = 0.75$, o método da calibração da regressão apresenta os maiores valores das medidas de predição e da qualidade do ajuste.

Modelo com erro de medida no preditor da dispersão

Neste cenário admite-se que o modelo de regressão beta não linear com erros de medida e dispersão variável, sob o enfoque estrutural, apresenta a seguinte estrutura:

$$\begin{aligned} g(\mu_t) &= \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}, \\ h(\phi_t) &= \gamma_0 + \lambda_1 x_{t1}, \\ w_t &= x_t + e_t, \end{aligned}$$

em que $t = 1, 2, \dots, n$. Para cada réplica de Monte Carlo, geraram-se ocorrências da variável sem erro de medida z_{t1} com distribuição uniforme (0.2, 1.2). Os seguintes valores para os parâmetros que permanecem fixos para todas as simulações foram escolhidos: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 2.5$, $\lambda_1 = 0.9$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$.

A Figura 5.4 apresenta os *boxplots* das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$) e R_{LR}^2 corrigida (R_{LRc}^2) corrigida considerando o coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado) e tamanho amostral $n = 160$, para os métodos da máxima verossimilhança aproximada (a), máxima pseudo-verossimilhança aproximada (b) e calibração da regressão (c), respectivamente.

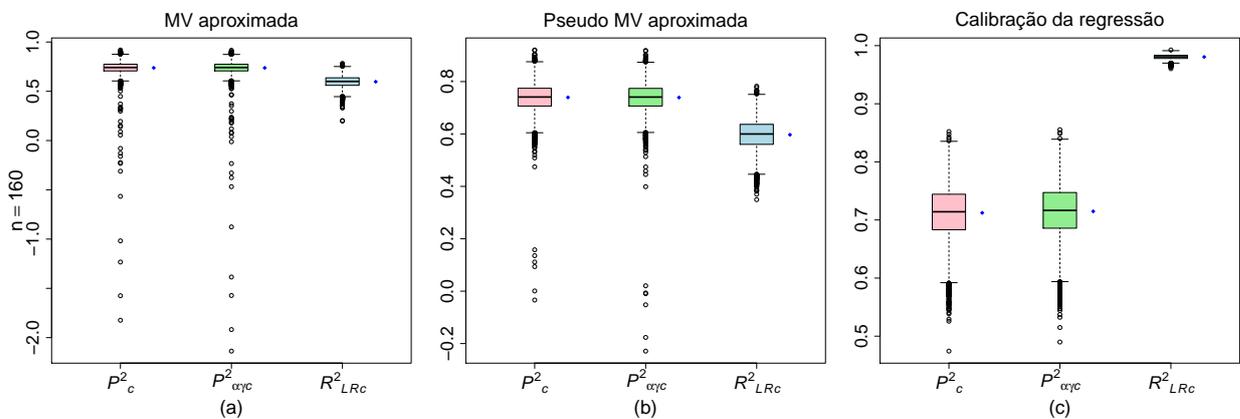


Figura 5.4: *Boxplots* das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$) e R_{LR}^2 corrigida (R_{LRc}^2) corrigida considerando o coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado) e tamanho amostral $n = 160$. Modelo com erro de medida no preditor da dispersão.

Com base na Figura 5.4 notamos que apesar do método da máxima verossimilhança aproximada (MV aproximada), Figura 5.4 (a), apresentar valores médios mais próximos de um para as medidas de predição também pode apresentar os maiores valores negativos para essas medidas, indicando que em muitas situações esse método apresentou dificuldades de fornecer boas predições. No geral quanto à predição, em um modelo com erro de medida simultaneamente na média e na dispersão e com não linearidade na média o

método da calibração da regressão (Calibração da regressão), foi o que apresentou o melhor desempenho considerando a distribuição global de seus valores. Neste mesmo cenário quanto a qualidade de ajuste o melhor método foi também o de calibração de regressão, Figura 5.4 (c).

5.2.3 Cenário 3: Não linearidade na dispersão

Neste cenário admite-se que o modelo de regressão beta não linear com erros de medida e não linearidade na dispersão, sob o enfoque estrutural, apresenta a estrutura:

$$\begin{aligned} g(\mu_t) &= \alpha_0 + z_{t1}^{\alpha_1} + \beta_1 x_{t1}, \\ h(\phi_t) &= \gamma_0 + v_{t1}^{\gamma_1} + \lambda_1 x_{t1}, \\ w_t &= x_t + e_t, \end{aligned}$$

em que $t = 1, 2, \dots, n$. Para cada réplica de Monte Carlo, geraram-se ocorrências das variáveis sem erro de medida z_{t1} e v_{t1} , ambas com distribuição uniforme em $(0.2, 1.2)$. Os seguintes valores para os parâmetros que permanecem fixos para todas as simulações foram escolhidos: $\alpha_0 = 0.7$, $\alpha_1 = 2.0$, $\beta_1 = -1.5$, $\gamma_0 = 1.5$, $\gamma_1 = 2.0$, $\lambda_1 = 1.3$, $\mu_x = 1.5$ e $\sigma_x^2 = 0.5$.

Tabela 5.3: Valores das estatísticas considerando os coeficientes de confiabilidade $k_x = 0.75$ e $k_x = 0.95$ e os diferentes métodos de estimação. Modelo com não linearidade na dispersão.

		Cenário 3: $g(\mu_t) = \alpha_0 + z_1^{\alpha_1} + \beta_1 x_1$ e $h(\phi_t) = \gamma_0 + v_{t1}^{\gamma_1} + \lambda_1 x_{t1}$					
		$k_x = 0.95$			$k_x = 0.75$		
		MVa	MPVa	RC	MVa	MPVa	RC
60	P^2	0.8576	0.8609	0.8615	0.5352	0.6229	0.8065
	P_c^2	0.8444	0.8480	0.8487	0.4922	0.5879	0.7886
	$P_{\alpha\gamma}^2$	0.8585	0.8615	0.8604	0.5574	0.6380	0.8032
	$P_{\alpha\gamma c}^2$	0.8454	0.8486	0.8474	0.5164	0.6045	0.7850
	ps- R^2	0.6622	0.6622	0.6621	0.6599	0.6600	0.6604
	ps- R_c^2	0.6309	0.6309	0.6308	0.6284	0.6285	0.6289
	R_{LR}^2	0.8745	0.8745	0.9905	0.7468	0.7458	0.9852
	R_{LRc}^2	0.8576	0.8576	0.9892	0.7126	0.7116	0.9833
	80	P^2	0.8739	0.8739	0.8602	0.7425	0.7628
P_c^2		0.8654	0.8653	0.8508	0.7251	0.7468	0.7986
$P_{\alpha\gamma}^2$		0.8741	0.8741	0.8591	0.7465	0.7690	0.8067
$P_{\alpha\gamma c}^2$		0.8656	0.8656	0.8495	0.7294	0.7534	0.7936
ps- R^2		0.6563	0.6563	0.6563	0.6549	0.6549	0.6552
ps- R_c^2		0.6331	0.6331	0.6330	0.6316	0.6315	0.6319
R_{LR}^2		0.8736	0.8736	0.9905	0.7420	0.7414	0.9852
R_{LRc}^2		0.8613	0.8613	0.9896	0.7169	0.7162	0.9837

A Tabela 5.3 mostra os valores médios das medidas mostradas na seção anterior considerando os coeficientes de confiabilidade $k_x = 0.75$ e $k_x = 0.95$ e os diferentes métodos de estimação para o modelo de regressão beta não linear com erros de medida e não

linearidade na dispersão. Nota-se que os métodos da máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa) são consideravelmente mais afetados quanto ao poder de predição do modelo que o método de calibração de regressão (RC), neste cenário em que consideramos erro de medida e não linearidade na média e na dispersão simultaneamente. O mesmo ocorre quanto a qualidade do ajuste.

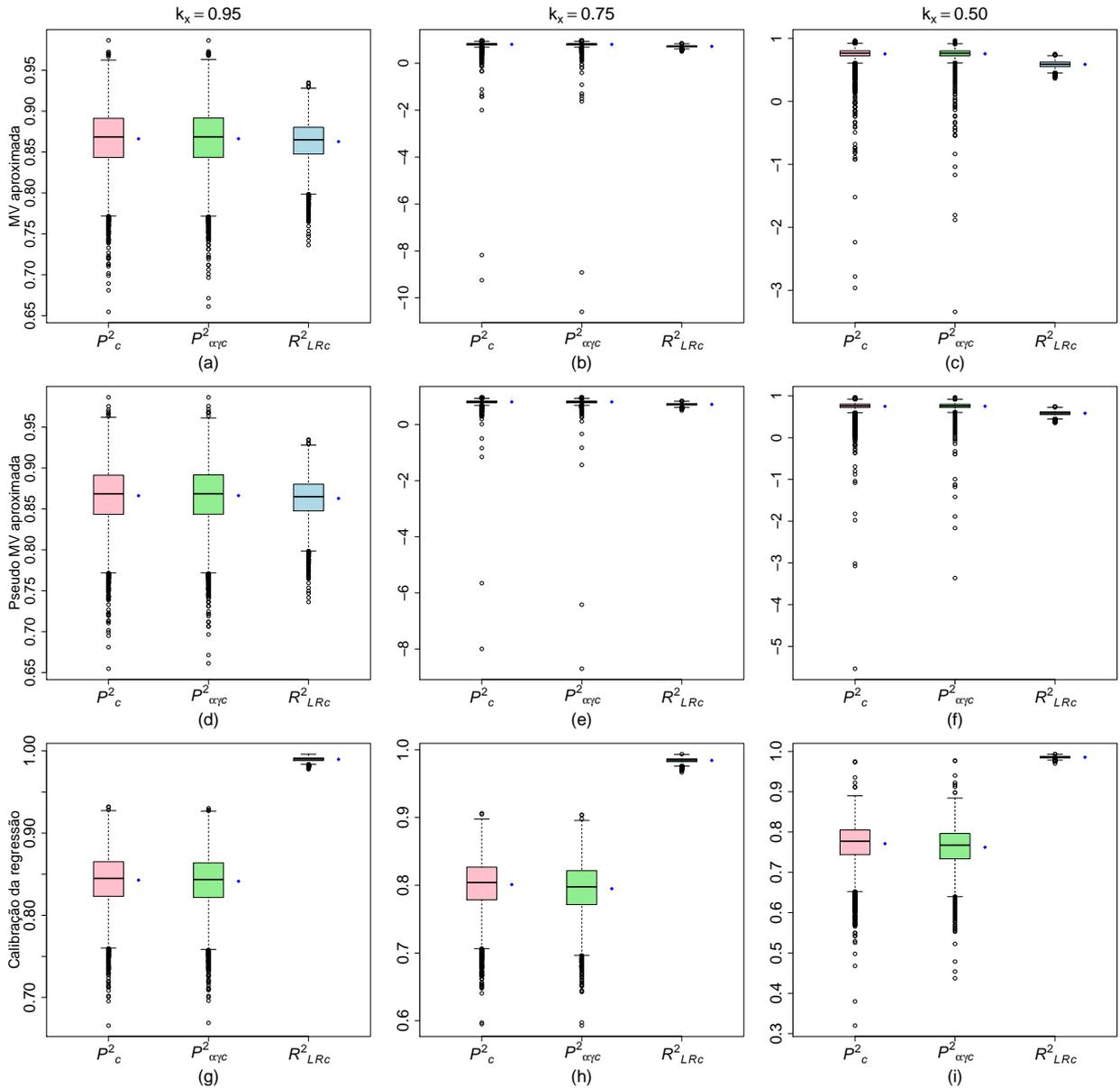


Figura 5.5: *Boxplots* das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$) e R_{LR}^2 corrigida (R_{LR}^2) obtidas pelos métodos avaliados considerando os coeficientes de confiabilidade $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Modelo com dispersão variável não linear. Tamanho amostral $n = 160$.

Uma análise mais detalhada das estatísticas de medidas de qualidade de ajuste propostas para o modelo de beta não linear com erros de medidas foi realizada através de gráficos *boxplots* considerando os métodos da máxima verossimilhança aproximada (MV

aproximada), pseudo-verossimilhança aproximada (Pseudo MV aproximada) e calibração da regressão (Calibração da regressão) apresentados na Figura 5.5, considerando um tamanho amostral $n = 160$ e os coeficientes de confiabilidade $k_x = 0.95$ (erro de medida baixo), $k_x = 0.75$ (erro de medida moderado) e $k_x = 0.50$ (erro de medida alto). Foram avaliadas as medidas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando resíduo combinado ($P_{\alpha\gamma c}^2$) e R_{LR}^2 corrigida (R_{LRc}^2).

As medidas P^2 e $P_{\alpha\gamma}^2$, novamente, indicam que à medida que o coeficiente de confiabilidade diminui a qualidade de predição do modelo é substancialmente afetada negativamente mesmo quando o modelo está bem especificado. Esse resultado corrobora com o desempenho dos estimadores do parâmetro associado a não linearidade o qual é prejudicado tanto quanto ao viés quanto às taxas de coberturas dos intervalos de confiança. Novamente, o método da calibração da regressão apresentou melhores resultados quando comparados aos outros métodos. Este fato pode estar associado ao melhor ajuste para o parâmetro associado a não linearidade do modelo, como mostra a Figura 5.5 (g)-(i).

5.2.4 Cenário 4 - Modelo Linear: Má especificação do modelo

Considere que os dados foram gerados através do modelo de regressão beta linear com erros de medida sob o enfoque estrutural com a seguinte estrutura:

$$\begin{aligned} g(\mu_t) &= \alpha_0 + \alpha_1 z_{t1} + \alpha_2 z_{t2} + \alpha_3 z_{t3} + \beta_1 x_{t1}, \\ h(\phi) &= \gamma_0, \\ w_t &= x_t + e_t, \end{aligned}$$

em que $t = 1, 2, \dots, n$. Para cada réplica de Monte Carlo, geraram-se ocorrências das variáveis sem erros de medida z_{t1} , z_{t2} e z_{t3} cada uma com distribuição uniforme (0.2, 1.2). Os seguintes valores para os parâmetros que permanecem fixos para todas as simulações foram escolhidos: $\alpha_0 = -2.2$, $\alpha_1 = 1.2$, $\alpha_2 = 1.3$, $\alpha_3 = 1.1$, $\beta_1 = 1.5$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Também avaliamos a influência do parâmetro de dispersão, considerando $\gamma_0 = (2.8, 5.0, 5.7)$, que fornecem $\phi \approx (20, 150, 300)$.

Com o objetivo de avaliar os desempenhos das estatísticas de predição e a qualidade do ajuste dos modelos considerados sob a omissão de covariáveis significativas para predição, foram geradas 5000 réplicas de Monte Carlo e considerou-se os seguintes cenários:

- Cenário 1: $g(\mu_t) = \alpha_0 + \beta_1 x_{t1}$;
- Cenário 2: $g(\mu_t) = \alpha_0 + \alpha_1 z_{t1} + \beta_1 x_{t1}$;
- Cenário 3: $g(\mu_t) = \alpha_0 + \alpha_1 z_{t1} + \alpha_2 z_{t2} + \beta_1 x_{t1}$;
- Cenário 4: $g(\mu_t) = \alpha_0 + \alpha_1 z_{t1} + \alpha_2 z_{t2} + \alpha_3 z_{t3} + \beta_1 x_{t1}$

Tabela 5.4: Valores das estatísticas. Modelo mal especificado: omissão de covariáveis. O modelo verdadeiro é $g(\mu_t) = \alpha_0 + \alpha_1 z_{t1} + \alpha_2 z_{t2} + \alpha_3 z_{t3} + \beta_1 x_{t1}, t = 1, \dots, n$ e ϕ constante. Método: Máxima verossimilhança aproximada. Coeficiente de confiabilidade: $k_x = 0.75$.

Cenários	Cenário 1			Cenário 2			Cenário 3			Cenário 4			
	$g(\mu_t) = \alpha_0 + \beta_1 x_t$			$g(\mu_t) = \alpha_0 + \alpha_1 z_{1t} + \beta_1 x_t$			$g(\mu_t) = \alpha_0 + \alpha_1 z_{1t} + \alpha_2 z_{2t} + \beta_1 x_t$			$g(\mu_t) = \alpha_0 + \alpha_1 z_{1t} + \alpha_2 z_{2t} + \alpha_3 z_{3t} + \beta_1 x_t$			
n	20	150	300	20	150	300	20	150	300	20	150	300	
40	P^2	0.70930	0.73669	0.72734	0.74654	0.82492	0.83602	0.74934	0.84680	0.86898	0.76962	0.89916	0.93306
	P_c^2	0.66655	0.69797	0.68724	0.70926	0.79917	0.81190	0.71247	0.82427	0.84971	0.73574	0.88433	0.92322
	$P_{\alpha\gamma}^2$	0.70939	0.73703	0.72727	0.74805	0.82579	0.83692	0.75070	0.84796	0.87016	0.77130	0.90006	0.93372
	$P_{\alpha\gamma c}^2$	0.66666	0.69836	0.68716	0.71100	0.80017	0.81294	0.71403	0.82561	0.85107	0.73767	0.88536	0.92397
	ps- R^2	0.71146	0.83045	0.84084	0.78139	0.90584	0.91959	0.80200	0.92528	0.94286	0.83709	0.95950	0.97996
	ps- R_c^2	0.66903	0.80552	0.81744	0.74924	0.89199	0.90777	0.77288	0.91429	0.93446	0.81313	0.95354	0.97701
	R_{LR}^2	0.64446	0.69168	0.69422	0.71218	0.77668	0.78579	0.74003	0.80976	0.81915	0.77777	0.85154	0.86008
	R_{LRc}^2	0.59217	0.64634	0.64925	0.66986	0.74383	0.75428	0.70179	0.78178	0.79255	0.74509	0.82971	0.83950
	P^2	0.68822	0.72555	0.70910	0.72781	0.79497	0.79056	0.77466	0.88135	0.89435	0.80415	0.93469	0.95738
	P_c^2	0.66715	0.70700	0.68944	0.70942	0.78111	0.77641	0.75943	0.87333	0.88722	0.79091	0.93027	0.95450
80	P^2	0.69252	0.73143	0.71519	0.73187	0.80013	0.79586	0.77671	0.88375	0.89685	0.80477	0.93499	0.95755
	$P_{\alpha\gamma}^2$	0.67174	0.71328	0.69594	0.71375	0.78663	0.78207	0.76162	0.87589	0.88989	0.79158	0.93059	0.95468
	$P_{\alpha\gamma c}^2$	0.69821	0.81879	0.82854	0.75423	0.87859	0.88992	0.80870	0.93632	0.95015	0.83895	0.96695	0.98288
	ps- R^2	0.67782	0.80654	0.81695	0.73763	0.87038	0.88248	0.79577	0.93202	0.94679	0.82806	0.96472	0.98172
	ps- R_c^2	0.62797	0.66359	0.66146	0.67891	0.72967	0.73445	0.72864	0.79112	0.79860	0.75738	0.82639	0.83386
	R_{LR}^2	0.60284	0.64086	0.63858	0.65722	0.71141	0.71651	0.71030	0.77701	0.78499	0.74098	0.81466	0.82263
	R_{LRc}^2	0.70072	0.76262	0.75690	0.73407	0.81346	0.81295	0.78078	0.89295	0.90341	0.81580	0.95215	0.96834
	P^2	0.69101	0.75491	0.74901	0.72544	0.80740	0.80688	0.77367	0.88948	0.90027	0.80982	0.95060	0.96731
	P_c^2	0.70503	0.76778	0.76230	0.73700	0.81690	0.81653	0.78183	0.89388	0.90430	0.81606	0.95221	0.96837
	$P_{\alpha\gamma}^2$	0.69545	0.76024	0.75459	0.72846	0.81096	0.81058	0.77475	0.89044	0.90120	0.81009	0.95066	0.96735
160	ps- R^2	0.71944	0.84395	0.85461	0.75881	0.88657	0.89768	0.80692	0.93818	0.95019	0.83781	0.97075	0.98383
	ps- R_c^2	0.71033	0.83889	0.84989	0.75098	0.88289	0.89436	0.80065	0.93617	0.94857	0.83255	0.96980	0.98331
	R_{LR}^2	0.63503	0.67455	0.67668	0.67058	0.71974	0.72419	0.71308	0.77237	0.77872	0.74100	0.80612	0.81293
	R_{LRc}^2	0.62318	0.66398	0.66619	0.65988	0.71064	0.71524	0.70376	0.76498	0.77154	0.73259	0.79982	0.80686

Tabela 5.6: Valores das estatísticas. Modelo mal especificado: omissão de covariáveis. O modelo verdadeiro é $g(\mu_t) = \alpha_0 + \alpha_1 z_{t1} + \alpha_2 z_{t2} + \alpha_3 z_{t3} + \beta_1 x_{t1}, t = 1, \dots, n$ e ϕ constante. Método: Calibração da regressão. Coeficiente de confiabilidade: $k_x = 0.75$.

Cenários	Cenário 1			Cenário 2			Cenário 3			Cenário 4			
	$g(\mu_t) = \alpha_0 + \beta_1 x_t$			$g(\mu_t) = \alpha_0 + \alpha_1 z_{1t} + \beta_1 x_t$			$g(\mu_t) = \alpha_0 + \alpha_1 z_{1t} + \alpha_2 z_{2t} + \beta_1 x_t$			$g(\mu_t) = \alpha_0 + \alpha_1 z_{1t} + \alpha_2 z_{2t} + \alpha_3 z_{3t} + \beta_1 x_t$			
n	ϕ	20	150	300	20	150	300	20	150	300	20	150	300
40	P^2	0.76181	0.82634	0.83144	0.80704	0.88266	0.88883	0.81447	0.89512	0.90226	0.77097	0.60633	0.56475
	P_c^2	0.72678	0.80080	0.80665	0.77867	0.86540	0.87248	0.78719	0.87970	0.88788	0.73728	0.54843	0.50074
	$P_{\alpha\gamma}^2$	0.76057	0.82376	0.82865	0.80723	0.88201	0.88814	0.81500	0.89508	0.90212	0.77902	0.63650	0.60026
	$P_{\alpha\gamma c}^2$	0.72536	0.79784	0.80346	0.77888	0.86465	0.87169	0.78779	0.87965	0.88772	0.74652	0.58304	0.54148
	ps- R^2	0.71146	0.83045	0.84084	0.78147	0.90553	0.91641	0.80190	0.92469	0.93587	0.83218	0.94823	0.95854
	ps- R_c^2	0.66903	0.80552	0.81744	0.74933	0.89164	0.90412	0.77277	0.91362	0.92644	0.80750	0.94062	0.95244
	R_{LR}^2	0.98168	0.98372	0.98385	0.98490	0.98736	0.98755	0.98619	0.98871	0.98887	0.98732	0.99006	0.99024
	R_{LRc}^2	0.97899	0.98133	0.98148	0.98268	0.98550	0.98572	0.98416	0.98705	0.98723	0.98546	0.98860	0.98880
	P^2	0.72810	0.80414	0.80961	0.77471	0.85723	0.86357	0.82329	0.91651	0.92461	0.79024	0.67700	0.64352
	P_c^2	0.70972	0.79090	0.79675	0.75948	0.84758	0.85435	0.81136	0.91087	0.91952	0.77607	0.65517	0.61943
80	P^2	0.73173	0.80724	0.81256	0.77770	0.85987	0.86610	0.82445	0.91727	0.92529	0.79564	0.69665	0.66649
	$P_{\alpha\gamma}^2$	0.71361	0.79421	0.79989	0.76267	0.85040	0.85706	0.81259	0.91168	0.92024	0.78184	0.67616	0.64395
	$P_{\alpha\gamma c}^2$	0.69821	0.81879	0.82854	0.75433	0.87872	0.88908	0.80868	0.93657	0.94778	0.83154	0.94880	0.95887
	ps- R^2	0.67782	0.80654	0.81695	0.73773	0.87052	0.88159	0.79575	0.93228	0.94425	0.82015	0.94534	0.95609
	ps- R_c^2	0.98195	0.98387	0.98394	0.98432	0.98655	0.98666	0.98664	0.98915	0.98925	0.98763	0.99031	0.99048
	R_{LR}^2	0.98073	0.98278	0.98285	0.98326	0.98564	0.98576	0.98574	0.98841	0.98852	0.98679	0.98965	0.98984
	P^2	0.73693	0.82246	0.82892	0.77409	0.86499	0.87206	0.82205	0.92267	0.93096	0.79979	0.71320	0.68502
	P_c^2	0.72839	0.81670	0.82337	0.76676	0.86061	0.86790	0.81628	0.92016	0.92872	0.79329	0.70389	0.67480
	$P_{\alpha\gamma}^2$	0.74087	0.82578	0.83223	0.77646	0.86679	0.87382	0.82262	0.92248	0.93072	0.80414	0.72821	0.70257
	$P_{\alpha\gamma c}^2$	0.73245	0.82012	0.82679	0.76921	0.86246	0.86972	0.81686	0.91996	0.92848	0.79778	0.71938	0.69291
160	ps- R^2	0.71944	0.84395	0.85461	0.75884	0.88667	0.89763	0.80688	0.93851	0.94991	0.83197	0.95485	0.96533
	ps- R_c^2	0.71033	0.83889	0.84989	0.75101	0.88299	0.89431	0.80061	0.93652	0.94828	0.82651	0.95338	0.96421
	R_{LR}^2	0.98276	0.98484	0.98493	0.98439	0.98667	0.98679	0.98634	0.98887	0.98901	0.98737	0.99008	0.99025
	R_{LRc}^2	0.98220	0.98434	0.98444	0.98388	0.98624	0.98636	0.98589	0.98851	0.98865	0.98696	0.98976	0.98993

No cenário 1 considerou-se apenas o intercepto e a variável mensurada com erros de medida, x_{t1} . No cenário 2 adicionou-se uma variável sem erros, z_{t1} , ao modelo do cenário 1. Para o cenário 3 acrescentou-se mais uma variável mensurada sem erros de medida, z_{t2} , ao modelo do cenário 2. Por fim, considerou-se o modelo proposto.

As Tabelas 5.4, 5.5 e 5.6 apresentam as estatísticas de predição P^2 considerando o resíduo ponderado (P^2) e sua versão corrigida (P_c^2), P^2 considerando o resíduo combinado ($P_{\alpha\gamma}^2$) e sua versão corrigida ($P_{\alpha\gamma c}^2$), bem como as medidas de qualidade de ajuste pseudo- R^2 (ps- R^2) e sua versão corrigida (ps- R_c^2) e a pseudo- R^2 baseada na razão de verossimilhança (R_{LR}^2) e sua versão corrigida (R_{LRc}^2).

Nota-se com base nas Tabelas 5.4 a 5.6 que considerando todos os métodos, a saber máxima verossimilhança aproximada (MVa), máxima pseudo-verossimilhança aproximada (MPVa) e calibração da regressão (RC), as estatísticas relacionadas à predição e à qualidade de ajuste aumentam em valores à medida que covariáveis importantes são consideradas corretamente na modelagem da média da variável resposta. Ressalta-se ainda a influência do valor de ϕ , tal que mesmo quando covariáveis importantes são desconsideradas na modelagem para os valores de ϕ maiores, maiores são os valores das estatísticas.

É possível notar também que o método da calibração da regressão (RC) apresentou os melhores resultados quando os modelos estão fora da especificação, ou seja, nos cenários 1, 2 e 3. Além disso, as medidas R_{LR}^2 e R_{LRc}^2 apresentam valores próximos de 1 e bem diferentes dos resultados apresentados nas medidas ps- R^2 e ps- R_c^2 (Tabela 5.6). Já os métodos da máxima verossimilhança aproximada e máxima pseudo-verossimilhança aproximada mostraram os melhores resultados no cenário 4 (Tabelas 5.4 e 5.5).

5.3 Avaliação da má especificação da distribuição mensurada com erro de medida

Nesta seção avaliamos o comportamento dos estimadores sob a má especificação da distribuição da variável mensurada com erro de medida. Inicialmente, é apresentada uma revisão das quatro distribuições comumente utilizadas na literatura que foram avaliadas. Para obtenção dos resultados, realizamos simulações de Monte Carlo com 5000 réplicas e os resultados são apresentados a seguir.

5.3.1 Distribuição *t-Student*

Seja X uma variável aleatória que assume valores reais. Se X segue distribuição *t-Student*, denotado por $X \sim t(\nu)$, então a função densidade de probabilidade é dada por

$$f(x; \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad (5.12)$$

em que ν é o número de graus de liberdade e Γ é a função gama.

A função densidade de probabilidade é simétrica e sua forma geral se assemelha à forma de sino de uma variável distribuída normalmente com média 0 e variância 1, exceto que é mais achatada e alcança os valores mais extremos. A Figura 5.6 apresenta a função de densidade da distribuição *t-Student* para diferentes valores de ν .

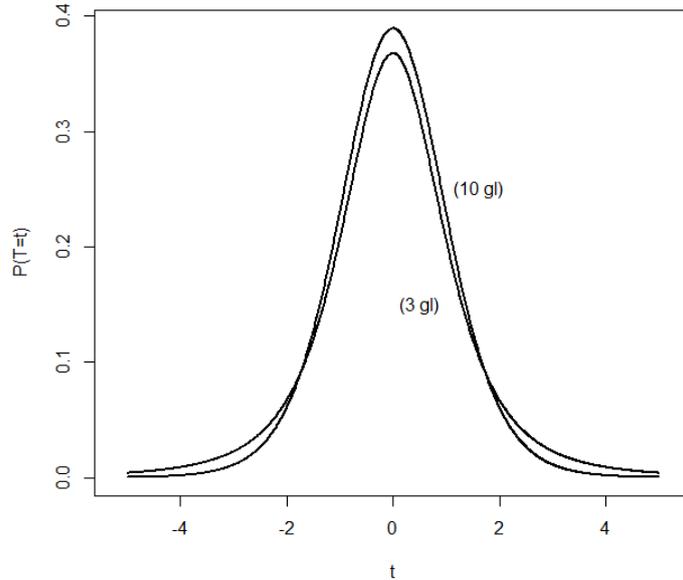


Figura 5.6: Função densidade de probabilidade da distribuição *t-Student* para diferentes graus de liberdade (gl).

A média e a variância são dadas, respectivamente, por

$$\begin{aligned} \mathbb{E}(X) &= 0 \quad \text{e} \\ \text{Var}(X) &= \begin{cases} \frac{\nu}{\nu-2}, & \nu > 2 \\ \infty, & 1 < \nu \leq 2. \end{cases} \end{aligned}$$

5.3.2 Distribuição Log-Normal

Se a variável aleatória X tem distribuição log-normal com parâmetros μ e σ^2 , então $X^* = \ln(X)$ segue distribuição normal com parâmetros μ e σ^2 . Bem como, se X^* segue distribuição normal com parâmetros μ e σ^2 , então a função exponencial de X^* , $X = \exp(X^*)$, segue uma distribuição log-normal com parâmetros μ e σ^2 .

Vale salientar que uma variável aleatória que segue distribuição log-normal assume apenas valores reais positivos. A função densidade de probabilidade é dada por

$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right]. \quad (5.13)$$

A Figura 5.7 apresenta a função densidade de probabilidade para diferentes parâmetros μ e σ^2 . Portanto, dizemos que se X é uma variável aleatória que tem a densidade apresentada na Equação (5.13), então $X \sim LN(\mu, \sigma^2)$.

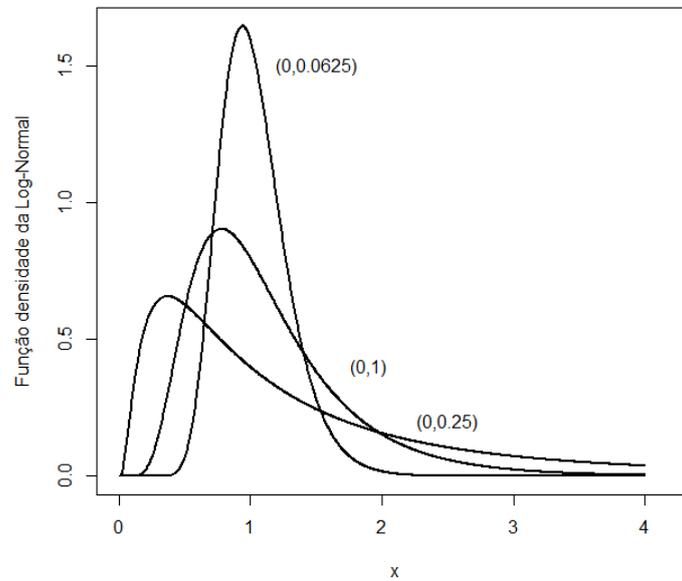


Figura 5.7: Função densidade de probabilidade da distribuição Log-Normal para diferentes parâmetros μ e σ^2 .

A média e a variância de X são dadas, respectivamente, por

$$\begin{aligned}\mathbb{E}(X) &= \exp\left(\mu + \frac{\sigma^2}{2}\right) \quad e \\ \text{Var}(X) &= [\exp(\sigma^2) - 1] \exp(2\mu + \sigma^2)\end{aligned}$$

5.3.3 Distribuição Weibull

Seja X uma variável aleatória que assume números reais positivos. Se X segue uma distribuição Weibull, então sua função densidade de probabilidade é dada por

$$f(x; k, \lambda) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, \quad (5.14)$$

em que $\lambda \in (0, +\infty)$ é um parâmetro de escala e $k \in (0, +\infty)$ denota o parâmetro de forma.

A Figura 5.8 apresenta a função densidade de probabilidade para diferentes parâmetros k e λ . Portanto, dizemos que se X é uma variável aleatória que tem a densidade apresentada na Equação (5.14), então $X \sim Weibull(k, \lambda)$.

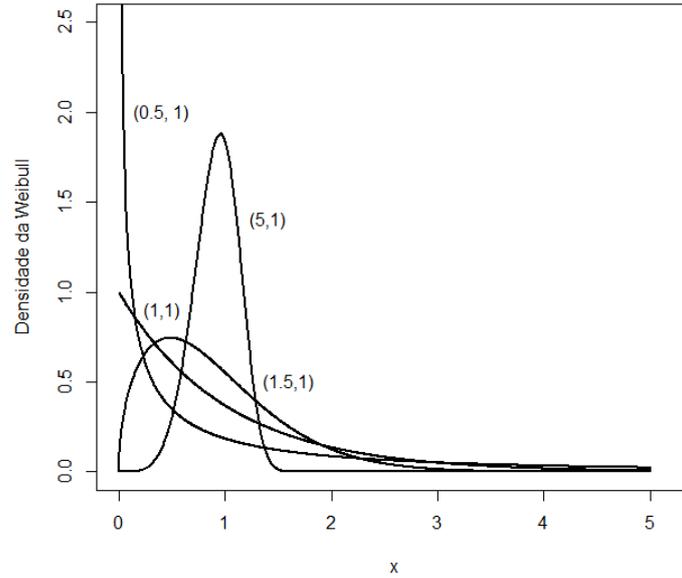


Figura 5.8: Função densidade de probabilidade para diferentes parâmetros k e λ .

A média e a variância de X são dadas, respectivamente, por

$$\begin{aligned}\mathbb{E}(X) &= \lambda \Gamma\left(1 + \frac{1}{k}\right) \quad \text{e} \\ \text{Var}(X) &= \lambda^2 \left[\Gamma\left(1 + \frac{2}{k}\right) - \left(\Gamma\left(1 + \frac{1}{k}\right)\right)^2 \right].\end{aligned}$$

Reparametrizando temos que $b = \lambda^{(-k)}$. Para $x \geq 0$ a função de risco é dada por

$$h(x; k, b) = b k x^{k-1}.$$

Desta forma, a função densidade é dada por

$$f(x; \lambda, k) = b k x^{k-1} \exp(-b x^k), \quad (5.15)$$

cuja média é dada por $\mathbb{E}(X) = b^{-1/k} \Gamma\left(1 + \frac{1}{k}\right)$.

5.3.4 Distribuição Beta

A distribuição beta é uma família de distribuições de probabilidade contínuas definidas no intervalo $(0,1)$ parametrizado por dois parâmetros positivos denotados por α e β , que aparecem como expoentes da variável aleatória e controlam o formato da distribuição.

A função densidade de probabilidade da distribuição beta, para $x \in (0, 1)$ e parâmetros α e $\beta > 0$, é dada por

$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad (5.16)$$

em que Γ denota a função gama.

A Figura 5.9 apresenta a função densidade de probabilidade para diferentes parâmetros α e β . Portanto, dizemos que se X é uma variável aleatória que tem a densidade apresentada na Equação (5.16), então $X \sim \text{Beta}(\alpha, \beta)$.

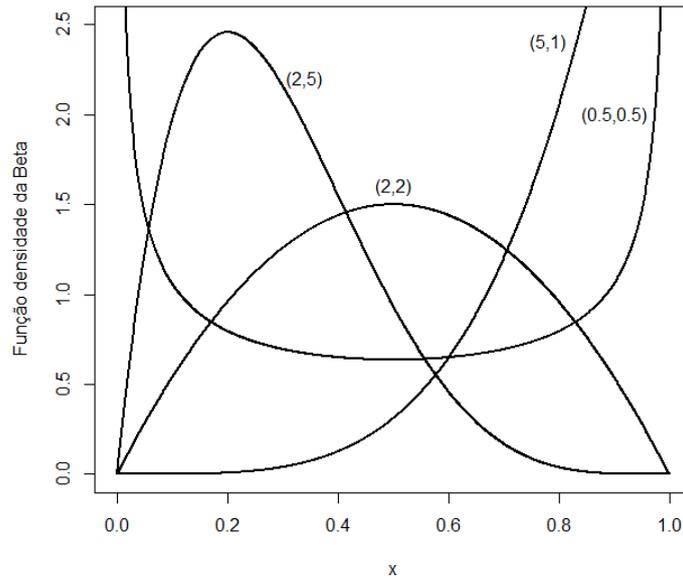


Figura 5.9: Função densidade de probabilidade da distribuição beta para diferentes valores dos parâmetros α e β .

A média e a variância de X são dadas, respectivamente, por

$$\mathbb{E}(X) = \frac{\alpha}{\alpha + \beta} \quad \text{e}$$

$$\text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Considerando a reparametrização $\phi = \alpha + \beta$ é possível reescrever a função de densidade da distribuição beta como segue. Sejam y_1, \dots, y_n variáveis aleatórias independentes. Para cada $t = 1, \dots, n$, y_t segue distribuição beta com função de densidade dada por:

$$f(y_t; \mu_t, \phi_t) = \frac{\Gamma(\phi_t)}{\Gamma(\mu_t\phi_t)\Gamma(1-\mu_t)\phi_t} y_t^{\mu_t\phi_t-1} (1-y_t)^{(1-\mu_t)\phi_t-1}, \quad 0 < y_t < 1, \quad (5.17)$$

em que $0 < \mu_t < 1$, $\phi_t > 0$.

5.4 Avaliação numérica - Má especificação da distribuição

Com o objetivo de avaliar o comportamento dos estimadores sob má especificação da distribuição da variável mensurada com erro de medida, realizamos simulações de Monte

Carlo com 5000 réplicas. Para $\Psi = (\Psi_1, \Psi_2, \Psi_3, \Psi_4)^\top = (\alpha, \beta, \gamma, \lambda)^\top$, tem-se que

$$\bar{\Psi}_j = \frac{1}{R} \sum_{i=1}^R \hat{\Psi}_j^{(i)}, \quad \text{Viés}(\Psi_j) = \bar{\Psi}_j - \Psi_j \quad \text{e} \quad \text{REQM}(\Psi_j) = \sqrt{\frac{1}{R} \sum_{i=1}^R (\hat{\Psi}_j^{(i)} - \Psi_j)^2},$$

em que $j = 1, \dots, 4$, são respectivamente, estimativas da média, do viés e da raiz do erro quadrático médio (REQM) dos estimadores dos parâmetros. Aqui $\hat{\Psi}_j^{(i)}$ denota a estimativa de Ψ_j obtida na i -ésima amostra simulada.

Também avaliamos as taxas de cobertura (TC) dos intervalos de confiança com coeficiente de confiança nominal de 95%. Esta medida foi estimada da seguinte forma

$$\text{TC} = \frac{\#\{\Psi_j \in \text{IC}[\Psi_j; 1 - \alpha]\}}{R},$$

em que Ψ_j é o j -ésimo elemento de $\Psi = (\Psi_1, \Psi_2, \Psi_3, \Psi_4)^\top = (\alpha, \beta, \gamma, \lambda)^\top$, R é o número de réplicas de Monte Carlo e $\text{IC}[\Psi_j; 1 - \alpha]$ é o intervalo de confiança para Ψ_j com coeficiente nominal $1 - \alpha$.

Neste estudo analisamos apenas o modelo de regressão beta não linear com erros de medida e dispersão constante dado por

$$\begin{aligned} g(\mu_t) &= \alpha_0 + z_t^{\alpha_1} + \beta_1 x_t, \\ h(\phi) &= \gamma_0 \\ w_t &= x_t + e_t. \end{aligned} \tag{5.18}$$

Gerou-se uma amostra de tamanho 100 e essa amostra foi tomada com valores verdadeiros de x considerando as quatro distribuições apresentadas na Seção 5.3.

As ligações logito e logarítmica foram escolhidas, respectivamente, para os parâmetros μ_t e ϕ . Ocorrências da variável dependente y_t foram geradas a partir de uma distribuição beta com média μ_t e precisão ϕ . Depois, geraram-se ocorrências da covariável w_t a partir de sua distribuição condicional dado x_t , ou seja, com uma distribuição normal com média x_t e variância σ_e^2 . O coeficiente de confiabilidade foi considerado conhecido e fixado em $k_x = 0.75$ (erro de medida moderado) que representa $\sigma_e^2 = \sigma_x^2/3$. A variável contaminada, w_t , foi utilizada para obtenção dos estimadores dos parâmetros de interesse descritos no Capítulo 2.

Para cada amostra simulada os parâmetros do modelo foram estimados sob dois enfoques: primeiro, assumimos (erroneamente) que não há erros de medida na covariável x_t , ou seja, utilizamos o método *naive* (*Naive*); segundo, reconhecemos que x_t é medida com erros e, nesse caso, utilizamos os três métodos descritos nas seções anteriores, a saber: máxima verossimilhança aproximada (MVa), máxima pseudo-verossimilhança aproximada (MPVa) e calibração da regressão (RC).

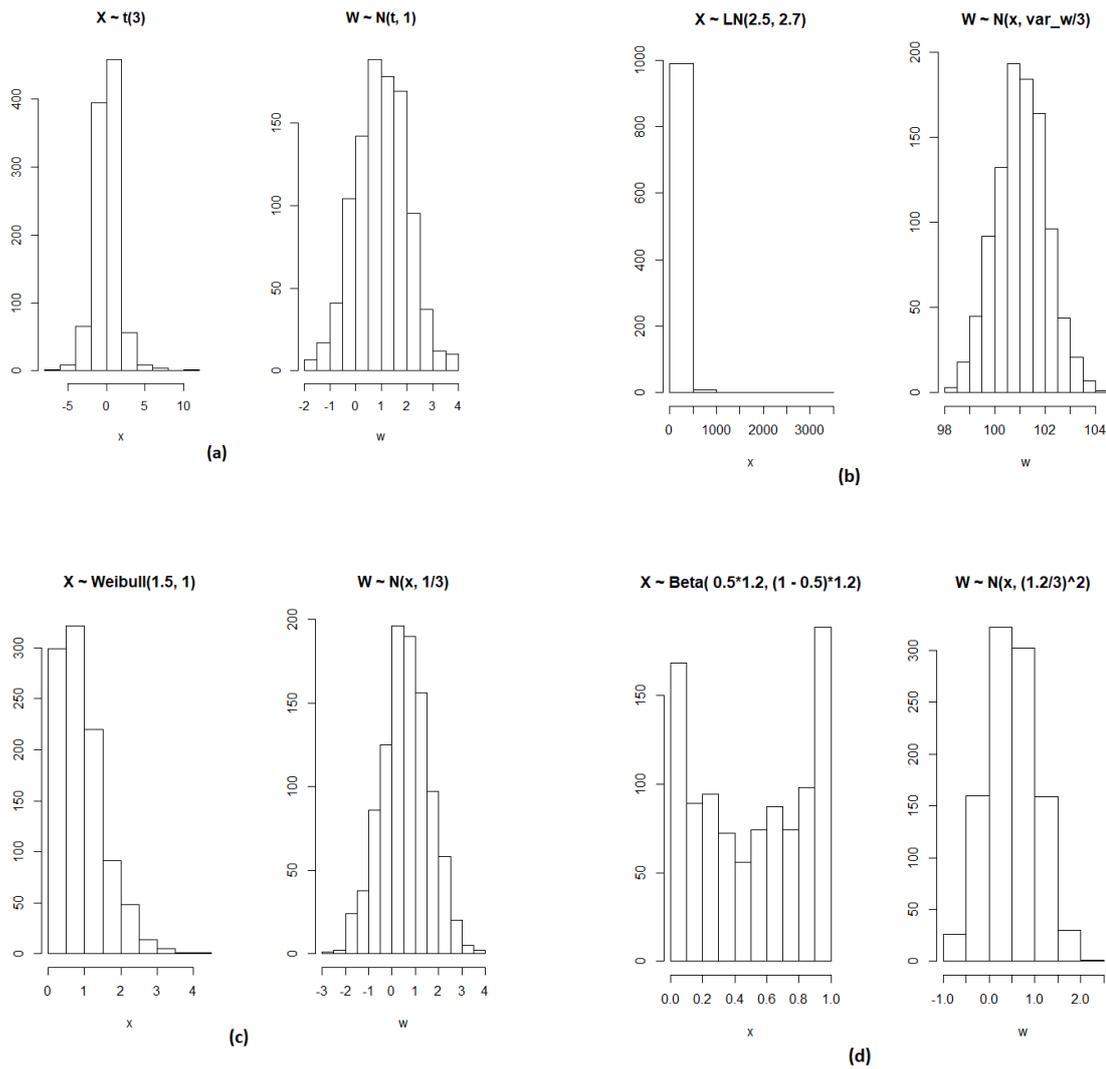


Figura 5.10: Histogramas considerando 100 ocorrências geradas da variável aleatória X que segue diferentes distribuições: (a) $X \sim t(3)$; (b) $X \sim LN(2.5, 2.7)$; (c) $X \sim Weibull(1.5, 1)$; (d) $X \sim Beta(0.6, 0.6)$.

A Figura 5.10 apresenta os histogramas de 100 ocorrências geradas considerando as variáveis aleatórias X e W . Na Figura 5.10-(a), X segue uma distribuição t -Student com 3 graus de liberdade e $W|X \sim N(x, 1)$. Na Figura 5.10-(b), X segue distribuição $LN(2.5, 2.7)$ e $W|X \sim N(x, 2.7/3)$. Na Figura 5.10-(c), temos que X segue distribuição Weibull com parâmetros $(1.5, 1.0)$ e $W|X \sim N(x, 1/3)$. Por fim, na Figura 5.10-(d), X segue distribuição $Beta(\mu_x \sigma_x^2, (1 - \mu_x) \sigma_x^2)$, com $\mu = 0.5$ e $\sigma^2 = 1.2$ e $W|X \sim N(x, 1.2/3)$.

A Tabela 5.7 apresenta os vieses, as raízes dos erros quadráticos médios (REQM) e as taxas de cobertura com 95% de confiança para os estimadores de α_0 , α_1 , β_1 e γ_0 considerando o modelo de regressão beta não linear com erros de medida com dispersão constante dado na Equação (5.18).

Considerando que $X \sim t(3)$ verificou-se que, em relação ao viés, os métodos de máxima

Tabela 5.7: Vieses, raízes do erros quadráticos médios (REQM) e taxas de cobertura com 95% de confiança para os estimadores de α_0 , α_1 , β_1 e γ_0 considerando o modelo com dispersão constante. Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$ e $\gamma_0 = 5.0$.

Cenário	Método		<i>Naive</i>	RC	MVa	MPVa
$X \sim t(\nu)$	Viés	$\hat{\alpha}_0$	0.02458	0.02591	0.00818	0.00814
		$\hat{\alpha}_1$	0.29415	0.29415	0.02077	0.01989
		$\hat{\beta}_1$	0.33941	0.10956	0.06973	0.06146
		$\hat{\gamma}_0$	2.51009	2.51009	0.61598	0.61953
	REQM	$\hat{\alpha}_0$	0.07741	0.08887	0.08751	0.08808
		$\hat{\alpha}_1$	0.63578	0.63578	0.62433	0.62302
		$\hat{\beta}_1$	0.34160	0.13937	0.10847	0.10509
		$\hat{\gamma}_0$	2.51416	2.51416	1.17420	1.14889
	TX Cobertura 95%	α_0	93.48000	88.38000	92.34000	92.78000
		α_1	84.82000	85.16000	84.22000	84.64000
		β_1	0.00000	44.08000	75.68000	77.32000
		γ_0	0.00000	0.00000	51.32000	48.90000
$X \sim LN(\mu_x, \sigma_x^2)$	Viés	$\hat{\alpha}_0$	-1.52351	-0.73267	-0.70486	0.77431
		$\hat{\alpha}_1$	0.56385	0.56385	0.19999	0.68162
		$\hat{\beta}_1$	0.41865	0.23591	0.19415	0.21103
		$\hat{\gamma}_0$	1.05920	1.05920	0.30781	0.48495
	REQM	$\hat{\alpha}_0$	1.53252	0.80871	0.68060	0.75911
		$\hat{\alpha}_1$	0.81898	0.81898	0.19948	0.68159
		$\hat{\beta}_1$	0.42056	0.24967	0.18837	0.20723
		$\hat{\gamma}_0$	1.07048	1.07048	0.30401	0.48245
	TX Cobertura 95%	α_0	0.00000	19.96000	42.62000	39.82000
		α_1	73.50000	73.58000	89.02000	86.52000
		β_1	0.00000	8.76000	36.68000	34.24000
		γ_0	0.00000	0.00000	76.78000	71.04000
$X \sim Weibull(a, b)$	Viés	$\hat{\alpha}_0$	-0.33657	0.00949	-0.03941	-0.02218
		$\hat{\alpha}_1$	0.09502	0.09502	-0.00071	-0.00143
		$\hat{\beta}_1$	0.40731	0.02116	0.05190	0.03211
		$\hat{\gamma}_0$	1.42802	1.42802	0.08842	0.06765
	REQM	$\hat{\alpha}_0$	0.34071	0.14563	0.11510	0.12161
		$\hat{\alpha}_1$	0.36466	0.36466	0.36832	0.36740
		$\hat{\beta}_1$	0.41005	0.15640	0.12317	0.13010
		$\hat{\gamma}_0$	1.43516	1.43516	0.83537	0.80419
	TX Cobertura 95%	α_0	0.02000	77.92000	89.96000	90.10000
		α_1	92.62000	89.66000	90.72000	90.32000
		β_1	0.00000	72.64000	88.00000	88.26000
		γ_0	0.00000	0.00000	81.04000	78.42000
$X \sim Beta(\mu_x \sigma_x^2, (1 - \mu_x) \sigma_x^2)$	Viés	$\hat{\alpha}_0$	-0.30194	0.75060	0.00572	0.02896
		$\hat{\alpha}_1$	0.04666	0.04666	-0.00618	-0.00504
		$\hat{\beta}_1$	0.62513	-1.44785	-0.00812	-0.05513
		$\hat{\gamma}_0$	1.07421	1.07421	-0.48907	-0.32263
	REQM	$\hat{\alpha}_0$	0.30433	36.03234	0.18670	0.40174
		$\hat{\alpha}_1$	0.27952	0.27952	0.28661	0.28202
		$\hat{\beta}_1$	0.62620	69.23198	0.35837	0.79878
		$\hat{\gamma}_0$	1.08181	1.08181	1.36354	1.14602
	TX Cobertura 95%	α_0	0.00000	52.54000	81.66000	88.66000
		α_1	94.72000	92.38000	86.82000	94.00000
		β_1	0.00000	45.88000	77.12000	87.20000
		γ_0	0.00000	0.00000	72.96000	76.12000

verossimilhança (MVa) e máxima pseudo-verossimilhança (MPVa) apresentam os melhores

desempenhos. Em relação a raiz do erro quadrático médio (REQM) tem-se que, para o estimador associada a variável com erro de medida β_1 , o método MPVa apresentou o melhor resultado (REQM = 0.10847). Em relação à taxa de cobertura, considerando o nível de confiança 95%, verificamos que métodos MVa e MPVa apresentam os melhores desempenhos para os parâmetros associados a variável com erros de medidas e a variável associada a dispersão, dadas respectivamente por, β_1 e γ_0 . Ou seja, mesmo a distribuição da variável com erro de medida sendo considerada erroneamente, os métodos MPVa e MVa apresentam os melhores resultados quanto à estimação.

