

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA
PROGRAMA DE PÓS GRADUAÇÃO EM ESTATÍSTICA

Modelos de Seleção Amostral Heckman Generalizado e
Birnbaum-Saunders

Fernando de Souza Bastos

BELO HORIZONTE
MINAS GERAIS - BRASIL

2018

Fernando de Souza Bastos

**Modelos de Seleção Amostral Heckman Generalizado e
Birnbaum-Saunders**

Tese apresentada ao Programa de Pós-graduação em Estatística do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Doutor em Estatística.

Orientador: Prof. Dr. Wagner Barreto de Souza

BELO HORIZONTE
MINAS GERAIS - BRASIL

2018

Modelos de Seleção Amostral Heckman Generalizado e Birnbaum-Saunders

Esta versão da tese contém as correções e alterações sugeridas pela banca durante a defesa do trabalho realizada em 03 de maio de 2018.

Banca Examinadora:

Prof. Dr. Wagner Barreto de Souza (Orientador) - UFMG

Profa. Dra. Rosangela Helena Loschi - UFMG

Prof. Dr. Vinícius Diniz Mayrink - UFMG

Prof. Dr. Alexandre de Bustamante Simas - UFPB

Prof. Dr. Caio Lucidius Naberezny Azevedo - UNICAMP

Agradecimentos

Sou grato a Deus por estar sempre presente em minha vida e por me abençoar com a oportunidade de poder agradecer a diversas pessoas, familiares e amigos, que sempre me apoiaram e colaboraram, direta ou indiretamente, para a concretização deste sonho. Assim, agradeço:

- A minha filha, Ana Beatriz dos Santos Bastos. Ela me ensina, todos os dias, que o amor não tem limites e que haverá sempre motivos para lutar por um mundo melhor;
- A minha esposa, Elizete Sélvio dos Santos Bastos. Obrigado por sua dedicação, confiança e presença em minha vida. Esta vitória é nossa!
- Ao meu orientador, Wagner Barreto de Souza. Sua competência teórica, seu apoio, amizade e disponibilidade são motivos de incentivo para todos, para mim foi uma grande honra ter trabalhado com você. Muito obrigado!
- Aos meus pais, que sempre me apoiaram e são incentivos para meu crescimento pessoal e profissional;
- A minha tia Tereza. Sua confiança, seu apoio e também seu investimento, com certeza, foram essenciais para que eu realizasse mais esse sonho. Muito obrigado!
- Ao meu irmão, Renato Luiz, que mostra todos os dias que dedicação é, em grande medida, o ingrediente para o sucesso.
- A minha avó, Maria de Lourdes Leite de Souza (in memoriam). Ela foi e sempre será meu maior exemplo de vida.
- A toda a minha família, tios e tias, primos e primas, todos muito presentes em minha vida. Obrigado a todos!
- Aos meus respectivos orientadores de iniciação científica e mestrado, Lucy Tiemi Takahashi e Carlos Henrique Osório Silva. O apoio, a orientação e a amizade de vocês me fizeram ir mais longe, dedico meu trabalho e meu sucesso também a vocês!
- Ao meu amigo e irmão Diogo da Silva Machado. Seu caráter, honestidade e competência sempre foram exemplos para mim;
- Ao meu amigo Rumenick Pereira da Silva. Sou muito grato a todo o seu apoio, sua ajuda e sua amizade. Tenho certeza que seu conhecimento e dedicação farão de você um dos grandes nomes da nossa área. Você vai longe!
- A todos os amigos da pós-graduação, em especial aos amigos Alejandro Guillermo Monzon Montoya, Wagner Pinheiro, Uriel Moreira Silva, Victor Schmidt Comitti e Frederico Machado.

- Aos amigos Alexandre Alvarenga Rocha, Justino Muniz Júnior, Brenno Santos Leite e Sibebe Augusta Leite. Durante esse processo de doutorado, a amizade de vocês foi um grande alento para mim!
- A minha querida amiga Elisângela Aparecida Oliveira (in memoriam). Espero continuar minha jornada, lutando também por alguns dos sonhos que você confiou a mim. Você foi embora muito cedo, mas deixou belos exemplos que ficarão sempre em minha memória e em meu coração!
- A sociedade brasileira. Aproximadamente, 200 milhões de pessoas, entre as quais, 52 milhões que vivem na pobreza. Sou Servidor Público, ligado a rede Federal de Ensino Superior, como tal, fiz uso de um direito descrito no Art. 30 da lei número 12.772/2012. Apesar de ter sido liberado por 48 meses, permaneci afastado por 38, com licença integral das minhas atividades profissionais e com total ônus a **sociedade brasileira**, pagadora de impostos altíssimos e que tem retorno pífio de tais recursos devido a má gestão, a privilégios injustificáveis e a corrupção. Tenho que ressaltar aqui, que nem o governo federal, nem a minha Instituição de origem (Universidade Federal de Viçosa) e também nenhum gestor de tais instituições, fizeram um favor para mim. Muito pelo contrário, usufruí de um direito, pelo qual, agradeço a **sociedade** e prometo a ela, honrar minha formação, sendo honesto, combatendo privilégios e a corrupção e lutando por mais justiça, por igualdade social e de direitos.
- Por fim, a todos que colaboraram direta ou indiretamente para a concretização deste sonho, meu muito obrigado.

Dedico, com todo o meu coração, aos meus pais, Adalberto de Aguiar Bastos e Maria José de Souza Bastos, a minha amada filha, Ana Beatriz dos Santos Bastos, e a minha querida esposa, Elizete Sélvio dos Santos Bastos.

Resumo

Em um processo de amostragem pode existir uma relação entre a variável de interesse e uma variável latente, de tal forma, que a variável de interesse é observada somente em um subconjunto da população sob estudo. Neste caso, dizemos que a variável de interesse está sujeita a um truncamento oculto e que possui problemas de seleção amostral, ou mais especificamente, viés de seleção amostral. Um dos primeiros pesquisadores a estudar modelos para ajustar dados com problemas de viés de seleção amostral foi o matemático e economista, James Joseph Heckman. Ele observou, que tal problema surge quando a variável de interesse é correlacionada com uma variável oculta e que era necessária uma modelagem conjunta destas variáveis, pois o ajuste de modelos de regressão somente para a parte observada da variável de interesse produz resultados viesados e inconsistentes. Nesse contexto, Heckman (1976) introduziu o modelo de seleção amostral, também conhecido como modelo de Heckman, que ajusta conjuntamente a variável de interesse e a variável latente e propôs um método de estimação por máxima verossimilhança sob a suposição de normalidade bivariada. Apesar dos avanços de seu modelo, o método de estimação por máxima verossimilhança logo foi criticado por pesquisadores da época, devido a necessidade da suposição de normalidade bivariada e também devido à dificuldade de sua implementação. Isto motivou Heckman (1979) a introduzir um método alternativo e de fácil implementação para a estimação dos parâmetros do modelo de seleção amostral, denominado método de dois passos. Várias generalizações foram também propostas ao longo dos anos para lidar com situações em que a normalidade é violada. Tais generalizações incluem, no âmbito paramétrico, o uso da distribuição t de Student por Marchenko e Genton (2012) e sua abordagem Bayesiana por Ding (2014), o uso da distribuição Skew-normal por Ogundimu e Hutton (2016) e também uma alternativa robusta para o método de dois passos apresentado por Zhelonkin et al. (2016). Existem também abordagens baseadas em cópula, semi-paramétricas e não-paramétrica. Neste trabalho, introduzimos dois novos modelos paramétricos. No Capítulo 2 acrescentamos covariáveis aos parâmetros de dispersão e correlação do modelo clássico e introduzimos o modelo de Heckman generalizado. Nossas simulações indicam, que na presença de heterocedasticidade e/ou correlação variável, o ajuste dos modelos, Heckman clássico, Skew-normal ou Heckman-t, pode nos levar a perda de eficiência e também a estimação incorreta dos parâmetros. Diferentemente do que ocorre com o ajuste do modelo Heckman generalizado a tais dados. Dessa forma, nosso modelo torna-se uma importante alternativa aos modelos de seleção conhecidos. No Capítulo 3, introduzimos mais uma abordagem do modelo de seleção de Heckman ao considerar a distribuição bivariada Birnbaum-Saunders para as variáveis de interesse primária e de seleção. Denominamos esse novo modelo por Heckman-BS. Além de ser uma alternativa para o ajuste de dados assimétricos positivos com problemas de seleção de amostra, nosso modelo tem a vantagem de possuir a mesma quantidade de parâmetros do modelo de Heckman clássico e de não ser necessária a transformação da variável de interesse para seu ajuste, diferentemente, dos modelos Heckman clássico, Skew-Normal e Heckman-t quando ajustados a tais dados.

Palavras-chave: Dispersão variável, correlação variável, seleção amostral, distribuição bivariada, normalidade, assimetria.

Abstract

In a sampling process there may be a relationship between the variable of interest and a latent variable, such that the variable of interest is observed only in a subset of the population under study. In this case, we say that the variable of interest is subject to a hidden truncation and that has problems of sample selection, or more specifically, sample selection bias. One of the first researchers to study models to fit data with sample selection bias problems was the mathematician and economist, James Joseph Heckman. He observed that such a problem arises when the variable of interest is correlated with a hidden variable and that a joint modeling of these variables was necessary since adjusting regression models only for the observed part of the variable of interest produces biased and inconsistent results. In this context, Heckman (1976) introduced the sample selection model, also known as the Heckman model, which together adjusts the variable of interest and the latent variable and proposed a maximum likelihood estimation method under the assumption of bivariate normality. Despite the advances of its model, the maximum likelihood estimation method was soon criticized by researchers of the time due to the need for the bivariate normality assumption and also due to the difficulty of its implementation. This motivated Heckman (1979) to introduce an alternative and easily implemented method for estimating the parameters of the sample selection model, called the two-step method. Several generalizations were also proposed over the years to deal with situations in which normality is violated. Such generalizations include, in the parametric framework, the use of the Student's t-distribution by Marchenko e Genton (2012) and its Bayesian approach by Ding (2014), the use of the Skew-normal distribution by Ogundimu e Hutton (2016) and also a robust alternative for the two-step method presented by Zhelonkin et al. (2016). There are also copula-based, semi-parametric, and non-parametric approaches. In this paper, we introduce two new parametric models. In Chapter 2 we added covariates to the dispersion and correlation parameters of the classical model and introduced the generalized Heckman model. Our simulations indicate that in the presence of heteroscedasticity and/or variable correlation, the adjustment of the models, classical Heckman, Skew-normal or Heckman-t, can lead to loss of efficiency and also incorrect estimation of parameters. Differently from what happens with the adjustment of the generalized Heckman model to such data. In this way, our model becomes an important alternative to known selection models. In Chapter 3, we introduce another approach to the Heckman selection model by considering the bivariate Birnbaum-Saunders distribution for the primary and selection interest variables. We call this new model by Heckman-BS. In addition to being an alternative for adjusting positive asymmetric data with sample selection problems, our model has the advantage of having the same number of parameters of the classical Heckman model and of not requiring the transformation of the variable of interest for its fit, differently of the classic Heckman, Skew-Normal and Heckman-t models when fitted to such data.

Keywords: Variable dispersion, variable correlation, sample selection, bivariate distribution, normality, asymmetry.

Resumo	7
Abstract	8
Lista de Figuras	11
Lista de Tabelas	14
1 Preliminares	19
1.1 Introdução	19
1.2 Modelo de Seleção Amostral de Heckman	21
1.3 Distribuição Birnbaum-Saunders	25
1.4 Distribuição Birnbaum-Saunders Bivariada	27
1.5 Objetivos do Trabalho	34
1.6 Suporte Computacional	35
1.7 Organização do Trabalho	35
2 Modelo Heckman Generalizado	36
Resumo do Capítulo	36
2.1 Introdução	36
2.2 Formulação do Modelo Heckman Generalizado	38
2.2.1 Verossimilhança e Vetor Escore	38
2.3 Comportamento dos Estimadores em Amostras Finitas	40

2.3.1	Cenário 1: Dispersão Fixa e Correlação Variável	42
2.3.2	Cenário 2: Dispersão Variável e Correlação Fixa	49
2.3.3	Cenário 3: Dispersão e Correlação Variáveis	56
2.3.4	Cenário 4: Dispersão e Correlação Variáveis Sem Restrição de Exclusão	63
2.3.5	Cenário 5: Dispersão e Correlação Variáveis com Alto Valor de Censura	65
2.4	Aplicação a Dados Reais	68
2.5	Conclusões	72
3	Modelo de Seleção Amostral Heckman Birnbaum-Saunders	73
	Resumo do Capítulo	73
3.1	Introdução	74
3.2	Formulação do Modelo Heckman-BS	75
3.2.1	Função de Verossimilhança e Vetor Escore	77
3.3	Comportamento dos Estimadores em Amostras Finitas	80
3.3.1	Cenário 1	81
3.3.2	Cenário 2	93
3.4	Tamanho e Poder dos Testes da Razão de Verossimilhança e Gradiente	99
3.5	Ajuste do Modelo Heckman-BS sob Falta de Especificação	102
3.6	Aplicação a Dados Reais	109
3.7	Conclusões	111
4	Considerações Finais	113
4.1	Conclusões	113
4.2	Trabalhos Futuros	114
	Referências Bibliográficas	114
	Apêndice	119
	Apêndice A Matriz Hessiana do Modelo Heckman Generalizado	120
	Apêndice B Matriz Hessiana do Modelo Heckman-BS	122

1.1	Razão inversa de Mills para $z^\top \gamma \in [-5, 5]$	25
1.2	Função densidade de $Y \sim BS(\mu, \phi)$, para (a) $\phi = 3$, (b) $\mu = 1$ e (c) variância de Y para $\mu = 3$ e alguns valores de ϕ	26
1.3	Distribuição de Probabilidade Conjunta de (Y_1, Y_2) em que $\mu_1 = 7$, $\mu_2 = 1$, $\phi_1 = 2$, $\phi_2 = 9$ e (a) $\rho = -0.9$, (b) $\rho = 0$, (c) $\rho = 0.5$ e (d) $\rho = 0.9$	34
2.1	Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 , (g) β_3 e (h) σ através do ajuste dos modelos (i) Heckman Generalizado, de (ii) Heckman Clássico, (iii) Skew-Normal e (iv) Heckman-t. Tamanho amostral $n = 500$	44
2.2	Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 , (g) β_3 e (h) σ através do ajuste dos modelos (i) Heckman Generalizado, de (ii) Heckman Clássico, (iii) Skew-Normal e (iv) Heckman-t. Tamanho amostral $n = 1.000$	45
2.3	Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 , (g) β_3 e (h) σ através do ajuste dos modelos (i) Heckman Generalizado, de (ii) Heckman Clássico, (iii) Skew-Normal e (iv) Heckman-t. Tamanho amostral $n = 2.000$	46
2.4	Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 , (g) β_3 e (h) ρ a partir do ajuste dos modelos (i) Heckman Generalizado, (ii) Heckman Clássico, (iii) Skew-Normal e (iv) Heckman-t. Tamanho amostral $n = 500$	51

2.5	Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 , (g) β_3 e (h) ρ a partir do ajuste dos modelos (i) Heckman Generalizado, (ii) Heckman Clássico, (iii) Skew-Normal e (iv) Heckman-t. Tamanho amostral $n = 1.000$.	52
2.6	Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 , (g) β_3 e (h) ρ a partir do ajuste dos modelos (i) Heckman Generalizado, (ii) Heckman Clássico, (iii) Skew-Normal e (iv) Heckman-t. Tamanho amostral $n = 2.000$.	53
2.7	Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 e (g) β_3 , a partir do ajuste dos modelos (i) Heckman Generalizado, de (ii) Heckman Clássico, (iii) Skew-Normal e (iv) Heckman-t. Tamanho amostral $n = 500$.	58
2.8	Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 e (g) β_3 , a partir do ajuste dos modelos (i) Heckman Generalizado, de (ii) Heckman Clássico, (iii) Skew-Normal e (iv) Heckman-t. Tamanho amostral $n = 1.000$.	59
2.9	Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 e (g) β_3 , a partir do ajuste dos modelos (i) Heckman Generalizado, de (ii) Heckman Clássico, (iii) Skew-Normal e (iv) Heckman-t. Tamanho amostral $n = 2.000$.	60
3.1	Histograma das estimativas de máxima verossimilhança do parâmetro β_1 , conforme Tabela 3.7, para os três tamanhos amostrais utilizados na simulação $n = 500$, $n = 1000$ e $n = 2000$.	89
3.2	Histograma das estimativas de máxima verossimilhança do parâmetro ϕ , conforme Tabela 3.7, para os três tamanhos amostrais utilizados na simulação $n = 500$, $n = 1000$ e $n = 2000$.	89
3.3	Histograma das estimativas de máxima verossimilhança do parâmetro ρ , conforme Tabela 3.7, para os três tamanhos amostrais utilizados na simulação, $n = 500$, $n = 1000$ e $n = 2000$.	89
3.4	Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 , (g) ϕ e (h) ρ obtidas a partir do ajuste dos modelos (i) Heckman BS, (ii) Heckman Clássico, (iii) Skew-normal e (iv) Heckman-t. Tamanho amostral $n = 500$.	90
3.5	Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 , (g) ϕ e (h) ρ obtidas a partir do ajuste dos modelos (i) Heckman BS, (ii) Heckman Clássico, (iii) Skew-normal e (iv) Heckman-t. Tamanho amostral $n = 1000$.	91
3.6	Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 , (g) ϕ e (h) ρ obtidas a partir do ajuste dos modelos (i) Heckman BS, (ii) Heckman Clássico, (iii) Skew-normal e (iv) Heckman-t. Tamanho amostral $n = 2000$.	92
3.7	Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) β_1 , (e) β_2 , (f) α_1 e (g) ρ obtidas a partir do ajuste dos modelos (i) Heckman BS, (ii) Heckman Clássico, (iii) Skew-normal e (iv) Heckman-t. Tamanho amostral $n = 500$.	106

- 3.8 Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) β_1 , (e) β_2 , (f) α_1 e (g) ρ obtidas a partir do ajuste dos modelos (i) Heckman BS, (ii) Heckman Clássico, (iii) Skew-normal e (iv) Heckman-t. Tamanho amostral $n = 1000$. . 107
- 3.9 Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) β_1 , (e) β_2 , (f) α_1 e (g) ρ obtidas a partir do ajuste dos modelos (i) Heckman BS, (ii) Heckman Clássico, (iii) Skew-normal e (iv) Heckman-t. Tamanho amostral $n = 2000$. . 108

2.1	Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman generalizado, Heckman clássico, Skew-Normal e Heckman-t aos dados gerados conforme Cenário 1. Tamanhos amostrais $n = 500, n = 1.000$ e $n = 2.000$ com $N = 1.000$ réplicas de Monte Carlo.	43
2.2	Nível de significância empírico dos testes da razão de verossimilhança e gradiente para $H_0 : \kappa_1 = \kappa_2 = 0$ e censura média de 33.2%.	47
2.3	Poder dos testes da razão de verossimilhança (S_{RV}) e gradiente (S_G), em porcentagem, para dados simulados de acordo com o Cenário 1 e com o ajuste dos modelos de Heckman generalizado, Heckman Clássico, Skew-Normal e Heckman-t, com nível nominal de 1%, 5% e 10%.	48
2.4	Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman generalizado, Heckman clássico, Skew-Normal e Heckman-t aos dados gerados conforme Cenário 2. Tamanhos amostrais $n = 500, n = 1.000$ e $n = 2.000$ com $N = 1.000$ réplicas de Monte Carlo.	50
2.5	Nível de significância empírico dos testes da razão de verossimilhança e gradiente para $H_0 : \rho = 0$	54
2.6	Tamanho empírico e poder dos testes da razão de verossimilhança (S_{RV}) e gradiente (S_G), em porcentagem, para dados simulados de acordo com o Cenário 2 e com o ajuste dos modelos de Heckman generalizado, Heckman Clássico, Skew-Normal e Heckman-t, com nível nominal de 1%, 5% e 10%.	55

2.7	Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman generalizado, Heckman clássico, Skew-Normal e Heckman-t aos dados gerados conforme Cenário 3. Tamanhos amostrais $n = 500, n = 1.000$ e $n = 2.000$ com $N = 1.000$ réplicas de Monte Carlo.	57
2.8	Nível de significância empírico dos testes da razão de verossimilhança e gradiente para $H_0 : \rho = 0$ e censura média de 30%.	61
2.9	Tamanho empírico e poder dos testes da razão de verossimilhança (S_{RV}) e gradiente (S_G), em porcentagem, para dados simulados de acordo com o Cenário 3 e sob o ajuste dos modelos Heckman generalizado, Heckman Clássico, Heckman Skew e Heckman-t, com nível nominal de 1%, 5% e 10%	62
2.10	Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman generalizado, Heckman clássico, Skew-Normal e Heckman-t aos dados gerados conforme Cenário 4. Tamanhos amostrais $n = 500, n = 1.000$ e $n = 2.000$ com $N = 1.000$ réplicas de Monte Carlo.	64
2.11	Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman generalizado, Heckman clássico, Skew-Normal e Heckman-t aos dados gerados conforme Cenário 5. Tamanhos amostrais $n = 500, n = 1.000$ e $n = 2.000$ com $N = 1.000$ réplicas de Monte Carlo.	66
2.12	Valor verdadeiro(VV) e média das estimativas de máxima verossimilhança com a respectiva estatística de raiz quadrada do erro quadrático médio (REQM) do ajuste dos modelos de Heckman generalizado, Heckman clássico, Skew-Normal e Heckman-t aos dados gerados conforme modelo 5. Tamanhos amostrais $n = 500, n = 1.000$ e $n = 2.000$ com $N = 1.000$ réplicas de Monte Carlo.	67
2.13	Estimativas do Modelo de Heckman clássico e os respectivos p -valores juntamente com as estimativas do Modelo de Heckman Generalizado com os respectivos valores de desvio-padrão (DP), valor-z, p -valor e limites inferior e superior para o intervalo de confiança de 95%.	70
2.14	Estimativas do Modelo de Heckman clássico e os respectivos p -valores juntamente com as estimativas do Modelo Heckman Generalizado com os respectivos valores de desvio-padrão (DP), valor-z, p -valor e limites inferior e superior para o intervalo de confiança de 95%.	71

3.1	Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas após ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 1 e $\rho = -0.7$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.	82
3.2	Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas após ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 1 e $\rho = -0.5$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.	83
3.3	Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas após ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 1 e $\rho = -0.2$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.	84
3.4	Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas após ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 1 e $\rho = 0$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.	85
3.5	Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas após ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 1 e $\rho = 0.2$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.	86
3.6	Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas após ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 1 e $\rho = 0.5$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.	87
3.7	Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas após ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 1 e $\rho = 0.7$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.	88

3.8	Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 2 com $\rho = -0.7$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.	93
3.9	Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 2 com $\rho = -0.5$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.	94
3.10	Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 2 com $\rho = -0.2$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.	95
3.11	Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 2 com $\rho = 0$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.	96
3.12	Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 2 com $\rho = 0.2$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.	97
3.13	Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 2 com $\rho = 0.5$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.	98
3.14	Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 2 com $\rho = 0.7$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.	99
3.15	Nível de significância empírico dos testes da razão de verossimilhança e gradiente para $H_0 : \rho = 0$ contra $H_1 : \rho \neq 0$ e censura média de 30%.	100

3.16 Poder (em percentagem) dos testes da razão de verossimilhança e gradiente para $\alpha = 1\%$, $\alpha = 5\%$ e $\alpha = 10\%$	101
3.17 Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) do ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados simulados com distribuição gama bivariada de Cherian. Tamanhos amostrais $n = 500$, $n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.	105
3.18 Poder (em percentagem) dos testes da razão de verossimilhança e gradiente para $\alpha = 1\%$, $\alpha = 5\%$ e $\alpha = 10\%$	109
3.19 Estimativas do Modelo Heckman-BS com os respectivos valores de desvio-padrão (DP), valor-z, p -valor, limite inferior e limite superior para o intervalo de confiança de 95%, juntamente com as estimativas do Modelo Heckman clássico e os respectivos p -valores. . .	111

1.1 Introdução

Em um processo de amostragem pode existir uma relação entre a variável de interesse e uma variável latente, de tal forma, que a variável de interesse é observada somente em um subconjunto da população sob estudo. Neste caso, dizemos que a variável de interesse está sujeita a um truncamento oculto e que possui problemas de seleção amostral, ou mais especificamente, viés de seleção amostral. De acordo com [Marchenko e Genton \(2012\)](#), seleção amostral é um caso especial, de um conceito mais geral, conhecido na literatura econométrica como variáveis dependentes limitadas e surge, na prática, em diversas áreas, tais como Estatística, Ciências Econômicas, Sociologia, Finanças, Ciências Políticas, dentre outras.

Um dos primeiros pesquisadores a estudar modelos para ajustar dados com problemas de viés de seleção amostral foi o matemático e economista, James Joseph Heckman. Seus trabalhos deram início a uma discussão teórica sobre a modelagem e o desenvolvimento de novos procedimentos estatísticos para tratar tal problema. Heckman observou, que viés de seleção amostral surge quando a variável de interesse é correlacionada com uma variável oculta e que era necessária uma modelagem conjunta destas variáveis, pois o ajuste de modelos de regressão somente para a parte observada da variável de interesse produz resultados viesados e inconsistentes.

Nesse contexto, [Heckman \(1976\)](#) introduziu o modelo de seleção amostral, também conhecido como modelo de Heckman, que ajusta conjuntamente a variável de interesse e a variável latente e propôs um método de estimação por máxima verossimilhança sob a suposição de normalidade bivariada. Apesar dos avanços de seu modelo, o método de estimação por máxima verossimilhança logo foi criticado por pesquisadores da época, devido a necessidade da suposição de normalidade bivariada e também devido a dificuldade de sua implementação. Isto motivou [Heckman \(1979\)](#) a introduzir um método alternativo e de fácil implementação para a estimação dos parâmetros do modelo de seleção amostral, denominado método de dois passos.

Após a introdução do modelo de Heckman, diversos estudos, tais como, [Nelson \(1984\)](#), [Paarsch \(1984\)](#), [Manning et al. \(1987\)](#), [Stolzenberg e Relles \(1990\)](#), [Hartman \(1991\)](#), [Nawata \(1993, 1994\)](#) e [Leung e Yu \(1996\)](#), foram realizados a fim de examinar seu ajuste e a performance de ambos os métodos de estimação. Os resultados sugerem que o modelo pode reduzir ou eliminar o viés de seleção quando seus pressupostos são atendidos, ou seja, quando o modelo está corretamente especificado. No entanto, também indicam que o desvio de normalidade, que pode ser causado pela presença de outliers, multimodalidade, misturas ou assimetria, é um problema que pode inviabilizar seu ajuste.

Várias generalizações foram então propostas para lidar com situações em que a normalidade é violada. Tais generalizações incluem, no âmbito paramétrico, o uso da distribuição t de Student por [Marchenko e Genton \(2012\)](#) e sua abordagem Bayesiana por [Ding \(2014\)](#), o uso da distribuição Skew-normal por [Ogundimu e Hutton \(2016\)](#) e também uma alternativa robusta para o método de dois passos apresentado por [Zhelonkin et al. \(2016\)](#). Além disso, abordagens baseadas em cópula foram discutidas em [Lee \(1983\)](#) e abordagens semi-paramétricas para a seleção amostral podem ser encontrados em [Ahn e Powell \(1993\)](#), [Powell \(1994\)](#) e [Newey \(2009\)](#). A abordagem não-paramétrica foi estudada por [Das et al. \(2003\)](#).

Mesmo com avanços nas abordagens semi e não-paramétrica, de acordo com [Ogundimu e Hutton \(2016\)](#), os modelos de seleção amostral paramétricos são preferíveis, pois permitem identificar com maior eficiência o intercepto do modelo de seleção. Este, além de ser de especial importância para a interpretação dos resultados, principalmente quando a predição é o objetivo da análise, é um parâmetro de interesse em muitas aplicações econômicas ([Marchenko e Genton, 2012](#)). Nesse sentido, os principais avanços na indicação de modelos paramétricos foram feitos por [Marchenko e Genton \(2012\)](#), que introduziu o modelo Heckman-t ao sugerir o uso da distribuição t de student bivariada em substituição a distribuição normal para o ajuste de dados com valores extremos, uma vez que a distribuição t de student apresenta caudas mais pesadas e permite ajustes mais robustos. [Zhelonkin et al. \(2016\)](#) que modificou o método de dois passos e introduziu um método mais robusto e [Ogundimu e Hutton \(2016\)](#) que apresentou o modelo Skew-normal ao sugerir o uso da distribuição Skew-normal bivariada em substituição a distribuição normal para o ajuste de dados assimétricos.

Contudo, apesar de suas vantagens, tais modelos apresentam limitações. O método de [Zhelonkin et al. \(2016\)](#), por exemplo, ainda necessita da suposição de normalidade bivariada. Os modelos Heckman-t e Skew-normal, apesar de flexíveis, não são parcimoniosos, pois necessitam, respectivamente, da estimação dos parâmetros de grau de liberdade e de assimetria, além dos parâmetros das equações de seleção e regressão e dos parâmetros de dispersão e correlação. Além disso, tal como o modelo clássico de Heckman, para o ajuste do modelo Heckman-t a dados assimétricos e positivos, é necessária a transformação da variável de interesse, de forma, a reduzir sua assimetria. Em geral, transformações da variável de interesse nem sempre são apropriadas e podem complicar a interpretação dos parâmetros do modelo resultante, principalmente, quando há problemas de viés de seleção. Ressaltamos que para o ajuste do modelo Skew-normal aos dados assimétricos e positivos considerados nesta tese, também foi necessária a transformação da variável de interesse com o mesmo objetivo anterior, apesar deste modelo ser indicado para o ajuste de dados assimétricos.

Neste trabalho mostramos que dados com dispersão e/ou correlação diferentes para cada variável também podem afetar a estimação dos parâmetros do modelo e podem ocultar a presença de viés de seleção quando o modelo de Heckman clássico, modelo Heckman-t ou modelo Skew-Normal são ajustados a tais dados. Assim, propomos no Capítulo 2, o modelo Heckman generalizado ao considerarmos covariáveis para os parâmetros de dispersão e correlação do modelo de Heckman clássico. Mostramos, via simulação Monte Carlo, que para dados com dispersão e correlação variáveis, o modelo introduzido é uma alternativa para uma estimação mais eficiente quando comparado ao ajuste dos demais modelos.

No Capítulo 3, introduzimos o modelo paramétrico Heckman-BS para o ajuste de dados assimétricos e positivos, ao considerarmos a distribuição Birnbaum-Saunders bivariada na modelagem conjunta da variável de interesse e variável latente. Nosso modelo é parcimonioso, possui a mesma quantidade de parâmetros do modelo de Heckman clássico e não necessita da transformação da variável de interesse para seu ajuste, diferentemente, dos modelos Heckman clássico, Skew-Normal e Heckman-t quando ajustados a tais dados.

Estudos Monte Carlo foram realizados para avaliarmos o comportamento dos estimadores de máxima verossimilhança obtidos para os modelos propostos, comparativamente aos modelos Heckman clássico, Heckman-t e Skew-Normal. Realizamos também o estudo dos testes da razão de verossimilhança e gradiente para compararmos o poder e o nível de significância empírico para todos os modelos. Simulamos dados assimétricos e positivos com distribuição gama e comparamos o ajuste dos modelos Heckman clássico, Heckman-t, Skew-Normal e Heckman-BS. Por fim, os modelos propostos foram ajustados aos dados de gastos ambulatoriais do conjunto de dados do *Medical Expenditure Panel Survey (MEPS)* de 2001, utilizados também por [Cameron e Trivedi \(2009\)](#), por [Marchenko e Genton \(2012\)](#) e por [Zhelonkin et al. \(2016\)](#). O banco de dados está disponível no *software* R via pacote *ssmrob* de [Zhelonkin et al. \(2014\)](#).

Antes de apresentarmos os modelos propostos nos Capítulos 2 e 3, detalhamos neste capítulo o modelo clássico de Heckman juntamente com o método de estimação por máxima verossimilhança e o método de dois passos, apresentamos as distribuições Birnbaum-Saunders e Birnbaum-Saunders Bivariada com as principais propriedades que serão necessárias para o desenvolvimento do modelo Heckman-BS.

1.2 Modelo de Seleção Amostral de Heckman

Considere o sistema de equações

$$Y_{1i}^* = \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_{1i}, \quad (1.1)$$

$$Y_{2i}^* = \mathbf{z}_i^\top \boldsymbol{\gamma} + \epsilon_{2i}, \quad i = 1, \dots, n, \quad (1.2)$$

em que a variável Y_{2i}^* é não observada, sabemos somente se Y_{2i}^* é maior ou menor que $a \in \mathbb{R}$, fixo. E, a variável Y_{1i}^* é observada somente se $Y_{2i}^* > a$. Ou seja, na prática, observamos

$$U_i = 1\{Y_{2i}^* > a\}, \quad (1.3)$$

$$Y_i = Y_{1i}^* U_i, \quad i = 1, \dots, n,$$

em que $1\{Y_{2i}^* > a\} = 1$, se $Y_{2i}^* > a$ ou $1\{Y_{2i}^* > a\} = 0$, caso contrário. Os vetores $\beta \in \mathbb{R}^p$ e $\gamma \in \mathbb{R}^q$ são de parâmetros desconhecidos, $\mathbf{x}_i \in \mathbb{R}^p$ e $\mathbf{z}_i \in \mathbb{R}^q$, são regressores não necessariamente mutuamente exclusivos. Para definir o modelo de seleção amostral, Heckman considerou os termos de erro $(\epsilon_{1i}, \epsilon_{2i})$ com distribuição normal bivariada dada por

$$\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \stackrel{iid.}{\sim} \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix} \right], \quad i = 1, \dots, n, \quad (1.4)$$

em que $\sigma_1^2 \equiv \sigma^2$ e, como não observamos Y_{2i}^* , a variância de Y_{2i}^* é não identificável, assim, Heckman fixou $\sigma_2^2 = 1$. No entanto, qualquer outro valor positivo pode ser considerado. Além disso, em geral, o valor de a é zero, pois qualquer outro valor seria absorvido pelo intercepto em (1.2). O modelo (1.1) a (1.4) é conhecido como modelo de Heckman, também denominado modelo Tobit tipo 2 ou apenas, modelo de seleção amostral. A equação (1.1) é denominada equação de interesse primária ou apenas como equação de regressão e a equação (1.2) é conhecida na literatura como equação de seleção. O parâmetro $\rho \in (-1, 1)$ é o parâmetro de maior interesse no modelo de Heckman, pois é ele que determina o viés de seleção que surge quando $\rho \neq 0$.

1.2.1 Estimação por Máxima Verossimilhança

Para a estimação dos parâmetros do modelo por máxima verossimilhança temos que encontrar a densidade da variável aleatória mista Y_i . Logo, notemos que

$$\begin{aligned} P(Y_i \leq y) &= P(Y_i \leq y | U_i = 1)P(U_i = 1) + P(Y_i \leq y | U_i = 0)P(U_i = 0) \\ &= 1\{Y_{2i}^* > 0\}P(Y_{1i}^* \leq y | Y_{2i}^* > 0)P(Y_{2i}^* > 0) + 1\{Y_{2i}^* \leq 0\}P(Y_{2i}^* \leq 0) \\ &= \begin{cases} P(Y_{1i}^* \leq y | Y_{2i}^* > 0) (1 - P(Y_{2i}^* \leq 0)), & \text{se } Y_{2i}^* > 0, \\ P(Y_{2i}^* \leq 0), & \text{se } Y_{2i}^* \leq 0. \end{cases} \end{aligned}$$

Segue que, para $\theta = (\beta^\top, \gamma^\top, \sigma, \rho)^\top$,

$$f_{Y_i}(y; \theta) = \begin{cases} f_{Y_{1i}^* | Y_{2i}^* > 0}(y; \theta) (1 - P(Y_{2i}^* \leq 0)), & \text{se } Y_{2i}^* > 0, \\ P(Y_{2i}^* \leq 0), & \text{se } Y_{2i}^* \leq 0, \end{cases}$$

ou seja, a densidade de Y_i possui um componente discreto e um componente contínuo. Seu componente discreto é definido pelo modelo Probit

$$P(U_i = u) = \Phi(\mathbf{z}_i^\top \gamma)^u \Phi(-\mathbf{z}_i^\top \gamma)^{1-u}, \quad \text{para } u = 0, 1,$$

e a partir do trabalho de [Arellano-Valle et al. \(2006\)](#), obtemos a densidade

$$f_{Y_{1i}^*|Y_{2i}^*>0}(y; \boldsymbol{\theta}) = f_{Y_{1i}^*}(y; \boldsymbol{\theta}) \frac{P(Y_{2i}^* > 0 | Y_{1i}^* = y, \boldsymbol{\theta})}{P(Y_{2i}^* > 0)}, \quad (1.5)$$

que representa o componente contínuo de Y_i . Sabendo que

$$Y_{2i}^* | Y_{1i}^* = y \sim \mathcal{N} \left(\mathbf{z}_i^\top \boldsymbol{\gamma} + \frac{\rho}{\sigma} (y - \mathbf{x}_i^\top \boldsymbol{\beta}), 1 - \rho^2 \right),$$

temos que,

$$P(Y_{2i}^* > 0 | Y_{1i}^* = y) = \Phi \left(\frac{\mathbf{z}_i^\top \boldsymbol{\gamma}}{\sqrt{1 - \rho^2}} + \frac{\rho(y - \mathbf{x}_i^\top \boldsymbol{\beta})}{\sigma \sqrt{1 - \rho^2}} \right). \quad (1.6)$$

Além disso,

$$f_{Y_{1i}^*}(y) = \frac{1}{\sigma} \phi \left(\frac{y - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right),$$

portanto, o componente contínuo de Y_i é dado por

$$f_{Y_{1i}^*|Y_{2i}^*>0}(y; \boldsymbol{\theta}) = \frac{1}{\sigma \Phi(\mathbf{z}_i^\top \boldsymbol{\gamma})} \phi \left(\frac{y - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right) \Phi \left(\frac{\mathbf{z}_i^\top \boldsymbol{\gamma}}{\sqrt{1 - \rho^2}} + \frac{\rho(y - \mathbf{x}_i^\top \boldsymbol{\beta})}{\sigma \sqrt{1 - \rho^2}} \right), \quad (1.7)$$

e a log-verossimilhança para um par de observações (Y_i, U_i) é

$$\begin{aligned} \mathcal{L}_i(\boldsymbol{\theta}) &= u_i \log f_{Y_{1i}^*|Y_{2i}^*>0}(y_i; \boldsymbol{\theta}) + u_i \log \Phi(\mathbf{z}_i^\top \boldsymbol{\gamma}) + (1 - u_i) \log \Phi(-\mathbf{z}_i^\top \boldsymbol{\gamma}) \\ &= u_i \left\{ \log \Phi \left(\frac{\mathbf{z}_i^\top \boldsymbol{\gamma} + \rho(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})/\sigma}{\sqrt{1 - \rho^2}} \right) + \log \phi \left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right) - \log \sigma \right\} \\ &\quad + (1 - u_i) \log \Phi(-\mathbf{z}_i^\top \boldsymbol{\gamma}). \end{aligned} \quad (1.8)$$

Quando a suposição de normalidade bivariada é válida, os estimadores de máxima verossimilhança (EMV) obtidos a partir da maximização de $\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^n \mathcal{L}_i(\boldsymbol{\theta})$, são consistentes, assintoticamente normal e eficientes. Porém, a função $\mathcal{L}(\boldsymbol{\theta})$ é não-linear e, assim, necessita de métodos iterativos para sua maximização, os quais, necessitam de um bom ponto de partida (chute inicial) para o algoritmo numérico, uma vez que $\mathcal{L}(\boldsymbol{\theta})$ pode possuir máximos locais.

1.2.2 Estimação de Parâmetros pelo Método de Dois Passos

Um segundo procedimento de estimação foi também proposto por [Heckman \(1979\)](#) e denominado método de dois passos. Esse método, foi sugerido como um bom estimador para pontos de partida confiáveis e eficientes na estimação por máxima verossimilhança ([Leung e Yu, 2000](#)). O método é baseado

no fato da média condicional $\tilde{\mu}_i = E(y_i | y_{1i}^* \text{ é observado, } \mathbf{x}_i, \mathbf{z}_i)$, para $i = 1, \dots, n$, ser dada por

$$\begin{aligned}
\tilde{\mu}_i &= E(y_i | y_{1i}^* \text{ é observado, } \mathbf{x}_i, \mathbf{z}_i) \\
&= E(y_i | y_{2i}^* > 0, \mathbf{x}_i, \mathbf{z}_i) \\
&= E(y_{1i}^* | \mathbf{z}_i^\top \boldsymbol{\gamma} + \epsilon_{2i} > 0, \mathbf{x}_i, \mathbf{z}_i) \\
&= E(\mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_{1i} | \epsilon_{2i} > -\mathbf{z}_i^\top \boldsymbol{\gamma}, \mathbf{x}_i, \mathbf{z}_i) \\
&= \mathbf{x}_i^\top \boldsymbol{\beta} + E(\epsilon_{1i} | \epsilon_{2i} > -\mathbf{z}_i^\top \boldsymbol{\gamma}, \mathbf{z}_i) \\
&= \mathbf{x}_i^\top \boldsymbol{\beta} + \rho \sigma \frac{\phi(\mathbf{z}_i^\top \boldsymbol{\gamma})}{\Phi(\mathbf{z}_i^\top \boldsymbol{\gamma})} \\
&= \mathbf{x}_i^\top \boldsymbol{\beta} + \lambda_i \beta_\lambda,
\end{aligned} \tag{1.9}$$

em que $\lambda_i = \frac{\phi(\mathbf{z}_i^\top \boldsymbol{\gamma})}{\Phi(\mathbf{z}_i^\top \boldsymbol{\gamma})}$ denota a razão inversa de Mills, $\beta_\lambda = \rho\sigma$, ρ é a correlação entre ϵ_{1i} e ϵ_{2i} e σ é o desvio padrão de ϵ_{1i} . A partir de (1.9) podemos reescrever a equação de interesse como

$$\tilde{y}_i = \tilde{\mu}_i + \varepsilon_i, \tag{1.10}$$

em que $\tilde{\mu}_i$ é dada em (1.9) e ε_i é um novo termo de erro de média zero e independente de \mathbf{z}_i e de \mathbf{x}_i . O termo $\lambda_i \beta_\lambda$ em (1.9) explica a inconsistência do estimador de mínimos quadrados ordinários (MQO), quando $\rho \neq 0$, e se MQO fosse utilizado para encontrar as estimativas dos parâmetros de (1.1). A partir daí, o primeiro passo do método é ajustar o modelo probit a equação de seleção (1.2) e estimar $\hat{\boldsymbol{\gamma}}$ e $\hat{\lambda}_i = \frac{\phi(\mathbf{z}_i^\top \hat{\boldsymbol{\gamma}})}{\Phi(\mathbf{z}_i^\top \hat{\boldsymbol{\gamma}})}$. Em um segundo passo, estimamos por MQO os parâmetros $\boldsymbol{\beta}$ e $\beta_\lambda = \rho\sigma$ de (1.10) usando os valores de y_{1i}^* observados. Um estimador para a variância de ϵ_1 é dado por

$$\hat{\sigma}^2 = \frac{1}{n_u} \left(\hat{\boldsymbol{\varepsilon}}^\top \hat{\boldsymbol{\varepsilon}} + \hat{\beta}_\lambda^2 \sum_{i=1}^{n_u} \hat{\delta}_i \right), \tag{1.11}$$

em que $\hat{\boldsymbol{\varepsilon}}$ é o vetor residual da estimação de MQO de (1.10), n_u é o número de observações nesta estimação e $\hat{\delta}_i = \hat{\lambda}_i (\hat{\lambda}_i + \mathbf{z}_i^\top \boldsymbol{\gamma})$. Finalmente, um estimador para a correlação entre ϵ_1 e ϵ_2 é dado por

$$\hat{\rho} = \frac{\hat{\beta}_\lambda}{\hat{\sigma}}, \tag{1.12}$$

nesse caso, $\hat{\rho}$ pode estar fora do intervalo $[-1, 1]$.

A maior vantagem do método de dois passos é sua simplicidade, pois é mais fácil de ajustar do que o método de máxima verossimilhança, não requer algoritmos complicados e é uma alternativa mais robusta. Porém, é menos eficiente e o uso da razão inversa de Mills (λ) pode ocasionar possíveis problemas de multicolinearidade devido a sua linearidade em grande parte do seu suporte, como é possível observar na Figura 1.1. Para diminuir este problema é sugerido a **restrição de exclusão**, de acordo com a qual, pelo menos uma variável, que é um bom preditor de Y_2^* e está incluída na equação de seleção, não deve aparecer na regressão primária.

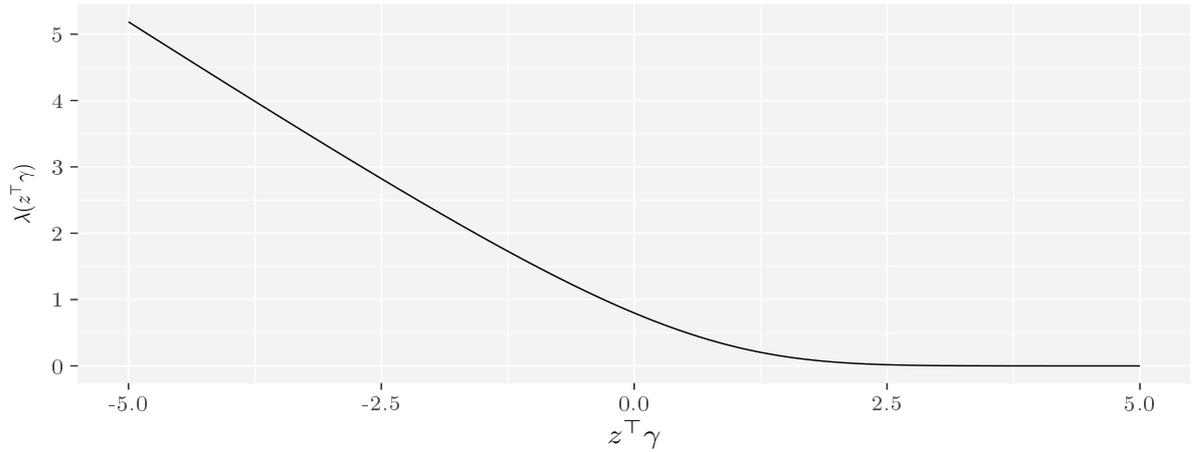


Figura 1.1: Razão inversa de Mills para $z^T \gamma \in [-5, 5]$.

1.3 Distribuição Birnbaum-Saunders

A distribuição Birnbaum-Saunders, denotada aqui por BS, foi introduzida por [Birnbaum e Saunders \(1969a,b\)](#) motivada por problemas de vibração encontrados em aviões comerciais e por problemas de falhas de materiais. Trata-se de uma distribuição que modela o tempo de vida de materiais e equipamentos submetidos a cargas dinâmicas e que foi derivada a partir de uma transformação monótona da distribuição normal. Desde sua introdução, uma quantidade considerável de trabalhos tem sido desenvolvidos a partir de seus diferentes aspectos, por exemplo, [Chang e Tang \(1993, 1994\)](#), [Johnson et al. \(1995\)](#), [Dupuis e Mills \(1998\)](#), [Ng et al. \(2003, 2006\)](#), [Leiva et al. \(2008\)](#), [From e Li \(2006\)](#), [Lemonte et al. \(2007, 2008\)](#), [Kundu et al. \(2010\)](#), [Jamalizadeh e Kundu \(2015\)](#), [Leiva \(2015\)](#), [Saulo et al. \(2017\)](#), [Fonseca e Cribari-Neto \(2018\)](#) e referências citadas nestes trabalhos.

Uma variável aleatória Y segue uma distribuição Birnbaum-Saunders com parâmetros $\alpha, \beta > 0$, denotada por $Y \sim BS(\alpha, \beta)$, se tem densidade dada por

$$f(y; \alpha, \beta) = \frac{1}{2\sqrt{2\pi\alpha\beta}} \left[\left(\frac{\beta}{y} \right)^{\frac{1}{2}} + \left(\frac{\beta}{y} \right)^{\frac{3}{2}} \right] \exp \left[-\frac{1}{2\alpha^2} \left(\frac{y}{\beta} + \frac{\beta}{y} - 2 \right) \right], \quad y > 0, \quad (1.13)$$

em que α e β são os parâmetros de forma e escala, respectivamente. A correspondente função de distribuição acumulada (fda) de Y é

$$F(y; \alpha, \beta) = \Phi \left\{ \frac{1}{\alpha} \left[\left(\frac{y}{\beta} \right)^{\frac{1}{2}} - \left(\frac{\beta}{y} \right)^{\frac{1}{2}} \right] \right\}, \quad \text{para } y > 0. \quad (1.14)$$

Para gerar variáveis aleatórias Y com distribuição BS basta considerar

$$Z \stackrel{D}{=} \frac{1}{\alpha} \left[\left(\frac{Y}{\beta} \right)^{\frac{1}{2}} - \left(\frac{\beta}{Y} \right)^{\frac{1}{2}} \right] \text{ ou } Y \stackrel{D}{=} \beta \left[\frac{\alpha Z}{2} + \sqrt{\left(\frac{\alpha Z}{2} \right)^2 + 1} \right]^2, \quad (1.15)$$

em que, $Z \sim \mathcal{N}(0, 1)$. A média e variância de $Y \sim BS(\alpha, \beta)$, são dadas por

$$E(Y) = \beta \left(1 + \frac{\alpha^2}{2}\right) \text{ e } Var(Y) = (\alpha\beta)^2 \left(1 + \frac{5}{4}\alpha^2\right). \quad (1.16)$$

Para nossos propósitos é interessante utilizar uma reparametrização da BS em função de sua média. Assim, consideramos $\alpha = \sqrt{\frac{2}{\phi}}$ e $\beta = \frac{\phi\mu}{\phi+1}$, e obtemos

$$\mu = \beta \left(1 + \frac{\alpha^2}{2}\right) \text{ e } \phi = \frac{2}{\alpha^2}, \quad (1.17)$$

em que $\mu, \phi > 0$ são os parâmetros de média e forma, respectivamente. Ao considerar esta nova reparametrização dizemos que Y tem distribuição Birnbaum-Saunders com parâmetros $\mu, \phi > 0$ e a denotamos por $Y \sim BS(\mu, \phi)$. Neste caso, a função densidade fica dada por

$$f(y; \alpha, \beta) = \frac{\exp\left(\frac{\phi}{2}\right)\sqrt{\phi+1}}{4\sqrt{\pi\mu}} y^{-\frac{3}{2}} \left[y + \frac{\phi\mu}{\phi+1}\right] \exp\left[-\frac{\phi}{4} \left(\frac{y(\phi+1)}{\phi\mu} + \frac{\phi\mu}{y(\phi+1)}\right)\right], \quad y > 0, \quad (1.18)$$

e a fda será

$$F(y; \alpha, \beta) = \Phi \left\{ \sqrt{\frac{\phi}{2}} \left[\left(\frac{y(\phi+1)}{\phi\mu}\right)^{\frac{1}{2}} - \left(\frac{\phi\mu}{y(\phi+1)}\right)^{\frac{1}{2}} \right] \right\}, \text{ para } y > 0. \quad (1.19)$$

A média e a variância são:

$$E(Y) = \mu \text{ e } Var(Y) = \frac{\mu^2(2\phi+5)}{(\phi+1)^2}. \quad (1.20)$$

Uma observação importante é que a distribuição $BS(\mu, \phi)$ satisfaz a propriedade de invariância por escala, ou seja, se $Y \sim BS(\mu, \phi)$, então $T = aY$, com $a > 0$, possui distribuição $BS(a\mu, \phi)$. Na Figura 1.2 apresentamos o gráfico da função densidade e da variância de $Y \sim BS(\mu, \phi)$ para alguns valores de μ e ϕ .

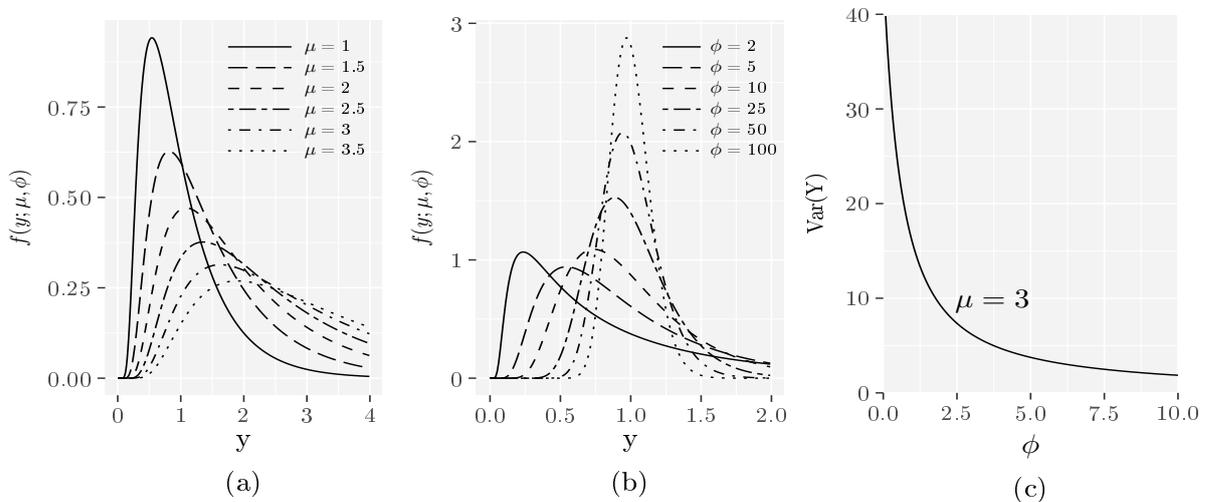


Figura 1.2: Função densidade de $Y \sim BS(\mu, \phi)$, para (a) $\phi = 3$, (b) $\mu = 1$ e (c) variância de Y para $\mu = 3$ e alguns valores de ϕ .

1.4 Distribuição Birnbaum-Saunders Bivariada

A distribuição Birnbaum-Saunders Bivariada ou, simplesmente distribuição BS bivariada, foi proposta por [Kundu et al. \(2010\)](#) como uma extensão da distribuição BS univariada e foi introduzida utilizando a mesma transformação monótona. Tal distribuição possui cinco parâmetros, é absolutamente contínua e possui distribuições marginais BS univariadas.

Dizemos que o vetor aleatório $\mathbf{Y} = (Y_1, Y_2)^\top$ possui distribuição BS bivariada com parâmetros $\alpha_1, \beta_1, \alpha_2, \beta_2 > 0$ e $-1 < \rho < 1$, se sua fda conjunta é

$$P(Y_1 \leq y_1, Y_2 \leq y_2) = \Phi_2 \left(\frac{1}{\alpha_1} \left(\sqrt{\frac{y_1}{\beta_1}} - \sqrt{\frac{\beta_1}{y_1}} \right), \frac{1}{\alpha_2} \left(\sqrt{\frac{y_2}{\beta_2}} - \sqrt{\frac{\beta_2}{y_2}} \right); \rho \right), \quad (1.21)$$

para $y_1, y_2 > 0$, em que a função $\Phi_2(\cdot, \cdot; \rho)$ denota a fda conjunta de uma normal padrão bivariada com coeficiente de correlação ρ . A função densidade conjunta de Y_1 e Y_2 é dada por

$$\begin{aligned} f_{Y_1, Y_2}(y_1, y_2) &= \phi_2 \left(\frac{1}{\alpha_1} \left(\sqrt{\frac{y_1}{\beta_1}} - \sqrt{\frac{\beta_1}{y_1}} \right), \frac{1}{\alpha_2} \left(\sqrt{\frac{y_2}{\beta_2}} - \sqrt{\frac{\beta_2}{y_2}} \right); \rho \right) \\ &\quad \times \frac{1}{2\alpha_1\beta_1} \left\{ \left(\frac{\beta_1}{y_1} \right)^{\frac{1}{2}} + \left(\frac{\beta_1}{y_1} \right)^{\frac{3}{2}} \right\} \frac{1}{2\alpha_2\beta_2} \left\{ \left(\frac{\beta_2}{y_2} \right)^{\frac{1}{2}} + \left(\frac{\beta_2}{y_2} \right)^{\frac{3}{2}} \right\}, \end{aligned} \quad (1.22)$$

em que $\phi_2(\cdot, \cdot; \rho)$ denota a fdp conjunta da normal bivariada padrão, dada por

$$\phi_2(z_1, z_2; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} (z_1^2 + z_2^2 - 2\rho z_1 z_2) \right\}, \quad Z \in \mathbb{R}^2. \quad (1.23)$$

Vamos denotar por $\mathbf{Y} \sim BS(\alpha_1, \alpha_2, \beta_1, \beta_2, \rho)$ ou por $\mathbf{Y} \sim BS_2(\boldsymbol{\alpha}, \boldsymbol{\beta}, \rho)$, em que $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^\top$ e $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$ uma variável \mathbf{Y} com densidade (1.22).

Considere,

$$a_{Y_i}(\alpha_i, \beta_i) = \frac{1}{\alpha_i} \left[\sqrt{\frac{Y_i}{\beta_i}} - \sqrt{\frac{\beta_i}{Y_i}} \right] \text{ e } A_{Y_i}(\alpha_i, \beta_i) = \frac{(Y_i + \beta_i)}{2\alpha_i\beta_i^{\frac{1}{2}}Y_i^{\frac{3}{2}}}, \quad i = 1, 2. \quad (1.24)$$

Assim, a fda conjunta de $\mathbf{Y} = (Y_1, Y_2)^\top$ com distribuição BS bivariada pode ser escrita de forma simplificada como

$$F_{\mathbf{Y}}(\mathbf{y}) = \Phi_2(a_{y_1}(\alpha_1, \beta_1), a_{y_2}(\alpha_2, \beta_2); \rho), \quad y_1, y_2 > 0, \quad (1.25)$$

em que $\Phi_2(\cdot; \rho)$ denota a fda da normal padrão bivariada $\mathbf{Z} = (Z_1, Z_2)^\top$ com coeficiente de correlação ρ .

A fdp de $\mathbf{Y} = (Y_1, Y_2)^\top$ pode ser expressa por

$$f_{\mathbf{Y}}(\mathbf{y}) = \phi_2(a_{\mathbf{y}}(\boldsymbol{\alpha}, \boldsymbol{\beta}); \rho) \mathbf{A}_{\mathbf{y}}(\boldsymbol{\alpha}, \boldsymbol{\beta}), \quad \mathbf{y} \in \mathbb{R}_+^2, \quad (1.26)$$

em que $a_{\mathbf{y}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = (a_{y_1}(\alpha_1, \beta_1), a_{y_2}(\alpha_2, \beta_2))$ e $\mathbf{A}_{\mathbf{y}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = A_{y_1}(\alpha_1, \beta_1)A_{y_2}(\alpha_2, \beta_2)$, com $a_{y_i}(\alpha_i, \beta_i)$ e

$A_{y_i}(\alpha_i, \beta_i)$ definidos em (1.24) para $i = 1, 2$, e $\phi_2(\cdot; \rho)$ é dado em (1.23). É fácil mostrar que,

$$\frac{d}{dy_i} a_{y_i}(\alpha_i, \beta_i) = A_{y_i}(\alpha_i, \beta_i), \text{ para } i = 1, 2. \quad (1.27)$$

As propriedades da distribuição BS bivariada apresentadas a seguir são de extrema importância para a obtenção de alguns resultados desta tese. Para maiores detalhes veja [Kundu et al. \(2010\)](#).

Proposição 1. *Seja $\mathbf{Y} = (Y_1, Y_2)^\top \sim BS_2(\boldsymbol{\alpha}, \boldsymbol{\beta}; \rho)$. Então*

- (a) $Y_i \sim BS(\alpha_i, \beta_i)$, para $i = 1, 2$;
- (b) $\mathbf{b} \odot \mathbf{Y} \sim BS_2(\boldsymbol{\alpha}, \mathbf{b} \odot \boldsymbol{\beta}; \rho)$, em que $\mathbf{b} = (b_1, b_2)^\top \in \mathbb{R}_+^2$ e \odot denota o produto de Hadamard¹;
- (c) $\mathbf{Y}^{-1} = (Y_1^{-1}, Y_2^{-1})^\top \sim BS_2(\boldsymbol{\alpha}, \boldsymbol{\beta}^{-1}; \rho)$, em que $\boldsymbol{\beta}^{-1} = (1/\beta_1, 1/\beta_2)^\top$;
- (d) $\mathbf{Y}_1^{-1} = (Y_1^{-1}, Y_2)^\top \sim BS_2(\boldsymbol{\alpha}, \boldsymbol{\beta}_1^{-1}; -\rho)$, em que $\boldsymbol{\beta}_1^{-1} = (1/\beta_1, \beta_2)^\top$;
- (e) $\mathbf{Y}_2^{-1} = (Y_1, Y_2^{-1})^\top \sim BS_2(\boldsymbol{\alpha}, \boldsymbol{\beta}_2^{-1}; -\rho)$, em que $\boldsymbol{\beta}_2^{-1} = (\beta_1, 1/\beta_2)^\top$;
- (f) Y_1 e Y_2 são independentes se, e somente se, $\rho = 0$;
- (g) A fdp condicional de Y_1 , dado $Y_2 = y_2$, é dada por:

$$f_{Y_1|Y_2=y_2}(y_1) = \frac{A_{y_1}(\alpha_1, \beta_1)}{\sqrt{2\pi}\sqrt{1-\rho^2}} \times \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[a_{y_1}(\alpha_1, \beta_1) - \rho a_{y_2}(\alpha_2, \beta_2) \right]^2 \right\} \quad (1.28)$$

- (h) A fda condicional de Y_1 , dado $Y_2 = y_2$, é dada por:

$$F_{Y_1|Y_2}(y_1|y_2) = \Phi \left\{ \frac{a_{y_1}(\alpha_1, \beta_1) - \rho a_{y_2}(\alpha_2, \beta_2)}{\sqrt{1-\rho^2}} \right\}. \quad (1.29)$$

Demonstração:

- (a) Seja $\mathbf{Y} = (Y_1, Y_2)^\top \sim BS_2(\boldsymbol{\alpha}, \boldsymbol{\beta}; \rho)$. Considerando $a_{y_i} = a_{y_i}(\alpha_i, \beta_i)$ e $A_{y_i} = A_{y_i}(\alpha_i, \beta_i)$, para $i = 1, 2$, temos que

$$\begin{aligned} f_{Y_1}(y_1) &= \int_0^\infty f_{\mathbf{Y}}(\mathbf{y}) dy_2 \\ &= \int_0^\infty \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left[-\frac{1}{2(1-\rho^2)} (a_{y_1}^2 + a_{y_2}^2 - 2\rho a_{y_1} a_{y_2}) \right] A_{y_1} A_{y_2} dy_2 \\ &= \int_0^\infty \frac{A_{y_1}}{\sqrt{2\pi}} \exp \left(-\frac{a_{y_1}^2}{2} \right) \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp \left[-\frac{1}{2(1-\rho^2)} (a_{y_2} - \rho a_{y_1})^2 \right] A_{y_2} dy_2 \\ &= \frac{A_{y_1}}{\sqrt{2\pi}} \exp \left(-\frac{a_{y_1}^2}{2} \right) \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{z^2}{2} \right) dz, \quad z = \frac{(a_{y_2} - \rho a_{y_1})}{\sqrt{1-\rho^2}} \\ &= \frac{A_{y_1}}{\sqrt{2\pi}} \exp \left(-\frac{a_{y_1}^2}{2} \right) \end{aligned}$$

¹**Definição** (Produto de Hadamard): Sejam A e B matrizes $m \times n$. O produto de Hadamard de A e B é definido por $[A \odot B]_{ij} = [A]_{ij}[B]_{ij}$, para todo $1 \leq i \leq m$, $1 \leq j \leq n$.

Ou seja,

$$f_{Y_1}(y_1) = \phi(a_{y_1})A_{y_1}.$$

Da mesma forma, mostra-se que $f_{Y_2}(y_2) = \phi(a_{y_2})A_{y_2}$. ■

- (b) Dado $\mathbf{Y} = (Y_1, Y_2)^\top \sim BS_2(\boldsymbol{\alpha}, \boldsymbol{\beta}; \rho)$, considere $\mathbf{T} = \mathbf{b} \odot \mathbf{Y} = (b_1 Y_1, b_2 Y_2) = (T_1, T_2)$, $b_1, b_2 > 0$. Queremos mostrar que $\mathbf{T} = \mathbf{b} \odot \mathbf{Y} \sim BS_2(\boldsymbol{\alpha}, \mathbf{b} \odot \boldsymbol{\beta}; \rho)$. Assim,

$$\begin{aligned} P(T_1 \leq t_1, T_2 \leq t_2) &= P(b_1 Y_1 \leq t_1, b_2 Y_2 \leq t_2) \\ &= P\left(Y_1 \leq \frac{t_1}{b_1}, Y_2 \leq \frac{t_2}{b_2}\right) \\ &= \Phi_2 \left[\frac{1}{\alpha_1} \left(\sqrt{\frac{t_1}{b_1 \beta_1}} - \sqrt{\frac{\beta_1 b_1}{t_1}} \right), \frac{1}{\alpha_2} \left(\sqrt{\frac{t_2}{b_2 \beta_2}} - \sqrt{\frac{\beta_2 b_2}{t_2}} \right); \rho \right] \\ &= \Phi_2 [a_{t_1}(\alpha_1, b_1 \beta_1), a_{t_2}(\alpha_2, b_2 \beta_2); \rho]. \end{aligned}$$
 ■

- (c) Dado que $\mathbf{Y} = (Y_1, Y_2)^\top \sim BS_2(\boldsymbol{\alpha}, \boldsymbol{\beta}; \rho)$, considere $T_1 = Y_1^{-1}$ e $T_2 = Y_2^{-1}$. Segue que $Y_1 = T_1^{-1}$, $Y_2 = T_2^{-1}$ e o jacobiano de \mathbf{Y} fica dado por $|J| = \begin{vmatrix} -t_1^{-2} & 0 \\ 0 & -t_2^{-2} \end{vmatrix} = t_1^{-2} t_2^{-2}$, $\mathbf{t} = (t_1, t_2) \in \mathbb{R}_+^2$. Como $f_{\mathbf{Y}}(\mathbf{y}) = \phi_2(a_{\mathbf{y}}(\boldsymbol{\alpha}, \boldsymbol{\beta}); \rho) \mathbf{A}_{\mathbf{y}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$, $\mathbf{y} \in \mathbb{R}_+^2$, e $\phi_2(\cdot, \cdot; \rho)$ é função par, temos que

$$\begin{aligned} f_{\mathbf{T}}(\mathbf{t}) &= f_{\mathbf{Y}}(\mathbf{t}^{-1})|J| \\ &= \phi_2(a_{\mathbf{t}^{-1}}(\boldsymbol{\alpha}, \boldsymbol{\beta}); \rho) \mathbf{A}_{\mathbf{t}^{-1}}(\boldsymbol{\alpha}, \boldsymbol{\beta})|J| \\ &= \phi_2 \left(\frac{1}{\alpha_1} \left[\sqrt{\frac{t_1^{-1}}{\beta_1}} - \sqrt{\frac{\beta_1}{t_1^{-1}}} \right], \frac{1}{\alpha_2} \left[\sqrt{\frac{t_2^{-1}}{\beta_2}} - \sqrt{\frac{\beta_2}{t_2^{-1}}} \right]; \rho \right) \frac{(t_1^{-1} + \beta_1)(t_2^{-1} + \beta_2)}{2\alpha_1 \beta_1^{\frac{1}{2}} t_1^{-\frac{3}{2}} 2\alpha_2 \beta_2^{\frac{1}{2}} t_2^{-\frac{3}{2}}} |J| \\ &= \phi_2 \left(\frac{1}{\alpha_1} \left[\sqrt{\frac{\beta_1^{-1}}{t_1}} - \sqrt{\frac{t_1}{\beta_1^{-1}}} \right], \frac{1}{\alpha_2} \left[\sqrt{\frac{\beta_2^{-1}}{t_2}} - \sqrt{\frac{t_2}{\beta_2^{-1}}} \right]; \rho \right) \frac{t_1^{\frac{1}{2}}(1 + t_1 \beta_1) t_2^{\frac{1}{2}}(1 + t_2 \beta_2)}{2\alpha_1 \beta_1^{\frac{1}{2}} 2\alpha_2 \beta_2^{\frac{1}{2}}} t_1^{-2} t_2^{-2} \\ &= \phi_2 \left(-\frac{1}{\alpha_1} \left[\sqrt{\frac{t_1}{\beta_1^{-1}}} - \sqrt{\frac{\beta_1^{-1}}{t_1}} \right], -\frac{1}{\alpha_2} \left[\sqrt{\frac{t_2}{\beta_2^{-1}}} - \sqrt{\frac{\beta_2^{-1}}{t_2}} \right]; \rho \right) \frac{t_1^2(\beta_1^{-1} + t_1) t_2^2(\beta_2^{-1} + t_2)}{2\alpha_1 \beta_1^{-\frac{1}{2}} 2\alpha_2 \beta_2^{-\frac{1}{2}}} t_1^{-2} t_2^{-2} \\ &= \phi_2 \left(\frac{1}{\alpha_1} \left[\sqrt{\frac{t_1}{\beta_1^{-1}}} - \sqrt{\frac{\beta_1^{-1}}{t_1}} \right], \frac{1}{\alpha_2} \left[\sqrt{\frac{t_2}{\beta_2^{-1}}} - \sqrt{\frac{\beta_2^{-1}}{t_2}} \right]; \rho \right) \frac{(\beta_1^{-1} + t_1)(\beta_2^{-1} + t_2)}{2\alpha_1 \beta_1^{-\frac{1}{2}} 2\alpha_2 \beta_2^{-\frac{1}{2}}} \\ &= \phi_2(a_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}^{-1}); \rho) \mathbf{A}_{\mathbf{t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}^{-1}). \end{aligned}$$
 ■

- (f) Se Y_1 e Y_2 são independentes é imediato que $\rho = 0$. Agora, se $\rho = 0$, temos

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= \phi_2(a_{\mathbf{y}}(\boldsymbol{\alpha}, \boldsymbol{\beta}); \rho) \mathbf{A}_{\mathbf{y}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \\ &= \frac{1}{2\pi} \exp \left\{ -\frac{1}{2} \left[\frac{1}{\alpha_1^2} \left(\sqrt{\frac{y_1}{\beta_1}} - \sqrt{\frac{\beta_1}{y_1}} \right)^2 + \frac{1}{\alpha_2^2} \left(\sqrt{\frac{y_2}{\beta_2}} - \sqrt{\frac{\beta_2}{y_2}} \right)^2 \right] \right\} \\ &\quad \times \frac{1}{2\alpha_1 \beta_1} \left\{ \left(\frac{\beta_1}{y_1} \right)^{\frac{1}{2}} + \left(\frac{\beta_1}{y_1} \right)^{\frac{3}{2}} \right\} \frac{1}{2\alpha_2 \beta_2} \left\{ \left(\frac{\beta_2}{y_2} \right)^{\frac{1}{2}} + \left(\frac{\beta_2}{y_2} \right)^{\frac{3}{2}} \right\} \end{aligned}$$

Ou seja,

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{Y_1}(y_1)f_{Y_2}(y_2).$$

■

(g) A densidade condicional de Y_1 dada a ocorrência do valor y_2 de Y_2 pode ser escrita como

$$f_{Y_1|Y_2=y_2}(y_1) = \frac{f_{Y_1, Y_2}(y_1, y_2)}{f_{Y_2}(y_2)}, \quad (1.30)$$

de onde obtemos facilmente (1.28).

■

(h)

$$\begin{aligned} P(Y_1 \leq y_1 | Y_2 = y_2) &= \int_0^{y_1} \frac{A_t(\alpha_1, \beta_1)}{\sqrt{2\pi}\sqrt{1-\rho^2}} \times \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[a_t(\alpha_1, \beta_1) - \rho a_{y_2}(\alpha_2, \beta_2) \right]^2 \right\} dt \\ &= \int_{-\infty}^{\frac{a_{y_1} - \rho a_{y_2}}{\sqrt{1-\rho^2}}} \frac{1}{\sqrt{2\pi}} \times \exp \left(-\frac{v^2}{2} \right) dv \\ &= \Phi \left\{ \frac{a_{y_1}(\alpha_1, \beta_1) - \rho a_{y_2}(\alpha_2, \beta_2)}{\sqrt{1-\rho^2}} \right\}, \end{aligned}$$

consideramos $v = \frac{a_t(\alpha_1, \beta_1) - \rho a_{y_2}(\alpha_2, \beta_2)}{\sqrt{1-\rho^2}}$, $dv = \frac{A_t(\alpha_1, \beta_1)}{\sqrt{1-\rho^2}} dt$ e $a_{y_i}(\alpha_i, \beta_i) = a_{y_i}$, $i = 1, 2$.

■

As demonstrações dos itens (d) e (e) seguem o mesmo raciocínio da demonstração em (c).

■

Corolário 1. *Seja $\mathbf{Y} = (Y_1, Y_2)^\top \sim BS_2(\boldsymbol{\alpha}, \boldsymbol{\beta}; \rho)$. Segue que, $(\alpha_1 Y_1 / \beta_1, \alpha_2 Y_2 / \beta_2)^\top \sim BS_2(\boldsymbol{\alpha}, \boldsymbol{\alpha}; \rho)$ e $(Y_1 / \beta_1, Y_2 / \beta_2)^\top \sim BS_2(\boldsymbol{\alpha}, \mathbf{1}_2; \rho)$, $\mathbf{1}_2 = (1, 1)^\top$.*

Demonstração: Aplicação direta da Proposição 1, item (b).

■

Outros resultados obtidos por Kundu et al. (2010) e importantes para o desenvolvimento de nosso trabalho são os seguintes momentos:

$$E \left[\frac{Y_1 Y_2}{(\beta_1 \beta_2)} \right] = 1 + \frac{(\alpha_1^2 + \alpha_2^2)}{2} + \frac{\alpha_1^2 \alpha_2^2 (1 + \rho^2)}{4} + \alpha_1 \alpha_2 I_1(\rho), \quad (1.31)$$

$$E \left[\frac{\sqrt{Y_1 Y_2}}{\sqrt{(\beta_1 \beta_2)}} \right] = \frac{\alpha_1 \alpha_2 \rho}{4} + I_2(\rho), \quad (1.32)$$

em que,

$$\begin{aligned} I_1(\rho) &= a_{0,0} + \frac{1}{2^3} a_{0,1}(\alpha_1^2 + \alpha_2^2) + \frac{1}{2^6} \alpha_1^2 \alpha_2^2 a_{1,1} + \sum_{i=2}^{\infty} (-1)^{i-1} \frac{1.3 \cdots (2i-3)}{2^{3i} i!} a_{0,i}(\alpha_1^{2i} + \alpha_2^{2i}) \\ &+ \sum_{i=2}^{\infty} (-1)^{i-1} \frac{1.3 \cdots (2i-3)}{2^{3i+3} i!} a_{1,i}(\alpha_1^2 \alpha_2^{2i} + \alpha_2^2 \alpha_1^{2i}) \\ &+ \sum_{i=2}^{\infty} \sum_{j=2}^{\infty} (-1)^{i+j} \frac{1.3 \cdots (2i-3)}{2^{3i} i!} \frac{1.3 \cdots (2j-3)}{2^{3j} j!} \alpha_1^{2i} \alpha_2^{2j} a_{i,j}, \end{aligned} \quad (1.33)$$

$$\begin{aligned}
I_2(\rho) &= 1 + \frac{1}{2^3}(\alpha_1^2 + \alpha_2^2) + \frac{1}{2^6}\alpha_1^2\alpha_2^2(1 + 2\rho^2) + \sum_{i=2}^{\infty}(-1)^{i-1}\frac{1.3 \cdots (2i-3)}{2^{3i}i!}b_{0,i}(\alpha_1^{2i} + \alpha_2^{2i}) \\
&\quad + \sum_{i=2}^{\infty}(-1)^{i-1}\frac{1.3 \cdots (2i-3)}{2^{3i+3}i!}b_{1,i}(\alpha_1^2\alpha_2^{2i} + \alpha_2^2\alpha_1^{2i}) \\
&\quad + \sum_{i=2}^{\infty}\sum_{j=2}^{\infty}(-1)^{i+j}\frac{1.3 \cdots (2i-3)}{2^{3i}i!}\frac{1.3 \cdots (2j-3)}{2^{3j}j!}\alpha_1^{2i}\alpha_2^{2j}b_{i,j},
\end{aligned} \tag{1.34}$$

e para m, n inteiros não negativos,

$$a_{m,n} = E(Z_1^{2m+1}Z_2^{2n+1}) = \frac{(2m+1)!(2n+1)!}{2^{m+n}} \sum_{i=0}^{\min\{m,n\}} \frac{(2\rho)^{2i+1}}{(m-i)!(n-i)!(2i+1)!},$$

$$b_{m,n} = E(Z_1^{2m}Z_2^{2n}) = \frac{(2m)!(2n)!}{2^{m+n}} \sum_{i=0}^{\min\{m,n\}} \frac{(2\rho)^{2i}}{(m-i)!(n-i)!(2i)!}.$$

Proposição 2. *Seja $\mathbf{Y} = (Y_1, Y_2)^\top \sim BS_2(\boldsymbol{\alpha}, \boldsymbol{\beta}; \rho)$. Então*

(a) $E(Y_i) = \frac{\beta_i}{2}(2 + \alpha_i^2);$

(b) $Var(Y_i) = \frac{\beta_i^2\alpha_i^2}{4}(4 + 5\alpha_i^2);$

(c) $Cov(Y_1, Y_2) = \beta_1\beta_2\alpha_1\alpha_2 \left(\frac{1}{4}\alpha_1\alpha_2\rho^2 + I_1(\rho) \right)$, com $I_1(\rho)$ definido em (1.33);

(d) $\rho_{\mathbf{Y}} = \frac{\alpha_1\alpha_2\rho^2 + 4I_1(\rho)}{\sqrt{(4 + 5\alpha_1^2)(4 + 5\alpha_2^2)}}$, em que $I_1(\rho)$ é definido em (1.33) e ρ é o coeficiente de correlação da normal bivariada.

Demonstração:

Os itens (a) e (b) da Proposição 2 seguem da definição da BS bivariada. ■

(c)

$$\begin{aligned}
Cov(Y_1, Y_2) &= E(Y_1Y_2) - E(Y_1)E(Y_2) \\
&= (\beta_1\beta_2) \left(1 + \frac{(\alpha_1^2 + \alpha_2^2)}{2} + \frac{\alpha_1^2\alpha_2^2(1 + \rho^2)}{4} + \alpha_1\alpha_2I_1(\rho) \right) - \left[\frac{\beta_1}{2}(2 + \alpha_1^2) \right] \left[\frac{\beta_2}{2}(2 + \alpha_2^2) \right] \\
&= \beta_1\beta_2\alpha_1\alpha_2 \left[\frac{1}{4}\alpha_1\alpha_2\rho^2 + I_1(\rho) \right].
\end{aligned}$$

■

(d)

$$\begin{aligned}\rho_{\mathbf{Y}} &= \frac{\text{Cov}(Y_1, Y_2)}{\sqrt{\text{Var}(Y_1)\text{Var}(Y_2)}} \\ &= \frac{\beta_1\beta_2\alpha_1\alpha_2 \left[\frac{1}{4}\alpha_1\alpha_2\rho^2 + I_1(\rho) \right]}{\sqrt{\frac{\beta_1^2\alpha_1^2}{4}(4+5\alpha_1^2) \frac{\beta_2^2\alpha_2^2}{4}(4+5\alpha_2^2)}} \\ &= \frac{[\alpha_1\alpha_2\rho^2 + 4I_1(\rho)]}{\sqrt{(4+5\alpha_1^2)(4+5\alpha_2^2)}}.\end{aligned}$$

■

Corolário 2. *Seja $\mathbf{Y} = (Y_1, Y_2)^\top \sim BS_2(\boldsymbol{\alpha}, \boldsymbol{\beta}; \rho)$. Considere $\rho_{\mathbf{Y}}$ o coeficiente de correlação entre Y_1 e Y_2 . Se $\rho = 0$, então $\rho_{\mathbf{Y}} = 0$.*

Demonstração:

Se $\rho = 0$, temos que $a_{m,n} = 0$ e $I_1(\rho) = 0$. Assim,

$$\begin{aligned}E(Y_1Y_2) &= \beta_1\beta_2 \left[1 + \frac{(\alpha_1^2 + \alpha_2^2)}{2} + \frac{(\alpha_1^2\alpha_2^2)}{4} \right] \\ &= \beta_1 \left(1 + \frac{\alpha_1^2}{2} \right) \beta_2 \left(1 + \frac{\alpha_2^2}{2} \right) \\ &= E(Y_1)E(Y_2),\end{aligned}$$

de onde, $\rho_{\mathbf{Y}} = 0$.

■

Proposição 3. *Seja $\mathbf{Y} = (Y_1, Y_2)^\top \sim BS_2(\boldsymbol{\alpha}, \boldsymbol{\beta}; \rho)$. Segue que Y_1 e Y_2 são independentes se, e somente se, $Z_1 = \frac{1}{\alpha_1} \left(\sqrt{\frac{Y_1}{\beta_1}} - \sqrt{\frac{\beta_1}{Y_1}} \right)$ e $Z_2 = \frac{1}{\alpha_2} \left(\sqrt{\frac{Y_2}{\beta_2}} - \sqrt{\frac{\beta_2}{Y_2}} \right)$ são independentes.*

Demonstração:

Consideremos $f_{Y_1Y_2}(y_1, y_2)$ a fdp de $\mathbf{Y} = (Y_1, Y_2)$ e $f_{Z_1Z_2}(z_1, z_2)$ a fdp normal padrão bivariada de $\mathbf{Z} = (Z_1, Z_2)$. Então para qualquer $(z_1, z_2) \in \mathbb{R}^2$ temos a partir de (1.15) que

$$f_{Z_i}(z_i) = f_{Y_i}(\varphi_i(z_i)) \text{ e } f_{Z_1Z_2}(z_1, z_2) = f_{Y_1Y_2}(\varphi_1(z_1), \varphi_2(z_2)),$$

em que $\varphi_i(\cdot)$ é uma função bijetora dada por $\varphi_i(z_i) = \beta_i \left[\frac{\alpha_i z_i}{2} + \sqrt{\left(\frac{\alpha_i z_i}{2} \right)^2 + 1} \right]$, $i = 1, 2$. Segue que, se Y_1 e Y_2 são independentes, então temos

$$f_{Z_1Z_2}(z_1, z_2) = f_{Y_1Y_2}(\varphi_1(z_1), \varphi_2(z_2)) = f_{Y_1}(\varphi_1(z_1))f_{Y_2}(\varphi_2(z_2)) = f_{Z_1}(z_1)f_{Z_2}(z_2),$$

de onde, Z_1 e Z_2 são independentes. Agora, se assumirmos que Z_1 e Z_2 são independentes, então temos

$$f_{Y_1Y_2}(y_1, y_2) = f_{Z_1Z_2}(\varphi_1^{-1}(y_1), \varphi_2^{-1}(y_2)) = f_{Z_1}(\varphi_1^{-1}(y_1))f_{Z_2}(\varphi_2^{-1}(y_2)) = f_{Y_1}(y_1)f_{Y_2}(y_2),$$

de onde, Y_1 e Y_2 são independentes. ■

De acordo com o item (b) da Proposição 1, a distribuição BS bivariada pertence a uma família tipo escala. Além disso, os itens (c) a (e) da Proposição 1 demonstram que a distribuição BS bivariada é fechada sob reciprocidade, pelo menos em uma de suas componentes, Saunders (1974). A partir dos resultados apresentados e da relação da distribuição BS bivariada com a distribuição normal padrão bivariada podemos construir um algoritmo para gerarmos variáveis aleatórias com distribuição BS bivariada.

Algoritmo:

- Passo 1: Geramos duas variáveis aleatórias independentes $U_1 \sim \mathcal{N}(0, 1)$ e $U_2 \sim \mathcal{N}(0, 1)$;
- Passo 2: Calculamos

$$\begin{aligned} Z_1 &= \frac{\sqrt{1+\rho} + \sqrt{1-\rho}}{2} U_1 + \frac{\sqrt{1+\rho} - \sqrt{1-\rho}}{2} U_2, \\ Z_2 &= \frac{\sqrt{1+\rho} - \sqrt{1-\rho}}{2} U_1 + \frac{\sqrt{1+\rho} + \sqrt{1-\rho}}{2} U_2; \end{aligned}$$

- Passo 3: Por fim, obtemos

$$Y_i = \beta_i \left[\frac{1}{2} \alpha_i Z_i + \sqrt{\left(\frac{1}{2} \alpha_i Z_i \right)^2 + 1} \right]^2, \quad i = 1, 2, \quad (1.35)$$

em que $\mathbf{Y} = (Y_1, Y_2)^\top \sim BS_2(\boldsymbol{\alpha}, \boldsymbol{\beta}; \rho)$. Se considerarmos $\alpha_i = \sqrt{\frac{2}{\phi_i}}$ e $\beta_i = \frac{\phi_i \mu_i}{(\phi_i + 1)}$, $i = 1, 2$, a média e variância de \mathbf{Y} ficam dadas por

$$E(Y_i) = \mu_i \text{ e } \text{Var}(Y_i) = \frac{\mu_i^2 (\phi_i + 1)^2}{(2\phi + 5)}. \quad (1.36)$$

Com esta parametrização, denote $\mathbf{Y} \sim BS_2(\boldsymbol{\mu}, \boldsymbol{\phi}; \rho)$, com $\boldsymbol{\mu} = (\mu_1, \mu_2)^\top$ e $\boldsymbol{\phi} = (\phi_1, \phi_2)^\top$. Neste caso, a densidade em (1.22) pode ser escrita como

$$\begin{aligned} f_{Y_1, Y_2}(y_1, y_2) &= \Phi_2 [a_{y_1}(\mu_1, \phi_1), a_{y_2}(\mu_2, \phi_2); \rho] \\ &\times \frac{\phi_1 + 1}{2\phi_1 \mu_1} \sqrt{\frac{\phi_1}{2}} \left\{ \left[\frac{\phi_1 \mu_1}{(\phi_1 + 1)y_1} \right]^{\frac{1}{2}} + \left[\frac{\phi_1 \mu_1}{(\phi_1 + 1)y_1} \right]^{\frac{3}{2}} \right\} \\ &\times \frac{\phi_2 + 1}{2\phi_2 \mu_2} \sqrt{\frac{\phi_2}{2}} \left\{ \left[\frac{\phi_2 \mu_2}{(\phi_2 + 1)y_2} \right]^{\frac{1}{2}} + \left[\frac{\phi_2 \mu_2}{(\phi_2 + 1)y_2} \right]^{\frac{3}{2}} \right\}, \end{aligned} \quad (1.37)$$

em que $a_{y_i}(\mu_i, \phi_i) = \sqrt{\frac{\phi_i}{2}} \left(\sqrt{\frac{(\phi_i + 1)y_i}{\phi_i \mu_i}} - \sqrt{\frac{\phi_i \mu_i}{(\phi_i + 1)y_i}} \right)$, $i = 1, 2$.

Todas as propriedades demonstradas anteriormente continuam válidas com a reparametrização. No caso da Proposição 1, item (b), se $\mathbf{Y} = (Y_1, Y_2)^\top \sim BS_2(\boldsymbol{\mu}, \boldsymbol{\phi}; \rho)$, então $\mathbf{b} \odot \mathbf{Y} \sim BS_2(\mathbf{b} \odot \boldsymbol{\mu}, \boldsymbol{\phi}; \rho)$, em que $\mathbf{b} = (b_1, b_2)^\top \in \mathbb{R}_+^2$. A função de distribuição de probabilidade conjunta de (Y_1, Y_2) é unimodal.

Gráficos de superfície de $f_{Y_1, Y_2}(y_1, y_2)$ para diferentes valores de ρ e valores fixos de μ_1, ϕ_1, μ_2 e ϕ_2 são apresentados na Figura 1.3.

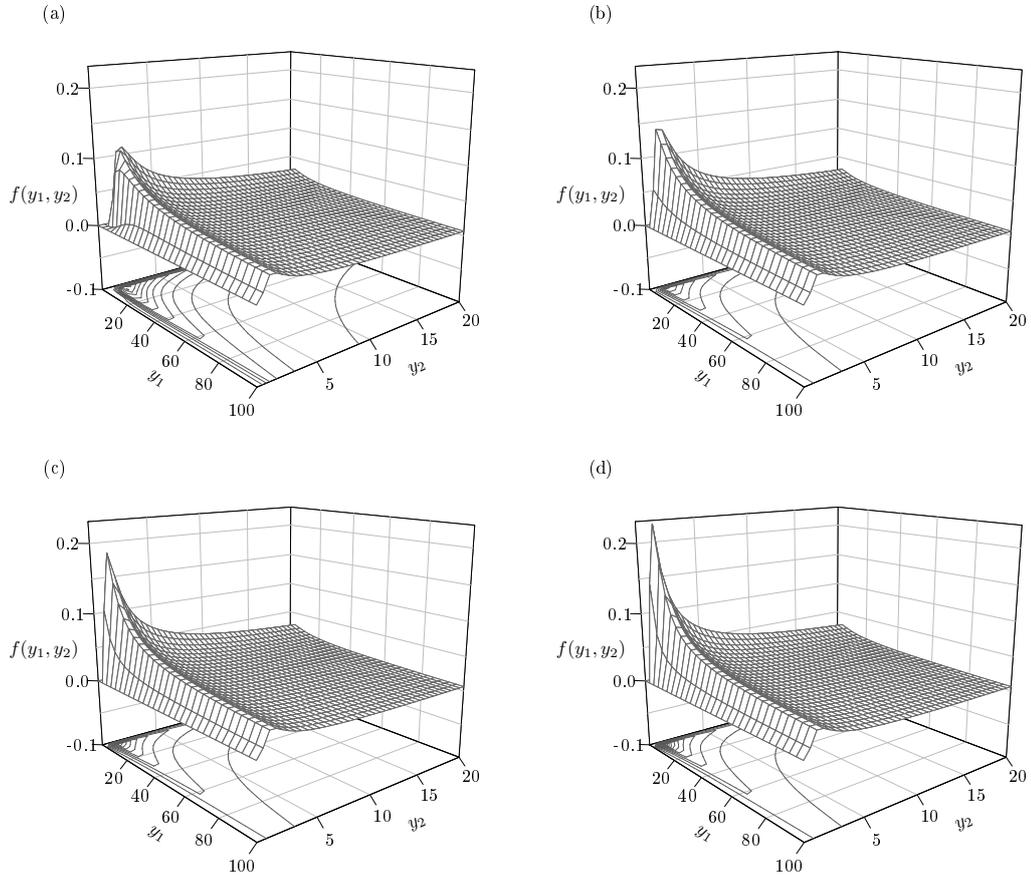


Figura 1.3: Distribuição de Probabilidade Conjunta de (Y_1, Y_2) em que $\mu_1 = 7$, $\mu_2 = 1$, $\phi_1 = 2$, $\phi_2 = 9$ e (a) $\rho = -0.9$, (b) $\rho = 0$, (c) $\rho = 0.5$ e (d) $\rho = 0.9$.

1.5 Objetivos do Trabalho

O objetivo deste trabalho é introduzir na literatura dois novos modelos de seleção amostral baseados no modelo clássico de Heckman. Os objetivos específicos podem ser resumidos em:

1. Descrever o modelo de seleção amostral de Heckman, destacando suas propriedades e os métodos de estimação por máxima verossimilhança e de dois passos.
2. Generalizar o modelo de Heckman ao acrescentar covariáveis aos parâmetros de dispersão e correlação e introduzir na literatura o modelo Heckman generalizado.
3. Introduzir um novo modelo de seleção amostral para dados assimétricos e positivos com o uso da distribuição Birnbaum-Saunders bivariada o qual será denominado modelo Heckman-BS;

Finalmente, investigar propriedades dos estimadores de máxima verossimilhança (EMV) dos modelos propostos por meio de simulações Monte Carlo e ilustrar o ajuste dos mesmos em dados reais.

1.6 Suporte Computacional

As avaliações numéricas realizadas ao longo deste trabalho, a geração dos gráficos e as aplicações foram feitas através do **Team (2016b)**, versão 3.4.3, para sistemas operacionais Windows, que se encontra disponível de forma gratuita no endereço www.rproject.org/. Usamos o compilador **Team (2016a)**, versão livre 1.1.383, que pode ser encontrado no endereço <https://www.rstudio.com/>. O trabalho foi digitado usando o sistema de tipografia LATEX no compilador TexnicCenter, que pode ser adquirido gratuitamente no endereço <http://www.texniccenter.org/>.

1.7 Organização do Trabalho

O presente trabalho encontra-se dividido em quatro capítulos. Neste primeiro capítulo apresentamos uma introdução e uma breve discussão sobre o modelo clássico de Heckman e algumas de suas propriedades, além de apresentarmos a distribuição Birnbaum-Saunders univariada e bivariada.

No capítulo 2, propomos uma generalização do modelo de Heckman ao considerarmos covariáveis para os parâmetros de variância e de correlação. Apresentamos estudos de simulação Monte Carlo em que comparamos os resultados obtidos a partir do ajuste do modelo Heckman generalizado com os modelos Heckman clássico, Skew-Normal e Heckman-t. Além de apresentar resultados do nível empírico e poder dos testes da razão de verossimilhança e gradiente e uma aplicação prática do modelo proposto.

No capítulo 3, propomos um novo modelo de seleção amostral para dados assimétricos e positivos com o uso da distribuição Birnbaum-Saunders bivariada, o qual denominamos modelo Heckman-BS. Fizemos um estudo de simulação Monte Carlo para comparar os resultados obtidos a partir do ajuste do modelo proposto com os modelos Heckman clássico, Skew-Normal e Heckman-t, além disso, apresentamos resultados do nível empírico e poder dos testes da razão de verossimilhança e gradiente e uma aplicação prática do modelo proposto.

O Capítulo 4 é dedicado às conclusões e alguns direcionamentos para estudos subsequentes.

Resumo do Capítulo

No modelo de Heckman clássico, a dispersão e a correlação são consideradas constantes. Dados reais, porém, mostram muito frequentemente, uma dispersão variável. Além disso, a correlação é inicialmente, o parâmetro de maior interesse nas diferentes aplicações dos modelos de seleção amostral. Logo, a fim de apresentar um modelo mais flexível, generalizamos o modelo de Heckman ao acrescentarmos covariáveis aos parâmetros de dispersão e de correlação, o que permite identificar as covariáveis responsáveis pela presença de vies de seleção e pela presença de heterocedasticidade. Avaliamos o desempenho do modelo proposto comparado ao desempenho dos modelos Heckman clássico, Skew-Normal e Heckman-t ajustados a dados gerados a partir de cinco cenários distintos. Os resultados indicam uma boa performance do nosso modelo quando a restrição de exclusão é imposta. Além disso, apresentamos uma aplicação aos dados *Medical Expenditure Panel Survey (MEPS)* de 2001, utilizados também por [Cameron e Trivedi \(2009\)](#), por [Marchenko e Genton \(2012\)](#) e por [Zhelonkin et al. \(2016\)](#).

Palavras-chave: Dispersão variável, correlação variável, seleção amostral.

2.1 Introdução

O problema mais discutido em relação ao ajuste do modelo de Heckman é a sua sensibilidade a suposição de distribuição de probabilidade dos erros. Pois, quando a forma paramétrica da função de verossimilhança é incorreta, os estimadores baseados em verossimilhança são, em geral, inconsistentes e produzem estimativas viesadas. Por outro lado, quando os termos de erro são corretamente ajustados, a estimação por máxima verossimilhança ou por procedimentos baseados em verossimilhança, produzem estimadores consistentes e eficientes. Conforme [Leung e Yu \(1996\)](#) e [Enders \(2010\)](#) por exemplo, modelos de seleção amostral podem reduzir o vies de seleção quando corretamente especificados.

No entanto, mesmo quando a forma da densidade dos erros é corretamente especificada, a heterocedasticidade dos termos de erro pode causar inconsistências nas estimativas dos parâmetros, como mostraram [Hurd \(1979\)](#) e [Arabmazar e Schmidt \(1981\)](#). Nesse sentido, [Donald \(1995\)](#) sugere que a heterocedasticidade em modelos de seleção amostral é relativamente negligenciada e cita dois motivos para suspeitar que sua modelagem é importante na prática. A primeira razão é que tipicamente os dados usados para ajustar modelos de seleção amostral são grandes bancos de dados, onde a heterogeneidade é comumente encontrada. A segunda razão, é que as estimativas dos parâmetros obtidas pelo ajuste dos modelos de seleção usuais, em alguns casos, podem ser mais severamente afetadas pela heterocedasticidade, do que pela distribuição incorreta dos termos de erro ([Powell, 1986](#)).

Logo, a modelagem da variância, que tem sido muito utilizado na literatura estatística, pode ser uma boa alternativa para reduzir o viés dos estimadores de máxima verossimilhança quando aplicamos modelos de seleção amostral. Não obstante, a correlação é, em geral, o parâmetro de maior interesse nas diferentes aplicações de tais modelos. Nesse sentido, generalizamos o modelo de Heckman e acrescentamos covariáveis aos parâmetros de dispersão e de correlação. Assim, introduzimos mais uma alternativa para o ajuste de dados com problemas de viés de seleção amostral, que permite ao pesquisador identificar, respectivamente, as fontes de variabilidade dos dados e as covariáveis responsáveis pelo viés de seleção, um resultado extremamente importante nas aplicações dos modelos de seleção.

Além disso, nossas simulações indicam, que na presença de heterocedasticidade e/ou correlação variável, o ajuste dos modelos, Heckman clássico, Skew-normal ou Heckman-t, pode nos levar a perda de eficiência e também a estimação incorreta dos parâmetros. Diferentemente do que ocorre com o ajuste do modelo Heckman generalizado a tais dados. Dessa forma, nosso modelo torna-se uma importante alternativa aos modelos de seleção conhecidos. Por outro lado, apesar das vantagens citadas, mostramos também que o modelo Heckman generalizado, assim como os demais, está sujeito a problemas de multicolinearidade e a sensibilidade dos estimadores a alta correlação entre a equação de regressão e seleção e que a restrição de exclusão é extremamente importante para uma estimação consistente e eficiente.

O presente capítulo encontra-se organizado da seguinte forma. Na Seção [2.2](#), introduzimos o modelo Heckman generalizado, com sua respectiva função de verossimilhança e vetor escore. Na Seção [2.3](#), apresentamos resultados de simulação Monte Carlo de cinco cenários distintos para avaliação dos estimadores de máxima verossimilhança do modelo Heckman generalizado comparado com os estimadores do modelo de Heckman clássico, Skew-Normal e Heckman-t. Apresentamos também o nível empírico e o poder dos testes da razão de verossimilhanças e do gradiente sob o ajuste dos quatro modelos. Na seção [2.4](#) apresentamos uma aplicação a dados reais. As conclusões do capítulo são apresentados na última seção.

2.2 Formulação do Modelo Heckman Generalizado

Considere o modelo de Heckman conforme descrito em (1.1) a (1.4), tal que

$$\begin{pmatrix} Y_{1i}^* \\ Y_{2i}^* \end{pmatrix} \underset{\text{ind.}}{\sim} \mathcal{N} \left[\begin{pmatrix} \mu_{1i} \\ \mu_{2i} \end{pmatrix}, \begin{pmatrix} \sigma_i^2 & \rho_i \sigma_i \\ \rho_i \sigma_i & 1 \end{pmatrix} \right], i = 1, \dots, n, \quad (2.1)$$

em que $\mu_{1i}, \mu_{2i}, \sigma_i$ e ρ_i são, respectivamente, parâmetros de média, de dispersão e correlação. Logo, consideramos a seguinte estrutura de regressão

$$\begin{aligned} g_1(\mu_{1i}) &= \sum_{j=1}^p x_{ji} \beta_j = \eta_{1i}, & g_2(\mu_{2i}) &= \sum_{j=1}^q w_{ji} \gamma_j = \eta_{2i}, \\ h_1(\sigma_i) &= \sum_{j=1}^r z_{ji} \phi_j = \eta_{3i} & \text{e } h_2(\rho_i) &= \sum_{j=1}^s v_{ji} \kappa_j = \eta_{4i}, \end{aligned} \quad (2.2)$$

em que $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top \in \mathbb{R}^p$, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)^\top \in \mathbb{R}^q$, $\boldsymbol{\phi} = (\phi_1, \dots, \phi_r)^\top \in \mathbb{R}^r$ e $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_s)^\top \in \mathbb{R}^s$ são vetores de parâmetros desconhecidos, independentes e $p+q+r+s = m < n$. Adicionalmente, $\boldsymbol{\eta}_1 = (\eta_{1i}, \dots, \eta_{1n})^\top$, $\boldsymbol{\eta}_2 = (\eta_{2i}, \dots, \eta_{2n})^\top$, $\boldsymbol{\eta}_3 = (\eta_{3i}, \dots, \eta_{3n})^\top$ e $\boldsymbol{\eta}_4 = (\eta_{4i}, \dots, \eta_{4n})^\top$ são preditores lineares e x_{1i}, \dots, x_{pi} , w_{1i}, \dots, w_{qi} , z_{1i}, \dots, z_{ri} e v_{1i}, \dots, v_{si} são observações conhecidas, não necessariamente exclusivas. Quando interceptos são incluídos nos submodelos da média, da dispersão e da correlação, temos que $x_{1i} = w_{1i} = z_{1i} = v_{1i} = 1$, para $i = 1, \dots, n$.

Para o modelo Heckman generalizado com a parametrização proposta vamos utilizar as funções de ligação canônica para a média e de forma a garantir $\rho \in [-1, 1]$ e $\sigma > 0$, considere arco-seno para a correlação e log para a dispersão, respectivamente. Assim, as funções de ligação $g_1(\cdot)$, $g_2(\cdot)$, $h_1(\cdot)$ e $h_2(\cdot)$ são estritamente monótonas e duas vezes diferenciáveis, $g_1(\cdot)$ e $g_2(\cdot)$ com domínio e imagem em \mathbb{R} , $h_1(\cdot)$ com domínio em \mathbb{R}^+ e imagem em \mathbb{R} e $h_2(\cdot)$ com domínio em $[-1, 1]$ e imagem em \mathbb{R} . Discussões mais aprofundadas sobre funções de ligação podem ser encontradas em [Atkinson \(1985\)](#) e [McCullagh e Nelder \(1989\)](#).

2.2.1 Verossimilhança e Vetor Escore

Dado um par de observações (Y_i, U_i) definido conforme (1.3), a densidade de $Y_i|U_i = 1$ é da forma dada na expressão (1.5), com $z_i^\top \boldsymbol{\phi}$ e $v_i^\top \boldsymbol{\kappa}$ substituindo σ e ρ , respectivamente. Para a estimação conjunta dos vetores de parâmetros $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\phi}$ e $\boldsymbol{\kappa}$ utilizamos o método da máxima verossimilhança. O logaritmo da função de verossimilhança para n vetores aleatórios e independentes (Y_i, U_i) e $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\phi}^\top, \boldsymbol{\kappa}^\top)^\top$ é

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^n \mathcal{L}_i(\mu_{1i}, \mu_{2i}, \sigma_i, \rho_i), \quad (2.3)$$

em que

$$\begin{aligned}\mathcal{L}_i(\mu_{1i}, \mu_{2i}, \sigma_i, \rho_i) &= u_i \log f(y_i|U=1) + u_i \log \Phi(\mu_{2i}) + (1-u_i) \log \Phi(-\mu_{2i}) \\ &= u_i \left\{ \log \Phi \left[\frac{\mu_{2i} + \rho_i(y_i - \mu_{1i})/\sigma_i}{\sqrt{1-\rho_i^2}} \right] + \log \phi \left(\frac{y_i - \mu_{1i}}{\sigma_i} \right) - \log \sigma_i \right\} \\ &\quad + (1-u_i) \log \Phi(-\mu_{2i}),\end{aligned}\tag{2.4}$$

com $\mu_{1i} = g_1^{-1}(\eta_{1i})$, $\mu_{2i} = g_2^{-1}(\eta_{2i})$, $\sigma_i = h_1^{-1}(\eta_{3i})$ e $\rho_i = h_2^{-1}(\eta_{4i})$, como definido em (2.2), funções de $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\phi}$ e $\boldsymbol{\kappa}$, respectivamente. Os componentes do vetor escore obtidos pela diferenciação da função log verossimilhança em relação aos parâmetros são

$$\begin{aligned}\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \gamma_j} &= \sum_{i=1}^n \frac{\partial \mathcal{L}_i(\boldsymbol{\theta})}{\partial \mu_{2i}} \frac{\partial \mu_{2i}}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\partial \gamma_j}, \quad j = 1, \dots, p, \\ \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \beta_k} &= \sum_{i=1}^n \frac{\partial \mathcal{L}_i(\boldsymbol{\theta})}{\partial \mu_{1i}} \frac{\partial \mu_{1i}}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \beta_k}, \quad k = 1, \dots, q, \\ \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \phi_l} &= \sum_{i=1}^n \frac{\partial \mathcal{L}_i(\boldsymbol{\theta})}{\partial \sigma_i} \frac{\partial \sigma_i}{\partial \eta_{3i}} \frac{\partial \eta_{3i}}{\partial \phi_l}, \quad l = 1, \dots, r, \\ \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \kappa_m} &= \sum_{i=1}^n \frac{\partial \mathcal{L}_i(\boldsymbol{\theta})}{\partial \rho_i} \frac{\partial \rho_i}{\partial \eta_{4i}} \frac{\partial \eta_{4i}}{\partial \kappa_m}, \quad m = 1, \dots, s,\end{aligned}$$

com,

$$\begin{aligned}\frac{\partial \mu_{1i}}{\partial \eta_{1i}} &= 1, & \frac{\partial \mu_{2i}}{\partial \eta_{2i}} &= 1, & \frac{\partial \sigma_i}{\partial \eta_{3i}} &= \sigma_i, & \frac{\partial \rho_i}{\partial \eta_{4i}} &= \cos(\eta_{4i}), \\ \frac{\partial \eta_{1i}}{\partial \boldsymbol{\beta}} &= \frac{\partial \mathbf{x}_i^\top \boldsymbol{\beta}}{\partial \boldsymbol{\beta}} = \mathbf{x}_i, & \frac{\partial \eta_{2i}}{\partial \boldsymbol{\gamma}} &= \frac{\partial \mathbf{w}_i^\top \boldsymbol{\gamma}}{\partial \boldsymbol{\gamma}} = \mathbf{w}_i, & \frac{\partial \eta_{3i}}{\partial \boldsymbol{\phi}} &= \frac{\partial \mathbf{z}_i^\top \boldsymbol{\phi}}{\partial \boldsymbol{\phi}} = \mathbf{z}_i, & \frac{\partial \eta_{4i}}{\partial \boldsymbol{\kappa}} &= \frac{\partial \mathbf{v}_i^\top \boldsymbol{\kappa}}{\partial \boldsymbol{\kappa}} = \mathbf{v}_i.\end{aligned}$$

Logo,

$$\begin{aligned}\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}} &= u_i \left\{ \frac{\phi \left[\frac{\mu_{2i} + \rho_i(y_i - \mu_{1i})/\sigma_i}{\sqrt{1-\rho_i^2}} \right]}{\Phi \left[\frac{\mu_{2i} + \rho_i(y_i - \mu_{1i})/\sigma_i}{\sqrt{1-\rho_i^2}} \right]} \frac{1}{\sqrt{1-\rho_i^2}} \right\} \mathbf{w}_i - (1-u_i) \frac{\phi(-\mu_{2i})}{\Phi(-\mu_{2i})} \mathbf{w}_i, \\ \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} &= u_i \left\{ -\frac{\phi \left[\frac{\mu_{2i} + \rho_i(y_i - \mu_{1i})/\sigma_i}{\sqrt{1-\rho_i^2}} \right]}{\Phi \left[\frac{\mu_{2i} + \rho_i(y_i - \mu_{1i})/\sigma_i}{\sqrt{1-\rho_i^2}} \right]} \frac{\rho_i}{\sigma_i \sqrt{1-\rho_i^2}} + \frac{y_i - \mu_{1i}}{\sigma_i^2} \right\} \mathbf{x}_i, \\ \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\phi}} &= u_i \left\{ -\frac{\phi \left[\frac{\mu_{2i} + \rho_i(y_i - \mu_{1i})/\sigma_i}{\sqrt{1-\rho_i^2}} \right]}{\Phi \left[\frac{\mu_{2i} + \rho_i(y_i - \mu_{1i})/\sigma_i}{\sqrt{1-\rho_i^2}} \right]} \frac{\rho_i}{\sqrt{1-\rho_i^2}} \frac{y_i - \mu_{1i}}{\sigma_i} + \left(\frac{y_i - \mu_{1i}}{\sigma_i} \right)^2 - 1 \right\} \mathbf{z}_i,\end{aligned}$$

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\kappa}} = u_i \left\{ \frac{\phi \left[\frac{\mu_{2i} + \rho_i(y_i - \mu_{1i})/\sigma_i}{\sqrt{1 - \rho_i^2}} \right]}{\Phi \left[\frac{\mu_{2i} + \rho_i(y_i - \mu_{1i})/\sigma_i}{\sqrt{1 - \rho_i^2}} \right]} \left[\frac{\sigma_i \mu_{2i} \rho_i + (y_i - \mu_{1i})}{\sigma_i \sqrt{(1 - \rho_i^2)^3}} \right] \right\} \cos(\eta_{4i}) v_i.$$

Os estimadores de máxima verossimilhança são obtidos pela solução do sistema $\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0$. Como não existe uma forma analítica fechada para encontrar os EMV's dos parâmetros do modelo, recorremos a métodos numéricos para resolver o sistema de equações. Dessa forma, as estimativas dos parâmetros são obtidas por meio da maximização numérica da função log verossimilhança, utilizando o algoritmo de otimização não-linear Broyden-Fletcher-Goldfarb-Shanno (BFGS), conhecido também como algoritmo quase-Newton BFGS. Para maiores detalhes sobre métodos de maximização numérica e o método BFGS, indicamos [Seiler e Seiler \(1989\)](#), [Press et al. \(1992\)](#) e [Wright e Nocedal \(1999\)](#). Para inferências intervalares e testes de hipóteses é necessário o conhecimento da matriz Hessiana que é apresentada no Apêndice [A](#).

2.3 Comportamento dos Estimadores em Amostras Finitas

Nesta seção desenvolvemos estudos de simulação Monte Carlo para avaliarmos e compararmos as propriedades e o desempenho dos estimadores de Máxima Verossimilhança obtidos usando os modelos Heckman generalizado, Heckman clássico, Skew-Normal e Heckman-t. Para isso, diferentes cenários foram considerados:

- Cenário 1: Simulamos dados considerando um modelo de dispersão fixa, correlação variável, com restrição de exclusão e censura média aproximada de 30%.
- Cenário 2: Simulamos dados considerando um modelo de dispersão variável, correlação fixa, com restrição de exclusão e censura média aproximada de 30%.
- Cenário 3: Simulamos dados considerando um modelo de dispersão e correlação variáveis, com restrição de exclusão e censura média aproximada de 30%.
- Cenário 4: Simulamos dados considerando um modelo de dispersão e correlação variáveis, sem restrição de exclusão e censura média aproximada de 30%.
- Cenário 5: Simulamos dados considerando um modelo de dispersão e correlação variáveis, com restrição de exclusão e censura média aproximada de 50%.

Os cinco cenários possuem características relevantes para uma avaliação mais detalhada. Os Cenários 1 e 2 objetivam justificar a importância da modelagem dos parâmetros de correlação e dispersão, respectivamente. O Cenário 3, tem o objetivo de avaliar o modelo Heckman generalizado e, juntamente com os resultados obtidos nos Cenários 1 e 2, confirmar sua contribuição para a modelagem de dados com problemas de viés de seleção. O Cenário 4 objetiva demonstrar a necessidade e a importância da

restrição de exclusão. Por fim, o Cenário 5 foi incluído para demonstrar a sensibilidade dos modelos de seleção a alta correlação e alto valor de censura. Todos os cenários foram baseados na seguinte estrutura:

$$Y_{1i}^* = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + \epsilon_{1i}, \quad (2.5)$$

$$Y_{2i}^* = \gamma_1 + \gamma_2 x_{1i} + \gamma_3 x_{2i} + \gamma_4 x_{3i} + \epsilon_{2i}, i = 1, \dots, n, \quad (2.6)$$

em que os termos de erros $(\epsilon_{1i}, \epsilon_{2i})$ foram gerados a partir da distribuição normal bivariada definida por,

$$\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \stackrel{\text{ind.}}{\sim} \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_i^2 & \rho_i \sigma_i \\ \sigma_i \rho_i & 1 \end{pmatrix} \right], i = 1, \dots, n, \quad (2.7)$$

com $\log(\sigma_i) = \phi_1 + \phi_2 x_{1i}$ e $\arcsen(\rho_i) = \kappa_1 + \kappa_2 x_{1i}$. Todas as covariáveis foram geradas como amostras aleatórias da distribuição normal padrão ou como amostras aleatórias da distribuição uniforme, dependendo do cenário, e foram mantidas constantes durante todo o experimento. Para cada um dos cenários considerados e que serão apresentados em seguida, simulamos n valores de Y_1^* e Y_2^* e definimos $Y_i = Y_{1i}^* U_i$, em que a variável $U_i = 1\{Y_{2i}^* > 0\}$. Os tamanhos amostrais considerados foram $n = 500, n = 1.000$ e $n = 2.000$ e o número de réplicas de Monte Carlo foi fixado em $N = 1.000$, totalizando 3.000 amostras para cada cenário. Consideramos amostras com tamanho $n \geq 500$, pois na prática, modelos de seleção são ajustados a banco de dados com tamanho superior a $n = 500$, e além disso, para tamanhos amostrais menores, nem sempre é possível realizar uma boa inferência por não haver informações suficientes ou por haver grande impacto na estimação dos parâmetros devido a heterocedasticidade. Os parâmetros foram escolhidos de forma a manter um percentual médio de censura em cada passo Monte Carlo de, aproximadamente, 30% nos 4 primeiros cenários e de 50% no Cenário 5. Simulações piloto mostraram que a escolha dos parâmetros utilizados nas simulações não afetam os resultados, desde que mantenham o mesmo percentual médio de censura.

Aos dados simulados, e que foram gerados a partir dos cinco cenários considerados, ajustamos os modelos Heckman generalizado, Heckman clássico, Skew-Normal e Heckman-t afim de comparar seus resultados. É importante ressaltar que não existem pacotes no *software* R para o ajuste dos modelos Heckman-t e Skew-Normal, portanto, tivemos que escrever todos os códigos (log verossimilhança e vetor escore) baseados nos artigos de [Marchenko e Genton \(2012\)](#) e [Ogundimu e Hutton \(2016\)](#), respectivamente.

Para a estimação dos parâmetros, utilizamos o método BFGS via função $optim(\cdot)$, que é implementada no pacote *stats* do *software* R. Tal função necessita de valores iniciais para o método iterativo. Assim, para cada amostra gerada, o valor inicial dos parâmetros $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^\top$, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)^\top$, ϕ_1 e κ_1 são obtidos a partir do método de dois passos de Heckman. Ressaltamos que para encontrar ϕ_1 e κ_1 utilizamos (1.11) e (1.12), respectivamente. Como não há forma de estimar os demais parâmetros $\phi_i, i = 2, \dots, r$ e $\kappa_j, j = 2, \dots, s$, com o método de dois passos, fixamos $\phi_i = 0, \forall i = 2, \dots, r$ e $\kappa_j = 0, \forall j = 2, \dots, s$. Os valores iniciais para os parâmetros de grau de liberdade (ν), do modelo Heckman-t, e assimetria (λ), do modelo Skew-Normal, foram fixados em $\lambda = 1$ e $\nu = 7$. Tais valores foram escolhidos após algumas simulações piloto.

Para todos os modelos considerados, a fim de examinar o desempenho dos EMV's, calculamos para cada amostra a média empírica das estimativas dos parâmetros e a raiz quadrada do erro quadrático médio (REQM), os quais foram utilizados para verificar a eficiência de cada estimador em relação ao valor verdadeiro dos parâmetros. Para auxiliar nossas conclusões sobre a importância do modelo Heckman generalizado, acrescentamos aos Cenários 1, 2 e 3, a aplicação dos testes da Razão de Verossimilhança (S_{RV}) e gradiente (S_G), com o objetivo de avaliar o poder e o tamanho empírico dos testes em relação a $H_0 : \rho = 0$ contra $H_1 : \rho \neq 0$, sob o ajuste dos quatro modelos considerados.

2.3.1 Cenário 1: Dispersão Fixa e Correlação Variável

Nosso objetivo neste Cenário é usar simulações de Monte Carlo para avaliar o desempenho, em amostras finitas, dos EMV's dos parâmetros do modelo Heckman generalizado comparado ao desempenhos dos modelos, Heckman clássico, Skew-normal e Heckman-t quando os dados possuem correlação variável e são homocedásticos. Avaliamos, principalmente, a importância da modelagem do parâmetro de correlação. Logo, consideramos o modelo com restrição de exclusão definido por

$$Y_{1i}^* = 1.1 + 0.7x_{1i} + 0.1x_{2i} + \epsilon_{1i}, \quad (2.8)$$

$$Y_{2i}^* = 0.8 + 0.5x_{1i} + 1.1x_{2i} + 0.6x_{3i} + \epsilon_{2i}, i = 1, \dots, n, \quad (2.9)$$

em que os termos de erros $(\epsilon_{1i}, \epsilon_{2i})$ foram gerados a partir da distribuição normal bivariada dada por

$$\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \stackrel{ind.}{\sim} \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho_i \sigma \\ \sigma \rho_i & 1 \end{pmatrix} \right], \text{ para } i = 1, \dots, n, \quad (2.10)$$

com $\sigma = 0.7$, $\arcsen(\rho_i) = 0.3 + 0.2x_{1i}$ e os regressores $x_{ji} \stackrel{iid}{\sim} \mathcal{N}(0, 1), j = 1, 2, 3$, para todo i .

A Tabela 2.1 apresenta, a média das estimativas obtidas e a raiz quadrada do erro quadrático médio, após o ajuste dos 4 modelos utilizados e para todos os tamanhos amostrais considerados. Notamos que quanto maior o tamanho da amostra menores são os valores da REQM, indicando uma consistência dos EMV's sob o ajuste de todos os modelos. Ainda nesse contexto, observamos que apesar dos valores de correlação obtidos pelos modelos Heckman clássico, Skew-normal e Heckman-t não serem comparáveis com κ_1 , os EMV's destes modelos também mostraram-se consistentes na estimação do parâmetro de correlação. O modelo Heckman generalizado apresenta estimativas não viesadas, o modelo Heckman clássico apresenta as estimativas β_2 e β_3 levemente viesadas, e os modelos Heckman-t e Skew-normal apresenta grande parte das estimativas viesadas. Este primeiro resultado indica que a correlação variável não afeta, significativamente, a estimação dos parâmetros e, portanto, um modelo mais parcimonioso, tal como o modelo de Heckman clássico pode ser preferível, em relação ao modelo Heckman generalizado. Ressaltamos porém, que o acréscimo de covariáveis ao parâmetro de correlação pode ser de grande utilidade para um pesquisador, pois permite identificar o fator gerador de viés. Além disso, notamos que as estimativas obtidas via modelo Heckman generalizado são similares ao modelo de Heckman clássico, o

que permite a mesma interpretação dos parâmetros.

As Figuras 2.1, 2.2 e 2.3 apresentam os box-plots das estimativas obtidas sob o ajuste dos modelos (i) Heckman generalizado, (ii) Heckman clássico, (iii) Skew-Normal e (iv) Heckman-t, observamos que os estimadores do modelo Skew-normal apresentam estimativas viesadas e com baixa precisão dos interceptos e dos parâmetros β_2 e σ . E os modelos Heckman clássico e Heckman-t apresentam estimativas viesadas do parâmetro β_2 . O modelo Heckman generalizado possui maior precisão e menor dispersão das estimativas em torno dos verdadeiros valores dos parâmetros. Não apresentamos o box-plot das estimativas do parâmetro κ_1 , pois as estimativas desse parâmetro obtidas sob o ajuste dos modelos Heckman clássico, Skew-Normal e Heckman-t não são comparáveis com as estimativas obtidas sob o ajuste do modelo Heckman generalizado.

Tabela 2.1: Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman generalizado, Heckman clássico, Skew-Normal e Heckman-t aos dados gerados conforme Cenário 1. Tamanhos amostrais $n = 500, n = 1.000$ e $n = 2.000$ com $N = 1.000$ réplicas de Monte Carlo.

Parâmetros	n	Heckman Generalizado		Heckman Clássico		Heckman Skew		Heckman t-Student		
		Média	REQM	Média	REQM	Média	REQM	Média	REQM	
γ_1	0.800	500	0.809	0.085	0.809	0.085	0.737	0.163	0.825	0.092
		1.000	0.804	0.062	0.803	0.061	0.733	0.128	0.816	0.065
		2.000	0.802	0.042	0.800	0.042	0.736	0.104	0.810	0.045
γ_2	0.500	500	0.505	0.082	0.501	0.083	0.498	0.083	0.512	0.086
		1.000	0.504	0.057	0.498	0.058	0.496	0.058	0.507	0.059
		2.000	0.503	0.038	0.497	0.039	0.495	0.040	0.503	0.040
γ_3	1.100	500	1.117	0.103	1.118	0.103	1.108	0.104	1.140	0.114
		1.000	1.109	0.076	1.109	0.077	1.103	0.077	1.127	0.083
		2.000	1.106	0.053	1.106	0.053	1.101	0.053	1.120	0.058
γ_4	0.600	500	0.609	0.082	0.610	0.082	0.605	0.082	0.623	0.088
		1.000	0.604	0.062	0.605	0.062	0.602	0.062	0.615	0.065
		2.000	0.604	0.041	0.606	0.041	0.603	0.041	0.614	0.044
β_1	1.100	500	1.104	0.074	1.106	0.078	0.891	0.313	1.105	0.078
		1.000	1.101	0.053	1.108	0.058	0.894	0.279	1.107	0.058
		2.000	1.101	0.036	1.107	0.038	0.913	0.252	1.106	0.038
β_2	0.700	500	0.699	0.054	0.738	0.058	0.738	0.058	0.737	0.058
		1.000	0.700	0.037	0.735	0.047	0.735	0.047	0.734	0.046
		2.000	0.699	0.026	0.735	0.041	0.735	0.041	0.735	0.041
β_3	0.100	500	0.098	0.058	0.088	0.061	0.087	0.061	0.089	0.061
		1.000	0.099	0.042	0.086	0.047	0.085	0.047	0.086	0.047
		2.000	0.099	0.029	0.088	0.033	0.087	0.033	0.088	0.033
σ	0.700	500	0.700	0.030	0.695	0.030	0.758	0.101	0.680	0.038
		1.000	0.700	0.020	0.696	0.020	0.747	0.079	0.684	0.027
		2.000	0.700	0.015	0.696	0.015	0.739	0.063	0.686	0.021
κ_1	0.300	500	0.297	0.219	0.244	0.233	0.256	0.245	0.247	0.231
		1.000	0.303	0.158	0.229	0.185	0.240	0.190	0.230	0.183
		2.000	0.302	0.104	0.226	0.136	0.237	0.136	0.227	0.136
κ_2	0.200	500	0.214	0.167	—	—	—	—	—	—
		1.000	0.206	0.109	—	—	—	—	—	—
		2.000	0.205	0.073	—	—	—	—	—	—

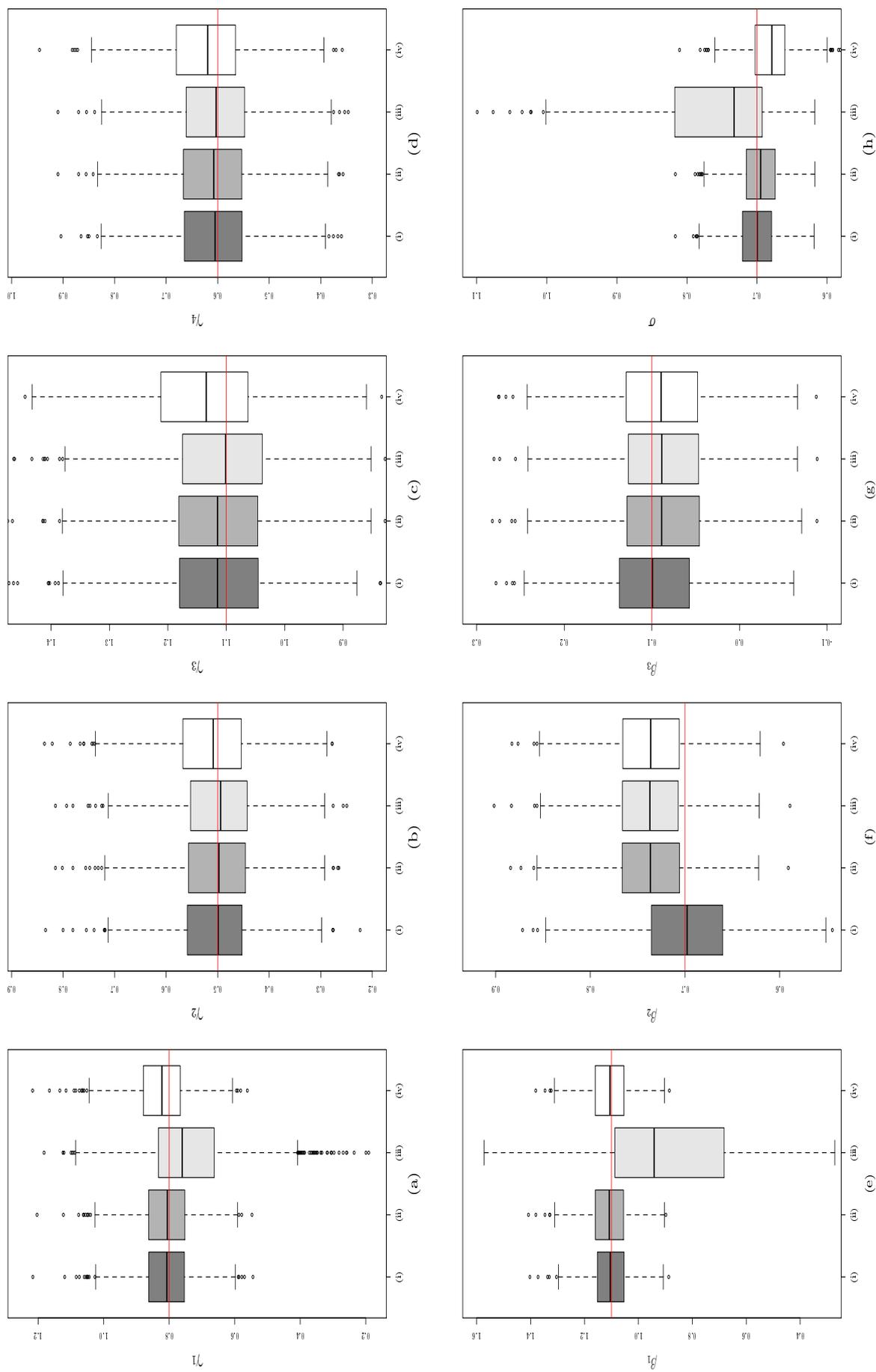


Figura 2.1: Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 , (g) β_3 e (h) β_4 através do ajuste dos modelos (i) Heckman Generalizado, (ii) Heckman Clássico, (iii) Skew-Normal e (iv) Heckman-t. Tamanho amostral $n = 500$.

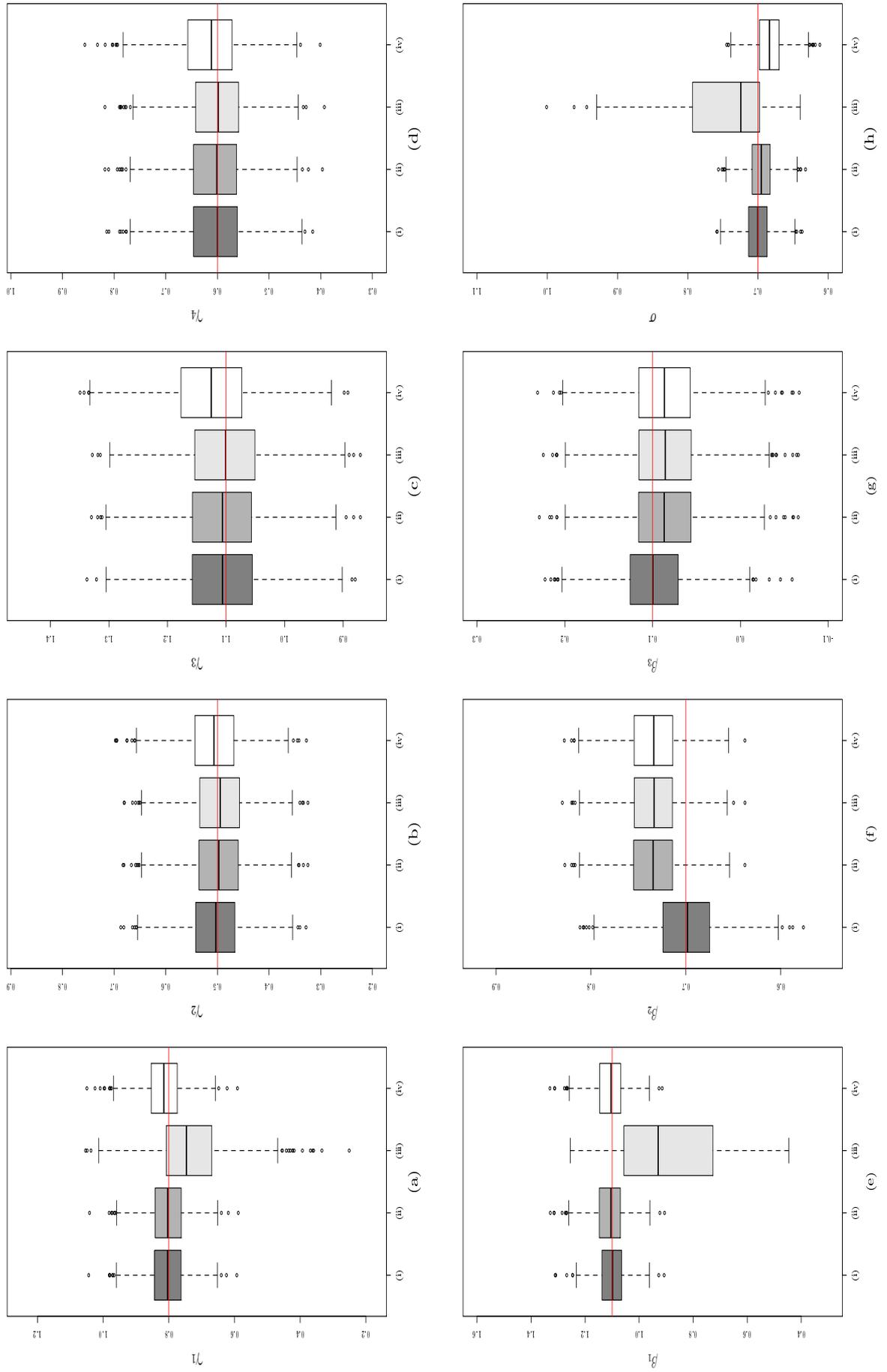


Figura 2.2: Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 , (g) β_3 e (h) β_4 através do ajuste dos modelos (i) Heckman Generalizado, (ii) Heckman Clássico, (iii) Skew-Normal e (iv) Heckman-t. Tamanho amostral $n = 1.000$.

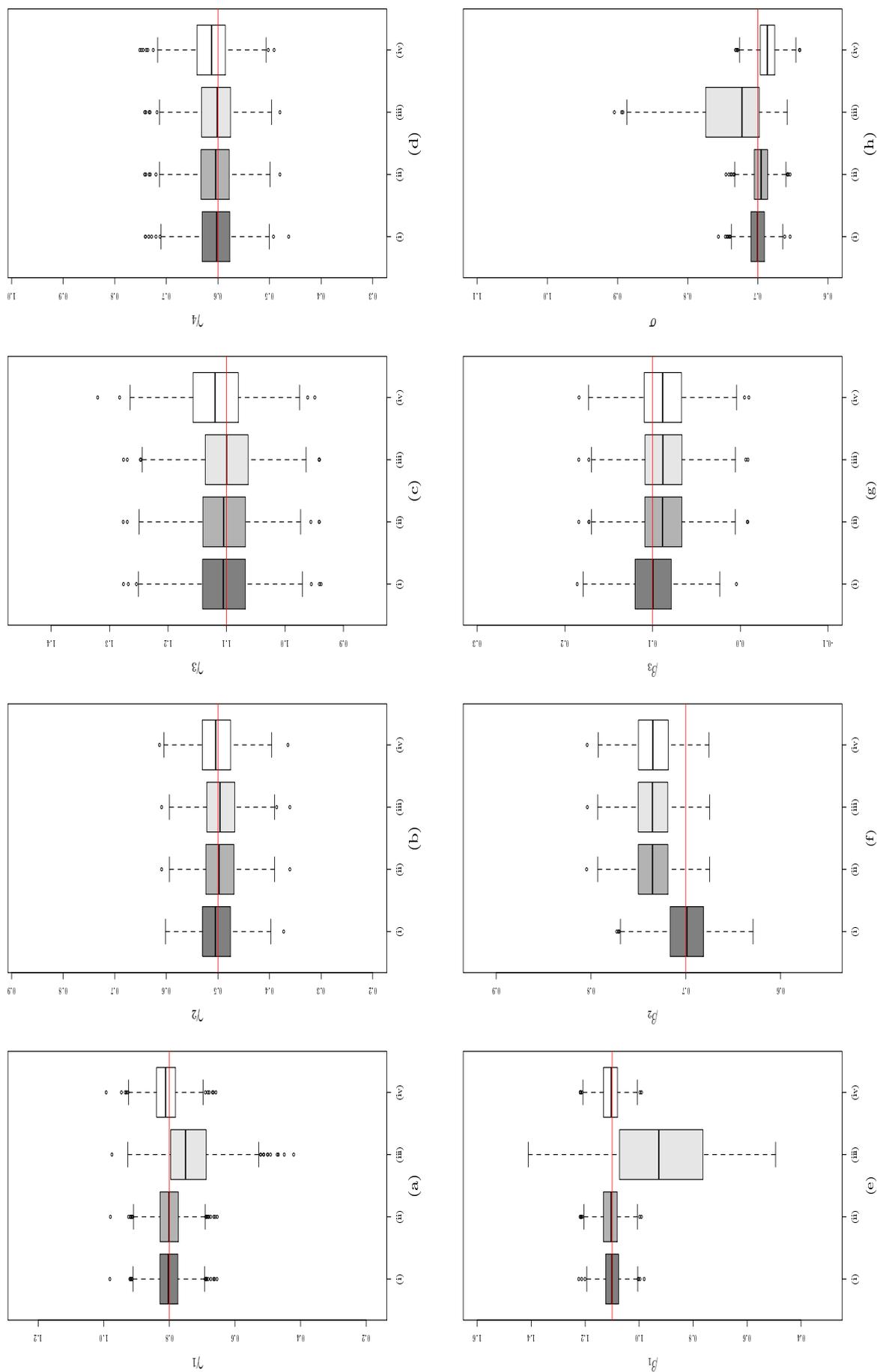


Figura 2.3: Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 , (g) β_3 e (h) σ através do ajuste dos modelos (i) Heckman Generalizado, (ii) Heckman Clássico, (iii) Skew-Normal e (iv) Heckman-t. Tamanho amostral $n = 2.000$.

Tamanho e Poder dos Testes da Razão de Verossimilhança e Gradiente

Apresentamos na Tabela 2.2, o cálculo do tamanho empírico dos testes razão de verossimilhança e gradiente, sob o ajuste dos modelos Heckman generalizado, Heckman clássico, Skew-Normal e Heckman-t. Foram geradas, sob a hipótese $H_0 : \rho_i = 0$ contra $H_1 : \rho_i \neq 0$, $N = 1.000$ amostras de tamanhos $n = 500, n = 1.000$ e $n = 2.000$. Para cada n e cada um dos modelos estudados, a proporção de vezes, entre as $N = 1.000$ amostras, em que os testes levam a falsa rejeição de H_0 , define o tamanho empírico dos testes. Consideramos níveis nominais de 1%, 5% e 10%. Ressaltamos que no caso do modelo Heckman generalizado, neste primeiro cenário, testar $H_0 : \rho_i = 0$ é equivalente a testar $H_0 : \kappa_1 = \kappa_2 = 0$.

Os valores das taxas de erro tipo I, dos testes da razão de verossimilhança e gradiente, sob o ajuste de todos os modelos, se aproximaram do nível de significância nominal, à medida que o tamanho amostral aumenta. Ressaltamos que sob H_0 , os dados são simulados com correlação fixa e igual a zero. Assim, para tamanhos amostrais menores e dados simulados conforme Cenário 1, resultados menos precisos do modelo Heckman generalizado são esperados, uma vez que nosso modelo parametriza a correlação e, por consequência, necessitamos estimar mais parâmetros que os demais modelos.

Tabela 2.2: Nível de significância empírico dos testes da razão de verossimilhança e gradiente para $H_0 : \kappa_1 = \kappa_2 = 0$ e censura média de 33.2%.

Valor nominal do teste	Tamanho da Amostra	Heckman Generalizado		Heckman Clássico		Heckman Skew		Heckman t-Student	
		S_{RV}	S_G	S_{RV}	S_G	S_{RV}	S_G	S_{RV}	S_G
1%	$n = 500$	2	0.5	1.1	0.3	1.1	0.3	1.1	0.2
	$n = 1.000$	1.4	0.7	1.2	0.4	1	0.6	1.2	0.5
	$n = 2.000$	1.1	0.5	1.1	0.5	1.1	0.5	1	0.6
5%	$n = 500$	8	4.2	5.3	3.2	5.6	3.2	5.2	3.4
	$n = 1.000$	6.7	4.2	5.2	4.5	5.3	4.4	5.3	4.5
	$n = 2.000$	5.2	4.7	4.6	4.3	4.5	3.9	4.1	4.3
10%	$n = 500$	13.1	9.6	10.6	9.1	10.8	8.8	10.4	9.1
	$n = 1.000$	12.4	10.3	10.6	9.4	10.4	9.5	10.2	9.6
	$n = 2.000$	10.9	10.2	9.7	9.2	9.8	9.4	9	9.3

Na Tabela 2.3 apresentamos o poder empírico dos testes. Notamos, que a medida que aumentamos o tamanho da amostra, os testes apresentam melhores resultados e as taxas de rejeição aumentam quando nos distanciamos de H_0 . Além disso, como é esperado, com o aumento do nível de significância houve um aumento do poder dos testes RV e gradiente. O teste da razão de verossimilhanças, porém, é visivelmente melhor sob o ajuste de todos os modelos. Sob o ajuste do modelo Heckman generalizado temos uma taxa de rejeição maior em todos os casos, ou seja, os testes se mostram mais poderosos sob o ajuste do nosso modelo. Tal resultado indica que o modelo Heckman generalizado é uma boa alternativa para avaliar a presença de viés de seleção nos dados e pode ser utilizado em conjunto com o modelo Heckman clássico para obtenção de resultados mais confiáveis.

Tabela 2.3: Poder dos testes da razão de verossimilhança (S_{RV}) e gradiente (S_G), em porcentagem, para dados simulados de acordo com o Cenário 1 e com o ajuste dos modelos de Heckman generalizado, Heckman Clássico, Skew-Normal e Heckman-t, com nível nominal de 1%, 5% e 10%.

Parâmetros	Tamanho da Amostra	Heckman Generalizado		Heckman Clássico		Heckman Skew		Heckman t-Student	
		S_{RV}	S_G	S_{RV}	S_G	S_{RV}	S_G	S_{RV}	S_G
$\alpha = 1\%$									
$\kappa_1 = -0.7$	$n = 500$	93.5	81.5	29.2	18.4	30	18.9	29.8	19.7
	$n = 1.000$	99.6	99.4	46	35.6	46.6	36.6	46.6	37.2
$\kappa_2 = -0.7$	$n = 2.000$	100	100	74.4	72.2	74.5	72.5	76	74.1
$\kappa_1 = -0.3$	$n = 500$	31.2	13.7	5.2	3.1	5.8	2.9	5.1	3
	$n = 1.000$	60.4	50.4	7.6	5.6	8.1	5.8	7.8	5.7
$\kappa_2 = -0.3$	$n = 2.000$	94.6	92	16.9	14.7	17.2	14.7	16.2	14.7
$\kappa_1 = 0.3$	$n = 500$	35.4	16.7	6.2	3.4	5.8	2.7	6.3	3.6
	$n = 1.000$	60.9	48.7	9.2	5.8	8.6	5.3	9	6.1
$\kappa_2 = 0.3$	$n = 2.000$	93.9	91.6	18	15.8	18	15.3	17.9	15.9
$\kappa_1 = 0.7$	$n = 500$	94.2	82	30	19.8	29	15.2	30.6	21.2
	$n = 1.000$	99.9	98.8	42.5	35.6	41.3	31.6	44.8	37.3
$\kappa_2 = 0.7$	$n = 2.000$	100	100	73.8	72	73.2	69.7	75.9	73.7
$\alpha = 5\%$									
$\kappa_1 = -0.7$	$n = 500$	98.1	94	51	45.7	51.4	45.6	52.3	47.6
	$n = 1.000$	100	99.8	66.5	62.9	66.3	63.5	67.6	64.7
$\kappa_2 = -0.7$	$n = 2.000$	100	100	89.4	88.5	89.3	88.6	90.5	90.3
$\kappa_1 = -0.3$	$n = 500$	56.9	40	16.8	12.8	17.3	13.2	16.6	13.2
	$n = 1.000$	78.7	73	21.2	17.9	21.5	18.5	21.3	18.4
$\kappa_2 = -0.3$	$n = 2.000$	98.9	98.5	36.7	35.3	36.9	35.5	35.7	35.7
$\kappa_1 = 0.3$	$n = 500$	57.2	43.4	16.8	13.8	17.2	12.9	16.2	13.5
	$n = 1.000$	81.6	75.2	21.7	19.7	21.4	19	21	19.5
$\kappa_2 = 0.3$	$n = 2.000$	98.4	97.9	37.9	36.3	37.3	35.3	36.6	35.9
$\kappa_1 = 0.7$	$n = 500$	98.4	95.2	51.9	45.2	50.6	42.1	53.4	46.5
	$n = 1.000$	100	99.9	66.7	63.2	65.3	60	67.6	64.8
$\kappa_2 = 0.7$	$n = 2.000$	100	100	88.5	87.8	87.8	86.7	89.7	89.7
$\alpha = 10\%$									
$\kappa_1 = -0.7$	$n = 500$	99.9	97.7	63.5	60.2	64.1	60.3	65.5	62
	$n = 1.000$	100	100	74.2	72.7	74.5	73	76	74.9
$\kappa_2 = -0.7$	$n = 2.000$	100	100	94	93.8	94	93.8	94.8	94.8
$\kappa_1 = -0.3$	$n = 500$	69.7	57	25.6	22.6	26.3	23.2	25.2	22.8
	$n = 1.000$	86.5	83.5	30.6	29.1	31.3	29.4	29.7	28.9
$\kappa_2 = -0.3$	$n = 2.000$	99.7	99.5	48.2	47.4	48	47	47.2	47.5
$\kappa_1 = 0.3$	$n = 500$	69.1	59	26.7	23.7	26.8	22.2	26.4	24.2
	$n = 1.000$	88.1	86	31	29.8	31.4	30.1	31.8	30.6
$\kappa_2 = 0.3$	$n = 2.000$	99.3	99.2	48.9	48.3	48.8	48	48.2	48.3
$\kappa_1 = 0.7$	$n = 500$	99.6	98.1	63.7	59.5	62.1	56.9	65	62.2
	$n = 1.000$	100	100	74.8	73.1	74	71.8	76.2	75
$\kappa_2 = 0.7$	$n = 2.000$	100	100	93.6	93.2	93.2	92.8	94.2	94.1

2.3.2 Cenário 2: Dispersão Variável e Correlação Fixa

Nosso objetivo aqui é mostrar que na presença de heterocedasticidade os modelos existentes na literatura produzem estimadores inconsistentes e estimativas viesadas, e a modelagem do parâmetro de dispersão com o acréscimo de covariáveis pode ser uma boa alternativa ao ajuste de tais dados. Para isso, consideramos as equações dadas por:

$$Y_{1i}^* = 1.1 + 0.7x_{1i} + 0.1x_{2i} + \epsilon_{1i}, \quad (2.11)$$

$$Y_{2i}^* = -0.2 + 0.5x_{1i} + 1.1x_{2i} + 0.6x_{3i} + \epsilon_{2i}, \text{ para } i = 1, \dots, n, \quad (2.12)$$

em que, os termos de erros $(\epsilon_{1i}, \epsilon_{2i})$ foram gerados a partir da distribuição normal bivariada dada por:

$$\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \underset{ind.}{\sim} \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_i^2 & \rho\sigma_i \\ \sigma_i\rho & 1 \end{pmatrix} \right], \text{ para } i = 1, \dots, n, \quad (2.13)$$

com $\log(\sigma_i) = 0.7 + 3x_{1i}$, $\rho = 0.7$ e x_{1i} , x_{2i} gerados independentemente da distribuição uniforme padrão, x_{3i} gerado a partir da normal padrão e todos mantidos fixos ao longo das simulações.

Neste segundo cenário, o parâmetro de correlação foi fixado e geramos amostras de Y_1^* e Y_2^* com dispersão variável. Da mesma forma como no Cenário 1, ajustamos os modelos Heckman generalizado, Heckman clássico, Skew-Normal e Heckman-t aos dados simulados com o objetivo de comparar seus resultados. Na Tabela 2.4 apresentamos resultados numéricos referentes às estimativas obtidas de acordo com o Cenário 2. Os EMV's obtidos sob ajuste do modelo Heckman generalizado se mostraram consistentes e suas estimativas apresentaram menor viés. Notamos ainda que o modelo Heckman generalizado é o único que estima bem os interceptos do modelo e o parâmetro de correlação. É importante ressaltar que todas as estimativas obtidas sob ajuste dos modelos Heckman clássico, Skew-Normal e Heckman-t foram altamente viesadas, o que indica um forte impacto negativo da heterocedasticidade na estimação de parâmetros sob ajuste de tais modelos, mesmo que os erros possuam distribuição normal bivariada e sejam corretamente especificados.

Além disso, verificamos nas Figuras 2.4, 2.5 e 2.6 que as estimativas de máxima verossimilhança do modelo Heckman generalizado tendem ao verdadeiro valor do parâmetro e a variabilidade das mesmas diminui quando aumentamos o tamanho amostral, o que sugere que as estimativas obtidas com o ajuste desse modelo apresentam boa precisão. Notamos que sob o ajuste dos demais modelos os parâmetros são sobre ou super-estimados e, portanto, para dados heterocedásticos os modelos de Heckman clássico, Skew-Normal e Heckman-t não são indicados. Não apresentamos o box-plot das estimativas do parâmetro ϕ_1 e ϕ_2 , pois as estimativas de ϕ_1 obtidas sob o ajuste dos modelos de Heckman clássico, Skew-Normal e Heckman-t não são comparáveis com as estimativas obtidas sob o ajuste do modelo Heckman generalizado e obtemos estimativas de ϕ_2 somente sob o ajuste deste último.

Tabela 2.4: Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman generalizado, Heckman clássico, Skew-Normal e Heckman-t aos dados gerados conforme Cenário 2. Tamanhos amostrais $n = 500, n = 1.000$ e $n = 2.000$ com $N = 1.000$ réplicas de Monte Carlo.

Parâmetros	n	Heckman Generalizado		Heckman Clássico		Heckman Skew		Heckman t-Student		
		Média	REQM	Média	REQM	Média	REQM	Média	REQM	
γ_1	-0.200	500	-0.205	0.169	-0.314	0.202	-0.445	0.327	-0.269	0.198
		1.000	-0.207	0.116	-0.313	0.162	-0.466	0.327	-0.267	0.146
		2.000	-0.199	0.081	-0.308	0.135	-0.419	0.258	-0.259	0.108
γ_2	0.500	500	0.508	0.232	0.747	0.352	0.702	0.329	0.666	0.322
		1.000	0.506	0.154	0.743	0.300	0.697	0.270	0.660	0.251
		2.000	0.498	0.104	0.754	0.282	0.731	0.267	0.657	0.206
γ_3	1.100	500	1.112	0.228	1.091	0.232	1.059	0.239	1.234	0.301
		1.000	1.111	0.152	1.083	0.153	1.047	0.167	1.230	0.220
		2.000	1.104	0.107	1.066	0.114	1.049	0.125	1.216	0.172
γ_4	0.600	500	0.609	0.071	0.583	0.077	0.568	0.087	0.676	0.118
		1.000	0.602	0.052	0.573	0.063	0.555	0.077	0.666	0.095
		2.000	0.603	0.034	0.568	0.049	0.560	0.059	0.667	0.081
β_1	1.100	500	1.101	0.222	0.468	0.741	0.002	1.318	0.833	0.444
		1.000	1.096	0.173	0.405	0.757	-0.146	1.415	0.804	0.411
		2.000	1.105	0.109	0.446	0.683	0.063	1.144	0.874	0.291
β_2	0.700	500	0.705	0.392	1.552	0.959	1.459	0.903	1.221	0.687
		1.000	0.707	0.254	1.567	0.923	1.484	0.859	1.211	0.596
		2.000	0.695	0.175	1.497	0.828	1.461	0.799	1.142	0.492
β_3	0.100	500	0.100	0.306	0.209	0.461	0.199	0.457	0.110	0.420
		1.000	0.102	0.223	0.247	0.361	0.234	0.358	0.136	0.311
		2.000	0.100	0.152	0.261	0.286	0.256	0.283	0.123	0.222
ϕ_1	0.700	500	0.684	0.110	2.431	1.739	2.553	1.874	1.983	1.297
		1.000	0.694	0.069	2.478	1.782	2.622	1.935	2.015	1.323
		2.000	0.694	0.045	2.462	1.763	2.528	1.835	1.968	1.271
ϕ_2	3.000	500	3.023	0.307	-	-	-	-	-	-
		1.000	3.009	0.198	-	-	-	-	-	-
		2.000	3.007	0.132	-	-	-	-	-	-
ρ	0.700	500	0.697	0.111	0.746	0.122	0.752	0.131	0.680	0.128
		1.000	0.699	0.078	0.757	0.095	0.767	0.104	0.688	0.095
		2.000	0.697	0.051	0.758	0.077	0.763	0.083	0.669	0.074

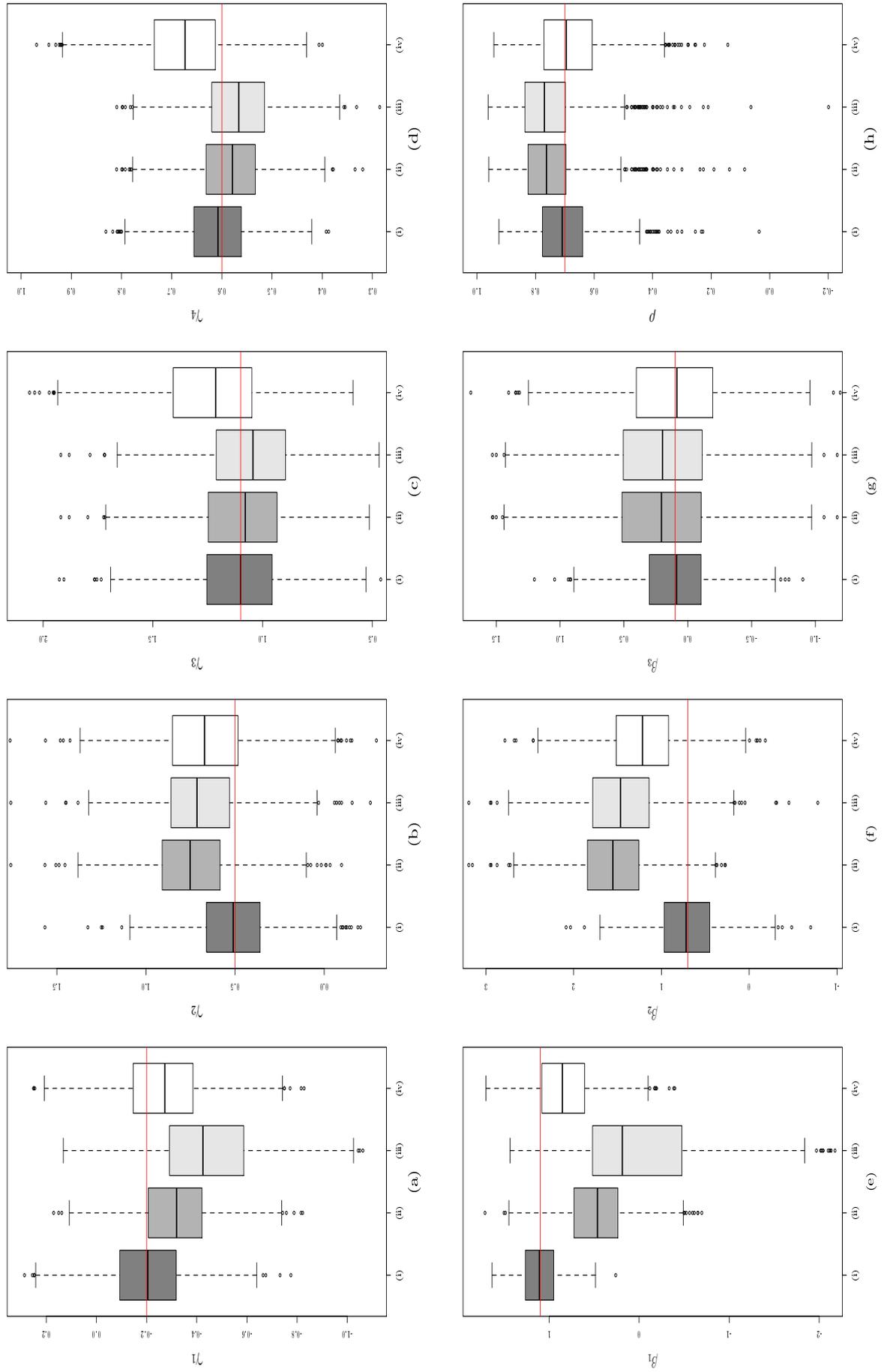


Figura 2.4: Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 , (g) β_3 e (h) β_4 a partir do ajuste dos modelos (i) Heckman Generalizado, (ii) Heckman Clássico, (iii) Skew-Normal e (iv) Heckman-t. Tamanho amostral $n = 500$.

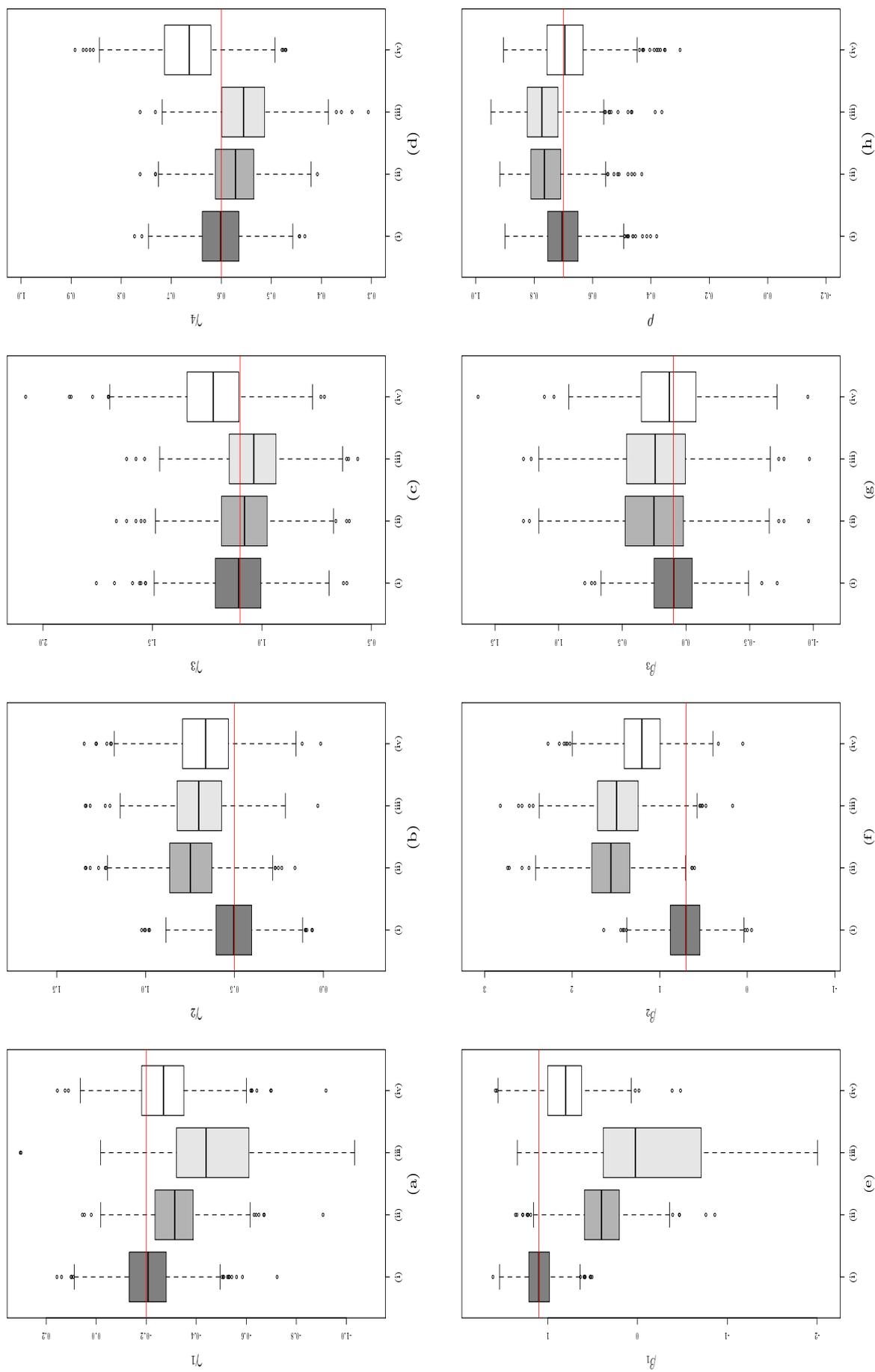


Figura 2.5: Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 , (g) β_3 e (h) β_4 a partir do ajuste dos modelos (i) Heckman Generalizado, (ii) Heckman Clássico, (iii) Skew-Normal e (iv) Heckman-t. Tamanho amostral $n = 1.000$.

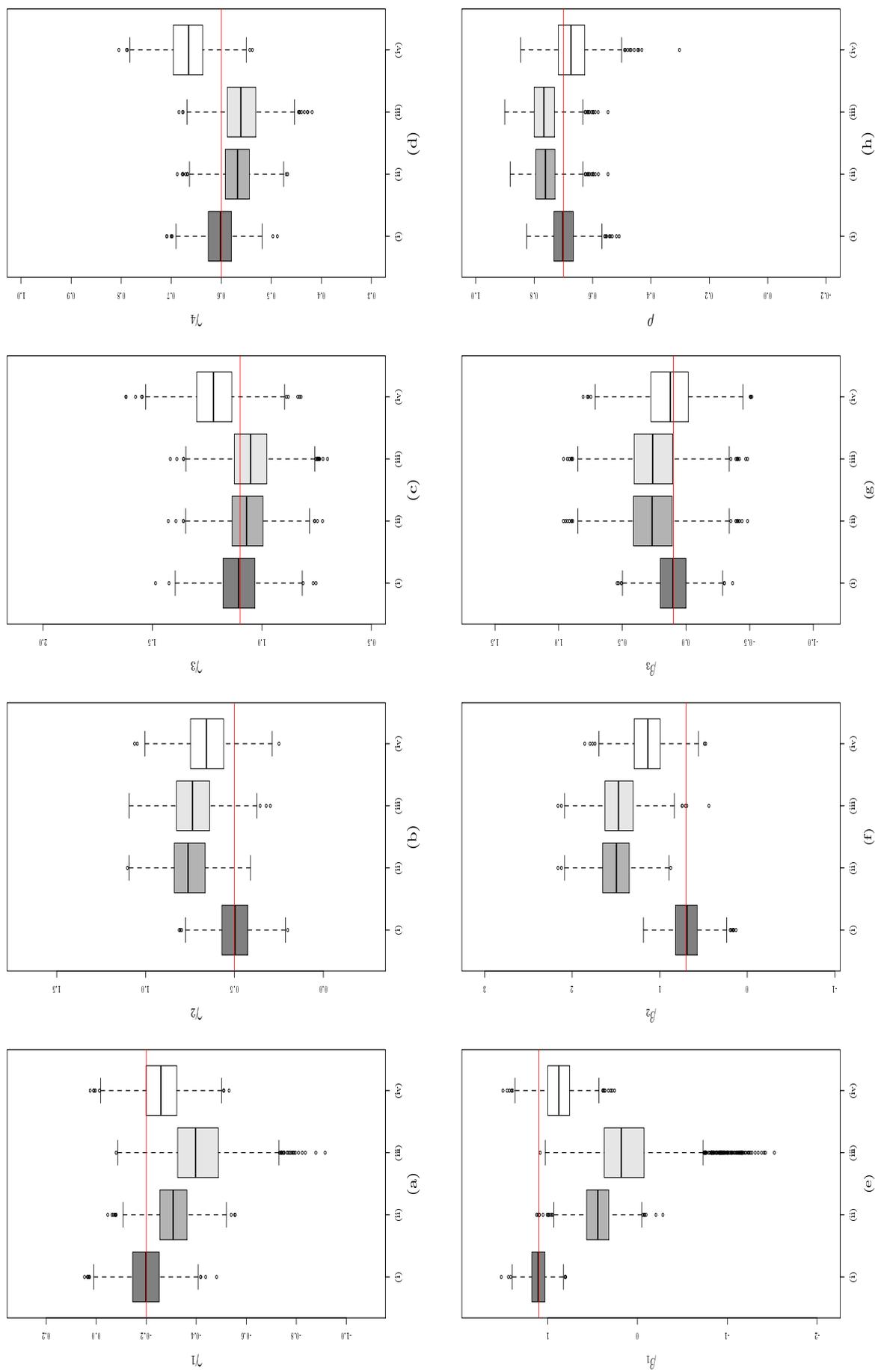


Figura 2.6: Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 , (g) β_3 e (h) β_4 a partir do ajuste dos modelos (i) Heckman Generalizado, (ii) Heckman Clássico, (iii) Skew-Normal e (iv) Heckman-t. Tamanho amostral $n = 2.000$.

Tamanho e Poder dos Testes da Razão de Verossimilhança e Gradiente

Apresentamos na Tabela 2.5 o cálculo do tamanho empírico dos testes razão de verossimilhança e gradiente sob ajuste dos modelos Heckman generalizado, Heckman clássico, Skew-Normal e Heckman-t. Novamente foram geradas, sob a hipótese $H_0 : \rho = 0$ contra $H_1 : \rho \neq 0$, $N = 1.000$ amostras de tamanhos $n = 500$, $n = 1.000$ e $n = 2.000$, e foram considerados níveis nominais de 1%, 5% e 10%.

No caso de erros não correlacionados ($\rho = 0$) e dispersão variável, observamos que os testes da razão de verossimilhança e gradiente mantiveram níveis nominais próximos dos valores corretos de tamanho empírico somente sob ajuste do modelo Heckman generalizado. Sob o ajuste dos demais modelos, os testes se mostraram mais sensíveis e inflacionam o erro tipo I. No entanto, mesmo sob ajuste do modelo Heckman generalizado, houve pequenas variações que contrariam a expectativa teórica, uma vez que, fixado o teste juntamente com o modelo utilizado e o tamanho empírico, observamos tamanho do teste próximo do valor nominal para $n = 500$ e tamanho empírico mais distante do valor nominal para $n = 1.000$ e $n = 2.000$. Entendemos que isso ocorre devido à sensibilidade dos testes diante da variação da censura, que apresentou valores médios de 31.56% para $n = 500$, de 32.14% para $n = 1.000$ e de 30.50% para tamanho amostral $n = 2.000$. E também devido a presença de heterocedasticidade. Ressaltamos que ambos os problemas podem afetar a estimação dos parâmetros e indicar, erroneamente, a presença ou ausência de viés de seleção.

Tabela 2.5: Nível de significância empírico dos testes da razão de verossimilhança e gradiente para $H_0 : \rho = 0$.

Valor nominal do teste	Tamanho da Amostra	Heckman Generalizado		Heckman Clássico		Heckman Skew		Heckman t-Student	
		S_{RV}	S_G	S_{RV}	S_G	S_{RV}	S_G	S_{RV}	S_G
1%	$n = 500$	1	0.5	4.8	1	9.5	0.8	2	1
	$n = 1.000$	1.2	1	3.9	2.9	6.2	5.1	2.7	2
	$n = 2.000$	1.4	1.4	3.9	3.4	2	2	2.3	1.8
5%	$n = 500$	5.5	4.2	11.6	7.9	19.3	8.2	7.6	6.3
	$n = 1.000$	5.5	4.8	12	10.9	13.9	13.1	8.4	7.2
	$n = 2.000$	5.5	5.1	11.7	11.5	6.6	8.8	7.8	7.1
10%	$n = 500$	11.4	9.9	19.6	16.6	28.7	17.1	13.5	12.4
	$n = 1.000$	9.5	9.2	19.6	19.1	21	20.2	14.8	14.3
	$n = 2.000$	11.9	11.7	18.9	18.5	9.9	15.3	13.4	13

Na Tabela 2.6 apresentamos resultados do poder dos testes da razão de verossimilhança e gradiente, sob o ajuste dos 4 modelos, nas quais observamos comportamento similares dos testes, com uma vantagem para o teste da razão de verossimilhança. Apesar de observarmos testes mais poderosos sob o ajuste dos modelos Heckman clássico, Skew-Normal e Heckman-t, tais resultados não parecem ser confiáveis, já que baseado nos resultados da Tabela 2.5, sob ajuste de tais modelos, os testes podem inflacionar o poder de rejeição. Portanto, consideramos os resultados dos testes mais confiáveis quando realizados sob ajuste do modelo Heckman generalizado.

Tabela 2.6: Tamanho empírico e poder dos testes da razão de verossimilhança (S_{RV}) e gradiente (S_G), em porcentagem, para dados simulados de acordo com o Cenário 2 e com o ajuste dos modelos de Heckman generalizado, Heckman Clássico, Skew-Normal e Heckman-t, com nível nominal de 1%, 5% e 10%.

Parâmetros	Tamanho da Amostra	Heckman Generalizado		Heckman Clássico		Heckman Skew		Heckman t-Student	
		S_{RV}	S_G	S_{RV}	S_G	S_{RV}	S_G	S_{RV}	S_G
$\alpha = 1\%$									
$\rho = -0.7$	$n = 500$	94.8	91.2	94.5	89.5	93.2	83.8	90.5	84.7
	$n = 1.000$	99.9	99.9	99.8	99.8	99.8	99.9	99.7	99.5
	$n = 2.000$	100	100	100	100	99.9	100	100	100
$\rho = -0.3$	$n = 500$	14.1	10	24.2	13.7	29.5	13.8	15.5	9.8
	$n = 1.000$	30.8	27.8	43.2	39	54.9	52	30.4	26.8
	$n = 2.000$	57.3	55.1	63.7	59.6	45.5	52.6	44.6	41.8
$\rho = 0.3$	$n = 500$	14.8	10.9	22.9	13.3	20.5	8.2	14.9	9
	$n = 1.000$	33.1	29.9	42.1	37	35.7	25.9	29.6	26.8
	$n = 2.000$	59.8	58.2	67.3	64.9	49.4	49.1	47.8	44.9
$\rho = 0.7$	$n = 500$	93	90.7	94.6	89.5	91.3	76.2	88.8	82.5
	$n = 1.000$	99.9	99.8	99.9	99.9	99.6	99	99.4	99.3
	$n = 2.000$	100	100	100	100	100	100	100	100
$\alpha = 5\%$									
$\rho = -0.7$	$n = 500$	99	99	98	97.6	98	96.9	96.8	95.8
	$n = 1.000$	99.9	99.9	99.9	100	99.9	100	99.8	99.8
	$n = 2.000$	100	100	100	100	100	100	100	100
$\rho = -0.3$	$n = 500$	31	28.6	40.2	34.9	47.6	35.9	29.7	26.2
	$n = 1.000$	54.8	53.4	61.6	59.7	72.1	71.8	49.4	47.6
	$n = 2.000$	78.9	78.3	79.7	79	61.1	77	66.9	65.9
$\rho = 0.3$	$n = 500$	34.1	30.8	41.6	35.3	41.3	32.1	30.1	27.9
	$n = 1.000$	55.7	54.1	62	60.1	54.2	50.4	50.5	49.2
	$n = 2.000$	80.8	80.3	83.5	82.9	65.2	73.2	70.4	69.3
$\rho = 0.7$	$n = 500$	98.2	97.8	98	97.3	96.4	94.6	96.6	95.8
	$n = 1.000$	100	100	100	100	99.6	99.6	99.9	99.9
	$n = 2.000$	100	100	100	100	100	100	100	100
$\alpha = 10\%$									
$\rho = -0.7$	$n = 500$	99.5	99.5	98.6	98.4	98.5	98.4	98.2	97.9
	$n = 1.000$	100	100	100	100	100	100	100	100
	$n = 2.000$	100	100	100	100	100	100	100	100
$\rho = -0.3$	$n = 500$	43.3	41.5	49	45	56.7	46.6	39.7	37.4
	$n = 1.000$	67.2	66.6	71.5	69.9	79.7	79.7	60.7	59.7
	$n = 2.000$	85.4	85.2	85.9	85.8	67.8	85.6	76.4	75.8
$\rho = 0.3$	$n = 500$	47.2	44.3	52.5	49.7	53.6	48.7	40.1	38
	$n = 1.000$	68	67.3	72	71.2	63.6	61	62	60.5
	$n = 2.000$	87.6	87.3	88.3	88.2	71.5	82	79.4	79
$\rho = 0.7$	$n = 500$	99	99	98.8	98.7	97.8	98	98.3	98
	$n = 1.000$	100	100	100	100	99.8	99.7	100	100
	$n = 2.000$	100	100	100	100	100	100	100	100

Os resultados desta seção indicaram que a heterocedasticidade pode afetar severamente a estimação de parâmetros via modelos de seleção amostral. Além disso, sob heterocedasticidade, os testes da razão de verossimilhança e gradiente, sob o ajuste de tais modelos, podem indicar a rejeição de $H_0 : \rho = 0$ mesmo quando não há evidências suficientes para rejeitar tal hipótese. Por outro lado, o modelo Heckman generalizado, reduziu o viés dos estimadores de máxima verossimilhança, mostrou-se consistente, eficiente, e apresentou resultados satisfatórios em relação ao nível de significância empírico para os testes da razão de verossimilhança e gradiente. Assim, o modelo Heckman generalizado é uma melhor alternativa quando comparado aos modelos Heckman clássico, Skew-normal e Heckman-t para dados com suspeita de viés de seleção e heterocedasticidade.

2.3.3 Cenário 3: Dispersão e Correlação Variáveis

Os Cenários 1 e 2 confirmam a importância da modelagem dos parâmetros de dispersão e correlação e sugerem que o modelo Heckman generalizado em alguns casos, pode ser uma boa alternativa aos modelos de seleção existentes. Nesta seção, nosso objetivo é avaliar o desempenho em amostras finitas dos EMV's do modelo Heckman generalizado comparado aos EMV's dos modelos Heckman clássico, Skew-normal e Heckman-t para dados com dispersão e correlação variáveis. Consideramos a mesma estrutura para Y_1^* e Y_2^* dada por:

$$Y_{1i}^* = 1.1 + 0.7x_{1i} + 0.1x_{2i} + \epsilon_{1i}, \quad (2.14)$$

$$Y_{2i}^* = -0.1 + 0.5x_{1i} + 1.1x_{2i} + 0.6x_{3i} + \epsilon_{2i}, \text{ para } i = 1, \dots, n, \quad (2.15)$$

em que os termos de erros $(\epsilon_{1i}, \epsilon_{2i})$ foram gerados a partir da distribuição normal bivariada dada por:

$$\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \underset{ind.}{\sim} \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_i^2 & \rho_i \sigma_i \\ \sigma_i \rho_i & 1 \end{pmatrix} \right], i = 1, \dots, n, \quad (2.16)$$

com $\log(\sigma_i) = 0.7 + 3x_{1i}$ e $\arcsen(\rho_i) = 0.7 + 0.7x_{1i}$. Os regressores são mantidos fixos ao longo das simulações com $x_{1i}, x_{2i} \stackrel{iid}{\sim} \mathcal{U}(0, 1)$ independente de $x_{3i} \sim \mathcal{N}(0, 1)$, para todo $i = 1, \dots, n$.

Na Tabela (2.7) apresentamos resultados numéricos referentes às estimativas de máxima verossimilhança de acordo com o Cenário 3. Observamos que conforme o tamanho amostral aumenta, o viés e a REQM diminuem, para todas as estimativas obtidas sob o ajuste do modelo Heckman generalizado, sugerindo novamente a consistência dos EMV. O mesmo resultado não foi observado sob o ajuste dos demais modelos, notamos que em alguns casos, tal como para o intercepto do modelo de regressão (β_1), a média das estimativas obtidas é altamente viesada, e a REQM não necessariamente diminui com o aumento do tamanho amostral.

Além disso, verificamos nas Figuras 2.7, 2.8 e 2.9 que as estimativas obtidas sob o ajuste do modelo Heckman generalizado, tendem ao verdadeiro valor do parâmetro e a variabilidade das mesmas diminui quando aumenta-se o tamanho amostral, indicando novamente a boa precisão das estimativas de máxima

verossimilhança. Sob o ajuste dos demais modelos observamos estimativas viesadas e imprecisas para quase todos os parâmetros. Não apresentamos o box-plot das estimativas dos parâmetros ϕ_1, ϕ_2, κ_1 e κ_2 , pois não obtemos estas estimativas sob o ajuste dos modelos de Heckman clássico, Skew-Normal e Heckman-t para comparar com as estimativas obtidas sob o ajuste do modelo Heckman generalizado.

Tabela 2.7: Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman generalizado, Heckman clássico, Skew-Normal e Heckman-t aos dados gerados conforme Cenário 3. Tamanhos amostrais $n = 500, n = 1.000$ e $n = 2.000$ com $N = 1.000$ réplicas de Monte Carlo.

Parâmetros	n	Heckman Generalizado		Heckman Clássico		Heckman Skew		Heckman t-Student	
		Média	REQM	Média	REQM	Média	REQM	Média	REQM
γ_1 -0.100	500	-0.090	0.162	-0.254	0.216	-0.496	0.458	-0.206	0.202
	1.000	-0.104	0.111	-0.252	0.186	-0.487	0.440	-0.206	0.161
	2.000	-0.099	0.076	-0.254	0.171	-0.552	0.479	-0.209	0.138
γ_2 0.500	500	0.479	0.219	0.821	0.386	0.724	0.330	0.757	0.355
	1.000	0.505	0.152	0.818	0.354	0.737	0.300	0.760	0.316
	2.000	0.497	0.101	0.844	0.359	0.745	0.281	0.781	0.307
γ_3 1.100	500	1.107	0.214	1.091	0.220	1.009	0.246	1.227	0.282
	1.000	1.106	0.143	1.079	0.148	1.005	0.185	1.215	0.204
	2.000	1.103	0.099	1.054	0.115	0.971	0.178	1.191	0.153
γ_4 0.600	500	0.607	0.064	0.584	0.066	0.542	0.099	0.668	0.104
	1.000	0.601	0.044	0.572	0.054	0.534	0.088	0.654	0.081
	2.000	0.604	0.032	0.561	0.051	0.518	0.094	0.644	0.061
β_1 1.100	500	1.122	0.247	0.346	0.816	-0.400	1.665	0.628	0.561
	1.000	1.105	0.179	0.292	0.840	-0.429	1.666	0.580	0.571
	2.000	1.110	0.117	0.348	0.769	-0.559	1.741	0.638	0.489
β_2 0.700	500	0.679	0.380	1.877	1.249	1.730	1.133	1.515	0.918
	1.000	0.697	0.254	1.893	1.228	1.788	1.139	1.522	0.874
	2.000	0.686	0.178	1.792	1.112	1.671	1.001	1.416	0.746
β_3 0.100	500	0.093	0.298	0.175	0.409	0.159	0.405	0.128	0.385
	1.000	0.099	0.214	0.213	0.314	0.202	0.310	0.169	0.289
	2.000	0.099	0.147	0.237	0.253	0.222	0.245	0.184	0.221
ϕ_1 0.700	500	0.682	0.124	2.390	1.696	2.608	1.930	2.047	1.356
	1.000	0.692	0.075	2.429	1.731	2.628	1.943	2.084	1.389
	2.000	0.693	0.052	2.409	1.710	2.638	1.948	2.053	1.356
ϕ_2 3.000	500	3.017	0.313	—	—	—	—	—	—
	1.000	3.009	0.201	—	—	—	—	—	—
	2.000	3.009	0.139	—	—	—	—	—	—
κ_1 0.700	500	0.686	0.334	1.153	0.463	1.173	0.483	1.096	0.412
	1.000	0.720	0.303	1.153	0.457	1.174	0.478	1.102	0.409
	2.000	0.688	0.114	1.143	0.445	1.168	0.470	1.085	0.388
κ_2 0.700	500	0.781	0.442	—	—	—	—	—	—
	1.000	0.711	0.344	—	—	—	—	—	—
	2.000	0.723	0.165	—	—	—	—	—	—

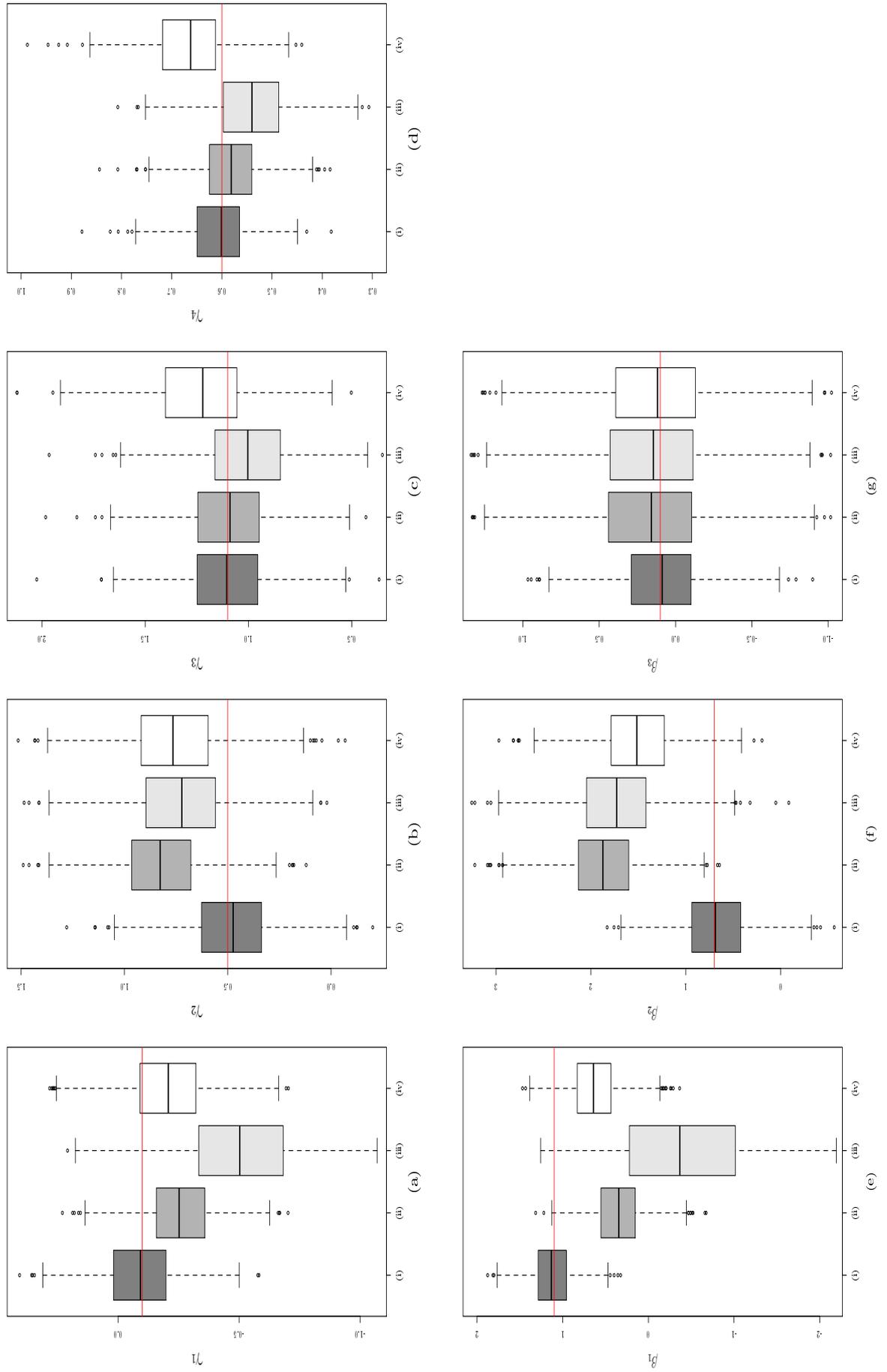


Figura 2.7: Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 e (g) β_3 , a partir do ajuste dos modelos (i) Heckman Generalizado, (ii) Skew-Normal, (iii) Heckman Clássico, (iv) Heckman-t. Tamanho amostral $n = 500$.

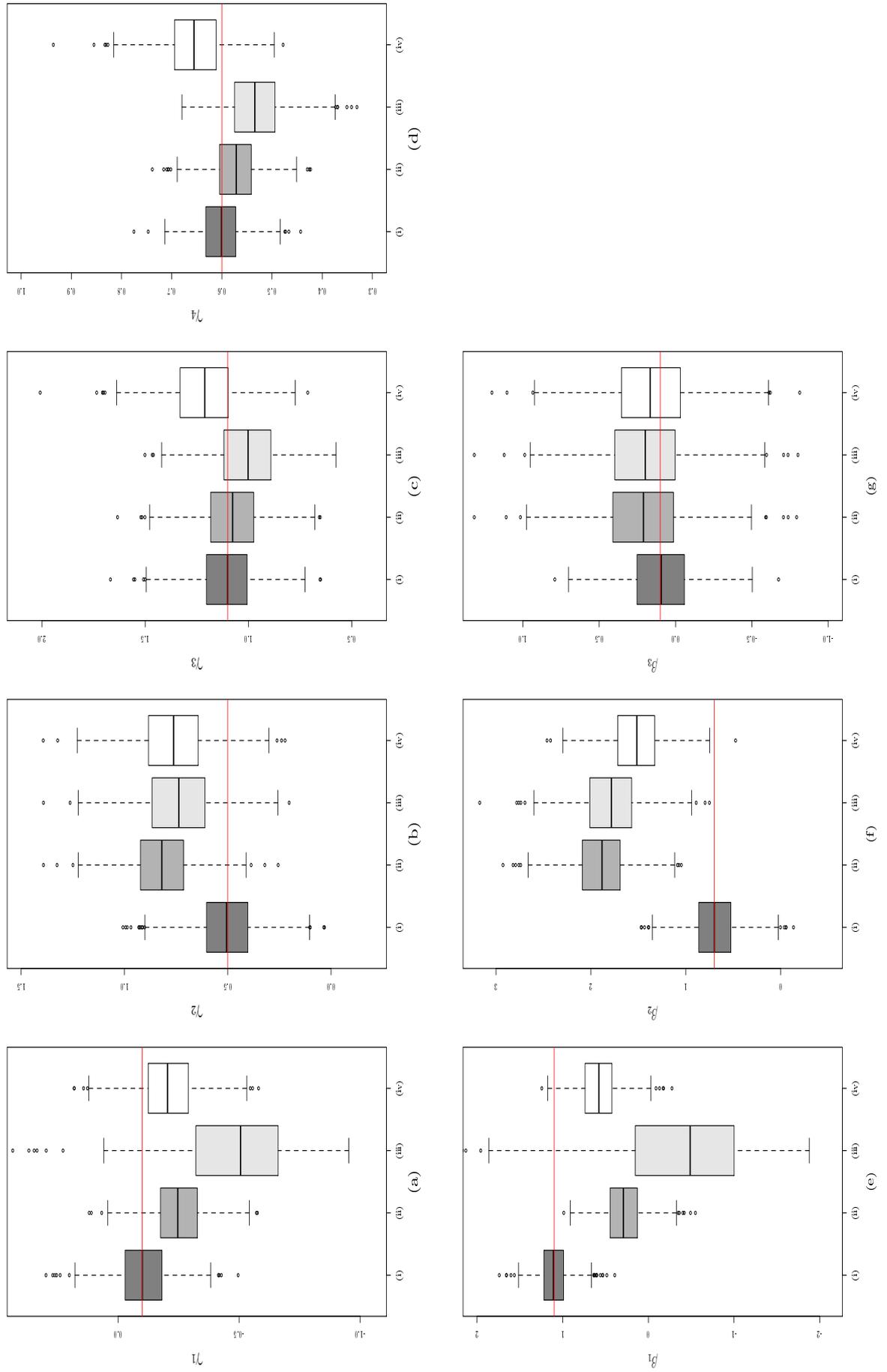


Figura 2.8: Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 e (g) β_3 , a partir do ajuste dos modelos (i) Heckman Generalizado, (ii) Skew-Normal, (iii) Heckman Clássico, (iv) Heckman-t. Tamanho amostral $n = 1.000$.

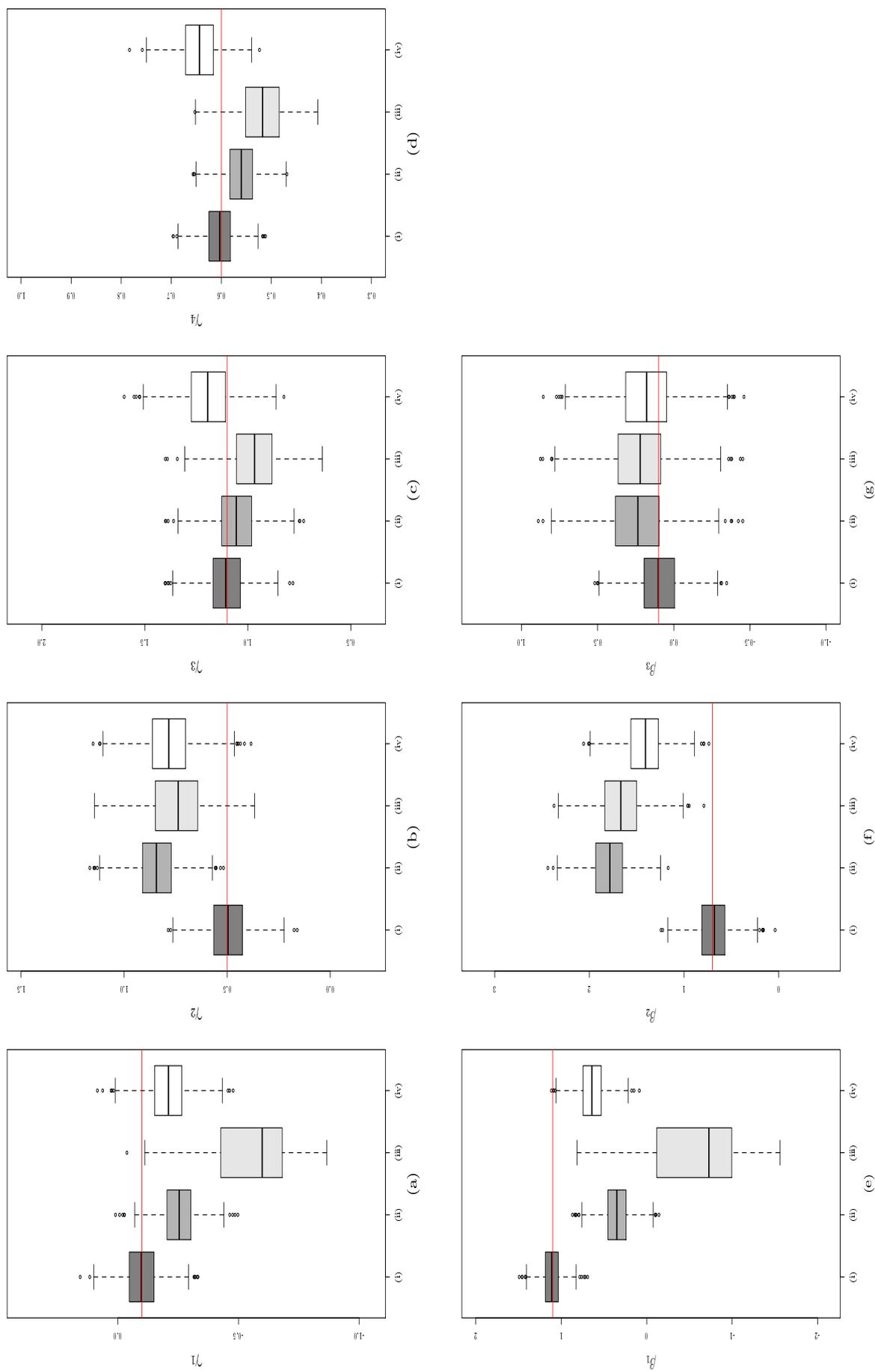


Figura 2.9: Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 e (g) β_3 , a partir do ajuste dos modelos (i) Heckman Clássico, (ii) Heckman Generalizado, (iii) Skew-Normal e (iv) Heckman-t. Tamanho amostral $n = 2.000$.

Tamanho e Poder dos Testes da Razão de Verossimilhança e Gradiente

Sob a hipótese nula ($H_0 : \rho = 0$), no caso de erros não correlacionados ($\kappa_0 = \kappa_1 = 0$), os testes da razão de verossimilhança e gradiente apresentaram valores empíricos próximos dos valores nominais somente sob o ajuste do modelo Heckman generalizado. Os testes sob o ajuste dos demais modelos inflacionam o erro tipo I e indicam a presença de viés de seleção devido a heterocedasticidade dos dados. Isso sugere, que os testes da razão de verossimilhança e gradiente, devem ser utilizados com cautela para testar parâmetros dos modelos de seleção amostral e que a heterocedasticidade pode ser a causa da presença de viés de seleção quando tais modelos são utilizados. É importante ressaltar ainda que novamente observamos variações do nível empírico sob o ajuste do modelo Heckman generalizado que contrariam nossa expectativa teórica, por exemplo, de acordo com o teste da razão de verossimilhança, para $n = 1.000$ e valor nominal igual a 5% o nível empírico é de 5.5% e para $n = 2.000$ o nível empírico é 5.8%. Novamente, acreditamos que essa variação é devido a variação da censura que foi de 28.65% para $n = 500$, 29.26% para $n = 1.000$ e 27.66% para $n = 2.000$ e também devido a variação da dispersão.

Tabela 2.8: Nível de significância empírico dos testes da razão de verossimilhança e gradiente para $H_0 : \rho = 0$ e censura média de 30%.

Valor nominal do teste	Tamanho da Amostra	Heckman Generalizado		Heckman Clássico		Heckman Skew		Heckman t-Student	
		S_{RV}	S_G	S_{RV}	S_G	S_{RV}	S_G	S_{RV}	S_G
1%	$n = 500$	1.5	0.7	5.8	0.9	10.8	0.7	2.2	0.9
	$n = 1.000$	1.7	0.8	4.7	2.4	7.3	4.8	2.5	1.9
	$n = 2.000$	1.2	0.9	5.4	4.2	3.1	2	2.5	2.3
5%	$n = 500$	6.1	4.5	12.9	8.5	21.1	8.9	8.3	6.4
	$n = 1.000$	5.5	4.2	13.6	11.2	16.1	14.5	8.2	7.5
	$n = 2.000$	5.8	4.9	14.5	12.9	8.7	8.8	8.4	8.1
10%	$n = 500$	12.6	9.2	20.9	16.7	30.7	17.5	14.5	12.7
	$n = 1.000$	10.4	8.8	21.1	19.8	22.4	21.1	14.4	14
	$n = 2.000$	11.7	11.2	20.5	20.4	12.9	15.5	14.8	14.5

Em relação ao poder dos testes da razão de verossimilhança e gradiente, sob o ajuste dos 4 modelos, observamos na Tabela 2.9, comportamentos similares ao Cenário 2, com ligeira vantagem para o teste da razão de verossimilhança. Novamente observamos testes mais poderosos sob o ajuste dos modelos Heckman clássico, Skew-Normal e Heckman-t. Porém, como observamos que os testes inflacionam o nível empírico sob o ajuste desses últimos modelos, conforme Tabela 2.8, acreditamos que o poder observado sob o ajuste do modelo Heckman generalizado produz resultados mais confiáveis.

Tabela 2.9: Tamanho empírico e poder dos testes da razão de verossimilhança (S_{RV}) e gradiente (S_G), em porcentagem, para dados simulados de acordo com o Cenário 3 e sob o ajuste dos modelos Heckman generalizado, Heckman Clássico, Heckman Skew e Heckman-t, com nível nominal de 1%, 5% e 10%

Parâmetros	Tamanho da Amostra	Heckman Generalizado		Heckman Clássico		Heckman Skew		Heckman t-Student	
		S_{RV}	S_G	S_{RV}	S_G	S_{RV}	S_G	S_{RV}	S_G
$\alpha = 1\%$									
$\kappa_1 = -0.7$ e $\kappa_2 = -0.7$	$n = 500$	100	99.9	100	99.9	100	96.6	100	100
	$n = 1.000$	100	100	100	100	100	100	100	100
	$n = 2.000$	100	100	100	100	100	100	100	100
$\kappa_1 = -0.3$ e $\kappa_2 = -0.3$	$n = 500$	31.5	21.4	51.8	35.8	56.7	34.3	35.9	27
	$n = 1.000$	60.4	54.5	83.5	79.6	89.3	87.3	68.2	63.8
	$n = 2.000$	89.7	87.9	95.8	95.7	88.4	91.1	88.3	87.2
$\kappa_1 = 0.3$ e $\kappa_2 = 0.3$	$n = 500$	30.2	19.4	53.6	37.9	50.3	27.8	37	26.8
	$n = 1.000$	63.4	58	81.9	78.4	75.2	66.5	67.8	63.1
	$n = 2.000$	89.4	88.9	97.3	96.3	90.9	92.1	89.7	88.3
$\kappa_1 = 0.7$ e $\kappa_2 = 0.7$	$n = 500$	99.9	99.5	100	99.7	99.8	96	100	99.9
	$n = 1.000$	100	100	100	100	100	100	100	100
	$n = 2.000$	100	100	100	100	100	100	100	100
$\alpha = 5\%$									
$\kappa_1 = -0.7$ e $\kappa_2 = -0.7$	$n = 500$	100	100	100	100	100	99.7	100	100
	$n = 1.000$	100	100	100	100	100	100	100	100
	$n = 2.000$	100	100	100	100	100	100	100	100
$\kappa_1 = -0.3$ e $\kappa_2 = -0.3$	$n = 500$	53.1	44.8	71.2	63.6	74.9	65.9	57.6	53.9
	$n = 1.000$	82.7	80.5	92.8	91.7	95.9	95.6	85.5	84.1
	$n = 2.000$	96.3	96	98.6	98.6	93.7	97.3	96.2	95.8
$\kappa_1 = 0.3$ e $\kappa_2 = 0.3$	$n = 500$	56	47.3	71.8	65.7	69.8	62.6	59.8	55.1
	$n = 1.000$	83.1	80.4	91.9	91	87.3	85.3	85.3	84.2
	$n = 2.000$	97.4	97.1	98.8	98.8	94.5	98.2	96.9	96.5
$\kappa_1 = 0.7$ e $\kappa_2 = 0.7$	$n = 500$	100	100	100	100	100	99.1	100	100
	$n = 1.000$	100	100	100	100	100	100	100	100
	$n = 2.000$	100	100	100	100	100	100	100	100
$\alpha = 10\%$									
$\kappa_1 = -0.7$ e $\kappa_2 = -0.7$	$n = 500$	100	100	100	100	100	99.8	100	100
	$n = 1.000$	100	100	100	100	100	100	100	100
	$n = 2.000$	100	100	100	100	100	100	100	100
$\kappa_1 = -0.3$ e $\kappa_2 = -0.3$	$n = 500$	63.7	59	79.1	76.7	82.2	78.4	70.1	66.6
	$n = 1.000$	89.5	88.4	95.7	95.5	97.5	97.4	90.5	89.7
	$n = 2.000$	98.1	98	99.6	99.6	95.6	99.1	98.2	98
$\kappa_1 = 0.3$ e $\kappa_2 = 0.3$	$n = 500$	66.9	61.5	79.9	77.5	78.6	76.3	69.8	67.7
	$n = 1.000$	90.3	89.1	95.6	95.2	91.3	90.2	90.7	90.1
	$n = 2.000$	98.8	98.7	99.3	99.3	96.1	98.9	98.3	98.3
$\kappa_1 = 0.7$ e $\kappa_2 = 0.7$	$n = 500$	100	100	100	100	100	99.8	100	100
	$n = 1.000$	100	100	100	100	100	100	100	100
	$n = 2.000$	100	100	100	100	100	100	100	100

O modelo Heckman generalizado mostrou-se uma boa alternativa aos modelos de seleção existentes quando há presença de heterocedasticidade e também correlação variável. Nosso modelo foi eficiente e consistente. Além disso, os testes da razão de verossimilhança e gradiente utilizados para avaliar $H_0 : \rho = 0$ contra $H_1 : \rho \neq 0$, revelaram maior precisão no cálculo do nível de significância empírico quando realizados sob o ajuste do modelo Heckman generalizado. Nesse sentido, com os resultados obtidos a partir das simulações dos Cenários 1, 2 e 3, esperamos reforçar a utilidade do modelo Heckman generalizado, seja em substituição aos demais modelos ou para uso conjunto, de forma a confirmar resultados relacionados a presença ou ausência de viés de seleção.

2.3.4 Cenário 4: Dispersão e Correlação Variáveis Sem Restrição de Exclusão

Nessa seção objetivamos mostrar a importância da restrição de exclusão. As simulações dos Cenários anteriores indicam que a heterocedasticidade tem papel relevante no estudo da presença ou ausência do viés de seleção, em todos os casos o modelo de Heckman generalizado mostrou-se consistente e eficiente na presença de heterocedasticidade e restrição de exclusão. Porém, é necessário avaliar se na ausência de restrição de exclusão iremos observar resultados semelhantes. Assim, consideramos o seguinte sistema de equações

$$Y_{1i}^* = 1.1 + 0.7x_{1i} + 0.1x_{2i} + \epsilon_{1i}, \quad (2.17)$$

$$Y_{2i}^* = -0.2 + 0.5x_{1i} + 1.1x_{2i} + \epsilon_{2i}, \text{ para } i = 1, \dots, n, \quad (2.18)$$

em que, os termos de erros $(\epsilon_{1i}, \epsilon_{2i})$ são gerados a partir da distribuição normal bivariada tal como em (2.7) com

$$\log(\sigma_i) = 0.7 + 3x_{1i} \text{ e } \arcsen(\rho_i) = 0.7 + 0.7x_{1i}, \text{ para } i = 1, \dots, n, \quad (2.19)$$

x_{1i} e x_{2i} são gerados independentemente com distribuição uniforme padrão e mantidos fixos ao longo das simulações. Nesse caso, a matriz de covariáveis para a equação de seleção é a mesma da equação de interesse primário. Além disso, a covariável x_{1i} também é utilizada para ajustar σ_i e ρ_i . Na Tabela 2.10, apresentamos resultados numéricos referentes às estimativas de máxima verossimilhança de acordo com o Cenário 4. Os modelos Heckman clássico, Skew-Normal e Heckman-t apresentaram resultados viesados e inconsistentes na estimação de todos os parâmetros, enquanto o modelo Heckman generalizado apresentou resultados viesados e inconsistentes na estimação dos parâmetros de correlação, β_1 e β_3 . Ressaltamos, dessa forma, a necessidade e importância da restrição de exclusão para o bom ajuste dos modelos de seleção amostral.

Tabela 2.10: Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman generalizado, Heckman clássico, Skew-Normal e Heckman-t aos dados gerados conforme Cenário 4. Tamanhos amostrais $n = 500, n = 1.000$ e $n = 2.000$ com $N = 1.000$ réplicas de Monte Carlo.

Parâmetros	n	Heckman Generalizado		Heckman Clássico		Heckman Skew		Heckman t-Student	
		Média	REQM	Média	REQM	Média	REQM	Média	REQM
		γ_1	-0.200						
	500	-0.197	0.160	-0.360	0.251	-0.166	0.405	-0.411	0.322
	1.000	-0.198	0.113	-0.311	0.177	-0.062	0.392	-0.395	0.265
	2.000	-0.200	0.078	-0.292	0.133	-0.089	0.333	-0.361	0.180
γ_2	0.500								
	500	0.494	0.209	0.767	0.397	0.818	0.409	0.922	0.515
	1.000	0.502	0.148	0.702	0.301	0.796	0.355	0.902	0.459
	2.000	0.505	0.102	0.672	0.232	0.764	0.315	0.880	0.401
γ_3	1.100								
	500	1.112	0.212	1.119	0.224	0.956	0.296	1.203	0.277
	1.000	1.105	0.145	1.121	0.156	0.921	0.289	1.187	0.210
	2.000	1.099	0.098	1.119	0.108	0.955	0.254	1.166	0.134
β_1	1.100								
	500	1.334	0.641	2.251	2.704	1.471	1.525	0.790	1.815
	1.000	1.228	0.472	2.309	1.761	1.577	1.353	0.626	1.321
	2.000	1.142	0.296	2.595	1.714	1.961	1.424	0.521	0.667
β_2	0.700								
	500	0.786	0.668	1.545	1.175	1.021	1.036	1.459	1.156
	1.000	0.724	0.445	1.627	1.009	0.842	0.938	1.548	0.995
	2.000	0.698	0.289	1.543	0.875	1.015	0.938	1.528	0.857
β_3	0.100								
	500	-0.073	0.529	-1.038	2.022	-1.137	1.623	0.017	1.515
	1.000	0.007	0.391	-1.121	1.542	-1.414	1.731	0.105	1.077
	2.000	0.075	0.238	-1.290	1.515	-1.391	1.657	0.220	0.326
ϕ_1	0.700								
	500	0.711	0.195	2.457	2.257	3.149	2.710	2.170	1.545
	1.000	0.703	0.138	2.161	1.522	3.392	2.975	2.162	1.502
	2.000	0.700	0.102	2.088	1.395	3.082	2.688	2.094	1.398
ϕ_2	3.000								
	500	2.866	0.494	—	—	—	—	—	—
	1.000	2.953	0.355	—	—	—	—	—	—
	2.000	2.996	0.238	—	—	—	—	—	—
κ_1	0.700								
	500	-0.426	2.348	-0.036	1.160	-0.329	1.429	1.010	0.647
	1.000	-0.045	1.816	-0.185	1.144	-0.585	1.542	1.098	0.555
	2.000	-0.674	2.507	-0.377	1.199	-0.565	1.476	1.137	0.483
κ_2	0.700								
	500	0.155	1.378	—	—	—	—	—	—
	1.000	0.564	0.883	—	—	—	—	—	—
	2.000	0.215	0.959	—	—	—	—	—	—

2.3.5 Cenário 5: Dispersão e Correlação Variáveis com Alto Valor de Censura

Nosso objetivo neste Cenário é avaliar o desempenho dos EMV's do modelo Heckman generalizado na presença da restrição de exclusão, com valores alto de censura e heterocedasticidade. Assim, consideramos o sistema de equações

$$Y_{1i}^* = 1.1 + 0.7x_{1i} + 0.1x_{2i} + \epsilon_{1i}, \quad (2.20)$$

$$Y_{2i}^* = -0.8 + 0.5x_{1i} + 1.1x_{2i} + 0.6x_{3i} + \epsilon_{2i}, \text{ para } i = 1, \dots, n, \quad (2.21)$$

e os termos de erros $(\epsilon_{1i}, \epsilon_{2i})$ gerados a partir da distribuição normal bivariada dada por:

$$\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \stackrel{ind.}{\sim} \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_i^2 & \rho_i \sigma_i \\ \sigma_i \rho_i & 1 \end{pmatrix} \right], \text{ para } i = 1, \dots, n, \quad (2.22)$$

em que $\log(\sigma_i) = 0.7 + 3x_{1i}$ e $\arcsen(\rho_i) = \kappa_0 + \kappa_1 x_{1i}$. Assim, simulamos dados com dispersão e correlação variáveis. Os regressores são mantidos fixos ao longo das simulações com $x_{1i}, x_{2i} \stackrel{iid}{\sim} \mathcal{U}(0, 1)$ independentes de $x_{3i} \sim \mathcal{N}(0, 1)$, para todo $i = 1, \dots, n$. A censura observada nas simulações deste cenário foram 50.6%, 51% e 49.6% para os tamanhos amostrais $n = 500, n = 1.000$ e $n = 2.000$, respectivamente. Apresentamos resultados para $\kappa_0 = \kappa_1 = -0.3$ e $\kappa_0 = \kappa_1 = -0.7$. Dessa forma temos, respectivamente, correlação moderada variando no intervalo $(-0.6, -0.29)$ e correlação forte variando no intervalo $(-0.99, -0.64)$.

Quando $\kappa_0 = \kappa_1 = -0.3$, observe na Tabela 2.11 que sob o ajuste do modelo Heckman generalizado a REQM e o viés diminuem para todos os parâmetros ao aumentar o tamanho amostral. Não percebemos, porém, os mesmos resultados para o ajuste dos demais modelos, nesses casos, observamos estimativas viesadas e inconsistentes. Tal resultado indica que mesmo sendo menos parcimonioso que os demais, nosso modelo é indicado para avaliar dados com heterocedasticidade, alto valor de censura e correlação moderada ou fraca. Por outro lado, na Tabela 2.12, quando $\kappa_0 = \kappa_1 = -0.7$, observamos problemas na estimação dos parâmetros sob o ajuste de todos os modelos. Apesar do modelo Heckman generalizado estimar bem os parâmetros das equações de seleção, regressão e da dispersão, os EMV's dos parâmetros de correlação se mostraram inconsistentes. Tal resultado indica que o modelo Heckman generalizado deve ser utilizado com cuidado na presença de heterocedasticidade e alto valor de censura, apesar de apresentar resultados melhores que os demais modelos de seleção.

Tabela 2.11: Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman generalizado, Heckman clássico, Skew-Normal e Heckman-t aos dados gerados conforme Cenário 5. Tamanhos amostrais $n = 500, n = 1.000$ e $n = 2.000$ com $N = 1.000$ réplicas de Monte Carlo.

Parâmetros	n	Heckman Generalizado		Heckman Clássico		Heckman Skew		Heckman t-Student	
		Média	REQM	Média	REQM	Média	REQM	Média	REQM
γ_1 -0.800	500	-0.813	0.165	-0.835	0.168	-0.840	0.214	-0.838	0.181
	1.000	-0.807	0.118	-0.827	0.121	-0.433	0.448	-0.829	0.130
	2.000	-0.804	0.080	-0.824	0.083	-0.818	0.138	-0.828	0.091
γ_2 0.500	500	0.514	0.219	0.566	0.235	0.559	0.233	0.502	0.246
	1.000	0.502	0.152	0.551	0.166	0.551	0.150	0.487	0.172
	2.000	0.500	0.102	0.554	0.120	0.553	0.119	0.486	0.116
γ_3 1.100	500	1.107	0.214	1.101	0.213	1.090	0.216	1.178	0.246
	1.000	1.111	0.148	1.105	0.148	1.000	0.190	1.182	0.181
	2.000	1.104	0.105	1.094	0.106	1.087	0.109	1.176	0.139
γ_4 0.600	500	0.608	0.067	0.597	0.071	0.592	0.075	0.646	0.091
	1.000	0.602	0.047	0.592	0.051	0.532	0.099	0.641	0.069
	2.000	0.604	0.035	0.592	0.038	0.588	0.042	0.645	0.061
β_1 1.100	500	1.098	0.430	2.002	1.109	2.113	1.294	1.643	0.767
	1.000	1.093	0.305	2.016	1.050	0.617	0.983	1.633	0.684
	2.000	1.103	0.194	1.982	0.956	1.985	1.090	1.554	0.535
β_2 0.700	500	0.732	0.773	-0.572	1.379	-0.545	1.381	-0.280	1.118
	1.000	0.708	0.516	-0.585	1.342	-1.034	1.811	-0.271	1.042
	2.000	0.675	0.349	-0.559	1.285	-0.583	1.316	-0.242	0.977
β_3 0.100	500	0.082	0.397	-0.008	0.577	-0.003	0.581	0.067	0.521
	1.000	0.102	0.296	-0.025	0.450	-0.087	0.476	0.059	0.404
	2.000	0.107	0.193	-0.009	0.325	-0.015	0.324	0.095	0.273
ϕ_1 0.700	500	0.693	0.154	2.437	1.748	2.550	1.888	2.055	1.371
	1.000	0.695	0.092	2.468	1.773	3.336	2.718	2.073	1.382
	2.000	0.698	0.058	2.449	1.753	2.517	1.835	2.020	1.324
ϕ_2 3.000	500	3.032	0.407	—	—	—	—	—	—
	1.000	3.014	0.270	—	—	—	—	—	—
	2.000	3.002	0.182	—	—	—	—	—	—
κ_1 -0.300	500	-0.280	0.347	-0.548	0.325	-0.558	0.341	-0.475	0.261
	1.000	-0.291	0.218	-0.539	0.289	-0.712	0.464	-0.464	0.221
	2.000	-0.302	0.161	-0.541	0.270	-0.555	0.284	-0.445	0.179
κ_2 -0.300	500	-0.341	0.553	—	—	—	—	—	—
	1.000	-0.311	0.344	—	—	—	—	—	—
	2.000	-0.290	0.248	—	—	—	—	—	—

Tabela 2.12: Valor verdadeiro(VV) e média das estimativas de máxima verossimilhança com a respectiva estatística de raiz quadrada do erro quadrático médio (*REQM*) do ajuste dos modelos de Heckman generalizado, Heckman clássico, Skew-Normal e Heckman-t aos dados gerados conforme modelo 5. Tamanhos amostrais $n = 500, n = 1.000$ e $n = 2.000$ com $N = 1.000$ réplicas de Monte Carlo.

Parâmetros	n	Heckman Generalizado		Heckman Clássico		Heckman Skew		Heckman t-Student	
		Média	<i>REQM</i>	Média	<i>REQM</i>	Média	<i>REQM</i>	Média	<i>REQM</i>
γ_1 -0.800	500	-0.808	0.151	-0.878	0.169	-0.670	0.439	-0.881	0.182
	1.000	-0.802	0.111	-0.863	0.127	-0.948	0.276	-0.864	0.135
	2.000	-0.803	0.077	-0.865	0.100	-0.827	0.283	-0.870	0.108
γ_2 0.500	500	0.502	0.203	0.664	0.260	0.621	0.229	0.600	0.246
	1.000	0.495	0.140	0.639	0.203	0.610	0.191	0.574	0.178
	2.000	0.503	0.099	0.668	0.194	0.647	0.181	0.602	0.150
γ_3 1.100	500	1.108	0.192	1.092	0.205	0.993	0.261	1.175	0.236
	1.000	1.108	0.135	1.093	0.144	1.051	0.165	1.174	0.173
	2.000	1.102	0.093	1.067	0.107	1.030	0.152	1.152	0.123
γ_4 0.600	500	0.607	0.059	0.586	0.061	0.532	0.118	0.642	0.079
	1.000	0.602	0.040	0.579	0.047	0.558	0.072	0.634	0.058
	2.000	0.602	0.031	0.570	0.044	0.549	0.078	0.629	0.047
β_1 1.100	500	1.085	0.355	2.384	1.354	1.896	1.432	2.016	1.006
	1.000	1.078	0.253	2.462	1.396	2.777	1.842	2.086	1.033
	2.000	1.093	0.155	2.387	1.304	2.326	1.450	1.999	0.924
β_2 0.700	500	0.727	0.514	-1.347	2.100	-1.485	2.261	-0.960	1.730
	1.000	0.735	0.341	-1.363	2.090	-1.341	2.075	-0.967	1.702
	2.000	0.699	0.216	-1.298	2.011	-1.320	2.038	-0.901	1.618
β_3 0.100	500	0.093	0.357	0.009	0.490	-0.003	0.494	0.084	0.460
	1.000	0.104	0.257	-0.052	0.371	-0.049	0.369	0.020	0.340
	2.000	0.108	0.169	-0.030	0.274	-0.033	0.275	0.051	0.237
ϕ_1 0.700	500	0.683	0.166	2.403	1.711	2.856	2.292	2.095	1.406
	1.000	0.687	0.098	2.446	1.750	2.575	1.901	2.137	1.442
	2.000	0.693	0.064	2.421	1.723	2.572	1.917	2.094	1.398
ϕ_2 3.000	500	3.014	0.372	—	—	—	—	—	—
	1.000	3.017	0.239	—	—	—	—	—	—
	2.000	3.006	0.159	—	—	—	—	—	—
κ_1 -0.700	500	-0.888	0.675	-1.130	0.440	-1.180	0.493	-1.078	0.391
	1.000	-1.364	1.107	-1.138	0.443	-1.154	0.459	-1.087	0.393
	2.000	-0.713	0.210	-1.135	0.438	-1.152	0.457	-1.076	0.380
κ_2 -0.700	500	-0.593	0.636	—	—	—	—	—	—
	1.000	-0.170	0.940	—	—	—	—	—	—
	2.000	-0.699	0.209	—	—	—	—	—	—

2.4 Aplicação a Dados Reais

Nesta seção, apresentamos uma aplicação do modelo proposto a um conjunto de dados reais. Considere os dados de despesas ambulatoriais do *Medical Expenditure Panel Survey (MEPS)* de 2001 disponíveis no *software* R via pacote *ssmrob* (Zhelonkin et al., 2014). Esses dados também foram utilizados por Cameron e Trivedi (2009), por Marchenko e Genton (2012) e por Zhelonkin et al. (2016) para ajuste dos modelos de Heckman clássico, Heckman-t e a versão robusta do método de dois passos, respectivamente. O MEPS é um conjunto de pesquisas em grande escala de famílias, indivíduos e seus provedores médicos (médicos, hospitais, farmácias, etc.) dos Estados Unidos. Possui dados sobre os serviços de saúde que os americanos usam, a frequência com que usam, o custo desses serviços e como eles são pagos, bem como dados sobre o custo e alcance do seguro de saúde disponível aos trabalhadores americanos.

A amostra é restrita a pessoas com idades entre 21 e 64 anos e contém uma variável resposta com 3328 observações dos valores de gastos ambulatoriais, dos quais 526 (15,8%) correspondem aos valores de despesas não observadas e identificadas como despesa zero para o ajuste dos modelos. