



Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Estatística

Dissertação de Mestrado

# **Regressão Gama Inversa Robusta**

**por**

**Josué Oliveira Lima Neto**

Brasília, 19 de agosto de 2025

# **Regressão Gama Inversa Robusta**

**por**

**Josué Oliveira Lima Neto**

Dissertação apresentada ao Departamento de Estatística da Universidade de Brasília, como requisito parcial para obtenção do título de Mestre em Estatística.

Orientadora: Profa. Dra. Terezinha  
Késsia de Assis Ribeiro

Brasília, 19 de agosto de 2025

Dissertação submetida ao Programa de Pós-Graduação em Estatística do Departamento de Estatística da Universidade de Brasília como parte dos requisitos para a obtenção do título de Mestre em Estatística.

Texto aprovado por:

Profa. Dra. Terezinha Késsia de Assis Ribeiro  
Orientadora, EST/UnB

Prof. Dr. Alan Ricardo da Silva  
Membro interno, EST/UnB

Prof. Dr. Francisco Felipe de Queiroz  
Membro externo, IME/USP

# Agradecimentos

Meus sinceros agradecimentos a todos os professores do PPGEST/UnB que contribuíram para minha formação pessoal e acadêmica, em especial a minha Orientadora, Professora Doutora Terezinha Késsia de Assis Ribeiro, por todo o apoio, paciência e dedicação ao longo desta jornada. Agradeço a ela especialmente pela confiança depositada em mim, mesmo nos momentos de dúvida, e por ter sido uma presença firme e generosa durante os desafios enfrentados. Agradeço aos colegas que estiveram presente nos momentos de insegurança em que a união foi essencial para superar os desafios. Por fim, agradeço especialmente a todos da minha família que contribuíram até aqui de alguma maneira para que essa caminhada na vida acadêmica fosse possível.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

# Resumo

## Regressão Gama Inversa Robusta

O modelo de regressão gama inverso é uma técnica estatística utilizada para modelar fenômenos aleatórios que assumem valores contínuos positivos tais como o tempo até a falha de um equipamento, o tempo entre chegadas em sistemas de fila, como em repartições públicas ou instituições financeiras, entre outros. A inferência para os parâmetros deste modelo é frequentemente baseada no procedimento de estimação por máxima verossimilhança. Entretanto, o método clássico de estimação por máxima verossimilhança pode apresentar resultados distorcidos na presença de observações atípicas. Neste trabalho, propõe-se um novo método de estimação robusto para os parâmetros do modelo de regressão gama inverso. Para motivar a necessidade de um novo estimador robusto sob esta classe de modelos, é demonstrado que a função de influência do estimador de máxima verossimilhança não é limitada. O novo procedimento inferencial que é proposto neste trabalho é baseado na classe de M-estimadores, a qual contém o estimador de máxima verossimilhança como um caso particular e é vastamente conhecida por produzir estimadores robustos. Para avaliar o desempenho e ilustrar a necessidade do novo estimador foram realizados estudos de simulações de Monte Carlo na ausência e na presença de contaminação nos dados. A partir dos estudos de simulação constatou-se que, na ausência de contaminação, o estimador de máxima verossimilhança é mais eficiente. Porém, na presença de contaminação, o estimador de máxima verossimilhança foi fortemente afetado, levando a estimativas distorcidas. Em contrapartida, o estimador proposto se mostrou superior

ao estimador de máxima verossimilhança no que diz respeito à estabilidade das estimativas, as quais não foram afetadas pela contaminação nos dados.

**Palavras-chave:** Estimador de máxima verossimilhança, Estimador robusto, Função de influência, *Outlier*, Regressão gama inversa.

# Abstract

## Robust Inverse Gamma Regression

The inverse gamma regression model is a statistical technique used for modeling random phenomena characterized by positive continuous values, such as the time until equipment failure and interarrival times in queuing systems—common in public service and financial institutions. The inference for the parameters of this model is typically performed using the maximum likelihood estimation method. However, this estimation approach can lead to distorted results in the presence of atypical observations. In this work, we propose a new robust estimation method for the parameters of the inverse gamma regression model. To motivate the importance of a robust alternative, we demonstrate that the influence function of the maximum likelihood estimator is unbounded. The new inference procedure is based on the class of M-estimators, which includes the maximum likelihood estimator as a special case and is well-known for producing robust estimators. To evaluate the performance and illustrate the necessity of the proposed estimator, we conducted Monte Carlo simulations in both presence and absence of contamination in the data. From the simulation studies, it was found that in the absence of contamination, the maximum likelihood estimator is more efficient. However, in the presence of contamination, the maximum likelihood estimator was significantly affected, leading to distorted estimates. In contrast, the proposed estimator demonstrated superiority over the maximum likelihood estimator in terms of the stability of the estimates, which were not affected by data contamination.

**Keywords:** Maximum likelihood estimator, Robust estimator, Influence function, Outlier, Inverse gamma regression.



# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Objetivos . . . . .	5
1.2	Organização do trabalho . . . . .	5
<b>2</b>	<b>Modelo de Regressão Gama Inverso</b>	<b>7</b>
2.1	Distribuição gama . . . . .	7
2.2	Regressão gama . . . . .	8
2.3	Distribuição gama inversa . . . . .	9
2.3.1	Parametrização original . . . . .	9
2.3.2	Reparametrização . . . . .	10
2.4	Regressão gama inversa . . . . .	13
2.5	Inferência por máxima verossimilhança . . . . .	13
<b>3</b>	<b>Medidas de Robustez</b>	<b>21</b>
3.1	Conceitos em inferência robusta . . . . .	21
3.2	M-Estimador . . . . .	24
3.3	Análise de Robustez do EMV . . . . .	26
<b>4</b>	<b>Inferência Robusta</b>	<b>29</b>
4.1	Estimação via maximização da $L_q$ -verossimilhança . . . . .	29
4.2	Propriedades . . . . .	33

<b>5</b>	<b>Estudos de Simulação</b>	<b>35</b>
<b>6</b>	<b>Conclusões</b>	<b>60</b>
6.1	Considerações Finais . . . . .	60
<b>A</b>	<b>Resultados Capítulo 4</b>	<b>63</b>
A.1	Transformação Potência . . . . .	63
	<b>Referências Bibliográficas</b>	<b>64</b>

# Lista de Tabelas

5.1	Razão entre os EQMTs dos estimadores sob os Cenários 1, 2, 3 e 4. . . . .	59
-----	---	----

# Lista de Figuras

2.1	Gráficos da função densidade da distribuição GIR fixando $\sigma = 0,5$ e variando $\mu = 1, 2, 3$ e $5$ , e fixando $\mu = 1$ e variando $\sigma = 0.3, 0.5, 0.75$ e $1$ . . . . .	12
3.1	Componentes do vetor escore versus $y$ ; $\mu = 1$ e $\sigma = 0,5$ . . . . .	28
5.1	Ilustração dos cenários de contaminação considerados para uma amostra de tamanho $n = 40$ . . . . .	38
5.2	<i>boxplots</i> das estimativas obtidas para o parâmetro $\beta_1$ sob o Cenário 1 via o EMV e via o EMVS variando $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro. . . . .	42
5.3	<i>boxplots</i> das estimativas obtidas para o parâmetro $\beta_2$ sob o Cenário 1 via o EMV e via o EMVS variando $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro. . . . .	43
5.4	<i>boxplots</i> das estimativas obtidas para o parâmetro $\gamma_1$ sob o Cenário 1 via o EMV e via o EMVS variando $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro. . . . .	44
5.5	<i>boxplots</i> das estimativas obtidas para o parâmetro $\beta_1$ sob o Cenário 2 via o EMV e via o EMVS variando $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro. . . . .	45

5.6	<i>boxplots</i> das estimativas obtidas para o parâmetro $\beta_2$ sob o Cenário 2 via o EMV e via o EMVS variando $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro. . . . .	46
5.7	<i>boxplots</i> das estimativas obtidas para o parâmetro $\gamma_1$ sob o Cenário 2 via o EMV e via o EMVS variando $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro. . . . .	47
5.8	<i>boxplots</i> das estimativas obtidas para o parâmetro $\gamma_2$ sob o Cenário 2 via o EMV e via o EMVS variando $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro. . . . .	48
5.9	<i>boxplots</i> das estimativas obtidas para o parâmetro $\beta_1$ sob o Cenário 3 via o EMV e via o EMVS variando $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro. . . . .	49
5.10	<i>boxplots</i> das estimativas obtidas para o parâmetro $\beta_2$ sob o Cenário 3 via o EMV e via o EMVS variando $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro. . . . .	50
5.11	<i>boxplots</i> das estimativas obtidas para o parâmetro $\beta_3$ sob o Cenário 3 via o EMV e via o EMVS variando $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro. . . . .	51
5.12	<i>boxplots</i> das estimativas obtidas para o parâmetro $\gamma_1$ sob o Cenário 3 via o EMV e via o EMVS variando $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro. . . . .	52
5.13	<i>boxplots</i> das estimativas obtidas para o parâmetro $\beta_1$ sob o Cenário 4 via o EMV e via o EMVS variando $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro. . . . .	53
5.14	<i>boxplots</i> das estimativas obtidas para o parâmetro $\beta_2$ sob o Cenário 4 via o EMV e via o EMVS variando $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro. . . . .	54

5.15	<i>boxplots</i> das estimativas obtidas para o parâmetro $\beta_3$ sob o Cenário 4 via o EMV e via o EMVS variando $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro. . . . .	55
5.16	<i>boxplots</i> das estimativas obtidas para o parâmetro $\gamma_1$ sob o Cenário 4 via o EMV e via o EMVS variando $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro. . . . .	56
5.17	<i>boxplots</i> das estimativas obtidas para o parâmetro $\gamma_2$ sob o Cenário 4 via o EMV e via o EMVS variando $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro. . . . .	57
5.18	<i>boxplots</i> das estimativas obtidas para o parâmetro $\gamma_3$ sob o Cenário 4 via o EMV e via o EMVS variando $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro. . . . .	58

# Abreviações e Siglas

EA	Eficiência Assintótica
EQMT	Erro Quadrático Médio Total
CV	Coefficiente de Variação
FDA	Função de Distribuição Acumulada
FMV	Função de Mudança de Variância
FDE	Função de Distribuição Empírica
IID	Independente e Identicamente Distribuído
FI	Função de Influência
DGI	Distribuição Gama Inversa
EMV	Estimador de Máxima Verossimilhança
EMVS	Estimador de Máxima Verossimilhança Substituto
FDP	Função Densidade de Probabilidade
SEGNP	Sensibilidade a Erro Grosseiro Não-Padronizada





# Capítulo 1

## Introdução

Fenômenos aleatórios que assumem valores contínuos positivos são frequentemente observados nas mais diversas áreas do conhecimento, tais como engenharia, economia, biologia e ciência dos materiais. Alguns exemplos de estudos sobre fenômenos que produzem dados no intervalo  $(0, \infty)$  são: o tempo até a falha de um equipamento, o tempo entre chegadas em sistemas de fila, como no transporte público ou redes de comunicação, fluxos de partículas ou radiação, intensidades de terremotos ou tempestades, crescimento de patrimônio financeiro ou dados econômicos, entre outros.

O modelo de regressão gama inverso é uma abordagem estatística usada para modelar variáveis dependentes contínuas e positivas, especialmente quando os dados apresentam assimetria positiva (Bourguignon e Gallardo, 2020). A base desse modelo é a distribuição gama inversa, amplamente utilizada em aplicações onde valores muito altos da variável resposta têm maior variabilidade, como na análise de sobrevivência e confiabilidade (Glen, 2017). A distribuição gama inversa é especialmente útil quando se quer modelar o tempo até o evento de falha, ou seja, o tempo que um sistema ou componente leva para falhar, sendo mais apropriada para sistemas onde a falha ocorre com maior probabilidade no início da operação (comumente associada ao conceito de “falha precoce” ou “falhas devido a defeitos de fabricação”).

Trabalhos recentes que utilizam modelos baseados na distribuição gama inversa podem ser

encontrados na literatura, como, por exemplo, os artigos citados a seguir. Lin et al. (1989) utilizaram o modelo gama inverso em estudos de confiabilidade para avaliar o tempo de vida de determinados equipamentos, descrevendo as principais causas de falhas desses equipamentos. Mead (2015) apresentou um novo modelo de confiabilidade utilizado em problemas de teoria da difração e problemas de corrosão em novas máquinas. Os autores propuseram uma generalização da distribuição gama inversa chamada de distribuição gama inversa generalizada (GIG). Mousa et al. (2016) utilizaram a abordagem não paramétrica de *kernel* para estimar a função densidade de probabilidade denominada de *kernel* gama inversa, com suporte pertencente ao intervalo  $(0, \infty)$ . Os autores mostraram que a densidade estimada tem as mesmas propriedades que a distribuição gama inversa, e que a mesma é livre de viés. Glen (2017) estudou algumas propriedades da distribuição gama inversa e como esta pode ser utilizada no contexto de análise de sobrevivência. Bourguignon e Gallardo (2020) propuseram um modelo de regressão onde a variável resposta segue uma distribuição gama inversa, reparametrizada por parâmetros que denotam média e precisão da distribuição. Magalhães et al. (2021) desenvolveram um esquema de correção de viés para reparametrizações do modelo de regressão gama inversa com precisão variável, derivando expressões para o viés de primeira ordem, obtendo assim estimadores corrigidos. Uma diferente parametrização para a distribuição gama inversa no contexto de modelos de regressão é apresentada por Rigby et al. (2019) na qual atribui-se estrutura de regressão para a moda da variável resposta. Esta última abordagem de regressão usando a distribuição gama inversa se torna mais adequada principalmente em situações nas quais os dados apresentam forte assimetria à direita. Neste caso, a modelagem da moda da resposta é mais adequada por ser mais representativa da tendência de concentração dos dados. Por esta razão, o modelo de regressão gama inverso a ser estudado neste trabalho é aquele apresentado por Rigby et al. (2019).

A inferência para o modelo de regressão usando a distribuição gama inversa (Bourguignon e Gallardo, 2020; Rigby et al., 2019) é normalmente baseada no estimador de máxima verossimilhança (EMV). Entretanto, este, pode ser altamente influenciado pela presença de observações

discrepantes nos dados. No Capítulo 3 mostra-se mediante o estudo da função de influência que o EMV sob a classe de modelos de regressão gama inverso proposto por Rigby et al. (2019) não é robusto.

As observações discrepantes ou *outliers* apresentam padrões distintos da maior parte dos dados e podem causar distorções graves em estimadores usuais para os parâmetros do modelo suposto. Por exemplo, geralmente os modelos de regressão linear são ajustados a partir do estimador de mínimos quadrados que são sensíveis a presença de *outliers*. Isto ocorre porque o estimador de mínimos quadrados é obtido a partir da minimização da soma dos quadrados das diferenças entre os valores observados e os valores previstos pelo modelo. Segundo Montgomery et al. (2021), quando há *outliers* nos dados, as grandes diferenças de valores geradas por esses pontos extremos fazem com que o modelo “se ajuste” para reduzir o erro para esses pontos, distorcendo as estimativas dos parâmetros para os dados. Segundo Maronna et al. (2019), é natural pensar que na presença de pontos atípicos é suficiente utilizar um bom método de rejeição de *outliers* para resolver o problema, mas isso geralmente não é verdade. De acordo com Rousseeuw e Leroy (2003), métodos para identificar *outliers* baseados em estimadores clássicos não são confiáveis devido à sensibilidade dos mesmos. Uma situação comum que pode ocorrer é que os resíduos podem passar despercebidos devido a pontos atípicos inflarem a estimativa da dispersão do modelo suposto, deixando de capturar pontos que deveriam ser classificados como *outliers*. Além disso, dado que uma observação é apontada como *outlier*, é necessário identificar qual é a melhor abordagem para tratá-la.

Existem várias abordagens tradicionais para tratar *outliers* em análise de dados, especialmente em modelos de regressão. A escolha da abordagem depende do contexto e do tipo de dado em estudo, mas todas possuem desvantagens quando comparadas a técnicas robustas. A primeira delas, segundo Ghosh (2019), consiste em identificar e então excluir esses pontos do conjunto de dados antes de ajustar o modelo, porém, ao excluir *outliers*, pode-se perder dados que, apesar de serem atípicos, são relevantes para entender a variabilidade do fenômeno em estudo. Além disso, se os *outliers* são comuns em um determinado contexto (por exemplo, em

vendas de produtos, onde ocorrem picos ocasionais), excluí-los pode resultar em uma amostra não representativa do comportamento real dos dados. Ribeiro (2020) cita que o melhor método de rejeição de *outliers* não se compara ao desempenho dos melhores procedimentos robustos, pois a estimação robusta transita suavemente entre a aceitação ou rejeição total de uma observação.

Huber (1964) desenvolveu uma nova abordagem de estimação sob modelos paramétricos baseado na minimização de uma generalização do logaritmo da função de verossimilhança. Tal classe de estimadores foi denominada M-estimadores. A generalização é feita através da substituição da contribuição individual  $-\log(f(y, \theta))$  no logaritmo da função de verossimilhança por uma função  $\rho(y, \theta)$ . Esta classe de estimadores é conhecida por produzir estimadores robustos que possuem propriedades ótimas como consistência e normalidade assintótica.

Um ramo de obtenção de estimadores robustos bastante utilizado e que pode ser visto como caso particular dos M-estimadores é o baseado no método de minimizar divergências entre densidades. Métodos de obtenção de estimadores baseados em minimizações de divergências empíricas podem ser facilmente encontrados na literatura. De acordo com Ghosh (2019) e Ribeiro (2020), ao realizar a minimização de uma divergência empírica entre duas densidades estamos minimizando a discrepância entre a distribuição de probabilidades postulada e a distribuição empírica dos dados observados com respeito ao parâmetro de interesse. Basu et al. (1998) propuseram um procedimento de estimação robusto obtido através da minimização empírica de um tipo de divergência potência entre duas densidades. Tal procedimento de estimação visa alcançar considerável robustez a pontos atípicos e ainda manter a eficiência assintótica próxima à do estimador de máxima verossimilhança, características as quais são controladas a partir de uma constante de afinação. Ferrari e La Vecchia (2012) propuseram um procedimento de estimação robusta baseado na maximização de uma  $L_q$ -verossimilhança reparametrizada, definida em termos da transformação de Box-Cox. O estimador resultante procura balancear a robustez e eficiência baseado em uma constante de afinação. Ribeiro e Ferrari (2023) desenvolveram um procedimento de estimação robusto sob modelos de regressão beta baseado no trabalho de

Ferrari e La Vecchia (2012) que é um caso especial dos M-estimadores. O estimador proposto, denominado de estimador de máxima verossimilhança substituto, oferece um balanço entre robustez e eficiência por meio de uma constante de afinação selecionada por um método orientado pelos dados que garante eficiência total na ausência de *outliers*.

## 1.1 Objetivos

O objetivo principal desse trabalho é propor um método de inferência robusta com base no procedimento inferencial desenvolvido por Ribeiro e Ferrari (2023), sob o modelo de regressão gama inverso reparametrizado, indexado pelos parâmetros da moda e forma como definido em Rigby et al. (2019).

Os objetivos específicos desse trabalho são:

- a) propor um método de estimação robusto para o modelo de regressão gama inverso reparametrizado;
- b) avaliar via estudos de simulação Monte Carlo o desempenho do estimador robusto em comparação ao estimador de máxima verossimilhança, na ausência e presença de contaminação nos dados.

## 1.2 Organização do trabalho

Esta dissertação está organizada em cinco Capítulos. No primeiro Capítulo, além da introdução, é apresentada uma revisão da literatura recente sobre o modelo de regressão gama inverso, assim como as principais referências relacionadas à teoria de robustez e os objetivos a serem alcançados. No segundo Capítulo, apresenta-se a distribuição gama inversa indexada pelos parâmetros da moda e forma, juntamente com a estrutura de regressão para os parâmetros, e a inferência por máxima verossimilhança. No terceiro Capítulo, é feita uma revisão dos principais conceitos e resultados que serão necessários para avaliar a robustez de estimadores, discutindo sobre o procedimento de M-estimação, e avaliando a robustez do EMV sob mode-

los de regressão gama inverso. No quarto Capítulo, é introduzido o estimador robusto baseado na maximização de uma  $L_q$ -verossimilhança reparametrizada sob o modelo de regressão gama inverso. No quinto Capítulo são apresentados estudos de simulação Monte Carlo para avaliar o desempenho do estimador robusto em comparação ao estimador de máxima verossimilhança, na ausência e presença de contaminação nos dados. No sexto Capítulo são apresentadas as conclusões com algumas considerações finais, assim como a indicação de trabalhos futuros. Detalhes técnicos referentes ao Capítulo 4 são deixados no Apêndice A.

# Capítulo 2

## Modelo de Regressão Gama Inverso

Neste Capítulo será introduzida a classe de modelos de regressão gama inverso. Primeiramente será caracterizada a distribuição gama e sua reparametrização no contexto de regressão. Em seguida, é apresentado o modelo gama inverso o qual é o objeto de estudo deste trabalho.

### 2.1 Distribuição gama

Em teoria de probabilidade, o modelo gama é uma distribuição assimétrica, unimodal de dois parâmetros, com suporte definido nos reais positivos, que pode modelar diversos tipos de fenômenos, como o tempo até a falha de um sistema ou o tempo de cura de um paciente submetido a um determinado tratamento (Feller, 1991).

Seja  $y$  a variável aleatória que segue uma distribuição gama com parâmetros  $\alpha, \beta > 0$ , denotada por  $y \sim G(\alpha, \beta)$ . A função de distribuição acumulada (FDA) é dada por (Colosimo e Giolo, 2021)

$$F(y; \alpha, \beta) = \frac{\gamma(\alpha, \beta y)}{\Gamma(\alpha)}, \quad y > 0, \quad (2.1)$$

em que  $\gamma(a, z) = \int_0^z t^{a-1} e^{-t} dt$  é a função gama incompleta,  $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$  é a função gama completa,  $\beta > 0$  é o parâmetro de taxa e  $\alpha > 0$  é o parâmetro de forma. A função

densidade de probabilidade (FDP) associada a (2.1) é dada por

$$f(y; \alpha, \beta) = \frac{\beta^\alpha y^{\alpha-1} e^{-y\beta}}{\Gamma(\alpha)}, \quad y > 0. \quad (2.2)$$

Particularmente, segundo Colosimo e Giolo (2021), a média e a variância de  $y$  são dadas por  $E(y) = \alpha/\beta$  e  $\text{Var}(y) = \alpha/\beta^2 = \mu^2/\alpha$ , respectivamente. Adicionalmente, a moda e o coeficiente de variação (CV) são dados por  $(\alpha - 1)\beta$ ,  $\alpha \geq 1$  e  $1/\sqrt{\alpha}$ , respectivamente. Quando  $\alpha = 1$ , a distribuição gama se reduz à distribuição exponencial, que é um caso especial da gama com um único parâmetro de forma. A distribuição gama pode ser parametrizada em termos da média e precisão. Segundo Paula (2023), a FDP da distribuição gama reparametrizada pode ser expressa por

$$f(y|(\mu, \phi)) = \frac{1}{\Gamma(\phi)} \left( \frac{\phi y}{\mu} \right)^\phi \exp \left\{ -\frac{\phi}{\mu} y \right\} \frac{1}{y}, \quad (2.3)$$

sendo  $y > 0$ ,  $\mu > 0$  parâmetro da média e  $\phi > 0$  parâmetro de precisão. A partir de agora, é usada a notação  $y \sim \text{GR}(\mu, \phi)$  para indicar que  $y$  é uma variável aleatória que segue uma distribuição gama reparametrizada com  $\mu > 0$  e  $\phi > 0$ , em que  $E(y) = \mu$  e  $\text{Var}(y) = \mu^2/\phi$ .

## 2.2 Regressão gama

Para definir a estrutura de regressão sob a distribuição gama, considere  $y_1, \dots, y_n$  variáveis aleatórias independentes tal que cada  $y_i$  possui a FDP dada em (2.3), com média  $\mu_i$  e parâmetro de precisão  $\phi_i$ . Suponha que a média e o parâmetro de precisão de  $y_i$  satisfazem as seguintes relações funcionais:

$$g_\mu(\mu_i) = \eta_{\mu_i} = \mathbf{x}_i^\top \boldsymbol{\beta} \quad \text{e} \quad g_\phi(\phi_i) = \eta_{\phi_i} = \mathbf{z}_i^\top \boldsymbol{\nu}, \quad (2.4)$$

em que  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top \in \mathbb{R}^p$  e  $\boldsymbol{\nu} = (\nu_1, \dots, \nu_q)^\top \in \mathbb{R}^q$  são os vetores dos coeficientes de



regressão desconhecidos, com  $p+q < n$ ,  $\eta_{\mu_i}$  e  $\eta_{\phi_i}$  são preditores lineares, e  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$  e  $\mathbf{z}_i = (z_{i1}, \dots, z_{iq})^\top$  são vetores de valores fixados das covariáveis dos submodelos da média e precisão, com dimensão  $p$  e  $q$ , respectivamente, para  $i = 1, \dots, n$ . As funções de ligação  $g_\mu : \mathbb{R}^+ \rightarrow \mathbb{R}$  e  $g_\phi : \mathbb{R}^+ \rightarrow \mathbb{R}$  em (2.4) devem ser estritamente monótonas, positivas e duas vezes diferenciáveis, de modo que  $\mu_i = g_\mu^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta})$  e  $\phi_i = g_\phi^{-1}(\mathbf{z}_i^\top \boldsymbol{\nu})$ , com  $g_\mu^{-1}(\cdot)$  e  $g_\phi^{-1}(\cdot)$  sendo as inversas das funções  $g_\mu(\cdot)$  e  $g_\phi(\cdot)$ , respectivamente.

## 2.3 Distribuição gama inversa

### 2.3.1 Parametrização original

Em teoria da probabilidade, a gama inversa (GI) é uma distribuição de dois parâmetros unimodal, positivamente assimétrica, com suporte definido nos reais positivos, que é obtida a partir do recíproco de uma variável distribuída conforme a distribuição gama. Em outras palavras, se  $y$  segue uma distribuição gama com parâmetros  $\alpha$  e  $\beta$ , então  $y^{-1}$  segue uma distribuição GI com parâmetros  $\alpha > 0$  parâmetro de forma, e  $\beta > 0$  parâmetro de escala (Bourguignon e Gallardo, 2020). Além disso, a recíproca também é verdadeira, se  $y$  segue uma distribuição GI,  $y^{-1}$  segue uma distribuição gama.

Seja a variável aleatória  $y$  que segue uma distribuição gama inversa com parâmetros  $\alpha, \beta > 0$ , denotada por  $y \sim \text{GI}(\alpha, \beta)$ . A sua função de distribuição acumulada é dada por

$$F(y; \alpha, \beta) = \frac{\gamma(\alpha, \beta/y)}{\Gamma(\alpha)}, \quad y > 0, \quad (2.5)$$

em que  $\beta > 0$  é o parâmetro de escala e  $\alpha > 0$  é o parâmetro de forma. A função densidade de probabilidade associada a (2.5) é dada por

$$f(y; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-\alpha-1} e^{-\frac{\beta}{y}}, \quad y > 0. \quad (2.6)$$

A média e a variância de  $y$  são dadas por

$$E(y) = \frac{\beta}{\alpha - 1}, \quad \alpha > 1, \quad \text{e} \quad \text{Var}(y) = \frac{\beta^2}{(\alpha - 2)(\alpha - 1)^2}, \quad \alpha > 2, \quad (2.7)$$

respectivamente. Adicionalmente, o coeficiente de variação é dado por  $(\alpha - 2)^{-1/2}$ . A moda de  $y \sim \text{GI}(\alpha, \beta)$  é obtida encontrando-se o ponto  $y$  que maximiza a função  $f(y; \alpha, \beta)$  ou equivalentemente, o valor de  $y$  que maximiza  $\log(f(y; \alpha, \beta))$ . Veja que

$$\log[f(y; \alpha, \beta)] = \alpha \log(\beta) - \log[\Gamma(\alpha)] - (\alpha + 1) \log(y) - \beta/y.$$

Derivando em relação à  $y$ , tem-se que

$$\frac{d \log[f(y; \alpha, \beta)]}{dy} = -\frac{\alpha + 1}{y} + \frac{\beta}{y^2}.$$

Igualando a derivada obtida a zero, tem-se:  $-(\alpha + 1)/y + \beta/y^2 = 0 \Leftrightarrow (\alpha + 1)/y = \beta/y^2 \Leftrightarrow y^2(\alpha + 1) = \beta y \Leftrightarrow y = \beta/(\alpha + 1)$ . A expressão da moda de  $y \sim \text{GI}(\alpha, \beta)$  será de grande importância para a definição do modelo de regressão GI que é considerado neste trabalho.

### 2.3.2 Reparametrização

Modelos de regressão usualmente são construídos para modelar um parâmetro que represente uma característica simples da variável resposta tais como média, mediana e moda. Contudo, a distribuição gama inversa com FDP (2.6) não está parametrizada em função de quantidades que representam características diretas da variável resposta. Nesta seção é apresentada a distribuição GI parametrizada em termos da moda e forma, seguindo a parametrização apresentada por Rigby et al. (2019). Considere  $\mu = \beta/(\alpha + 1)$  e  $\sigma = 1/\sqrt{\alpha}$ , isto é,  $\beta = \mu(1 + \sigma^{-2})$  e  $\alpha = \sigma^{-2}$ . Conforme esta nova parametrização, substituindo as expressões de  $\mu$  e  $\sigma$  em (2.7)

tem-se que

$$E(y) = \frac{(1 + \sigma^2)\mu}{(1 - \sigma^2)}, \text{ se } \sigma^2 < 1, \text{ e, } \text{Var}(y) = \frac{(1 + \sigma^2)^2 \mu^2 \sigma^2}{(1 - \sigma^2)^2 (1 - 2\sigma^2)}, \text{ se } \sigma^2 < 1/2.$$

Estas expressões de média e variância também estão apresentadas em Rigby et al. (2019).

A partir de agora a notação usada será  $y \sim \text{GIR}(\mu, \sigma)$  para indicar que  $y$  é uma variável aleatória que segue uma distribuição GI reparametrizada com  $\mu > 0$  parâmetro que denota a moda e  $\sigma > 0$  parâmetro de forma. Uma vez que a distribuição GIR preserva a propriedade de reciprocidade da distribuição IG original, isto é,  $1/y$  segue uma distribuição gama com parâmetros  $\sigma^{-2}$  e  $\mu(1 + \sigma^{-2})$  segue que

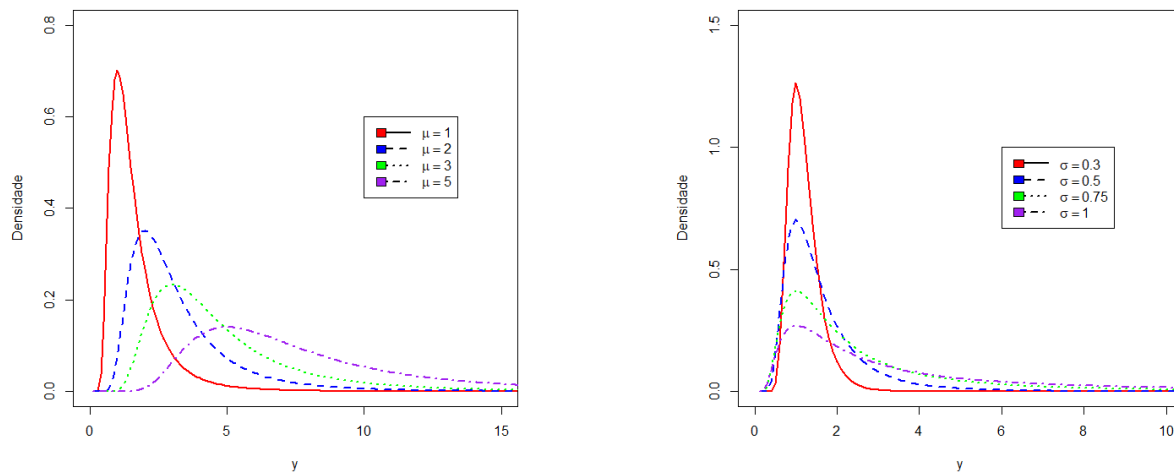
$$E[1/y] = \frac{\sigma^{-2}}{\mu(1 + \sigma^{-2})} \text{ e } \text{Var}[1/y] = \frac{\sigma^{-2}}{\mu^2(1 + \sigma^{-2})^2}. \quad (2.8)$$

Usando a reparametrização proposta, a função densidade da distribuição GIR fica reescrita como

$$f(y; \mu, \sigma) = \frac{[\mu(1 + \sigma^{-2})]^{\sigma^{-2}}}{\Gamma(\sigma^{-2})} y^{-(\sigma^{-2}+1)} e^{-\frac{\mu(1+\sigma^{-2})}{y}}, \quad y > 0. \quad (2.9)$$

Na Figura 2.1 são apresentados gráficos da função densidade da distribuição GIR variando valores de  $\mu$  e  $\sigma$  na qual percebe-se que à medida que  $\mu$  aumenta, a variância da distribuição também aumenta. Isso ocorre porque a variância da distribuição gama inversa tem uma dependência quadrática de  $\mu$ , o que implica que um aumento na moda tende a aumentar a variância de maneira mais acentuada. Além disso, também pode-se notar que o parâmetro  $\sigma$  controla a assimetria, ou seja, à medida que  $\sigma$  aumenta (diminui), a distribuição se torna menos (mais) simétrica em torno da moda, aumentando a variância. Sendo assim,  $\sigma$  que é um parâmetro de forma parece afetar a assimetria e precisão da distribuição.

Adicionalmente, segundo Rigby et al. (2019), tem-se que o CV, assimetria  $\gamma_1$  e curtose  $\gamma_2$



**Figura 2.1:** Gráficos da função densidade da distribuição GIR fixando  $\sigma = 0,5$  e variando  $\mu = 1, 2, 3$  e  $5$ , e fixando  $\mu = 1$  e variando  $\sigma = 0.3, 0.5, 0.75$  e  $1$ .

são dados, respectivamente por

$$CV = \sigma / \sqrt{1 + 2\sigma^2}, \quad \gamma_1 = \frac{4\sigma(1 - 2\sigma^2)^{1/2}}{1 - 3\sigma^2}, \quad \sigma^2 < 1/3, \quad \text{e} \quad \gamma_2 = \frac{3\sigma^2(10 - 22\sigma^2)}{(1 - 3\sigma^2)(1 - 4\sigma^2)}, \quad \sigma^2 < 1/4.$$

A distribuição gama e a distribuição GIR têm propriedades comuns, como unimodalidade em sua função de densidade e função de variância quadrática. A seguir são listadas algumas motivações para o uso prático da distribuição GIR em comparação com a distribuição gama, assim como a principal vantagem de se modelar a moda ao invés da média. A primeira delas é que a função taxa de falha da distribuição GI pode ter formato de “U”, dependendo do valor do parâmetro de forma (Glen, 2017). Em contrapartida, segundo Bourguignon e Gallardo (2020), a maioria das distribuições clássicas de dois parâmetros, como as distribuições Weibull e gama, possuem funções monótonas de taxa de risco. A segunda vantagem é que, apesar de a distribuição GI e a distribuição gama possuírem o mesmo CV, a distribuição GI permite obter maiores coeficientes de assimetria e curtose, possibilitando o ajuste de dados com caudas mais pesadas. Uma vantagem da parametrização em termos da moda em relação à média é que ao modelar a

moda, significa que se está sempre modelando a região onde a distribuição é mais concentrada. Em situações em que os dados possuem uma distribuição assimétrica, a modelagem da moda é mais adequada do que a da média, pois refletirá o comportamento mais representativo dos dados. Além disso, ao se trabalhar com a moda, espera-se que os efeitos de valores atípicos sejam menores do que os efeitos causados na estimação da média.

## 2.4 Regressão gama inversa

Para definir a estrutura de regressão, considere  $y_1, \dots, y_n$  variáveis aleatórias independentes tal que cada  $y_i$  possui a FDP dada em (2.9), com moda  $\mu_i$  e parâmetro de forma  $\sigma_i$ . Suponha que a moda e o parâmetro de forma satisfazem as seguintes relações funcionais

$$g_\mu(\mu_i) = \eta_{\mu_i} = \mathbf{x}_i^\top \boldsymbol{\beta} \quad \text{e} \quad g_\sigma(\sigma_i) = \eta_{\sigma_i} = \mathbf{z}_i^\top \boldsymbol{\nu}, \quad (2.10)$$

em que  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top \in \mathbb{R}^p$  e  $\boldsymbol{\nu} = (\nu_1, \dots, \nu_q)^\top \in \mathbb{R}^q$  são os vetores dos coeficientes de regressão desconhecidos, com  $p + q < n$ ,  $\eta_{\mu_i}$  e  $\eta_{\sigma_i}$  são preditores lineares, e  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top \in \mathbb{R}^p$  e  $\mathbf{z}_i = (z_{i1}, \dots, z_{iq})^\top \in \mathbb{R}^q$  são vetores de valores fixados das covariáveis dos submodelos da moda e forma, respectivamente, para  $i = 1, \dots, n$ . As funções de ligação  $g_\mu : \mathbb{R}^+ \rightarrow \mathbb{R}$  e  $g_\sigma : \mathbb{R}^+ \rightarrow \mathbb{R}$  em (2.10) devem ser estritamente monótonas, positivas e duas vezes diferenciáveis, de modo que  $\mu_i = g_\mu^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta})$  e  $\sigma_i = g_\sigma^{-1}(\mathbf{z}_i^\top \boldsymbol{\nu})$ , com  $g_\mu^{-1}(\cdot)$  e  $g_\sigma^{-1}(\cdot)$  sendo as inversas das funções  $g_\mu(\cdot)$  e  $g_\sigma(\cdot)$ , respectivamente.

## 2.5 Inferência por máxima verossimilhança

Os vetores de parâmetros  $\boldsymbol{\beta}$  e  $\boldsymbol{\nu}$  devem ser estimados com base numa amostra. Um procedimento clássico para a estimação de tais parâmetros é o método da máxima verossimilhança, o qual visa encontrar o valor de  $\boldsymbol{\theta}$  que provavelmente produziu os dados observados.

A função de verossimilhança para  $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\nu}^\top)^\top$  sob o modelo GIR é expressa por

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^n f(y_i; \mu_i, \sigma_i) \\ &= \prod_{i=1}^n \left\{ \frac{[\mu_i(1 + \sigma_i^{-2})]^{\sigma_i^{-2}}}{\Gamma(\sigma_i^{-2})} y_i^{-(\sigma_i^{-2}+1)} e^{-\frac{\mu_i(1+\sigma_i^{-2})}{y_i}} \right\}. \end{aligned}$$

Assim, o logaritmo da função de verossimilhança para  $\boldsymbol{\theta}$  é dado por

$$\begin{aligned} \ell(\boldsymbol{\theta}) &= \log(L(\boldsymbol{\theta})) \\ &= \sum_{i=1}^n \log(f(y_i; \mu_i, \sigma_i)) \\ &= \sum_{i=1}^n \log \left\{ \frac{[\mu_i(1 + \sigma_i^{-2})]^{\sigma_i^{-2}}}{\Gamma(\sigma_i^{-2})} y_i^{-(\sigma_i^{-2}+1)} e^{-\frac{\mu_i(1+\sigma_i^{-2})}{y_i}} \right\} \\ &= \sum_{i=1}^n \ell(\mu_i, \sigma_i), \end{aligned}$$

em que

$$\ell(\mu_i, \sigma_i) = \sigma_i^{-2} \log[\mu_i(1 + \sigma_i^{-2})] - (\sigma_i^{-2} + 1) \log(y_i) - \mu_i(1 + \sigma_i^{-2})/y_i - \log[\Gamma(\sigma_i^{-2})].$$

A estimativa de máxima verossimilhança é definida por  $\hat{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta})$ , com  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\nu}}^\top)^\top$ .

Assim, para se obter os respectivos EMVs para os parâmetros  $\boldsymbol{\beta}$  e  $\boldsymbol{\nu}$ , nos casos em que o logaritmo da função de verossimilhança é diferenciável, pode-se calcular as derivadas parciais de  $\ell(\boldsymbol{\theta})$  com relação a cada um dos parâmetros, obtendo assim os vetores escore para  $\boldsymbol{\beta}$  e  $\boldsymbol{\nu}$  que serão representados por  $\mathbf{U}_\beta(\boldsymbol{\theta})$  e  $\mathbf{U}_\nu(\boldsymbol{\theta})$ , respectivamente. Adicionalmente, os estimadores para a moda  $\mu_i$  e  $\sigma_i$  são obtidos indiretamente por meio de  $\hat{\boldsymbol{\beta}}$  e  $\hat{\boldsymbol{\nu}}$  através das estruturas de regressão definidas em (2.10). O vetor escore para  $\boldsymbol{\theta}$  é definido por

$$\mathbf{U}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{U}_\beta(\boldsymbol{\theta}) \\ \mathbf{U}_\nu(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\nu}} \end{bmatrix}.$$

As entradas do vetor escore associado à  $\beta$ ,  $U_{\beta_j}(\theta)$ ,  $j = 1, 2, \dots, p$  são dadas por

$$\begin{aligned} U_{\beta_j}(\theta) &= \frac{\partial \ell(\theta)}{\partial \beta_j} \\ &= \sum_{i=1}^n \frac{d\ell(\mu_i, \sigma_i)}{d\mu_i} \frac{d\mu_i}{d\eta_{\mu_i}} \frac{\partial \eta_{\mu_i}}{\partial \beta_j} \\ &= \sum_{i=1}^n \left\{ \frac{\sigma_i^{-2}}{\mu_i} - \frac{1 + \sigma_i^{-2}}{y_i} \right\} \frac{d\mu_i}{d\eta_{\mu_i}} \frac{\partial \eta_{\mu_i}}{\partial \beta_j} \\ &= \sum_{i=1}^n (y_i^* - \mu_i^*) \frac{1}{g'_\mu(\mu_i)} x_{ij}, \end{aligned}$$

em que  $y_i^* = -(1 + \sigma_i^{-2})/y_i$  e  $\mu_i^* = -\sigma_i^{-2}/\mu_i$ , com  $E(y_i^*) = \mu_i^*$  obtido usando o resultado em (2.8). Considere  $\mathbf{X}$  uma matriz de dimensão  $n \times p$  em que cada uma de suas colunas representa os valores da  $j$ -ésima covariável associada ao submodelo da moda,  $j = 1, 2, \dots, p$ ,  $\mathbf{y}^* = (y_1^*, \dots, y_n^*)^\top$ ,  $\boldsymbol{\mu}^* = (\mu_1^*, \dots, \mu_n^*)^\top$ ,  $\mathbf{T}_\beta = \text{diag}(t_{\beta_1}, \dots, t_{\beta_n}) = \text{diag}\{1/g'_\mu(\mu_1), \dots, 1/g'_\mu(\mu_n)\}$ ,  $t_{\beta_i} = d\mu_i/d\eta_{\mu_i}$  e  $\partial \eta_{\mu_i}/\partial \beta_j = x_{ij}$ . Logo, o vetor escore para  $\beta$  é

$$U_\beta(\theta) = \mathbf{X}^\top \mathbf{T}_\beta (\mathbf{y}^* - \boldsymbol{\mu}^*).$$

As entradas do vetor escore associado à  $\nu$ ,  $U_{\nu_j}(\theta)$ ,  $j = 1, 2, \dots, q$  são dadas por

$$\begin{aligned} U_{\nu_j}(\theta) &= \frac{\partial \ell(\theta)}{\partial \nu_j} \\ &= \sum_{i=1}^n \frac{d\ell(\mu_i, \sigma_i)}{d\sigma_i} \frac{d\sigma_i}{d\eta_{\sigma_i}} \frac{\partial \eta_{\sigma_i}}{\partial \nu_j} \\ &= \sum_{i=1}^n \left\{ -2\sigma_i^{-3} \log[\mu_i(1 + \sigma_i^{-2})] - \frac{2\sigma_i^{-5}}{1 + \sigma_i^{-2}} + 2\sigma_i^{-3} \log(y_i) \right. \\ &\quad \left. + 2\sigma_i^{-3} \frac{\mu_i}{y_i} + 2\sigma_i^{-3} \psi(\sigma_i^{-2}) \right\} \frac{d\sigma_i}{d\eta_{\sigma_i}} \frac{\partial \eta_{\sigma_i}}{\partial \nu_j} \\ &= \sum_{i=1}^n (y_i^\dagger - \mu_i^\dagger) \frac{1}{g'_\sigma(\sigma_i)} z_{ij}, \end{aligned}$$

em que  $y_i^\dagger = 2\sigma_i^{-3}[\log(y_i) + \mu_i/y_i]$  e  $\mu_i^\dagger = 2\sigma_i^{-3}[\log(\mu_i) + \log(1 + \sigma_i^{-2}) + \frac{\sigma_i^{-2}}{(1 + \sigma_i^{-2})} - \psi(\sigma_i^{-2})]$ , com  $E[\log(y_i)] = \log[\mu_i(1 + \sigma_i^{-2})] - \psi(\sigma_i^{-2})$ , e  $\psi(z) = d \ln[\Gamma(z)]/dz$  é a função digama. Considere  $\mathbf{Z}$  uma matriz de dimensão  $n \times q$  onde cada uma de suas colunas representa os valores da  $i$ -ésima covariável associada ao submodelo de  $\sigma$ ,  $j = 1, 2, \dots, q$ ,  $\mathbf{y}^\dagger = (y_1^\dagger, \dots, y_n^\dagger)^\top$ ,  $\boldsymbol{\mu}^\dagger = (\mu_1^\dagger, \dots, \mu_n^\dagger)^\top$ ,  $\mathbf{T}_\nu = \text{diag}(t_{\nu_1}, \dots, t_{\nu_n}) = \text{diag}\{1/g'_\sigma(\sigma_1), \dots, 1/g'_\sigma(\sigma_n)\}$ ,  $t_{\nu_i} = \partial\sigma_i/\partial\eta_{\sigma_i}$  e  $z_{ij} = \partial\eta_{\sigma_i}/\partial\nu_j$ . Assim, o vetor escore  $\mathbf{U}_\nu(\boldsymbol{\theta})$  pode ser representado por

$$\mathbf{U}_\nu(\boldsymbol{\theta}) = \mathbf{Z}^\top \mathbf{T}_\nu (\mathbf{y}^\dagger - \boldsymbol{\mu}^\dagger).$$

Sendo assim, o vetor escore de  $\boldsymbol{\theta}$  correspondente à  $i$ -ésima observação é dado por

$$U(y_i, \boldsymbol{\theta}) = \left( (y_i^* - \mu_i^*) \frac{1}{g'_\mu(\mu_i)} \mathbf{x}_i^\top, (y_i^\dagger - \mu_i^\dagger) \frac{1}{g'_\sigma(\sigma_i)} \mathbf{z}_i^\top \right)^\top. \quad (2.11)$$

A matriz de informação de Fisher é definida por

$$\mathbf{K}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \begin{bmatrix} \mathbf{K}_{\beta\beta} & \mathbf{K}_{\beta\nu} \\ \mathbf{K}_{\nu\beta} & \mathbf{K}_{\nu\nu} \end{bmatrix} = \begin{bmatrix} -E\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\beta\partial\beta^\top}\right) & -E\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\beta\partial\nu^\top}\right) \\ -E\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\nu\partial\beta^\top}\right) & -E\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\nu\partial\nu^\top}\right) \end{bmatrix}.$$

Para  $j, l = 1, 2, \dots, p$ , segue que

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\beta_j \partial\beta_l} &= \frac{\partial}{\partial\beta_l} \left\{ \sum_{i=1}^n \left[ \frac{\sigma_i^{-2}}{\mu_i} - \frac{1 + \sigma_i^{-2}}{y_i} \right] \frac{d\mu_i}{d\eta_{\mu_i}} \frac{\partial\eta_{\mu_i}}{\partial\beta_j} \right\} \\ &= \sum_{i=1}^n \frac{\partial}{\partial\beta_l} \left[ \frac{\sigma_i^{-2}}{\mu_i} - \frac{1 + \sigma_i^{-2}}{y_i} \right] \frac{d\mu_i}{d\eta_{\mu_i}} \frac{\partial\eta_{\mu_i}}{\partial\beta_j} \\ &\quad + \sum_{i=1}^n \left[ \frac{\sigma_i^{-2}}{\mu_i} - \frac{1 + \sigma_i^{-2}}{y_i} \right] \frac{d^2\mu_i}{d\eta_{\mu_i}^2} \frac{\partial\eta_{\mu_i}}{\partial\beta_j} \frac{\partial\eta_{\mu_i}}{\partial\beta_l} \\ &= - \sum_{i=1}^n \frac{\sigma_i^{-2}}{\mu_i^2} \frac{d\mu_i}{d\eta_{\mu_i}} \frac{\partial\eta_{\mu_i}}{\partial\beta_j} \frac{d\mu_i}{d\eta_{\mu_i}} \frac{\partial\eta_{\mu_i}}{\partial\beta_l} + \sum_{i=1}^n (y_i^* - \mu_i^*) \frac{d^2\mu_i}{d\eta_{\mu_i}^2} x_{ij} x_{il}. \end{aligned}$$



Daí, tem-se

$$-E \left( \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_l} \right) = \sum_{i=1}^n \frac{\sigma_i^{-2}}{\mu_i^2} \left( \frac{d\mu_i}{d\eta_{\mu_i}} \right)^2 x_{ij} x_{il}.$$

Daí, o componente da matriz de informação de Fisher para  $\boldsymbol{\beta}$  é escrita como

$$\mathbf{K}_{\boldsymbol{\beta}\boldsymbol{\beta}} = -E \left( \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \right) = \mathbf{X}^\top \boldsymbol{\Sigma} \mathbf{T}_{\boldsymbol{\beta}}^2 \mathbf{X},$$

em que  $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^{-2}/\mu_1^2, \dots, \sigma_n^{-2}/\mu_n^2)$ . Também, tem-se que

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \nu_j \partial \nu_l} &= \frac{\partial}{\partial \nu_l} \left\{ \sum_{i=1}^n \left\{ -2\sigma_i^{-3} \log[\mu_i(1 + \sigma_i^{-2})] - 2\sigma_i^{-3} \frac{\sigma_i^{-2}}{1 + \sigma_i^{-2}} \right. \right. \\ &\quad \left. \left. + 2\sigma_i^{-3} \log(y_i) + 2\sigma_i^{-3} \frac{\mu_i}{y_i} + 2\sigma_i^{-3} \psi(\sigma_i^{-2}) \right\} \frac{d\sigma_i}{d\eta_{\sigma_i}} \frac{\partial \eta_{\sigma_i}}{\partial \nu_j} \right. \\ &\quad \left. + \sum_{i=1}^n \left\{ -2\sigma_i^{-3} \log[\mu_i(1 + \sigma_i^{-2})] - 2\sigma_i^{-3} \frac{\sigma_i^{-2}}{1 + \sigma_i^{-2}} + 2\sigma_i^{-3} \log(y_i) \right. \right. \\ &\quad \left. \left. + 2\sigma_i^{-3} \frac{\mu_i}{y_i} + 2\sigma_i^{-3} \psi(\sigma_i^{-2}) \right\} \frac{\partial}{\partial \nu_l} \left\{ \frac{d\sigma_i}{d\eta_{\sigma_i}} \frac{\partial \eta_{\sigma_i}}{\partial \nu_j} \right\} \right. \\ &= \sum_{i=1}^n \left[ 6\sigma_i^{-4} \log[\mu_i(1 + \sigma_i^{-2})] + \frac{4\sigma_i^{-6}}{\sigma_i^{-2} + 1} + \frac{10\sigma_i^{-6} + 6\sigma_i^{-8}}{(1 + \sigma_i^{-2})^2} - 6\sigma_i^{-4} \log(y_i) - 6\sigma_i^{-4} \frac{\mu_i}{y_i} \right. \\ &\quad \left. - 6\sigma_i^{-4} \psi(\sigma_i^{-2}) - 4\sigma_i^{-6} \psi'(\sigma_i^{-2}) \right] \left( \frac{d\sigma_i}{d\eta_{\sigma_i}} \right)^2 z_{ij} z_{il} \\ &\quad + \sum_{i=1}^n \left\{ -2\sigma_i^{-3} \log[\mu_i(1 + \sigma_i^{-2})] - 2\sigma_i^{-3} \frac{\sigma_i^{-2}}{1 + \sigma_i^{-2}} + 2\sigma_i^{-3} \log(y_i) \right. \\ &\quad \left. + 2\sigma_i^{-3} \frac{\mu_i}{y_i} + 2\sigma_i^{-3} \psi(\sigma_i^{-2}) \right\} \frac{d^2 \sigma_i}{d\eta_{\sigma_i}^2} z_{ij} z_{il} \\ &= \sum_{i=1}^n \left[ 6\sigma_i^{-4} \log[\mu_i(1 + \sigma_i^{-2})] + \frac{4\sigma_i^{-6}}{\sigma_i^{-2} + 1} + \frac{10\sigma_i^{-6} + 6\sigma_i^{-8}}{(1 + \sigma_i^{-2})^2} - 6\sigma_i^{-4} \log(y_i) - 6\sigma_i^{-4} \frac{\mu_i}{y_i} \right. \\ &\quad \left. - 6\sigma_i^{-4} \psi(\sigma_i^{-2}) - 4\sigma_i^{-6} \psi'(\sigma_i^{-2}) \right] \left( \frac{d\sigma_i}{d\eta_{\sigma_i}} \right)^2 z_{ij} z_{il} \\ &\quad + \sum_{i=1}^n (y_i^\dagger - \mu_i^\dagger) \frac{d^2 \sigma_i}{d\eta_{\sigma_i}^2} z_{ij} z_{il}. \end{aligned}$$

Então, tem-se que

$$-E \left( \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\nu}_j \partial \boldsymbol{\nu}_l} \right) = - \sum_{i=1}^n \left[ \frac{8\sigma_i^{-6} + 4\sigma_i^{-8}}{(1 + \sigma_i^{-2})^2} - 4\sigma_i^{-6} \psi'(\sigma_i^{-2}) \right] \left( \frac{d\sigma_i}{d\eta_{\sigma_i}} \right)^2 z_{ij} z_{il}.$$

Logo, o componente da matriz de informação de Fisher para  $\boldsymbol{\nu}$  é

$$\mathbf{K}_{\boldsymbol{\nu}\boldsymbol{\nu}} = -E \left( \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\nu} \partial \boldsymbol{\nu}^\top} \right) = \mathbf{Z}^\top \mathbf{T}_{\boldsymbol{\nu}}^2 \mathbf{W} \mathbf{Z},$$

em que  $\mathbf{W} = \text{diag}\{w_1, \dots, w_n\}$ ,  $w_i = -\frac{8\sigma_i^{-6} + 4\sigma_i^{-8}}{(1 + \sigma_i^{-2})^2} + 4\sigma_i^{-6} \psi'(\sigma_i^{-2})$ . Por fim, tem-se que

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \nu_l} &= \frac{\partial}{\partial \nu_l} \left\{ \sum_{i=1}^n \left[ \frac{\sigma_i^{-2}}{\mu_i} - \frac{1 + \sigma_i^{-2}}{y_i} \right] \frac{d\mu_i}{d\eta_{\mu_i}} \frac{\partial \eta_{\mu_i}}{\partial \beta_j} \right\} \\ &= \sum_{i=1}^n \left[ -\frac{2\sigma_i^{-3}}{\mu_i} + \frac{2\sigma_i^{-3}}{y_i} \right] \frac{d\mu_i}{d\eta_{\mu_i}} \frac{d\sigma_i}{d\eta_{\sigma_i}} \frac{\partial \eta_{\mu_i}}{\partial \beta_j} \frac{\partial \eta_{\sigma_i}}{\partial \nu_l} \\ &= \sum_{i=1}^n \left[ -\frac{2\sigma_i^{-3}}{\mu_i} + \frac{2\sigma_i^{-3}}{y_i} \right] \frac{1}{g'_\mu(\mu_i)} \frac{1}{g'_\sigma(\sigma_i)} z_{il} x_{ij}. \end{aligned}$$

Portanto,

$$\begin{aligned} -E \left( \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \nu_l} \right) &= \sum_{i=1}^n \left[ \frac{2\sigma_i^{-3}}{\mu_i} - 2\sigma_i^{-3} E \left( \frac{1}{y_i} \right) \right] \frac{1}{g'_\mu(\mu_i)} \frac{1}{g'_\sigma(\sigma_i)} x_{ij} z_{il} \\ &= \sum_{i=1}^n \left[ \frac{2\sigma_i^{-3}}{\mu_i} - \frac{2\sigma_i^{-3} \sigma_i^{-2}}{\mu_i (1 + \sigma_i^{-2})} \right] \frac{1}{g'_\mu(\mu_i)} \frac{1}{g'_\sigma(\sigma_i)} x_{ij} z_{il} \\ &= \sum_{i=1}^n 2\sigma_i^{-3} \left[ \frac{1}{\mu_i} - \frac{\sigma_i^{-2}}{\mu_i (1 + \sigma_i^{-2})} \right] \frac{1}{g'_\mu(\mu_i)} \frac{1}{g'_\sigma(\sigma_i)} x_{ij} z_{il} \\ &= \sum_{i=1}^n \left[ \frac{2\sigma_i^{-3}}{\mu_i (1 + \sigma_i^{-2})} \right] \frac{1}{g'_\mu(\mu_i)} \frac{1}{g'_\sigma(\sigma_i)} x_{ij} z_{il}. \end{aligned}$$

Logo,

$$\mathbf{K}_{\beta\boldsymbol{\nu}} = -E \left( \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta \partial \boldsymbol{\nu}^\top} \right) = \mathbf{X}^\top \boldsymbol{\Phi}_{\beta\boldsymbol{\nu}} \mathbf{T}_\beta \mathbf{T}_{\boldsymbol{\nu}} \mathbf{Z},$$

em que  $\Phi_{\beta\nu} = \text{diag}(\phi_{\beta\nu_1}, \dots, \phi_{\beta\nu_n})$ ,  $\phi_{\beta\nu_i} = \frac{2\sigma_i^{-3}}{\mu_i(1+\sigma_i^{-2})}$ .

Assim, a matriz de informação de Fisher para o modelo é dada por

$$K_{\theta\theta} = \begin{bmatrix} K_{\beta\beta} & K_{\beta\nu} \\ K_{\nu\beta} & K_{\nu\nu} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top \Sigma \mathbf{T}_\beta^2 \mathbf{X} & \mathbf{X}^\top \Phi_{\beta\nu} \mathbf{T}_\beta \mathbf{T}_\nu \mathbf{Z} \\ \mathbf{Z}^\top \Phi_{\beta\nu} \mathbf{T}_\beta \mathbf{T}_\nu \mathbf{X} & \mathbf{Z}^\top \mathbf{T}_\nu^2 \mathbf{W} \mathbf{Z} \end{bmatrix}.$$

As estimativas de máxima verossimilhança  $\hat{\beta}$  e  $\hat{\nu}$  de  $\beta$  e  $\nu$ , respectivamente, são obtidos resolvendo o sistema de equações não linear  $\mathbf{U}(\theta) = \mathbf{0}_{p+q}$ , onde  $\mathbf{0}_{p+q}$  denota o vetor de zeros de dimensão  $p + q$ . Contudo, não é possível obter uma forma fechada para  $\hat{\beta}$  e  $\hat{\nu}$ . Para  $n$  grande, sob certas condições de regularidade,  $\hat{\theta}$  (o estimador de máxima verossimilhança de  $\theta$ ) é consistente. Além disso,  $\hat{\theta}$  segue uma distribuição aproximadamente normal, isto é,

$$\hat{\theta} \stackrel{a}{\sim} N_{p+q}(\theta, K_{\theta\theta}^{-1}).$$

sendo  $K_{\theta\theta}^{-1}$  a inversa da matriz  $K_{\theta\theta}$ , e  $\stackrel{a}{\sim}$  denota distribuição aproximada.

Dados  $j \in \{1, \dots, p\}$  e  $l \in \{1, \dots, q\}$ , um intervalo de confiança para  $\beta_j$  e  $\nu_l$  respectivamente, com  $(1 - \gamma)100\%$  de confiança são aproximadamente dados por

$$\left[ \hat{\beta}_j - z_{\gamma/2}(\hat{K}_{\beta_j\beta_j}^{-1})^{1/2} \ ; \ \hat{\beta}_j + z_{\gamma/2}(\hat{K}_{\beta_j\beta_j}^{-1})^{1/2} \right],$$

e

$$\left[ \hat{\nu}_l - z_{\gamma/2}(\hat{K}_{\nu_l\nu_l}^{-1})^{1/2} \ ; \ \hat{\nu}_l + z_{\gamma/2}(\hat{K}_{\nu_l\nu_l}^{-1})^{1/2} \right],$$

em que as quantidades  $\hat{K}_{\beta_j\beta_j}^{-1}$  e  $\hat{K}_{\nu_l\nu_l}^{-1}$  são, respectivamente, as variâncias assintóticas dos estimadores  $\hat{\beta}_j$  e  $\hat{\nu}_l$ , e  $z_{\gamma/2}$  representa o quantil da distribuição normal padrão tal que  $P(-z_{\gamma/2} \leq Z \leq z_{\gamma/2}) = 1 - \gamma$ , em que  $0 < \gamma < 1$  e  $Z \sim N(0, 1)$ .

Para realizar o procedimento inferencial por meio de máxima verossimilhança como des-

critério acima pode-se utilizar a função `gamlss` da biblioteca `gamlss` do *software* R informando `family = IGAMMA` que corresponde a distribuição gama inversa; para mais detalhes veja <http://127.0.0.1:25156/library/gamlss.dist/html/IGAMMA.html>. Esta foi a forma utilizada para a obtenção das estimativas de máxima verossimilhança dos modelos de regressão gama inverso considerados nos estudos de simulações discutidos no Capítulo 5.

# Capítulo 3

## Medidas de Robustez

Neste Capítulo são apresentadas algumas das principais medidas utilizadas para avaliar a robustez de estimadores, assim como a classe de M-estimadores que é a base para procedimento inferencial que será proposto. Primeiramente serão definidas as duas principais medidas que avaliam a robustez de estimadores, e em seguida, é caracterizado o procedimento de M-estimação.

### 3.1 Conceitos em inferência robusta

Nesta seção serão discutidos os principais conceitos necessários para o desenvolvimento do procedimento inferencial robusto que será proposto. Considere  $y_1, \dots, y_n$  variáveis aleatórias independentes e identicamente distribuídas (IID). Supondo que essas variáveis são provenientes da família de distribuições paramétrica  $\mathcal{F}_\Theta = \{F_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$ ,  $p \geq 1$ , com funções densidade  $f_\theta$ , serão considerados estimadores  $T$  para o parâmetro  $\theta$  tal que

$$T(y_1, \dots, y_n) = T(F_n), \quad (3.1)$$

e que dependam dos dados  $y_1, \dots, y_n$  apenas através da função de distribuição empírica (FDE)

definida por

$$F_n(Z) = \frac{1}{n} \sum_{i=1}^n I(y_i < Z),$$

em que  $I(\cdot)$  denota a função indicadora. Estimadores que satisfazem (3.1) são chamados de estimadores funcionais. Estimadores funcionais  $\mathbf{T}(F_n)$  são comumente consistentes (Serfling, 2009).

**Definição 3.1.1.** Um estimador  $\mathbf{T}(F_n)$  é dito consistente, se para  $\epsilon > 0$ ,

$$\lim_{n \rightarrow \infty} P(|\mathbf{T}(F_n) - \boldsymbol{\theta}| > \epsilon) = 0, \forall \boldsymbol{\theta} \in \Theta.$$

Em outras palavras,  $\mathbf{T}(F_n) \xrightarrow{P} \boldsymbol{\theta}, \forall \boldsymbol{\theta} \in \Theta$ , em que  $\xrightarrow{P}$  denota convergência em probabilidade. Tal estimador funcional  $\mathbf{T}(F_n)$  é dito ser consistente em  $F_\theta$ .

**Definição 3.1.2.** O estimador  $\mathbf{T}(F_n)$  é Fisher-consistente (Kallianpur e Rao, 1955) se é funcional e satisfaz

$$\mathbf{T}(F_\theta) = \boldsymbol{\theta}, \quad \forall \boldsymbol{\theta} \in \Theta. \quad (3.2)$$

A definição de Fisher-consistência difere da definição de consistência de um estimador  $\mathbf{T}(F_n)$ . Segundo Fasano et al. (2012), a propriedade de Fisher-consistência garante que o estimador irá atingir o verdadeiro valor do parâmetro  $\boldsymbol{\theta}$  quando calculado sob a distribuição populacional dos dados  $F_\theta$ . Em termos simples, a Fisher-consistência assegura que não haja viés sistemático no processo de estimação. Segundo Casella e Berger (2024), viés sistemático é um tipo de erro constante que ocorre de forma repetida e previsível em um processo de medição, estimativa ou inferência. Ele desloca os resultados consistentemente em uma direção (para mais ou para menos), levando a estimativas distorcidas do valor verdadeiro. A consistência de um estimador garante que, a medida que  $n$  cresce, o estimador converge em probabilidade para

o verdadeiro valor  $\theta$ . Logo, não se pode dizer que Fisher-consistência implica em consistência e vice-versa.

**Definição 3.1.3.** Seja  $F_{h,y} = (1 - h)F_\theta + h\Delta_y$  a função de distribuição contaminada obtida após a introdução de uma contaminação infinitesimal  $h$  no ponto  $y$ . A função de influência (FI) de  $\mathbf{T}(\cdot)$  em  $F_\theta$  é dada por

$$\begin{aligned} \text{FI}(y; \mathbf{T}, F_\theta) &= \frac{\partial}{\partial h} [\mathbf{T}(F_{h,y})] \Big|_{h=0} \\ &= \lim_{h \rightarrow 0} \frac{\mathbf{T}(F_{h,y}) - \mathbf{T}(F_\theta)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\mathbf{T}((1 - h)F_\theta + h\Delta_y) - \mathbf{T}(F_\theta)}{h}, \end{aligned} \quad (3.3)$$

em que  $\Delta_y$  é a medida de probabilidade que coloca toda massa em  $y$ . A Definição 3.1.3 é bastante intuitiva. A FI descreve o efeito causado no estimador  $\mathbf{T}(\cdot)$  após uma contaminação infinitesimal  $h$  no ponto  $y$ . De acordo com Hampel et al. (2011), a FI avalia o comportamento assintótico do estimador, quantificando o viés causado pela contaminação nos dados. Dessa forma, se um estimador  $\mathbf{T}(\cdot)$  possui FI não limitada para todo  $y$  pertencente ao suporte, isto será um indicativo de que  $\mathbf{T}(\cdot)$  não é robusto. Uma medida baseada na FI é a sensibilidade a erro grosseiro não-padronizada (SEGNP), apresentada por Hampel (1974) e definida como

$$\gamma_u^* = \sup_y \|\text{FI}(y; \mathbf{T}, F_\theta)\|, \quad (3.4)$$

com  $\|\cdot\|$  denotando a norma euclidiana. A medida  $\gamma_u^*$  mensura o valor máximo do viés causado na estimativa de  $\mathbf{T}(F_n)$  conforme a contaminação introduzida. Logo, a medida de SEGNP pode ser vista como um limite superior para o viés do estimador  $\mathbf{T}(F_n)$  sob o cenário de contaminação. Dessa forma, deseja-se que  $\gamma_u^*$  seja finita. Observa-se que, se existir pelo menos uma das entradas do estimador  $\mathbf{T}(\cdot)$  cuja respectiva FI divirja, então  $\gamma_u^*$  será infinito. Estimadores que possuem tal medida finita são denominados B-robustos (Rousseeuw, 1981).

Outra característica de interesse relacionada à robustez de estimadores é estudar o com-

portamento da variância assintótica de um estimador após uma contaminação infinitesimal nos dados. Rousseeuw (1981) propôs a função de mudança de variância (FMV). Tal medida será definida seguindo a notação apresentada por Genton e Rousseeuw (1995).

**Definição 3.1.4.** Seja  $F_{h,y} = (1 - h)F_{\theta} + h\Delta_y$  a função de distribuição contaminada obtida após a introdução de uma contaminação infinitesimal  $h$  no ponto  $y$ . A FMV do estimador  $\mathbf{T}(\cdot)$  em  $F_{\theta}$  é definida por

$$\begin{aligned} \text{FMV}(y; \mathbf{T}, F_{\theta}) &= \frac{\partial}{\partial h} [V(\mathbf{T}(F_{h,y}), F_{h,y})] \Big|_{h=0} \\ &= \lim_{h \rightarrow 0} \frac{V(\mathbf{T}(F_{h,y}), F_{h,y}) - V(\mathbf{T}, F_{\theta})}{h} \\ &= \lim_{h \rightarrow 0} \frac{V(\mathbf{T}(F_{h,y}), (1 - h)F_{\theta} + h\Delta_y) - V(\mathbf{T}, F_{\theta})}{h}, \end{aligned} \quad (3.5)$$

em que  $V(\mathbf{T}(F_{h,y}), F_{h,y})$  é a matriz de covariâncias do estimador  $\mathbf{T}(\cdot)$  sob o modelo contaminado  $F_{h,y}$ . Analogamente a FI a definição da medida FMV é bastante intuitiva. Esta avalia o efeito causado na matriz de covariâncias assintótica do estimador  $\mathbf{T}(\cdot)$  após uma contaminação  $h$  no ponto  $y$ .

### 3.2 M-Estimador

Uma das principais classes de estimadores que são estudadas a fim de obter estimadores robustos é a classe dos M-estimadores. Introduzida inicialmente por Huber (1964), esta classe de estimadores contém o estimador de máxima verossimilhança como caso particular. Considere  $y_1, \dots, y_n$  variáveis aleatórias independentes, cada uma com densidade  $f_{\theta}(y_i)$  com  $\theta \in \Theta$ . O logaritmo da função de verossimilhança é dado por

$$\ell(\theta) = \sum_{i=1}^n \log(f_{\theta}(y_i)).$$

A estimativa de máxima verossimilhança para  $\theta$  é obtida maximizando a função  $\ell(\theta)$ , isto é,



$\hat{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta})$ , ou, equivalentemente,  $\hat{\boldsymbol{\theta}} = \operatorname{argmin}_{\boldsymbol{\theta} \in \Theta} [-\ell(\boldsymbol{\theta})]$ . A ideia proposta por Huber (1964) é generalizar o procedimento de estimação por máxima verossimilhança substituindo a contribuição individual  $-\log(f_{\boldsymbol{\theta}}(y_i))$  por uma função  $\rho(y_i, \boldsymbol{\theta})$ , resultando no estimador

$$\mathbf{T}(F_n) = \operatorname{argmin}_{\boldsymbol{\theta} \in \Theta} \sum_{i=1}^n \rho(y_i, \boldsymbol{\theta}), \quad (3.6)$$

em que  $\rho(y_i, \boldsymbol{\theta})$  é uma função diferenciável em seus argumentos. A equação de estimação associada aos M-estimadores desde que  $\rho$  seja diferenciável é dada por

$$\sum_{i=1}^n \psi(y_i, \mathbf{T}(F_n)) = \mathbf{0}, \quad (3.7)$$

em que  $\psi(\cdot, \mathbf{T}(F_n)) = (\partial/\partial \boldsymbol{\theta})\rho(\cdot, \boldsymbol{\theta})|_{\boldsymbol{\theta}=\mathbf{T}(F_n)}$ . Qualquer estimador que satisfaz (3.6) é chamado de M-estimador. Ao considerar  $\rho(y_i, \boldsymbol{\theta}) = -\log(f_{\boldsymbol{\theta}}(y_i))$ , tem-se que  $\psi(y, \boldsymbol{\theta})$  é menos a função escore, e portanto, o EMV é um caso particular.

De acordo com Hampel (1974), a FI para o M-estimador é expressa por

$$\begin{aligned} \text{FI}(y, \mathbf{T}, F_{\boldsymbol{\theta}}) &= \left[ - \int \frac{\partial}{\partial \boldsymbol{\theta}} [\psi(y, \boldsymbol{\theta})] |_{\boldsymbol{\theta}=\mathbf{T}(F_{\boldsymbol{\theta}})} dF_{\boldsymbol{\theta}}(y) \right]^{-1} \psi(y, \mathbf{T}(F_{\boldsymbol{\theta}})) \\ &= M(\psi, F_{\boldsymbol{\theta}})^{-1} \psi(y, \mathbf{T}(F_{\boldsymbol{\theta}})), \end{aligned} \quad (3.8)$$

em que

$$M(\psi, F_{\boldsymbol{\theta}}) = - \int \frac{\partial}{\partial \boldsymbol{\theta}} [\psi(y, \boldsymbol{\theta})] |_{\boldsymbol{\theta}=\mathbf{T}(F_{\boldsymbol{\theta}})} dF_{\boldsymbol{\theta}}(y).$$

Portanto, a FI de um M-estimador é limitada se o produto da inversa da matriz  $M(\psi, F_{\boldsymbol{\theta}})$  pelo vetor  $\psi(y, \boldsymbol{\theta})$  for limitado. Se houver algum componente do vetor  $\psi(y, \boldsymbol{\theta})$  que não seja limitado, então este estimador não possui FI não limitada, e, portanto, não será B-robusto. Particu-

larmente, tomando  $\mathbf{T}(\cdot) = \hat{\boldsymbol{\theta}}$  como sendo o EMV, obtemos que

$$\text{FI}(y, \hat{\boldsymbol{\theta}}, F_{\boldsymbol{\theta}}) = K(\boldsymbol{\theta}, F_{\boldsymbol{\theta}})^{-1} U(y, \boldsymbol{\theta}). \quad (3.9)$$

Sendo assim, se algum componente do vetor escore não for limitado, a função de influência do EMV não será limitada, e, portanto, o estimador não será B-robusto. De acordo com Hampel (1974), para  $n$  grande, vale que

$$\sqrt{n}(\mathbf{T}(F_n) - \boldsymbol{\theta}) \xrightarrow[n \rightarrow \infty]{D} \mathbf{N}(\mathbf{0}, \mathbf{V}(\mathbf{T}, F_{\boldsymbol{\theta}})),$$

em que  $\mathbf{T}(F_n)$  é o M-estimador em questão e  $\mathbf{V}(\mathbf{T}, F_{\boldsymbol{\theta}})$  é a matriz de covariâncias de  $T$  definida como

$$\mathbf{V}(\mathbf{T}, F_{\boldsymbol{\theta}}) = M(\psi, F_{\boldsymbol{\theta}})^{-1} Q(\psi, F_{\boldsymbol{\theta}}) M(\psi, F_{\boldsymbol{\theta}})^{-1^{\top}}, \quad (3.10)$$

em que

$$Q(\psi, F_{\boldsymbol{\theta}}) = \int \psi(y, \mathbf{T}(F_{\boldsymbol{\theta}})) \psi(y, \mathbf{T}(F_{\boldsymbol{\theta}}))^{\top} dF_{\boldsymbol{\theta}}(y).$$

### 3.3 Análise de Robustez do EMV

Nessa Seção será provado que o procedimento de estimação por máxima verossimilhança não é robusto sob a classe de modelos de regressão gama inverso definidos por (2.9) e (2.10), usando os resultados obtidos na Seção 3.2. Considere o vetor  $U(y, \boldsymbol{\theta})$  como o escore para  $\boldsymbol{\theta}$  referente a observação  $y$ . Tome a partição  $U(y, \boldsymbol{\theta}) = (U_{\beta}(y, \boldsymbol{\theta})^{\top}, U_{\nu}(y, \boldsymbol{\theta})^{\top})^{\top}$ , em que  $U_{\beta}(y, \boldsymbol{\theta})$  e  $U_{\nu}(y, \boldsymbol{\theta})$  são vetores para  $\beta$  e  $\nu$  referentes a  $y$ . Assumindo  $y_0 > 0$  como um valor

fixado, tem-se que

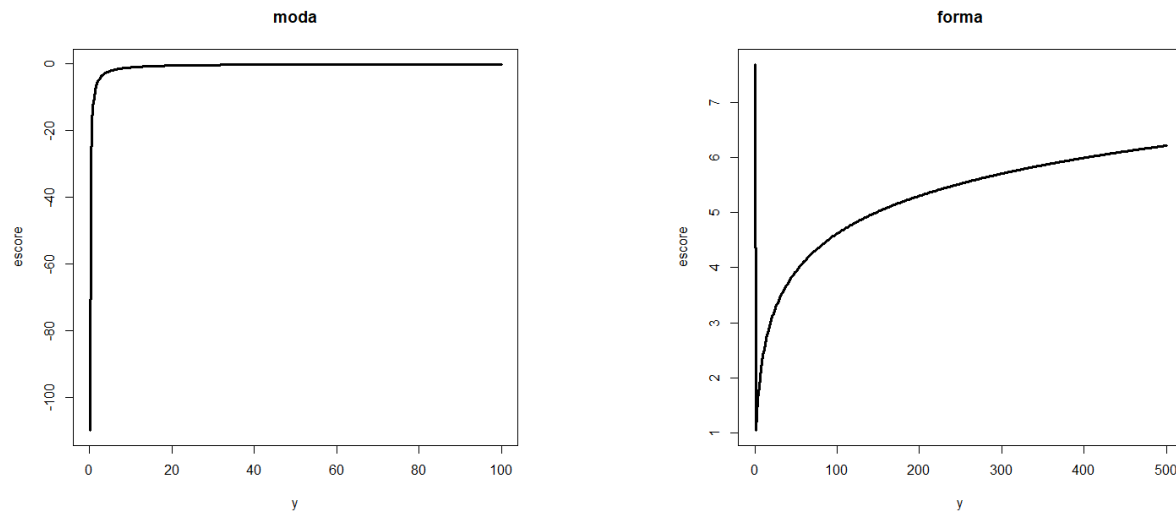
$$\lim_{y \rightarrow y_0} U_{\beta}(y, \theta) = \lim_{y \rightarrow y_0} (y^* - \mu^*) \frac{1}{g'_{\mu}(\mu)} \mathbf{X}^{\top} = \frac{1}{g'_{\mu}(\mu)} \left\{ \lim_{y \rightarrow y_0} y^* - \mu^* \right\} \mathbf{X}^{\top},$$

$$\lim_{y \rightarrow y_0} U_{\nu}(y, \theta) = \lim_{y \rightarrow y_0} (y^{\dagger} - \mu^{\dagger}) \frac{1}{g'_{\sigma}(\sigma)} \mathbf{Z}^{\top} = \frac{1}{g'_{\sigma}(\sigma)} \left\{ \lim_{y \rightarrow y_0} y^{\dagger} - \mu^{\dagger} \right\} \mathbf{Z}^{\top}.$$

Observa-se que, a menos de constantes,  $\lim_{y \rightarrow y_0} U_{\beta}(y, \theta) \propto \lim_{y \rightarrow y_0} y^*$  e  $\lim_{y \rightarrow y_0} U_{\nu}(y, \theta) \propto \lim_{y \rightarrow y_0} y^{\dagger}$ . Além disso,  $\lim_{y \rightarrow y_0} y^* = -\lim_{y \rightarrow 0} (1 + \sigma^{-2})/y$  e  $\lim_{y \rightarrow y_0} y^{\dagger} = 2\sigma^{-3} \lim_{y \rightarrow y_0} [\log(y) + \mu/y]$ . Quando  $y \rightarrow 0$ ,  $y^* \rightarrow -\infty$  e  $y^{\dagger} \rightarrow +\infty$  pois,  $\lim_{y \rightarrow 0} y^* = -(1 + \sigma^{-2}) \lim_{y \rightarrow 0} 1/y = -\infty$ . Agora, analisando separadamente os termos da soma  $\log(y) + \mu/y$ , quando  $y \rightarrow 0$ , temos que  $\log(y) \rightarrow -\infty$  e  $\mu/y \rightarrow +\infty$ . O termo  $\log(y)$  decresce para  $-\infty$ , mas de forma mais lenta em comparação com  $\frac{\mu}{y}$ , que tende a  $+\infty$  de forma muito mais rápida à medida que  $y$  se aproxima de zero.

Em contrapartida, quando  $y \rightarrow +\infty$ ,  $y^* \rightarrow 0$  e  $y^{\dagger} \rightarrow +\infty$  pois, quando  $y \rightarrow +\infty$ ,  $\lim_{y \rightarrow \infty} y^* = -(1 + \sigma^{-2}) \lim_{y \rightarrow \infty} 1/y = 0$  e  $\lim_{y \rightarrow \infty} y^{\dagger} = 2\sigma^{-3} \lim_{y \rightarrow \infty} [\log(y) + \mu/y] = +\infty$ . Portanto, o vetor escore não é limitado quando  $y \rightarrow 0$  ou  $y \rightarrow +\infty$ . Sendo assim, a FI do estimador de máxima verossimilhança não é limitada e, portanto o mesmo sob o modelo de regressão gama inverso (2.9) e (2.10) não é  $B$ -robusto. Também, por consequência, o EMV não é  $V$ -robusto. Na Figura 3.1 estão apresentadas as curvas do vetor escore para o caso IID. A partir da Figura 3.1 percebe-se que os componentes do vetor escore crescem e decrescem de forma ilimitada quando  $y$  se aproxima de zero ou tende a infinito, exceto para o escore de  $\mu$  quando  $y \rightarrow +\infty$  que é limitado pelo zero. Portanto, do estudo dos limites das entradas do vetor escore e da análise gráfica pode-se concluir que observações que assumem valores próximos de zero ou que tendem a infinito (valores muito altos) podem ser influentes no ajuste do EMV. Em outras palavras, em situações que tivermos dados observados  $y$  muito próximos a zero ou a valores muito altos de  $y$ , as estimativas obtidas pelo estimador de máxima verossimilhança sob

o modelo GIR podem não ser confiáveis.



**Figura 3.1:** Componentes do vetor escore versus  $y$ ;  $\mu = 1$  e  $\sigma = 0,5$ .

# Capítulo 4

## Inferência Robusta

Neste Capítulo será apresentado o procedimento inferencial robusto que será proposto neste trabalho. A ideia inicial do procedimento inferencial que será adotado aqui é maximizar uma função alternativa ao logaritmo da função de verossimilhança que conduz à estimação robusta.

### 4.1 Estimação via maximização da $L_q$ -verossimilhança

Considere  $y_1, \dots, y_n$  observações independentes, em que  $y_i \sim \text{GIR}(\mu_i, \sigma_i)$ . A função densidade postulada para  $y_i$  é dada por (2.9) e será denotada a partir de agora por  $f_{\theta}(y_i; \mu_i, \sigma_i)$ . Seja  $\ell_q(\theta)$  a função de  $L_q$ -verossimilhança (Ferrari e Yang, 2010), definida por

$$\ell_q(\theta) = \sum_{i=1}^n L_q(f_{\theta}(y_i; \mu_i, \sigma_i)), \quad (4.1)$$

em que  $0 < q \leq 1$  é uma constante denominada constante de afinação, e

$$L_q(u) = \begin{cases} (u^{1-q} - 1)/(1 - q), & \text{se } q \neq 1 \\ \log(u), & \text{se } q = 1, \end{cases}$$

é a transformação de Box-Cox (Box e Cox, 1964). Para  $q = 1$ ,  $\ell_q(\theta)$  é o logaritmo da função de

verossimilhança e maximizando-o obtém-se o EMV. Portanto, o estimador de máxima verossimilhança é um caso particular. O estimador obtido pela maximização de (4.1) será denotado por  $\tilde{\theta}_q$ . Uma vez que

$$f_{\theta}(y_i; \mu_i, \sigma_i)^{1-q} = \exp\{(1-q)\log(f_{\theta}(y_i; \mu_i, \sigma_i))\},$$

segue que a equação de estimação associada a  $\tilde{\theta}_q$  é dada por

$$\sum_{i=1}^n U(y_i, \theta) f_{\theta}(y_i; \mu_i, \sigma_i)^{1-q} = \mathbf{0}, \quad (4.2)$$

em que  $U(y_i, \theta)$  é o vetor escore associado à  $i$ -ésima observação dado em (2.11). Note que o vetor escore é ponderado por  $f_{\theta}(y_i; \mu_i, \sigma_i)^{1-q}$ , que depende do modelo postulado e da constante de afinação  $q$ . Se  $y_i$  é uma observação discrepante com relação ao modelo suposto aos dados,  $f_{\theta}$  se aproxima de zero e uma escolha de  $q$  menor que 1 produz um procedimento de estimação robusto, uma vez que a observação  $y_i$  receberá uma ponderação pequena. Além disso, (4.2) é um procedimento de M-estimação, como descrito na Seção 3.2, com  $\psi(y_i, \theta) = U(y_i, \theta) f_{\theta}(y_i; \mu_i, \sigma_i)^{1-q}$ . Com isso, as propriedades do estimador  $\tilde{\theta}_q$  podem ser obtidas dos resultados presentes na literatura de M-estimação. Uma desvantagem do estimador  $\tilde{\theta}_q$  é que este é obtido a partir de uma função de estimação viciada, ou seja,  $E(U_q(y, \theta)) \neq 0$ , com  $U_q(y, \theta) = \sum_{i=1}^n U(y_i, \theta) f_{\theta}(y_i; \mu_i, \sigma_i)^{1-q}$ , exceto para  $q = 1$ . O vício da equação de estimação associada a  $\tilde{\theta}_q$  equivale a falta de Fisher-consistência. Sob modelos de regressão beta, Ribeiro e Ferrari (2023) contornam o problema da falta de Fisher-consistência reparametrizando a função de  $L_q$ -verossimilhança com base em ideias de Ferrari e La Vecchia (2012). Será adotada a mesma estratégia sob modelos de regressão GIR. Sob o modelo GIR (2.9), o interesse residirá em encontrar o valor de  $\theta$  que maximiza a função de  $L_q$ -verossimilhança reparametrizada

definida por

$$\ell_q^*(\boldsymbol{\theta}) = \sum_{i=1}^n L_q \left( f_{\boldsymbol{\theta}}^{(1/q)}(y_i; \mu_i, \sigma_i) \right),$$

em que  $f_{\boldsymbol{\theta}}^{(1/q)}(y_i; \mu_i, \sigma_i) = f_{\boldsymbol{\theta}}(y_i; \mu_i, \sigma_{i,q^{-1}})$ , para  $0 < q \leq 1$ , com  $\sigma_{i,q} = [q(\sigma_i^{-2} + 1) - 1]^{-1/2}$ , desde que  $\sigma_{i,q^{-1}} > 0$  ou equivalentemente,  $\sigma_i^{-2} > q - 1$ . Desde que  $0 < q \leq 1$  e  $\sigma_i > 0$ , esta condio segue diretamente. A notaco  $f_{\boldsymbol{\theta}}^{(1/q)}(y_i; \mu_i, \sigma_i)$  denota que foi aplicada a transformaco potncia, definida a seguir, sob a densidade  $f_{\boldsymbol{\theta}}(y_i; \mu_i, \sigma_i)$ . Dada uma densidade  $h$  e uma constante  $\alpha > 0$ , a transformaco potncia  definida como

$$h^{(\alpha)}(y) = \frac{h(y)^\alpha}{\int h(y)^\alpha dy} \propto h(y)^\alpha, \forall y \text{ no suporte}, \quad (4.3)$$

sendo  $\int h(y)^\alpha dy < \infty$  para  $0 < \alpha < \infty$ . Para a famlia de densidades postuladas  $\{h_{\boldsymbol{\theta}}(\cdot), \boldsymbol{\theta} \in \Theta\}$  fechada sob (4.3), considere  $\tau_\alpha(\boldsymbol{\theta}) : \Theta \rightarrow \Theta$  uma funo inversvel que satisfaz

$$h_{\tau_\alpha(\boldsymbol{\theta})}(y) = h_{\boldsymbol{\theta}}^{(\alpha)}(y),$$

para todo  $y$  no suporte, supondo que esse no dependa de  $\boldsymbol{\theta}$ . Isso significa que, ao aplicar a transformaco potncia em  $h_{\boldsymbol{\theta}}$ , a densidade resultante  $h_{\boldsymbol{\theta}}^{(\alpha)}(\cdot)$  pertence  mesma famlia de distribues a que  $h_{\boldsymbol{\theta}}(\cdot)$  pertence, agora indexada por uma parametrizaco  $\tau_\alpha(\boldsymbol{\theta})$  ao invs de  $\boldsymbol{\theta}$ . O modelo gama inverso  fechado sob a transformaco potncia; veja a seo A.1. A quantidade  $f_{\boldsymbol{\theta}}^{(1/q)}(y_i; \mu_i, \sigma_i)$  corresponde ao modelo gama inverso modificado com submodelos da moda e forma definidos por

$$g_\mu^*(\mu_i) = g_\mu(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta}, \quad g_\sigma^*(\sigma_i) = g_\sigma(\sigma_{i,q}) = \mathbf{z}_i^\top \boldsymbol{\nu}.$$

Diferentemente do que ocorre no procedimento inferencial sob modelos de regresso beta robustos (Ribeiro e Ferrari, 2023), aqui o nico submodelo que sofre alteraco  o submodelo

referente ao parâmetro de forma  $\sigma$ . A densidade modificada será denotada por  $f_{\theta}^*(y_i; \mu_i, \sigma_i)$ .

Dessa forma, o estimador proposto é obtido através da maximização de

$$\ell_q^*(\theta) = \sum_{i=1}^n L_q(f_{\theta}^*(y_i; \mu_i, \sigma_i)), \quad (4.4)$$

que conduz a equação de estimação

$$\sum_{i=1}^n U^*(y_i, \theta) f_{\theta}^*(y_i; \mu_i, \sigma_i)^{1-q} = \mathbf{0}, \quad (4.5)$$

em que  $U^*(y_i, \theta) = \nabla_{\theta} \log[f_{\theta}^*(y_i; \mu_i, \sigma_i)]$  é o vetor escore modificado para  $\theta$  correspondente a  $i$ -ésima observação dado por

$$U^*(y_i, \theta) = \begin{pmatrix} \frac{\mathbf{X}_i(y_i^* - \mu_i^*)}{g_{\mu}(\mu_i)} \\ \frac{q^{-1} \mathbf{Z}_i(y_i^{\dagger} - \mu_i^{\dagger})}{g_{\sigma}(\sigma_{i,q})} \left( \frac{\sigma_i}{\sigma_{i,q}} \right)^3 \end{pmatrix}.$$

O estimador final  $\hat{\theta}_q$  é chamado de estimador de máxima verossimilhança substituto (EMVS).

A equação de estimação (4.5) associada ao EMVS depende de um valor pré-especificado para a constante de afinação  $q \in (0, 1]$ . Observe que o valor de  $q$  não é estimado conjuntamente com  $\theta$ . Por esta razão, antes de encontrar o valor de  $\theta$  que soluciona o sistema de equações (4.5) deve-se fixar um valor para  $q$ . Esta constante  $q$  controla a robustez e eficiência assintótica do EMVS. Valores de  $q$  muito próximos de 1 privilegiam a eficiência assintótica em detrimento de robustez, enquanto valores de  $q$  mais distantes de 1 acentuam a robustez do estimador. Ghosh (2019) propõe fixar valores para  $q$  ao obter um estimador robusto similar ao EMVS com base em estudos de simulação, sob modelos de regressão beta. Nestes estudos, o autor compara os valores das estimativas dos parâmetros da regressão de acordo com diferentes valores de  $q$  na presença e na ausência de contaminação. De forma geral, o autor sugere fixar valores de  $q \in (0, 6; 0, 7)$ . Estratégia similar será adotada neste trabalho. Serão fixados valores da constante



$q$  menores do que 1, entretanto não muito distantes de 1 para garantir eficiência assintótica do EMVS próxima ao EMV na ausência de contaminação e robustez do EMVS na presença de contaminação nos dados.

Analogamente ao EMV, o EMVS requer um algoritmo de otimização não-linear. Nesse trabalho, todos os resultados numéricos foram obtidos usando o método de otimização não-linear de Broyden-Fletcher-Goldfarb-Shanno (BFGS). Como estimativas iniciais para o modelo de regressão gama inverso com  $\sigma$  constante ou variável foram utilizadas as estimativas obtidas por meio do EMV. Em outras palavras, o algoritmo para encontrar as estimativas do EMVS é iniciado com base nas estimativas de máxima verossimilhança que são obtidas através do pacote `gamlss` como anteriormente citado na Seção 2.5.

## 4.2 Propriedades

Para provar a propriedade de Fisher-consistência do EMVS seguindo La Vecchia et al. (2015), é necessário mostrar que o valor esperado da equação de estimação (4.5) sob o modelo postulado  $f_{\theta}$  é igual a zero para todo  $\theta$ . Para tanto, mostrar-se-á que a média individual de  $U^*(y, \theta) f_{\theta}^*(y; \mu, \sigma)^{1-q}$  é nula para todo  $\theta$ . Assim, tem-se que

$$\begin{aligned} E_{\theta} (U^*(y, \theta) f_{\theta}^*(y; \mu, \sigma)^{1-q}) &= \int_0^{\infty} \nabla_{\theta} \log[f_{\theta}^*(y; \mu, \sigma)] f_{\theta}^*(y; \mu, \sigma)^{1-q} f_{\theta}(y; \mu, \sigma) dy \\ &= \int_0^{\infty} \frac{\nabla_{\theta} f_{\theta}^*(y; \mu, \sigma)}{f_{\theta}^*(y; \mu, \sigma)} f_{\theta}^*(y; \mu, \sigma)^{1-q} f_{\theta}(y; \mu, \sigma) dy \\ &= \int_0^{\infty} \frac{\nabla_{\theta} f_{\theta}^*(y; \mu, \sigma)}{f_{\theta}^*(y; \mu, \sigma)^q} f_{\theta}(y; \mu, \sigma) dy. \end{aligned}$$

Desde que  $f_{\theta}(y; \mu, \sigma) = c_q(\theta) f_{\theta}^*(y; \mu, \sigma)^q$ , em que  $c_q(\theta) = [\int_0^{\infty} f_{\theta}(y; \mu, \sigma)^{1/q} dy]^q$ , tem-se  $E_{\theta} (U^*(y, \theta) f_{\theta}^*(y; \mu, \sigma)^{1-q}) = c_q(\theta) \int_0^{\infty} \nabla_{\theta} f_{\theta}^*(y; \mu, \sigma) dy = c_q(\theta) \nabla_{\theta} \int_0^{\infty} f_{\theta}^*(y; \mu, \sigma) dy = 0, \forall \theta$ . Portanto, o EMVS é Fisher-consistente para  $\theta$ .

Outra propriedade de interesse que pode ser derivada para o EMVS sob os modelos de regressão gama inverso é a distribuição de probabilidades assintótica. Da teoria de M-estimação,

$\hat{\boldsymbol{\theta}}_q \approx \mathbf{N}(\boldsymbol{\theta}, \mathbf{V}_q(\boldsymbol{\theta}))$ , e

$$\mathbf{V}_q(\boldsymbol{\theta}) = \mathbf{J}_q(\boldsymbol{\theta})^{-1} \mathbf{K}_q(\boldsymbol{\theta}) \mathbf{J}_q(\boldsymbol{\theta})^{-1^\top}, \quad (4.6)$$

em que

$$\mathbf{J}_q(\boldsymbol{\theta}) = \sum_{i=1}^n \mathbf{E}\{\nabla_{\boldsymbol{\theta}^\top} [U^*(y_i, \boldsymbol{\theta}) f_{\boldsymbol{\theta}}^*(y_i; \mu_i, \sigma_i)^{1-q}]\}$$

e

$$\mathbf{K}_q(\boldsymbol{\theta}) = \sum_{i=1}^n \mathbf{E}\{U^*(y_i, \boldsymbol{\theta}) U^*(y_i, \boldsymbol{\theta})^\top f_{\boldsymbol{\theta}}^*(y_i; \mu_i, \sigma_i)^{2(1-q)}\}.$$

Neste trabalho, não serão apresentadas as expressões das matrizes  $\mathbf{J}_q$  e  $\mathbf{K}_q$ . A obtenção analítica da matriz de covariância assintótica do EMVS fará parte de trabalhos futuros provenientes da abordagem proposta neste trabalho.

## Capítulo 5

### Estudos de Simulação

Para avaliar o desempenho do EMVS em comparação ao desempenho do EMV na ausência e na presença de contaminação nos dados, foram realizados estudos de simulação de Monte Carlo, considerando os tamanhos amostrais de  $n = 40, 80, 160$  e  $320$ . Inicialmente foram obtidos valores das covariáveis para  $n = 40$ , e depois replicados duas, quatro e oito vezes, obtendo-se assim os valores das covariáveis correspondentes aos demais tamanhos amostrais. Esse esquema garante que o grau de heteroscedasticidade  $\max(\sigma_i)/\min(\sigma_i)$  seja constante para todos os tamanhos amostrais sob os cenários em que ambos  $\mu$  e  $\sigma$  são modelados por meio de estruturas de regressão. Para todos os cenários, a função de ligação utilizada foi a logarítmica nos dois submodelos de regressão. Os submodelos contêm intercepto e as covariáveis utilizadas no submodelo da moda são geradas a partir de variáveis aleatórias com distribuição uniforme padrão e permanecem constantes em todas as amostras simuladas. Nos cenários em que o parâmetro de forma  $\sigma$  é variável, as mesmas covariáveis utilizadas no submodelo da moda  $\mu$  são também empregadas no submodelo do  $\sigma$ . Em todos os cenários considerados, a proporção de contaminação nas amostras foi fixada em 5%. A constante de afinação  $q$  para o EMVS foi fixada em 0, 7; 0, 8; 0, 9, e 1 (EMV), conforme discutido no Capítulo 4. Todos os resultados são baseados em 1000 réplicas Monte Carlo que foram conduzidas utilizando o *software* R. Diferentes combinações dos valores dos parâmetros foram considerados. A seguir são descritos os

cenários de simulação considerados.

**Cenário 1: Modelo de regressão gama inverso: uma covariável no submodelo da moda e  $\sigma$  constante, valores da moda da resposta em torno de 0,99.** Os valores dos parâmetros foram fixados em  $\beta_1 = 1,0$ ,  $\beta_2 = -2,0$ , e  $\gamma_1 = -1,5$ , que conduzem a  $\mu \in (0,38; 2,65)$  com  $\text{Med}(\mu) = 0,99$  e  $\sigma = \exp(-1,5) = 0,22$ . A amostra contaminada substitui as observações independentes geradas com as 5% maiores modas da resposta por observações geradas a partir de um modelo de regressão gama inverso com as menores modas geradas. Por exemplo, para  $n = 40$ , 5% corresponde a duas observações contaminadas. Se as duas maiores modas foram 2,65 e 2,60, e as duas menores foram 0,38 e 0,39, então troca-se os valores gerados de  $y$  com  $\mu = 2,65$  e  $\mu = 2,60$  por duas novas observações de  $y$  geradas com  $\mu = 0,38$  e  $\mu = 0,39$ .

**Cenário 2: Modelo de regressão gama inverso duplo: uma covariável em cada submodelo, valores da moda da resposta em torno de 5,9.** Os valores dos parâmetros foram fixados em  $\beta_1 = 1,0$ ,  $\beta_2 = 1,5$ ,  $\gamma_1 = -1,8$  e  $\gamma_2 = -1,0$ , resultando em  $\mu \in (2,77; 12,01)$  com  $\text{Med}(\mu) = 5,92$  e  $\sigma \in (0,06; 0,16)$  com  $\text{Med}(\sigma) = 0,10$ . A contaminação substitui 5% da amostra como se segue: as observações geradas com os 2,5% maiores valores de  $\mu$  e aqueles gerados com os 2,5% menores valores são substituídos por observações independentes obtidas por meio de um modelo de regressão gama inverso com média  $\mu_i^{(1)} = \min(\mu)$  e  $\mu_i^{(2)} = \max(\mu)$ , respectivamente. Por exemplo, se o mínimo e o máximo de  $\mu$  são 2,77 e 12,01, respectivamente, então o valor gerado de  $y$  com  $\mu = 2,77$  será substituído por uma nova observação gerada para  $y$  com  $\mu = 12,01$ . Enquanto que, o valor gerado de  $y$  gerado com  $\mu = 12,01$  será substituído por uma nova observação gerada para  $y$  com  $\mu = 2,77$ .

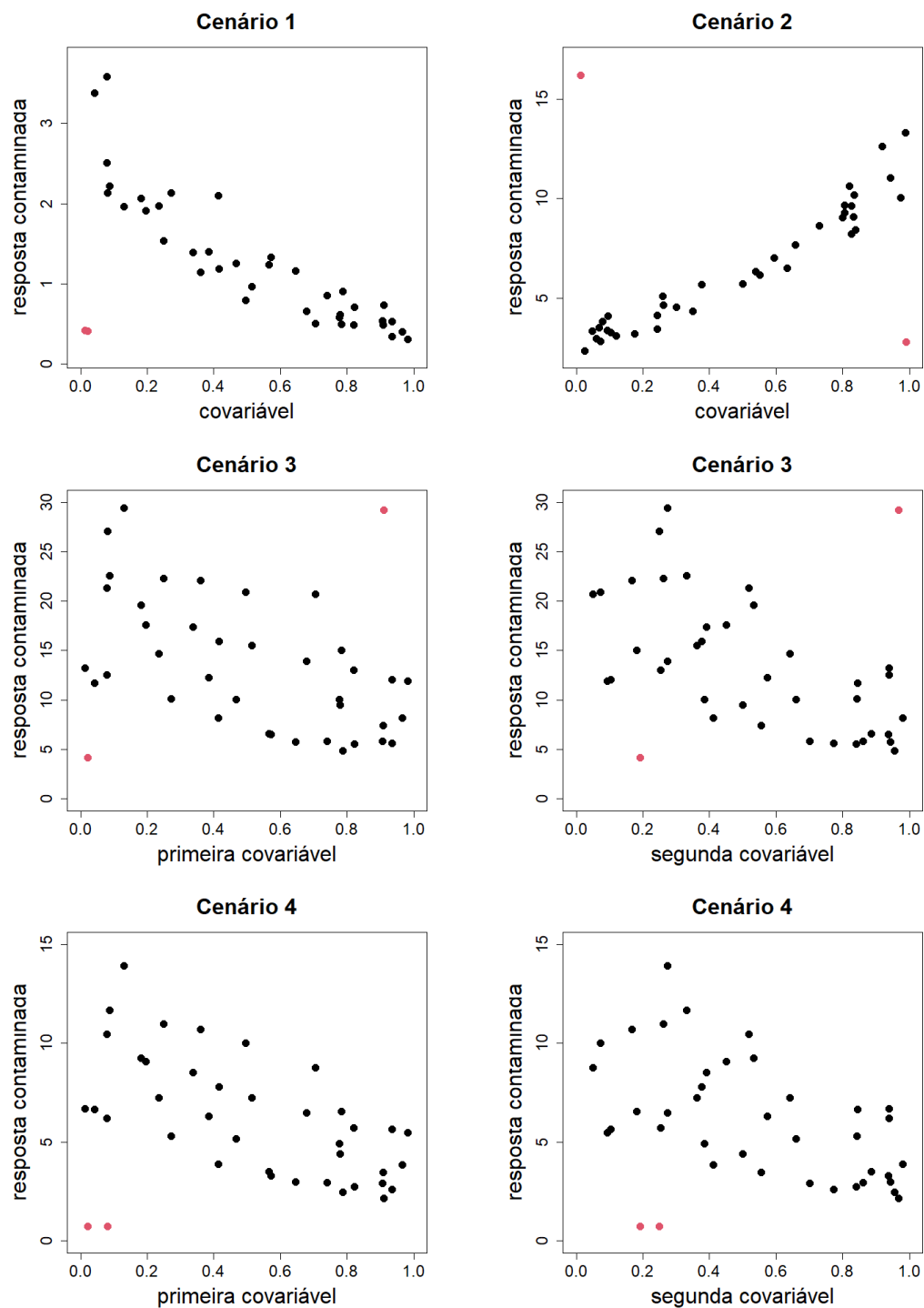
**Cenário 3: Modelo de regressão gama inverso: duas covariáveis no submodelo da moda e  $\sigma$  constante, valores da moda da resposta em torno de 12,37.** Os valores dos parâmetros foram fixados em  $\beta_1 = 3,7$ ,  $\beta_2 = -1,1$ ,  $\beta_3 = -1,3$ , e  $\gamma_1 = -2,5$ , que conduzem a

$\mu \in (4, 22; 30, 75)$  com  $\text{Med}(\mu) = 12,37$  e  $\sigma = \exp(-2,5) = 0,08$ . A contaminação substitui 5% da amostra como se segue: as observações geradas com os 2,5% maiores valores de  $\mu$  e aqueles gerados com os 2,5% menores valores são substituídos por observações independentes obtidas por meio de um modelo de regressão gama inverso com média  $\mu_i^{(1)} = \min(\mu)$  e  $\mu_i^{(2)} = \max(\mu)$ , respectivamente. Por exemplo, se o mínimo e o máximo de  $\mu$  são 4,22 e 30,75, respectivamente, então o valor gerado de  $y$  com  $\mu = 4,22$  será substituído por uma nova observação gerada para  $y$  com  $\mu = 30,75$ . Enquanto que, o valor gerado de  $y$  gerado com  $\mu = 30,75$  será substituído por uma nova observação gerada para  $y$  com  $\mu = 4,22$ .

**Cenário 4: Modelo de regressão gama inverso duplo: duas covariáveis em cada submodelo, valores da moda da resposta em torno de 6,11.** Os valores dos parâmetros foram fixados em  $\beta_1 = 3,0$ ,  $\beta_2 = -1,2$ ,  $\beta_3 = -1,2$ ,  $\gamma_1 = -2,5$ ,  $\gamma_2 = -1,0$ , e  $\gamma_3 = -1,0$ , que conduzem a  $\mu \in (2,11; 15,53)$  com  $\text{Med}(\mu) = 6,11$  e  $\sigma \in (0,01; 0,07)$ . A amostra contaminada substitui 5% da amostra como se segue: as observações geradas com os 5% maiores valores de  $\mu$  são substituídos por observações iguais a 0,6. Por exemplo, se as duas maiores modas foram 14 e 15, respectivamente, então o valor gerado de  $y$  com  $\mu = 14$  e  $\mu = 15$  serão ambos substituídos por 0,6.

A Figura 5.1 ilustra os cenários de contaminação descritos acima para uma única amostra de tamanho 40. As observações contaminadas destacam-se em vermelho. Visualmente, a contaminação introduzida nos Cenários 1 e 2 foi muito mais abrupta em comparação aos demais Cenários, dado que as observações contaminadas se destoam muito das demais observações. O grau de contaminação introduzido nos Cenários 2 e 3 foi um pouco mais suave porém suficiente para atingir o efeito desejado.

Nas Figuras 5.2 a 5.18 são apresentados os *boxplots* das estimativas dos parâmetros obtidas por meio do EMV e do EMVS para os dados com e sem contaminação, de acordo com os Cenários 1, 2, 3 e 4, respectivamente. A partir destas Figuras 5.2 a 5.18 é possível observar algumas tendências gerais com relação ao comportamento dos dois estimadores. Primeiro, nota-



**Figura 5.1:** Ilustração dos cenários de contaminação considerados para uma amostra de tamanho  $n = 40$ .

se que, na ausência de contaminação, o EMV apresenta o comportamento esperado, ou seja, apresenta distribuição centrada em torno do verdadeiro valor com variabilidade diminuindo de acordo com que o tamanho amostral aumenta. Este mesmo desempenho é observado para o EMVS para todos os valores de  $q$  fixados na ausência de contaminação com um leve aumento da variabilidade das estimativas deste para tamanhos amostrais pequenos. Este resultado é esperado desde que, para  $q < 1$ , o EMVS possui eficiência assintótica menor do que o EMV. Em contrapartida, na presença de contaminação, as estimativas de máxima verossimilhança para os parâmetros foram altamente afetadas pela contaminação introduzida nos dados. Em outras palavras, o EMV apresenta viés severo quando existem observações atípicas nos dados, levando a estimativas centradas em torno dos valores errados. Por exemplo, no Cenário 1, para o parâmetro  $\beta_2$ , o EMV sob contaminação produziu estimativas em torno do valor  $-1,10$ , porém o verdadeiro valor de  $\beta_2$  neste Cenário é  $-2,0$ . Analogamente, observando os resultados do Cenário 2 para o parâmetro  $\gamma_2$ , nota-se que o EMV sob contaminação produziu estimativas em torno do valor  $0,2$ , porém o verdadeiro valor de  $\gamma_2$  neste Cenário é  $-1,0$ .

Por outro lado, na presença de contaminação, as estimativas dos parâmetros obtidas via o EMVS variando o valor fixado de  $q$  permanecem centradas nos verdadeiros valores para a grande maioria dos cenários. Entretanto, é possível notar uma variabilidade extra que fica mais evidente em cenários com tamanhos amostrais menores, sobretudo para o Cenário de simulação 2. Por exemplo, observando a Figura 5.5 é possível perceber que para o tamanho amostral  $n = 40$ , mesmo estando centradas nos verdadeiros valores, as estimativas dos parâmetros obtidas via EMVS apresentam uma maior variabilidade na presença de contaminação nos dados em comparação às obtidas sem a presença de contaminação. Ainda, é possível identificar pontos atípicos nos *boxplots* das estimativas obtidas via o EMVS sob contaminação, em que estes tendem a se concentrar ao redor dos valores das estimativas obtidas via EMV sob contaminação. Este resultado é esperado para uma minoria de amostras, desde que o valor inicial utilizado no algoritmo de estimação do EMVS é a própria estimativa de máxima verossimilhança, fortemente enviesada sob contaminação.

O valor fixado para a constante  $q$  de afinação exerce um papel fundamental no procedimento de estimação proposto, desde que a depender do valor de  $q$ , atribui-se menores ponderações para observações discrepantes. Por exemplo, a partir das Figuras 5.7 e 5.8, que apresentam as estimativas dos parâmetros  $\gamma_1$  e  $\gamma_2$  sob o Cenário 2, é possível perceber que fixar  $q = 0,9$  não foi suficiente para garantir que as estimativas obtidas via o EMVS para  $\gamma_1$  e  $\gamma_2$  não fossem afetadas pela presença de contaminação. Em contrapartida, neste mesmo cenário, os valores  $q = 0,7$  e  $q = 0,8$  foram suficientes para garantir estimativas confiáveis de todos os parâmetros. Além disso, percebe-se alguma falta de estabilidade da estimação robusta relacionada ao tamanho amostral  $n$  nos Cenários 1 e 3 com  $q = 0,8$  e  $n = 160$ , e Cenário 4 com  $q = 0,7$  e  $n = 320$ . De forma geral, para a grande maioria dos cenários, observa-se que o valor  $q = 0,7$  foi adequado para garantir estimativas confiáveis dos parâmetros na presença e na ausência de contaminação.

Visando avaliar a eficiência assintótica em especial na presença de contaminação nos dados, o erro quadrático médio (EQM) é uma medida adequada, pois permite quantificar a discrepância entre as estimativas obtidas e os verdadeiros valores dos parâmetros levando em conta o viés e a variabilidade empírica das estimativas. O EQM do estimador  $\hat{\theta}_t$  via simulação é obtido por meio da expressão

$$\text{EQM}(\hat{\theta}_t) = \frac{1}{1000} \sum_{j=1}^{1000} \left( \hat{\theta}_t^j - \theta_t \right)^2,$$

em que  $\hat{\theta}_t^j$  é o EMV do  $t$ -ésimo elemento  $\theta_t$  do vetor de parâmetros  $\theta$  na  $j$ -ésima réplica de Monte Carlo, com  $j \in \{1, \dots, 1000\}$ . Para calcular o respectivo EQM do EMVS basta substituir  $\hat{\theta}_t^j$  por  $\hat{\theta}_{tq}^j$  que denota o EMVS do  $t$ -ésimo elemento  $\theta_t$  na  $j$ -ésima réplica.

Para comparar os EQMs obtidos via EMV e EMVS, calculou-se a razão entre os erros



quadráticos médios totais (REQMT) definido por

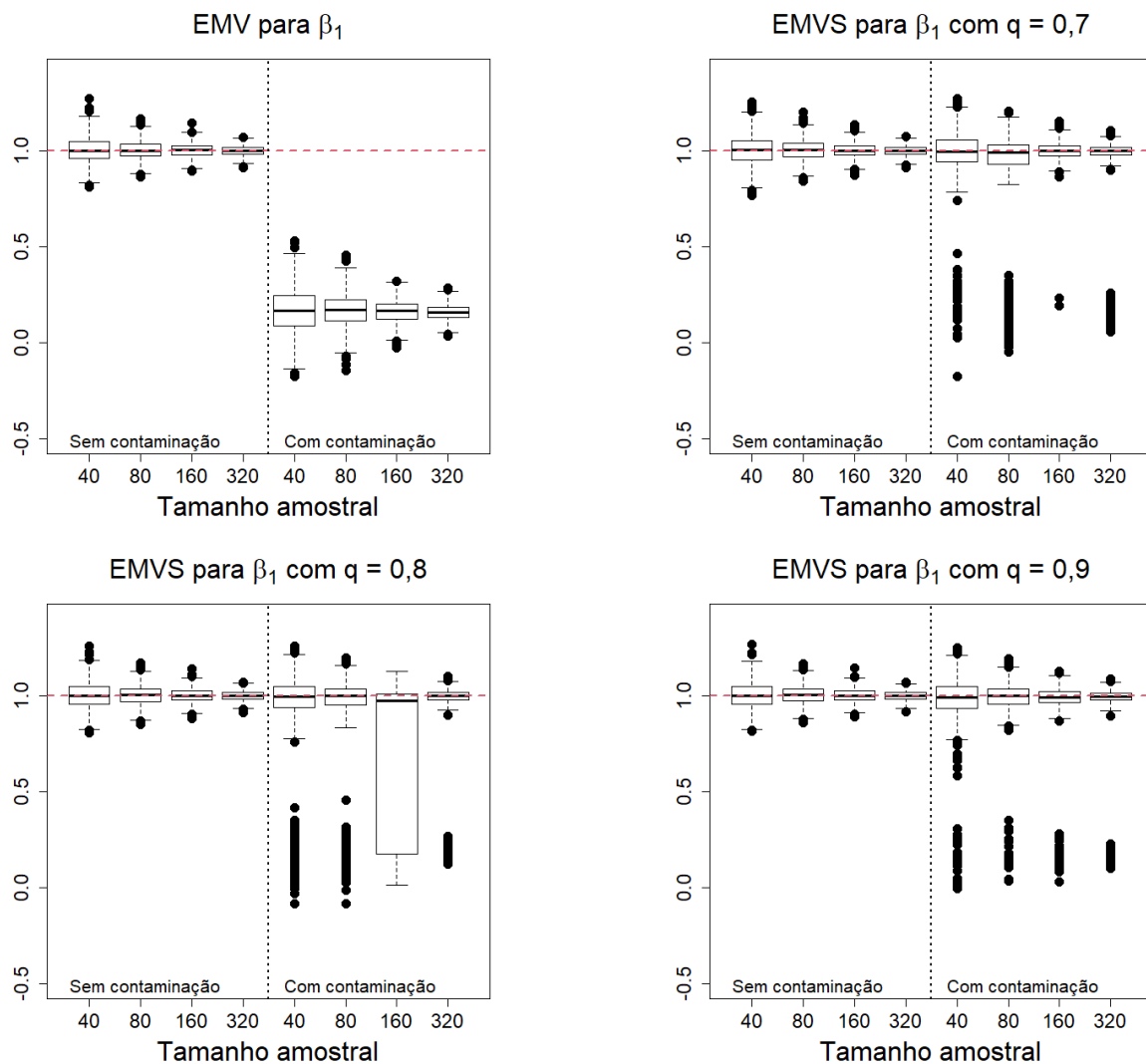
$$\text{REQMT} = \frac{\text{EQMT}(\hat{\theta})}{\text{EQMT}(\hat{\theta}_q)} = \frac{\sum_{t=1}^{p+q} \text{EQM}(\hat{\theta}_t)}{\sum_{t=1}^{p+q} \text{EQM}(\hat{\theta}_{tq})},$$

em que  $\text{EQMT}(\hat{\theta})$  e  $\text{EQMT}(\hat{\theta}_q)$  denotam a soma (o total) dos EQMs para todos os parâmetros do modelo considerando o EMV e o EMVS, respectivamente.

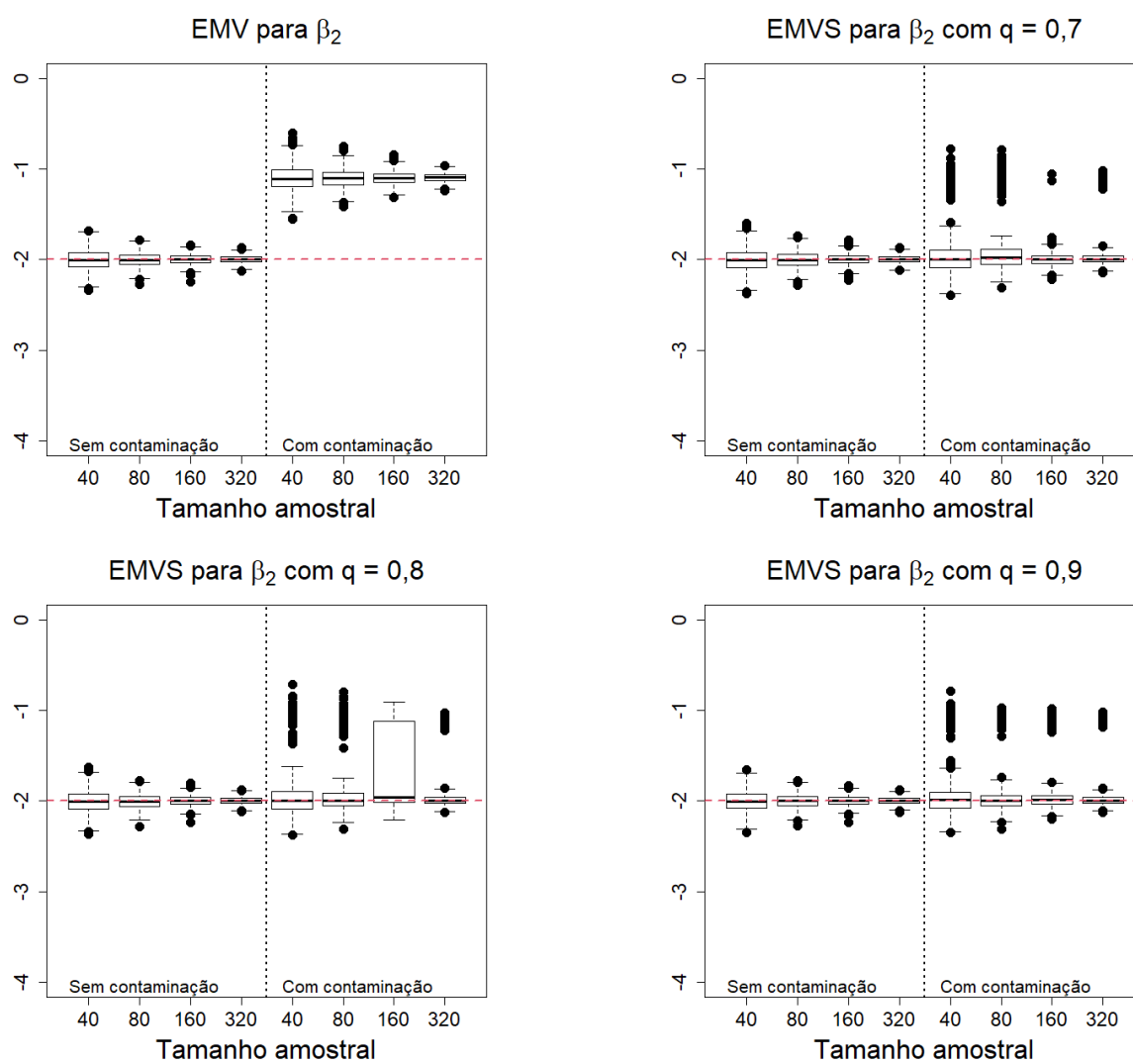
Na Tabela 5.1 são apresentadas as REQMTs sob todos os cenários considerados. Na ausência de contaminação nos dados, a eficiência dos estimadores é próxima (EQMT próximo a 1), em particular para tamanhos amostrais maiores e valor de  $q$  mais próximo de 1. Isso significa que quando não há presença de observações discrepantes o EMV é mais eficiente desde que  $\text{REQMT} < 1$ .

Na presença de contaminação, o EQMT do estimador robusto foi consideravelmente menor do que o do EMV, indicando um melhor desempenho do estimador proposto. Em alguns cenários essa discrepância cresce de acordo com que o tamanho da amostra aumenta. Em outros casos, a depender do valor fixado para  $q$ , a discrepância aumenta de um tamanho amostral para o outro, diminui, e depois volta a aumentar. Mesmo assim, em todos os casos, o EMVS apresentou um menor erro quadrático médio total na presença de contaminação. Por exemplo, para o tamanho amostral  $n = 40$  e considerando o Cenário 2, o EQMT do EMV foi 6,88; 8,22 e 3,75 vezes maior que o do EMVS, com  $q = 0,7$ ;  $q = 0,8$  e  $q = 0,9$  respectivamente. Agora, considerando o mesmo cenário, quando  $n = 160$ , o EQMT do EMV é 38; 43 e 6,8 vezes maior. Nos Cenários 2 e 4, na presença de contaminação, as razões entre o EQMT do EMV e o do estimador robusto são menores do que nos demais cenários. Nestes dois cenários, o parâmetro de forma  $\sigma$  é modelado por meio de estruturas de regressão. No Cenário 2 considera-se uma covariável nos submodelos (moda e forma); já no Cenário 4 consideram-se duas covariáveis nos submodelos. Segundo Ribeiro (2020), sob modelos de regressão beta, isto pode ocorrer pois

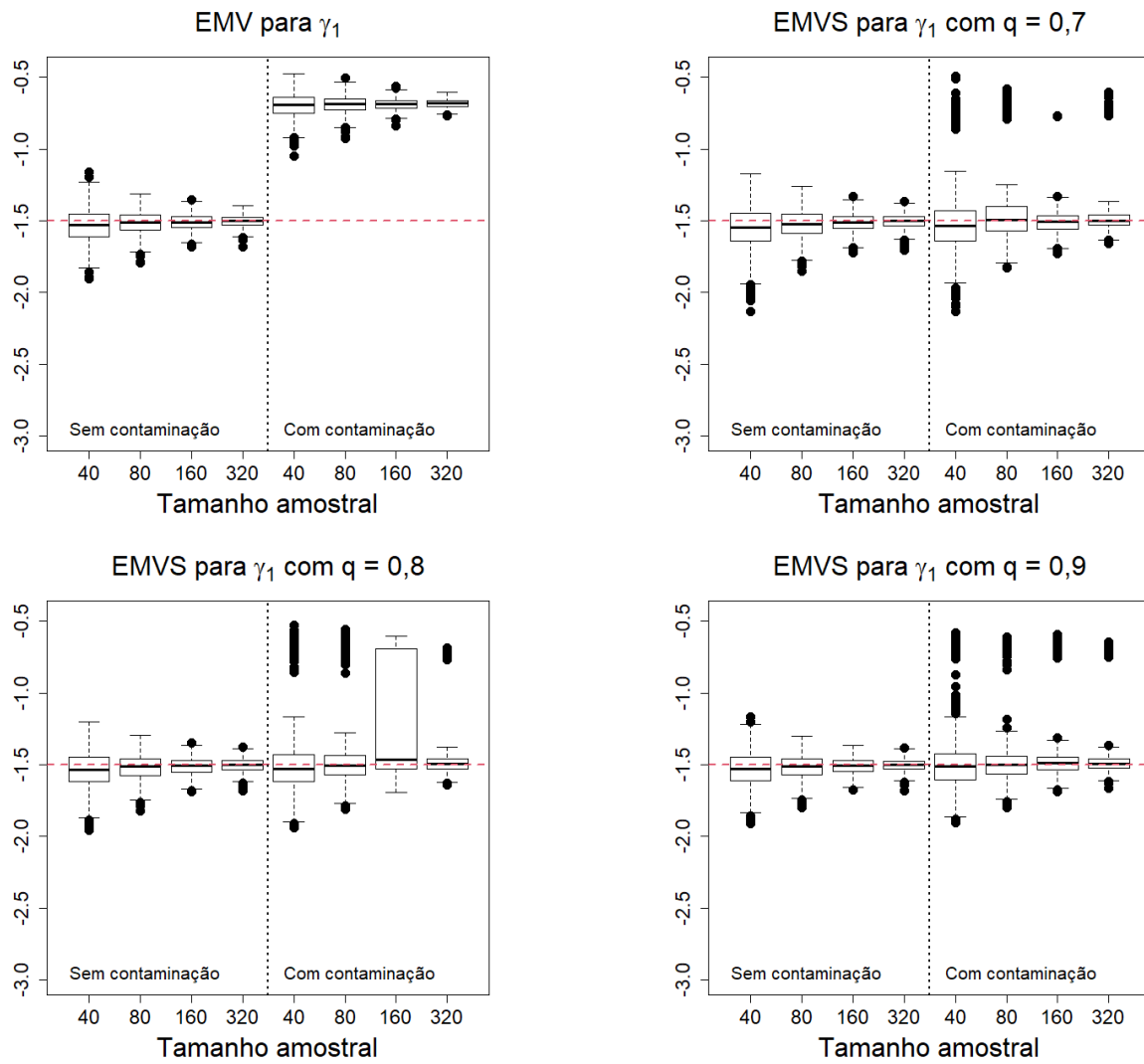
o viés produzido nas estimativas dos parâmetros pela contaminação nos dados é menor quando a precisão é modelada, reduzindo a discrepância entre os EQMTs. O mesmo comportamento é esperado sob regressão gama inversa, isto é, ao supor o modelo de regressão duplo modelando  $\mu$  e  $\sigma$  (mais flexível), pode-se tentar acomodar observações discrepantes, minimizando assim o impacto do viés no EMV e, conseqüentemente, no EQMT.



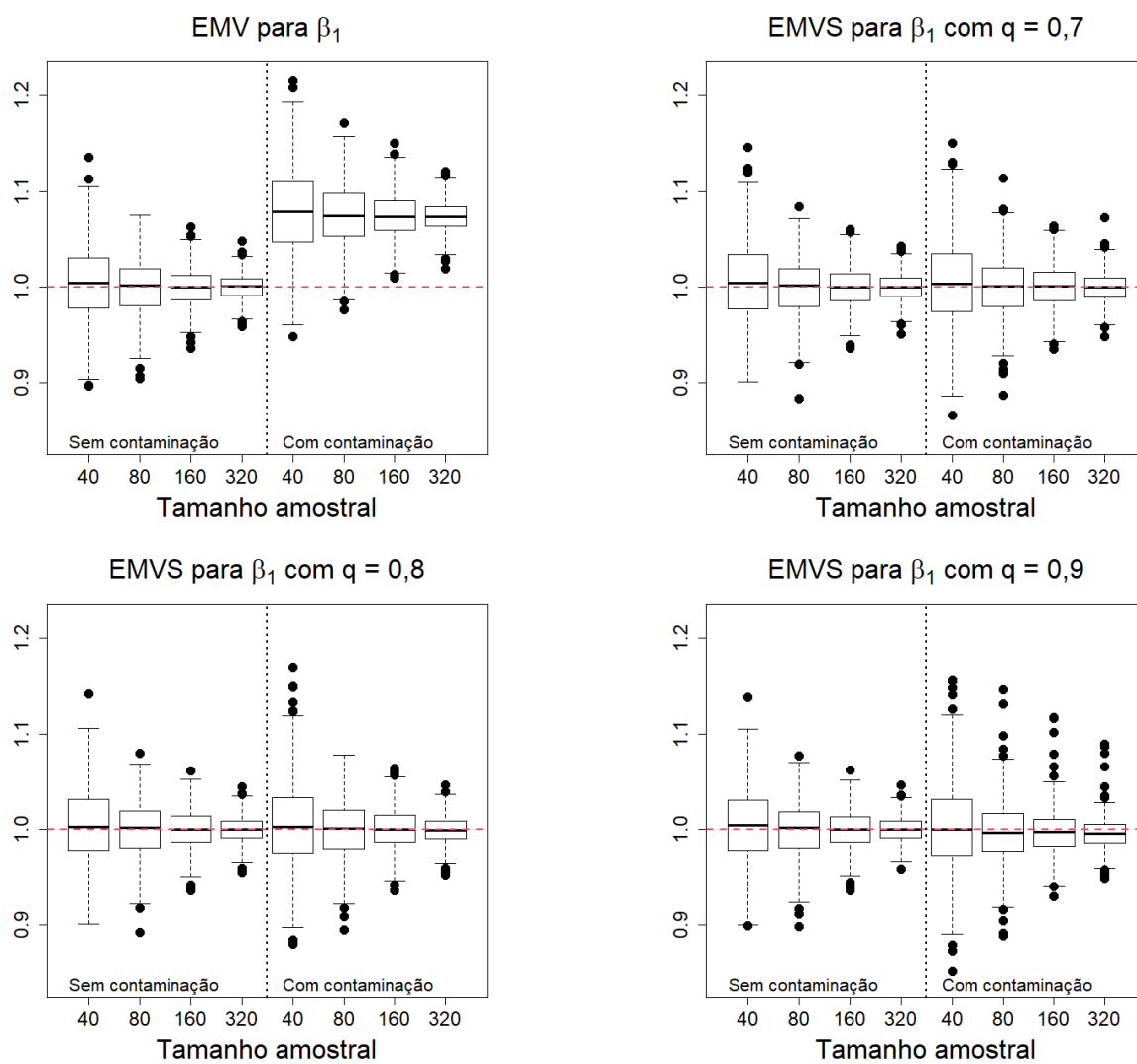
**Figura 5.2:** *boxplots* das estimativas obtidas para o parâmetro  $\beta_1$  sob o Cenário 1 via o EMV e via o EMVS variando  $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro.



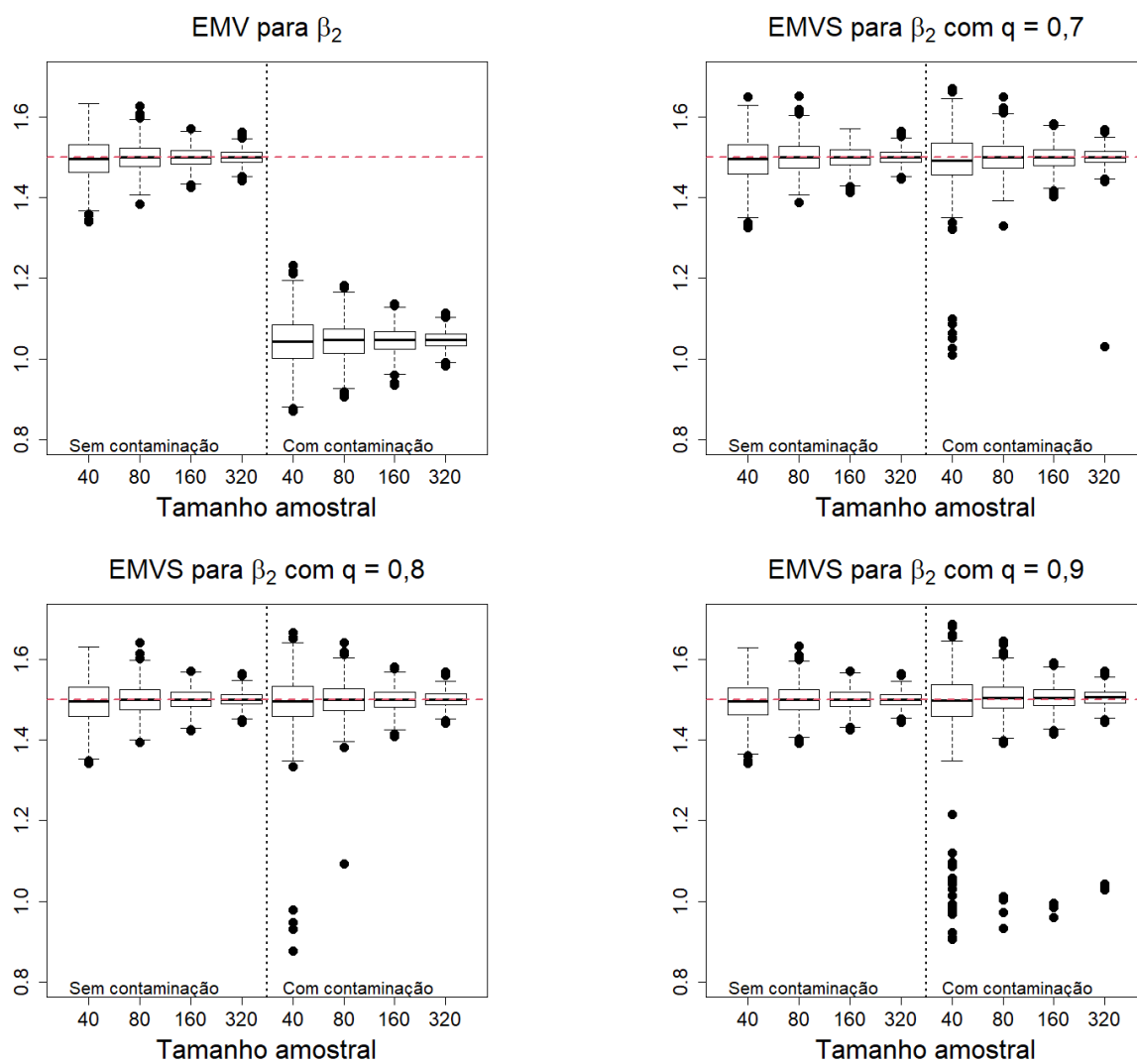
**Figura 5.3:** *boxplots* das estimativas obtidas para o parâmetro  $\beta_2$  sob o Cenário 1 via o EMV e via o EMVS variando  $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro.



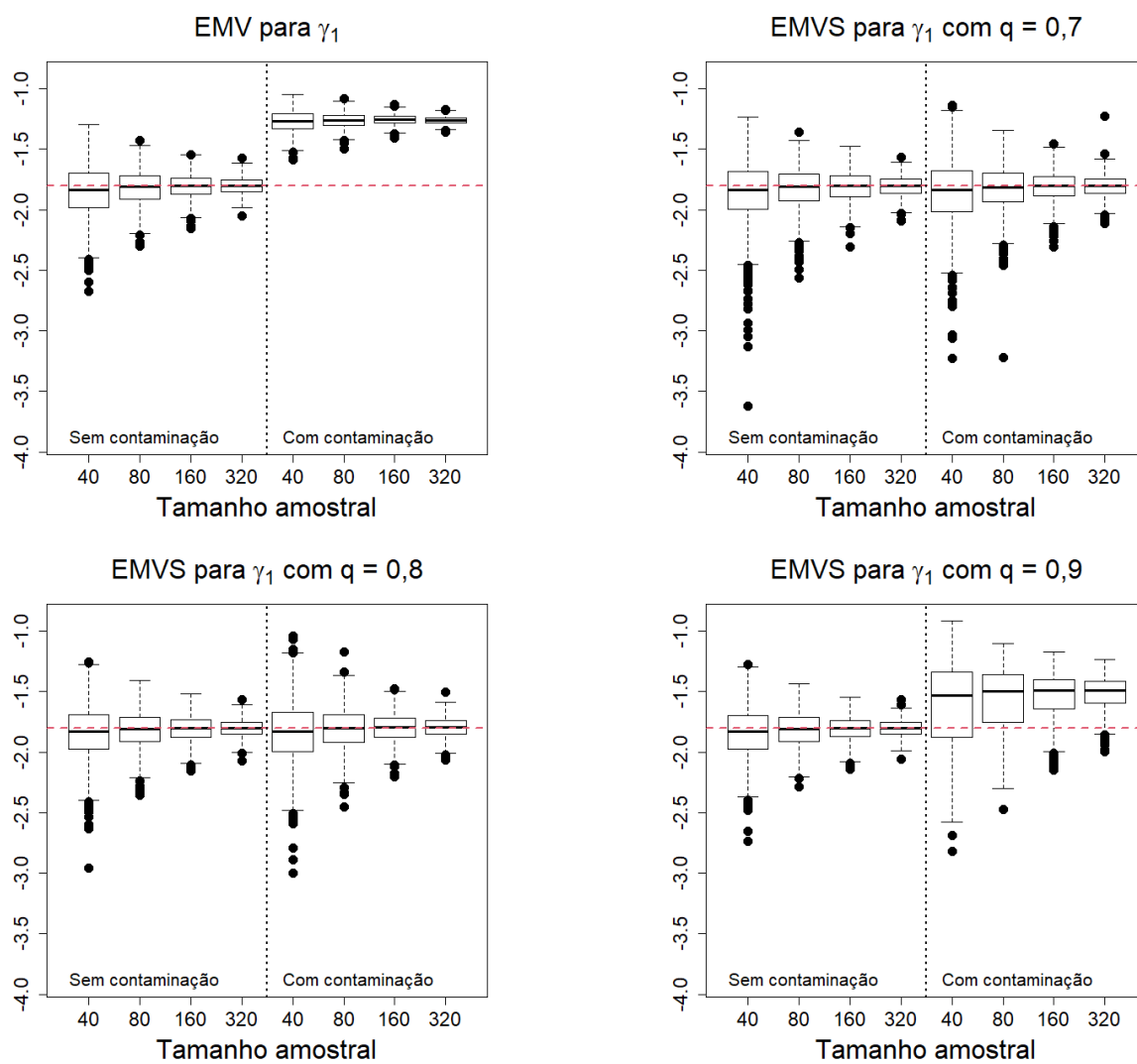
**Figura 5.4:** boxplots das estimativas obtidas para o parâmetro  $\gamma_1$  sob o Cenário 1 via o EMV e via o EMVS variando  $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro.



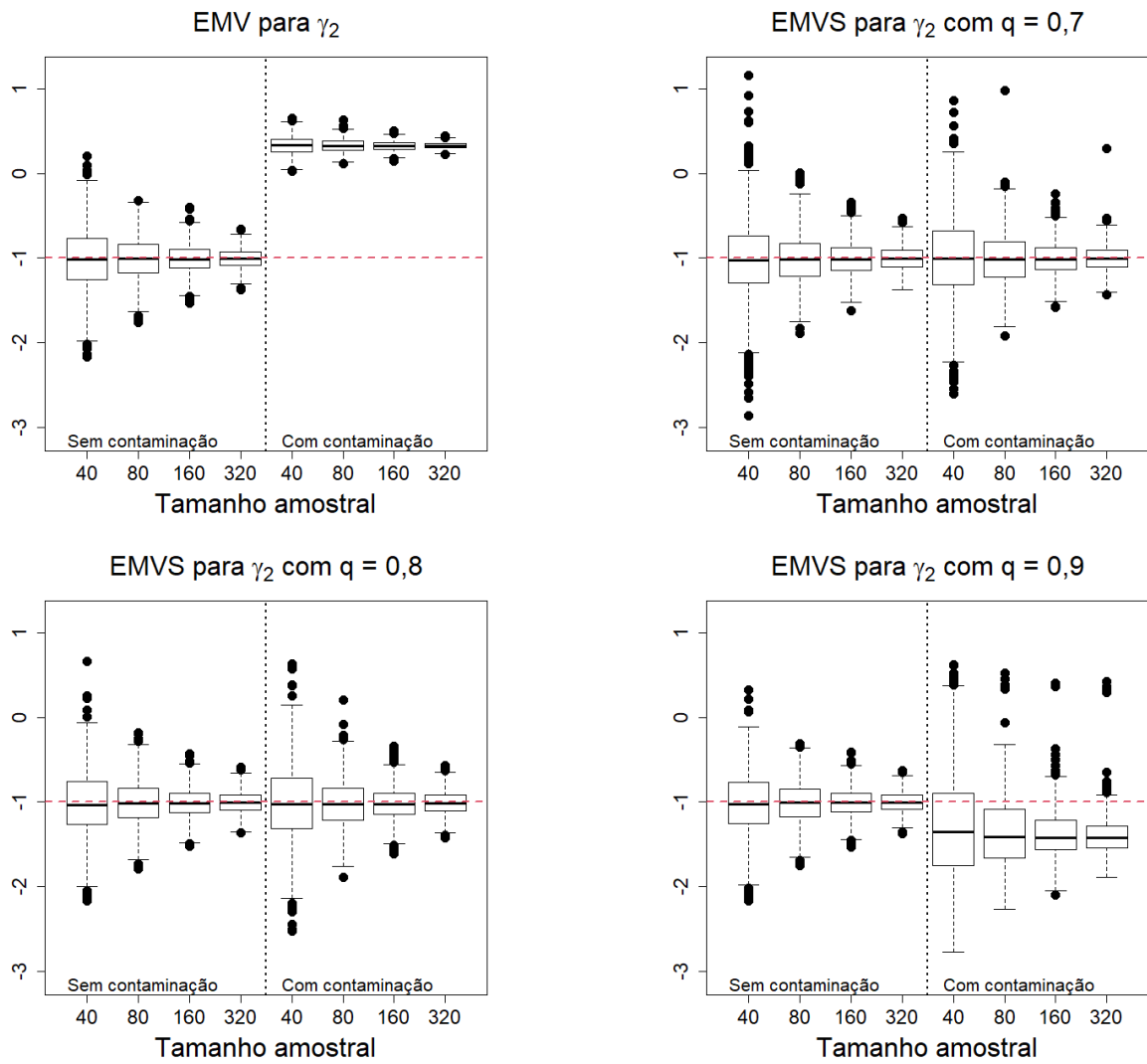
**Figura 5.5:** *boxplots* das estimativas obtidas para o parâmetro  $\beta_1$  sob o Cenário 2 via o EMV e via o EMVS variando  $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro.



**Figura 5.6:** boxplots das estimativas obtidas para o parâmetro  $\beta_2$  sob o Cenário 2 via o EMV e via o EMVS variando  $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro.

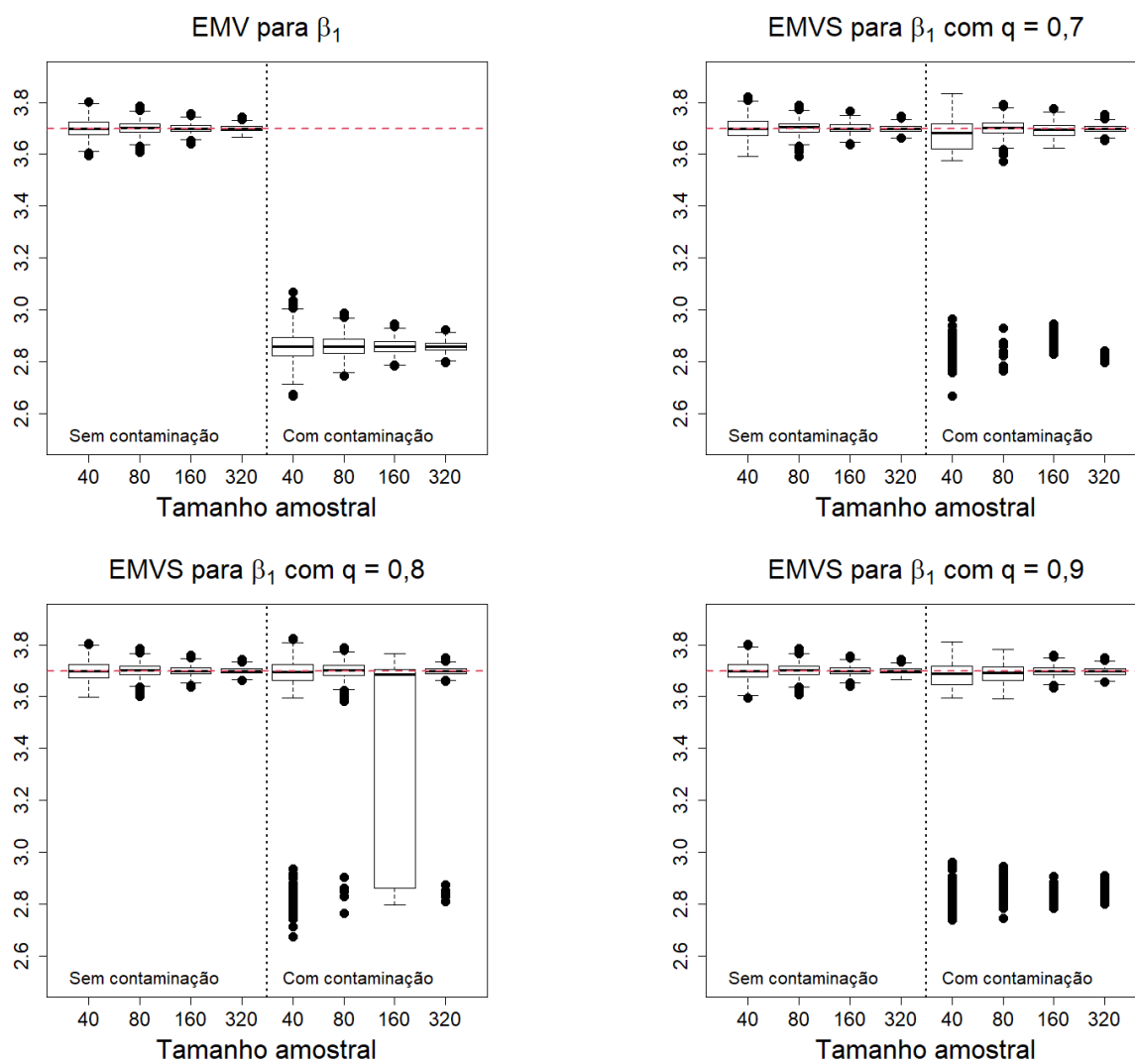


**Figura 5.7:** *boxplots* das estimativas obtidas para o parâmetro  $\gamma_1$  sob o Cenário 2 via o EMV e via o EMVS variando  $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro.

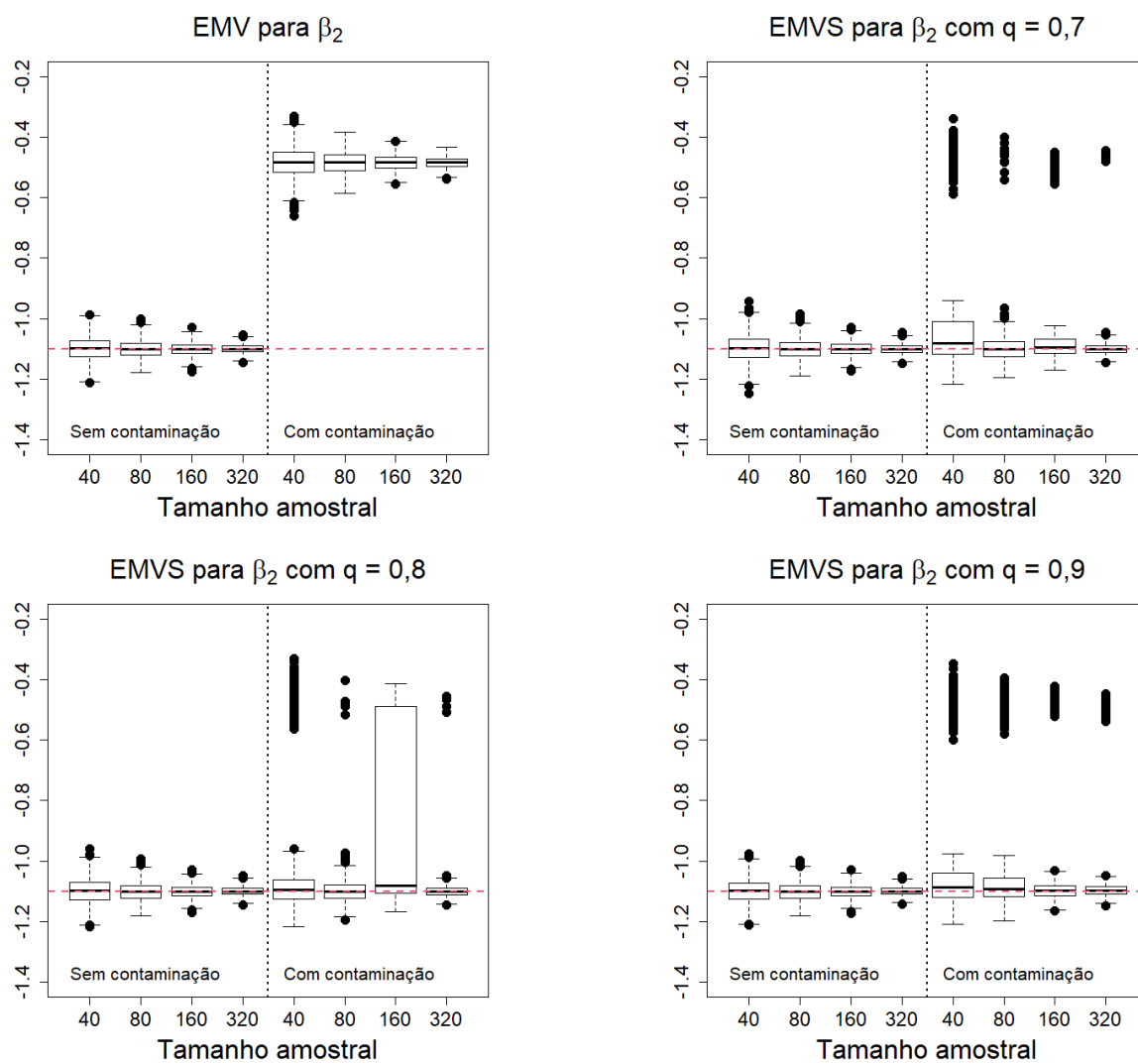


**Figura 5.8:** boxplots das estimativas obtidas para o parâmetro  $\gamma_2$  sob o Cenário 2 via o EMV e via o EMVS variando  $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro.

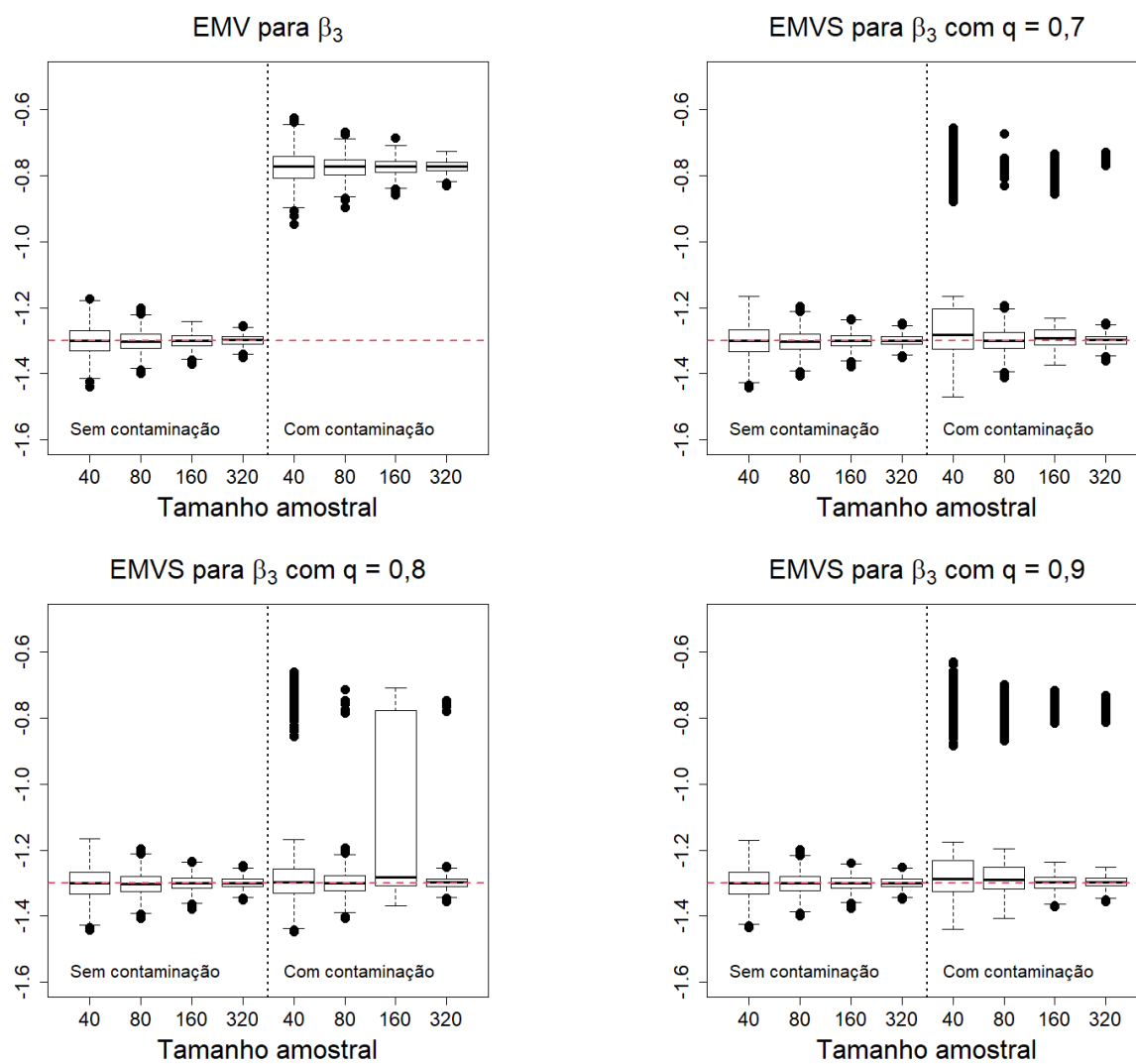




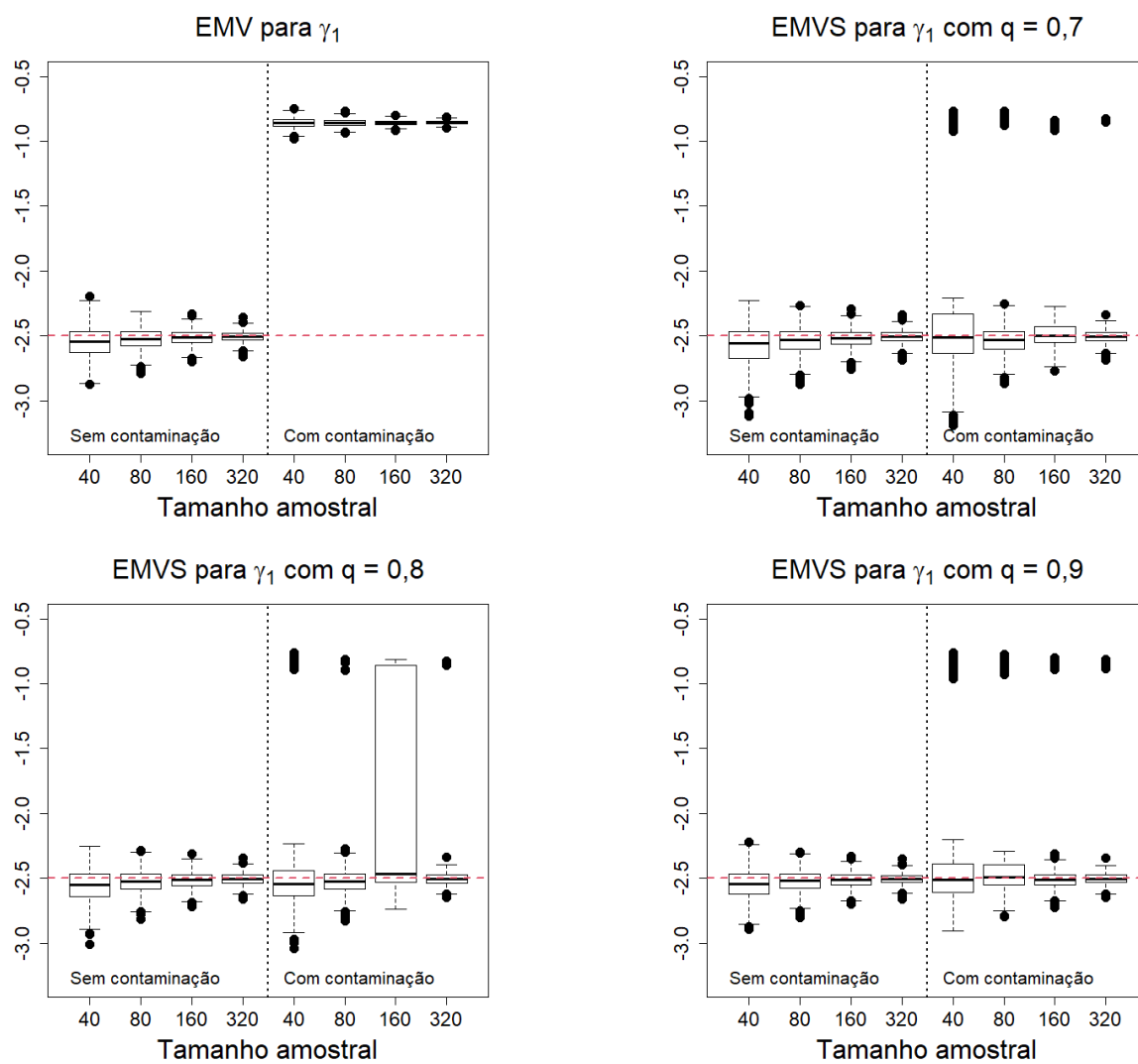
**Figura 5.9:** *boxplots* das estimativas obtidas para o parâmetro  $\beta_1$  sob o Cenário 3 via o EMV e via o EMVS variando  $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro.



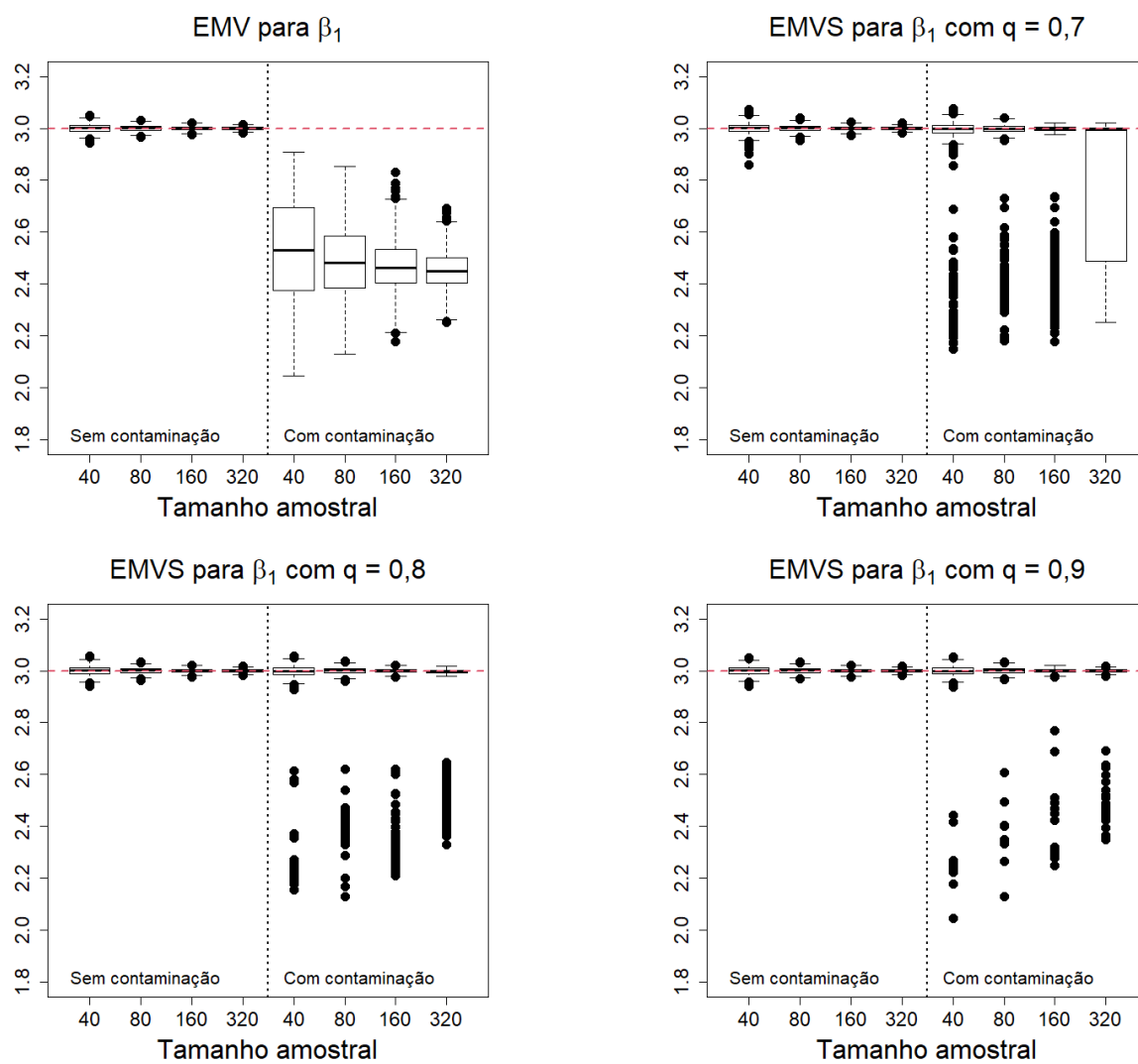
**Figura 5.10:** *boxplots* das estimativas obtidas para o parâmetro  $\beta_2$  sob o Cenário 3 via o EMV e via o EMVS variando  $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro.



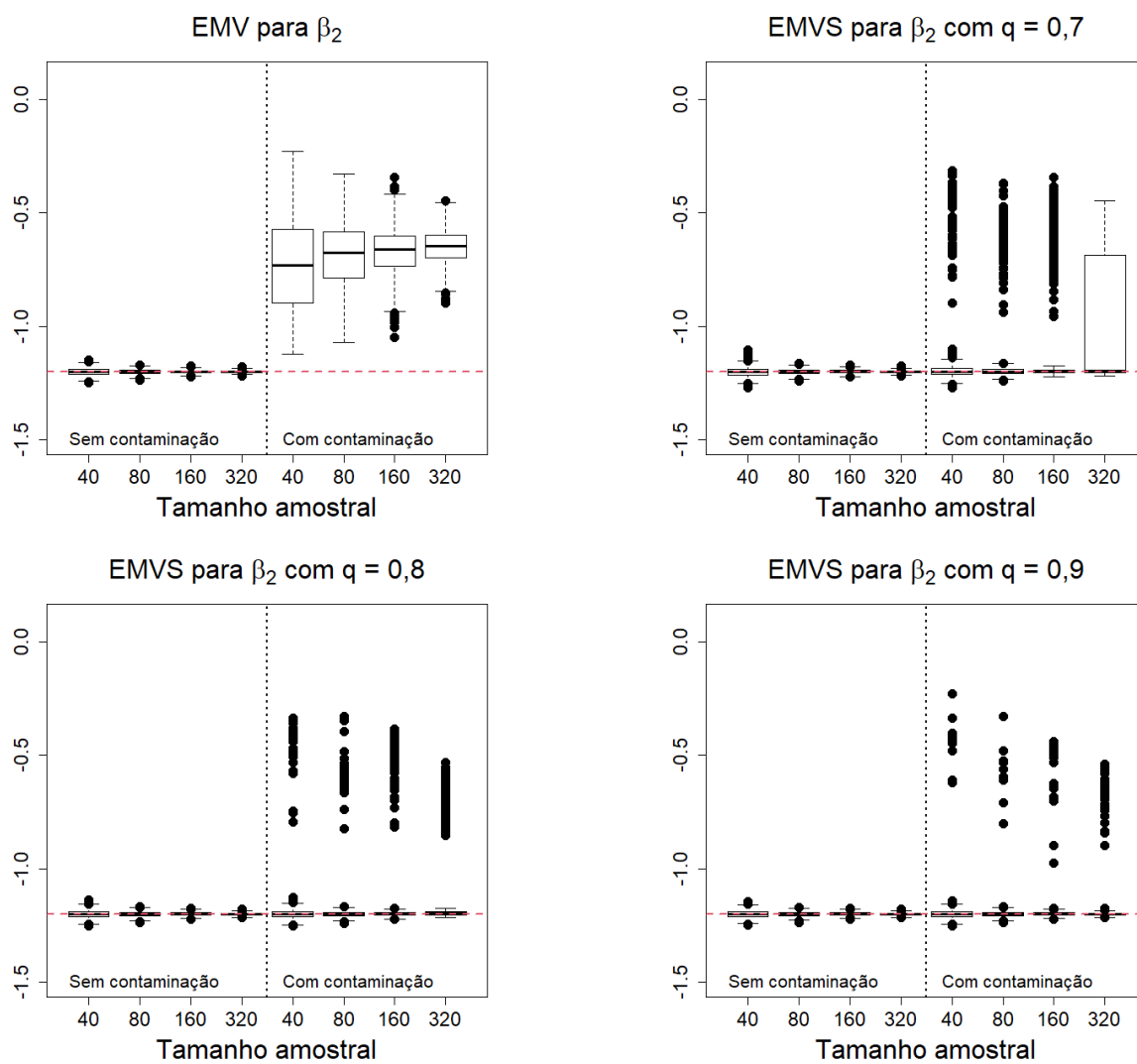
**Figura 5.11:** *boxplots* das estimativas obtidas para o parâmetro  $\beta_3$  sob o Cenário 3 via o EMV e via o EMVS variando  $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro.



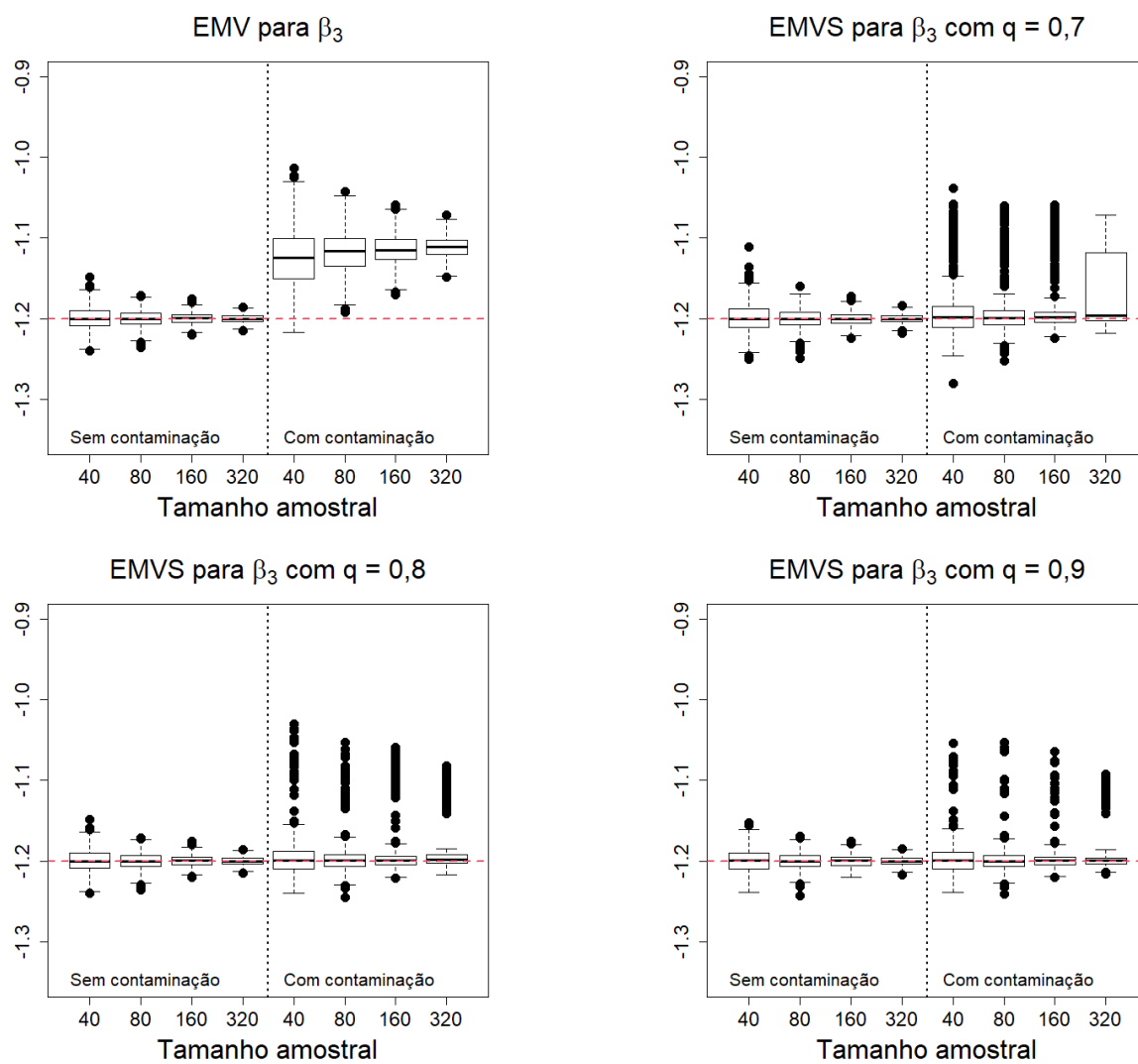
**Figura 5.12:** *boxplots* das estimativas obtidas para o parâmetro  $\gamma_1$  sob o Cenário 3 via o EMV e via o EMVS variando  $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro.



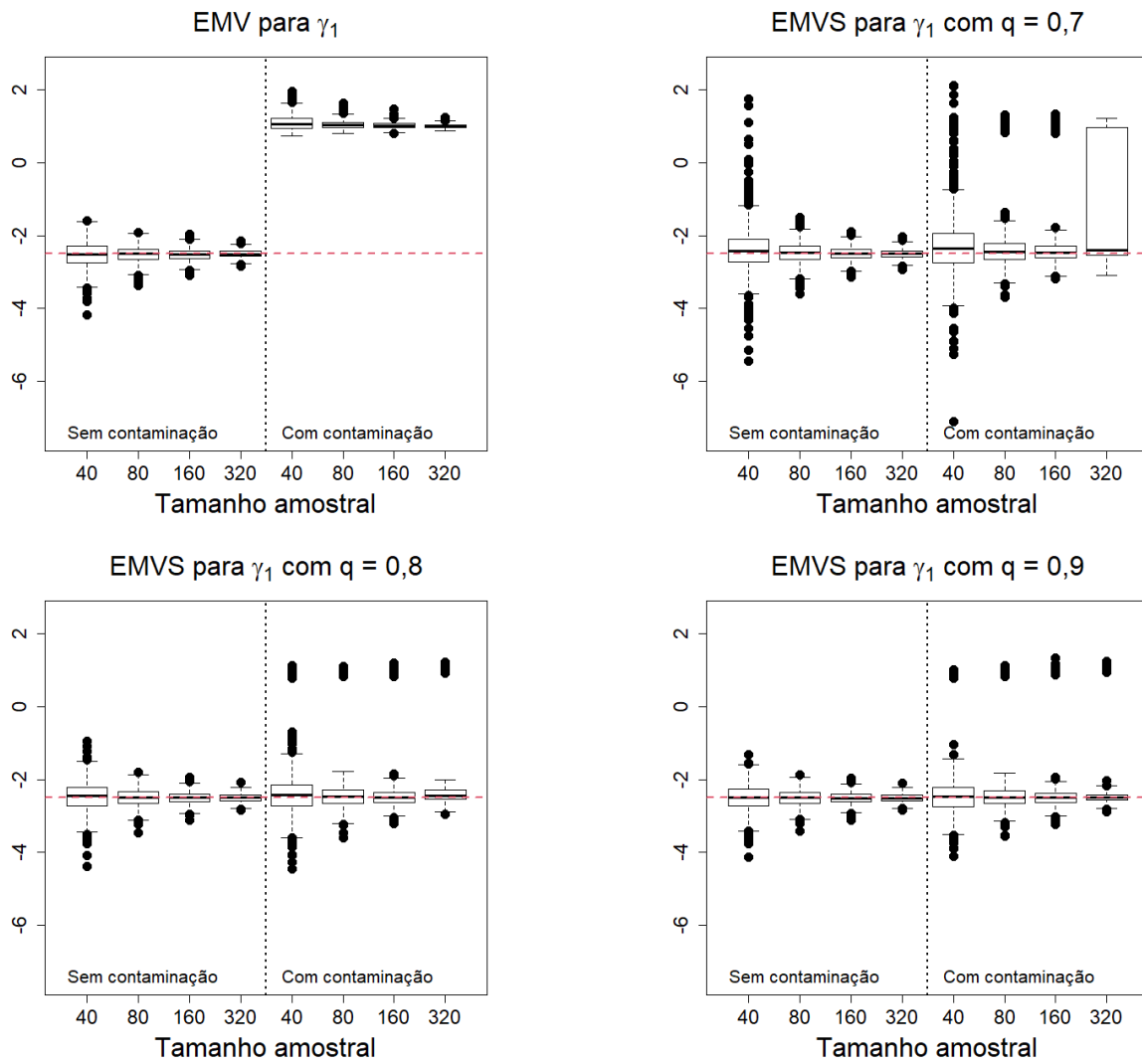
**Figura 5.13:** *boxplots* das estimativas obtidas para o parâmetro  $\beta_1$  sob o Cenário 4 via o EMV e via o EMVS variando  $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro.



**Figura 5.14:** *boxplots* das estimativas obtidas para o parâmetro  $\beta_2$  sob o Cenário 4 via o EMV e via o EMVS variando  $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro.

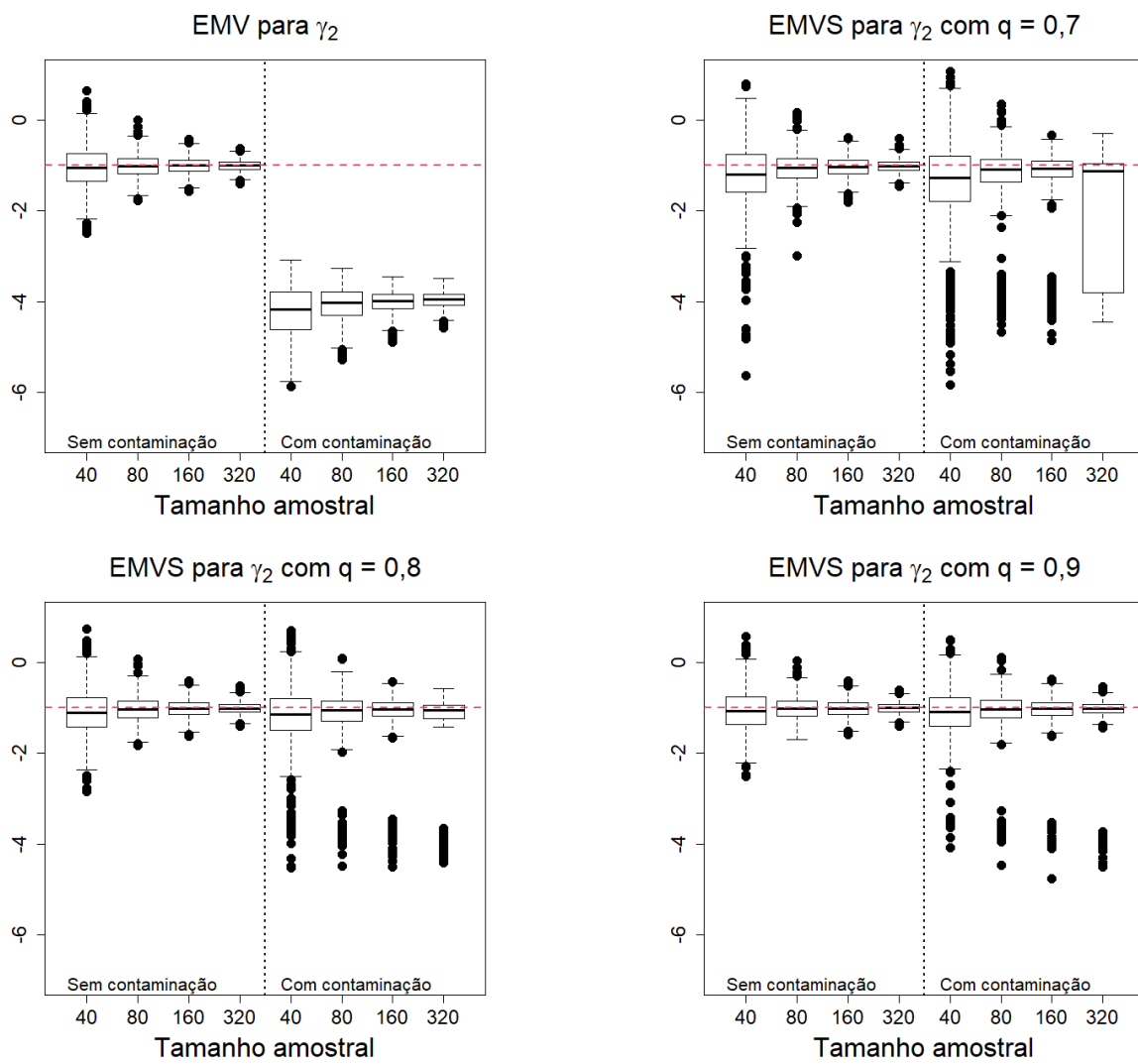


**Figura 5.15:** *boxplots* das estimativas obtidas para o parâmetro  $\beta_3$  sob o Cenário 4 via o EMV e via o EMVS variando  $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro.

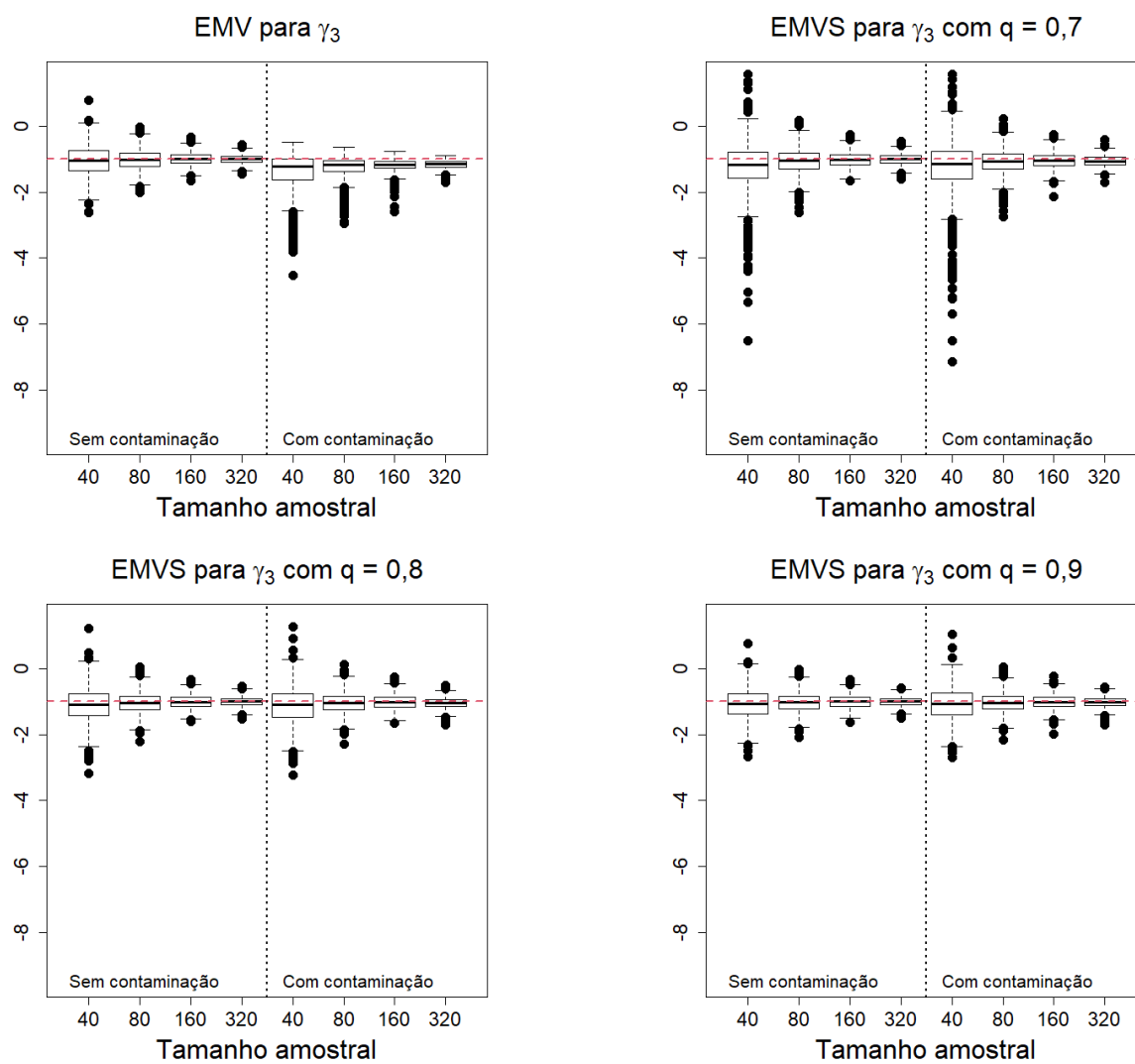


**Figura 5.16:** *boxplots* das estimativas obtidas para o parâmetro  $\gamma_1$  sob o Cenário 4 via o EMV e via o EMVS variando  $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro.





**Figura 5.17:** *boxplots* das estimativas obtidas para o parâmetro  $\gamma_2$  sob o Cenário 4 via o EMV e via o EMVS variando  $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro.



**Figura 5.18:** *boxplots* das estimativas obtidas para o parâmetro  $\gamma_3$  sob o Cenário 4 via o EMV e via o EMVS variando  $q$ . A linha vermelha tracejada representa o verdadeiro valor do parâmetro.

**Tabela 5.1:** Razão entre os EQMTs dos estimadores sob os Cenários 1, 2, 3 e 4.

<b>Cenário 1</b>						
$n$	<b>Ausência de contaminação</b>			<b>Presença de contaminação</b>		
	$\frac{EMV}{EMVS}$	$\frac{EMV}{EMVS}$	$\frac{EMV}{EMVS}$	$\frac{EMV}{EMVS}$	$\frac{EMV}{EMVS}$	$\frac{EMV}{EMVS}$
	$q = 0,7$	$q = 0,8$	$q = 0,9$	$q = 0,7$	$q = 0,8$	$q = 0,9$
40	0,70	0,87	0,97	16,94	13,30	19,64
80	0,75	0,88	0,97	5,79	10,26	35,89
160	0,78	0,91	0,98	152,17	2,72	14,96
320	0,76	0,89	0,97	17,30	13,88	24,80

<b>Cenário 2</b>						
$n$	<b>Ausência de contaminação</b>			<b>Presença de contaminação</b>		
	$\frac{EMV}{EMVS}$	$\frac{EMV}{EMVS}$	$\frac{EMV}{EMVS}$	$\frac{EMV}{EMVS}$	$\frac{EMV}{EMVS}$	$\frac{EMV}{EMVS}$
	$q = 0,7$	$q = 0,8$	$q = 0,9$	$q = 0,7$	$q = 0,8$	$q = 0,9$
40	0,63	0,89	0,98	6,88	8,22	3,75
80	0,69	0,88	0,97	17,00	19,76	5,45
160	0,71	0,86	0,97	38,00	43,00	6,80
320	0,73	0,87	0,96	74,22	87,74	7,45

<b>Cenário 3</b>						
$n$	<b>Ausência de contaminação</b>			<b>Presença de contaminação</b>		
	$\frac{EMV}{EMVS}$	$\frac{EMV}{EMVS}$	$\frac{EMV}{EMVS}$	$\frac{EMV}{EMVS}$	$\frac{EMV}{EMVS}$	$\frac{EMV}{EMVS}$
	$q = 0,7$	$q = 0,8$	$q = 0,9$	$q = 0,7$	$q = 0,8$	$q = 0,9$
40	0,63	0,86	0,98	4,26	11,83	5,32
80	0,68	0,86	0,97	62,47	126,55	5,19
160	0,72	0,88	0,97	5,82	2,63	23,65
320	0,72	0,86	0,96	123,08	146,33	13,43

<b>Cenário 4</b>						
$n$	<b>Ausência de contaminação</b>			<b>Presença de contaminação</b>		
	$\frac{EMV}{EMVS}$	$\frac{EMV}{EMVS}$	$\frac{EMV}{EMVS}$	$\frac{EMV}{EMVS}$	$\frac{EMV}{EMVS}$	$\frac{EMV}{EMVS}$
	$q = 0,7$	$q = 0,8$	$q = 0,9$	$q = 0,7$	$q = 0,8$	$q = 0,9$
40	0,31	0,74	0,93	7,39	17,52	26,51
80	0,60	0,83	0,96	13,22	21,38	49,60
160	0,68	0,86	0,97	9,33	20,24	42,94
320	0,69	0,86	0,96	3,21	4,76	20,42

# Capítulo 6

## Conclusões

### 6.1 Considerações Finais

Neste trabalho foi realizada uma revisão sobre os modelos de regressão gama inverso no contexto de modelagem de dados contínuos estritamente positivos que apresentam forte assimetria à direita. A parametrização da distribuição gama inversa que foi considerada depende de dois parâmetros denotados por  $\mu$  e  $\sigma$  que representam moda e forma da distribuição, respectivamente, assim como previamente apresentado por Rigby et al. (2019). A modelagem de ambos os parâmetros da distribuição gama inversa foi definida por meio de estruturas de regressão lineares, e a tradicional inferência pelo método de máxima verossimilhança foi discutida.

A partir da análise da função de influência, foi demonstrado que o procedimento de estimação por máxima verossimilhança sob modelos de regressão gama inverso não é robusto, isto é, o EMV é influenciado pela presença de *outliers* nos dados. Este resultado teórico motivou o desenvolvimento de um novo procedimento de estimação para os parâmetros do modelo de regressão gama inverso que seja robusto na presença de *outliers*. Vale ressaltar que, de acordo com nosso conhecimento, não existem propostas de estimação robusta sob regressão gama inversa.

O procedimento inferencial robusto proposto neste trabalho foi baseado nas ideias propostas

por Ribeiro e Ferrari (2023) sob modelos de regressão beta. O novo estimador denominado de EMVS é um M-estimador que contém o EMV como caso particular, e sua construção parte da ideia de substituir o logaritmo da função de verossimilhança por uma função alternativa denominada de função  $L_q$ -verossimilhança reparametrizada. Este procedimento depende de uma constante de afinação  $0 < q \leq 1$  que controla a troca entre eficiência e robustez do estimador. Aqui, a inferência proposta foi desenvolvida fixando valores para a constante  $q$  assim como feito por Ghosh (2019) sob regressão beta, considerando um estimador robusto similar ao EMVS.

Para avaliar o desempenho do EMVS comparado ao EMV na presença e ausência de contaminação nos dados, foram realizados estudos de simulação de Monte Carlo. A partir dos resultados apresentados, foi ilustrada a clara vantagem do uso do procedimento robusto em relação ao método de estimação tradicional na presença de *outliers*. Na ausência de contaminação, o EMV e o EMVS apresentaram comportamentos similares, com o estimador robusto apresentando uma maior variância para amostras pequenas. Entretanto, na presença de contaminação, o EMV apresentou altos vieses, enquanto o EMVS se manteve centrado em torno dos verdadeiros valores dos parâmetros na maioria dos casos. Ainda, observou-se que o valor de  $q = 0,7$  se mostrou mais adequado para garantir estabilidade das estimativas na grande maioria dos cenários considerados.

Por fim, destacam-se as limitações e possíveis extensões dos resultados aqui obtidos. Primeiro, neste trabalho, o valor da constante de afinação  $q$  foi fixado. Do ponto de vista prático, é de interesse que o valor  $q$  seja escolhido com base nos dados. Nesse sentido, o próximo passo é propor a escolha do valor de  $q$  de forma automática, como por exemplo, adaptando o método orientado pelos dados proposto por Ribeiro e Ferrari (2023) sob regressão beta. Segundo, é de suma importância mensurar os erros-padrão dos estimadores robustos. Assim, se torna importante a obtenção da matriz de covariâncias assintótica do EMVS com base nas propriedades dos M-estimadores ou ainda por métodos de reamostragem como *bootstrap* paramétrico. Por fim, a aplicação do EMVS sob regressão gama inversa considerando conjuntos de dados reais é rele-

vante para ilustração da aplicabilidade do novo estimador. Estas extensões foram inicialmente consideradas neste trabalho, entretanto, por delimitação de tempo, não foram incluídas.

# Apêndice A

## Resultados Capítulo 4

### A.1 Transformação Potência

Aqui será mostrado que o modelo gama inverso definido em (2.9) é fechado sob a transformação potência sob certas condições.

A função densidade de probabilidade é dada por

$$f_{\theta}(y; \mu, \sigma) = \frac{[\mu(1 + \sigma^{-2})]^{\sigma^{-2}}}{\Gamma(\sigma^{-2})} y^{-(\sigma^{-2}+1)} e^{-\frac{\mu(1+\sigma^{-2})}{y}}, y > 0. \quad (\text{A.1})$$

Para  $\alpha > 0$ , tem-se que

$$f_{\theta}(y; \mu, \sigma)^{\alpha} = \frac{[\mu(1 + \sigma^{-2})]^{\alpha\sigma^{-2}}}{[\Gamma(\sigma^{-2})]^{\alpha}} y^{-\alpha(\sigma^{-2}+1)} e^{-\frac{\alpha\mu(1+\sigma^{-2})}{y}}.$$

Daí,

$$\int_0^{\infty} f_{\theta}(y; \mu, \sigma)^{\alpha} dy = \frac{[\mu(1 + \sigma^{-2})]^{\alpha\sigma^{-2}}}{[\Gamma(\sigma^{-2})]^{\alpha}} \int_0^{\infty} y^{-\alpha(\sigma^{-2}+1)} e^{-\frac{\alpha\mu(1+\sigma^{-2})}{y}} dy.$$

Seja  $\sigma_\alpha^{-2} = \alpha(\sigma^{-2} + 1) - 1$ ,

$$\begin{aligned} \int_0^\infty f_\theta(y; \mu, \sigma)^\alpha dy &= \frac{[\mu(1 + \sigma^{-2})]^{\alpha\sigma^{-2}}}{[\Gamma(\sigma^{-2})]^\alpha} \int_0^\infty y^{-[\alpha(\sigma^{-2}+1)-1+1]} e^{-\frac{\mu[\alpha(1+\sigma^{-2})-1+1]}{y}} dy \\ &= \frac{[\mu(1 + \sigma^{-2})]^{\alpha\sigma^{-2}}}{[\Gamma(\sigma^{-2})]^\alpha} \int_0^\infty y^{-(\sigma_\alpha^{-2}+1)} e^{-\frac{\mu(1+\sigma_\alpha^{-2})}{y}} dy. \end{aligned}$$

Como  $\sigma_\alpha^{-2} = \alpha(\sigma^{-2} + 1) - 1 \Leftrightarrow \sigma_\alpha = 1/\sqrt{\alpha(\sigma^{-2} + 1) - 1}$ , com  $\alpha > 0$  e  $\sigma > 0$ . Daí,

$$\begin{aligned} \int_0^\infty f_\theta(y; \mu, \sigma)^\alpha dy &= \frac{[\mu(1 + \sigma^{-2})]^{\alpha\sigma^{-2}}}{[\Gamma(\sigma^{-2})]^\alpha} \frac{\Gamma(\sigma_\alpha^{-2})}{[\mu(1 + \sigma_\alpha^{-2})]^{\sigma_\alpha^{-2}}} \int_0^\infty \frac{[\mu(1 + \sigma_\alpha^{-2})]^{\sigma_\alpha^{-2}}}{\Gamma(\sigma_\alpha^{-2})} y^{-(\sigma_\alpha^{-2}+1)} e^{-\frac{\mu(1+\sigma_\alpha^{-2})}{y}} dy \\ &= \frac{[\mu(1 + \sigma^{-2})]^{\alpha\sigma^{-2}}}{[\Gamma(\sigma^{-2})]^\alpha} \frac{\Gamma(\sigma_\alpha^{-2})}{[\mu(1 + \sigma_\alpha^{-2})]^{\sigma_\alpha^{-2}}} \int_0^\infty f_\theta(y; \mu, \sigma_\alpha) dy \\ &= \frac{[\mu(1 + \sigma^{-2})]^{\alpha\sigma^{-2}}}{[\Gamma(\sigma^{-2})]^\alpha} \frac{\Gamma(\sigma_\alpha^{-2})}{[\mu(1 + \sigma_\alpha^{-2})]^{\sigma_\alpha^{-2}}}, \end{aligned}$$

desde que a densidade  $f_\theta(y; \mu, \sigma_\alpha)$  esteja bem definida para  $\alpha > 0$ . Logo, Pela definição (4.3)

$$\begin{aligned} f_\theta^{(\alpha)}(y; \mu, \sigma) &= \frac{f_\theta(y; \mu, \sigma)^\alpha}{\int f_\theta(y; \mu, \sigma)^\alpha dy} \\ &= \frac{[\mu(1 + \sigma_\alpha^{-2})]^{\sigma_\alpha^{-2}}}{\Gamma(\sigma_\alpha^{-2})} y^{-[\alpha(\sigma^{-2}+1)-1+1]} e^{-\frac{\mu[\alpha(1+\sigma^{-2})-1+1]}{y}} dy \\ &= \frac{[\mu(1 + \sigma_\alpha^{-2})]^{\sigma_\alpha^{-2}}}{\Gamma(\sigma_\alpha^{-2})} y^{-(\sigma_\alpha^{-2}+1)} e^{-\frac{\mu(\sigma_\alpha^{-2}+1)}{y}} dy \end{aligned}$$

Assim, a densidade gama inversa de parâmetros  $\mu$  e  $\sigma$  com  $\mu = \text{moda}(y)$  é fechada sob a transformação potência, isto é,  $y^{(\alpha)} \sim \text{GI}(\mu, \sigma_\alpha)$ . Para obter o EMVS, deve-se considerar  $\alpha = 1/q$ , de onde segue que  $\sigma_{1/q} = 1/\sqrt{q^{-1}(\sigma^{-2} + 1) - 1}$ . A condição necessária para que o estimador esteja bem definido é que  $\sigma_{1/q} > 0$ , sendo  $0 < q \leq 1$  e  $\sigma > 0$ . Mas  $\sigma_{1/q} > 0 \Leftrightarrow \sqrt{(1/q)(\sigma^{-2} + 1) - 1} > 0 \Leftrightarrow (1/q)(\sigma^{-2} + 1) > 1 \Leftrightarrow \sigma^{-2} + 1 > q \Leftrightarrow \sigma^{-2} > q - 1$ . Para  $0 < q \leq 1$ , a condição  $\sigma^{-2} > q - 1$  sempre é válida, desde que  $\sigma^{-2} > 0$ , para  $q = 0$  deve-se ter  $\sigma^{-2} > -1$  e para  $q = 1$  deve-se ter  $\sigma^{-2} > 0$ . Este resultado implica que o EMVS sob a distribuição gama inversa reparametrizada estará bem definido para todo  $0 < q \leq 1$ .



## Referências Bibliográficas

- Basu, A.; Harris, I. R., Hjort, N. L.; e Jones, M. C. (1998). “Robust and efficient estimation by minimising a density power divergence”. *Biometrika*, Oxford University, v.85, n.3, p.549–559.
- Bourguignon, M.; e Gallardo, D. I. (2020). “Reparameterized inverse gamma regression models with varying precision”. *Statistica Neerlandica*, v.74, n.4, p.611–627.
- Box, G. E. P.; e Cox, D. R. (1964). “An analysis of transformations”. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, Oxford University, v.26, n.2, p.211–243.
- Casella, G.; e Berger, R. (2024). *Statistical inference*. Chapman e Hall/CRC.
- Colosimo, E. A.; e Giolo, S. R. (2021). *Análise de Sobrevivência Aplicada*. 2.ed. São Paulo: Editora Blucher.
- Fasano, M. V.; Maronna, R. A.; Sued, M.; e Yohai, V. (2012). “Continuity and differentiability of regression M functionals”.
- Feller, W. (1991). *An Introduction to Probability Theory and its Applications*. 2.ed. New York: John Wiley & Sons.
- Ferrari, D.; e La Vecchia, D. (2012). “On robust estimation via pseudo-additive information”. *Biometrika*, Oxford University, v.99, n.1, p.238–244.
- Ferrari, D.; e Yang, Y. (2010). “Maximum lq-likelihood method”. *Annals of Statistics*, New York, v.38, n.2, p.573–583.

- Genton, M. G.; e Rousseeuw, P. J. (1995). “The change-of-variance function of M-estimators of scale under general contamination”. *Journal of Computational and Applied Mathematics*, Amsterdam, v.64,n.1-2, p.69–80.
- Ghosh, A. (2019). “Robust inference under the beta regression model with application to health care studies”. *Statistical Methods in Medical Research*, London, v.28,n.3, p.871–888.
- Glen, A. G. (2017). “On the inverse gamma as a survival distribution”. *Journal of Quality Technology*, Oxfordshire, v.43,n.2, pp. 158–166.
- Hampel, F. R. (1974). “The influence curve and its role in robust estimation”. *Journal of the American Statistical Association*, Oxfordshire, v.69,n.346, p.383–393.
- Hampel, F. R.; Ronchetti, E. M.; Rousseeuw, P. J.; e Stahel, W. A. (2011). *Robust Statistics: The Approach Based on Influence Functions*. New York: John Wiley & Sons.
- Huber, P. J. (1964). “Robust estimation of a location parameter”. *Ann. Math. Statist*, Beachwood, v.35,n.4, p.73–101.
- Kallianpur, G.; e Rao, C. R. (1955). “On fisher’s lower bound to asymptotic variance of a consistent estimate”. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, Kolkata, v.15,n.4, p.331–342.
- La Vecchia, D.; Camponovo, L.; e Ferrari, D. (2015). “Robust heart rate variability analysis by generalized entropy minimization”. *Computational Statistics & Data Analysis*, Amsterdam, v.82,n.1, p.137–151.
- Lin, C. T.; Duran, B. S.; e Lewis, T. O. (1989). “Inverted gamma as a life distribution”. *Microelectronics Reliability*, Amsterdam, v.29,n.4, p.619–626.
- Magalhães, T. M.; Gallardo, D. I. e Bourguignon, M. (2021). “Improved point estimation for inverse gamma regression models”. *Journal of Statistical Computation and Simulation*, Oxfordshire, v.91,n.12, p.2444–2456.
- Maronna, R. A.; Martin, R. D.; Yohai, V. J. e Salibián-Barrera, M. (2019). *Robust Statistics: Theory and Methods (with R)*. New York: John Wiley & Sons.

- Mead, M. E. (2015). “Generalized inverse gamma distribution and its application in reliability”. *Communications in Statistics-Theory and Methods*, Oxfordshire, v.44,n.7, p.1426–1435.
- Montgomery, D. C.; Peck, E. A.; e Vining, G. G. (2021). *Introduction to linear regression analysis*. 5.ed. John Wiley & Sons.
- Mousa, A. M.; Hassan, M. K.; e Fathi, A. (2016). “A new non parametric estimator for Pdf based on inverse gamma distribution”. *Communications in Statistics-Theory and Methods*, Oxfordshire, v.45,n.23, p.7002–7010.
- Paula, G. A. (2023). *Modelos de Regressão: com Apoio Computacional*. 2.ed.São Paulo: Instituto de Matemática e Estatística – USP.
- Ribeiro, T. K. A. (2020). “Regressão Beta Robusta. Tese (Doutorado)”. *Universidade de São Paulo*.
- Ribeiro, T. K. A.; e Ferrari, S. L. P. (2023). “Robust estimation in beta regression via maximum L q-likelihood”. *Statistical Papers*, New York, v.64,n.1, p.321–353.
- Rigby, R. A.; Stasinopoulos, M. D.; Heller, G. Z.; e De Bastiani, F. (2019). *Distributions for Modeling Location, Scale, and Shape: Using GAMLSS in R*. New York: Chapman e Hall/CRC Press.
- Rousseeuw, P. J. (1981). “A new infinitesimal approach to robust estimation”. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, Oxfordshire, v.56,n.1, p.127–132.
- Rousseeuw, P. J.; e Leroy, A. M. (2003). *Robust regression and outlier detection*. John wiley & sons.
- Serfling, R. J. (2009). *Approximation theorems of mathematical statistics*. John Wiley & Sons.