Para $X \sim LN(2.5, 2.7)$ avaliando o viés é possível notar que o método de MVa apresenta os menores valores, especialmente para os parâmetros associados às variáveis não linear, com erro de medida e dispersão, cujos resultados foram $Viés(\alpha_1) = 0.19999$, $Viés(\beta_1) = 0.19415$ e $Viés(\gamma_0) = 0.30781$, respectivamente. Em relação a raiz do erro quadrático médio (REQM), o método MVa apresentou melhores resultados. Em relação à taxa de cobertura, considerando o nível de confiança 95%, verificamos que métodos MVa e MPVa apresentam os melhores desempenhos para todos os parâmetros. De modo geral, verificou-se que o método *naive* apresentou os piores desempenhos considerando os métodos de avaliação dos estimadores (viés, REQM, TX cobertura).

Quando $X \sim Weibull(1.5, 1.0)$ observou-se que, em geral, os métodos MVa e MPVa apresentaram os melhores desempenhos. Vale salientar também que, em relação à taxa de cobertura com 95% de confiança, o método *naive* apresenta os piores desempenhos, especialmente para os parâmetros associados às variáveis com erro de medida (β_1) e dispersão (γ).

Considerando que $X \sim Beta(0.6, 0.6)$, verificou-se novamente de forma geral melhores desempenhos para os métodos MVa e MPVa. Em relação à taxa de cobertura considerando 95% de confiança, é possível observar que o método *naive* apresentou os piores resultados, há uma superestimação em relação aos parâmetros associados as variáveis com erros de medida e dispersão, β_1 e γ_0 , respectivamente.

A Tabela 5.8 apresenta os vieses, raízes do erro quadrático médio (REQM) e taxas de cobertura com 95% de confiança para os estimadores de α_0 , α_1 , β_1 e γ_0 considerando o modelo com dispersão constante, utilizando a distribuição normal. Avaliando o viés é possível notar que o método da MPVa apresenta os menores valores, especialmente para os parâmetros associados às variáveis não linear, com erro de medida e dispersão, cujos resultados foram $Viés(\alpha_1) = -0.02155$, $Viés(\beta_1) = -0.00345$ e $Viés(\gamma_0) = -0.46810$, respectivamente. Em relação a raiz do erro quadrático médio (REQM), o método MPVa apresentou melhores resultados. Em relação à taxa de cobertura, considerando o nível de confiança 95%, verificamos que métodos MVa e MPVa apresentam os melhores desempenhos para os parâmetros associados a não linearidade (β_1) e dispersão (γ_0). De modo geral,

verificou-se que o método *naive* apresentou os piores desempenhos considerando os métodos de avaliação dos estimadores (viés, REQM).

Com o objetivo de avaliar a qualidade dos valores preditos utilizou-se as estatísticas P^2 e $PRESS$. Além disso a pseudo- R^2 corrigida (R_c^2) e o R_{LR}^2 corrigido (R_{LRc}^2) foram usados para avaliar a qualidade do ajuste do modelo.

Tabela 5.8: Vieses, raízes dos erros quadráticos médios (REQM) e taxas de cobertura com 95% de confiança para os estimadores de α_0 , α_1 , β_1 e γ_0 considerando o modelo com dispersão constante. Verdadeiros valores de parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$ e $\gamma_0 = 5.0$.

Cenário	Método		<i>Naive</i>	RC	MVa	MPVa
$X \sim N(\mu_x, \sigma_x^2)$	Viés	$\hat{\alpha}_0$	0.01044	0.01058	0.00279	0.00278
		$\hat{\alpha}_1$	0.11788	0.11788	-0.02219	-0.02155
		$\hat{\beta}_1$	0.22541	0.02545	-0.00379	-0.00345
		$\hat{\gamma}_0$	1.77119	1.77119	-0.48565	-0.46810
	REQM	$\hat{\alpha}_0$	0.04988	0.05512	0.05661	0.05645
		$\hat{\alpha}_1$	0.42165	0.42165	0.42611	0.42490
		$\hat{\beta}_1$	0.22854	0.06878	0.06539	0.06376
		$\hat{\gamma}_0$	1.77680	1.77680	1.19652	1.16958
	TX Cobertura 95%	α_0	94.76000	91.16000	88.90000	90.60000
		α_1	90.60000	89.22000	82.64000	83.60000
		β_1	0.00000	83.94000	89.52000	92.38000
		γ_0	0.00000	0.00000	72.64000	69.04000

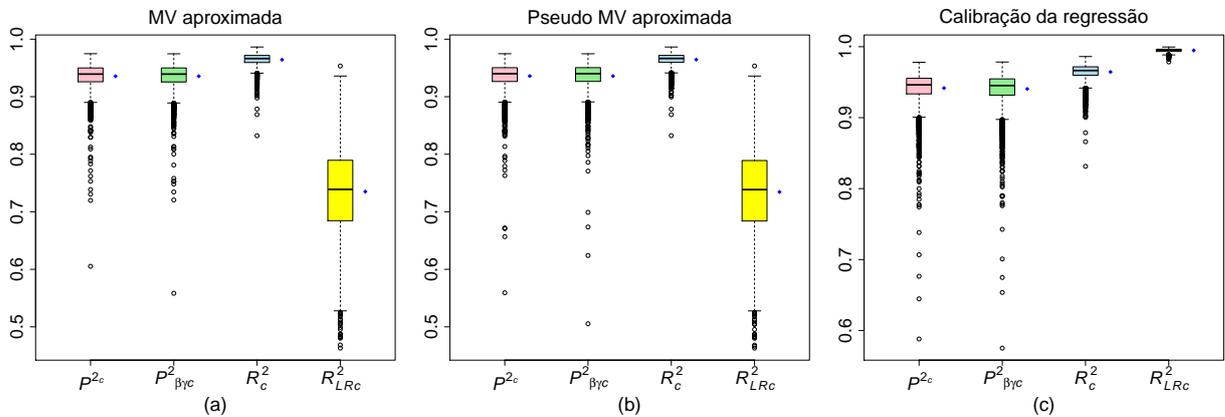


Figura 5.11: *Boxplots* das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$), R_c^2 e R_{LR}^2 corrigida (R_{LRc}^2) obtidas pelos métodos: (a) máxima verossimilhança aproximada, (b) pseudo-verossimilhança aproximada e (c) calibração da regressão. Coeficiente de confiabilidade $k_x = 0.75$. Modelo com dispersão constante. $X \sim t(3)$

As Figuras 5.11, 5.12, 5.13 e 5.14 apresentam os *boxplots* das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$), R_c^2 e R_{LR}^2 corrigida (R_{LRc}^2) obtidas pelos métodos da máxima verossimilhança aproximada (MV aproximada), pseudo-verossimilhança aproximada (Pseudo MV aproximada) e calibração da regressão (RC), considerando que $X \sim t(3)$, $X \sim LN(2.5, 2.7)$, $X \sim Weibull(1.5, 1.0)$ e $X \sim Beta(0.6, 0.6)$. Aqui, o principal objetivo foi realizar uma

análise mais detalhada das estatísticas de medidas de qualidade de ajuste propostas para o modelo de regressão beta não linear com erros de medida quando existe má especificação no modelo.

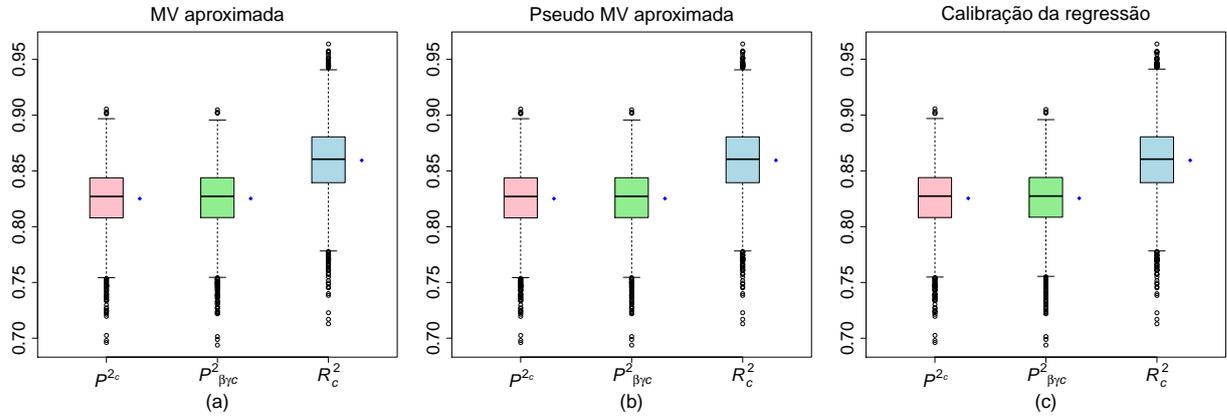


Figura 5.12: *Boxplots* das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$), R_c^2 e R_{LR}^2 corrigida (R_{LRc}^2) obtidas pelos métodos: (a) máxima verossimilhança aproximada, (b) pseudo-verossimilhança aproximada e (c) calibração da regressão. Coeficiente de confiabilidade $k_x = 0.75$. Modelo com dispersão constante. $X \sim LN(2.5, 2.7)$

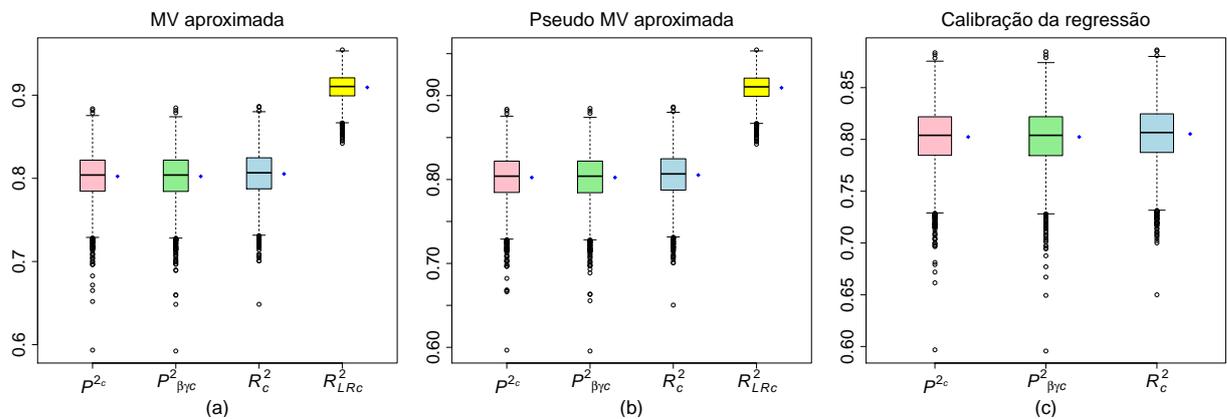


Figura 5.13: *Boxplots* das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$), R_c^2 e R_{LR}^2 corrigida (R_{LRc}^2) obtidas pelos métodos: (a) máxima verossimilhança aproximada, (b) pseudo-verossimilhança aproximada e (c) calibração da regressão. Coeficiente de confiabilidade $k_x = 0.75$. Modelo com dispersão constante. $X \sim Weibull(1.5, 1.0)$

Com base nas Figuras de 5.11 a 5.12 notamos que mais uma vez as medidas de predição assumem os menores valores. No entanto, esses menores valores ocorrem em especial quando a verdadeira distribuição é a distribuição *t-Student* com três graus de liberdade. Quanto às outras distribuições tanto o poder preditivo do modelo quanto a qualidade do ajuste são menos afetadas.

A Figura 5.15 apresenta os *boxplots* das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$), R_c^2 e R_{LR}^2 corrigida

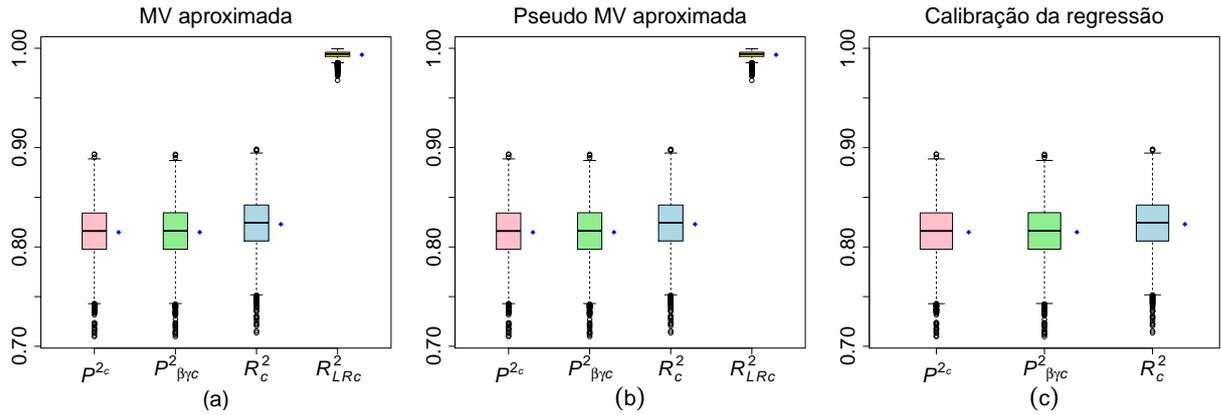


Figura 5.14: *Boxplots* das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$), R_c^2 e R_{LR}^2 corrigida (R_{LRc}^2) obtidas pelos métodos: (a) máxima verossimilhança aproximada, (b) pseudo-verossimilhança aproximada e (c) calibração da regressão. Coeficiente de confiabilidade $k_x = 0.75$. Modelo com dispersão constante. $X \sim \text{Beta}(0.6, 0.6)$

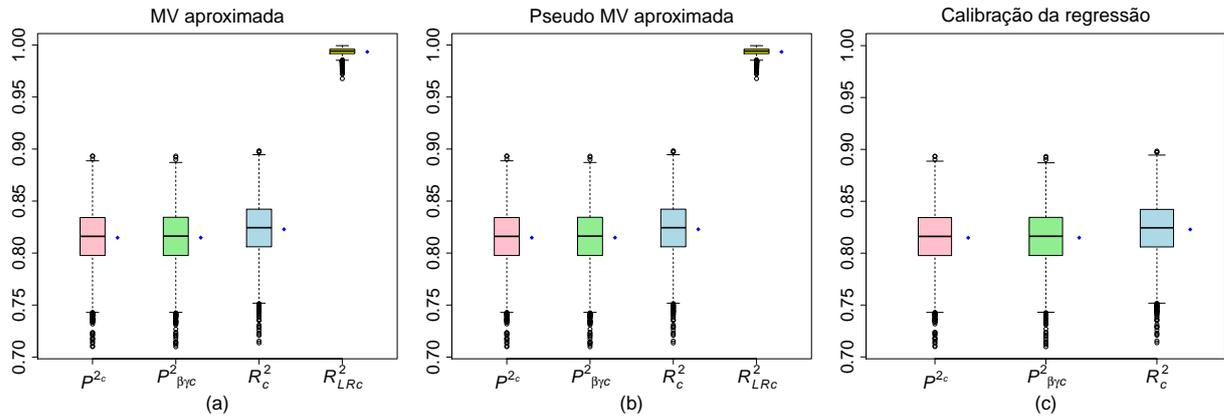


Figura 5.15: *Boxplots* das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$), R_c^2 e R_{LR}^2 corrigida (R_{LRc}^2) obtidas pelos métodos: (a) máxima verossimilhança aproximada, (b) pseudo-verossimilhança aproximada e (c) calibração da regressão. Coeficiente de confiabilidade $k_x = 0.75$. Modelo com dispersão constante. $X \sim N(0.0, 1.0)$

(R_{LRc}^2) obtidas pelos métodos da máxima verossimilhança aproximada (MV aproximada), pseudo-verossimilhança aproximada (Pseudo MV aproximada) e calibração da regressão (RC), considerando que $X \sim N(0.0, 1.0)$. É possível notar que as medidas de predição estão similares aos resultados verificados anteriormente. Em relação à medida de qualidade de ajuste, é possível notar que os métodos da MV aproximada e Pseudo MV aproximada são métodos que ajustam bem o modelo.

Avaliamos também os desempenhos dos diferentes métodos considerando heteroscedasticidade para os erros, ou seja, $\text{Var}(e_t|x_t) = |x_t - \mu|^{1/2}\sigma_e^2$. Ou seja, a inferência é baseada na suposição errônea de erros homoscedásticos.

A Tabela 5.9 apresenta os resultados dos vieses, das raízes do erro quadrático médio

(REQM) e as taxas de cobertura com 95% de confiança para os estimadores de α_0 , α_1 , β_1 e γ_0 considerando o modelo de regressão beta não linear com erros de medida com dispersão constante dado na Equação (5.18), considerando heteroscedasticidade nos erros. Em relação ao viés, o método da máxima verossimilhança aproximada (MVa) tem o melhor desempenho. Considerando a taxa de cobertura, o método *naive* apresentou o pior desempenho quando se considera os parâmetros associados a variável com erros de medida e a dispersão.

Tabela 5.9: Vieses, raízes dos erros quadráticos médios (REQM) e taxas de cobertura com 95% de confiança para os estimadores de α_0 , α_1 , β_1 e γ_0 considerando o modelo com dispersão constante. Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$ e $\gamma_0 = 5.0$.

		Método			
		Naive	RC	MVa	MPVa
Viés	$\hat{\alpha}_0$	0.00739	0.00639	-0.00141	-0.00124
	$\hat{\alpha}_1$	0.09143	0.09143	-0.06367	-0.06258
	$\hat{\beta}_1$	0.18385	-0.04794	-0.03288	-0.04038
	$\hat{\gamma}_0$	1.63978	1.63978	-1.23717	-1.21482
REQM	$\hat{\alpha}_0$	0.04710	0.05321	0.05524	0.05529
	$\hat{\alpha}_1$	0.39749	0.39749	0.42679	0.42605
	$\hat{\beta}_1$	0.18807	0.09574	0.07538	0.08114
	$\hat{\gamma}_0$	1.64741	1.64741	1.60203	1.56369
TX Cobertura 95%	α_0	94.80000	90.30000	84.80000	87.90000
	α_1	92.00000	90.30000	71.00000	74.70000
	β_1	0.90000	81.90000	80.50000	88.20000
	γ_0	0.00000	0.00000	51.70000	48.50000

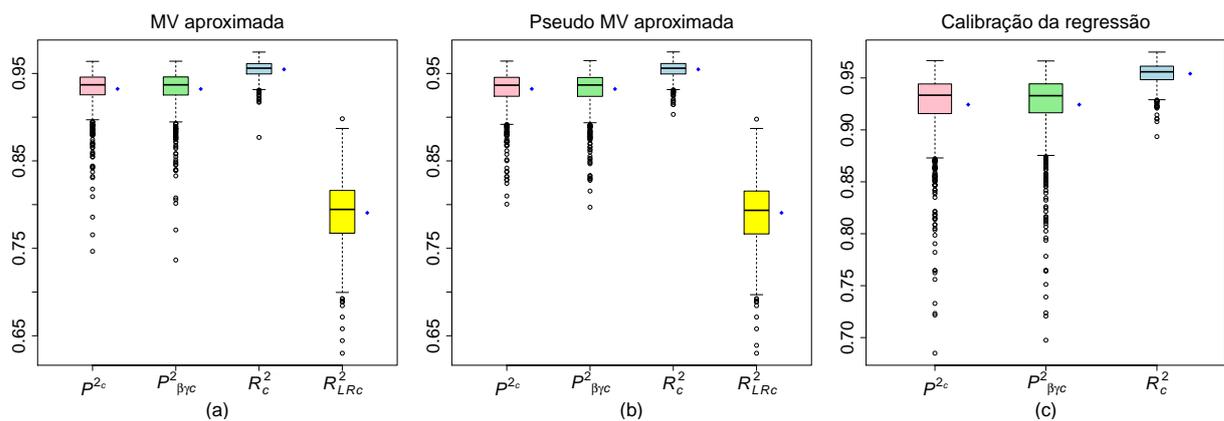


Figura 5.16: Boxplots das estatísticas P^2 corrigida considerando o resíduo ponderado (P_c^2), P^2 corrigida considerando o resíduo combinado ($P_{\alpha\gamma c}^2$), R_c^2 e R_{LR}^2 corrigida (R_{LRc}^2) obtidas pelos métodos: (a) máxima verossimilhança aproximada, (b) pseudo-verossimilhança aproximada e (c) calibração da regressão. Coeficiente de confiabilidade $k_x = 0.75$. Avaliação da homoscedasticidade.

A Figura 5.16 apresenta os resultados das estatísticas P^2 e pseudo- R^2 e suas versões corrigidas. Podemos verificar que o modelo está bem ajustado, mesmo com a suposição de

homoscedasticidade sendo violada. Um comportamento similar foi verificado no modelo de regressão beta linear com erros de medida (Carrasco *et al.*, 2014).

5.5 Conclusões

Neste capítulo propomos as estatísticas de predição P^2 considerando o resíduo ponderado (P^2), P^2 considerando o resíduo combinado ($P_{\alpha\gamma}^2$) e suas versões corrigidas P_c^2 e $P_{\alpha\gamma c}^2$, respectivamente, para o modelo de regressão beta não linear com erros de medida. Também avaliamos as medidas de qualidade de ajuste pseudo- R^2 (ps- R^2) e pseudo- R^2 baseada na razão de verossimilhança (ps- R_{LR}^2) e suas versões corrigidas ps- R_c^2 e ps- R_{LRc}^2 , respectivamente.

Realizamos simulações de Monte Carlo para avaliar os desempenhos das estatísticas de predição e das medidas de qualidade de ajuste considerando os métodos estudados. Concluimos que as estatísticas P^2 e $P_{\alpha\gamma}^2$ e suas versões corrigidas indicam que à medida que o coeficiente de confiabilidade diminui a qualidade da predição do modelo é substancialmente afetada negativamente mesmo quando o modelo está bem especificado. Verificamos também que a medida R_{LR}^2 e sua versão corrigida não detecta nenhum tipo de problema. Além disso, no modelo com dispersão constante as duas medidas de qualidade de ajuste pseudo- R^2 e R_{LR}^2 apresentam comportamento equivalentes.

6 APLICAÇÕES

6.1 Aplicação 1 - Modelo Linear: Sobrepeso em crianças considerando o peso mensurado duas vezes.

O conjunto de dados utilizado é oriundo de uma pesquisa da Faculdade de Medicina e do Departamento de Nutrição da Universidade Federal da Bahia. Nesta pesquisa 1439 crianças de 9 escolas públicas foram avaliadas. Foram analisadas duas medidas de altura, peso e circunferência da cintura para todas as crianças.

Com o objetivo de avaliar o sobrepeso em crianças, analisamos uma única escola, sorteada aleatoriamente, Colégio Estadual Arlete Magalhães, totalizando uma amostra de 166 crianças. A variável de interesse é a razão entre a circunferência da cintura (cm) e a altura (cm) que assume valores de forma contínua no intervalo (0,1), denotada por y_t . Esta variável é considerada um indicador de obesidade abdominal.

Foram consideradas as segundas medidas mensuradas.

Para este conjunto de dados foram avaliadas conjuntamente com a variável de interesse as variáveis: idade, em anos (z_{t1}), a indicadora do estado nutricional: 0 = magreza/eutrofia e 1 = sobrepeso/obesidade (z_{t2}), gênero da criança: 0 = feminino e 1 = masculino (z_{t3}) e peso (x_t). A variável peso é considerada uma variável mensurada com erro de medida.

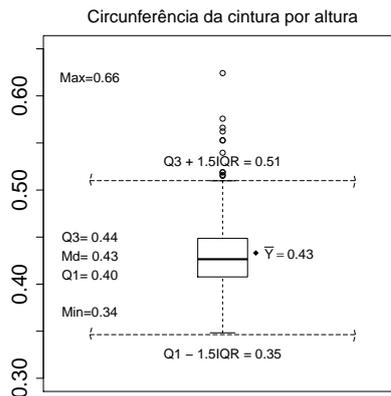


Figura 6.1: *Boxplot* da variável de interesse (y_t) razão entre a circunferência da cintura (cm) e a altura (cm).

Com base na Figura 6.1, nota-se que a variável resposta apresenta valores concentrados próximos ao valor médio do intervalo unitário, 0.50, sendo que 75% estão acima de 0.40 e o valor máximo é de 0.66. Note que o valor médio e a mediana são idênticos, $\bar{Y} = Md = 0.43$. Vale salientar que existem várias observações *outliers* no conjunto de dados.

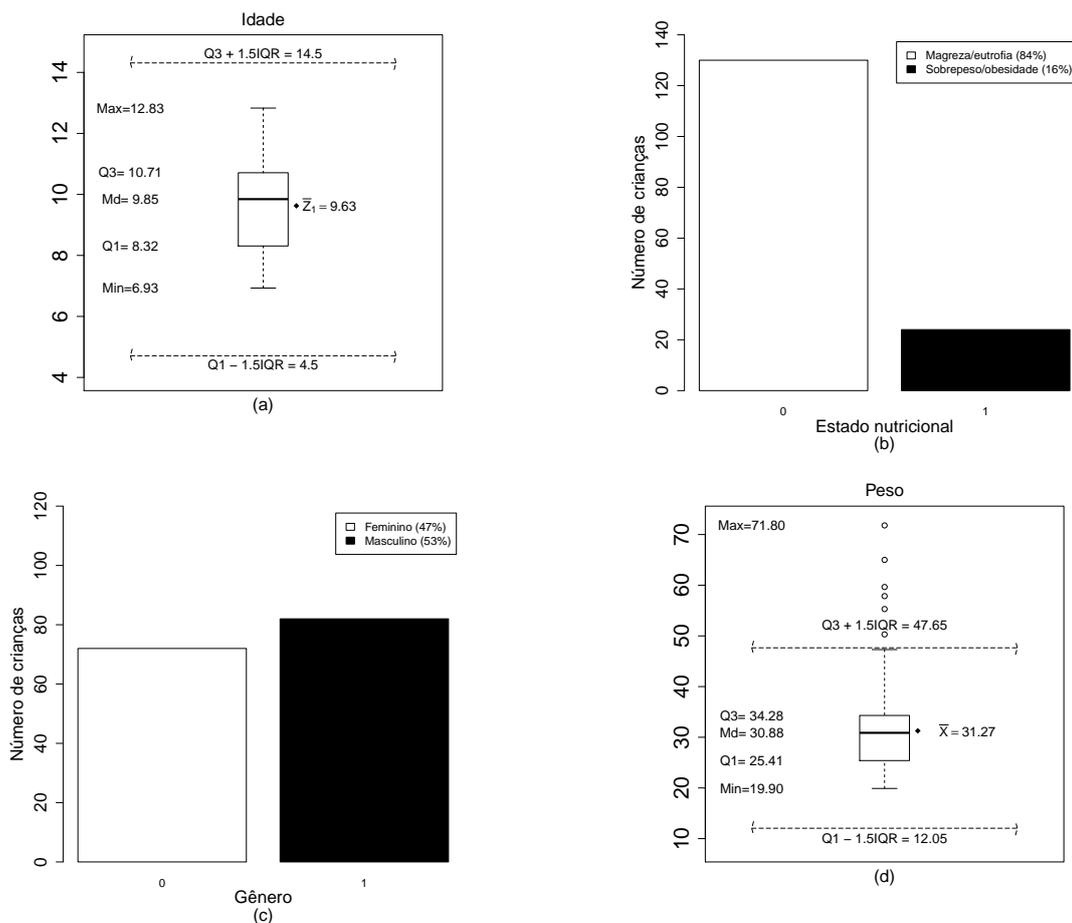


Figura 6.2: *Boxplots* com as medidas resumo das covariáveis idade (a) e peso (d) e gráficos de colunas com as porcentagens das covariáveis estado nutricional (b) e gênero (c) das 166 crianças.

A Figura 6.2 apresenta as medidas resumo das variáveis analisadas. Note que 25% das crianças tem idade acima de 8.32 anos, variando de 6.93 a 12.83 anos, Figura 6.22 (a). 84% das crianças apresentam o estado nutricional de magreza/eutrofia e 53% são do gênero masculino, como mostram as Figuras 6.22 (b) e 6.22 (c), respectivamente. Note que 25% das crianças tem peso acima de 34.26 kg, com mínimo e máximo iguais 19.90 kg e 71.80 kg, respectivamente. Vale salientar que, existem várias observações *outliers*.

Como dito anteriormente, a variável peso é uma variável mensurada com erro de medida. Além disso, a mesma foi mensurada duas vezes. Portanto, a variância dos erros de medidas é estimada pela Equação (2.7), ou seja,

$$\hat{\sigma}_{ei}^2 = \frac{1}{\sum_{t=1}^n (n_{ti} - 1)} \sum_{t=1}^n \sum_{j=1}^{n_{ti}} (w_{tij} - \bar{w}_{ti.})^2,$$

em que

$$\bar{w}_{ti.} = \frac{1}{n_{ti}} \sum_{j=1}^{n_{ti}} w_{tij}.$$

Desta forma, $\hat{\sigma}_e^2 = 66.838 \text{ kg}^2$. Assim, o coeficiente de confiabilidade é dado por $k_x = \sigma_x^2 / (\sigma_x^2 + \sigma_e^2)$. Considerando os dados analisados temos que $\hat{\sigma}_x^2 = 69.786 \text{ kg}^2$, que se refere à variância dos pesos mensurados na primeira vez. Portanto $\hat{k}_x = 0.512$. Ou seja, o erro de medida é considerado alto.

Com o objetivo de analisar o sobrepeso das crianças em função das covariáveis estudadas, ajustamos modelos de regressão beta lineares com erros de medida.

6.1.1 Modelos ajustados

Inicialmente, ajustou-se o modelo de regressão beta com erros de medidas considerado todas as covariáveis dado por

$$\begin{aligned} \log\left(\frac{\mu_t}{1 - \mu_t}\right) &= \alpha_0 + \alpha_1 z_{t1} + \alpha_2 z_{t2} + \alpha_3 z_{t3} + \beta_1 x_t, \\ \log(\phi_t) &= \gamma_0 + \gamma_1 z_{t1} + \gamma_2 z_{t2} + \gamma_3 z_{t3} + \lambda_1 x_t \end{aligned}$$

em que $t = 1, \dots, 166$, z_{t1} denota a idade da criança, z_{t2} é a indicadora do estado nutricional da criança, z_{t3} denota o gênero da criança e x_t representa o peso. Consideramos as funções logito e logarítmica para os submodelos da média e da dispersão, respectivamente, e utilizamos $Q = 50$ pontos de quadratura.

As estimativas, erros-padrão e p -valores relativos ao processo de inferência sobre os parâmetros obtidos a partir dos métodos de *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa), pseudo-verossimilhança aproximada (MPVa) são apresentadas na Tabela 6.1.

É possível observar que a variável gênero das crianças (z_{t3}) não foi significativa, ao nível de 10%, no submodelo da média em nenhum dos métodos ajustados, exceto no método da calibração da regressão (RC) em que o p -valor foi 0.038. Em relação ao submodelo da dispersão temos que as variáveis idade (z_{t1}) e peso (x_t) não foram significativas, ao nível de 10%, quando consideramos os métodos *naive*, MVa e MPVa. Note que há diferenças nas estimativas dos parâmetros de acordo com os métodos utilizados, especialmente para o método RC.

Baseado nos resultados da Tabela 6.1, consideramos um novo modelo de regressão beta linear com erros de medidas com as variáveis significativas, ao nível de 10%. Assim, o modelo é dado por

$$\begin{aligned} \log\left(\frac{\mu_t}{1 - \mu_t}\right) &= \alpha_0 + \alpha_1 z_{t1} + \alpha_3 z_{t3} + \beta_1 x_t \\ \log(\phi_t) &= \gamma_0 + \gamma_2 z_{t2} + \gamma_3 z_{t3}, \end{aligned} \tag{6.1}$$

em que $t = 1, \dots, 166$, z_{t1} denota a idade da criança, z_{t2} é a indicadora do estado nutricional da criança, z_{t3} denota o gênero das crianças e x_t representa o peso.

Tabela 6.1: Estimativas, erros-padrão e p -valores obtidos pelos métodos *Naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa). Considerando o modelo de regressão beta linear com erros de medida com todas as covariáveis.

Método	Parâmetro	Estimativa	Erro-padrão	p -valor
<i>Naive</i>	α_0	-0.148	0.051	0.004
	α_1	-0.036	0.007	0.000
	α_2	0.260	0.041	0.000
	α_3	0.007	0.016	0.652
	β_1	0.006	0.002	0.001
	γ_0	6.854	0.833	0.000
	γ_1	-0.032	0.097	0.740
	γ_2	-1.204	0.391	0.002
	γ_3	0.466	0.231	0.044
	λ_1	-0.017	0.018	0.333
RC	α_0	-0.740	0.128	0.000
	α_1	0.048	0.017	0.005
	α_2	-0.035	0.069	0.611
	α_3	0.042	0.020	0.038
	β_1	0.016	0.003	0.000
	γ_0	7.033	0.460	0.000
	γ_1	0.038	0.085	0.650
	γ_2	-0.106	0.449	0.814
	γ_3	-1.686	0.569	0.003
	λ_1	0.492	0.114	0.000
MVa	α_0	-0.173	0.054	0.001
	α_1	-0.036	0.007	0.000
	α_2	0.261	0.041	0.000
	α_3	0.007	0.016	0.652
	β_1	0.007	0.002	0.001
	γ_0	7.005	0.886	0.000
	γ_1	-0.046	0.100	0.648
	γ_2	-1.266	0.404	0.002
	γ_3	0.479	0.241	0.047
	λ_1	-0.017	0.021	0.417
MPVa	α_0	-0.173	0.051	0.001
	α_1	-0.036	0.007	0.000
	α_2	0.261	0.043	0.000
	α_3	0.007	0.016	0.656
	β_1	0.007	0.002	0.007
	γ_0	7.006	0.766	0.000
	γ_1	-0.046	0.094	0.629
	γ_2	-1.266	0.304	0.000
	γ_3	0.479	0.228	0.035
	λ_1	-0.017	0.019	0.383

A Tabela 6.2 apresenta as estimativas, erros-padrão e p -valores relativos ao processo de inferência sobre os parâmetros obtidos a partir dos métodos de *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa), pseudo-verossimilhança aproximada (MPVa) do Modelo (6.1).

Tabela 6.2: Estimativas, erros-padrão e p -valores obtidos pelos métodos *Naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa). Modelo (6.1).

Método	Parâmetro	Estimativa	Erro-padrão	p -valor
<i>Naive</i>	α_0	-0.153	0.052	0.003
	α_1	-0.036	0.006	0.000
	α_2	0.268	0.044	0.000
	β_1	0.006	0.002	0.000
	γ_0	6.049	0.176	0.000
	γ_2	-1.469	0.324	0.000
	γ_3	0.442	0.232	0.057
RC	α_0	-0.765	0.129	0.000
	α_1	0.051	0.017	0.002
	α_2	-0.037	0.072	0.611
	β_1	0.017	0.003	0.000
	γ_0	7.904	0.531	0.000
	γ_2	-0.110	0.361	0.760
	γ_3	-1.481	0.462	0.001
MVa	α_0	-0.180	0.054	0.001
	α_1	-0.036	0.006	0.000
	α_2	0.268	0.044	0.000
	β_1	0.007	0.002	0.000
	γ_0	6.087	0.181	0.000
	γ_2	-1.505	0.325	0.000
	γ_3	0.460	0.243	0.058
MPVa	α_0	-0.180	0.052	0.001
	α_1	-0.036	0.007	0.000
	α_2	0.268	0.041	0.000
	β_1	0.007	0.002	0.001
	γ_0	6.087	0.184	0.000
	γ_2	-1.505	0.275	0.000
	γ_3	0.460	0.232	0.048

Note que todas as covariáveis foram significantes, ao nível de 10% de significância, exceto no método da calibração da regressão (RC) que apresentou os piores erros-padrão.

As Figuras 6.3, 6.4 e 6.5 apresentam os resíduos e os envelopes simulados considerando os métodos calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa), respectivamente.

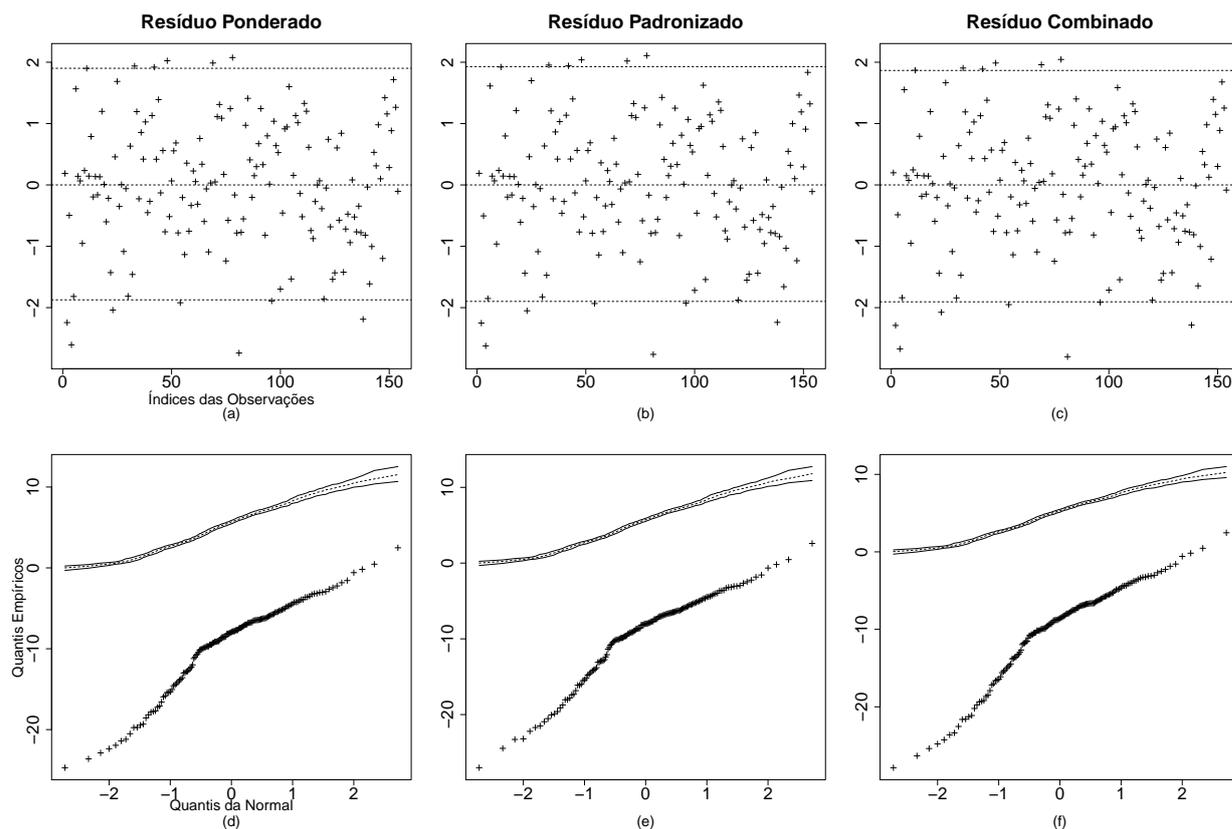


Figura 6.3: Resíduos e envelopes simulados considerando os resíduos ponderados (a) e (d), padronizados (b) e (e) e combinados (c) e (f) utilizando o método da **calibração da regressão** para o ajuste do modelo de regressão beta com erros de medidas. Modelo (6.1).

É interessante notar com base nos gráficos de resíduos que o ajuste realizado no método da calibração de regressão não é adequado aos dados. Ou seja, é possível observar que os resíduos ponderados, padronizados e combinados estão se comportando de forma aleatória, com várias observações fora dos limites considerados, visto nas Figuras 6.3 (a) - (c). Mas, os envelopes simulados apresentados nas Figuras 6.3 (d) - (f) mostram que o modelo de regressão linear com erros de medidas utilizando o método da calibração da regressão não ajusta bem o conjunto de dados. Vale salientar que os demais métodos, MVA e MPVa, apresentam um excelente ajuste.

As Figuras 6.4 (a) - (c) apresentam os resíduos ponderados, padronizados e combinados, respectivamente, considerando o método da máxima verossimilhança aproximada. É possível verificar que os resíduos estão se comportando de forma aleatória, com várias observações fora dos limites considerados. Os envelopes simulados apresentados nas Figuras 6.4 (d) - (f) mostram que o modelo de regressão linear com erros de medidas utilizando o método da máxima verossimilhança aproximada ajusta bem o conjunto de dados.

As Figuras 6.5 (a) - (c) apresentam os resíduos ponderados, padronizados e combinados, respectivamente, considerando o método da máxima pseudo-verossimilhança aproximada. É possível verificar que os resíduos estão se comportando de forma aleatória, com várias

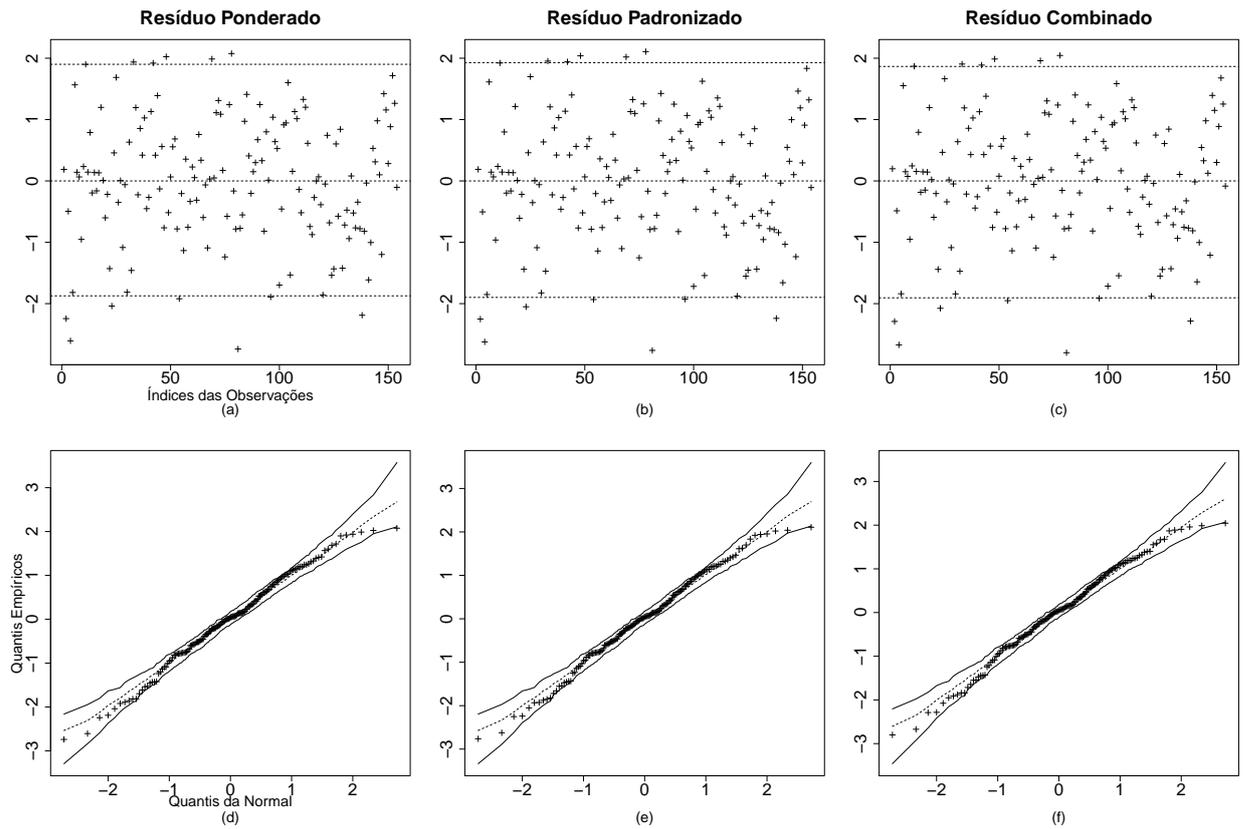


Figura 6.4: Resíduos e envelopes simulados considerando os resíduos ponderados (a) e (d), padronizados (b) e (e) e combinados (c) e (f) utilizando o método da **máxima verossimilhança aproximada** para o ajuste do modelo de regressão beta com erros de medidas. Modelo (6.1).

observações fora dos limites considerados. Os envelopes simulados apresentados nas Figuras 6.5 (d) - (f) mostram que o modelo de regressão linear com erros de medidas utilizando o método da máxima pseudo-verossimilhança aproximada ajusta bem o conjunto de dados.

A Tabela 6.3 apresenta as estatísticas de predição e medidas de qualidade do ajuste dos modelos ajustado considerando os métodos calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa).

Tabela 6.3: Estatísticas de predição e medidas de qualidade do ajuste dos modelos ajustado considerando os métodos calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa).

Medidas	RC	MVa	MPVa
P^2	-7.22870	0.71254	0.71254
P_c^2	-7.56460	0.70080	0.70080
$P_{\alpha\gamma}^2$	-6.73750	0.71154	0.71154
$P_{\alpha\gamma c}^2$	-7.05330	0.69977	0.69977
ps- R^2	0.05573	0.60405	0.60405
ps- R_c^2	0.01719	0.58789	0.58789
R_{LR}^2	0.99984	0.67998	0.67998
R_{LRc}^2	0.99983	0.66139	0.66139

Tanto com base nos gráficos de resíduos quanto com base nas medidas de predição e nas medidas de qualidade de ajuste vemos que o modelo com erro de medida proposto está adequado aos dados em especial quando os métodos de estimação são de máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa), os quais apresentam desempenhos equivalentes. É interessante notar que a falta de qualidade de ajuste do modelo baseado na estimação por calibração de regressão evidenciada pelos gráficos normais de probabilidade com envelopes simulados é confirmada pela medida $ps-R_2$. Também o modelo preditivo do modelo é consideravelmente afetado quando o método de estimação utilizado é o de calibração de regressão ao ponto da medida P_2 ser aproximadamente igual a -7.23 .

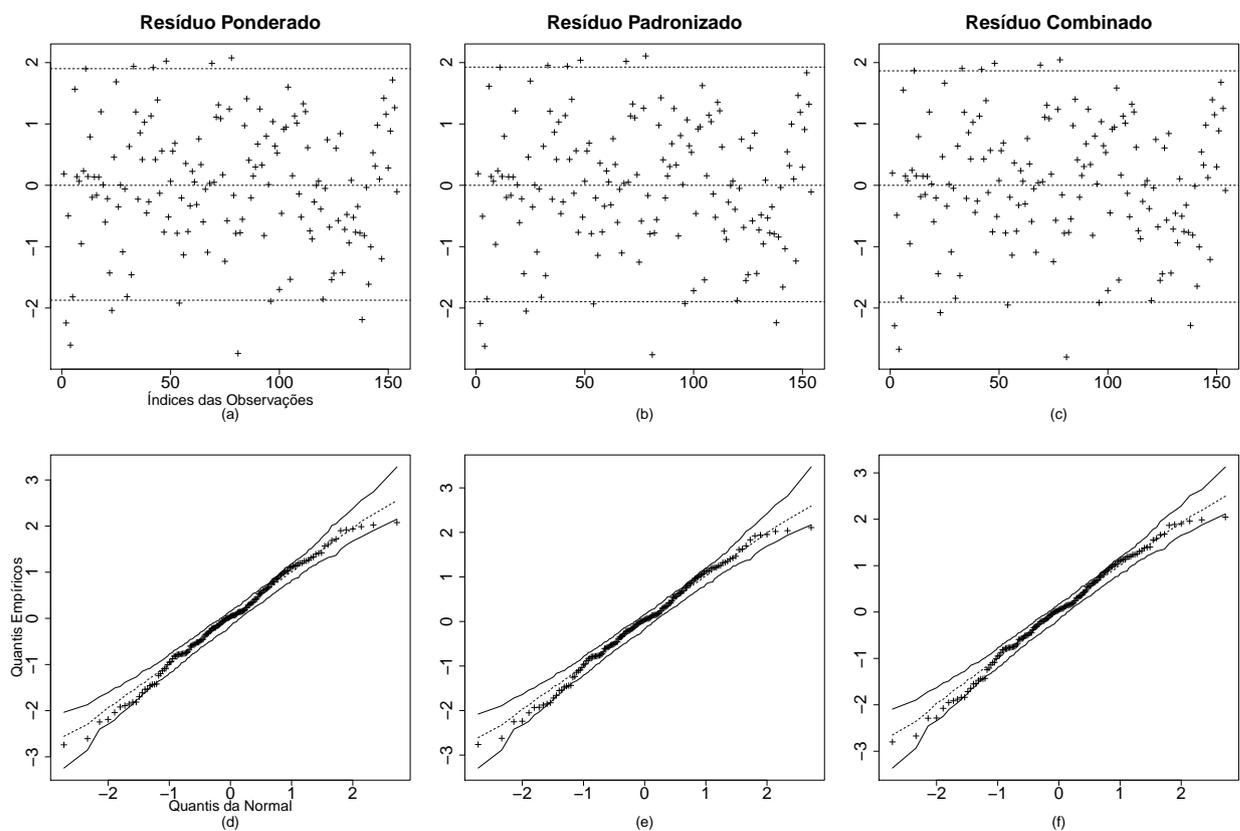


Figura 6.5: Resíduos e envelopes simulados considerando os resíduos ponderados (a) e (d), padronizados (b) e (e) e combinados (c) e (f) utilizando o método da **máxima pseudo-verossimilhança aproximada** para o ajuste do modelo de regressão beta com erros de medidas. Modelo (6.1).

6.2 Aplicação 2 - Modelo Não Linear: Dados de craqueamento catalítico fluido (FCC)

O processo craqueamento catalítico fluido, conhecido como “*Fluid Catalytic Cracking*” (FCC), é usado para converter hidrocarbonetos de alto peso molecular em pequenas moléculas de maior valor comercial, através do contato destes com um catalisador. O processo FCC é muitas vezes considerado o coração de uma refinaria, uma vez que permite que a produção seja adaptada aos produtos de maior demanda e/ou alta rentabilidade. O catalisador do processo consiste em partículas finas de 10 a 150 microns, facilmente fluidizável tendo como componente principal o zeólito Y incorporado em uma matriz amorfa de aluminossilicato e argila (Salazar, 2005).

Sabe-se que cada 1000 ppm de vanádio no catalisador a produção de gasolina diminui em cerca de 2.3%. Além disso, este componente químico é conhecido por participar da destruição do catalisador, reduzindo a superfície ativa, a seletividade e a cristalinidade do zeólito Y especialmente na presença de vapor. O vanádio é depositado na superfície externa das partículas do catalisador no reator da unidade FCC, esses complexos sofrem decomposição parcial e são transferidos para o regenerador onde eles são queimados com a coca e o vanádio é oxidado. Esta reação depende da temperatura do regenerador que deve ser próximo a $720\text{ }^{\circ}\text{C}$ (Salazar, 2005).

O vanádio é um elemento químico, símbolo V, número atômico 23 (23 prótons e 23 elétrons) de massa atômica 51 u que, nas condições ambientes, é encontrado no estado sólido. Ele é um metal de transição mole, dúctil de cor cinzenta e brilhante. Apresenta alta resistência ao ataque das bases, ao ácido sulfúrico (H_2SO_4) e ao ácido clorídrico (HCl). É obtido a partir de diversos minerais, até do petróleo. Também pode ser obtido da recuperação do óxido de vanádio em pó procedente de processos de combustão.

Como todo componente químico, o vanádio pode ser considerado como uma variável mensurada com erros de medida. Para tanto, é necessário determinar os parâmetros de perturbação μ_x , σ_x^2 e σ_e^2 . Salazar (2015) mostra que a determinação do vanádio nas amostras impregnadas foi realizada pelo método espectrofotométrico, com a formação de um complexo entre vanádio e ácido fosfotungstíco, denotado por método 1. Aqui o vanádio denota a variável mensurada com erros de medida x_t e a medição foi feita a 370 nm.

Uma curva de calibração foi preparada utilizando metavanadato de amônio NH_4VO_3 como padrão (Tabela 6.4). O estudo realizado considerou 5 amostras de concentração de vanádio (w_t), denotado aqui por método 2, como mostra a Tabela 6.5. Estes resultados foram obtidos do anexo 5 em Salazar (2005) que apresenta a determinação da concentração de vanádio em amostras impregnadas.

O cálculo utilizado para a obtenção do mmol vanádio/ mmol alumínio nas amostras é

Tabela 6.4: Curva de calibração. Método 1.

Concentração ppm vanádio ± 0.2022	Absorvância
0.8720	0.0447
2.1700	0.1067
4.3500	0.2189
6.5300	0.3118
8.7100	0.4174

Tabela 6.5: Concentração de vanádio, absorvância, peso do zeólito sem água e relação entre vanádio e alumínio (mmol) das 5 amostras estudadas. Método 2.

Amostra	Absorvância	Concentração ppm vanádio ± 0.2022	Peso de zeólito sem água	mmol vanádio/ mmol alumínio
1	0.0406	0.753	0.3826	0.191
2	0.1229	2.480	0.3774	0.640
3	0.2097	4.320	0.3782	1.100
4	0.2749	5.690	0.3788	1.450
5	0.3092	6.420	0.3705	1.630

dado por

$$v_1 = con_vanadio \times 20 \times 0.1000, \quad v_2 = \frac{v_1}{50.94},$$

$$v_3 = \frac{v_2}{peso}, \quad v_4 = \frac{v_3}{0.4048},$$

em que *con_vanadio* denota a concentração ppm de vanádio ± 0.2022 , *peso* é o peso de zeólito sem água e v_4 representa mmol vanádio/mmol alumínio. A Tabela 6.6 apresenta os resultados da conversão da concentração ppm vanádio ± 0.2022 para mmol vanádio/mmol alumínio considerando os dois métodos avaliados, em que *mmol1* e *mmol2* representam mmol vanádio/mmol alumínio para os métodos 1 e 2, respectivamente.

Tabela 6.6: Conversão da concentração ppm vanádio ± 0.2022 para mmol vanádio/mmol alumínio.

Método 1	Método 2	<i>mmol1</i>	<i>mmol2</i>
0.872	0.753	0.221	0.191
2.170	2.480	0.558	0.640
4.350	4.320	1.116	1.100
5.530	5.690	1.672	1.450
8.710	6.420	2.280	1.630

A Figura 6.6 apresenta a relação entre a mmol vanádio/mmol alumínio (método 1) e mmol vanádio/mmol alumínio (método 2). É possível notar uma relação linear entre as duas variáveis com correlação de 0.98. Assim, parece razoável fazer a regressão linear

considerando o modelo

$$w_t = \tau_0 + \tau_1 x_t + e_t, \quad (6.2)$$

em que $t = 1, \dots, n$, $x_t \sim N(\mu_x, \sigma_x^2)$ e $e_t \sim N(0, \sigma_e^2)$. As estimativas e os respectivos erros-padrão de τ_0 , τ_1 e σ_e^2 são obtidos a partir do modelo (6.2) e são dados por: $\hat{\tau}_0 = 0.181$ (0.123), $\hat{\tau}_1 = 0.713$ (0.089) e $\hat{\sigma}_e^2 = 0.022$, respectivamente.

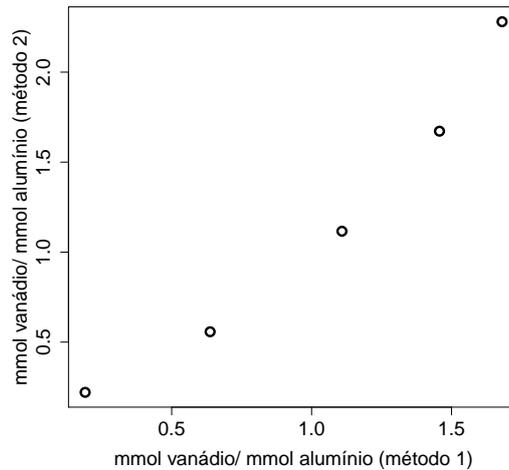


Figura 6.6: Gráfico de dispersão da concentração de mmol vanádio/mmól alumínio dos dois métodos.

Como visto anteriormente, o coeficiente de confiabilidade é dado por $k_x = \sigma_x^2 / (\sigma_x^2 + \sigma_e^2)$. Considerando os dados analisados temos que $\hat{\sigma}_x^2 = 0.691$ e $\hat{\sigma}_e^2 = 0.022$, portanto $\hat{k}_x = 0.97$. Ou seja, o erro de medida é considerado baixo.

Aqui, o interesse é modelar a porcentagem de cristalinidade do zeólito Y (y_t) com base em diferentes concentrações de vanádio (x_1) e de vapor d'água (z_1), considerando dois valores para a temperatura do processo (z_2). Espera-se que quanto maior as concentrações de vanádio e vapor d'água a porcentagem de cristalinidade seja menor. O conjunto de dados é composto por 28 observações.

Os *boxplots* apresentam as medidas descritivas das variáveis contínuas analisadas. Com base na Figura 6.7 (a) nota-se que a resposta concentra-se próxima ao extremo superior do intervalo unitário padrão. De fato, 75% das observações são iguais ou superiores a 0.77, primeiro quartil. Adicionalmente, percebe-se a presença de um *outlier*, a observação 28 que assume o menor valor da resposta igual a 0.64. Este *outlier* confirma que a natureza desta variável aleatória é apresentar valores mais próximos de um. Na Figura 6.7 (b) nota-se como é alta a dispersão dos valores de vapor d'água, com ocorrências de valores extremos como o zero associado ao fato que 50% das observações são maiores que 35.8. Além disso, nota-se que 25% da concentração de vanádio é 0%, ou seja, não há concentração de vanádio,

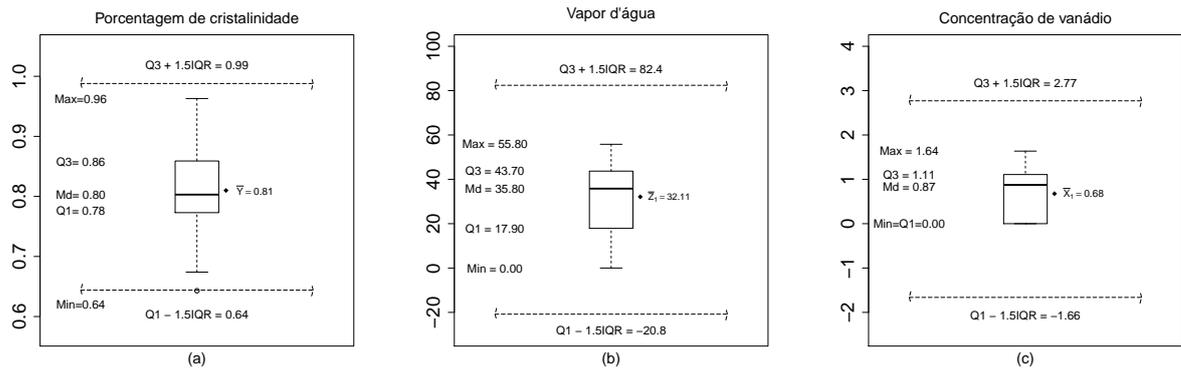


Figura 6.7: *Boxplots* com as medidas descritivas da percentagem de cristalinidade (a), do vapor d'água (b) e da concentração de vanádio (c). Dados de craqueamento catalítico fluido (FCC).

como mostra a Figura 6.7 (c). Vale salientar que 50% das temperaturas são de 700 °C e 50% das temperaturas são de 760 °C.

6.2.1 Modelos ajustados

Inicialmente, ajustou-se o modelo de regressão beta linear com erros de medida com dispersão variável considerando os métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa).

O modelo de regressão beta com erros de medida considerando a dispersão variável é dado por

$$\begin{aligned}
 y_t &\sim \text{Beta}(\mu_t, \phi_t) \\
 \log\left(\frac{\mu_t}{1 - \mu_t}\right) &= \alpha_0 + \alpha_1 z_{1t} + \alpha_2 z_{2t} + \beta_1 x_{1t}, \\
 \log(\phi_t) &= \gamma_0 + \gamma_1 z_{1t}, \quad t = 1, \dots, 28,
 \end{aligned} \tag{6.3}$$

em que x_{1t} representa a concentração de vanádio, z_{1t} denota vapor d'água e z_{2t} é a variável indicadora de temperatura ($0 = 700$ °C e $1 = 760$ °C). Aqui, as funções de ligação associadas à média e à dispersão são logito e logaritmo, respectivamente, e utilizamos $Q = 50$ pontos de quadratura.

A Tabela 6.7 apresenta as estimativas, os erros-padrão e os p -valores obtidos pelos métodos avaliados. É possível notar que vapor d'água (z_{1t}) não foi significativa, ao nível de 5%, no parâmetro associado à dispersão em todos os métodos estudados. Também foram ajustados dois modelos com dispersão variável: o primeiro considerando a variável indicadora de temperatura (z_{2t}) e o segundo considerando a variável com erro de medida vanádio (x_{1t}). Os resultados obtidos nos dois modelos com dispersão variável citados acima foram os mesmos do modelo apresentado na Equação (Apli2:Eq1), ou seja, não houve significância das variáveis z_{2t} e x_{1t} no submodelo associado à dispersão, ao nível de 5% de significância.

Tabela 6.7: Estimativas, erros-padrão e p -valores obtidos pelos métodos *Naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa). Modelo beta linear com erros de medida e dispersão variável.

Método	Parâmetro	Estimativa	Erro-padrão	p -valor
<i>Naive</i>	α_0	2.740	0.130	0.000
	α_1	-0.019	0.003	0.000
	α_2	-0.361	0.084	0.000
	β_1	-0.618	0.080	0.000
	γ_0	5.294	0.693	0.000
	γ_1	-0.011	0.020	0.584
RC	α_0	2.789	0.217	0.000
	α_1	-0.019	0.004	0.000
	α_2	-0.361	0.113	0.001
	β_1	-0.693	0.126	0.000
	γ_0	5.294	1.006	0.000
	γ_1	-0.011	0.033	0.743
MVa	α_0	2.771	0.136	0.000
	α_1	-0.019	0.003	0.000
	α_2	-0.368	0.085	0.000
	β_1	-0.652	0.084	0.000
	γ_0	5.352	0.788	0.000
	γ_1	-0.007	0.022	0.747
MPVa	α_0	2.771	0.135	0.000
	α_1	-0.019	0.002	0.000
	α_2	-0.368	0.082	0.000
	β_1	-0.652	0.082	0.000
	γ_0	5.352	0.568	0.000
	γ_1	-0.007	0.017	0.675

Como a variável vapor d'água, associada ao modelo de dispersão, não foi significativa ao nível de 5% de significância, ajustou-se um novo modelo de regressão beta linear com erros de medida com dispersão constante dado por

$$\begin{aligned}
 y_t &\sim \text{Beta}(\mu_t, \phi) \\
 \log\left(\frac{\mu_t}{1-\mu_t}\right) &= \alpha_0 + \alpha_1 z_{1t} + \alpha_2 z_{2t} + \beta_1 x_{1t} \\
 \log(\phi) &= \gamma_0, \quad t = 1, \dots, 28,
 \end{aligned} \tag{6.4}$$

em que x_{1t} representa a concentração de vanádio, z_{1t} denota vapor d'água e z_{2t} é a variável indicadora de temperatura ($0 = 700$ °C e $1 = 760$ °C). Aqui, as funções de ligação associadas à média e à dispersão são logito e logaritmo, respectivamente, e os pontos de quadratura foram $Q = 50$.

A Tabela 6.8 apresenta as estimativas, erros-padrão e p -valores obtidos pelos métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa) considerando o Modelo 1, Equação (6.4).

Tabela 6.8: Estimativas, erros-padrão e p -valores obtidos pelos métodos *Naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa). Modelo de regressão beta linear com erros de medidas com dispersão constante (Modelo 1).

Método	Parâmetro	Estimativa	Erro-padrão	p -valor
Naive	α_0	2.705	0.124	0.000
	α_1	-0.018	0.003	0.000
	α_2	-0.358	0.083	0.000
	β_1	-0.599	0.071	0.000
	γ_0	4.934	0.267	0.000
RC	α_0	2.752	0.160	0.000
	α_1	-0.018	0.003	0.000
	α_2	-0.358	0.107	0.001
	β_1	-0.671	0.086	0.000
	γ_0	4.934	0.284	0.000
MVa	α_0	2.751	0.128	0.000
	α_1	-0.019	0.003	0.000
	α_2	-0.367	0.084	0.000
	β_1	-0.641	0.078	0.000
	γ_0	5.118	0.325	0.000
MPVa	α_0	2.751	0.130	0.000
	α_1	-0.019	0.002	0.000
	α_2	-0.367	0.084	0.000
	β_1	-0.641	0.077	0.000
	γ_0	5.118	0.253	0.000

Note que todas as variáveis são significativas, ao nível de 5% de significância, em todos os métodos avaliados.

As Figuras 6.8, 6.9, 6.10 e 6.11 apresentam os resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados para os métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa), respectivamente. Note que, os resíduos não se comportam de forma aleatória, apresentando um comportamento sinuoso, como uma função seno (cosseno). Este resultado é verificado em todos os métodos de estimação utilizados. Desta forma, é evidente a tendência não linear nos gráficos dos resíduos contra os índices das observações (Ver Figuras 6.8 (a)-(c), 6.9 (a)-(c), 6.10 (a)-(c), 6.11 (a)-(c)).

Em relação aos envelopes simulados, há evidências de que o modelo está sendo bem ajustado, mesmo apresentando quebra do pressuposto da linearidade. Isto é verificado em todos os métodos de estimação avaliados (Ver Figuras 6.8 (d)-(f), 6.9 (d)-(f), 6.10 (d)-(f), 6.11 (d)-(f)).

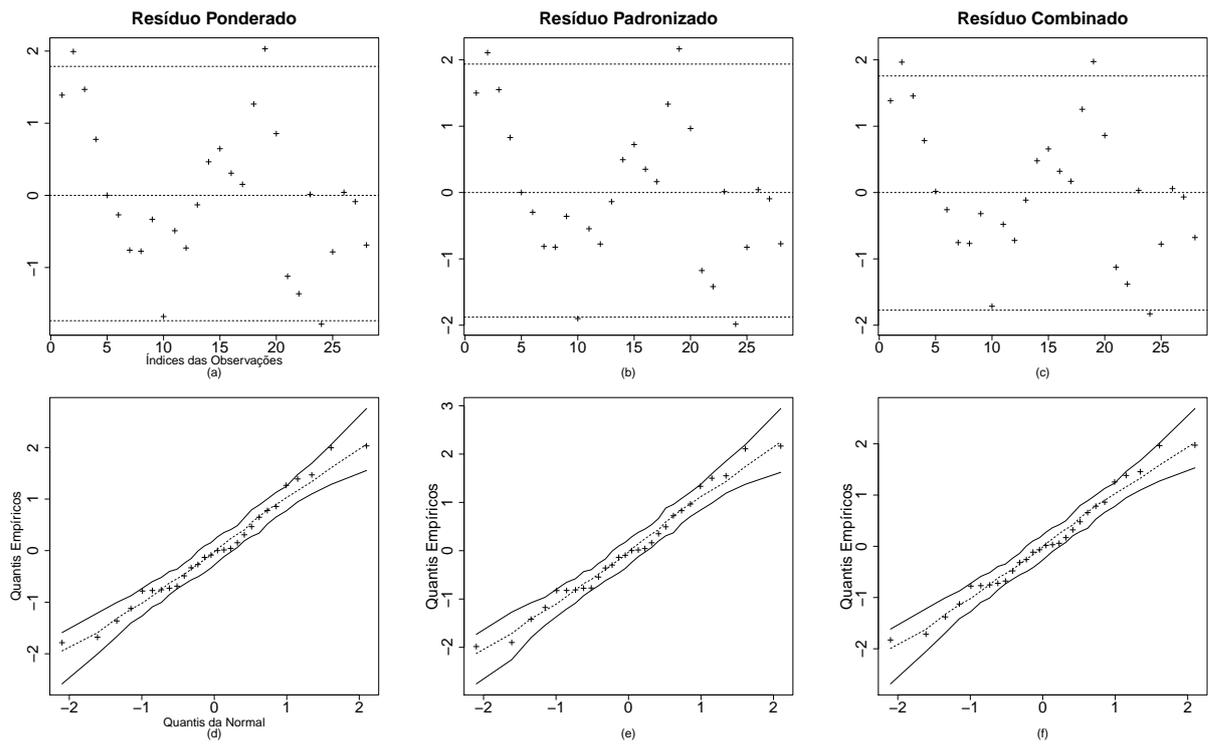


Figura 6.8: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o **método naive** para o ajuste do modelo de regressão beta linear com erros de medidas e dispersão constante, Equação (6.4).

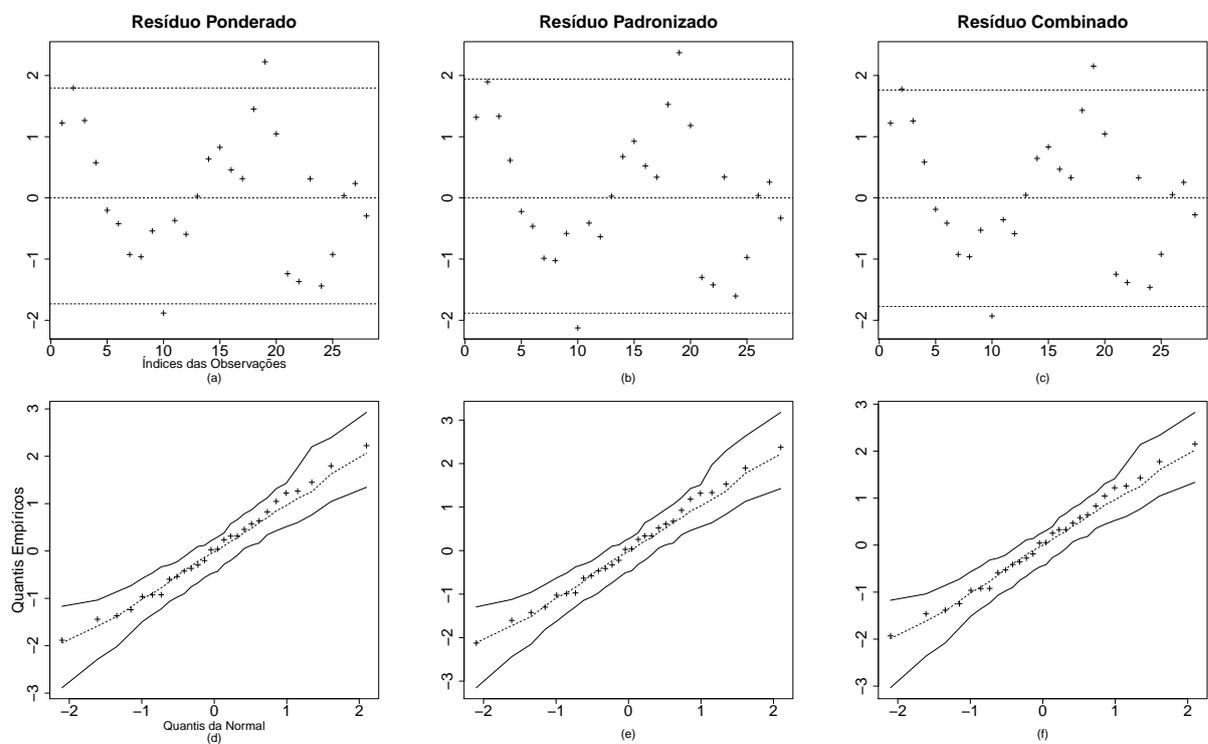


Figura 6.9: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o **método da calibração da regressão (RC)** para o ajuste do modelo de regressão beta linear com erros de medidas e dispersão constante, Equação (6.4).

Para tentar melhorar os resultados anteriores, outros modelos foram testados e analisados considerando todos os métodos de estimação apresentados. Um dos modelos considerou, por exemplo $\sqrt{V_{anadio}}$. No entanto, a análise de resíduos não foi tão satisfatória, especialmente em relação aos envelopes simulados.

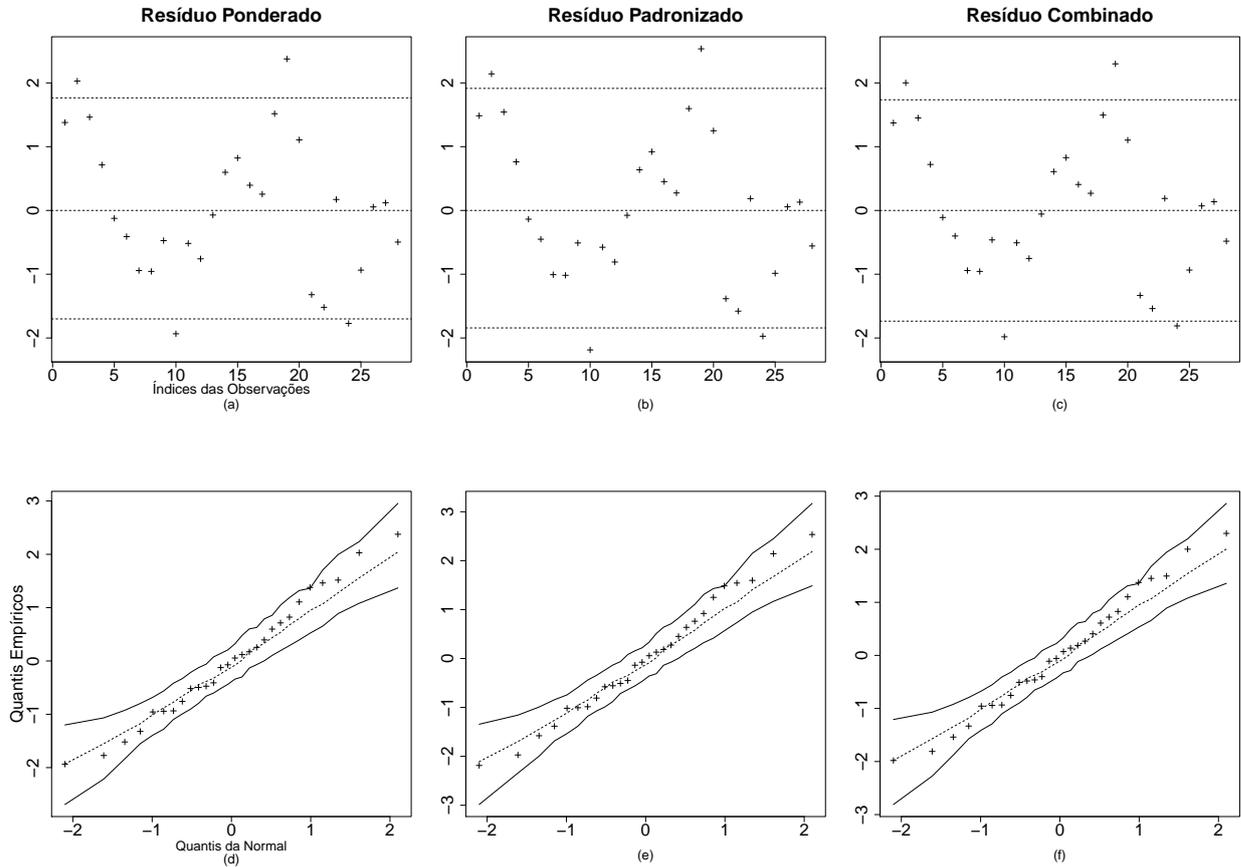


Figura 6.10: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o **método da máxima verossimilhança aproximada (MVA)** para o ajuste do modelo de regressão beta com erros de medidas e dispersão constante, Equação (6.4).

Com base nos gráficos de resíduos contra os índices das observações, os indícios da necessidade de se considerar preditores não lineares são grandes. A questão é como definir o preditor não linear. Com base no *boxplot* da variável vapor d'água nota-se uma grande dispersão e ocorrência de valores extremos como 0 e ao mesmo tempo 55.8. Assim, o passo inicial é pensar em uma função que estabilize essa dispersão.

Uma função plausível é o logaritmo. No entanto, a presença de zeros implica que teríamos que considerar, por exemplo, $\log(\text{vapor} + \alpha)$, ou seja uma função não linear no preditor. Outra possibilidade é considerar funções do tipo $1/(\text{vapor} + \alpha)$ ou $\text{vapor}/(\text{vapor} + \alpha)$, também não lineares.

Avaliamos o ajuste considerando as três funções não lineares citadas e o ajuste que apresentou os melhores gráficos de resíduos foi o que considera a função $\text{vapor}/(\text{vapor} + \alpha)$

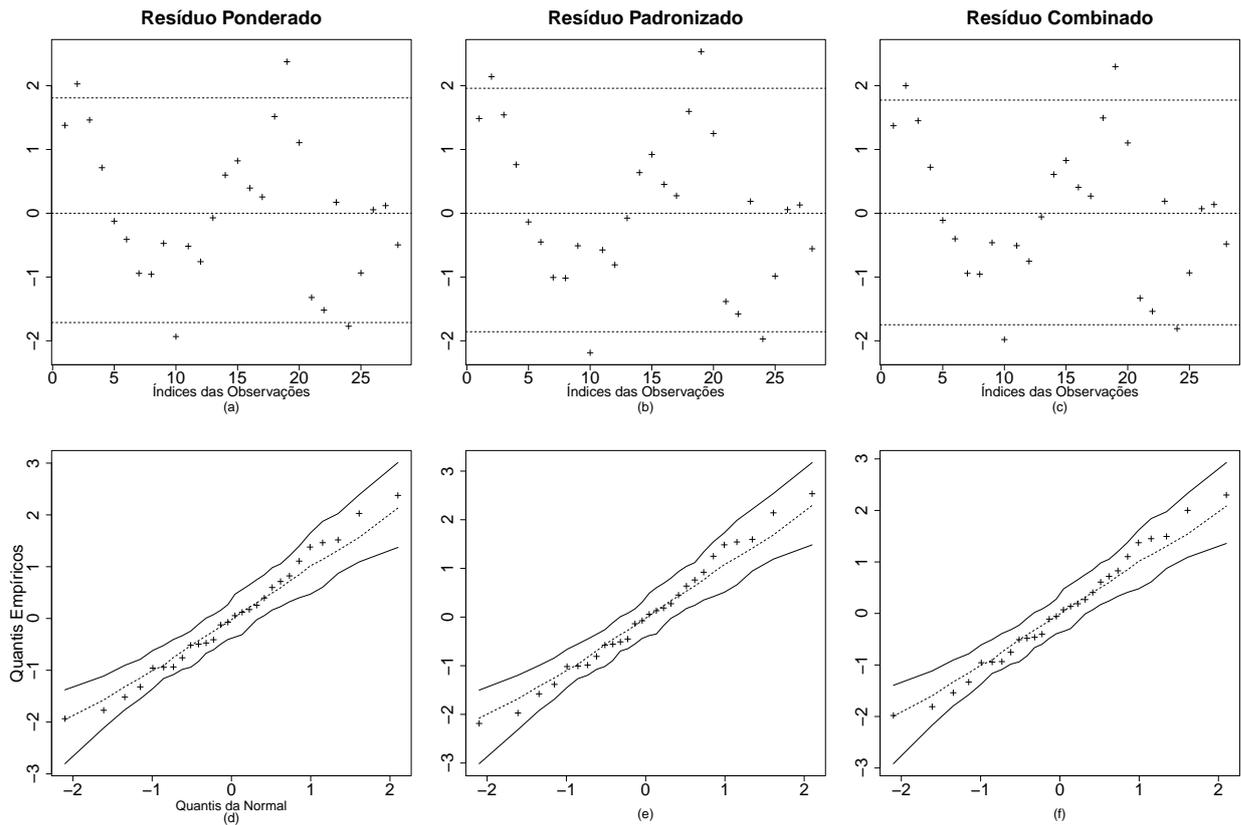


Figura 6.11: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o **método da máxima pseudo-verossimilhança aproximada (MPVa)** para o ajuste do modelo de regressão beta linear com erros de medidas e dispersão constante, Equação (6.4).

no predictor. Desta forma, o modelo de regressão beta não linear com erros de medida (Modelo 2) considerado é dado por

$$\begin{aligned}
 y_t &\sim \text{Beta}(\mu_t, \phi) \\
 \log\left(\frac{\mu_t}{1 - \mu_t}\right) &= \alpha_0 + \alpha_1 \frac{z_{1t}}{z_{1t} + \alpha_2} + \alpha_3 z_{2t} + \beta_1 x_{1t} \\
 \log(\phi) &= \gamma_0, \quad t = 1, \dots, 28,
 \end{aligned} \tag{6.5}$$

em que x_{1t} representa a concentração de vanádio, z_{1t} denota vapor d'água e z_{2t} é a variável indicadora de temperatura ($0 = 700 \text{ } ^\circ\text{C}$ e $1 = 760 \text{ } ^\circ\text{C}$). Aqui, as funções de ligação associadas à média e à dispersão são logito e logaritmo, respectivamente. Utilizamos $Q = 50$ pontos de quadratura.

A Tabela 6.9 apresenta as estimativas, erros-padrão e p -valores obtidos pelos métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa) considerando o Modelo 2, Equação (6.5). Note que todas as variáveis foram significativas, ao nível de 5% de significância, em todos os métodos avaliados. Também é possível que as estimativas estão bem consistentes nos quatro métodos de estimação avaliados. Além disso, observamos que os métodos da

máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa) apresentam todas as estimativas dos parâmetros idênticas, diferindo apenas nos erros-padrão. Já os métodos *naive* e calibração da regressão (RC) apresentam estimativas dos parâmetros semelhantes e erros-padrão similares.

Tabela 6.9: Estimativas, erros-padrão e p -valores obtidos pelos métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa). Modelo de regressão beta não linear com erros de medidas com dispersão constante, considerando a função de ligação logito associado à media (Modelo 2).

Método	Parâmetro	Estimativa	Erro-padrão	p -valor
<i>Naive</i>	α_0	2.265	0.129	0.000
	α_1	-0.103	0.033	0.002
	α_2	-26.136	2.867	0.000
	α_3	-0.350	0.115	0.002
	β_1	-0.558	0.100	0.000
	γ_0	4.261	0.266	0.000
RC	α_0	2.310	0.132	0.000
	α_1	-0.103	0.048	0.030
	α_2	-26.136	2.559	0.000
	α_3	-0.350	0.163	0.032
	β_1	-0.626	0.094	0.000
	γ_0	4.261	0.342	0.000
MVa	α_0	2.291	0.134	0.000
	α_1	-0.104	0.033	0.002
	α_2	-26.101	2.875	0.000
	α_3	-0.355	0.116	0.002
	β_1	-0.590	0.108	0.000
	γ_0	4.324	0.287	0.000
MPVa	α_0	2.291	0.147	0.000
	α_1	-0.104	0.028	0.000
	α_2	-26.101	1.950	0.000
	α_3	-0.355	0.108	0.001
	β_1	-0.590	0.105	0.000
	γ_0	4.324	0.278	0.000

As Figuras 6.12, 6.13, 6.14 e 6.15 apresentam os resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados para os métodos *naive*, calibração da regressão, máxima verossimilhança aproximada e máxima pseudo-verossimilhança aproximada, respectivamente, considerando o Modelo 2 dado na Equação 6.5.

É possível notar que os resíduos se comportam de forma aleatória dentro do intervalo considerado, sendo este resultado verificado em todos os métodos de estimação utilizados. Portanto, o modelo de regressão beta não linear com erros de medida considerado está ajustando bem os dados (Ver Figuras 6.12 (a)-(c), 6.13 (a)-(c), 6.14 (a)-(c), 6.15 (a)-(c)).

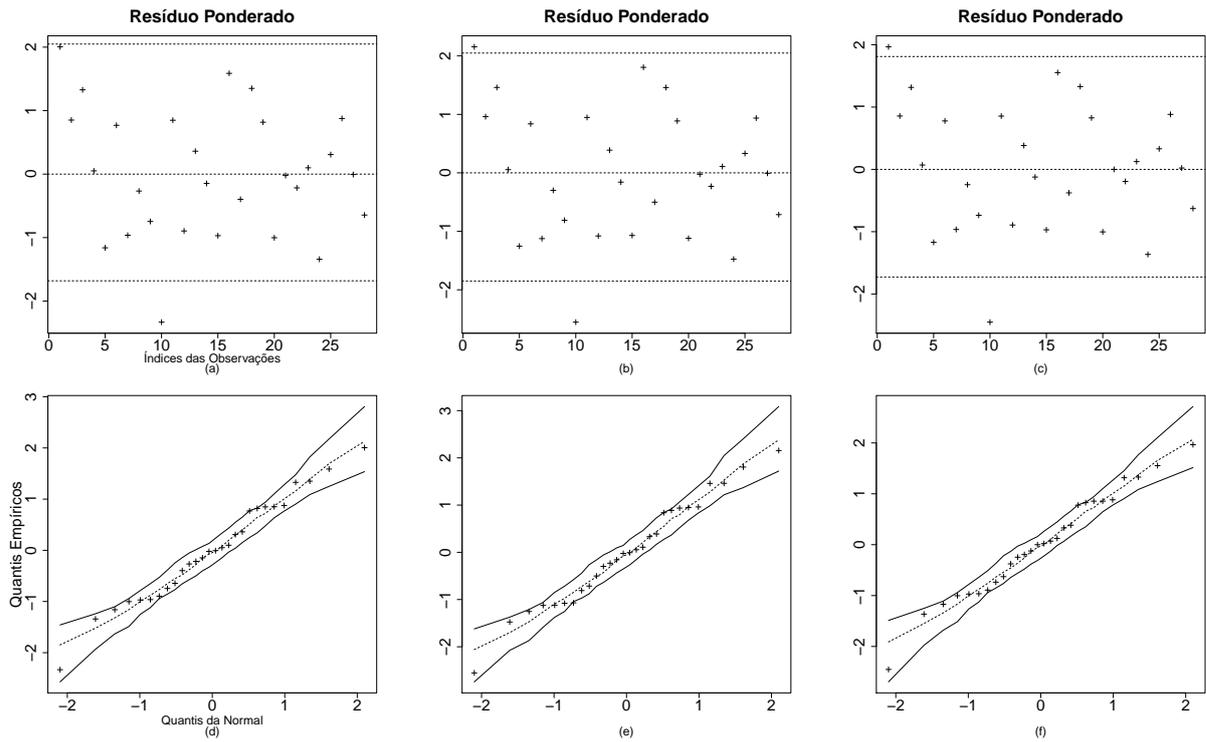


Figura 6.12: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o **método naive** para o ajuste do modelo de regressão beta não linear com erros de medidas. Modelo 2.

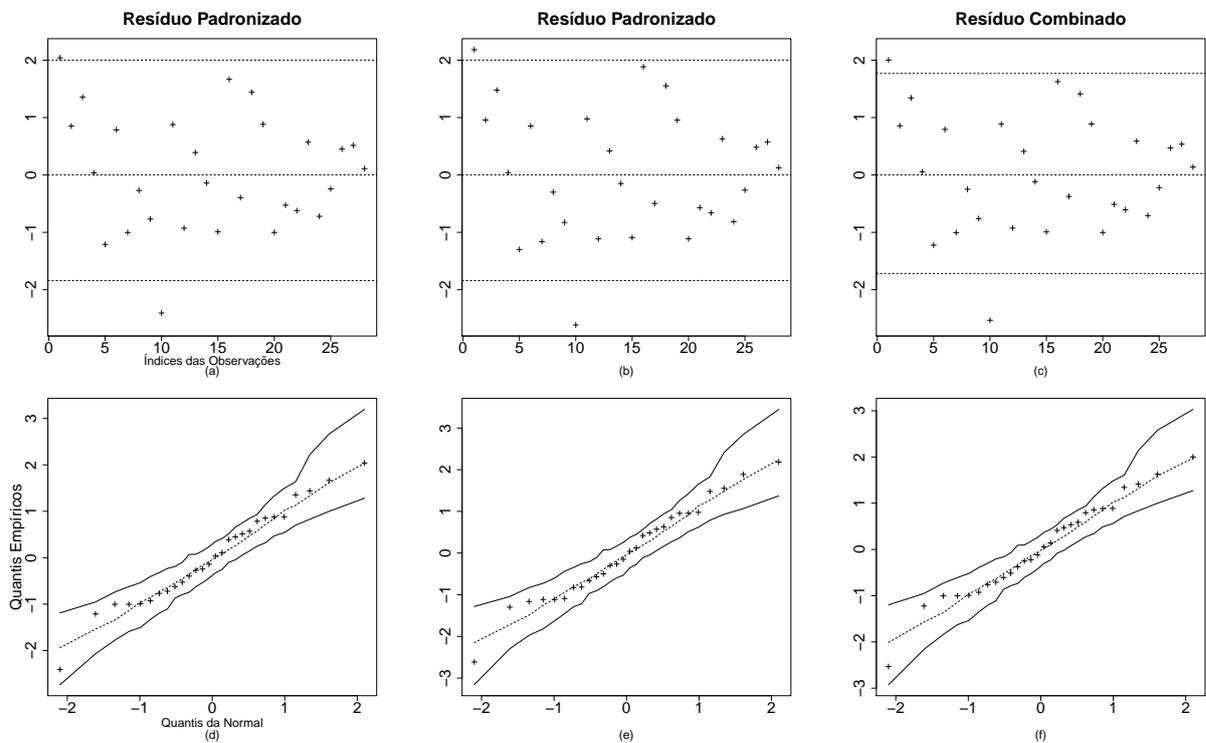


Figura 6.13: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o **método da calibração da regressão** para o ajuste do modelo de regressão beta não linear com erros de medidas. Modelo 2.

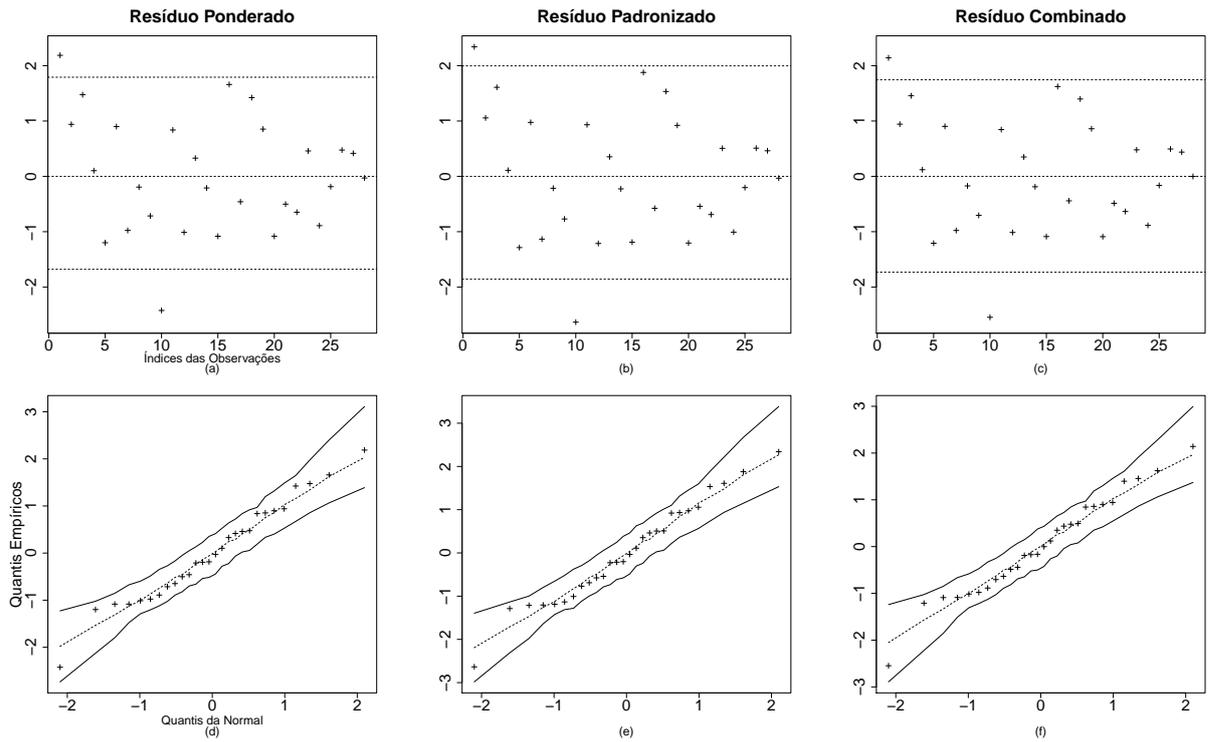


Figura 6.14: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o **método da máxima verossimilhança aproximada** para o ajuste do modelo de regressão beta não linear com erros de medidas. Modelo 2.

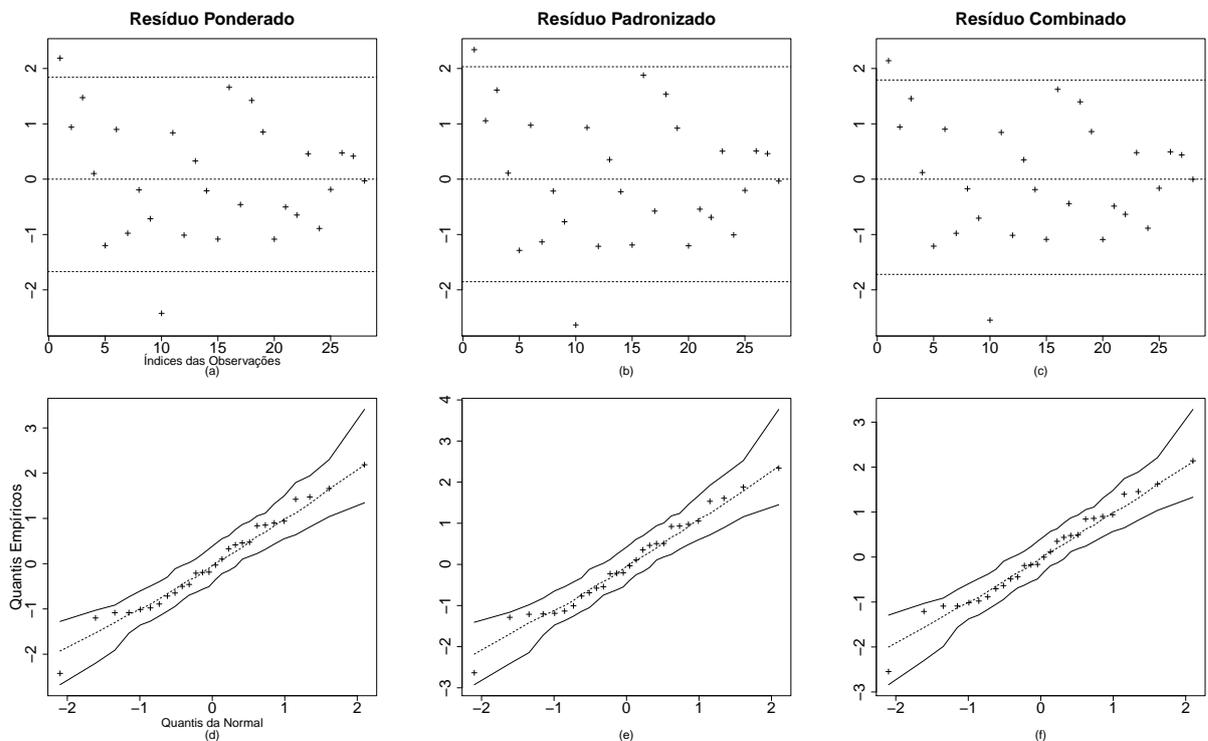


Figura 6.15: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o **método da máxima pseudo-verossimilhança aproximada** para o ajuste do modelo de regressão beta não linear com erros de medidas. Modelo 2.

Em relação aos envelopes simulados, é possível notar que os métodos de máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa) apresentam melhores resultados (Ver Figuras 6.14 (d)-(f), 6.15 (d)-(f)). Portanto, há evidências de que o modelo de regressão beta não linear dado na Equação (6.5) está sendo bem ajustado.

Tabela 6.10: Estimativas, erros-padrão e p -valores obtidos pelos métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa). Modelos de regressão beta não linear com erros de medida considerando a função de ligação complemento log-log associada à média. Modelo 3.

Método	Parâmetro	Estimativa	Erro-padrão	p -valor
<i>Naive</i>	α_0	0.898	0.056	0.000
	α_1	-0.049	0.015	0.001
	α_2	-26.497	2.717	0.000
	α_3	-0.181	0.054	0.001
	β_1	-0.274	0.046	0.000
	γ_0	4.319	0.267	0.000
RC	α_0	0.919	0.060	0.000
	α_1	-0.049	0.018	0.008
	α_2	-26.497	2.537	0.000
	α_3	-0.181	0.084	0.031
	β_1	-0.307	0.047	0.000
	γ_0	4.319	0.322	0.000
MVa	α_0	0.910	0.057	0.000
	α_1	-0.049	0.015	0.001
	α_2	-26.451	2.724	0.000
	α_3	-0.182	0.054	0.001
	β_1	-0.292	0.050	0.000
	γ_0	4.405	0.293	0.000
MPVa	α_0	0.910	0.063	0.000
	α_1	-0.049	0.011	0.000
	α_2	-26.451	1.891	0.000
	α_3	-0.182	0.054	0.001
	β_1	-0.292	0.050	0.000
	γ_0	4.405	0.286	0.000

Com o objetivo de melhorar o modelo consideramos outras funções de ligação para o modelo da média e a que apresentou resultados mais interessantes foi a complemento log-log. Assim, o modelo não linear com erros de medidas, denotado por Modelo 3, é dado por

$$\begin{aligned}
 y_t &\sim \text{Beta}(\mu_t, \phi), \\
 \log[-\log(1 - \mu_t)] &= \alpha_0 + \alpha_1 \frac{z_{1t}}{z_{1t} + \alpha_2} + \alpha_3 z_{2t} + \beta_1 x_{1t}, \\
 \log(\phi) &= \gamma_0, \quad t = 1, \dots, 28,
 \end{aligned} \tag{6.6}$$

em que x_{1t} representa a concentração de vanádio, z_{1t} denota vapor d'água e z_{2t} é a variável indicadora de temperatura ($0 = 700 \text{ }^{\circ}\text{C}$ e $1 = 760 \text{ }^{\circ}\text{C}$). Aqui as funções de ligação associadas à média e à dispersão são complemento log-log e logaritmo, respectivamente, e os pontos de quadratura considerados é $Q = 50$.

A Tabela 6.10 apresenta as estimativas, erros-padrão e p -valores obtidos pelos métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa) do modelo apresentado na Equação (6.6). Note que todas as variáveis foram significativas, ao nível de 5% de significância, em todos os métodos avaliados. É possível verificar que estimativas estão bem consistentes nos quatro métodos de estimação.

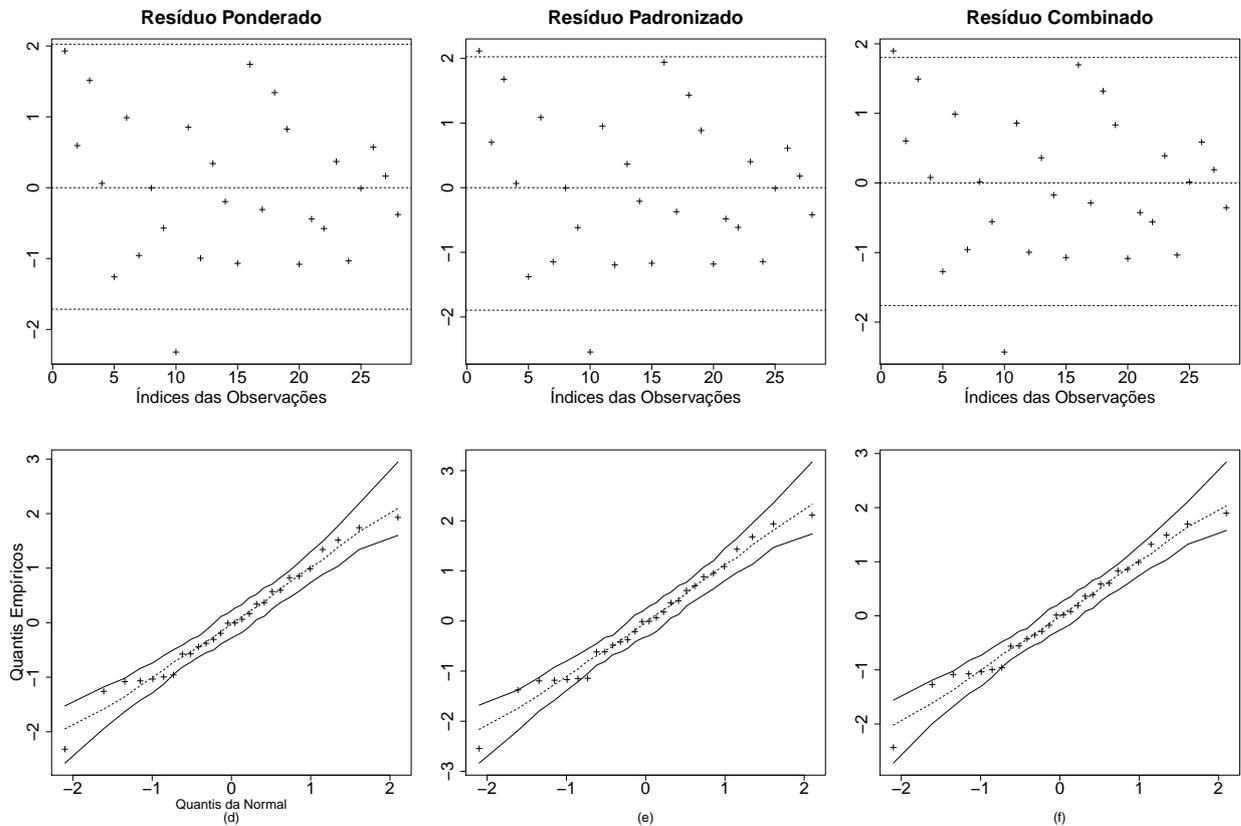


Figura 6.16: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método *naive* para o ajuste do modelo de regressão beta não linear com erros de medidas. Modelo 3.

As Figuras 6.16, 6.17, 6.18 e 6.19 apresentam os resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados para os métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa), respectivamente, considerando o Modelo 3. É possível notar que os resíduos se comportam de forma aleatória dentro do intervalo considerado.

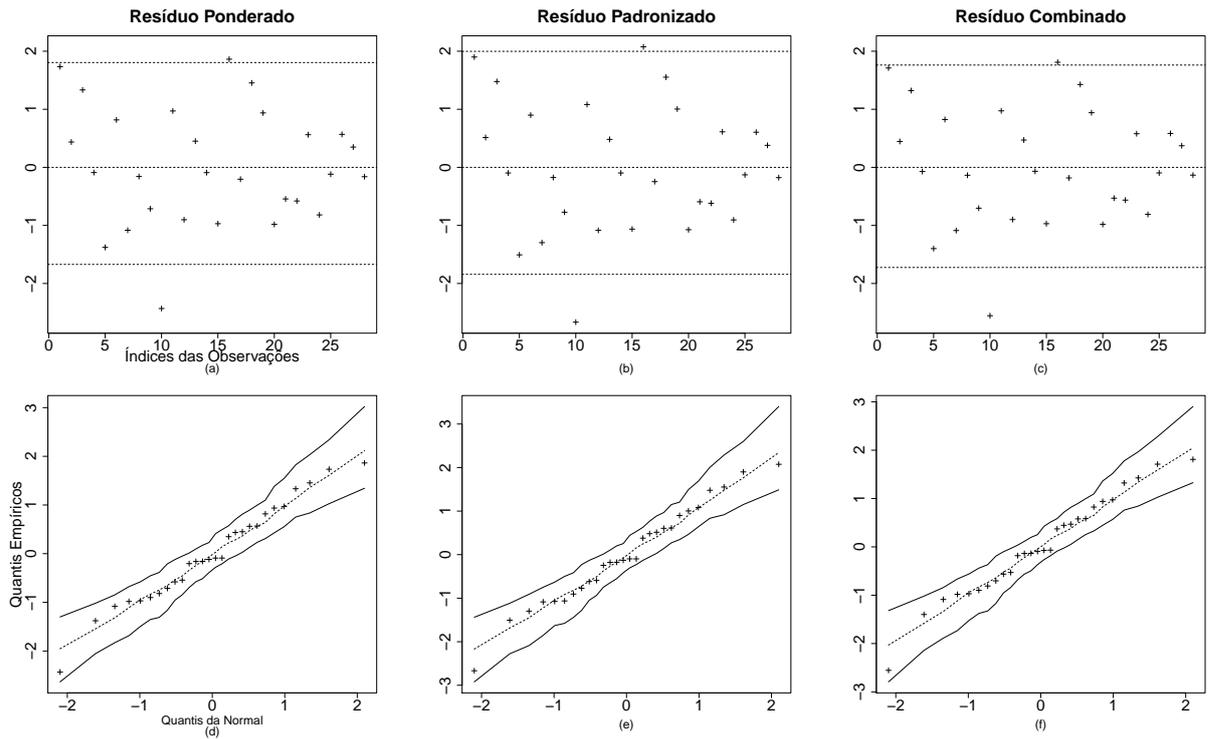


Figura 6.17: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da calibração da regressão para o ajuste do modelo de regressão beta não linear com erros de medidas. Modelo 3.

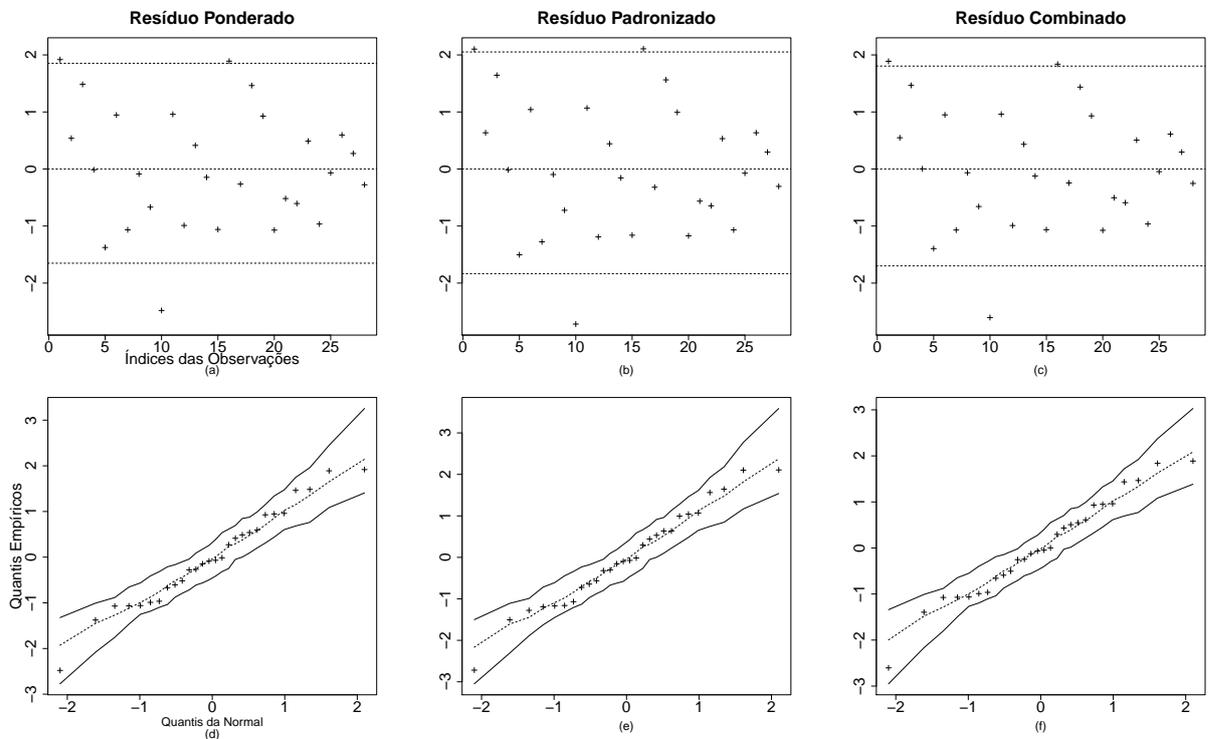


Figura 6.18: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da máxima verossimilhança aproximada para o ajuste do modelo de regressão beta não linear com erros de medidas. Modelo 3.

Portanto, o modelo não linear considerado está ajustando bem os dados. Em relação aos envelopes, é possível notar que os métodos de máxima verossimilhança aproximada e máxima pseudo-verossimilhança aproximada apresentam os melhores resultados. Vale salientar também que o gráfico de envelopes adquirido pelo método *naive* não está bom, especialmente quando é considerado o resíduo padronizado 6.16 (d)-(f).

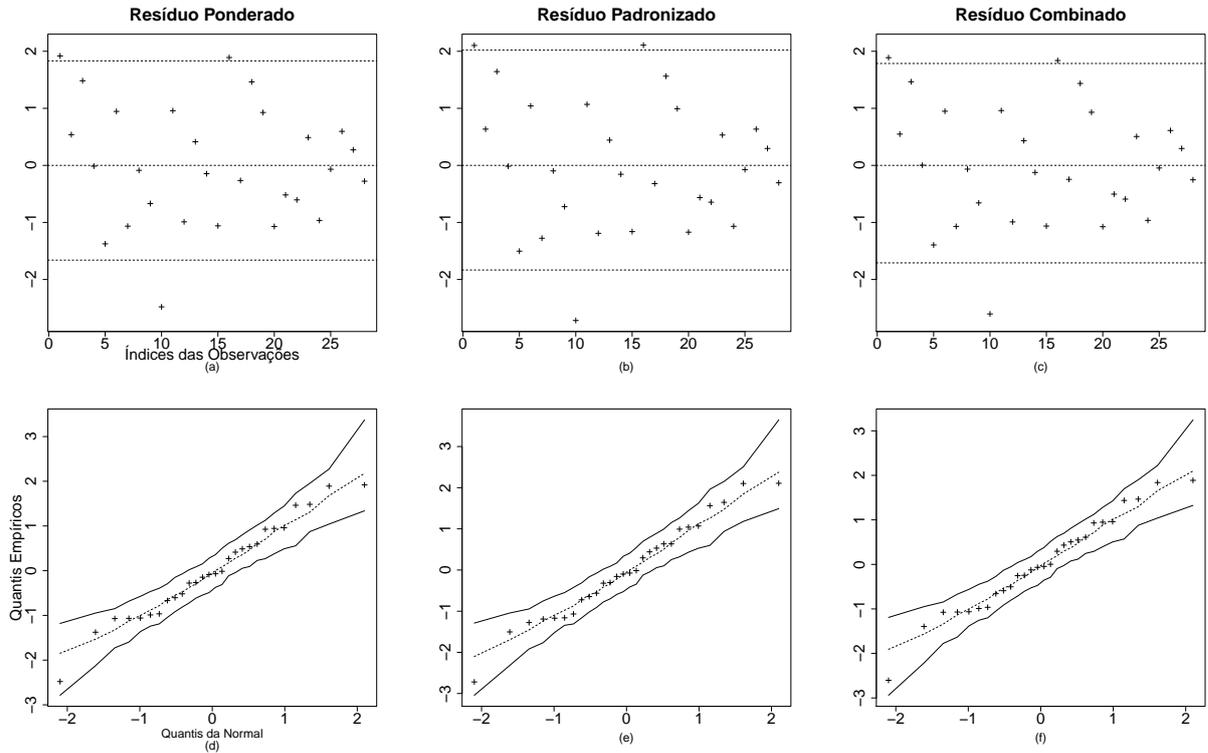


Figura 6.19: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da máxima pseudo-verossimilhança aproximada para o ajuste do modelo de regressão beta não linear com erros de medidas. Modelo 3.

Tabela 6.11: Estatísticas de predição e medidas para avaliação do ajuste do modelo considerando o modelo de regressão linear beta com erros de medida (Modelo 1).

Medidas	Modelo 1 - Modelo linear			
	<i>Naive</i>	RC	MVa	MPVa
P^2	0.61058	0.56961	0.58242	0.58241
P_c^2	0.54286	0.49476	0.50980	0.50978
$P_{\alpha\gamma}^2$	0.61118	0.57098	0.58395	0.58394
$P_{\alpha\gamma c}^2$	0.54356	0.49637	0.51159	0.51158
ps- R^2	0.84724	0.84283	0.84599	0.84599
ps- R_c^2	0.82067	0.81550	0.81921	0.81920
R_{LR}^2	0.85132	0.97556	0.85074	0.85074
R_{LRc}^2	0.82546	0.97131	0.82478	0.82478

As Tabelas 6.11 e 6.12 apresentam as estatísticas de predição e as medidas para avaliação do ajuste dos modelos avaliados considerando os métodos *naive*, calibração da regressão

(RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa).

Tabela 6.12: Estatísticas de predição e medidas para avaliação do ajuste do modelo considerando os modelos de regressão linear beta não linear com erros de medida com função de ligação logito (Modelo 2) e função de ligação complemento log-log (Modelo 3).

Medidas	Modelo 2 - Logito				Modelo 3 - complemento log-log			
	<i>Naive</i>	RC	MVa	MPVa	<i>Naive</i>	RC	MVa	MPVa
P^2	0.46553	0.38485	0.40053	0.40052	0.72661	0.69720	0.69915	0.69915
P_c^2	0.34405	0.27786	0.29628	0.29627	0.66448	0.64454	0.64683	0.64683
$P_{\alpha\gamma}^2$	0.45990	0.37784	0.39410	0.39409	0.72359	0.69312	0.69557	0.69556
$P_{\alpha\gamma c}^2$	0.33715	0.26964	0.28873	0.28872	0.66077	0.63975	0.64262	0.64262
ps- R^2	0.68613	0.68249	0.68516	0.68515	0.71186	0.70969	0.71117	0.71117
ps- R_c^2	0.61480	0.62727	0.63040	0.63040	0.64637	0.65920	0.66094	0.66094
R_{LR}^2	0.70882	0.95215	0.70623	0.70623	0.72422	0.95468	0.72379	0.72379
R_{LRc}^2	0.64264	0.94382	0.65514	0.65514	0.67626	0.94680	0.67575	0.67575

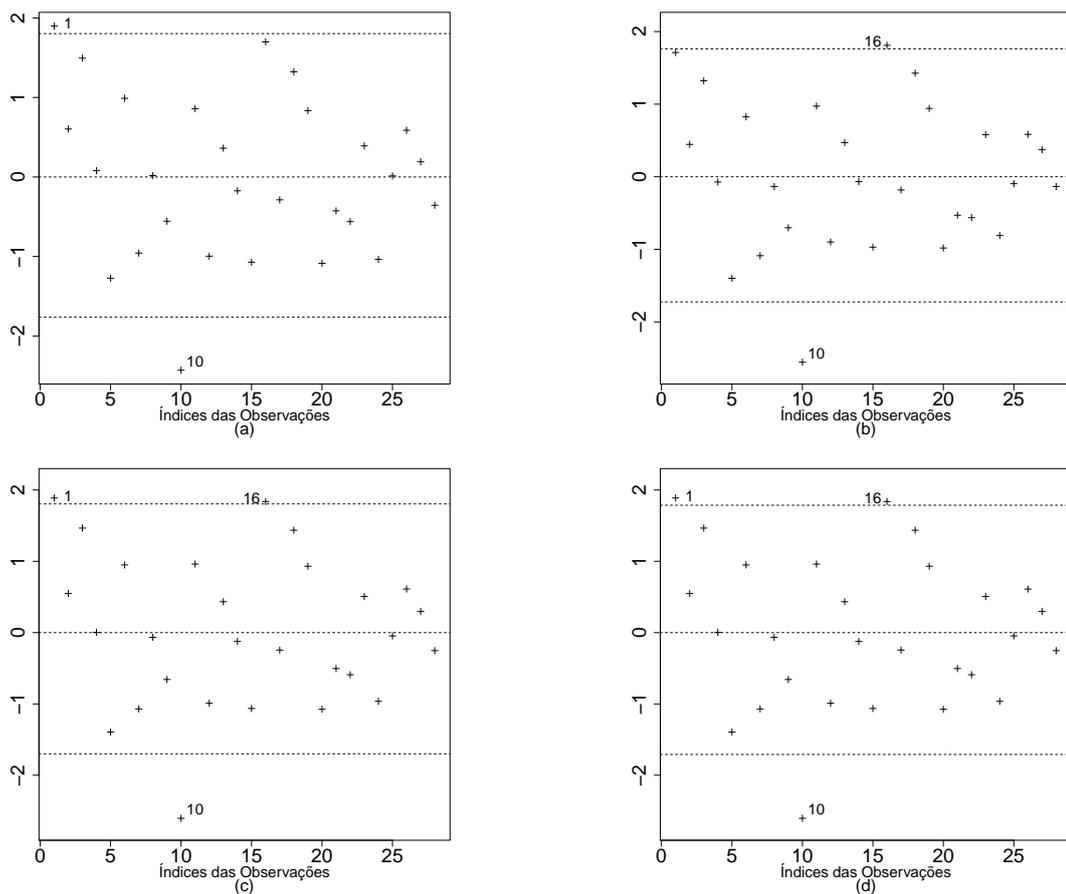


Figura 6.20: Identificação dos pontos aberrantes considerando os resíduos combinados utilizando os métodos naive (a), calibração da regressão (b), máxima verossimilhança aproximada (c) e máxima pseudo-verossimilhança aproximada (d) para o ajuste do modelo de regressão beta não linear com erros de medidas. Modelo 3.

Considerando o Modelo 1 verificamos que o método *naive* apresenta os maiores valores das estatísticas de predição. No entanto, o modelo 1 não é adequado para ajustar os dados analisados, uma vez que a análise de resíduos indicou a necessidade de um modelo não linear (Tabela 6.11). Em relação ao ajuste do modelo, é verificado que todos os resultados estão próximos e acima de 0.80. Vale salientar que, considerando o R_{LR}^2 e R_{LRc}^2 , o método da calibração da regressão apresentou os melhores resultados.

Tabela 6.13: Estimativas, erros-padrão e p -valores obtidos pelo método da calibração da regressão (RC).

Exclusão	Parâmetro	Estimativa	Erro-padrão	p -valor
Nenhuma	α_0	0.919	0.060	0.000
	α_1	-0.049	0.018	0.008
	α_2	-26.497	2.537	0.000
	α_3	-0.181	0.084	0.031
	β_1	-0.307	0.047	0.000
	γ_0	4.319	0.322	0.000
#10	α_0	0.958	0.074	0.000
	α_1	-0.045	0.100	0.654
	α_2	-27.628	10.582	0.009
	α_3	-0.154	0.052	0.003
	β_1	-0.339	0.047	0.000
	γ_0	4.587	0.245	0.000
#16	α_0	0.919	0.071	0.000
	α_1	-0.045	0.013	0.001
	α_2	-25.497	2.607	0.000
	α_3	-0.203	0.064	0.001
	β_1	-0.327	0.050	0.000
	γ_0	4.410	0.378	0.000

A Tabela 6.12 apresentam as estatísticas de predição e medidas de avaliação do ajuste do modelo considerando os modelos não lineares apresentados. Note que o modelo que utiliza a função de ligação complemento log-log apresenta melhores resultados tanto quanto ao poder preditivo quanto a qualidade de ajuste relativamente ao modelo que se baseia no uso da função de ligação logito. Por exemplo, considerando a estatística de predição P^2 baseada no resíduo ponderado ($P_{\alpha\gamma c}^2$) e o método da máxima verossimilhança aproximada (MVA), os valores são $P_{\alpha\gamma c}^2 = 0.28873$ para o Modelo 2 e $P_{\alpha\gamma c}^2 = 0.64262$ para o Modelo 3. O mesmo ocorre para as medidas de qualidade de ajuste, por exemplo, considerando a pseudo- R^2 corrigida ($ps-R_c^2$) e o método da máxima pseudo-verossimilhança aproximada (MPVa), os valores são $ps-R_c^2 = 0.63040$ para o Modelo 2 e $ps-R_c^2 = 0.66094$ para o Modelo 3.

Vale ressaltar que o método da calibração da regressão tem mais facilidade no processo de estimação. Já os outros métodos carregam a dificuldade da não linearidade e dos

métodos. Assim, não é possível comparar o modelo linear com os não lineares.

Para verificar se os pontos aberrantes estão influenciando nos resultados, os mesmos foram identificados, retirados e novas análises foram realizadas. Considerando o modelo de regressão beta não linear com erros de medidas com função de ligação complemento log-log, identificamos os pontos aberrantes considerando os resíduos combinados utilizando os métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa) e os resultados encontram-se na Figura 6.20 (a)-(d), respectivamente.

As Tabelas 6.13, 6.14 e 6.15 apresentam as estimativas, erros-padrão e p -valores obtidos pelos métodos da calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa), respectivamente, após a retirada dos pontos identificados. Ao retirar as observações aberrantes não houve mudanças inferenciais. Além disso, observamos que estimativas são bem consistentes quando retiramos os pontos.

Tabela 6.14: Estimativas, erros-padrão e p -valores obtidos pelo método da máxima verossimilhança aproximada (MVa).

Exclusão	Parâmetro	Estimativa	Erro-padrão	p -valor
Nenhuma	α_0	0.910	0.057	0.000
	α_1	-0.049	0.015	0.001
	α_2	-26.451	2.724	0.000
	α_3	-0.182	0.054	0.001
	β_1	-0.292	0.050	0.000
	γ_0	4.405	0.293	0.000
#1	α_0	0.871	0.063	0.000
	α_1	-0.046	0.013	0.000
	α_2	-25.517	2.931	0.000
	α_3	-0.160	0.053	0.002
	β_1	-0.273	0.048	0.000
	γ_0	4.601	0.331	0.000
#10	α_0	0.950	0.055	0.000
	α_1	-0.045	0.014	0.001
	α_2	-27.555	2.606	0.000
	α_3	-0.154	0.049	0.002
	β_1	-0.322	0.046	0.000
	γ_0	4.886	0.379	0.000
#16	α_0	0.909	0.059	0.000
	α_1	-0.045	0.013	0.001
	α_2	-25.370	3.022	0.000
	α_3	-0.204	0.053	0.000
	β_1	-0.310	0.049	0.000
	γ_0	4.632	0.349	0.000

Tabela 6.15: Estimativas, erros-padrão e p -valores obtidos pelo método da máxima pseudo-verossimilhança aproximada (MPVa).

Exclusão	Parâmetro	Estimativa	Erro-padrão	p -valor
Nenhuma	α_0	0.910	0.063	0.000
	α_1	-0.049	0.011	0.000
	α_2	-26.451	1.891	0.000
	α_3	-0.182	0.054	0.001
	β_1	-0.292	0.050	0.000
	γ_0	4.405	0.286	0.000
#1	α_0	0.871	0.063	0.000
	α_1	-0.046	0.011	0.000
	α_2	-25.517	2.200	0.000
	α_3	-0.160	0.051	0.002
	β_1	-0.273	0.050	0.000
	γ_0	4.601	0.342	0.000
#10	α_0	0.950	0.061	0.000
	α_1	-0.045	0.011	0.000
	α_2	-27.555	1.885	0.000
	α_3	-0.154	0.048	0.001
	β_1	-0.322	0.044	0.000
	γ_0	4.886	0.274	0.000
#16	α_0	0.909	0.066	0.000
	α_1	-0.045	0.010	0.000
	α_2	-25.370	2.157	0.000
	α_3	-0.204	0.053	0.000
	β_1	-0.310	0.046	0.000
	γ_0	4.632	0.350	0.000

6.2.2 Intervalos de confiança

As Tabelas 6.16, 6.17 e 6.18 apresentam os limites inferiores (LI), limites superiores (LS) e comprimento (Comp) dos intervalos de confiança considerando os níveis 90%, 95% e 99% de confiança, para os métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa), considerando os Modelos 1, 2 e 3, respectivamente.

Considerando o modelo de regressão beta linear com erros de medida, Tabela 6.16, é possível verificar que o método da máxima pseudo-verossimilhança aproximada apresenta os menores comprimentos para os intervalos de confiança, ou seja, este método apresenta os intervalos mais precisos.

O mesmo pode ser observado nos modelos não lineares. Vale salientar que o modelo de regressão beta não linear com erro de medidas considerando a função de ligação complemento log-log apresenta os melhores resultados, ou seja, apresenta intervalos de menor comprimento, como visto na Tabela 6.18.

Tabela 6.16: Limites inferiores (LI), limites superiores (LS) e comprimento (Comp) dos intervalos de confiança considerando os níveis 90%, 95% e 99% de confiança, para os métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MPVa) e máxima pseudo-verossimilhança aproximada (MPVa). Modelo de regressão beta linear com erros de medida (Modelo Linear).

Métodos	Parâmetros	IC 90%			IC 95%			IC 99%		
		LI	LS	Comp	LI	LS	Comp	LI	LS	Comp
<i>Naive</i>	α_0	2.50129	2.90783	0.40654	2.46235	2.94677	0.48442	2.38624	3.02288	0.63664
	α_1	-0.02283	-0.01400	0.00883	-0.02368	-0.01316	0.01052	-0.02533	-0.01150	0.01383
	α_2	-0.49507	-0.22164	0.27343	-0.52126	-0.19545	0.32581	-0.57245	-0.14426	0.42819
	β_1	-0.71513	-0.48194	0.23320	-0.73747	-0.45960	0.27787	-0.78113	-0.41594	0.36519
	γ_0	4.49536	5.37311	0.87776	4.41128	5.45719	1.04591	4.24695	5.62151	1.37456
RC	α_0	2.46848	2.99351	0.52503	2.41818	3.04380	0.62562	2.31989	3.14209	0.82220
	α_1	-0.02349	-0.01334	0.01015	-0.02447	-0.01237	0.01210	-0.02637	-0.01047	0.01590
	α_2	-0.53428	-0.18243	0.35185	-0.56798	-0.14873	0.41925	-0.63385	-0.08286	0.55099
	β_1	-0.77768	-0.49759	0.28009	-0.80451	-0.47076	0.33374	-0.85694	-0.41833	0.43861
	γ_0	4.46689	5.40158	0.93468	4.37736	5.49111	1.11374	4.20238	5.66609	1.46371
MPVa	α_0	2.54072	2.96135	0.42063	2.50043	3.00164	0.50121	2.42169	3.08039	0.65870
	α_1	-0.02315	-0.01429	0.00887	-0.02400	-0.01344	0.01056	-0.02566	-0.01178	0.01388
	α_2	-0.50467	-0.22826	0.27641	-0.53114	-0.20178	0.32936	-0.58289	-0.15003	0.43286
	β_1	-0.76842	-0.51346	0.25496	-0.79284	-0.48904	0.30380	-0.84057	-0.44131	0.39927
	γ_0	4.58245	5.64989	1.06744	4.48020	5.75214	1.27194	4.28037	5.95197	1.67161
MPVa	α_0	2.53784	2.96426	0.42642	2.49699	3.00510	0.50811	2.41716	3.08493	0.66777
	α_1	-0.02247	-0.01497	0.00750	-0.02319	-0.01425	0.00894	-0.02459	-0.01285	0.01174
	α_2	-0.50402	-0.22891	0.27511	-0.53037	-0.20255	0.32782	-0.58187	-0.15105	0.43083
	β_1	-0.76735	-0.51457	0.25278	-0.79156	-0.49036	0.30120	-0.83888	-0.44303	0.39584
	γ_0	4.70025	5.53210	0.83184	4.62057	5.61178	0.99120	4.46484	5.76750	1.30266

Tabela 6.17: Limites inferiores (LI), limites superiores (LS) e comprimento (Comp) dos intervalos de confiança considerando os níveis 90%, 95% e 99% de confiança, para os métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MPVa) e máxima pseudo-verossimilhança aproximada (MPVa). Modelo de regressão beta não linear com erros de medida considerando a função de ligação logito (Modelo 2).

Métodos	Parâmetros	IC 90%			IC 95%			IC 99%		
		LI	LS	Comp	LI	LS	Comp	LI	LS	Comp
<i>Naive</i>	α_0	2.05259	2.47829	0.42570	2.01182	2.51907	0.50725	1.93212	2.59876	0.66664
	α_1	-0.15852	-0.04847	0.11005	-0.16906	-0.03793	0.13113	-0.18966	-0.01733	0.17234
	α_2	-30.85159	-21.42127	9.43032	-31.75489	-20.51797	11.23692	-33.52034	-18.75252	14.76782
	α_3	-0.53894	-0.16029	0.37865	-0.57521	-0.12402	0.45119	-0.64610	-0.05314	0.59296
	β_1	-0.72235	-0.39372	0.32863	-0.75383	-0.36225	0.39159	-0.81536	-0.30072	0.51463
RC	γ_0	3.82316	4.69967	0.87650	3.73921	4.78362	1.04442	3.57512	4.94771	1.37260
	α_0	2.07344	2.50673	0.43329	2.03194	2.54823	0.51630	1.95082	2.62935	0.67853
	α_1	-0.18173	-0.02527	0.15646	-0.19671	-0.01028	0.18643	-0.22600	0.01901	0.24502
	α_2	-30.34481	-21.92805	8.41676	-31.15103	-21.12183	10.02919	-32.72673	-19.54613	13.18059
	β_1	-0.61799	-0.08124	0.53675	-0.66940	-0.02983	0.63957	-0.76989	0.07065	0.84054
MPVa	γ_0	-0.74831	-0.44067	0.30764	-0.77778	-0.41120	0.36658	-0.83537	-0.35361	0.48176
	α_0	3.69926	4.82357	1.12431	3.59157	4.93126	1.33969	3.38109	5.14174	1.76066
	α_1	2.07119	2.51090	0.43971	2.02907	2.55302	0.52395	1.94675	2.63534	0.68859
	α_2	-0.15887	-0.04902	0.10985	-0.16939	-0.03850	0.13089	-0.18996	-0.01793	0.17202
	β_1	-30.83006	-21.37182	9.45824	-31.73603	-20.46584	11.27019	-33.50671	-18.69517	14.81154
MPVa	α_3	-0.54651	-0.16333	0.38319	-0.58322	-0.12662	0.45660	-0.65496	-0.05489	0.60007
	β_1	-0.76658	-0.41297	0.35361	-0.80045	-0.37910	0.42135	-0.86665	-0.31290	0.55375
	γ_0	3.85072	4.79551	0.94480	3.76022	4.88601	1.12579	3.58334	5.06288	1.47954
	α_0	2.04896	2.53316	0.48420	2.00258	2.57954	0.57696	1.91193	2.67018	0.75825
	α_1	-0.14984	-0.05805	0.09179	-0.15863	-0.04926	0.10938	-0.17582	-0.03207	0.14375
MPVa	α_2	-29.30830	-22.89357	6.41473	-29.92275	-22.27912	7.64362	-31.12365	-21.07822	10.04542
	α_3	-0.53200	-0.17784	0.35415	-0.56592	-0.14392	0.42200	-0.63222	-0.07762	0.55460
	β_1	-0.76217	-0.41743	0.34474	-0.79519	-0.38441	0.41078	-0.85973	-0.31987	0.53986
	γ_0	3.86546	4.78077	0.91531	3.77779	4.86844	1.09066	3.60643	5.03980	1.43337

Tabela 6.18: Limites inferiores (LI), limites superiores (LS) e comprimento (Comp) dos intervalos de confiança considerando os níveis 90%, 95% e 99% de confiança, para os métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MPVa) e máxima pseudo-verossimilhança aproximada (MPVa). Modelo de regressão não linear com erros de medida considerando a função de ligação complemento log-log (Modelo 3).

Métodos	Parâmetros	IC 90%			IC 95%			IC 99%		
		LI	LS	Comp	LI	LS	Comp	LI	LS	Comp
<i>Naive</i>	α_0	0.80588	0.98914	0.18326	0.78833	1.00670	0.21837	0.75402	1.04101	0.28698
	α_1	-0.07247	-0.02467	0.04781	-0.07705	-0.02009	0.05696	-0.08600	-0.01114	0.07486
	α_2	-30.96676	-22.02752	8.93923	-31.82302	-21.17126	10.65176	-33.49653	-19.49775	13.99878
	α_3	-0.27015	-0.09198	0.17818	-0.28722	-0.07491	0.21231	-0.32058	-0.04155	0.27902
	β_1	-0.35046	-0.19772	0.15274	-0.36509	-0.18309	0.18200	-0.39368	-0.15450	0.23919
	γ_0	3.88034	4.75715	0.87681	3.79636	4.84114	1.04478	3.63221	5.00528	1.37307
RC	α_0	0.81184	1.00739	0.19555	0.79311	1.02612	0.23301	0.75651	1.06273	0.30623
	α_1	-0.07846	-0.01868	0.05979	-0.08419	-0.01295	0.07124	-0.09539	-0.00176	0.09363
	α_2	-30.67078	-22.32350	8.34728	-31.47034	-21.52395	9.94640	-33.03303	-19.96125	13.07178
	α_3	-0.31937	-0.04276	0.27661	-0.34586	-0.01627	0.32960	-0.39765	0.03552	0.43316
	β_1	-0.36819	-0.21580	0.15240	-0.38279	-0.20120	0.18159	-0.41132	-0.17267	0.23865
	γ_0	3.78844	4.84905	1.06060	3.68685	4.95064	1.26379	3.48830	5.14919	1.66090
MPVa	α_0	0.81567	1.00411	0.18844	0.79762	1.02216	0.22454	0.76235	1.05744	0.29509
	α_1	-0.07258	-0.02485	0.04773	-0.07715	-0.02028	0.05687	-0.08609	-0.01135	0.07474
	α_2	-30.93200	-21.97158	8.96042	-31.79029	-21.11329	10.67700	-33.46777	-19.43581	14.03196
	α_3	-0.27103	-0.09222	0.17881	-0.28815	-0.07509	0.21306	-0.32163	-0.04162	0.28001
	β_1	-0.37361	-0.20958	0.16403	-0.38932	-0.19387	0.19546	-0.42003	-0.16316	0.25687
	γ_0	3.92316	4.88551	0.96235	3.83098	4.97769	1.14671	3.65082	5.15786	1.50703
MPVa	α_0	0.80695	1.01283	0.20588	0.78723	1.03256	0.24533	0.74869	1.07110	0.32241
	α_1	-0.06756	-0.02987	0.03769	-0.07117	-0.02626	0.04492	-0.07823	-0.01920	0.05903
	α_2	-29.56210	-23.34146	6.22064	-30.15796	-22.74561	7.41235	-31.32252	-21.58104	9.74148
	α_3	-0.26970	-0.09354	0.17616	-0.28658	-0.07667	0.20991	-0.31956	-0.04369	0.27587
	β_1	-0.37307	-0.21012	0.16296	-0.38868	-0.19451	0.19417	-0.41919	-0.16400	0.25519
	γ_0	3.93373	4.87495	0.94122	3.84357	4.96511	1.12153	3.66737	5.14131	1.47395

6.3 Aplicação 3 - Modelo Não Linear: sobrepeso em adolescentes

Essa aplicação é referente a uma pesquisa da Faculdade de Medicina e do Departamento de Nutrição da Universidade Federal da Bahia cujo objetivo era identificar a prevalência do fenótipo cintura hipertrigliceridêmica (CHT) e avaliar sua associação com alterações metabólicas em adolescentes de baixa condição econômica.

Para tanto, o estudo transversal era composto por uma amostra probabilística de 1076 adolescentes entre 11 e 17 anos, de ambos os sexos, estudantes de escolas públicas. Os participantes foram submetidos à avaliação antropométrica (peso, altura e circunferência da cintura) e à dosagem dos níveis de colesterol total, LDL-C, HDL-C, colesterol não HDL, triglicérides (TG) e glicemia de jejum. Foram obtidas informações referentes às condições econômicas das famílias dos participantes. O fenótipo CHT foi definido pela presença simultânea da circunferência da cintura aumentada (\geq percentil 90 por idade e sexo) e dos níveis séricos de triglicérides elevados (≥ 100 mg/dL) (Conceição-Machado *et al.*, 2013).

Com o objetivo de avaliar o sobrepeso em adolescentes, analisamos uma única escola, sorteada aleatoriamente, denominada por Colégio Estadual Carneiro Ribeiro Classe IV, totalizando uma amostra 54 adolescentes. A variável de interesse é a razão entre a circunferência da cintura (cm) e a altura (cm) à qual se distribui de forma contínua no intervalo (0,1), denotada por y_t . Como dito na aplicação 1, esta variável é considerada um indicador de obesidade abdominal.

Para este conjunto de dados foram avaliadas conjuntamente com a variável de interesse as variáveis: índice de massa corporal, IMC, (z_{t1}), idade, em anos (z_{t2}), gênero do adolescente (z_{t3}), triglicérides (z_{t4}), colesterol total (w_t) e colesterol LDL (x_t).

A variável colesterol total (w_t) é considerada como uma substituta do colesterol LDL (x_t). Portanto, parece adequado considerar um modelo de regressão beta com erros de medida para ajustar os dados. Note que o erro de medição de w_t não é de interesse primário, mas sim das quantidades desconhecidas dos outros componentes (HDL e VLDL) que produzem erro de medição. Consideramos as transformações $CT/100$ e $LDL/100$ para w_t e x_t , respectivamente.

Com base na Figura 6.21, nota-se que a variável resposta apresenta valores concentrados próximos ao valor médio do intervalo unitário, 0.50, sendo que 75% estão acima de 0.40 e o valor máximo é de 0.66. Note que o valor médio e a mediana são bem próximos, $\bar{Y} = 0.45$ e $Md = 0.43$, respectivamente. Vale salientar que existem três observações *outliers*.

A Figura 6.22 apresenta as medidas resumo das variáveis analisadas. Note que 25% dos adolescente apresentam o índice de massa corporal (IMC) acima de 22.16, além disso, o IMC apresentou três *outliers*, Figura 6.22(a). A idade dos adolescentes variou de 11.41 a

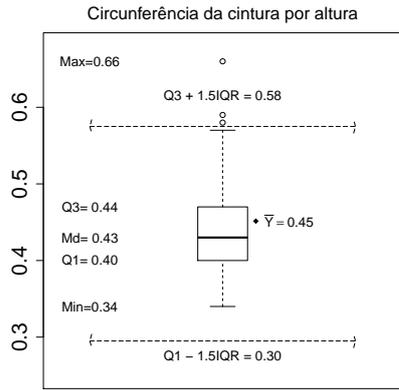


Figura 6.21: *Boxplot* da variável de interesse é a razão entre a circunferência da cintura (cm) e a altura (cm).

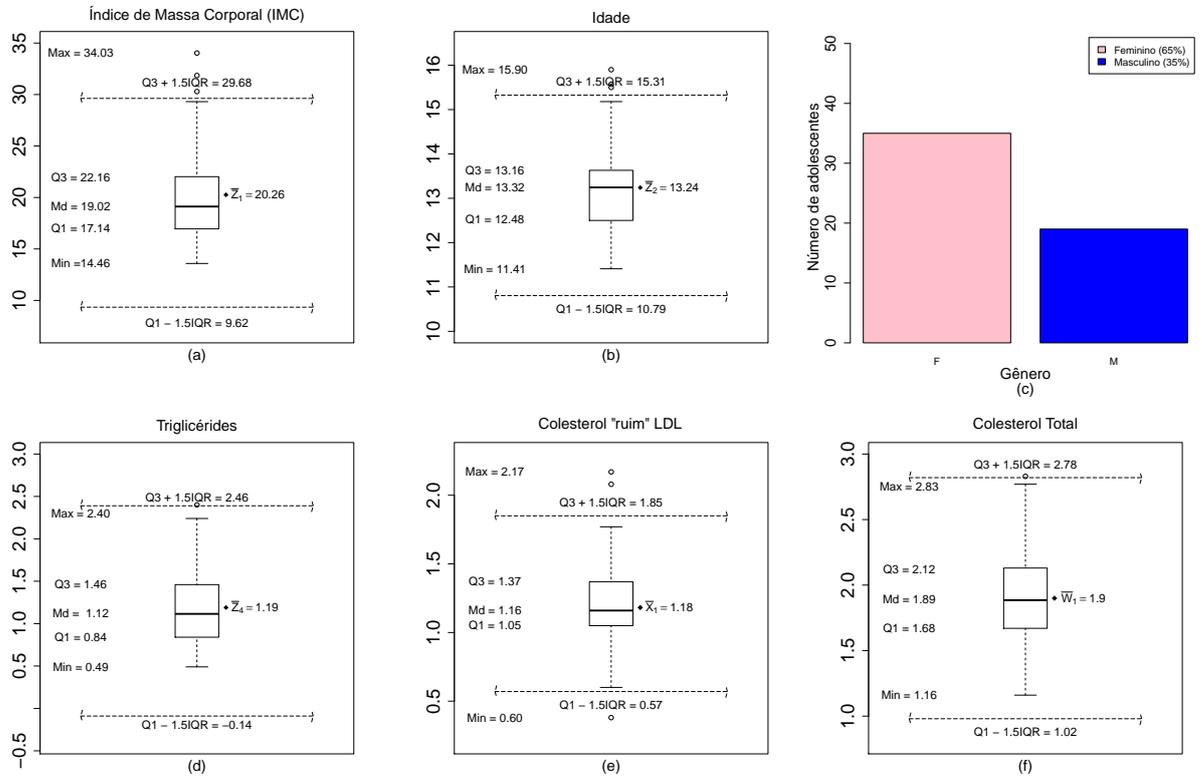


Figura 6.22: *Boxplots* com as medidas resumo das covariáveis (a),(b),(d)-(e) e gráfico de colunas do gênero dos adolescentes (c).

15.90 anos, com a mediana e a média da idade bem próximas, $Md = 13.32$ e $\bar{Z}_2 = 13.24$ anos. 65% dos adolescentes neste estudo são do gênero feminino, como mostra a Figura 6.22(c). É interessante notar na Figura 6.22(d) que a variável triglicérides apresenta uma dispersão considerável, com ocorrências de valores extremos tal que o mínimo é 0.49 e o máximo é 2.40. Em relação ao colesterol LDL, é verificado que a média e a mediana encontram-se bem próximas, $\bar{X} = 1.18$ e $Md = 1.16$, respectivamente, apresentando dois

outliers. Em relação ao colesterol total, verifica-se que está variando de 116.0 a 283.0 mg/dl.

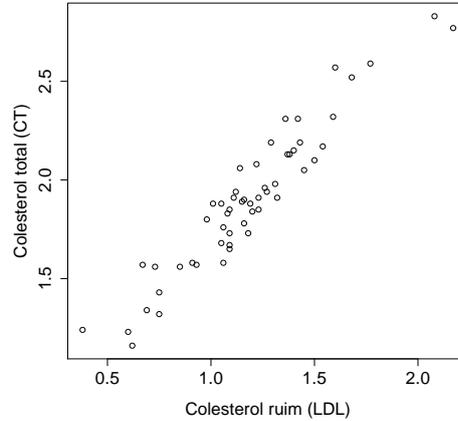


Figura 6.23: Gráfico de dispersão do colesterol total (w_t) e colesterol “ruim” LDL (x_t) ambas medições em mg/dl.

A Figura 6.23 mostra a existência de uma relação linear entre o colesterol total e o colesterol “ruim” LDL. Essa característica é útil para a escolha do modelo com erros de medida.

Admitimos agora que y_1, y_2, \dots, y_n são variáveis aleatórias independentes tais que y_t , para todo $t = 1, \dots, n$, tem distribuição beta com média μ_t e parâmetro de dispersão ϕ_t .

Inicialmente, ajustamos o modelo beta com erros de medida com dispersão variável

$$w_t = \tau_0 + \tau_1 x_t + e_t \quad (6.7)$$

em que $t = 1, \dots, n$, $x_t \sim N(\mu_x, \sigma_x^2)$ e $e_t \sim N(0, \sigma_e^2)$. Aqui, o vetor de parâmetros de interesse $\theta = (\alpha_0, \alpha_1, \beta_1, \gamma_0)^\top$ e o vetor de parâmetros de perturbação é $\xi = (\tau_0, \tau_1, \mu_x, \sigma_x^2, \sigma_e^2)^\top$. É possível determinar as estimativas de τ_0 , τ_1 e σ_e^2 a partir do modelo (6.7), dado que nosso conjunto de dados contém as medições de colesterol total e colesterol “ruim” LDL para todos os indivíduos. As estimativas dos parâmetros e respectivos erros-padrão são: $\hat{\tau}_0 = 0.68861$ (0.05739) e $\hat{\tau}_1 = 1.02766$ (0.04669). Considerando os dados analisados temos que $\hat{\sigma}_x^2 = 0.117$ e $\hat{\sigma}_e^2 = 0.013$, portanto o coeficiente de confiabilidade, dado por $k_x = \sigma_x^2 / (\sigma_x^2 + \sigma_e^2)$, é $\hat{k}_x = 0.90$. Ou seja, o erro de medida é considerado baixo.

6.3.1 Modelos ajustados

Inicialmente, ajustou-se o modelo de regressão beta com erros de medidas considerando todas as covariáveis dado por

Tabela 6.19: Estimativas, erros-padrão e p -valores obtidos pelos métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa).

Método	Parâmetro	Estimativa	Erro-padrão	p -valor
<i>Naive</i>	α_0	-1.078	0.167	0.000
	α_1	0.053	0.003	0.000
	α_2	-0.031	0.012	0.007
	α_3	0.019	0.020	0.353
	α_4	-0.018	0.037	0.626
	β_1	0.114	0.031	0.000
	γ_0	8.515	3.956	0.031
	γ_1	-0.065	0.053	0.217
	γ_2	0.041	0.268	0.879
	γ_3	0.927	0.502	0.065
RC	α_0	-1.102	0.232	0.000
	α_1	0.053	0.004	0.000
	α_2	-0.031	0.017	0.062
	α_3	0.019	0.023	0.402
	α_4	-0.018	0.053	0.738
	β_1	0.127	0.056	0.025
	γ_0	8.515	6.853	0.214
	γ_1	-0.065	0.079	0.409
	γ_2	0.041	0.469	0.931
	γ_3	0.927	0.802	0.248
MVa	α_0	-1.097	0.172	0.000
	α_1	0.053	0.003	0.000
	α_2	-0.031	0.012	0.007
	α_3	0.019	0.021	0.367
	α_4	-0.015	0.036	0.677
	β_1	0.123	0.034	0.000
	γ_0	8.354	4.055	0.039
	γ_1	-0.066	0.054	0.220
	γ_2	0.056	0.274	0.837
	γ_3	0.930	0.519	0.073
MPVa	α_0	-1.097	0.145	0.000
	α_1	0.053	0.004	0.000
	α_2	-0.031	0.011	0.003
	α_3	0.019	0.018	0.304
	α_4	-0.015	0.040	0.712
	β_1	0.123	0.036	0.001
	γ_0	8.354	4.271	0.050
	γ_1	-0.066	0.046	0.157
	γ_2	0.056	0.285	0.844
	γ_3	0.930	0.573	0.105
γ_4	-1.222	0.604	0.043	

$$\begin{aligned}\log\left(\frac{\mu_t}{1-\mu_t}\right) &= \alpha_0 + \alpha_1 z_{t1} + \alpha_2 z_{t2} + \alpha_3 z_{t3} + \alpha_4 z_{t4} + \beta_1 x_t, \\ \log(\phi_t) &= \gamma_0 + \gamma_1 z_{t1} + \gamma_2 z_{t2} + \gamma_3 z_{t3} + \gamma_4 z_{t4}, \\ w_t &= \tau_0 + \tau_1 x_t + e_t,\end{aligned}$$

em que $t = 1, \dots, 54$, z_{t1} denota o índice de massa corporal (IMC), z_{t2} é idade mensurada em anos, z_{t3} representa o gênero dos adolescentes, z_{t4} é a variável triglicérides, x_t e w_t denotam o colesterol “ruim” LDL e colesterol total, respectivamente, e utilizamos $Q = 50$ pontos de quadratura.

As estimativas, erros-padrão e p -valores relativos ao processo de inferência sobre os parâmetros obtidos a partir dos métodos de máxima verossimilhança aproximada (MVa), máxima pseudo-verossimilhança aproximada (MPVa), calibração da regressão (RC) e *naive* e o modelo sem erros são apresentadas nas Tabelas 6.19 e 6.20. Para a comparação, utilizamos o modelo considerado sem erros, ou seja, quando a covariável medida com erro é o próprio colesterol “ruim” LDL (x_t).

Tabela 6.20: Estimativas, erros-padrão e p -valores obtidos pelo modelo de regressão beta tradicional.

Método	Parâmetro	Estimativa	Erro-padrão	p -valor
Sem erros	α_0	-0.9925	0.1484	0.0000
	α_1	0.0516	0.0030	0.0000
	α_2	-0.0319	0.0103	0.0021
	α_3	0.0207	0.0194	0.2851
	α_4	0.0056	0.0275	0.8388
	β_1	0.1190	0.0264	0.0000
	γ_0	8.5753	2.7267	0.0017
	γ_1	-0.0686	0.0491	0.1619
	γ_2	0.0383	0.2008	0.8486
	γ_3	0.8079	0.4345	0.0630
	γ_4	-1.1404	0.4535	0.0119

Note que as variáveis sexo e triglicérides, associadas ao submodelo da média, e as variáveis IMC e idade, associadas ao submodelo da dispersão, não são significantes ao nível de 10%, tanto no modelo de regressão beta quanto nos métodos utilizados para estimação dos estimadores associados ao modelo de regressão beta linear com erros de medidas. Verifica-se também, que as estimativas estão bem consistentes nos quatro métodos de estimação apresentados na Tabela 6.19. Porém, quando comparamos com os resultados da Tabela 6.20 verificamos que há uma diferença na estimativa do parâmetro α_4 considerando o ajuste do modelo de regressão sem erros.

Após identificar as variáveis significativas, ao nível de 10% de significância, no modelo linear com erros de medida, ajustamos um novo modelo considerando apenas estas variáveis.

Tabela 6.21: Estimativas, erros-padrão e p -valores obtidos pelo modelo de regressão beta com covariáveis significativas.

Método	Parâmetro	Estimativa	Erro-padrão	p -valor
Sem erros	α_0	-0.9876	0.1468	0.0000
	α_1	0.0536	0.0025	0.0000
	α_2	-0.0328	0.0100	0.0011
	β_1	0.1072	0.0269	0.0001
	γ_0	7.9769	0.5262	0.0000
	γ_3	0.8516	0.4027	0.0344
	γ_4	-1.4391	0.3930	0.0003

Tabela 6.22: Estimativas, erros-padrão e p -valores obtidos pelos métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa).

Método	Parâmetro	Estimativa	Erro-padrão	p -valor
<i>Naive</i>	α_0	-1.037	0.163	0.000
	α_1	0.054	0.003	0.000
	α_2	-0.034	0.010	0.001
	β_1	0.096	0.027	0.000
	γ_0	7.847	0.549	0.000
	γ_3	0.955	0.450	0.034
	γ_4	-1.383	0.455	0.002
RC	α_0	-1.057	0.205	0.000
	α_1	0.054	0.003	0.000
	α_2	-0.034	0.014	0.018
	β_1	0.106	0.041	0.009
	γ_0	7.847	0.558	0.000
	γ_3	0.955	0.447	0.033
	γ_4	-1.383	0.459	0.003
MVa	α_0	-1.051	0.166	0.000
	α_1	0.054	0.003	0.000
	α_2	-0.034	0.010	0.001
	β_1	0.105	0.030	0.000
	γ_0	7.883	0.568	0.000
	γ_3	0.965	0.466	0.038
	γ_4	-1.395	0.465	0.003
MPVa	α_0	-1.051	0.159	0.000
	α_1	0.054	0.003	0.000
	α_2	-0.034	0.011	0.002
	β_1	0.105	0.030	0.000
	γ_0	7.883	0.518	0.000
	γ_3	0.965	0.491	0.049
	γ_4	-1.395	0.426	0.001

Assim, considere o seguinte modelo de regressão beta linear com erros de medida:

$$\begin{aligned}\log\left(\frac{\mu_t}{1-\mu_t}\right) &= \alpha_0 + \alpha_1 z_{t1} + \alpha_2 z_{t2} + \beta_1 x_t, \\ \log(\phi_t) &= \gamma_0 + \gamma_3 z_{t3} + \gamma_4 z_{t4}, \\ w_t &= \tau_0 + \tau_1 x_t + e_t,\end{aligned}$$

em que $t = 1, \dots, 54$, z_{t1} denota o índice de massa corporal (IMC), z_{t2} é idade mensurada em anos, z_{t3} representa o gênero dos adolescentes e z_{t4} é a variável triglicérides, x_t e w_t denotam o colesterol “ruim” LDL e colesterol total, respectivamente, sendo utilizado $Q = 50$ pontos de quadratura.

Novamente avaliamos as estimativas, erros-padrão e p -valores obtidos pelo modelo de regressão beta tradicional (sem erros) e pelos métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa). Os resultados são apresentados nas Tabelas 6.21 e 6.22, respectivamente. Com base nestas tabelas verifica-se que todas as covariadas são consideradas significativas, ao nível de 5%. Note que as estimativas estão bem consistentes tanto nos quatro métodos de estimação apresentados na Tabela 6.22 quanto no ajuste do modelo sem erros dado na Tabela 6.21.

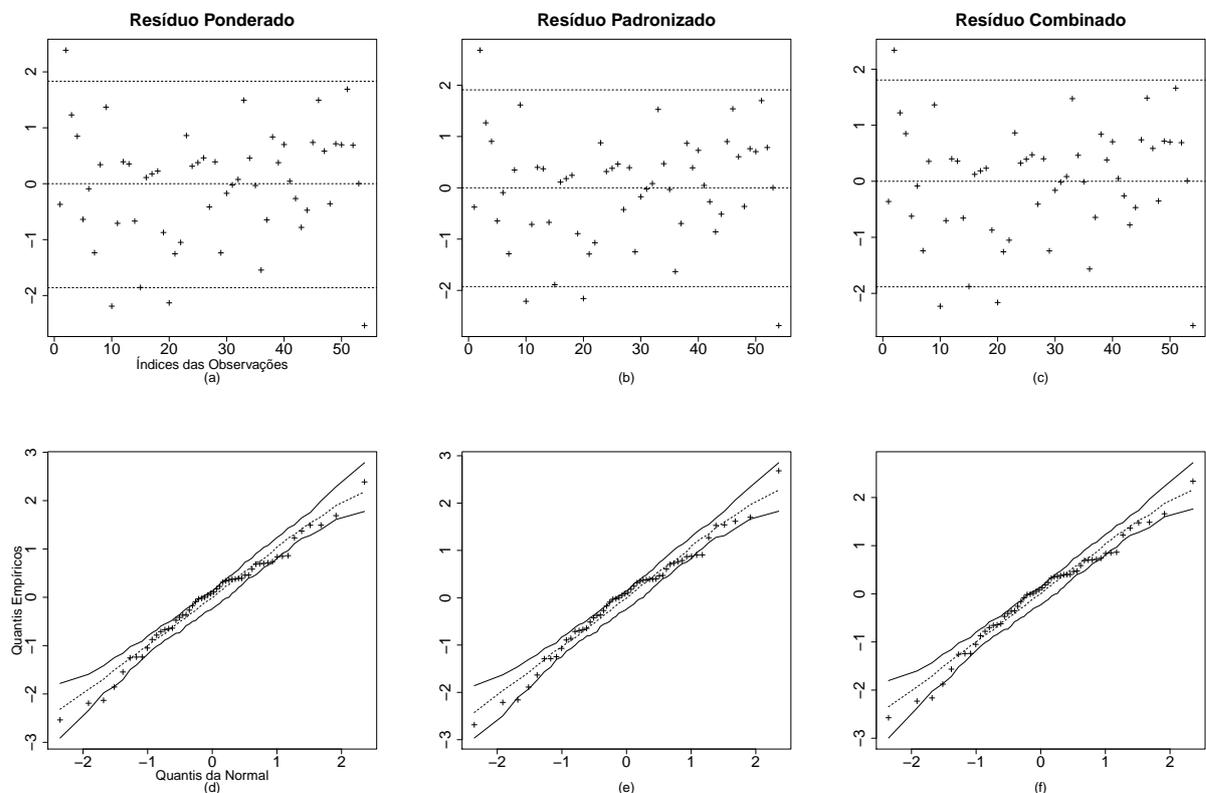


Figura 6.24: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o **método padrão** para o ajuste do modelo de regressão beta tradicional. Modelo 2.

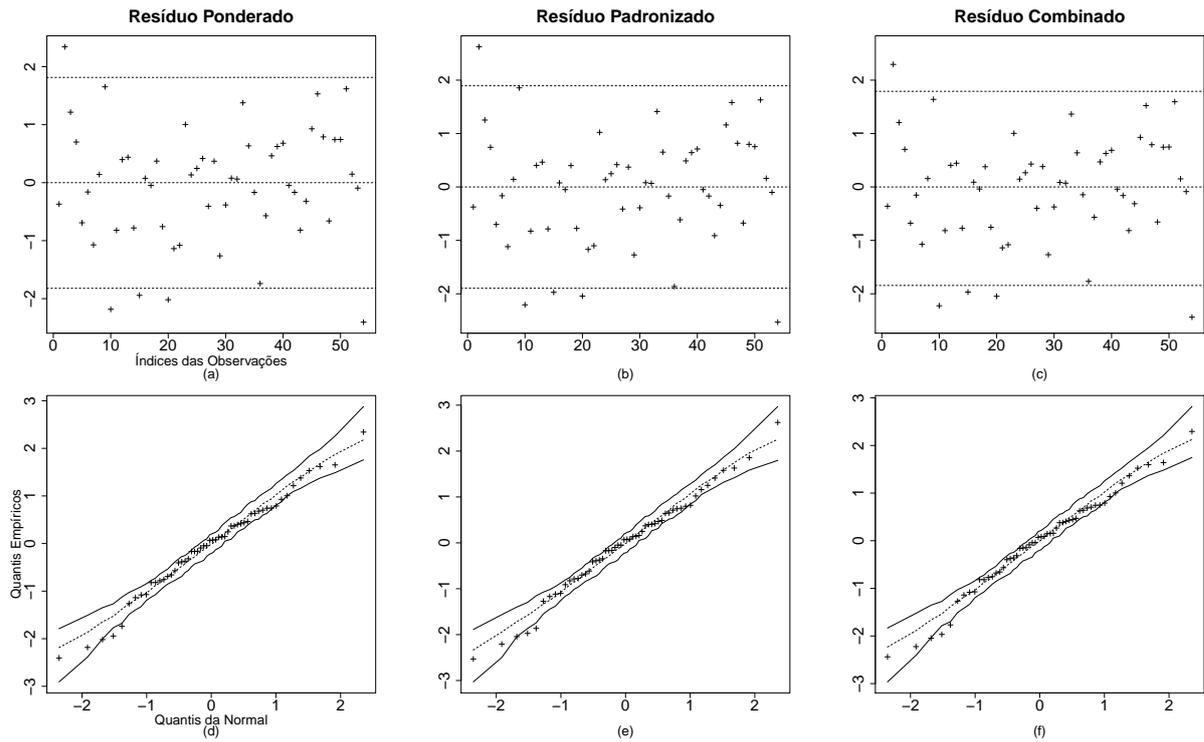


Figura 6.25: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o **método naive** para o ajuste do modelo de regressão beta com erros de medidas. Modelo 2.

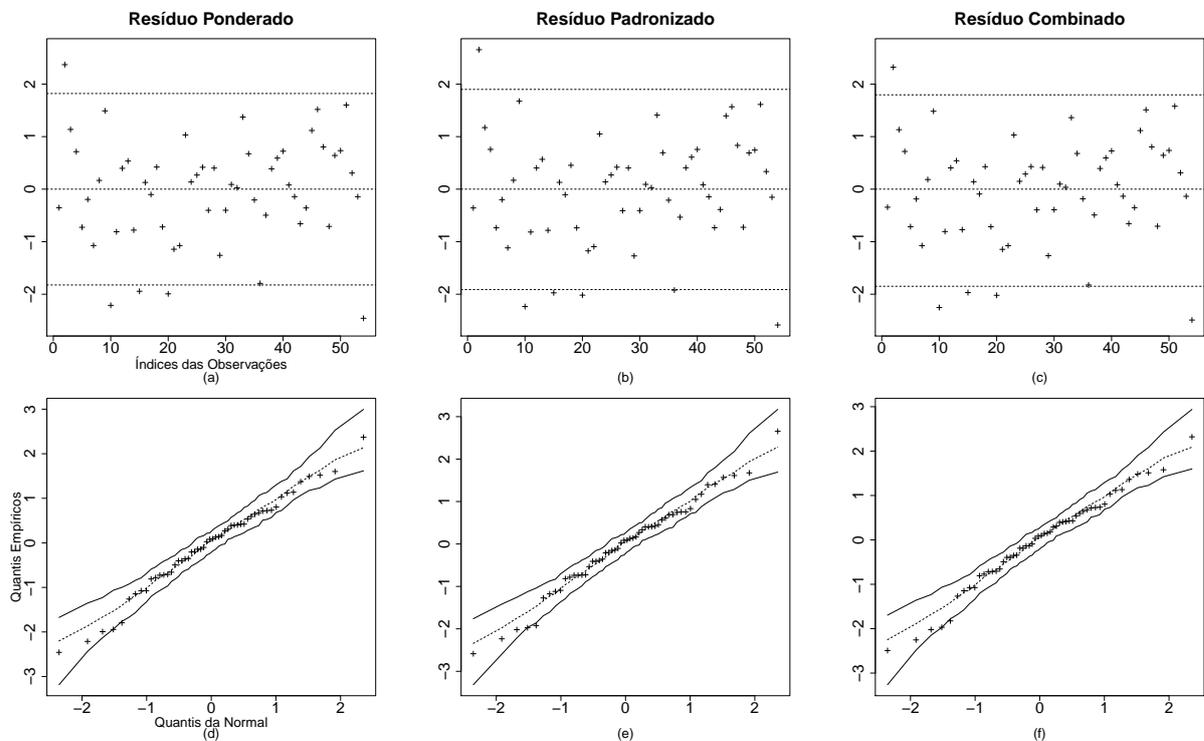


Figura 6.26: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o **método da calibração da regressão** para o ajuste do modelo de regressão beta com erros de medidas. Modelo 2.

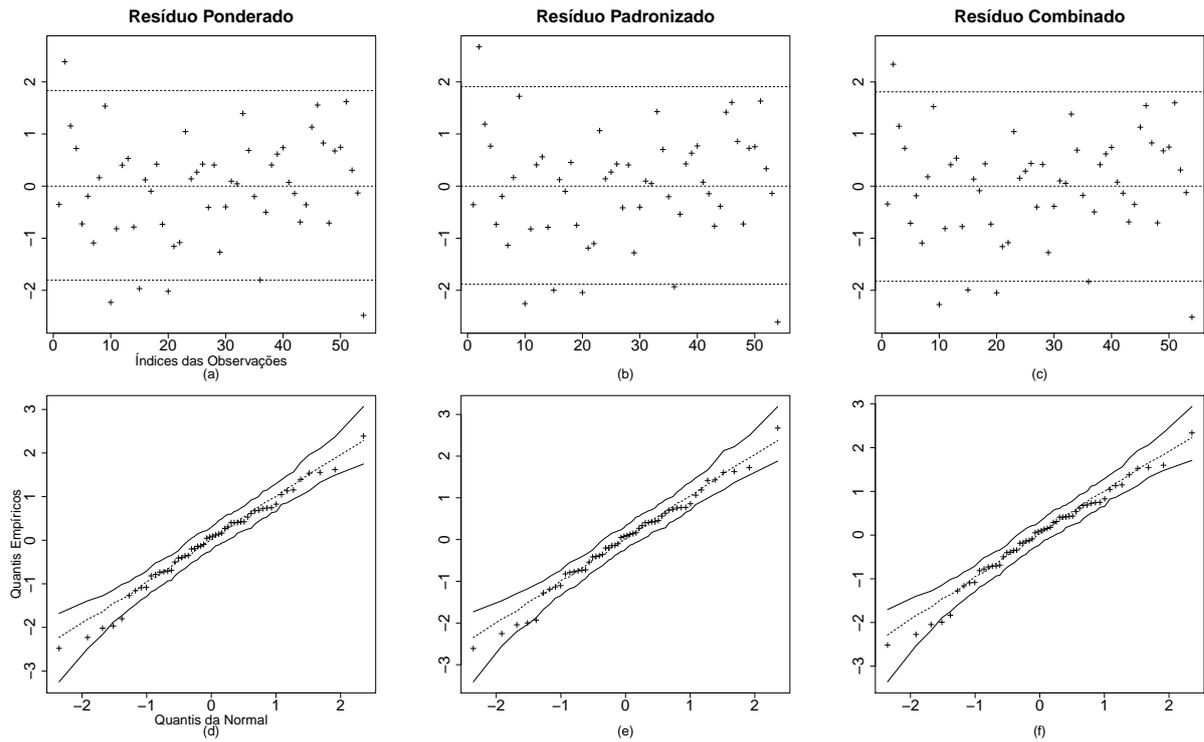


Figura 6.27: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o **método da máxima verossimilhança aproximada** para o ajuste do modelo de regressão beta com erros de medidas. Modelo 2.

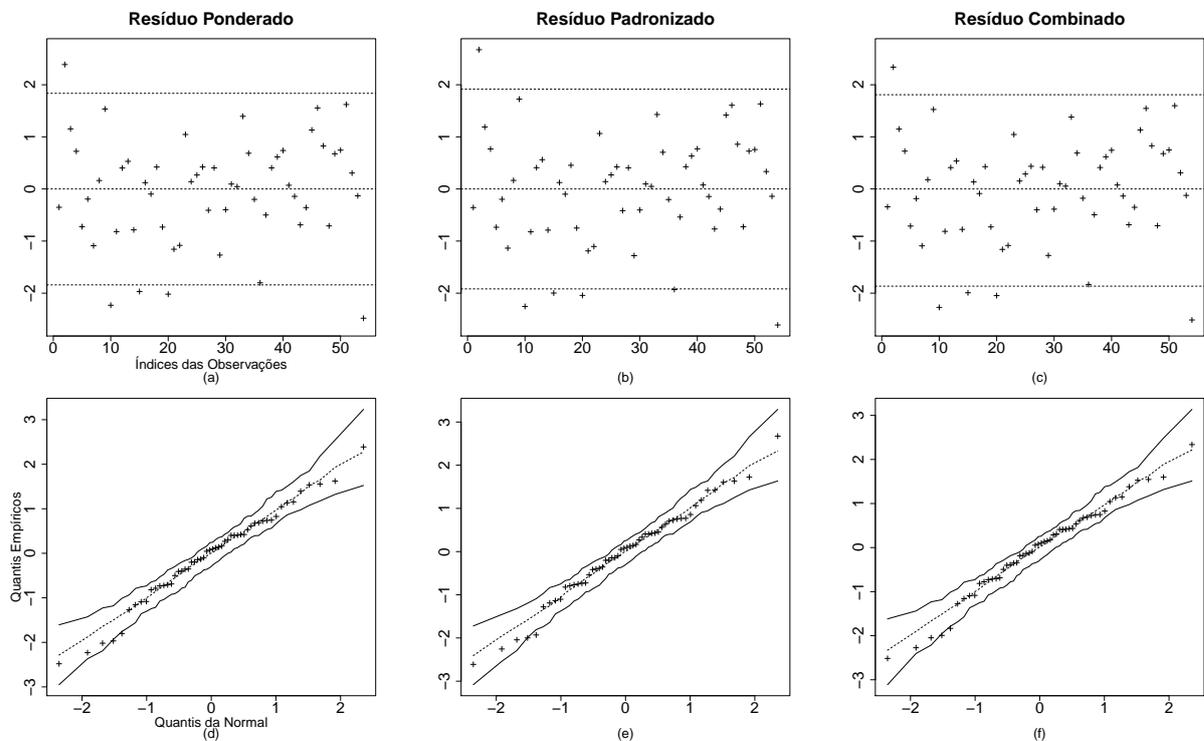


Figura 6.28: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o **método da máxima pseudo-verossimilhança aproximada** para o ajuste do modelo de regressão beta com erros de medidas. Modelo 2.

As Figuras 6.24,6.25, 6.26, 6.27 e 6.28 apresentam os resíduos e os envelopes simulados considerando os métodos padrão, *naive*, calibração da regressão, máxima verossimilhança aproximada e máxima pseudo-verossimilhança aproximada, respectivamente. De uma forma geral, os gráficos de resíduos, em especial os gráficos de envelopes, evidenciam que o modelo beta com erros nas variáveis apresenta um ajuste melhor que o modelo beta tradicional.

Tabela 6.23: Estimativas, erros-padrão e p -valores obtidos pelos métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa). Modelo 6.8.

Método	Parâmetro	Estimativa	Erro-padrão	p -valor
<i>Naive</i>	α_0	-1.189	0.173	0.000
	α_1	0.056	0.003	0.000
	α_2	-0.029	0.010	0.004
	α_4	-0.059	0.034	0.084
	β_1	0.121	0.025	0.000
	γ_0	6.458	0.245	0.000
	γ_3	1.019	0.474	0.031
RC	α_0	-1.215	0.265	0.000
	α_1	0.056	0.004	0.000
	α_2	-0.029	0.015	0.051
	α_4	-0.059	0.042	0.156
	β_1	0.135	0.048	0.005
	γ_0	6.458	0.300	0.000
	γ_3	1.019	0.562	0.070
MVa	α_0	-1.189	0.173	0.000
	α_1	0.056	0.003	0.000
	α_2	-0.029	0.010	0.004
	α_4	-0.059	0.034	0.084
	β_1	0.121	0.025	0.000
	γ_0	6.458	0.245	0.000
	γ_3	1.019	0.474	0.031
MPVa	α_0	-1.189	0.154	0.000
	α_1	0.056	0.003	0.000
	α_2	-0.029	0.010	0.004
	α_4	-0.059	0.036	0.102
	β_1	0.121	0.026	0.000
	γ_0	6.458	0.234	0.000
	γ_3	1.019	0.550	0.064
	γ_4	-2.335	0.726	0.001

No entanto, pesquisadores da área de Nutrição argumentam que a variável triglicérides é importante para explicar a média da variável resposta. Desta forma, avaliamos um novo

modelo que considera como covariável o logaritmo da covariável triglicérides. Assim, o modelo de regressão beta linear é dado por

$$\begin{aligned}\log\left(\frac{\mu_t}{1-\mu_t}\right) &= \alpha_0 + \alpha_1 z_{t1} + \alpha_2 z_{t2} + \alpha_4 \log(z_{t4}) + \beta_1 x_t, \\ \log(\phi_t) &= \gamma_0 + \gamma_3 z_{t3} + \gamma_4 \log(z_{t4}), \\ w_t &= \tau_0 + \tau_1 x_t + e_t,\end{aligned}\tag{6.8}$$

em que $t = 1, \dots, 54$, z_{t1} denota o índice de massa corporal (IMC), z_{t2} é idade mensurada em anos, z_{t3} representa o gênero do adolescente, z_{t4} denota a variável triglicérides, x_t e w_t denotam o colesterol “ruim” LDL e colesterol total, respectivamente. Utilizamos $Q = 50$ pontos de quadratura.

Os resultados inferenciais deste modelo são apresentados nas Tabelas 6.23 e 6.24, considerando os métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa) e o modelo de regressão beta tradicional (sem erros), respectivamente. Com base nestas tabelas, nota-se que o logaritmo da covariável triglicérides possa ser significativo ao nível de 10% em especial quando consideramos o método da máxima pseudo-verossimilhança aproximada.

Tabela 6.24: Estimativas, erros-padrão e p -valores obtidos pelo modelo de regressão beta. Modelo 6.8.

Método	Parâmetro	Estimativa	Erro-padrão	p -valor
Sem erros	α_0	-1.0512	0.1446	0.0000
	α_1	0.0543	0.0028	0.0000
	α_2	-0.0304	0.0097	0.0017
	α_4	-0.0334	0.0295	0.2577
	β_1	0.1204	0.0238	0.0000
	γ_0	6.4956	0.2424	0.0000
	γ_3	0.8849	0.4028	0.0280
	γ_4	-2.1945	0.4810	0.0000

É interessante notar que no modelo padrão a covariável logaritmo de triglicérides, $\log(z_{t4})$, não é significativa (Tabela 6.24) ao nível de 10% de significância. Note que as estimativas estão bem consistentes nos quatro métodos de estimação, como mostra a Tabela 6.23. É interessante notar que os métodos da máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa) apresentam valores das estimativas dos parâmetros idênticas. Já os métodos *naive* e calibração da regressão (RC) apresentam resultados similares para as estimativas dos parâmetros.

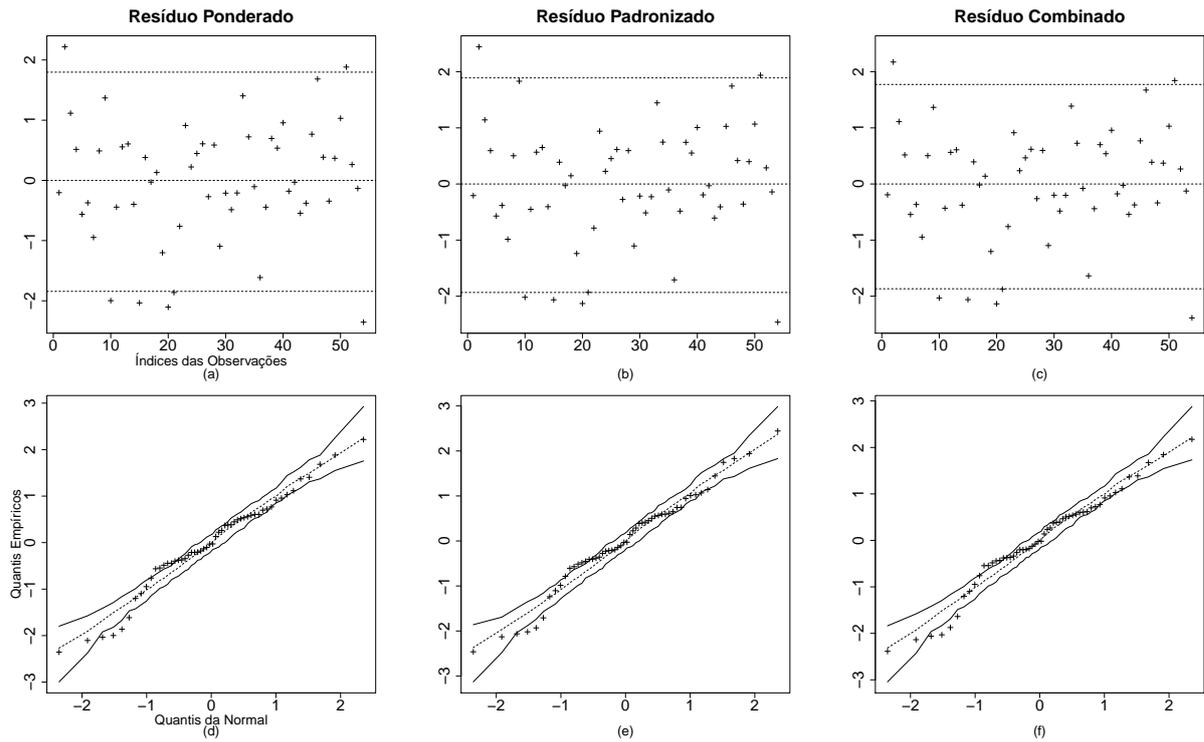


Figura 6.29: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o **método naive** para o ajuste do modelo de regressão beta com erros de medidas.

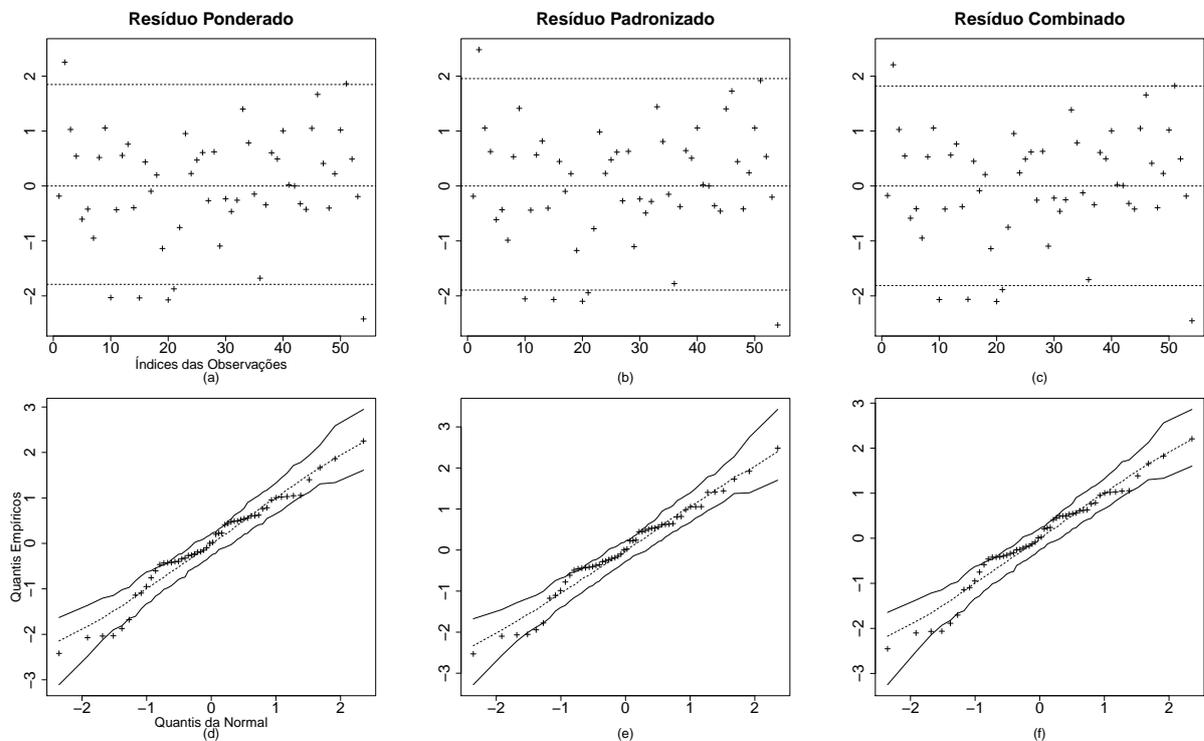


Figura 6.30: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o **método da calibração da regressão** para o ajuste do modelo de regressão beta com erros de medidas.

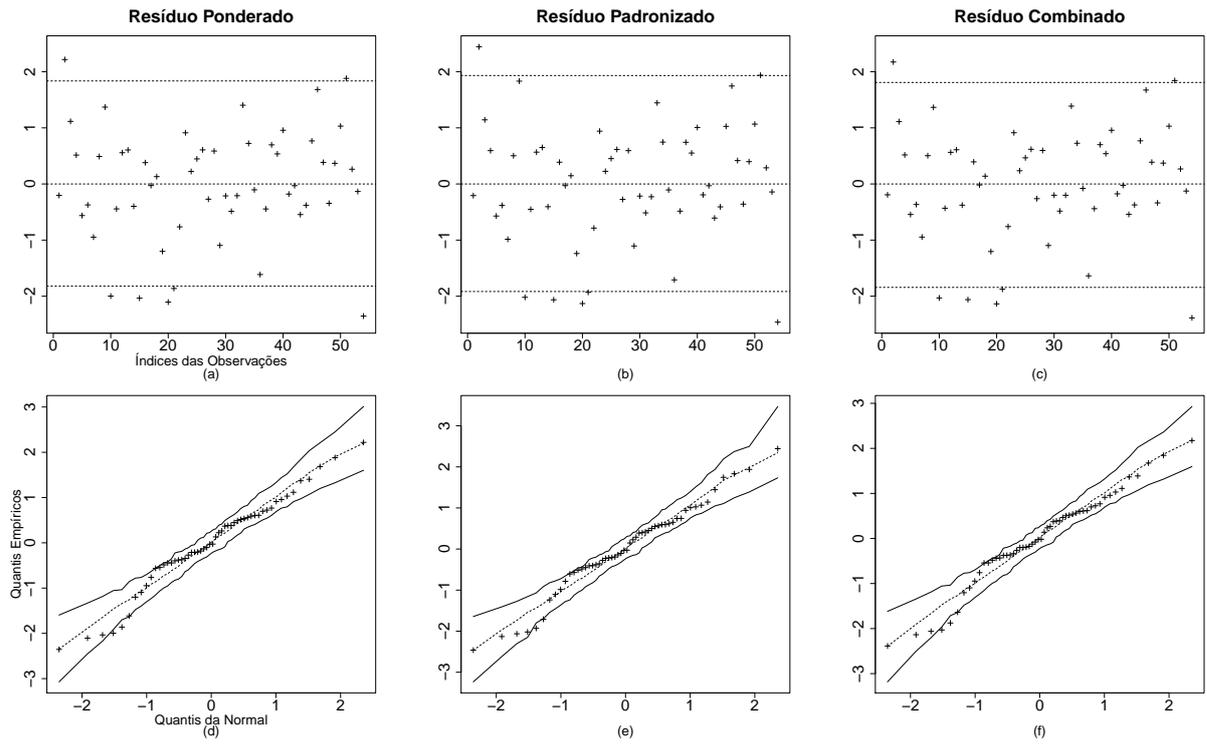


Figura 6.31: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o **método da máxima verossimilhança aproximada** para o ajuste do modelo de regressão beta com erros de medidas.

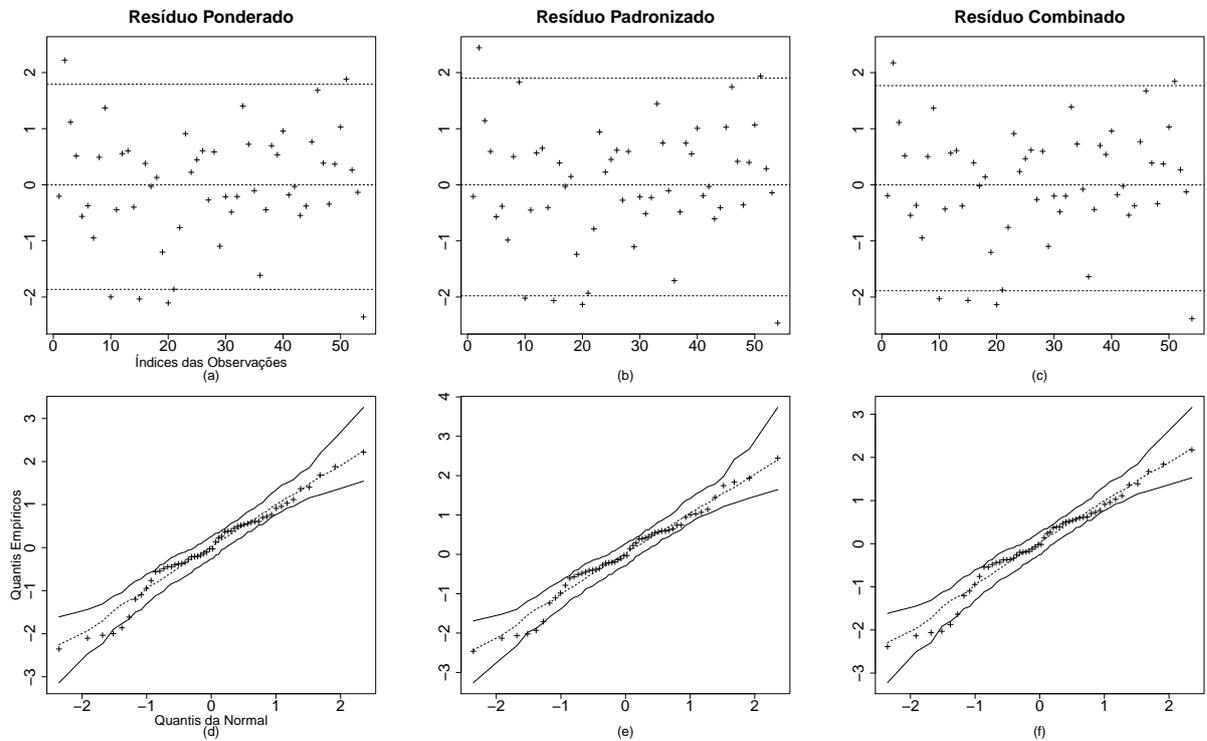


Figura 6.32: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o **método da máxima pseudo-verossimilhança aproximada** para o ajuste do modelo de regressão beta com erros de medidas.

Tabela 6.25: Estimativas, erros-padrão e p -valores obtidos pelos métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa).

Método	Parâmetro	Estimativa	Erro-padrão	p -valor
<i>Naive</i>	α_0	-1.260	0.178	0.000
	α_1	0.057	0.003	0.000
	α_2	-0.026	0.010	0.012
	α_4	-0.024	0.015	0.101
	α_5	-0.471	0.041	0.000
	β_1	0.118	0.024	0.000
	γ_0	6.526	0.249	0.000
	γ_3	0.928	0.479	0.053
	γ_4	-2.382	0.618	0.000
	RC	α_0	-1.285	0.222
α_1		0.057	0.004	0.000
α_2		-0.026	0.013	0.054
α_4		-0.024	0.018	0.185
α_5		-0.471	0.131	0.000
β_1		0.131	0.032	0.000
γ_0		6.526	0.264	0.000
γ_3		0.928	0.729	0.203
γ_4		-2.382	0.796	0.003
MVa		α_0	-1.264	0.182
	α_1	0.057	0.003	0.000
	α_2	-0.027	0.010	0.009
	α_4	-0.023	0.014	0.107
	α_5	-0.473	0.038	0.000
	β_1	0.127	0.027	0.000
	γ_0	6.576	0.263	0.000
	γ_3	0.903	0.495	0.068
	γ_4	-2.384	0.642	0.000
	MPVa	α_0	-1.264	0.148
α_1		0.057	0.003	0.000
α_2		-0.027	0.009	0.004
α_4		-0.023	0.013	0.062
α_5		-0.473	0.022	0.000
β_1		0.127	0.024	0.000
γ_0		6.576	0.247	0.000
γ_3		0.903	0.530	0.088
γ_4		-2.384	0.607	0.000

As Figuras 6.29, 6.33, 6.35 e 6.35 apresentam os resíduos e os envelopes simulados considerando os métodos *naive*, calibração da regressão, máxima verossimilhança aproximada e máxima pseudo-verossimilhança aproximada, respectivamente. Os gráficos dos resíduos evidenciam uma boa qualidade do ajuste do modelo que considera o $\log(z_{t4})$ tanto na média quanto na dispersão. Com base nestas conclusões ainda é possível considerar

uma função não linear para o modelo da média com base no $\log(z_{t4})$.

O modelo de regressão beta não linear com erro de medidas é dado por

$$\begin{aligned}\log\left(\frac{\mu_t}{1-\mu_t}\right) &= \alpha_0 + \alpha_1 z_{t1} + \alpha_2 z_{t2} + \alpha_4 \log(z_{t4} + \alpha_5) + \beta_1 x_t, \\ \log(\phi_t) &= \gamma_0 + \gamma_3 z_{t3} + \gamma_4 \log(z_{t4}), \\ w_t &= \tau_0 + \tau_1 x_t + e_t,\end{aligned}\tag{6.9}$$

em que $t = 1, \dots, 54$, z_{t1} denota o índice de massa corporal (IMC), z_{t2} é idade mensurada em anos, z_{t3} representa o gênero do adolescente e z_{t4} é a variável triglicérides, x_t e w_t denotam o colesterol “ruim” LDL e colesterol total, respectivamente. Utilizamos $Q = 50$ pontos de quadratura.

A Tabela 6.25 apresenta as estimativas, erros-padrão e p -valores obtidos pelos métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa).

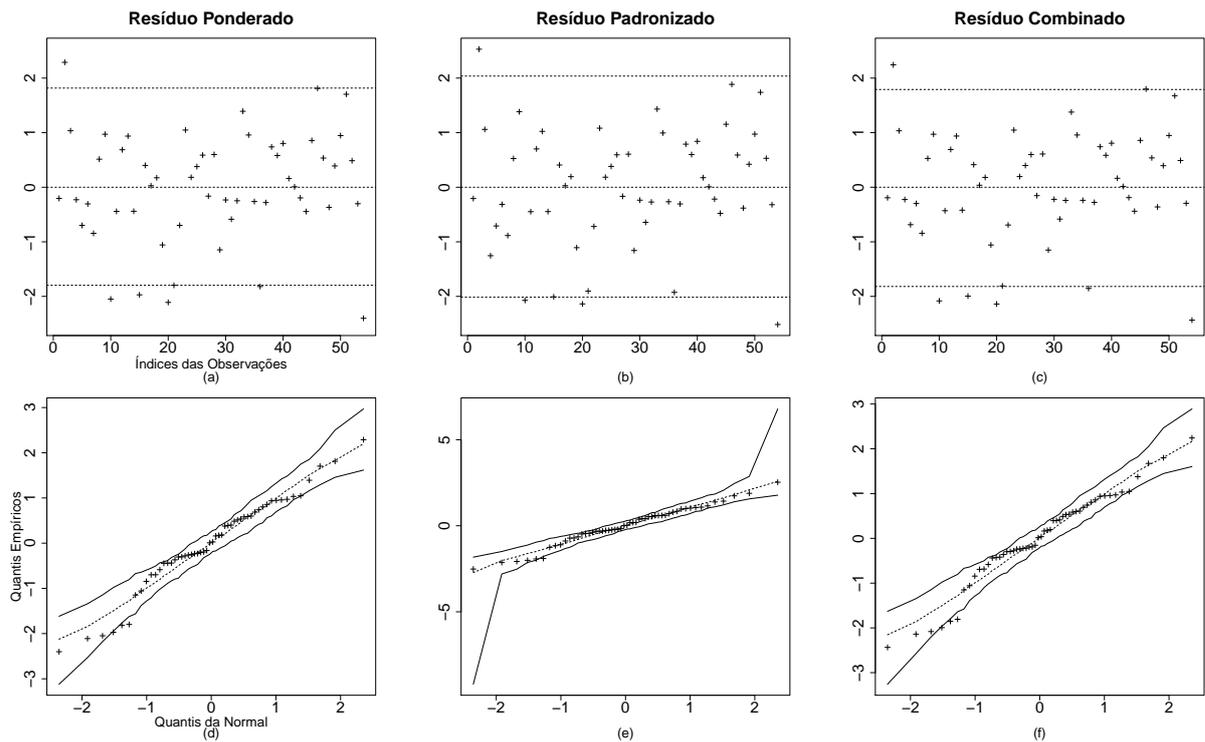


Figura 6.33: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o método da **calibração da regressão** para o ajuste do modelo de regressão beta com erros de medidas.

Com base nesta tabela podemos notar que o parâmetro associado à não linearidade (α_5) é altamente significativo considerando todos os métodos de estimação. É interessante notar também que o parâmetro α_4 relacionada a covariada triglicérides é mais significativo considerando o método MPVa que no modelo linear que utiliza o logaritmo de triglicérides.

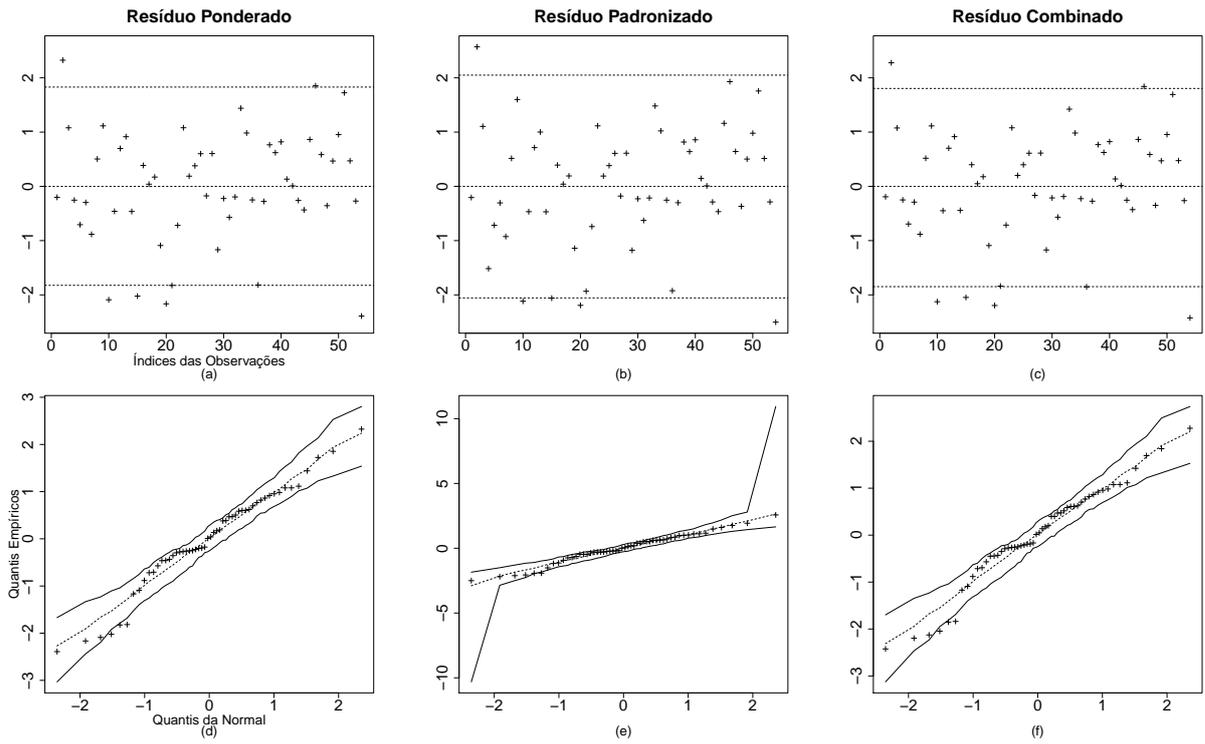


Figura 6.34: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o **método da máxima verossimilhança aproximada** para o ajuste do modelo de regressão beta com erros de medidas.

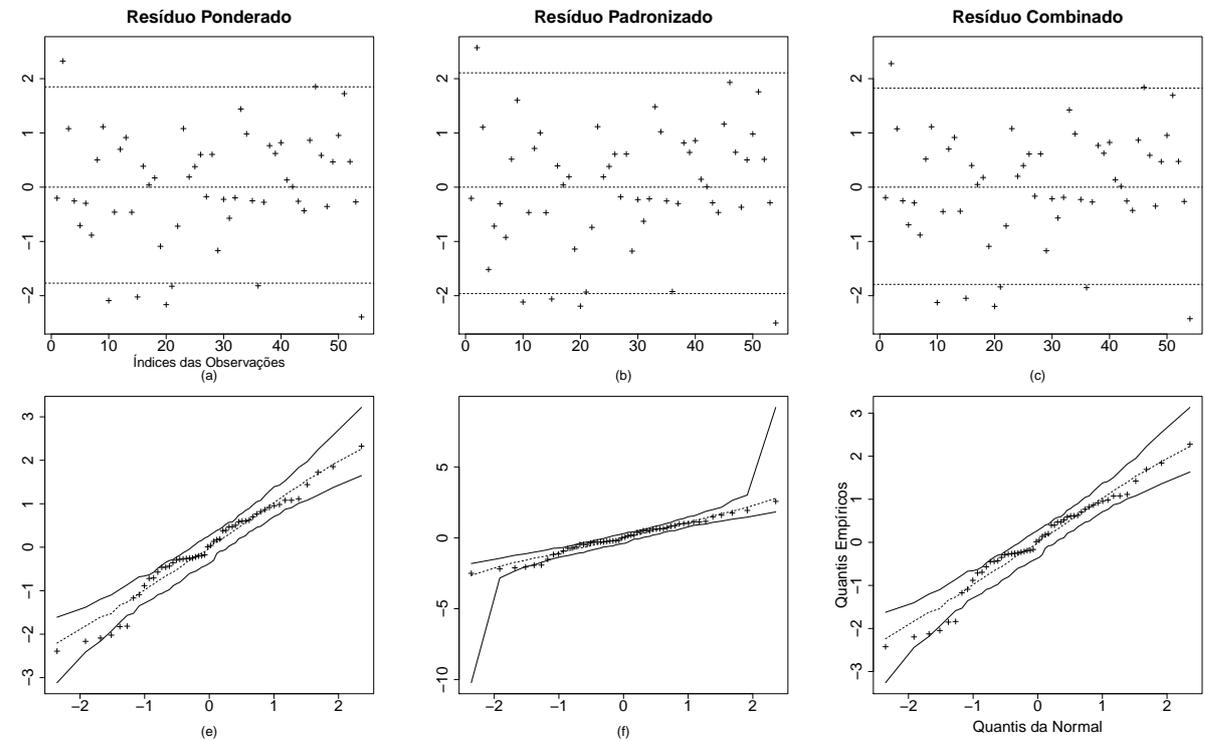


Figura 6.35: Resíduos e envelopes simulados considerando os resíduos ponderados, padronizados e combinados utilizando o **método da máxima pseudo-verossimilhança aproximada** para o ajuste do modelo de regressão beta com erros de medidas.

O contrário ocorre com o método MVA. Vale salientar que as estimativas são bem consistentes nos quatro métodos de estimação utilizados. As Figuras 6.33, 6.34 e 6.35 apresentam os gráficos de resíduos e envelopes simulados do modelo de regressão beta não linear com erros de medidas utilizando os métodos da calibração da regressão (RC), máxima verossimilhança aproximada (MVA) e máxima pseudo-verossimilhança aproximada (MPVa), respectivamente.

Devemos salientar a diferença entre os resíduos padronizado e combinado. Em geral eles apresentam comportamento parecido, mas esta aplicação não linear revelou que o resíduo padronizado que considera os elementos da matriz \mathbf{H} pode apresentar dificuldade para a construção do envelope simulado. Adicionalmente, com base nos gráficos de resíduos notamos que o ajuste do modelo não linear mostrou-se adequado aos dados considerando todos os métodos de estimação.

A Tabela 6.26 apresenta as estatísticas de predição e medidas de qualidade do ajuste dos modelos ajustado considerando os métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVA) e máxima pseudo-verossimilhança aproximada (MPVa). Os modelos são:

- MODELO 1:

$$\begin{aligned}\log\left(\frac{\mu_t}{1-\mu_t}\right) &= \alpha_0 + \alpha_1 z_{t1} + \alpha_2 z_{t2} + \beta_1 x_t \\ \log(\phi_t) &= \gamma_0 + \gamma_3 z_{t3} + \gamma_4 z_{t4},\end{aligned}$$

- MODELO 2:

$$\begin{aligned}\log\left(\frac{\mu_t}{1-\mu_t}\right) &= \alpha_0 + \alpha_1 z_{t1} + \alpha_2 z_{t2} + \alpha_4 \log(z_{t4}) + \beta_1 x_t \\ \log(\phi_t) &= \gamma_0 + \gamma_3 z_{t3} + \gamma_4 \log(z_{t4}),\end{aligned}$$

- MODELO 3:

$$\begin{aligned}\log\left(\frac{\mu_t}{1-\mu_t}\right) &= \alpha_0 + \alpha_1 z_{t1} + \alpha_2 z_{t2} + \alpha_4 \log(z_{t4} + \alpha_5) + \beta_1 x_t \\ \log(\phi_t) &= \gamma_0 + \gamma_3 z_{t3} + \gamma_4 \log(z_{t4}),\end{aligned}$$

em que $w_t = \tau_0 + \tau_1 x_t + e_t$, $t = 1, \dots, 54$, z_{t1} denota o índice de massa corporal (IMC), z_{t2} é idade mensurada em anos, z_{t3} representa o gênero do adolescente e z_{t4} é a variável triglicérides, x_t e w_t denotam o colesterol “ruim” LDL e colesterol total, respectivamente.

Tabela 6.26: Estatísticas de predição e medidas de qualidade do ajuste dos modelos ajustado considerando os métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa)

Medidas	MODELO 1			
	<i>Naive</i>	RC	MVa	MPVa
P^2	0.93136	0.93078	0.93112	0.93112
P_c^2	0.92260	0.92194	0.92233	0.92233
$P_{\alpha\gamma}^2$	0.93110	0.93046	0.93082	0.93082
$P_{\alpha\gamma c}^2$	0.92230	0.92159	0.92198	0.92198
ps- R^2	0.89432	0.89351	0.89364	0.89364
ps- R_c^2	0.88083	0.87991	0.88006	0.88006
R_{LR}^2	0.91585	0.96317	0.91565	0.91565
R_{LRc}^2	0.90000	0.95623	0.89976	0.89976
Medidas	MODELO 2			
	<i>Naive</i>	RC	MVa	MPVa
P^2	0.94981	0.94971	0.94981	0.94981
P_c^2	0.94217	0.94206	0.94217	0.94217
$P_{\alpha\gamma}^2$	0.94967	0.94953	0.94967	0.94967
$P_{\alpha\gamma c}^2$	0.94201	0.94185	0.94201	0.94201
ps- R^2	0.88969	0.88814	0.88969	0.88969
ps- R_c^2	0.87291	0.87111	0.87291	0.87291
R_{LR}^2	0.92056	0.96523	0.92056	0.92056
R_{LRc}^2	0.90254	0.95735	0.90254	0.90254
Medidas	MODELO 3			
	<i>Naive</i>	RC	MVa	MPVa
P^2	0.91265	0.91810	0.89530	0.89531
P_c^2	0.89712	0.90564	0.87937	0.87938
$P_{\alpha\gamma}^2$	0.91363	0.91895	0.89666	0.89667
$P_{\alpha\gamma c}^2$	0.89827	0.90662	0.88094	0.88095
ps- R^2	0.89201	0.89051	0.89130	0.89130
ps- R_c^2	0.87281	0.87385	0.87476	0.87476
R_{LR}^2	0.92303	0.96631	0.92242	0.92242
R_{LRc}^2	0.90240	0.95867	0.90482	0.90482

Com base nesta tabela notamos que entre os modelo lineares os que apresentam melhor modelo preditivo e melhor qualidade de ajuste é o modelo 2 o qual considera o logaritimo de triglicerides com covariada tanto no modelo da média quanto no modelo da dispersão.

Em relação ao modelo não linear considerado temos uma boa qualidade de ajuste tanto baseado nos valores do pseudo- R_2 quanto nos valores de R_{LR}^2 . As medidas de predição também estão próximas de um, apesar de estarem inferiores ao do modelo linear, mas não podemos compará-las diretamente. Desta forma, tanto o modelo 2 quanto o modelo 3 são boas opções para o ajuste dos dados de sobrepeso em adolescentes, cuja variável de interesse é a circunferência da cintura por altura e as covariáveis são IMC, idade, gênero do adolescente, triglicérides e colesterol total.

Tabela 6.27: Limites inferiores (LI), limites superiores (LS) e comprimento (Comp) dos intervalos de confiança considerando os níveis 90%, 95% e 99% de confiança, para os métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MPVa) e máxima pseudo-verossimilhança aproximada (MPVa). Modelo 3.

Métodos	Parâmetros	IC 90%			IC 95%			IC 99%		
		LI	LS	Comp	LI	LS	Comp	LI	LS	Comp
<i>Naive</i>	α_0	-1.55353	-0.96655	0.58699	-1.60976	-0.91032	0.69944	-1.71965	-0.80043	0.91922
	α_1	0.05173	0.06160	0.00987	0.05079	0.06254	0.01176	0.04894	0.06439	0.01545
	α_2	-0.04240	-0.00894	0.03346	-0.04560	-0.00573	0.03987	-0.05187	0.00053	0.05240
	α_4	-0.04836	0.00005	0.04841	-0.05300	0.00469	0.05769	-0.06206	0.01376	0.07582
	α_5	-0.53889	-0.40378	0.13511	-0.55183	-0.39084	0.16099	-0.57713	-0.36555	0.21158
	β_1	0.07837	0.15800	0.07964	0.07074	0.16563	0.09489	0.05583	0.18054	0.12471
RC	γ_0	6.11726	6.93483	0.81757	6.03895	7.01315	0.97420	5.88589	7.16620	1.28031
	γ_3	0.14057	1.71493	1.57436	-0.01023	1.86573	1.87597	-0.30497	2.16047	2.46544
	γ_4	-3.39873	-1.36577	2.03295	-3.59346	-1.17104	2.42241	-3.97405	-0.79045	3.18359
	α_0	-1.65041	-0.92002	0.73039	-1.72037	-0.85006	0.87031	-1.85711	-0.71332	1.14379
	α_1	0.05057	0.06276	0.01219	0.04941	0.06392	0.01452	0.04712	0.06621	0.01908
	α_2	-0.04762	-0.00371	0.04391	-0.05183	0.00049	0.05232	-0.06005	0.00871	0.06876
	α_4	-0.05414	0.00583	0.05997	-0.05988	0.01157	0.07145	-0.07111	0.02280	0.09390
α_5	-0.68660	-0.25608	0.43052	-0.72784	-0.21484	0.51300	-0.80843	-0.13424	0.67419	
β_1	0.07839	0.18443	0.10604	0.06823	0.19459	0.12635	0.04838	0.21444	0.16606	
MPVa	γ_0	6.09106	6.96103	0.86997	6.00773	7.04436	1.03663	5.84486	7.20723	1.36236
	γ_3	-0.27138	2.12688	2.39826	-0.50110	2.35660	2.85771	-0.95008	2.80558	3.75566
	γ_4	-3.69175	-1.07275	2.61901	-3.94262	-0.82188	3.12074	-4.43292	-0.33158	4.10134
	α_0	-1.56329	-0.96556	0.59773	-1.62054	-0.90831	0.71224	-1.73244	-0.79640	0.93604
	α_1	0.05167	0.06164	0.00997	0.05071	0.06259	0.01188	0.04885	0.06446	0.01561
	α_2	-0.04336	-0.00984	0.03353	-0.04658	-0.00662	0.03995	-0.05285	-0.00035	0.05251
	α_4	-0.04714	0.00050	0.04764	-0.05170	0.00507	0.05677	-0.06062	0.01399	0.07461
α_5	-0.53622	-0.41014	0.12608	-0.54830	-0.39807	0.15023	-0.57191	-0.37446	0.19744	
β_1	0.08314	0.17115	0.08801	0.07471	0.17958	0.10487	0.05823	0.19605	0.13782	
MPVa	γ_0	6.14313	7.00946	0.86633	6.06015	7.09244	1.03229	5.89797	7.25463	1.35666
	γ_3	0.08886	1.71797	1.62911	-0.06719	1.87401	1.94120	-0.37218	2.17900	2.55117
	γ_4	-3.44064	-1.32747	2.11317	-3.64305	-1.12506	2.51799	-4.03865	-0.72945	3.30920
	α_0	-1.50803	-1.02086	0.48716	-1.55469	-0.97420	0.58049	-1.64589	-0.88300	0.76289
	α_1	0.05120	0.06210	0.01089	0.05016	0.06314	0.01298	0.04812	0.06518	0.01706
	α_2	-0.04199	-0.01121	0.03078	-0.04494	-0.00826	0.03668	-0.05070	-0.00250	0.04820
	α_4	-0.04389	-0.00275	0.04114	-0.04783	0.00119	0.04902	-0.05553	0.00889	0.06442
α_5	-0.50953	-0.43684	0.07269	-0.51649	-0.42988	0.08662	-0.53010	-0.41627	0.11383	
β_1	0.08703	0.16728	0.08025	0.07935	0.17497	0.09562	0.06432	0.18999	0.12567	
MPVa	γ_0	6.16935	6.98325	0.81390	6.09139	7.06121	0.96982	5.93902	7.21358	1.27456
	γ_3	0.03211	1.77471	1.74261	-0.13481	1.94163	2.07644	-0.46104	2.26786	2.72891
	γ_4	-3.38269	-1.38542	1.99726	-3.57400	-1.19411	2.37989	-3.94790	-0.82020	3.12770

A Tabela 6.27 apresenta os limites inferiores (LI), limites superiores (LS) e comprimento (Comp) dos intervalos de confiança considerando os níveis 90%, 95% e 99% de confiança, para os métodos *naive*, calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa), considerando o modelo 3.

É possível verificar que o método da máxima pseudo-verossimilhança aproximada apresenta os menores comprimentos para os intervalos de confiança, ou seja, este método apresenta os intervalos mais precisos.

7 CONSIDERAÇÕES FINAIS

Nesta tese propusemos alguns métodos de estimação e análise de resíduos para o modelo de regressão beta não linear com erros de medida nas covariáveis. A seguir apresentaremos as principais conclusões de cada capítulo.

- **Capítulo 2:** Neste capítulo apresentamos o modelo de regressão beta não linear com erros de medidas nas covariáveis. Os métodos da máxima verossimilhança aproximada (MVa), pseudo-verossimilhança aproximada (MPVa) e calibração da regressão (RC) foram utilizados na estimação dos parâmetros. A função escore dos parâmetros e o método utilizado para obter o chute inicial foram apresentados. A avaliação numérica foi realizada através de um estudo de Monte Carlo com 10000 réplicas cujo objetivo foi verificar o desempenho dos estimadores estudados através da média, do viés e da raiz do erro quadrático médio. Podemos concluir que os métodos da máxima verossimilhança aproximada e pseudo-verossimilhança aproximada apresentaram os melhores resultados. Além disso, o método *naive* apresentou o pior desempenho, como esperado, uma vez que neste método ajustamos um modelo de regressão beta não linear desconsiderando o erro de mensuração. Notamos ainda que o desempenho dos estimadores ligados a não linearidade e dispersão variável são afetados negativamente quando o coeficiente de confiabilidade diminui.
- **Capítulo 3:** Definimos as distribuições assintóticas dos estimadores obtidos pelos métodos de máxima verossimilhança aproximada, pseudo-verossimilhança aproximada e calibração da regressão e apresentamos os respectivos intervalos de confiança. Novamente, pode-se concluir que os métodos da máxima verossimilhança aproximada e pseudo-verossimilhança aproximada apresentam melhores desempenhos quando comparados com os métodos *naive* e calibração da regressão. Além disso, vale salientar que quanto maior o erro de medida é necessário um número de medidas maior para que os resultados alcancem as taxas de cobertura fixadas, principalmente para os parâmetros associados à dispersão do modelo. Vale salientar que o método da pseudo-verossimilhança aproximada apresentam os melhores resultados, porém o comprimento do intervalo é muito superior ao obtido pelo método da máxima verossimilhança aproximada. Afirmamos, através da estimação intervalar, que para o coeficiente de confiabilidade pequeno os métodos de estimação apresentam dificuldades para estimar parâmetros associados à não linearidade.

- **Capítulo 4:** Aqui desenvolvemos o resíduo ponderado, ponderado padronizado e combinado, propostos por Espinheira *et al.* (2008) e Espinheira *et al.* (2017), para os modelos de regressão beta não linear com erros de medida. Avaliamos numericamente o desempenho dos resíduos propostos na presença de erros de medida através de um estudo de simulação de Monte Carlo com 10000 réplicas. Verificou-se que as distribuições dos resíduos não podem ser aproximadas pela distribuição normal padrão, portanto não se deve considerar os intervalos $[-2.0, 2.0]$ para identificar pontos aberrantes. Vale salientar também os problemas de convergência associados aos métodos da máxima pseudo-verossimilhança aproximada e calibração da regressão quando se tem um coeficiente de confiabilidade e tamanho amostral pequenos.
- **Capítulo 5:** Propomos as estatísticas de predição PRESS e P^2 para o modelo de regressão beta não linear com erros de medida e avaliamos os seus desempenhos considerando os métodos estudados através de um estudo de simulação de Monte Carlo com 10000 réplicas. Concluimos que as estatísticas P^2 e $P_{\alpha\gamma}^2$ e suas versões corrigidas indicam que à medida que o coeficiente de confiabilidade diminui a qualidade da predição do modelo é substancialmente afetada negativamente mesmo quando o modelo está bem especificado. Vale salientar também que a medida R_{LR}^2 e sua versão corrigida não detecta nenhum tipo de problema. Além disso, no modelo com dispersão constante as duas medidas de qualidade de ajuste pseudo- R^2 e R_{LR}^2 apresentam comportamento equivalentes.
- **Capítulo 6:** Neste capítulo apresentamos três aplicações a dados reais. A primeira aplicação estava relacionada a avaliação do sobrepeso em crianças, em que o peso foi mensurado duas vezes. A segunda avaliou dados de craqueamento catalítico fluido e a terceira verificou o sobrepeso em adolescentes. Na primeira aplicação ajustamos apenas modelos lineares com erros de medidas e verificamos que os métodos da máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança aproximada (MPVa) apresentaram melhores desempenhos. Já o método da calibração da regressão (RC) mostrou muita dificuldade tanto no ajuste quanto no poder preditivo. Na aplicação 2 verificamos a necessidade de um modelo não linear, através dos gráficos dos resíduos. Considerando duas funções de ligação, verificamos que o melhor modelo não linear para ajustar o conjuntos de dados foi considerando a função de ligação complemento log-log. Além disso, verificamos que ao retirar os pontos aberrantes não houve mudanças inferenciais. Em relação aos intervalos de confiança, concluimos que o método da máxima pseudo-verossimilhança aproximada (MPVa) apresenta os menores, ou seja, o método apresenta os intervalos mais precisos. A última aplicação, também ajustamos um modelo de regressão beta não linear com

erros de medida. Verificamos que mesmo não sendo significativo para o modelo, a variável triglicérides deve estar no modelo. Os resíduos e envelopes simulados evidenciam que o modelo 3 apresenta resultados satisfatórios, bem como o poder de predição é alto em todos os métodos analisados. Novamente o método da MPVa apresentou os intervalos de confiança mais precisos.

7.1 Trabalhos futuros

As principais ideias de trabalhos futuros são:

- realizar testes de hipóteses para os parâmetros associados ao modelo de regressão beta não linear com erros de medida;
- utilizar o enfoque funcional para contornar a presença de erros de medida;
- utilizar o método de estimação EM-Monte Carlo como alternativa de estimação dos parâmetros;
- considerar uma função não linear na relação entre x_t e w_t ;
- considerar uma relação multiplicativa entre x_t e e_t , ou seja, $w_t = x_t \times e_t$.

REFERÊNCIAS

- ABRAMOWITZ, M.; STEGUN, I. Handbook of Mathematical Functions. *Dover*, New York, 1964.
- AKAIKE, H. Information theory and an extension of the maximum likelihood principle. *2nd, International Symposium on Information Theory*, pp. 267-281, 1973.
- ALLEN, D. M. *The prediction sum of squares as a criterion for selecting predictor variables*. Lexington: University of Kentucky, 1971.
- BAYER, F. M.; Cribari-Neto, F. Model selection criteria in beta regression with varying dispersion. *Communications in Statistics - Simulation and Computation*, **46**, n.1, pp. 729-746, 2017.
- BOZDONGAN, H. Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*. **52**, n.3, pp. 345-370, 1987.
- BOUNACCORSI, J. P. *Measurement Error: Models, Methods and Applications*. London: Chapman and Hall, 2010.
- BOUNACCORSI, J. P.; TOSTESON, T. D. Correcting for nonlinear measurement errors in the dependent variable in the general linear model. *Communications in Statistics - Theory and Methods*, **22**, pp. 2687-2702, 1993.
- CARRASCO, J. M. F. *Modelos de regressão beta com erro nas variáveis*. Tese apresentada ao Instituto de Matemática e Estatística, IME-USP, 2012.
- CARRASCO, J. M. F.; FERRARI, S. L. P.; ARELLANO-VALLE, R. B. Errors-in-variables beta regression models, *Journal of Applied Statistics*, **41**, n.7, pp. 1530-1547, 2014.
- CARROLL, R. J.; RUPPERT, D. Transformation and Weighting in Regression. *Chapman and Hall*, London, 1988.
- CARROLL, R.; RUPPERT, D.; STEFANSKI, L.; e CRAINICEANU, C. M. Measurement Error in Nonlinear Models: A Modern Perspective. New York: *Chapman and Hall*, 2006.
- CARROLL, R. J.; SPIEGELMAN, C. H. Diagnostics for nonlinearity and heteroscedasticity in errors-in-variables regression. *Technometrics*, **34**, pp. 186-196, 1992.

- CHENG, C; VAN NESS, J. Statistical Regression with Measurement Error. *Oxford University Press*, London, 1999.
- CHESHER, A. The effect of measurement error. *Biometrika*, **78**, pp. 451-162, 1991.
- CORDEIRO, G. M. *Modelos Lineares Generalizados*. VII Simpósio Nacional de Probabilidade e Estatística, Campinas, SP. 1986.
- COX, D. R. Partial likelihood. *Biometrika*, **62**, pp. 269-276, 1975.
- COX, D.; HINKLEY, D. *Theoretical Statistics*. London: Chapman & Hall. 1974.
- COX, D.; SNELL, E. A general definition of residuals. *Journal of the Royal Statistical Society B*, **30**, pp. 248-275, 1968.
- COOK, R. D.; WEISBERG, S. Residuals and Influence in Regression. *Chapman and Hall*, 1982.
- CRAMÉR, H. Mathematical Methods of Statistics. *Princeton University Press*, New York, 1999.
- CRIBARI-NETO, F.; ZARKOS, S. G. R: yet another econometric programming environment. *Journal of Applied Econometrics*, **14**, pp. 319-329, 1999.
- DALGAARD, P. Introductory Statistics with R. *Springer*, New York, 2002.
- DOORNIK, J. A. *Object-Oriented Matrix Programming using Ox*. 5th ed. London: Timberlake Consultants Press, 2006.
- DOORNIK, J. A.; OOMS, M. *An introduction to Ox*. Technical report, <http://www.doornik.com/ox/OxIntro.pdf>, 2007.
- DOORNIK, J. A. *Ox: an Object-oriented matrix Programming Language*. 6th ed. London: Timberlake Consultants Ltd, 2009.
- DOORNIK, J. A. Ox 6.2: *Object-Oriented Matrix Programming Language*. Timberlake Consultants and Oxford, London. Available at <http://www.doornik.com>, 2011.
- DRAPER, N.R.; SMITH, H. (1981) Applied Regression Analysis. *Wiley*, New York, 1981.
- ESPINHEIRA, P. L.; FERRARI, S. L. P.; CRIBARI-NETO, F. On beta regression residuals. *Journal of Applied Statistics*, **35**, n.4, pp. 407-419, 2008.
- ESPINHEIRA, P. L.; SANTOS, E. G.; CRIBARI-NETO, F. On nonlinear beta regression residuals. *Biometrical Journal*. **59** pp.445-461, 2017.

- FAHRMEIR, L.; TUTZ, G. Multivariate Statistical Modelling Based on Generalized Linear Models, 2nd ed. *Springer*, New York, 2001.
- FERRARI, S.; CRIBARI-NETO, F. Beta regression for modelling rates and proportion. *Journal of Applied Statistics*, **31**, pp. 799-815, 2004.
- FERRARI, S. L. P., ESPINHEIRA, P. L.; CRIBARI-NETO, F. Diagnostic tools in beta regression with varying dispersion. *Statistica Neerlandica*, **65**, pp. 337-351, 2011.
- FREEDMAN, L.; MIDTHUNE, D.; CARROLL, R.; KIPNIS, V. A comparison of regression calibration, moment reconstruction and imputation for adjusting for covariate measurement error in regression. *Statistics in Medicine*, **27**, pp. 5195-5216, 2008.
- FULLER, W. A. Measurement Error Models. *John Wiley*, New York, 1987.
- GARSIDE, M. J. The best subset in multiple regression analysis. *Journal of the Royal Statistical Society C*, **14**, pp. 196-200, 1965.
- GODAMBE, V. P. Estimating Functions. *Clarendon Press*, New York, 1991.
- GOLUB, G. H.; WELSCH, J. Calculation of Gauss quadrature rules. *Mathematics of Computation*, **23**, pp. 221-221, 1969.
- GONG, G.; SAMANIEGO, F. Pseudo maximum likelihood estimation: theory and applications. *The Annals of Statistics*, **9**, pp. 861-869, 1981.
- GUOLO, A. Pseudo-likelihood inference for regression models with misclassified and mismeasured variables. *Statistica Sinica*, **21**, pp. 1639-1663, 2011.
- HUWANG, L.; HUANG, Y. H. S. On errors-in-variables in polynomial regression - Berkson case. *Statistics Sinica*, **10**, pp. 923-936, 2000.
- KNUTH, D. *The TEXbook*. Addison-Wesley, New York, 1986.
- LAMPORT, L. *A Document Preparation System*. 2nd ed, Addison-Wesley, Massachusetts, 1994.
- LITTELL, R. C.; MILLIKEN, G. A.; STROUP, W. W.; WOLFINGER, R. D. *SAS System for Mixed Models*. Cary: Statistical Analysis System Institute, 2002.
- LIU, Q.; PIERCE, D. A note on Gauss-Hermite quadrature. *Biometrika*, **81**, pp. 624-629, 1994.

- LIU, H.; WEISSB, R. E.; JENNRICH, R. I.; WENGER, N. S. PRESS model selection in repeated measures data. *Computational Statistics & Data Analysis*, **30**, pp. 169-184, 1999.
- MALLOWS, C. L. Some comments on CP. *Technometrics*, **15**, pp. 661-675, 1973.
- MANCO, O. C. U. *Modelos de regressão beta com efeitos aleatórios normais e não normais para dados longitudinais*. Tese de doutorado. IME-USP, 2013.
- MURRELL, P. R. Graphics. *Chapman and Hall*, London, 2006.
- PREGIBON, D. Logistic regression diagnostics. *Ann. Statist.* **9**, pp. 705-724, 1981.
- RABE-HESKETH, S.; SKRONDAL, A.; PICKLES, A. Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, **128**, 301-323, 2005.
- ROCHA, A. V.; SIMAS, A. B. Influence diagnostics in a general class of beta regression models. *Test*, **20**, n.1, pp. 95-119, 2011.
- ROSNER, W. B.; SPIEGELMAN, D. Correction of logistic regression relative risk estimate and confidence intervals for systematic within-person measurement error. *Statistics in Medicine*, **8**, pp. 1075-1093, 1989.
- SANTOS, E. G. *Um novo resíduo para classes do modelo de regressão beta - linear e não linear*. Dissertação de Mestrado. Programa de Pós-graduação em Estatística - UFPE, 2015.
- SCHENNACH, S. M. *Measurement error in nonlinear models - a review*. The Institute for Fiscal Studies. Department of Economics, UCL, 2012.
- SCHWARZ, G. Estimating the dimension of a model. *The Annals of Statistics*, **6**, pp. 461-464, 1978.
- SILVA, L. C. M. *Coefficientes de predição para os modelos de regressão beta e simplex*. Dissertação de Mestrado. Programa de Pós-graduação em Estatística - UFPE, 2015.
- SIMAS, A.; BARRETO-SOUZA, W.; ROCHA, A. V. . Improved estimators for a general class of beta regression models. *Computational Statistics and Data Analysis*, **54**, pp. 348-366, 2010.
- STEFANSKI, L. The effects of measurement error on parameter estimation. *Biometrika*, **72**, pp. 583-592, 1985.

STEFANSKI, L.; CARROLL, R. Covariate measurement error in logistic regression. *The Annals of Statistics*, **13**, pp. 1335-1351, 1985.

STONE, M. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society B*, **36**, pp. 111-147, 1974.

THURSTON, S.; WILLIAMS, P.; HAUSER, R.; HU, H.; HERNANDEZ-AVILA, M.; SPIEGELMAN, D. A comparison of regression calibration approaches for designs with internal validation data. *Journal of Statistical Planning and Inference*, **131**, pp. 175-190, 2005.

VENABLES, W. N.; RIPLEY, B. D. Modern Applied Statistics with S, 4th ed. *Springer*, New York, 2002.

VENABLES, W. N.; Smith, D. M. *The R Development Core Team. An Introduction to R*. 2nd ed. Network Theory Limited, 2009.

VILLEGAS, C.; PAULA, G. A.; LEIVA, V. Birnbaum-Saunders mixed models for censored reliability data analysis. *IEEE Transactions on Reliability*, **60**, pp. 748-758, 2011.

WANG, L. Estimation of Berkson-type measurement error models. *Statistics Sinica*, **13**, pp. 1201-1210, 2003.

WANG, L. Estimation of nonlinear models with Berkson measurement errors. *The Annals of Statistics*, **32**, pp. 2559-2597, 2004.

WOLFINGER, R. D. Covariance structure selection in general mixed models. *Communications in Statistics*, **22**, pp. 1079-1106, 1993.

APÊNDICE A - FUNÇÃO ESCORE E MATRIZ DE INFORMAÇÃO OBSERVADA

Apêndice A.1 - Função escore

Apêndice A.1.1 - Máxima verossimilhança aproximada

O logaritmo da função de verossimilhança aproximada, admitindo que uma única covariável é medida com erro, é dado por

$$\ell_a(\Psi) = \sum_{t=1}^n \ell_{1t}(\xi, \sigma_e^2) + \sum_{t=1}^n \ell_{2t}(\theta, \xi, \sigma_e^2), \quad (1)$$

sendo

$$\begin{aligned} \ell_{1t}(\xi, \sigma_e^2) &= \log f(w_t; \xi, \sigma_e^2) = -\frac{1}{2} \log[2\pi(\sigma_x^2 + \sigma_e^2)] - \frac{1}{2(\sigma_x^2 + \sigma_e^2)}(w_t - \mu_x)^2, \\ \ell_{2t}(\theta, \xi, \sigma_e^2) &\simeq \log \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp[\ell_{tq}(\mu_{tq}, \phi_{tq})] \end{aligned}$$

e

$$\begin{aligned} \ell_{tq}(\mu_{tq}, \phi_{tq}) &= \log \Gamma(\phi_{tq}) - \log \Gamma(\mu_{tq}\phi_{tq}) - \log \Gamma[(1 - \mu_{tq})\phi_{tq}] \\ &+ (\mu_{tq}\phi_{tq} - 1) \log y_t + [(1 - \mu_{tq})\phi_{tq} - 1] \log(1 - y_t), \end{aligned} \quad (2)$$

com

$$\begin{aligned} g(\mu_{tq}) &= \eta_{1t} = f_1(\mathbf{z}_t^\top, x_t^*; \boldsymbol{\alpha}, \beta), \\ h(\phi_{tq}) &= \eta_{2t} = f_2(\mathbf{v}_t^\top, x_t^*; \boldsymbol{\gamma}, \lambda), \\ x_t^* &= \mu_{x_t|w_t} + \sqrt{2\sigma_{x_t|w_t}^2} s_q, \\ \mu_{x_t|w_t} &= \mu_x + k_x(w_t - \mu_x), \\ \sigma_{x_t|w_t}^2 &= \sigma_e^2 k_x, \\ k_x &= \frac{\sigma_x^2}{(\sigma_x^2 + \sigma_e^2)}. \end{aligned}$$

Diferenciando o logaritmo da função de verossimilhança aproximada (1) em relação a cada um dos parâmetros de interesse $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \beta, \boldsymbol{\gamma}^\top, \lambda)^\top$ é obtida a função escore aproximada dada por $U_a(\boldsymbol{\theta}^\top) = (U_{\boldsymbol{\alpha}}^\top, U_\beta, U_{\boldsymbol{\gamma}}^\top, U_\lambda)^\top$ que se encontram detalhadas abaixo.

Para $i = 1, 2, \dots, p$, a função escore aproximada para cada α_i é dada por

$$U_{\alpha_i} = \frac{\partial \ell_a(\Psi)}{\partial \alpha_i} = \sum_{t=1}^n b_t a_{t\alpha_i},$$

sendo

$$b_t = \left(\sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \right)^{-1}, \quad (3)$$

$$a_{t\alpha_i} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \dot{\ell}_{tq(\alpha_i)}(\mu_{tq}, \phi_{tq}), \quad (4)$$

$$\dot{\ell}_{tq(\alpha_i)}(\mu_{tq}, \phi_{tq}) = \phi_{tq} [y_t^* - \mu_{tq}^*] \frac{1}{g'(\mu_{tq})} \frac{\partial \eta_{1t}}{\partial \alpha_i}, \quad (5)$$

$$y_t^* = \log \left(\frac{y_t}{1 - y_t} \right) \quad \text{e} \quad \mu_{tq}^* = \psi(\mu_{tq} \phi_{tq}) - \psi[(1 - \mu_{tq}) \phi_{tq}],$$

em que $\ell_{tq}(\mu_{tq}, \phi_{tq})$ é definido em (2) e $\psi(\cdot)$ é a função digama.

A função escore para o parâmetro correspondente à covariável medida com erro denotado por β , é dada por

$$U_\beta = \frac{\partial \ell_a(\Psi)}{\partial \beta} = \sum_{t=1}^n b_t a_{t\beta},$$

em que

$$a_{t\beta} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \dot{\ell}_{tq(\beta)}(\mu_{tq}, \phi_{tq}) \quad \text{e} \quad (6)$$

$$\dot{\ell}_{tq(\beta)}(\mu_{tq}, \phi_{tq}) = \phi_{tq} [y_t^* - \mu_{tq}^*] \frac{1}{g'(\mu_{tq})} \frac{\partial \eta_{1t}}{\partial \beta}. \quad (7)$$

Em forma matricial, as funções escores para os parâmetros α e β são dados, respectivamente, por

$$\begin{aligned} U_\alpha(\theta) &= \mathbf{F}_1^\top \left[(\Phi \circ \mathbf{G})(\mathbf{P} \circ \mathbf{V})^\top \circ \mathbf{B} \right] \mathbf{L} \\ U_\beta(\theta) &= \mathbf{F}_2^\top \left[(\Phi \circ \mathbf{G})(\mathbf{P} \circ \mathbf{V})^\top \circ \mathbf{B} \right] \mathbf{L} \end{aligned}$$

em que \circ denota o produto de Hadamard (elemento a elemento) e \mathbf{L} é um vetor $n \times 1$ de uns. $\mathbf{F}_{1t} = \partial \eta_{1t} / \partial \alpha$ é uma matriz de derivadas de ordem $(n \times p)$ e $\mathbf{F}_{2t} = \partial \eta_{1t} / \partial \beta$ é um vetor de derivadas r -dimensional. As matrizes de ordem $(n \times Q)$ Φ , \mathbf{G} , \mathbf{P} , $\mathbf{V} = (\mathbf{y}^* - \mu_{tq})$ e a matriz $n \times n$ \mathbf{B} são definidos a seguir.

A matriz de ordem $(n \times Q)$ Φ definida em $U_\alpha(\theta)$ e $U_\beta(\theta)$ é dada por

$$\Phi = \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1Q} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2Q} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{n1} & \phi_{n2} & \cdots & \phi_{nQ} \end{bmatrix}.$$

em que ϕ denota a dispersão do modelo.

As matrizes de ordem $(n \times Q)$ \mathbf{G} , \mathbf{P} e \mathbf{V} definidas em $U_{\alpha}(\boldsymbol{\theta})$ e $U_{\beta}(\boldsymbol{\theta})$ são definidas por

$$\mathbf{G} = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1Q} \\ g_{21} & g_{22} & \cdots & g_{2Q} \\ \vdots & \vdots & \ddots & \vdots \\ g_{n1} & g_{n2} & \cdots & g_{nQ} \end{bmatrix},$$

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1Q} \\ p_{21} & p_{22} & \cdots & p_{2Q} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nQ} \end{bmatrix}$$

e

$$\mathbf{V} = \begin{bmatrix} y_1^* - \mu_{11}^* & y_1^* - \mu_{12}^* & \cdots & y_1^* - \mu_{1Q}^* \\ y_2^* - \mu_{21}^* & y_2^* - \mu_{22}^* & \cdots & y_2^* - \mu_{2Q}^* \\ \vdots & \vdots & \ddots & \vdots \\ y_n^* - \mu_{n1}^* & y_n^* - \mu_{n2}^* & \cdots & y_n^* - \mu_{nQ}^* \end{bmatrix},$$

em que $g_{tq} = 1/g'(\mu_{tq})$, $p_{tq} = \frac{\nu_q}{\sqrt{\pi}} \exp[\ell(\mu_{tq}, \phi_{tq})]$, $y_t^* = \log\left(\frac{y_t}{1-y_t}\right)$ e $\mu_{tq}^* = \psi(\mu_{tq}\phi_{tq}) - \psi[(1-\mu_{tq})\phi_{tq}]$, $t = 1, \dots, n$ e $q = 1, \dots, Q$.

O vetor n -dimensional \mathbf{B} é definido por

$$\mathbf{B} = \begin{bmatrix} b_1 & 0 & \cdots & 0 \\ 0 & b_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_n \end{bmatrix},$$

em que b_t , para $t = 1, \dots, n$ é definido em 3. Assim, $U_{\alpha}(\boldsymbol{\theta})$ e $U_{\beta}(\boldsymbol{\theta})$ possuem dimensões $(n \times 1)$ e (1×1) , respectivamente.

Agora derivando $\ell_a(\boldsymbol{\Psi})$ com relação aos parâmetros de precisão, γ_j , para todo $j = 1, \dots, \check{q}$ e λ temos que

$$U_{\gamma_j} = \frac{\partial \ell_a(\boldsymbol{\Psi})}{\partial \gamma_j} = \sum_{t=1}^n b_t a_{t\gamma_j} \quad \text{e} \quad U_{\lambda} = \frac{\partial \ell_a(\boldsymbol{\Psi})}{\partial \lambda} = \sum_{t=1}^n b_t a_{t\lambda},$$

em que

$$a_{t\gamma_j} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \dot{\ell}_{tq(\gamma_j)}(\mu_{tq}, \phi_{tq}), \quad (8)$$

$$a_{t\lambda} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \dot{\ell}_{tq(\lambda)}(\mu_{tq}, \phi_{tq}), \quad (9)$$

$$\dot{\ell}_{tq(\gamma_j)}(\mu_{tq}, \phi_{tq}) = a_{tq} \frac{1}{h'(\phi_{tq})} \frac{\partial \eta_{2t}}{\partial \gamma_i}, \quad (10)$$

$$\dot{\ell}_{tq(\lambda)}(\mu_{tq}, \phi_{tq}) = a_{tq} \frac{1}{h'(\phi_{tq})} \frac{\partial \eta_{2t}}{\partial \lambda} \quad e \quad (11)$$

$$a_{tq} = \mu_t(y_t^* - \mu_{tq}^*) + \log(1 - y_t) + \psi((1 - \mu_{tq})\phi_{tq}) + \psi(\phi_t). \quad (12)$$

Em forma matricial, as funções escores para os parâmetros $\boldsymbol{\gamma}$ e λ são dados, respectivamente, por

$$\begin{aligned} U_{\boldsymbol{\gamma}}(\boldsymbol{\theta}) &= \mathbf{F}_3^\top [\mathbf{H}\mathbf{a}^{*\top} \circ \mathbf{B}] L, \\ U_{\lambda}(\boldsymbol{\theta}) &= \mathbf{F}_4^\top [\mathbf{H}\mathbf{a}^{*\top} \circ \mathbf{B}] L, \end{aligned}$$

em que $\mathbf{F}_{3t} = \partial \eta_{2t} / \partial \boldsymbol{\gamma}$ é uma matriz de derivadas de ordem $(n \times \check{q})$ e $\mathbf{F}_{4t} = \partial \eta_{2t} / \partial \lambda$ é um vetor de derivadas s -dimensional. As matrizes de ordem $(n \times Q)$ \mathbf{H} e \mathbf{a}^* definidas em $U_{\boldsymbol{\gamma}}(\boldsymbol{\theta})$ e $U_{\lambda}(\boldsymbol{\theta})$ são dadas, respectivamente, por

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1Q} \\ h_{21} & h_{22} & \dots & h_{2Q} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \dots & h_{nQ} \end{bmatrix}$$

e

$$\mathbf{a}^* = \begin{bmatrix} a^{*11} & a^{*12} & \dots & a^{*1q} \\ a^{*21} & a^{*22} & \dots & a^{*2q} \\ \vdots & \vdots & \ddots & \vdots \\ a^{*n1} & a^{*n2} & \dots & a^{*nq} \end{bmatrix},$$

em que $h_{tq} = 1/h'(\phi_{tq})$ e $a^{*tq} = p_{tq} a_{tq}$, com $a_{tq} = \mu_{tq}(y_t^* - \mu_{tq}^*) + \log(1 - y_t) - \psi[(1 - \mu_{tq})\phi_{tq}] + \psi(\phi_{tq})$.

Também é possível obter a função escore para os parâmetros de perturbação $\boldsymbol{\xi} = (\mu_x, \sigma_x^2)^\top$ diferenciando o logaritmo da função de verossimilhança aproximada (1) em relação aos mesmos. Assim, a função escore aproximada para os parâmetros de perturbação é $U_a(\boldsymbol{\xi}) = (U_{\mu_x}, U_{\sigma_x^2})^\top$, em que

$$U_{\mu_x} = \frac{1}{(\sigma_x^2 + \sigma_e^2)} \sum_{t=1}^n (w_t - \mu_x) + \sum_{t=1}^n b_t a_{t\mu_x} \quad e \quad (13)$$

$$U_{\sigma_x^2} = -\frac{1}{2(\sigma_x^2 + \sigma_e^2)} \sum_{t=1}^n \left[1 - \frac{(w_t - \mu_x)^2}{(\sigma_x^2 - \sigma_e^2)} \right] + \sum_{t=1}^n b_t a_{t\sigma_x^2} \quad (14)$$

sendo

$$a_{t\mu_x} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \dot{\ell}_{tq(\mu_x)}(\mu_{tq}, \phi_{tq}), \quad (15)$$

$$a_{t\sigma_x^2} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \dot{\ell}_{tq(\sigma_x^2)}(\mu_{tq}, \phi_{tq}), \quad (16)$$

$$\dot{\ell}_{tq(\mu_x)}(\mu_{tq}, \phi_{tq}) = \left[\frac{\phi_{tq}(y_t^* - \mu_{tq}^*)}{g'(\mu_{tq})} \frac{\partial \eta_{1t}}{\partial \mu_x} + \frac{a_{tq}}{h'(\phi_{tq})} \frac{\partial \eta_{2t}}{\partial \mu_x} \right] \text{ e} \quad (17)$$

$$\dot{\ell}_{tq(\sigma_x^2)}(\mu_{tq}, \phi_{tq}) = \left[\frac{\phi_{tq}(y_t^* - \mu_{tq}^*)}{g'(\mu_{tq})} \frac{\partial \eta_{1t}}{\partial \sigma_x^2} + \frac{a_{tq}}{h'(\phi_{tq})} \frac{\partial \eta_{2t}}{\partial \sigma_x^2} \right]. \quad (18)$$

As equações (13) e (14) podem ser escritas matricialmente da seguinte forma: $U_{\mu_x} = \mathbf{W}_{\mu_x} + \mathbf{B}^\top \mathbf{A}_{\mu_x}$ e $U_{\sigma_x^2} = \mathbf{W}_{\sigma_x^2} + \mathbf{B}^\top \mathbf{A}_{\sigma_x^2}$, em que \mathbf{W}_{μ_x} e $\mathbf{W}_{\sigma_x^2}$ são vetores que dependem da covariável medida com erro w_t cujos t -ésimos elementos são dados por $(w_t - \mu_x)/(\sigma_x^2 + \sigma_e^2)$ e $-[2(\sigma_x^2 + \sigma_e^2)]^{-1}[1 - (w_t - \mu_x)^2/(\sigma_x^2 + \sigma_e^2)]$, respectivamente. Ademais, $\mathbf{A}_{\mu_x} = (a_{1\mu_x}, \dots, a_{n\mu_x})^\top$ e $\mathbf{A}_{\sigma_x^2} = (a_{1\sigma_x^2}, \dots, a_{n\sigma_x^2})^\top$ são vetores, ambos com dimensão $n \times 1$.

Apêndice A.1.2 - Máxima pseudo-verossimilhança aproximada

Considere $(\boldsymbol{\xi}^\top, \sigma_e^2)^\top$, em que $\boldsymbol{\xi}^\top = (\mu_x, \sigma_x^2)$ é o vetor de parâmetros de perturbação e $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \beta, \boldsymbol{\gamma}^\top, \lambda)^\top$ é o vetor de parâmetros de interesse. Como não é possível eliminar o vetor $\boldsymbol{\xi}$ em $\ell(\boldsymbol{\Psi})$ através de integração, utiliza-se a abordagem de pseudo-verossimilhança para estimar os parâmetros (Gong e Samaniego, 1981; Guolo, 2011). Desta forma, o logaritmo da função verossimilhança (2.15) é maximizado em duas etapas. Na primeira etapa, estima-se o vetor de parâmetros de perturbação $\boldsymbol{\xi}$ a partir da maximização do logaritmo da função de verossimilhança reduzida, como sugerem Buonaccorsi e Tosteson (1993) e Guolo (2011). Ou seja,

$$\ell_r(\boldsymbol{\xi}, w) = \sum_{t=1}^n \log f(w; \boldsymbol{\xi}, \sigma_e^2) \quad (19)$$

denota o logaritmo da função de verossimilhança reduzida, em que a variância do erro de medição, dada por σ_e^2 , é assumida conhecida ou é estimada quando existem réplicas para w_t . Na segunda, a estimativa de $\boldsymbol{\xi}$, denotada por $\hat{\boldsymbol{\xi}}$, obtida da maximização de (19) é substituída no logaritmo da função de verossimilhança original (1) o que resulta no logaritmo da pseudo-verossimilhança

$$\ell_p(\boldsymbol{\theta}, \hat{\boldsymbol{\xi}}, \hat{\sigma}_e^2) = \sum_{t=1}^n \ell_{1t}(\hat{\boldsymbol{\xi}}, \hat{\sigma}_e^2) + \sum_{t=1}^n \ell_{2t}(\boldsymbol{\theta}, \hat{\boldsymbol{\xi}}, \hat{\sigma}_e^2), \quad (20)$$

sendo $\hat{\sigma}_e^2$ representa o valor conhecido ou estimado de σ_e^2 .

Desta forma, o logaritmo da pseudo-verossimilhança aproximada para o modelo de regressão beta não linear com erros de medida unidimensional é dado por (20), sendo

$$\ell_{1t}(\hat{\boldsymbol{\xi}}, \hat{\sigma}_e^2) = -\frac{1}{2} \log[2\pi(\hat{\sigma}_x^2 + \hat{\sigma}_e^2)] - \frac{1}{2(\hat{\sigma}_x^2 + \hat{\sigma}_e^2)} (w_t - \hat{\mu}_x)^2$$

e

$$\ell_{2t}(\boldsymbol{\theta}, \hat{\boldsymbol{\xi}}, \hat{\sigma}_e^2) \simeq \sum_{t=1}^n \log \left(\sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp[\ell_{tq}(\boldsymbol{\theta}; \hat{\boldsymbol{\xi}}, \hat{\sigma}_e^2)] \right),$$

em que

$$\begin{aligned} \ell_{tq}(\boldsymbol{\theta}; \hat{\boldsymbol{\xi}}, \hat{\sigma}_e^2) &= \log \Gamma(\phi_{tq}) - \log \Gamma(\mu_{tq}\phi_{tq}) - \log \Gamma[(1 - \mu_{tq})\phi_{tq}] + (\mu_{tq}\phi_{tq} - 1) \log y_t + \\ &\quad [(1 - \mu_{tq})\phi_{tq} - 1] \log(1 - y_t), \\ g(\mu_{tq}) &= f_1(\mathbf{z}_t^\top, \hat{x}_t^*; \boldsymbol{\alpha}, \beta), \\ h(\phi_{tq}) &= f_2(\mathbf{v}_t^\top, \hat{x}_t^*; \boldsymbol{\gamma}, \lambda), \\ x_t^* &= \hat{\mu}_{x_t|w_t} + \sqrt{2\hat{\sigma}_{x_t|w_t}^2} s_q, \\ \hat{\mu}_{x_t|w_t} &= \hat{\mu}_x + \hat{k}_x(w_t - \hat{\mu}_x), \\ \hat{\sigma}_{x_t|w_t}^2 &= \hat{\sigma}_e^2 \hat{k}_x, \\ \hat{k}_x &= \frac{\hat{\sigma}_x^2}{(\hat{\sigma}_x^2 + \hat{\sigma}_e^2)}. \end{aligned}$$

Diferenciando o logaritmo da função de máxima pseudo-verossimilhança aproximada (20) em relação a cada um dos parâmetros de interesse $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \beta, \boldsymbol{\gamma}^\top, \lambda)^\top$ é obtida a função escore aproximada dada por $U_a(\boldsymbol{\theta}^\top) = (U_{\boldsymbol{\alpha}}^\top, U_{\beta}, U_{\boldsymbol{\gamma}}^\top, U_{\lambda})^\top$:

$$\begin{aligned} U_{p\boldsymbol{\alpha}}(\boldsymbol{\theta}) &= \mathbf{F}_1^\top \left[(\boldsymbol{\Phi} \circ \mathbf{G})(\mathbf{P} \circ \mathbf{V})^\top \circ \mathbf{B} \right] \mathbf{L}, \\ U_{p\beta}(\boldsymbol{\theta}) &= \mathbf{F}_2^\top \left[(\boldsymbol{\Phi} \circ \mathbf{G})(\mathbf{P} \circ \mathbf{V})^\top \circ \mathbf{B} \right] \mathbf{L}, \\ U_{p\boldsymbol{\gamma}}(\boldsymbol{\theta}) &= \mathbf{F}_3^\top \left[\mathbf{H}\mathbf{a}^{*\top} \circ \mathbf{B} \right] L, \\ U_{p\lambda}(\boldsymbol{\theta}) &= \mathbf{F}_4^\top \left[\mathbf{H}\mathbf{a}^{*\top} \circ \mathbf{B} \right] L. \end{aligned}$$

As matrizes são obtidas da mesma forma que as apresentadas para a máxima verossimilhança aproximada, considerando \hat{x}_t^* .

Apêndice A.1.3 - Calibração da regressão

O método de estimação por calibração da regressão vem sendo muito utilizado na última década por sua facilidade na estimação de parâmetros quando se evidencia erros de mensuração nas covariáveis. A principal ideia deste método é substituir a variável

não observável, x_t por uma estimativa da esperança condicional de x_t dado w_t , $\mathbb{E}(x_t|w_t)$, denotada por *função de calibração*, que depende de um vetor de parâmetros $\boldsymbol{\psi}$. A troca da variável não observada x_t por $r(w_t, \hat{\boldsymbol{\psi}}) = \mathbb{E}(x_t|w_t)$ estabelece um modelo modificado para os dados. Aqui, $\hat{\boldsymbol{\psi}}$ é uma estimativa de $\boldsymbol{\psi}$.

Considere a função densidade da distribuição beta proposto por Ferrari e Cribari-Neto (2004) dada por

$$f(y_t; \mu_t, \phi_t) = \frac{\Gamma(\phi_t)}{\Gamma(\mu_t \phi_t) \Gamma((1 - \mu_t) \phi_t)} y_t^{\mu_t \phi_t - 1} (1 - y_t)^{(1 - \mu_t) \phi_t - 1}, \quad (21)$$

em que $0 < y_t < 1$, $0 < \mu_t < 1$ e $\phi_t > 0$.

Sejam y_1, \dots, y_n variáveis independentes, em que cada y_t segue a distribuição (21). No modelo de regressão beta não linear a média e o parâmetro de precisão satisfazem as seguintes relações funcionais

$$g(\mu_t) = \eta_{1t} = f_1(\mathbf{z}_t^T; \boldsymbol{\alpha}) \quad \text{e} \quad h(\phi_t) = \eta_{2t} = f_2(\mathbf{v}_t^T; \boldsymbol{\gamma}),$$

em que $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^\top$ e $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_{\check{q}})^\top$ são vetores de parâmetros desconhecidos, com $p + \check{q} < n$, η_{1t} e η_{2t} são os preditores e $\mathbf{z}_t^\top = (z_{t1}, \dots, z_{tp_1})$ e $\mathbf{v}_t^\top = (v_{t1}, \dots, v_{t\check{q}_1})$ são observações em p_1 e \check{q}_1 covariáveis fixas e conhecidas, com $t = 1, \dots, n$, $p_1 \leq p$ e $\check{q}_1 \leq \check{q}$. Geralmente as covariáveis utilizadas para modelar a precisão é um subconjunto das que foram utilizadas na modelagem da média. As funções de ligação, $g(\cdot)$ e $h(\cdot)$, são estritamente monótonas e duas vezes diferenciáveis e $f(\cdot)$ é uma função contínua e diferenciável, tal que as matrizes de derivadas $\mathbf{F}_1 = \partial \eta_{1t} / \partial \boldsymbol{\alpha}$ e $\mathbf{F}_3 = \partial \eta_{2t} / \partial \boldsymbol{\gamma}$ têm postos p e \check{q} , respectivamente.

O modelo de regressão beta não linear com erros de medida baseado em (21) assume que a média e o parâmetro de precisão de y_t satisfazem as seguintes relações funcionais

$$\begin{aligned} g(\mu_t) &= \eta_{1t} = f_1(\mathbf{z}_t^\top, \mathbf{x}_t^\top; \boldsymbol{\alpha}, \boldsymbol{\beta}), \quad \text{e} \\ h(\phi_t) &= \eta_{2t} = f_2(\mathbf{v}_t^\top, \mathbf{m}_t^\top; \boldsymbol{\gamma}, \boldsymbol{\lambda}), \end{aligned}$$

em que $\boldsymbol{\beta} = (\beta_1, \dots, \beta_r)^\top \in \mathbb{R}^r$, e $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_s)^\top \in \mathbb{R}^s$ são parâmetros desconhecidos, $\mathbf{x}_t^\top = (x_{t1}, \dots, x_{tr_1})$ e $\mathbf{m}_t^\top = (m_{t1}, \dots, m_{ts_1})$ são vetores de covariáveis não observadas diretamente ou associadas a erros de medição, com $r_1 \leq r$ e $s_1 \leq s$. As funções $f_1(\cdot)$ e $f_2(\cdot)$ são funções contínuas e diferenciáveis, tais que as matrizes de derivadas $\mathbf{F}_2 = \partial \eta_{1t} / \partial \boldsymbol{\beta}$, e $\mathbf{F}_4 = \partial \eta_{2t} / \partial \boldsymbol{\lambda}$ têm postos r e s , respectivamente.

Através da função de calibração estimada na função densidade de probabilidade de y_t dado x_t , obtém-se o logaritmo da função de verossimilhança modificada, que é

$$\ell_{rc}(\boldsymbol{\theta}) = \sum_{t=1}^n \ell(\mu_t, \phi_t),$$

sendo

$$\begin{aligned}\ell_{tq}(\mu_t, \phi_t) &= \log \Gamma(\phi_{tq}) - \log \Gamma(\mu_{tq}\phi_{tq}) - \log \Gamma[(1 - \mu_{tq})\phi_{tq}] + (\mu_{tq}\phi_{tq} - 1) \log y_t + \\ &\quad [(1 - \mu_{tq})\phi_{tq} - 1] \log(1 - y_t), \\ g(\mu_{tq}) &= f_1(\mathbf{z}_t^\top, x_t^*; \boldsymbol{\alpha}, \beta), \\ h(\phi_{tq}) &= f_2(\mathbf{v}_t^\top, x_t^*; \boldsymbol{\gamma}, \lambda), \\ x_t^* &= \bar{w} + \hat{k}_x(w_t - \bar{w}),\end{aligned}$$

em que \hat{k}_x é o estimador de k_x .

Observe que o logaritmo da função de verossimilhança modificada depende apenas do parâmetro de interesse $\boldsymbol{\theta}$ e coincide com o logaritmo da função de verossimilhança do modelo de regressão beta usual, uma vez que x_t^* atua como uma variável explicativa medida sem erro.

O vetor escore é denotado por $\mathbf{U}_{rc} = (\mathbf{U}_{rc\boldsymbol{\alpha}}(\boldsymbol{\theta}), \mathbf{U}_{rc\beta}(\boldsymbol{\theta}), \mathbf{U}_{rc\boldsymbol{\gamma}}(\boldsymbol{\theta}), \mathbf{U}_{rc\lambda}(\boldsymbol{\theta}))^\top$ é apresentado a seguir. Os componentes do vetor escore para $\boldsymbol{\alpha}$, vetor de dimensão p , são dados por

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\alpha}, \beta, \boldsymbol{\gamma}, \lambda)}{\partial \alpha_i} &= \sum_{t=1}^n \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \alpha_i} = \sum_{t=1}^n \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} \frac{d\mu_t}{d\eta_{1t}} \frac{\partial \eta_{1t}}{\partial \beta_i}, \quad i = 1, \dots, p \\ &= \sum_{t=1}^n \phi_t (y_t^* - \mu_t^*) \frac{1}{g'(\mu_t)} \frac{\partial \eta_{1t}}{\eta \alpha_i},\end{aligned}$$

$$y_t^* = \log \left(\frac{y_t}{1 - y_t} \right) \quad \text{e} \quad \mu_{tq}^* = \psi(\mu_{tq}\phi_{tq}) - \psi[(1 - \mu_{tq})\phi_{tq}].$$

A função escore para o parâmetro correspondente à covariável medida com erro denotado por β , é dada por

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\alpha}, \beta, \boldsymbol{\gamma}, \lambda)}{\partial \beta} &= \sum_{t=1}^n \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \beta} = \sum_{t=1}^n \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \mu_t} \frac{d\mu_t}{d\eta_{1t}} \frac{\partial \eta_{1t}}{\partial \beta} \\ &= \sum_{t=1}^n \phi_t (y_t^* - \mu_t^*) \frac{1}{g'(\mu_t)} \frac{\partial \eta_{1t}}{\eta \beta}.\end{aligned}$$

Em forma matricial,

$$\begin{aligned}\mathbf{U}_{rc\boldsymbol{\alpha}}(\boldsymbol{\theta}) &= \mathbf{F}_1^\top \boldsymbol{\Phi} \mathbf{T} (\mathbf{y}^* - \boldsymbol{\mu}^*) \\ \mathbf{U}_{rc\beta}(\boldsymbol{\theta}) &= \mathbf{F}_2^\top \boldsymbol{\Phi} \mathbf{T} (\mathbf{y}^* - \boldsymbol{\mu}^*),\end{aligned}$$

em que $\mathbf{F}_1^\top = \partial \eta_1 / \partial \boldsymbol{\alpha}$ é uma matriz de ordem $n \times p$, $\boldsymbol{\Phi} = \text{diag}\{\phi_1, \dots, \phi_n\}$, $\mathbf{T} = \text{diag}\{1/g'(\mu_1), \dots, 1/g'(\mu_n)\}$.

Os componentes do vetor escore para $\boldsymbol{\gamma}$, vetor de dimensão q e λ são dados por

$$\frac{\partial \ell(\boldsymbol{\alpha}, \beta, \boldsymbol{\gamma}, \lambda)}{\partial \gamma_j} = \sum_{t=1}^n \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \gamma_j} = \sum_{t=1}^n \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \phi_t} \frac{d\phi_t}{d\eta_{2t}} \frac{\partial \eta_{2t}}{\partial \gamma_j}$$

e

$$\frac{\partial \ell(\boldsymbol{\alpha}, \beta, \boldsymbol{\gamma}, \lambda)}{\partial \lambda} = \sum_{t=1}^n \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \lambda} = \sum_{t=1}^n \frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \phi_t} \frac{d\phi_t}{d\eta_{2t}} \frac{\partial \eta_{2t}}{\partial \lambda}, \quad j = 1, \dots, \check{q},$$

em que

$$\frac{\partial \ell_t(\mu_t, \phi_t)}{\partial \phi_t} = \mu_t(y_t^* - \mu_t^*) + \log(1 - y_t) = \psi((1 - \mu_t)\phi_t) + \psi(\phi_t) = a_t.$$

Além disso, $d\phi_t/d\eta_{2t} = 1/h'(\phi_t)$. Em forma matricial

$$\begin{aligned} \mathbf{U}_{rc\boldsymbol{\gamma}}(\boldsymbol{\theta}) &= \mathbf{F}_3^\top \mathbf{H} \mathbf{a} \\ \mathbf{U}_{rc\lambda}(\boldsymbol{\theta}) &= \mathbf{F}_4^\top \mathbf{H} \mathbf{a}, \end{aligned}$$

em que $\mathbf{F}_3^\top = \partial \eta_{2j} / \partial \boldsymbol{\gamma}$ é uma matriz de ordem $n \times \check{q}$ e $\mathbf{F}_4^\top = \partial \eta_{2j} / \partial \lambda$ é um vetor s -dimensional, $\mathbf{H} = \text{diag}\{1/h'(\phi_1), \dots, 1/h'(\phi_n)\}$ e $\mathbf{a} = (a_1, \dots, a_n)^\top$.

Apêndice A.2 - Matriz de Informação Observada - MVa

Nesta seção são apresentados os elementos das matriz de informação observada denotada por

$$J(\boldsymbol{\Psi}) = -\frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{\partial \boldsymbol{\Psi} \partial \boldsymbol{\Psi}^\top}.$$

A representação matricial da matriz de informação observada tem a forma

$$J(\boldsymbol{\Psi}) = \begin{bmatrix} \frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^\top} & \frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{\partial \boldsymbol{\alpha} d\beta} & \frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\gamma}^\top} & \frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{\partial \boldsymbol{\alpha} d\lambda} & \frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{\partial \boldsymbol{\alpha} d\mu_x} & \frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{\partial \boldsymbol{\alpha} d\sigma_x^2} \\ & \frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{d\beta d\beta} & \frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{d\beta \partial \boldsymbol{\gamma}^\top} & \frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{d\beta d\lambda} & \frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{d\beta d\mu_x} & \frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{d\beta d\sigma_x^2} \\ & & \frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top} & \frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{\partial \boldsymbol{\gamma} d\lambda} & \frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{\partial \boldsymbol{\gamma} d\mu_x} & \frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{\partial \boldsymbol{\gamma} d\sigma_x^2} \\ & & & \frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{d\lambda d\lambda} & \frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{d\lambda d\mu_x} & \frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{\partial \lambda \partial \sigma_x^2} \\ & & & & \frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{d\mu_x d\mu_x} & \frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{d\mu_x d\sigma_x^2} \\ & & & & & \frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{d\sigma_x^2 d\sigma_x^2} \end{bmatrix}_{n(\boldsymbol{\Psi}) \times n(\boldsymbol{\Psi})}$$

em que $n(\boldsymbol{\Psi}) = (p + 1) + (\check{q} + 1) + 2 < n$. Considerando as primeiras derivadas de $\ell_a(\boldsymbol{\Psi})$ mostradas na seção anterior, calcula-se as expressões das segundas derivadas de $\ell_a(\boldsymbol{\Psi})$ como é mostrado a seguir. Para $i, j = 1, 2, \dots, p$ temos que

$$\frac{\partial^2 \ell_a(\boldsymbol{\Psi})}{\partial \alpha_i \partial \alpha_j} = J_{a(\alpha_i \alpha_j)} = \sum_{t=1}^n \left[\dot{b}_{t(\alpha_j)} a_{t\alpha_i} + b_t \dot{a}_{t\alpha_i(\alpha_j)} \right],$$

em que

$$\dot{b}_{t(\alpha_j)} = -b_t^2 \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \dot{\ell}_{tq(\alpha_j)}(\mu_{tq}, \phi_{tq}),$$

$$\dot{a}_{t\alpha_i(\alpha_j)} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \left[\dot{\ell}_{tq(\alpha_j)}(\mu_{tq}, \phi_{tq}) \dot{\ell}_{tq(\alpha_i)}(\mu_{tq}, \phi_{tq}) + \ddot{\ell}_{tq(\alpha_i)(\alpha_j)}(\mu_{tq}, \phi_{tq}) \right] \quad e$$

$$\ddot{\ell}_{tq(\alpha_i)(\alpha_j)}(\mu_{tq}, \phi_{tq}) = -\phi_{tq} \left[u_{tq} + (y_t^* - \mu_{tq}^*) \frac{g''(\mu_{tq})}{[g'(\mu_{tq})]^3} \right] \frac{\partial \eta_{1t}}{\partial \alpha_i} \frac{\partial \eta_{1t}}{\partial \alpha_j},$$

sendo $u_{tq} = \phi_{tq} \{ \psi'(\mu_{tq} \phi_{tq}) + \psi'[(1 - \mu_{tq}) \phi_{tq}] \} [g'(\mu_{tq})]^{-2}$, b_t dado em (3), $a_{t\alpha_i}$ dado em (4) e $\dot{\ell}_{tq(\alpha_i)}(\mu_{tq}, \phi_{tq})$ dado em (5).

Temos que $\frac{\partial^2 \ell_a(\Psi)}{\partial \alpha_i \partial \beta} = J_{a(\alpha_i \beta)} = \sum_{t=1}^n \left[\dot{b}_{t(\beta)} a_{t\alpha_i} + b_t \dot{a}_{t\alpha_i(\beta)} \right]$, em que:

$$\dot{b}_{t(\beta)} = -b_t^2 \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \dot{\ell}_{tq(\beta)}(\mu_{tq}, \phi_{tq}),$$

$$\dot{a}_{t\alpha_i(\beta)} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \left[\dot{\ell}_{tq(\beta)}(\mu_{tq}, \phi_{tq}) \dot{\ell}_{tq(\alpha_i)}(\mu_{tq}, \phi_{tq}) + \ddot{\ell}_{tq(\alpha_i)(\beta)}(\mu_{tq}, \phi_{tq}) \right] \quad e$$

$$\ddot{\ell}_{tq(\alpha_i)(\beta)}(\mu_{tq}, \phi_{tq}) = -\phi_{tq} \left[u_{tq} + (y_t^* - \mu_{tq}^*) \frac{g''(\mu_{tq})}{[g'(\mu_{tq})]^3} \right] \frac{\partial \eta_{1t}}{\partial \alpha_i} \frac{\partial \eta_{1t}}{\partial \beta},$$

sendo $\dot{\ell}_{tq(\beta)}(\mu_{tq}, \phi_{tq})$ dado em (7).

Para $j = 1, 2, \dots, \check{q}$ temos que $\frac{\partial^2 \ell_a(\Psi)}{\partial \alpha_i \partial \gamma_j} = J_{a(\alpha_i \gamma_j)} = \sum_{t=1}^n \left[\dot{b}_{t(\gamma_j)} a_{t\alpha_i} + b_t \dot{a}_{t\alpha_i(\gamma_j)} \right]$, em que

$$\dot{b}_{t(\gamma_j)} = -b_t^2 \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \dot{\ell}_{tq(\gamma_j)}(\mu_{tq}, \phi_{tq}),$$

$$\dot{a}_{t\alpha_i(\gamma_j)} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \left[\dot{\ell}_{tq(\gamma_j)}(\mu_{tq}, \phi_{tq}) \dot{\ell}_{tq(\alpha_i)}(\mu_{tq}, \phi_{tq}) + \ddot{\ell}_{tq(\alpha_i)(\gamma_j)}(\mu_{tq}, \phi_{tq}) \right] \quad e$$

$$\ddot{\ell}_{tq(\alpha_i)(\gamma_j)}(\mu_{tq}, \phi_{tq}) = [(y_t^* - \mu_{tq}^*) - c_{tq}] \frac{1}{g'(\mu_{tq})} \frac{1}{h'(\phi_{tq})} \frac{\partial \eta_{1t}}{\partial \alpha_i} \frac{\partial \eta_{2t}}{\partial \gamma_j},$$

sendo $c_{tq} = \phi_{tq} [\psi'(\mu_{tq} \phi_{tq}) \mu_{tq} + \psi'[(1 - \mu_{tq}) \phi_{tq}] (1 - \mu_{tq})]$ e $\dot{\ell}_{tq(\gamma_j)}(\mu_{tq}, \phi_{tq})$ dado em (10).

Temos que $\frac{\partial^2 \ell_a(\Psi)}{\partial \alpha_i \partial \lambda} = J_{a(\alpha_i \lambda)} = \sum_{t=1}^n \left[\dot{b}_t(\lambda) a_{t\alpha_i} + b_t \dot{a}_{t\alpha_i(\lambda)} \right]$, em que

$$\dot{b}_t(\lambda) = -b_t^2 \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \dot{\ell}_{tq(\lambda)}(\mu_{tq}, \phi_{tq}),$$

$$\dot{a}_{t\alpha_i(\lambda)} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \left[\dot{\ell}_{tq(\lambda)}(\mu_{tq}, \phi_{tq}) \dot{\ell}_{tq(\alpha_i)}(\mu_{tq}, \phi_{tq}) + \ddot{\ell}_{tq(\alpha_i)(\lambda)}(\mu_{tq}, \phi_{tq}) \right] \quad e$$

$$\ddot{\ell}_{tq(\alpha_i)(\lambda)}(\mu_{tq}, \phi_{tq}) = \left[(y_t^* - \mu_{tq}^*) - c_{tq} \right] \frac{1}{g'(\mu_{tq})} \frac{1}{h'(\phi_{tq})} \frac{\partial \eta_{1t}}{\partial \alpha_i} \frac{\partial \eta_{2t}}{\partial \lambda},$$

sendo $\dot{\ell}_{tq(\lambda)}(\mu_{tq}, \phi_{tq})$ dado em (11).

$\frac{\partial^2 \ell_a(\Psi)}{\partial \alpha_i \partial \mu_x} = J_{a(\alpha_i \mu_x)} = \sum_{t=1}^n \left[\dot{b}_t(\mu_x) a_{t\alpha_i} + b_t \dot{a}_{t\alpha_i(\mu_x)} \right]$, em que

$$\dot{b}_t(\mu_x) = -b_t^2 \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \dot{\ell}_{tq(\mu_x)}(\mu_{tq}, \phi_{tq}),$$

$$\dot{a}_{t\alpha_i(\mu_x)} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \left[\dot{\ell}_{tq(\mu_x)}(\mu_{tq}, \phi_{tq}) \dot{\ell}_{tq(\alpha_i)}(\mu_{tq}, \phi_{tq}) + \ddot{\ell}_{tq(\alpha_i)(\mu_x)}(\mu_{tq}, \phi_{tq}) \right] \quad e$$

$$\begin{aligned} \ddot{\ell}_{tq(\alpha_i)(\mu_x)}(\mu_{tq}, \phi_{tq}) &= -\phi_{tq} \left\{ u_{tq} + (y_t^* - \mu_{tq}^*) \frac{g''(\mu_{tq})}{[g'(\mu_{tq})]^3} \right\} \frac{\partial \eta_{1t}}{\partial \alpha_i} \frac{\partial \eta_{1t}}{\partial \mu_x} \\ &\quad + \left\{ (y_t^* - \mu_{tq}^*) - c_{tq} \right\} \frac{1}{g'(\mu_{tq})} \frac{1}{h'(\phi_{tq})} \frac{\partial \eta_{1t}}{\partial \alpha_i} \frac{\partial \eta_{2t}}{\partial \mu_x}. \end{aligned}$$

$\frac{\partial^2 \ell_a(\Psi)}{\partial \alpha_i \partial \sigma_x^2} = J_{a(\alpha_i \sigma_x^2)} = \sum_{t=1}^n \left[\dot{b}_t(\sigma_x^2) a_{t\alpha_i} + b_t \dot{a}_{t\alpha_i(\sigma_x^2)} \right]$, em que

$$\dot{b}_t(\sigma_x^2) = -b_t^2 \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \dot{\ell}_{tq(\sigma_x^2)}(\mu_{tq}, \phi_{tq}),$$

$$\dot{a}_{t\alpha_i(\sigma_x^2)} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \left[\dot{\ell}_{tq(\sigma_x^2)}(\mu_{tq}, \phi_{tq}) \dot{\ell}_{tq(\alpha_i)}(\mu_{tq}, \phi_{tq}) + \ddot{\ell}_{tq(\alpha_i)(\sigma_x^2)}(\mu_{tq}, \phi_{tq}) \right] \quad e$$

$$\begin{aligned} \ddot{\ell}_{tq(\alpha_i)(\sigma_x^2)}(\mu_{tq}, \phi_{tq}) &= \left[-\phi_{tq} \left(u_{tq} + (y_t^* - \mu_{tq}^*) \frac{g''(\mu_{tq})}{[g'(\mu_{tq})]^3} \right) \right] \frac{\partial \eta_{1t}}{\partial \alpha_i} \frac{\partial \eta_{1t}}{\partial \sigma_x^2} \\ &\quad + \left[(y_t^* - \mu_{tq}^*) - c_{tq} \right] \frac{1}{g'(\mu_{tq})} \frac{1}{h'(\phi_{tq})} \frac{\partial \eta_{1t}}{\partial \alpha_i} \frac{\partial \eta_{2t}}{\partial \sigma_x^2}. \end{aligned}$$

Em relação ao parâmetro associado a covariável mensurada com erro, tem-se que

$$\frac{\partial^2 \ell_a(\Psi)}{d\beta d\beta} = J_{a(\beta\beta)} = \sum_{t=1}^n \left[\dot{b}_{t(\beta)} a_{t\beta} + b_t \dot{a}_{t\beta(\beta)} \right],$$

em que

$$\dot{a}_{t\beta(\beta)} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \left[\dot{\ell}_{tq(\beta)}(\mu_{tq}, \phi_{tq}) \dot{\ell}_{tq(\beta)}(\mu_{tq}, \phi_{tq}) + \ddot{\ell}_{tq(\beta)(\beta)}(\mu_{tq}, \phi_{tq}) \right] \quad e$$

$$\ddot{\ell}_{tq(\beta)(\beta)}(\mu_{tq}, \phi_{tq}) = -\phi_{tq} \left[u_{tq} + (y_t^* - \mu_{tq}^*) \frac{g''(\mu_{tq})}{[g'(\mu_{tq})]^3} \right] \left(\frac{\partial \eta_{1t}}{\partial \beta} \right)^2.$$

Para $j = 1, 2, \dots, \check{q}$ temos que $\frac{\partial^2 \ell_a(\Psi)}{d\beta d\gamma_j} = J_{a(\beta\gamma_j)} = \sum_{t=1}^n \left[\dot{b}_{t(\gamma_j)} a_{t\beta} + b_t \dot{a}_{t\beta(\gamma_j)} \right]$, em que

$$\dot{a}_{t\beta(\gamma_j)} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \left[\dot{\ell}_{tq(\gamma_j)}(\mu_{tq}, \phi_{tq}) \dot{\ell}_{tq(\beta)}(\mu_{tq}, \phi_{tq}) + \ddot{\ell}_{tq(\beta)(\gamma_j)}(\mu_{tq}, \phi_{tq}) \right] \quad e$$

$$\ddot{\ell}_{tq(\beta)(\gamma_j)}(\mu_{tq}, \phi_{tq}) = [(y_t^* - \mu_{tq}^*) - c_{tq}] \frac{1}{g'(\mu_{tq})} \frac{1}{h'(\phi_{tq})} \frac{\partial \eta_{1t}}{\partial \beta} \frac{\partial \eta_{2t}}{\partial \gamma_j},$$

sendo $\dot{\ell}_{tq(\mu_x)}(\mu_{tq}, \phi_{tq})$ e $\dot{\ell}_{tq(\sigma_x^2)}(\mu_{tq}, \phi_{tq})$ dados, respectivamente, por (17) e (18).

Agora $\frac{\partial^2 \ell_a(\Psi)}{d\beta d\lambda} = J_{a(\beta\lambda)} = \sum_{t=1}^n \left[\dot{b}_{t(\lambda)} a_{t\beta} + b_t \dot{a}_{t\beta(\lambda)} \right]$, em que

$$\dot{a}_{t\beta(\lambda)} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \left[\dot{\ell}_{tq(\lambda)}(\mu_{tq}, \phi_{tq}) \dot{\ell}_{tq(\beta)}(\mu_{tq}, \phi_{tq}) + \ddot{\ell}_{tq(\beta)(\lambda)}(\mu_{tq}, \phi_{tq}) \right] \quad e$$

$$\ddot{\ell}_{tq(\beta)(\lambda)}(\mu_{tq}, \phi_{tq}) = [(y_t^* - \mu_{tq}^*) - c_{tq}] \frac{1}{g'(\mu_{tq})} \frac{1}{h'(\phi_{tq})} \frac{\partial \eta_{1t}}{\partial \beta} \frac{\partial \eta_{2t}}{\partial \lambda},$$

sendo $a_{t\beta}$ dado em (6).

Agora $\frac{\partial^2 \ell_a(\Psi)}{d\beta d\mu_x} = J_{a(\beta\mu_x)} = \sum_{t=1}^n \left[\dot{b}_{t(\mu_x)} a_{t\beta} + b_t \dot{a}_{t\beta(\mu_x)} \right]$, em que

$$\dot{a}_{t\beta(\mu_x)} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \left[\dot{\ell}_{tq(\mu_x)}(\mu_{tq}, \phi_{tq}) \dot{\ell}_{tq(\beta)}(\mu_{tq}, \phi_{tq}) + \ddot{\ell}_{tq(\beta)(\mu_x)}(\mu_{tq}, \phi_{tq}) \right] \quad e$$

$$\begin{aligned} \ddot{\ell}_{tq(\beta)(\mu_x)}(\mu_{tq}, \phi_{tq}) &= -\phi_{tq} \left\{ u_{tq} + (y_t^* - \mu_{tq}^*) \frac{g''(\mu_{tq})}{[g'(\mu_{tq})]^3} \right\} \frac{\partial \eta_{1t}}{\partial \beta} \frac{\partial \eta_{1t}}{\partial \mu_x} \\ &\quad + \{(y_t^* - \mu_{tq}^*) - c_{tq}\} \frac{1}{g'(\mu_{tq})} \frac{1}{h'(\phi_{tq})} \frac{\partial \eta_{1t}}{\partial \beta} \frac{\partial \eta_{2t}}{\partial \mu_x}. \end{aligned}$$

Temos que $\frac{\partial^2 \ell_a(\Psi)}{\partial \beta \partial \sigma_x^2} = J_{a(\beta \sigma_x^2)} = \sum_{t=1}^n \left[\dot{b}_{t(\sigma_x^2)} a_{t\beta} + b_t \dot{a}_{t\beta(\sigma_x^2)} \right]$, em que

$$\dot{a}_{t\beta(\sigma_x^2)} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \left[\dot{\ell}_{tq(\sigma_x^2)}(\mu_{tq}, \phi_{tq}) \dot{\ell}_{tq(\beta)}(\mu_{tq}, \phi_{tq}) + \ddot{\ell}_{tq(\beta)(\sigma_x^2)}(\mu_{tq}, \phi_{tq}) \right] \quad e$$

$$\begin{aligned} \ddot{\ell}_{tq(\beta)(\sigma_x^2)}(\mu_{tq}, \phi_{tq}) &= -\phi_{tq} \left(u_{tq} + (y_t^* - \mu_{tq}^*) \frac{g''(\mu_{tq})}{[g'(\mu_{tq})]^3} \right) \frac{\partial \eta_{1t}}{\partial \beta} \frac{\partial \eta_{1t}}{\partial \sigma_x^2} \\ &\quad + \lambda [(y_t^* - \mu_{tq}^*) - c_{tq}] \frac{1}{g'(\mu_{tq})} \frac{1}{h'(\phi_{tq})} \frac{\partial \eta_{1t}}{\partial \beta} \frac{\partial \eta_{2t}}{\partial \sigma_x^2}. \end{aligned}$$

Para $k, j = 1, 2, \dots, \check{q}$ temos que

$$\frac{\partial^2 \ell_a(\Psi)}{\partial \gamma_k \partial \gamma_j} = J_{a(\gamma_k \gamma_j)} = \sum_{t=1}^n \left[\dot{b}_{t(\gamma_j)} a_{t\gamma_k} + b_t \dot{a}_{t\gamma_k(\gamma_j)} \right],$$

em que

$$\dot{a}_{t\gamma_k(\gamma_j)} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \left[\dot{\ell}_{tq(\gamma_j)}(\mu_{tq}, \phi_{tq}) \dot{\ell}_{tq(\gamma_k)}(\mu_{tq}, \phi_{tq}) + \ddot{\ell}_{tq\gamma_k(\gamma_j)}(\mu_{tq}, \phi_{tq}) \right] \quad e$$

$$\ddot{\ell}_{tq(\gamma_k)(\gamma_j)}(\mu_{tq}, \phi_{tq}) = \left\{ d_{tq} - a_{tq} \frac{h''(\phi_{tq})}{[h'(\phi_{tq})]^3} \right\} \frac{\partial \eta_{2t}}{\partial \gamma_k} \frac{\partial \eta_{2t}}{\partial \gamma_j},$$

sendo $d_{tq} = \mu_{tq} \{-\mu_{tq} \psi'(\mu_{tq} \phi_{tq}) - (1 - \mu_{tq}) \psi'[(1 - \mu_{tq}) \phi_{tq}]\} + \psi'(\phi_{tq}) - (1 - \mu_{tq}) \psi'[(1 - \mu_{tq}) \phi_{tq}]$, $a_{t\gamma_k}$ e a_{tq} são definidos em 8 e 12.

Agora $\frac{\partial^2 \ell_a(\Psi)}{\partial \gamma_k \partial \lambda} = J_{a(\gamma_k \lambda)} = \sum_{t=1}^n \left[\dot{b}_{t(\lambda)} a_{t\gamma_k} + b_t \dot{a}_{t\gamma_k(\lambda)} \right]$, sendo

$$\dot{a}_{t\gamma_k(\lambda)} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \left[\dot{\ell}_{tq(\lambda)}(\mu_{tq}, \phi_{tq}) \dot{\ell}_{tq(\gamma_k)}(\mu_{tq}, \phi_{tq}) + \ddot{\ell}_{tq(\gamma_k)(\lambda)}(\mu_{tq}, \phi_{tq}) \right] \quad e$$

$$\ddot{\ell}_{tq(\gamma_k)(\lambda)}(\mu_{tq}, \phi_{tq}) = \left\{ d_{tq} - a_{tq} \frac{h''(\phi_{tq})}{[h'(\phi_{tq})]^3} \right\} \frac{\partial \eta_{2t}}{\partial \gamma_k} \frac{\partial \eta_{2t}}{\partial \lambda}.$$

Agora $\frac{\partial^2 \ell_a(\Psi)}{\partial \gamma_k \partial \mu_x} = J_{a(\gamma_k \mu_x)} = \sum_{t=1}^n \left[\dot{b}_{t(\mu_x)} a_{t\gamma_k} + b_t \dot{a}_{t\gamma_k(\mu_x)} \right]$, em que

$$\dot{a}_{t\gamma_k(\mu_x)} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \left[\dot{\ell}_{tq(\mu_x)}(\mu_{tq}, \phi_{tq}) \dot{\ell}_{tq(\gamma_k)}(\mu_{tq}, \phi_{tq}) + \ddot{\ell}_{tq(\gamma_k)(\mu_x)}(\mu_{tq}, \phi_{tq}) \right] \quad e$$

$$\begin{aligned} \ddot{\ell}_{tq(\gamma_k)(\mu_x)}(\mu_{tq}, \phi_{tq}) &= \{(y_t^* - \mu_{tq}^*) - c_{tq}\} \frac{1}{g'(\mu_{tq})} \frac{1}{h'(\phi_{tq})} \frac{\partial \eta_{2t}}{\partial \gamma_k} \frac{\partial \eta_{1t}}{\partial \mu_x} \\ &\quad + \left\{ d_{tq} - a_{tq} \frac{h''(\phi_{tq})}{[h'(\phi_{tq})]^3} \right\} \frac{\partial \eta_{2t}}{\partial \gamma_k} \frac{\partial \eta_{2t}}{\partial \mu_x}. \end{aligned}$$

Temos que $\frac{\partial^2 \ell_a(\Psi)}{\partial \gamma_k \partial \sigma_x^2} = J_{a(\gamma_k \sigma_x^2)} = \sum_{t=1}^n \left[\dot{b}_{t(\sigma_x^2)} a_{t\gamma_k} + b_t \dot{a}_{t\gamma_k(\sigma_x^2)} \right]$, sendo

$$\dot{a}_{t\gamma_k(\sigma_x^2)} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \left[\dot{\ell}_{tq(\sigma_x^2)}(\mu_{tq}, \phi_{tq}) \dot{\ell}_{tq(\gamma_k)}(\mu_{tq}, \phi_{tq}) + \ddot{\ell}_{tq(\gamma_k)(\sigma_x^2)}(\mu_{tq}, \phi_{tq}) \right] \quad e$$

$$\begin{aligned} \ddot{\ell}_{tq(\gamma_k)(\sigma_x^2)}(\mu_{tq}, \phi_{tq}) &= \left[(y_t^* - \mu_{tq}^*) - c_{tq} \right] \frac{1}{g'(\mu_{tq})} \frac{1}{h'(\phi_{tq})} \frac{\partial \eta_{2t}}{\partial \gamma_k} \frac{\partial \eta_{2t}}{\partial \sigma_x^2} \\ &\quad + \left[d_{tq} - a_{tq} \frac{h''(\phi_{tq})}{[h'(\phi_{tq})]^3} \right] \frac{\partial \eta_{2t}}{\partial \gamma_k} \frac{\partial \eta_{2t}}{\partial \sigma_x^2}. \end{aligned}$$

Para o parâmetro associado a variável mensurada com erro, tem-se que

$$\frac{\partial^2 \ell_a(\Psi)}{d\lambda d\lambda} = J_{a(\lambda\lambda)} = \sum_{t=1}^n \left[\dot{b}_{t(\lambda)} a_{t\lambda} + b_t \dot{a}_{t\lambda(\lambda)} \right],$$

em que

$$\dot{a}_{t\lambda(\lambda)} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \left[\dot{\ell}_{tq(\lambda)}(\mu_{tq}, \phi_{tq}) \dot{\ell}_{tq(\lambda)}(\mu_{tq}, \phi_{tq}) + \ddot{\ell}_{tq(\lambda)(\lambda)}(\mu_{tq}, \phi_{tq}) \right] \quad e$$

$$\ddot{\ell}_{tq(\lambda)(\lambda)}(\mu_{tq}, \phi_{tq}) = \left\{ d_{tq} - a_{tq} \frac{h''(\phi_{tq})}{[h'(\phi_{tq})]^3} \right\} \frac{\partial \eta_{2t}}{\partial \lambda} \frac{\partial \eta_{2t}}{\partial \lambda},$$

sendo $a_{t\lambda}$ dado em (9).

Agora $\frac{\partial^2 \ell_a(\Psi)}{d\lambda d\mu_x} = J_{a(\lambda\mu_x)} = \sum_{t=1}^n \left[\dot{b}_{t(\mu_x)} a_{t\lambda} + b_t \dot{a}_{t\lambda(\mu_x)} \right]$, em que

$$\dot{a}_{t\lambda(\mu_x)} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \left[\dot{\ell}_{tq(\mu_x)}(\mu_{tq}, \phi_{tq}) \dot{\ell}_{tq(\lambda)}(\mu_{tq}, \phi_{tq}) + \ddot{\ell}_{tq(\lambda)(\mu_x)}(\mu_{tq}, \phi_{tq}) \right] \quad e$$

$$\begin{aligned} \ddot{\ell}_{tq(\lambda)(\mu_x)}(\mu_{tq}, \phi_{tq}) &= \left\{ (y_t^* - \mu_{tq}^*) - c_{tq} \right\} \frac{1}{g'(\mu_{tq})} \frac{1}{h'(\phi_{tq})} \frac{\partial \eta_{2t}}{\partial \lambda} \frac{\partial \eta_{1t}}{\partial \mu_x} \\ &\quad + \left\{ d_{tq} - a_{tq} \frac{h''(\phi_{tq})}{[h'(\phi_{tq})]^3} \right\} \frac{\partial \eta_{2t}}{\partial \lambda} \frac{\partial \eta_{2t}}{\partial \mu_x}. \end{aligned}$$

Agora $\frac{\partial^2 \ell_a(\Psi)}{d\lambda d\sigma_x^2} = J_{a(\lambda\sigma_x^2)} = \sum_{t=1}^n \left[\dot{b}_{t(\sigma_x^2)} a_{t\lambda} + b_t \dot{a}_{t\lambda(\sigma_x^2)} \right]$, sendo

$$\dot{a}_{t\lambda(\sigma_x^2)} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \left[\dot{\ell}_{tq(\sigma_x^2)}(\mu_{tq}, \phi_{tq}) \dot{\ell}_{tq(\lambda)}(\mu_{tq}, \phi_{tq}) + \ddot{\ell}_{tq(\lambda)(\sigma_x^2)}(\mu_{tq}, \phi_{tq}) \right] \quad e$$

$$\begin{aligned} \ddot{\ell}_{tq(\lambda)(\sigma_x^2)}(\mu_{tq}, \phi_{tq}) &= \left[(y_t^* - \mu_{tq}^*) - c_{tq} \right] \frac{1}{g'(\mu_{tq})} \frac{1}{h'(\phi_{tq})} \frac{\partial \eta_{2t}}{\partial \lambda} \frac{\partial \eta_{1t}}{\partial \sigma_x^2} \\ &\quad + \left[d_{tq} - a_{tq} \frac{h''(\phi_{tq})}{[h'(\phi_{tq})]^3} \right] \frac{\partial \eta_{2t}}{\partial \lambda} \frac{\partial \eta_{2t}}{\partial \sigma_x^2}. \end{aligned}$$

Temos que $\frac{\partial^2 \ell_a(\Psi)}{d\mu_x d\mu_x} = J_{a(\mu_x \mu_x)} = -\frac{n}{(\sigma_x^2 + \sigma_e^2)} + \sum_{t=1}^n \left[\dot{b}_{t(\mu_x)} a_{t\mu_x} + b_t \dot{a}_{t\mu_x(\mu_x)} \right]$, sendo

$$\dot{a}_{t\mu_x(\mu_x)} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \left[\dot{\ell}_{tq(\mu_x)}(\mu_{tq}, \phi_{tq}) \dot{\ell}_{tq(\mu_x)}(\mu_{tq}, \phi_{tq}) + \ddot{\ell}_{tq(\mu_x)(\mu_x)}(\mu_{tq}, \phi_{tq}) \right] \quad e$$

$$\begin{aligned} \ddot{\ell}_{tq(\mu_x)(\mu_x)}(\mu_{tq}, \phi_{tq}) &= -\phi_{tq} \left[u_{tq} + (y_t^* - \mu_{tq}^*) \frac{g''(\mu_{tq})}{[g'(\mu_{tq})]^3} \right] \frac{\partial \eta_{1t}}{\partial \mu_x} \frac{\partial \eta_{1t}}{\partial \mu_x} \\ &+ 2 \left[(y_t^* - \mu_{tq}^*) - c_{tq} \right] \frac{1}{h'(\phi_{tq}) g'(\mu_{tq})} \frac{\partial \eta_{1t}}{\partial \mu_x} \frac{\partial \eta_{2t}}{\partial \mu_x} \\ &+ \left[d_{tq} - a_{tq} \frac{h''(\phi_{tq})}{[h'(\phi_{tq})]^3} \right] \frac{\partial \eta_{2t}}{\partial \mu_x} \frac{\partial \eta_{2t}}{\partial \mu_x}, \end{aligned}$$

em que $a_{t\mu_x}$ é apresentado em (15).

Agora $\frac{\partial^2 \ell_a(\Psi)}{d\mu_x d\sigma_x^2} = J_{a(\mu_x \sigma_x^2)} = -\frac{n}{(\sigma_x^2 + \sigma_e^2)} + \sum_{t=1}^n \left[\dot{b}_{t(\sigma_x^2)} a_{t\mu_x} + b_t \dot{a}_{t\mu_x(\sigma_x^2)} \right]$, em que

$$\dot{a}_{t\mu_x(\sigma_x^2)} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \left[\dot{\ell}_{tq(\sigma_x^2)}(\mu_{tq}, \phi_{tq}) \dot{\ell}_{tq(\sigma_x^2)}(\mu_{tq}, \phi_{tq}) + \ddot{\ell}_{tq(\mu_x)(\sigma_x^2)}(\mu_{tq}, \phi_{tq}) \right] \quad e$$

$$\begin{aligned} \ddot{\ell}_{tq(\mu_x)(\sigma_x^2)}(\mu_{tq}, \phi_{tq}) &= -\phi_{tq} \left[u_{tq} + (y_t^* - \mu_{tq}^*) \frac{g''(\mu_{tq})}{[g'(\mu_{tq})]^3} \right] \frac{\partial \eta_{1t}}{\partial \mu_x} \frac{\partial \eta_{1t}}{\partial \sigma_x^2} \\ &+ \left[(y_t^* - \mu_{tq}^*) - c_{tq} \right] \frac{1}{h'(\phi_{tq}) g'(\mu_{tq})} \frac{\partial \eta_{2t}}{\partial \mu_x} \frac{\partial \eta_{1t}}{\partial \sigma_x^2} \\ &+ \left[(y_t^* - \mu_{tq}^*) - c_{tq} \right] \frac{1}{h'(\phi_{tq}) g'(\mu_{tq})} \frac{\partial \eta_{1t}}{\partial \mu_x} \frac{\partial \eta_{2t}}{\partial \sigma_x^2} \\ &+ \left[d_{tq} - a_{tq} \frac{h''(\phi_{tq})}{[h'(\phi_{tq})]^3} \right] \frac{\partial \eta_{2t}}{\partial \mu_x} \frac{\partial \eta_{2t}}{\partial \sigma_x^2}. \end{aligned}$$

Temos que $\frac{\partial^2 \ell_a(\Psi)}{d\sigma_x^2 d\sigma_x^2} = J_{a(\sigma_x^2 \sigma_x^2)} = -\frac{n}{(\sigma_x^2 + \sigma_e^2)} + \sum_{t=1}^n \left[\dot{b}_{t(\sigma_x^2)} a_{t\mu_x} + b_t \dot{a}_{t\sigma_x^2(\sigma_x^2)} \right]$, sendo

$$\dot{a}_{t\sigma_x^2(\sigma_x^2)} = \sum_{q=1}^Q \frac{\nu_q}{\sqrt{\pi}} \exp\{\ell_{tq}(\mu_{tq}, \phi_{tq})\} \left[\dot{\ell}_{tq(\sigma_x^2)}(\mu_{tq}, \phi_{tq}) \dot{\ell}_{tq(\sigma_x^2)}(\mu_{tq}, \phi_{tq}) + \ddot{\ell}_{tq(\sigma_x^2)(\sigma_x^2)}(\mu_{tq}, \phi_{tq}) \right] \quad e$$

$$\begin{aligned} \ddot{\ell}_{tq(\sigma_x^2)(\sigma_x^2)}(\mu_{tq}, \phi_{tq}) &= -\phi_{tq} \left[u_{tq} + (y_t^* - \mu_{tq}^*) \frac{g''(\mu_{tq})}{[g'(\mu_{tq})]^3} \right] \frac{\partial \eta_{1t}}{\partial \sigma_x^2} \frac{\partial \eta_{1t}}{\partial \sigma_x^2} \\ &+ 2 \left[(y_t^* - \mu_{tq}^*) - c_{tq} \right] \frac{\lambda \beta}{h'(\phi_{tq}) g'(\mu_{tq})} \frac{\partial \eta_{1t}}{\partial \sigma_x^2} \frac{\partial \eta_{2t}}{\partial \sigma_x^2} \\ &+ \left[d_{tq} - a_{tq} \frac{h''(\phi_{tq})}{[h'(\phi_{tq})]^3} \right] \frac{\partial \eta_{2t}}{\partial \sigma_x^2} \frac{\partial \eta_{1t}}{\partial \sigma_x^2}, \end{aligned}$$

em que $a_{t\sigma^2}$ é apresentado em (16).

APÊNDICE B - RESULTADOS DOS ESTUDOS DE SIMULAÇÃO DOS CAPÍTULOS 2 E 3

Apêndice B.1 - Cenário 2: Avaliação do parâmetro de precisão ϕ

Tabela 1: Limites inferiores médios (LI) e superiores médios (LS) e suas respectivas medianas (Md) para a média da variável resposta (μ_t) considerando o modelo de regressão beta não linear com erros de medida com dispersão constante. Coeficiente de confiabilidade $k_x = 0.75$.

ϕ	n	μ_t		
		LI	LS	Md
16.44	40	0.1363	0.8824	0.5136
	80	0.1090	0.8965	0.4971
	160	0.0877	0.9183	0.4943
54.60	40	0.1348	0.8831	0.5143
	80	0.1075	0.8970	0.4971
	160	0.0881	0.9190	0.4943
298.87	40	0.1354	0.8829	0.5138
	80	0.1081	0.8971	0.4968
	160	0.0878	0.9192	0.4943

Tabela 2: Limites inferiores médios (LI) e superiores médios (LS) e suas respectivas medianas (Md) para a média da variável resposta (μ_t) considerando o modelo de regressão beta não linear com erros de medida com dispersão constante. Coeficiente de confiabilidade $k_x = 0.95$.

ϕ	n	μ_t		
		LI	LS	Md
16.44	40	0.1363	0.8824	0.5136
	80	0.1090	0.8965	0.4971
	160	0.0877	0.9183	0.4943
54.60	40	0.1350	0.8830	0.5142
	80	0.1077	0.8970	0.4970
	160	0.0878	0.9185	0.4943
298.87	40	0.1354	0.8823	0.5131
	80	0.1090	0.8970	0.4968
	160	0.0875	0.9186	0.4939

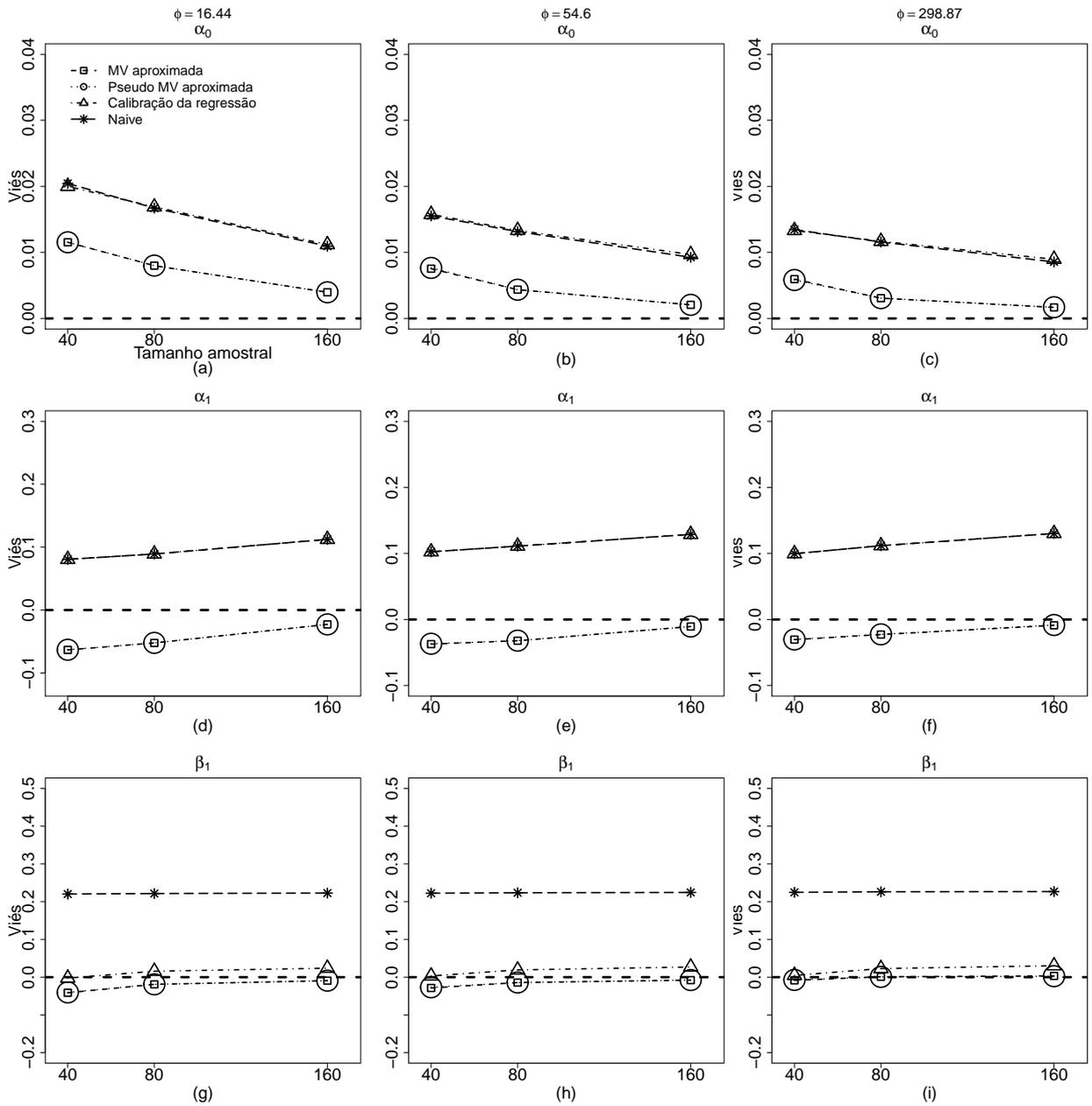


Figura 1: Vieses estimados dos estimadores de α_0 , α_1 e β_1 para $\phi = 16.44$, $\phi = 54.60$ e $\phi = 298.87$ com o coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado). Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 2: Modelo com dispersão constante.

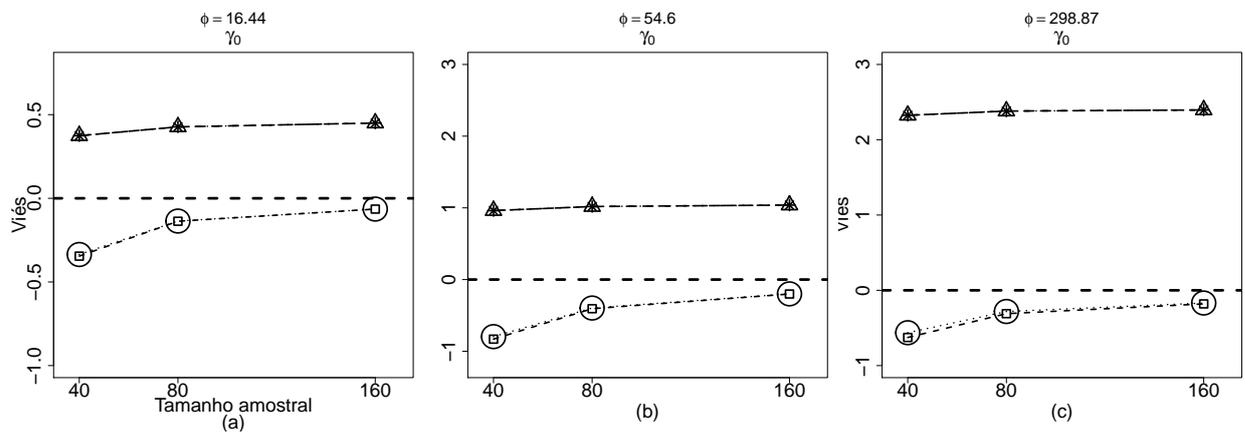


Figura 2: Vieses estimados dos estimadores de γ_0 para $\phi = 16.44$, $\phi = 54.60$ e $\phi = 298.87$ com o coeficiente de confiabilidade $k_x = 0.75$ (erro de medida moderado). Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 2: Modelo com dispersão constante.

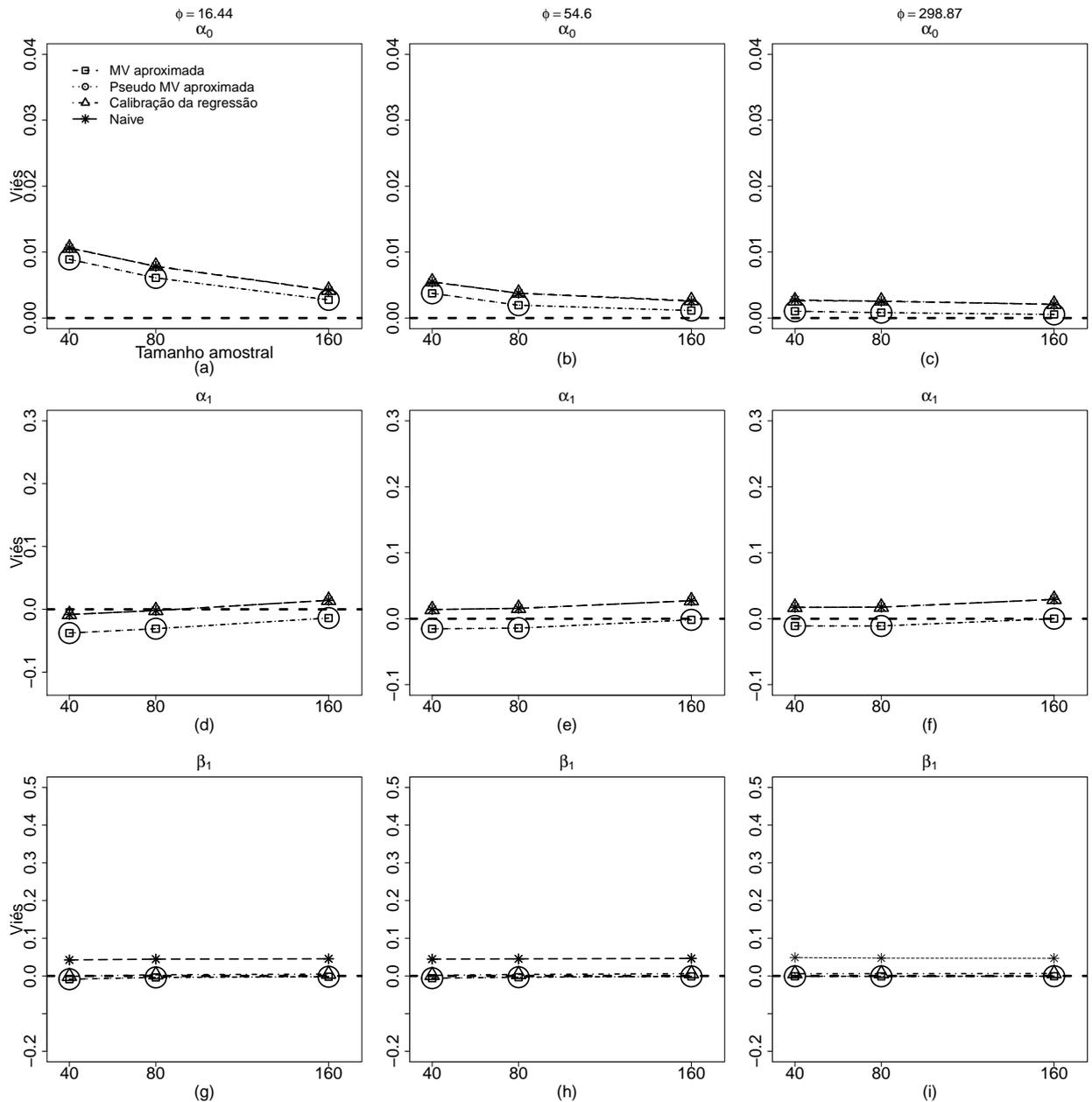


Figura 3: Vieses estimados dos estimadores de α_0 , α_1 e β_1 para $\phi = 16.44$, $\phi = 54.60$ e $\phi = 298.87$ com o coeficiente de confiabilidade $k_x = 0.95$ (erro de medida baixo). Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 2: Modelo com dispersão constante.

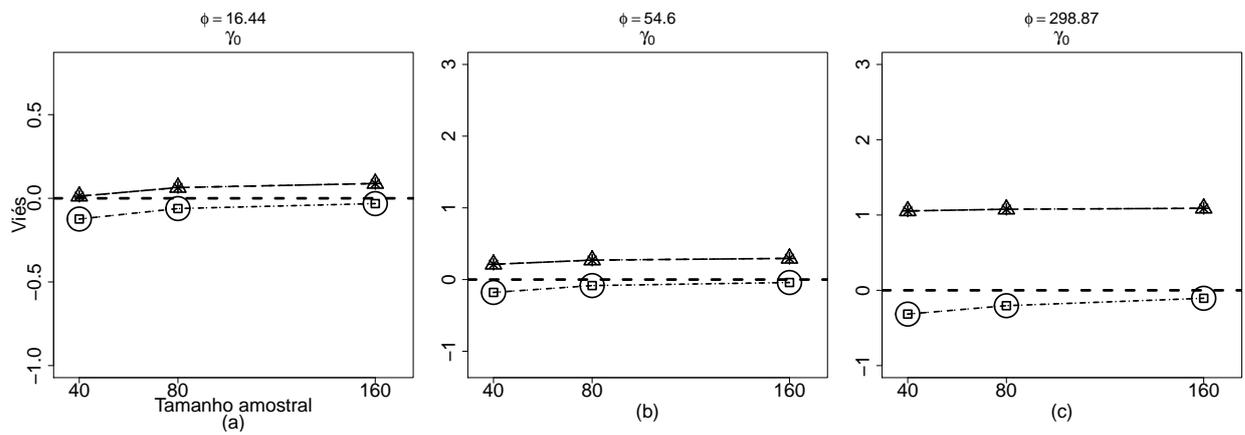


Figura 4: Vieses estimados dos estimadores de γ_0 para $\phi = 16.44$, $\phi = 54.60$ e $\phi = 298.87$ com o coeficiente de confiabilidade $k_x = 0.95$ (erro de medida baixo). Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 2: Modelo com dispersão constante.

Tabela 3: Raízes dos erros quadráticos médios (REQM) para os estimadores de α_0 , α_1 , β_1 e γ_0 considerando o modelo com dispersão constante com o coeficiente de confiabilidade $k_x = 0.75$. Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 2.

ϕ	n		$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_0$
16.44	40	<i>Naive</i>	0.11282	0.82153	0.24027	0.44025
		RC	0.12339	0.82153	0.16066	0.44025
		MVa	0.12662	0.86048	0.18623	0.86490
		MPVa	0.12626	0.86030	0.18254	0.83099
	80	<i>Naive</i>	0.08799	0.69056	0.23136	0.45550
		RC	0.09259	0.69056	0.10344	0.45550
		MVa	0.09299	0.71664	0.11725	0.39100
		MPVa	0.09298	0.71656	0.11658	0.38459
	160	<i>Naive</i>	0.05848	0.45680	0.22752	0.46336
		RC	0.06135	0.45680	0.07365	0.46336
		MVa	0.06152	0.46184	0.07857	0.23006
		MPVa	0.06151	0.46183	0.07849	0.22975
54.60	40	<i>Naive</i>	0.08516	0.63061	0.23372	0.98917
		RC	0.09603	0.63061	0.12979	0.98917
		MVa	0.09834	0.65179	0.13896	1.62137
		MPVa	0.09788	0.65086	0.13305	1.55799
	80	<i>Naive</i>	0.06615	0.53462	0.22888	1.03099
		RC	0.07204	0.53462	0.08413	1.03099
		MVa	0.07256	0.54799	0.08908	1.01159
		MPVa	0.07248	0.54770	0.08814	0.98959
	160	<i>Naive</i>	0.04419	0.36442	0.22708	1.04472
		RC	0.04793	0.36442	0.06143	1.04472
		MVa	0.04810	0.35609	0.06231	0.63719
		MPVa	0.04807	0.35605	0.06165	0.62982
298.87	40	<i>Naive</i>	0.07092	0.08357	0.08600	0.08558
		RC	0.53792	0.53792	0.56829	0.56674
		MVa	0.23287	0.12007	0.11116	0.10585
		MPVa	2.33637	2.33637	1.54866	1.47917
	80	<i>Naive</i>	0.05631	0.45635	0.22995	2.38605
		RC	0.06300	0.45635	0.07761	2.38605
		MVa	0.06446	0.47834	0.06999	1.19798
		MPVa	0.06431	0.47681	0.06831	1.16410
	160	<i>Naive</i>	0.03763	0.31477	0.22859	2.39808
		RC	0.04163	0.31477	0.05819	2.39808
		MVa	0.04271	0.31192	0.04791	0.93463
		MPVa	0.04262	0.31131	0.04814	0.92032

Tabela 4: Raízes dos erros quadráticos médios (REQM) para os estimadores de α_0 , α_1 , β_1 e γ_0 considerando o modelo com dispersão constante com o coeficiente de confiabilidade $k_x = 0.95$. Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 5.0$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 2.

ϕ	n		$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_0$
16.44	40	<i>Naive</i>	0.09452	0.71185	0.10515	0.23190
		RC	0.09510	0.71185	0.10274	0.23190
		MVa	0.09551	0.71986	0.10599	0.29819
		MPVa	0.09551	0.71986	0.10599	0.29819
	80	<i>Naive</i>	0.07325	0.59183	0.08010	0.17217
		RC	0.07351	0.59183	0.07069	0.17217
		MVa	0.07357	0.59785	0.07256	0.19405
		MPVa	0.07357	0.59785	0.07256	0.19405
	160	<i>Naive</i>	0.04849	0.38055	0.06507	0.14112
		RC	0.04860	0.38055	0.04958	0.14112
		MVa	0.04868	0.38393	0.05068	0.12946
		MPVa	0.04868	0.38393	0.05068	0.12946
54.60	40	<i>Naive</i>	0.05878	0.45462	0.07496	0.31493
		RC	0.05939	0.45462	0.06499	0.31493
		MVa	0.05951	0.45803	0.06685	0.40332
		MPVa	0.05951	0.45803	0.06684	0.40330
	80	<i>Naive</i>	0.04524	0.37535	0.06113	0.31511
		RC	0.04560	0.37535	0.04416	0.31511
		MVa	0.04557	0.37782	0.04516	0.25022
		MPVa	0.04557	0.37782	0.04516	0.25022
	160	<i>Naive</i>	0.03019	0.24377	0.05463	0.31617
		RC	0.03038	0.24377	0.03122	0.31617
		MVa	0.03035	0.24398	0.03138	0.16635
		MPVa	0.03035	0.24398	0.03138	0.16635
298.87	40	<i>Naive</i>	0.03702	0.28716	0.06130	1.07419
		RC	0.03770	0.28716	0.04112	1.07419
		MVa	0.03767	0.28723	0.04096	0.82174
		MPVa	0.03767	0.28723	0.04097	0.82134
	80	<i>Naive</i>	0.02873	0.24350	0.05361	1.08723
		RC	0.02909	0.24350	0.02831	1.08723
		MVa	0.02874	0.24287	0.02789	0.58904
		MPVa	0.02874	0.24287	0.02789	0.58896
	160	<i>Naive</i>	0.01941	0.15606	0.05025	1.09494
		RC	0.01965	0.15606	0.02082	1.09494
		MVa	0.01953	0.15368	0.01996	0.38759
		MPVa	0.01953	0.15368	0.01996	0.38756

Apêndice B.2 - Cenário 4: Avaliação dos pontos de quadratura “Q”

Tabela 5: Limites inferiores médios (LI) e superiores médios (LS) e suas respectivas medianas (Md_{μ_t}) para a média da variável resposta (μ_t) e médias dos mínimos (Mín), máximos (Máx) e medianas (Md_{ϕ_t}) de ϕ_t e seus respectivos graus de heterogeneidade (δ) para o modelo de regressão beta não linear com erros de medida com dispersão variável. Coeficiente de confiabilidade $k_x = 0.75$.

Q	n	μ_t			ϕ_t			
		LI	LS	Md_{μ_t}	Mín	Máx	Md_{ϕ_t}	δ
12	40	0.1499	0.8829	0.5152	2.1152	93.2300	12.5038	50.5855
	80	0.1245	0.8970	0.4989	1.7204	117.7495	12.4217	76.0749
	160	0.1041	0.9188	0.4962	1.4163	145.6003	12.3718	112.9395
30	40	0.1497	0.8828	0.5152	2.1135	92.9549	12.5035	50.4677
	80	0.1245	0.8969	0.4989	1.7202	117.7619	12.4216	76.0982
	160	0.1041	0.9188	0.4962	1.4162	145.6183	12.3721	112.9577
80	40	0.1498	0.8829	0.5151	2.1139	93.2379	12.5007	50.5286
	80	0.1245	0.8969	0.4989	1.7200	117.6542	12.4213	76.0498
	160	0.1041	0.9188	0.4962	1.4161	145.5574	12.3720	112.9185

Tabela 6: Limites inferiores médios (LI) e superiores médios (LS) e suas respectivas medianas (Md_{μ_t}) para a média da variável resposta (μ_t) e médias dos mínimos (Mín), máximos (Máx) e medianas (Md_{ϕ_t}) de ϕ_t e seus respectivos graus de heterogeneidade (δ) para o modelo de regressão beta não linear com erros de medida com dispersão variável. Coeficiente de confiabilidade $k_x = 0.95$.

Q	n	μ_t			ϕ_t			
		LI	LS	Md_{μ_t}	Mín	Máx	Md_{ϕ_t}	δ
12	40	0.1494	0.8826	0.5159	2.1197	94.1837	12.5145	50.4305
	80	0.1245	0.8972	0.4988	1.7196	118.1482	12.4231	76.4595
	160	0.1041	0.9188	0.4962	1.4163	145.6003	12.3718	112.9395
30	40	0.1494	0.8826	0.5159	2.1197	94.2031	12.5147	50.4424
	80	0.1245	0.8972	0.4988	1.7196	118.1482	12.4231	76.4595
	160	0.1041	0.9188	0.4962	1.4163	145.6003	12.3718	112.9395
80	40	0.1494	0.1494	0.5159	2.1198	94.2291	12.5144	50.4504
	80	0.1245	0.8972	0.4988	1.7196	118.1482	12.4231	76.4595
	160	0.1041	0.9188	0.4962	1.4163	145.6003	12.3718	112.9395

Tabela 7: Raiz do erro quadrático médio (REQM) para os estimadores de α_0 , α_1 , β_1 , γ_0 e λ_1 considerando o modelo de regressão beta não linear com dispersão variável. Coeficiente de confiabilidade $k_x = 0.75$. Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 2.5$, $\lambda_1 = 0.9$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 4.

Q	n		$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_0$	$\hat{\lambda}_1$
12	40	<i>Naive</i>	0.13419	0.89080	0.25209	0.42066	0.47427
		RC	0.14400	0.89080	0.18142	0.42894	0.42535
		MVa	0.14517	0.92270	0.20072	0.79540	0.66746
		MPVa	0.14394	0.92025	0.18903	0.76422	0.61780
	80	<i>Naive</i>	0.10397	0.74461	0.23951	0.45416	0.45453
		RC	0.10954	0.74461	0.11578	0.45881	0.33523
		MVa	0.10416	0.74297	0.12193	0.41252	0.35611
		MPVa	0.10410	0.74238	0.12083	0.40709	0.34851
	160	<i>Naive</i>	0.07056	0.49993	0.23473	0.46617	0.44805
		RC	0.07498	0.49993	0.08025	0.47033	0.30523
		MVa	0.07035	0.48460	0.08063	0.22068	0.20770
		MPVa	0.07036	0.48460	0.08023	0.21913	0.20551
30	40	<i>Naive</i>	0.13402	0.88706	0.25347	0.42286	0.47559
		RC	0.14350	0.88706	0.17995	0.43101	0.42583
		MVa	0.14207	0.90478	0.18865	0.75993	0.59748
		MPVa	0.14153	0.90286	0.18373	0.72755	0.56121
	80	<i>Naive</i>	0.10400	0.74577	0.23965	0.45427	0.45481
		RC	0.10959	0.74577	0.11556	0.45892	0.33542
		MVa	0.10382	0.73867	0.11968	0.40333	0.34081
		MPVa	0.10378	0.73851	0.11871	0.39718	0.33109
	160	<i>Naive</i>	0.07056	0.50005	0.23473	0.46616	0.44802
		RC	0.07496	0.50005	0.08025	0.47032	0.30519
		MVa	0.07035	0.48418	0.08070	0.21652	0.20268
		MPVa	0.07036	0.48417	0.08057	0.21601	0.20215
80	40	<i>Naive</i>	0.13396	0.88741	0.25408	0.42418	0.47566
		RC	0.14294	0.88741	0.17705	0.43221	0.42453
		MVa	0.14034	0.89364	0.17675	0.66044	0.48161
		MPVa	0.14038	0.89351	0.17558	0.64054	0.46915
	80	<i>Naive</i>	0.10400	0.74481	0.23983	0.45464	0.45485
		RC	0.10960	0.74481	0.11534	0.45929	0.33539
		MVa	0.10349	0.73661	0.11507	0.35626	0.29677
		MPVa	0.10350	0.73666	0.11527	0.35300	0.29505
	160	<i>Naive</i>	0.07056	0.49980	0.23474	0.46619	0.44800
		RC	0.07497	0.49980	0.08024	0.47035	0.30517
		MVa	0.07033	0.48322	0.07854	0.20609	0.19442
		MPVa	0.07034	0.48315	0.07877	0.20689	0.19483

Tabela 8: Raízes dos erros quadráticos médios (REQM) para os estimadores de α_0 , α_1 , β_1 , γ_0 e λ_1 considerando o modelo de regressão beta não linear com dispersão variável. Coeficiente de confiabilidade $k_x = 0.95$. Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\gamma_0 = 2.5$, $\lambda_1 = 0.9$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 4.

Q	n		$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\gamma}_0$	$\hat{\lambda}_1$
12	40	<i>Naive</i>	0.11384	0.78233	0.11805	0.24057	0.28912
		RC	0.11425	0.78233	0.11329	0.24122	0.29383
		MVa	0.11423	0.78660	0.11430	0.33300	0.31699
		MPVa	0.11423	0.78660	0.11430	0.33299	0.31695
	80	<i>Naive</i>	0.08554	0.63553	0.08659	0.16965	0.21617
		RC	0.08585	0.63553	0.07410	0.17011	0.20151
		MVa	0.08556	0.63761	0.07450	0.20468	0.19815
		MPVa	0.08556	0.63761	0.07450	0.20468	0.19815
	160	<i>Naive</i>	0.05815	0.40976	0.07239	0.13680	0.18564
		RC	0.05834	0.40976	0.05290	0.13742	0.15914
		MVa	0.05856	0.41187	0.05311	0.13433	0.13798
		MPVa	0.05856	0.41187	0.05311	0.13433	0.13798
30	40	<i>Naive</i>	0.11383	0.78223	0.11810	0.24059	0.28938
		RC	0.11424	0.78223	0.11339	0.24125	0.29417
		MVa	0.11418	0.78623	0.11436	0.32922	0.31078
		MPVa	0.11418	0.78623	0.11436	0.32921	0.31079
	80	<i>Naive</i>	0.08554	0.63553	0.08659	0.16965	0.21617
		RC	0.08585	0.63553	0.07410	0.17011	0.20151
		MVa	0.08556	0.63761	0.07450	0.20462	0.19807
		MPVa	0.08556	0.63761	0.07450	0.20462	0.19807
	160	<i>Naive</i>	0.05815	0.40976	0.07239	0.13680	0.18564
		RC	0.05834	0.40976	0.05290	0.13742	0.15914
		MVa	0.05856	0.41187	0.05311	0.13433	0.13798
		MPVa	0.05856	0.41187	0.05311	0.13433	0.13798
80	40	<i>Naive</i>	0.11389	0.78223	0.11810	0.24075	0.28941
		RC	0.11429	0.78223	0.11341	0.24140	0.29429
		MVa	0.11424	0.78614	0.11437	0.32887	0.31098
		MPVa	0.11424	0.78613	0.11437	0.32887	0.31100
	80	<i>Naive</i>	0.08554	0.63553	0.08659	0.16965	0.21617
		RC	0.08585	0.63553	0.07410	0.17011	0.20151
		MVa	0.08555	0.63758	0.07449	0.20405	0.19803
		MPVa	0.08555	0.63758	0.07449	0.20405	0.19804
	160	<i>Naive</i>	0.05815	0.40976	0.07239	0.13680	0.18564
		RC	0.05834	0.40976	0.05290	0.13742	0.15914
		MVa	0.05856	0.41184	0.05311	0.13387	0.13803
		MPVa	0.05856	0.41184	0.05311	0.13387	0.13803

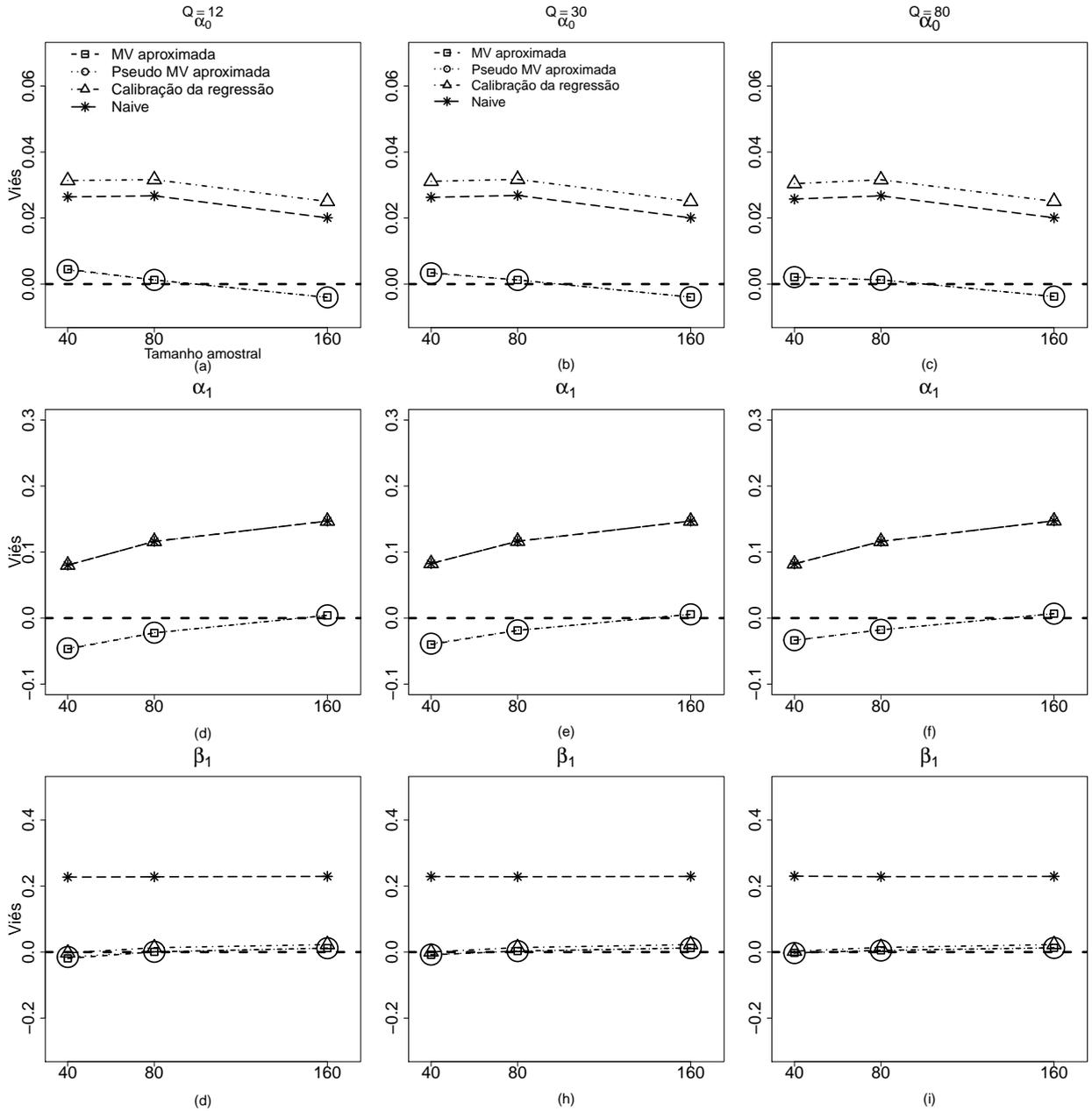


Figura 5: Vieses estimados dos estimadores de α_0 , α_1 e β_1 para $Q = 12$ (coluna esquerda), $Q = 30$ (coluna central) e $Q = 80$ (coluna direita) com o coeficiente de confiabilidade $k_x = 0.75$. Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 4: Modelo com dispersão variável.

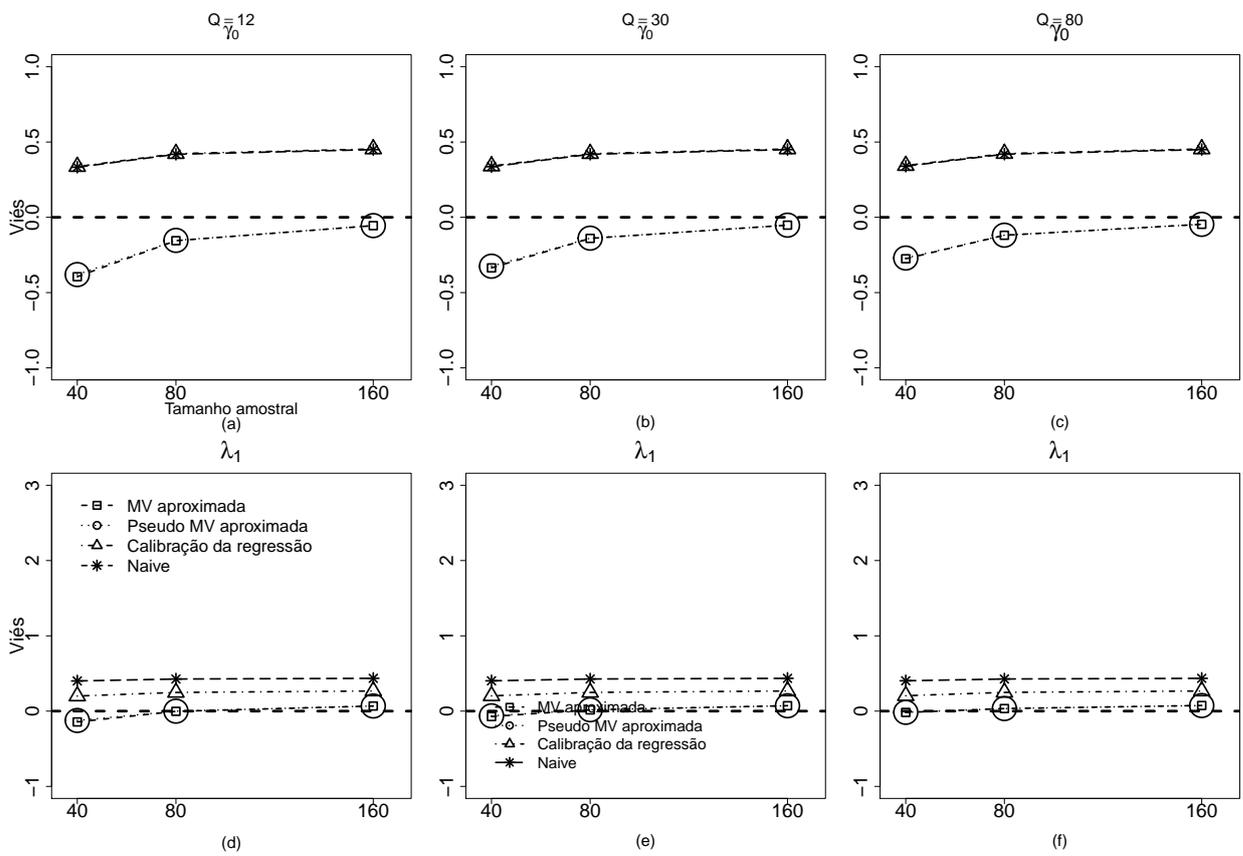


Figura 6: Vieses estimados dos estimadores de γ_0 e λ_1 para $Q = 12$ (coluna esquerda), $Q = 30$ (coluna central) e $Q = 80$ (coluna direita) com o coeficiente de confiabilidade $k_x = 0.75$. Verdadeiros valores dos parâmetros: $\gamma_0 = 2.5$, $\lambda_1 = 0.9$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 4: Modelo com dispersão variável.

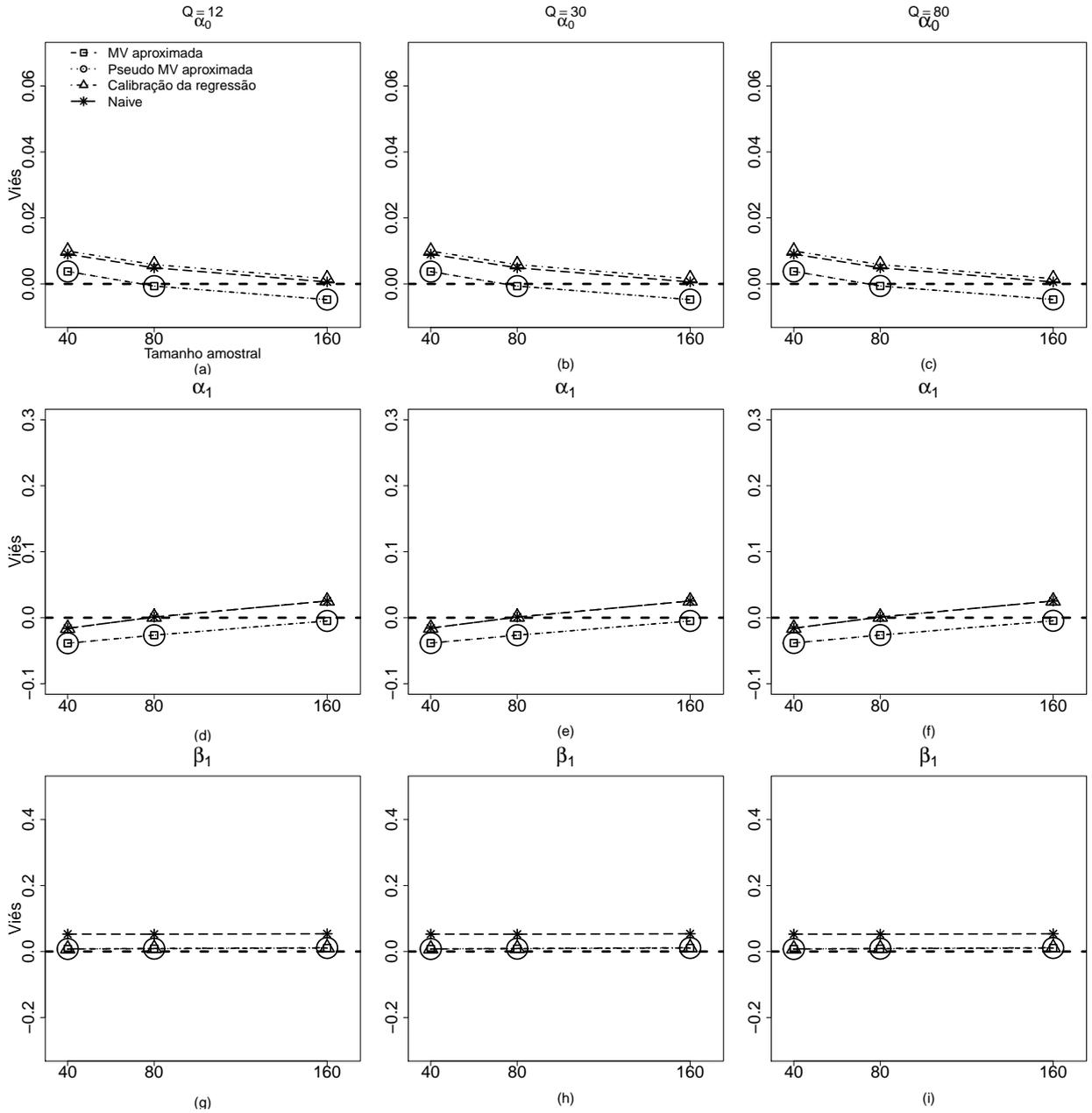


Figura 7: Vieses estimados dos estimadores de α_0 , α_1 e β_1 para $Q = 12$ (coluna esquerda), $Q = 30$ (coluna central) e $Q = 80$ (coluna direita) com o coeficiente de confiabilidade $k_x = 0.95$. Verdadeiros valores dos parâmetros: $\alpha_0 = -0.6$, $\alpha_1 = 2.4$, $\beta_1 = 0.8$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 4: Modelo com dispersão variável.

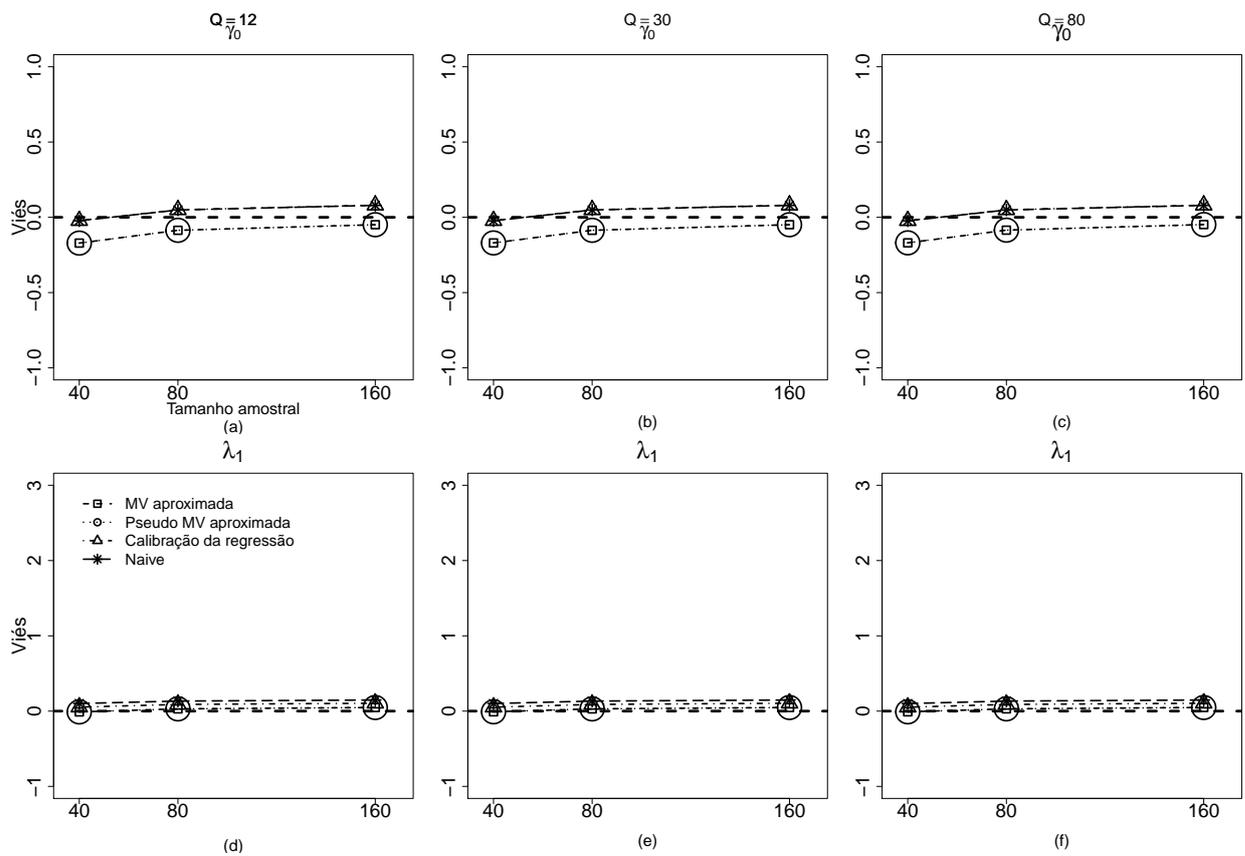


Figura 8: Vieses estimados dos estimadores de γ_0 e λ_1 para $Q = 12$ (coluna esquerda), $Q = 30$ (coluna central) e $Q = 80$ (coluna direita) com o coeficiente de confiabilidade $k_x = 0.95$. Verdadeiros valores dos parâmetros: $\gamma_0 = 2.5$, $\lambda_1 = 0.9$, $\mu_x = 0.0$ e $\sigma_x^2 = 1.0$. Cenário 4: Modelo com dispersão variável.

Tabela 9: Taxas de cobertura e comprimentos médios dos intervalos de 99% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão constante, utilizando os estimadores obtidos pelos métodos *naive* (*Naive*), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.95$. Cenário 1.

k_x	$\hat{\theta}$	Intervalo	$n = 40$						$n = 80$						$n = 160$							
			$Md_{\mu_t} = 0.5135$			$Md_{\mu_t} = 0.4968$			$Md_{\mu_t} = 0.4942$			$Md_{\mu_t} = 0.4968$			$Md_{\mu_t} = 0.4942$			$Md_{\mu_t} = 0.4942$				
			inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp
0.95	α_0	<i>Naive</i>	0.69	95.79	0.97	0.21	0.71	96.22	0.61	0.17	0.52	96.41	0.60	0.11	0.71	96.22	0.61	0.17	0.52	96.41	0.60	0.11
		RC	0.95	97.75	1.30	0.22	1.24	97.56	1.20	0.17	0.91	97.96	1.13	0.11	1.24	97.56	1.20	0.17	0.91	97.96	1.13	0.11
		MVa	0.74	98.41	0.85	0.22	0.75	98.73	0.52	0.17	0.62	98.88	0.50	0.12	0.75	98.73	0.52	0.17	0.62	98.88	0.50	0.12
		MPVa	0.88	98.12	1.00	0.21	0.97	98.17	0.86	0.17	0.67	98.76	0.57	0.11	0.97	98.17	0.86	0.17	0.67	98.76	0.57	0.11
	α_1	<i>Naive</i>	0.72	94.56	1.63	1.64	0.48	95.15	1.49	1.39	0.28	95.22	1.18	0.91	0.48	95.15	1.49	1.39	0.28	95.22	1.18	0.91
		RC	1.13	96.69	2.18	1.69	0.87	97.06	2.07	1.39	0.57	97.73	1.70	0.91	0.87	97.06	2.07	1.39	0.57	97.73	1.70	0.91
		MVa	1.03	97.62	1.35	1.64	0.67	98.30	1.03	1.39	0.52	98.75	0.73	0.91	0.67	98.30	1.03	1.39	0.52	98.75	0.73	0.91
		MPVa	1.84	96.15	2.01	1.58	1.29	97.20	1.51	1.34	0.59	98.49	0.92	0.90	1.29	97.20	1.51	1.34	0.59	98.49	0.92	0.90
	β_1	<i>Naive</i>	0.04	79.96	7.75	0.22	0.00	66.35	15.11	0.16	0.00	42.03	34.13	0.11	0.00	66.35	15.11	0.16	0.00	42.03	34.13	0.11
		RC	0.94	97.21	1.85	0.23	0.75	97.13	2.12	0.16	0.37	97.70	1.93	0.11	0.75	97.13	2.12	0.16	0.37	97.70	1.93	0.11
		MVa	0.84	98.42	0.74	0.24	0.86	98.42	0.72	0.17	0.53	98.97	0.50	0.12	0.86	98.42	0.72	0.17	0.53	98.97	0.50	0.12
		MPVa	1.36	97.66	0.98	0.23	1.03	98.14	0.83	0.16	0.63	98.77	0.60	0.12	1.03	98.14	0.83	0.16	0.63	98.77	0.60	0.12
γ_0	<i>Naive</i>	0.00	22.46	55.85	1.15	0.00	1.98	93.34	0.81	0.00	0.02	99.95	0.57	0.00	1.98	93.34	0.81	0.00	0.02	99.95	0.57	
	RC	0.00	47.43	52.57	1.19	0.00	8.91	91.09	0.82	0.00	0.18	99.82	0.57	0.00	8.91	91.09	0.82	0.00	0.18	99.82	0.57	
	MVa	0.32	99.36	0.32	3.09	0.00	99.60	0.40	1.89	0.01	99.47	0.52	1.23	0.00	99.60	0.40	1.89	0.01	99.47	0.52	1.23	
	MPVa	0.71	98.72	0.57	2.94	0.04	99.50	0.46	1.85	0.16	99.17	0.67	1.22	0.04	99.50	0.46	1.85	0.16	99.17	0.67	1.22	

Tabela 10: Taxas de cobertura e comprimentos médios dos intervalos de 99% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão constante, utilizando os estimadores obtidos pelos métodos *naive* (*Naive*), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.75$. Cenário 1.

k_x	$\hat{\theta}$	Intervalo	$n = 40$						$n = 80$						$n = 160$					
			$Md_{\mu_t} = 0.5136$			$Md_{\mu_t} = 0.4967$			$Md_{\mu_t} = 0.4944$			$Md_{\mu_t} = 0.4967$			$Md_{\mu_t} = 0.4944$					
			inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp		
	α_0	<i>Naive</i>	0.72	95.12	1.13	0.36	0.51	95.70	0.74	0.28	0.28	95.74	0.97	0.20	0.20	95.74	0.97	0.20		
		RC	1.61	95.78	2.61	0.38	1.46	96.35	2.19	0.30	0.30	97.01	2.10	0.20	0.20	97.01	2.10	0.20		
		MVa	4.65	90.49	4.86	0.37	3.35	92.93	3.72	0.29	0.29	96.56	1.67	0.21	0.21	96.56	1.67	0.21		
		MPVa	2.85	94.04	3.11	0.61	2.38	94.80	2.82	0.35	0.35	97.41	1.31	0.22	0.22	97.41	1.31	0.22		
0.75	α_1	<i>Naive</i>	0.49	92.27	3.10	2.73	0.28	91.83	3.10	2.31	2.31	91.51	3.07	1.51	1.51	91.51	3.07	1.51		
		RC	0.84	95.46	3.70	2.81	0.57	95.80	3.63	2.33	2.33	96.15	3.69	1.51	1.51	96.15	3.69	1.51		
		MVa	13.87	73.04	13.09	2.03	7.80	84.23	7.97	1.98	1.98	92.93	3.83	1.44	1.44	92.93	3.83	1.44		
		MPVa	10.63	79.10	10.27	2.92	6.95	86.00	7.05	2.21	2.21	93.69	3.39	1.52	1.52	93.69	3.39	1.52		
	β_1	<i>Naive</i>	0.00	6.73	82.26	0.32	0.00	0.14	98.96	0.23	0.00	0.00	100.00	0.16	0.16	0.00	0.00	100.00		
		RC	3.08	90.32	6.60	0.39	0.94	92.07	6.99	0.30	0.30	94.10	5.82	0.24	0.24	94.10	5.82	0.24		
		MVa	11.80	85.30	2.90	0.44	5.78	92.70	1.52	0.34	0.34	95.82	0.87	0.25	0.25	95.82	0.87	0.25		
		MPVa	2.14	95.92	1.94	0.80	2.27	96.37	1.36	0.43	0.43	97.28	0.90	0.27	0.27	97.28	0.90	0.27		
	γ_0	<i>Naive</i>	0.00	0.00	100.00	1.13	0.00	0.00	100.00	0.80	0.00	0.00	100.00	0.57	0.57	0.00	0.00	100.00		
		RC	0.00	0.00	100.00	1.18	0.00	0.00	100.00	0.81	0.00	0.00	100.00	0.57	0.57	0.00	0.00	100.00		
		MVa	38.05	59.58	2.37	4.74	26.42	71.36	2.22	4.24	4.24	81.35	2.28	3.51	3.51	81.35	2.28	3.51		
		MPVa	35.18	61.73	3.09	8.71	26.33	71.11	2.56	5.33	5.33	80.19	2.49	3.90	3.90	80.19	2.49	3.90		

Tabela 11: Taxas de cobertura e comprimentos médios dos intervalos de 99% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão constante, utilizando os estimadores obtidos pelos métodos *naive* (*Naive*), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.95$. Cenário 1.

k_x	$\hat{\theta}$	Intervalo	$n = 40$						$n = 80$						$n = 160$																																				
			$Md_{\mu_t} = 0.5135$			$Md_{\mu_t} = 0.4969$			$Md_{\mu_t} = 0.4941$			$Md_{\mu_t} = 0.4969$			$Md_{\mu_t} = 0.4941$			$Md_{\mu_t} = 0.4941$																																	
			inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp																													
	α_0	<i>Naive</i>	0.43	94.75	1.27	0.48	0.29	95.42	1.00	0.39	0.20	94.89	1.11	0.26	4.52	88.66	6.82	0.53	2.70	92.27	5.03	0.40	1.72	93.88	4.40	0.27	2.22	94.95	2.83	0.65	2.33	94.99	2.68	0.46	1.70	96.12	2.18	0.33	0.67	98.26	1.07	8.92	1.03	97.48	1.49	1.18	0.75	97.94	1.31	0.51	
		RC	0.42	88.84	5.65	3.46	0.18	88.43	5.48	2.94	0.02	86.83	5.46	1.94	0.79	93.94	5.27	3.57	0.44	93.93	5.63	2.99	0.10	93.43	6.47	1.95	20.39	60.09	19.52	2.09	14.08	71.78	14.14	2.11	8.49	82.59	8.92	1.63	10.02	79.88	10.10	11.82	8.15	83.77	8.08	5.38	5.10	89.34	5.56	2.48	
		MVa	0.00	0.01	99.99	0.33	0.00	0.00	100.00	0.24	0.00	0.00	100.00	0.17	0.17	12.12	70.43	17.45	0.69	5.35	78.14	16.51	0.48	1.00	86.10	12.90	0.41	19.89	71.40	8.71	0.65	11.87	82.91	5.22	0.52	6.15	91.13	2.72	0.41	0.08	96.73	3.19	34.13	0.24	97.47	2.29	1.96	0.25	98.27	1.48	0.75
		MPVa	0.00	0.00	100.00	1.12	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	0.56	0.00	0.00	100.00	1.16	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	54.79	39.13	6.08	4.09	46.27	48.53	5.20	3.87	38.52	56.59	4.89	3.42	34.15	58.83	7.02	57.33	32.27	62.05	5.68	15.01	28.08	66.67	5.25	7.10
0.50	α_1	<i>Naive</i>	0.00	0.01	99.99	0.33	0.00	0.00	100.00	0.24	0.00	0.00	100.00	0.17	12.12	70.43	17.45	0.69	5.35	78.14	16.51	0.48	1.00	86.10	12.90	0.41	19.89	71.40	8.71	0.65	11.87	82.91	5.22	0.52	6.15	91.13	2.72	0.41	0.08	96.73	3.19	34.13	0.24	97.47	2.29	1.96	0.25	98.27	1.48	0.75	
		RC	0.00	0.00	100.00	1.12	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	0.00	0.00	100.00	1.16	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	54.79	39.13	6.08	4.09	46.27	48.53	5.20	3.87	38.52	56.59	4.89	3.42	34.15	58.83	7.02	57.33	32.27	62.05	5.68	15.01	28.08	66.67	5.25	7.10	
		MVa	0.00	0.00	100.00	1.12	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	0.00	0.00	100.00	1.16	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	54.79	39.13	6.08	4.09	46.27	48.53	5.20	3.87	38.52	56.59	4.89	3.42	34.15	58.83	7.02	57.33	32.27	62.05	5.68	15.01	28.08	66.67	5.25	7.10	
		MPVa	0.00	0.00	100.00	1.12	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	0.00	0.00	100.00	1.16	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	54.79	39.13	6.08	4.09	46.27	48.53	5.20	3.87	38.52	56.59	4.89	3.42	34.15	58.83	7.02	57.33	32.27	62.05	5.68	15.01	28.08	66.67	5.25	7.10	
	β_1	<i>Naive</i>	0.00	0.00	100.00	1.12	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	0.00	0.00	100.00	1.16	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	54.79	39.13	6.08	4.09	46.27	48.53	5.20	3.87	38.52	56.59	4.89	3.42	34.15	58.83	7.02	57.33	32.27	62.05	5.68	15.01	28.08	66.67	5.25	7.10	
		RC	0.00	0.00	100.00	1.12	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	0.00	0.00	100.00	1.16	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	54.79	39.13	6.08	4.09	46.27	48.53	5.20	3.87	38.52	56.59	4.89	3.42	34.15	58.83	7.02	57.33	32.27	62.05	5.68	15.01	28.08	66.67	5.25	7.10	
		MVa	0.00	0.00	100.00	1.12	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	0.00	0.00	100.00	1.16	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	54.79	39.13	6.08	4.09	46.27	48.53	5.20	3.87	38.52	56.59	4.89	3.42	34.15	58.83	7.02	57.33	32.27	62.05	5.68	15.01	28.08	66.67	5.25	7.10	
		MPVa	0.00	0.00	100.00	1.12	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	0.00	0.00	100.00	1.16	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	54.79	39.13	6.08	4.09	46.27	48.53	5.20	3.87	38.52	56.59	4.89	3.42	34.15	58.83	7.02	57.33	32.27	62.05	5.68	15.01	28.08	66.67	5.25	7.10	
	γ_0	<i>Naive</i>	0.00	0.00	100.00	1.12	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	0.00	0.00	100.00	1.16	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	54.79	39.13	6.08	4.09	46.27	48.53	5.20	3.87	38.52	56.59	4.89	3.42	34.15	58.83	7.02	57.33	32.27	62.05	5.68	15.01	28.08	66.67	5.25	7.10	
		RC	0.00	0.00	100.00	1.12	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	0.00	0.00	100.00	1.16	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	54.79	39.13	6.08	4.09	46.27	48.53	5.20	3.87	38.52	56.59	4.89	3.42	34.15	58.83	7.02	57.33	32.27	62.05	5.68	15.01	28.08	66.67	5.25	7.10	
		MVa	0.00	0.00	100.00	1.12	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	0.00	0.00	100.00	1.16	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	54.79	39.13	6.08	4.09	46.27	48.53	5.20	3.87	38.52	56.59	4.89	3.42	34.15	58.83	7.02	57.33	32.27	62.05	5.68	15.01	28.08	66.67	5.25	7.10	
		MPVa	0.00	0.00	100.00	1.12	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	0.00	0.00	100.00	1.16	0.00	0.00	100.00	0.79	0.00	0.00	100.00	0.56	54.79	39.13	6.08	4.09	46.27	48.53	5.20	3.87	38.52	56.59	4.89	3.42	34.15	58.83	7.02	57.33	32.27	62.05	5.68	15.01	28.08	66.67	5.25	7.10	

Tabela 12: Taxas de cobertura e comprimentos médios dos intervalos de 99% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão variável, utilizando os estimadores obtidos pelos métodos *naive* (*Naive*), calibração da regressão (RC), máxima verossimilhança aproximada (MPVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0,95$. Cenário 2.

k_x	$\hat{\theta}$	Intervalo	$n = 40$				$n = 80$				$n = 160$			
			$Md_{\mu_t} = 0, 5156$		$Md_{\mu_t} = 0, 4989$		$Md_{\mu_t} = 0, 4989$		$Md_{\mu_t} = 0, 4962$		$\phi_t \in [1, 4157; 145, 5675]$		$\phi_t \in [1, 4157; 145, 5675]$	
			inf	sup	comp	inf	sup	comp	inf	sup	comp	inf	sup	comp
0,95	α_0	<i>Naive</i>	0,90	95,50	1,20	0,55	0,72	96,32	0,69	0,43	0,68	96,70	0,51	0,29
		RC	0,82	98,19	0,99	0,60	0,91	98,08	1,01	0,45	0,88	98,27	0,85	0,30
		MVa	1,06	97,90	1,04	0,55	0,84	98,53	0,63	0,43	0,88	98,75	0,37	0,29
		MPVa	1,11	97,77	1,12	0,56	0,94	98,32	0,74	0,42	0,88	98,63	0,49	0,29
	α_1	<i>Naive</i>	0,87	91,28	4,21	3,66	0,52	93,26	3,18	3,15	0,30	95,02	1,81	2,08
		RC	1,08	95,82	3,10	4,01	1,10	95,96	2,94	3,25	0,71	97,31	1,98	2,08
		MVa	1,09	95,02	3,89	3,65	0,61	96,49	2,90	3,13	0,38	98,16	1,46	2,07
		MPVa	1,93	93,31	4,76	3,55	1,26	95,29	3,45	3,06	0,68	97,63	1,69	2,05
	β_1	<i>Naive</i>	0,26	89,05	4,09	0,50	0,05	86,18	5,04	0,34	0,03	77,94	9,11	0,24
		RC	0,95	97,84	1,21	0,58	0,90	97,33	1,77	0,36	0,67	97,11	2,22	0,24
		MVa	0,65	97,83	1,52	0,54	0,41	98,47	1,12	0,37	0,38	98,30	1,32	0,26
		MVPa	1,22	96,42	2,36	0,52	0,55	97,59	1,86	0,36	0,45	97,76	1,79	0,25
γ_0	<i>Naive</i>	1,83	95,66	0,39	1,12	0,40	93,91	1,19	0,78	0,12	87,66	3,69	0,55	
	RC	1,81	97,67	0,52	1,26	0,64	97,71	1,65	0,82	0,16	96,04	3,80	0,57	
	MVa	2,61	97,30	0,09	1,35	1,86	98,00	0,14	0,92	1,35	98,51	0,14	0,64	
	MVpa	5,13	94,68	0,19	1,26	2,94	96,83	0,23	0,89	1,79	98,05	0,16	0,63	
λ_1	<i>Naive</i>	0,39	91,76	2,45	1,26	0,09	85,56	4,83	0,83	0,00	72,02	11,92	0,57	
	RC	0,49	98,55	0,96	1,67	0,30	96,25	3,45	0,91	0,11	91,76	8,13	0,58	
	MVa	0,30	98,90	0,80	1,54	0,13	98,99	0,88	1,00	0,11	98,29	1,60	0,67	
	MPVa	1,54	96,32	2,14	1,35	0,60	97,06	2,34	0,90	0,27	96,75	2,98	0,62	

Tabela 13: Taxas de cobertura e comprimentos médios dos intervalos de 99% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão variável, utilizando os estimadores obtidos pelos métodos *naive* (*Naive*), calibração da regressão (RC), máxima verossimilhança aproximada (MPVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0,75$. Cenário 2.

k_x	$\hat{\theta}$	Intervalo	$n = 40$				$n = 80$				$n = 160$			
			$Md_{\mu_t} = 0, 5157$		$Md_{\mu_t} = 0, 4989$		$Md_{\mu_t} = 0, 4962$		$\phi_t \in [1, 4161; 145, 6058]$		$\phi_t \in [1, 4161; 145, 6058]$		$\phi_t \in [1, 4161; 145, 6058]$	
			inf	sup	comp	comp	inf	sup	comp	comp	inf	sup	comp	comp
0,75	α_0	<i>Naive</i>	0,83	94,97	0,99	0,63	0,43	95,57	0,77	0,50	0,39	95,36	0,82	0,34
		RC	0,83	97,59	1,58	0,69	0,52	97,80	1,68	0,52	0,58	97,59	1,83	0,34
		MVa	1,23	97,60	1,17	0,67	0,93	98,64	0,43	0,51	1,01	98,54	0,45	0,35
		MPVa	1,03	98,17	0,80	1,05	0,96	98,48	0,56	0,53	0,99	98,55	0,46	0,35
	α_1	<i>Naive</i>	0,50	89,22	6,06	4,27	0,32	90,40	5,06	3,69	0,05	92,20	3,57	2,46
		RC	1,10	94,97	3,93	4,52	0,80	95,29	3,91	3,73	0,51	95,83	3,66	2,41
		MVa	2,38	92,70	4,92	4,15	0,93	95,42	3,65	3,62	0,41	97,63	1,96	2,42
		MPVa	2,60	91,78	5,62	5,31	1,44	94,48	4,08	3,57	0,73	97,17	2,10	2,39
	β_1	<i>Naive</i>	0,00	37,69	40,81	0,52	0,00	11,89	72,41	0,37	0,00	0,60	97,08	0,25
		RC	5,11	90,18	4,71	0,60	4,44	88,82	6,74	0,38	3,03	87,53	9,44	0,26
		MVa	2,77	95,14	2,09	0,82	1,01	97,42	1,57	0,56	0,36	98,13	1,51	0,39
		MPVa	1,64	95,61	2,75	2,13	1,01	96,85	2,14	0,54	0,47	97,48	2,05	0,37
γ_0	<i>Naive</i>	0,08	61,26	20,46	1,11	0,00	24,12	57,22	0,78	0,00	2,57	92,11	0,54	
	RC	0,09	85,33	14,58	1,30	0,00	50,86	49,14	0,87	0,00	12,23	87,77	0,61	
	MVa	3,85	95,97	0,18	2,31	1,35	98,38	0,27	1,50	0,40	99,20	0,40	1,00	
	MPVa	3,42	96,14	0,44	5,92	1,69	97,79	0,52	1,49	0,76	98,77	0,47	0,97	
λ_1	<i>Naive</i>	0,01	50,43	29,25	1,11	0,00	17,46	65,96	0,73	0,00	1,30	95,69	0,49	
	RC	1,15	90,15	8,70	1,50	0,43	73,55	26,02	0,83	0,03	48,71	51,26	0,53	
	MVa	3,09	95,56	1,35	2,23	1,24	97,08	1,68	1,44	0,25	97,19	2,56	0,94	
	MPVa	2,43	94,67	2,90	3,86	1,07	95,40	3,53	1,29	0,31	95,26	4,43	0,84	

Tabela 14: Taxas de cobertura e comprimentos médios dos intervalos de 99% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com dispersão variável, utilizando os estimadores obtidos pelos métodos *naive* (*Naive*), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0,50$. Cenário 2.

k_x	$\hat{\theta}$	Intervalo	$n = 40$				$n = 80$				$n = 160$			
			$Md_{\mu_t} = 0, 5160$		$Md_{\mu_t} = 0, 4988$		$Md_{\mu_t} = 0, 4988$		$Md_{\mu_t} = 0, 4962$		$\phi_t \in [2, 1182; 92, 8067]$		$\phi_t \in [1, 7201; 117, 2303]$	
			inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp
0,50	α_0	<i>Naive</i>	0,57	93,97	1,37	0,72	0,28	94,04	1,45	0,57	0,19	92,51	1,64	0,38
		RC	2,85	91,66	5,49	0,78	0,91	94,54	4,55	0,60	0,65	93,97	5,38	0,40
		MVa	1,10	97,28	1,62	0,93	1,00	98,15	0,85	0,67	0,93	98,65	0,42	0,45
		MPVa	0,55	98,88	0,57	17,13	0,65	98,95	0,40	1,41	0,82	98,79	0,39	0,51
	α_1	<i>Naive</i>	0,34	85,68	9,08	4,69	0,18	86,28	8,07	4,07	0,03	87,95	6,24	2,74
		RC	0,79	93,91	5,30	4,90	0,68	93,90	5,42	4,10	0,33	94,11	5,56	2,70
		MVa	8,06	81,51	10,43	4,07	3,14	90,54	6,32	3,84	1,19	95,73	3,08	2,69
		MPVa	3,89	88,33	7,78	35,21	2,15	92,14	5,71	7,99	1,11	95,96	2,93	2,96
	β_1	<i>Naive</i>	0,00	1,24	95,66	0,48	0,00	0,00	100,00	0,33	0,00	0,00	100,00	0,27
		RC	21,38	61,44	17,18	0,54	18,97	58,60	22,43	0,35	14,33	57,28	28,39	0,24
		MVa	13,33	82,52	4,15	1,24	6,91	90,33	2,76	0,90	4,01	93,80	2,19	0,64
		MPVa	0,55	94,98	4,47	80,77	1,25	95,57	3,18	3,16	1,97	95,19	2,84	0,76
γ_0	<i>Naive</i>	0,01	20,28	64,40	1,10	0,00	1,24	96,16	0,77	0,00	0,00	99,99	0,54	
	RC	0,38	49,94	49,68	1,33	0,01	8,53	91,46	0,90	0,00	0,09	99,91	0,64	
	MVa	16,47	82,59	0,94	3,21	9,71	89,36	0,93	2,30	5,06	93,90	1,04	1,55	
	MPVa	7,62	91,02	1,36	115,61	5,25	93,43	1,32	11,86	3,62	95,08	1,30	2,22	
λ_1	<i>Naive</i>	0,00	9,63	80,57	0,89	0,00	0,26	99,74	0,59	0,00	0,00	100,00	0,40	
	RC	7,87	63,31	28,82	1,23	3,58	42,12	54,30	0,69	0,84	20,87	78,29	0,45	
	MVa	9,91	86,28	3,81	2,80	7,91	88,97	3,12	1,86	4,35	91,46	4,19	1,24	
	MPVa	2,52	93,08	4,40	156,70	3,09	92,02	4,89	6,82	2,71	90,69	6,60	1,53	

Tabela 17: Taxas de cobertura e comprimentos médios dos intervalos de 99% de confiança para os parâmetros do modelo de regressão beta não linear com erros de medidas com não linearidade na dispersão, utilizando os estimadores obtidos pelos métodos *naive* (*Naive*), calibração da regressão (RC), máxima verossimilhança aproximada (MVa) e máxima pseudo-verossimilhança (MPVa). Coeficiente de confiabilidade $k_x = 0.50$. Cenário 3.

k_x	$\hat{\theta}$	Intervalo	$n = 40$						$n = 80$						$n = 160$							
			$Md_{\mu_t} = 0.2623$			$Md_{\mu_t} = 0.2603$			$Md_{\mu_t} = 0.2687$			$\phi_t \in [8.9631; 1353.5230]$			$\phi_t \in [7.9631; 1511.6661]$			$\phi_t \in [6.0751; 1746.4286]$				
			inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp	inf	certo	sup	comp
	α_0	<i>Naive</i>	0.00	0.03	99.74	1.02	0.00	0.00	100.00	0.90	0.00	0.00	100.00	0.62	0.00	0.00	100.00	0.62	0.00	0.00	100.00	0.62
		RC	15.78	62.89	21.33	1.14	15.12	61.98	22.90	0.96	15.12	61.98	22.90	0.96	10.94	59.87	29.19	0.64	10.94	59.87	29.19	0.64
		MVa	11.54	73.77	14.69	1.48	11.30	75.63	13.07	1.29	11.30	75.63	13.07	1.29	6.91	83.23	9.86	1.01	6.91	83.23	9.86	1.01
		MPVa	0.14	94.05	5.81	13.02	0.17	95.15	4.68	9.38	0.17	95.15	4.68	9.38	0.26	95.61	4.13	3.05	0.26	95.61	4.13	3.05
	α_1	<i>Naive</i>	0.98	80.04	11.76	3.51	0.52	79.46	12.13	3.14	0.52	79.46	12.13	3.14	0.27	75.84	13.70	2.21	0.27	75.84	13.70	2.21
		RC	0.55	91.78	7.67	4.39	0.37	91.27	8.36	3.80	0.37	91.27	8.36	3.80	0.12	88.70	11.18	2.54	0.12	88.70	11.18	2.54
		MVa	25.11	54.15	20.74	2.06	29.30	50.90	19.80	1.75	29.30	50.90	19.80	1.75	28.08	57.83	14.09	1.44	28.08	57.83	14.09	1.44
		MPVa	14.62	72.12	13.26	12.82	18.88	68.77	12.35	8.03	18.88	68.77	12.35	8.03	20.88	69.69	9.43	3.60	20.88	69.69	9.43	3.60
0.50	β_1	<i>Naive</i>	99.99	0.01	0.00	0.51	100.00	0.00	0.45	100.00	0.00	0.00	0.45	100.00	0.00	0.00	0.31	100.00	0.00	0.00	0.31	
		RC	33.93	48.95	17.12	0.57	36.43	47.04	16.53	0.48	36.43	47.04	16.53	0.48	46.86	42.62	10.52	0.32	46.86	42.62	10.52	0.32
		MVa	17.42	67.43	15.15	0.86	15.77	70.53	13.70	0.76	15.77	70.53	13.70	0.76	12.52	79.25	8.23	0.60	12.52	79.25	8.23	0.60
		MPVa	6.35	93.62	0.03	8.52	5.38	94.49	0.13	6.19	5.38	94.49	0.13	6.19	4.82	95.02	0.16	2.00	4.82	95.02	0.16	2.00
	γ_0	<i>Naive</i>	0.00	47.11	33.64	1.78	0.01	33.33	47.13	1.51	0.01	33.33	47.13	1.51	0.00	6.25	84.66	1.03	0.00	6.25	84.66	1.03
		RC	0.42	39.25	60.33	2.11	0.18	24.67	75.15	1.70	0.18	24.67	75.15	1.70	0.00	4.93	95.07	1.13	0.00	4.93	95.07	1.13
		MVa	15.33	83.14	1.53	5.66	12.41	85.70	1.89	5.15	12.41	85.70	1.89	5.15	6.92	91.62	1.46	3.87	6.92	91.62	1.46	3.87
		MPVa	6.84	90.72	2.44	40.72	5.69	91.82	2.49	26.48	5.69	91.82	2.49	26.48	4.02	93.90	2.08	9.28	4.02	93.90	2.08	9.28
	γ_1	<i>Naive</i>	0.00	26.87	53.97	3.16	0.00	11.47	75.49	2.57	0.00	11.47	75.49	2.57	0.00	1.93	94.06	1.92	0.00	1.93	94.06	1.92
		RC	0.00	60.51	39.49	3.86	0.00	35.70	64.30	2.98	0.00	35.70	64.30	2.98	0.00	10.82	89.18	2.10	0.00	10.82	89.18	2.10
		MVa	22.78	57.29	19.93	5.29	21.50	55.24	23.26	4.73	21.50	55.24	23.26	4.73	24.72	56.53	18.75	4.34	24.72	56.53	18.75	4.34
		MPVa	14.36	72.59	13.05	40.25	13.50	68.95	17.55	29.34	13.50	68.95	17.55	29.34	18.15	66.65	15.20	12.46	18.15	66.65	15.20	12.46
	λ_1	<i>Naive</i>	0.00	1.20	96.06	1.05	0.00	0.18	99.14	0.88	0.00	0.18	99.14	0.88	0.00	0.00	100.00	0.60	0.00	0.00	100.00	0.60
		RC	2.63	47.00	50.37	1.29	1.96	37.29	60.75	1.02	1.96	37.29	60.75	1.02	0.43	16.68	82.89	0.66	0.43	16.68	82.89	0.66
		MVa	4.11	82.96	12.93	3.24	4.54	83.79	11.67	2.89	4.54	83.79	11.67	2.89	5.03	86.33	8.64	2.20	5.03	86.33	8.64	2.20
		MPVa	1.39	89.70	8.91	21.40	1.78	89.66	8.56	15.14	1.78	89.66	8.56	15.14	2.69	89.48	7.83	5.67	2.69	89.48	7.83	5.67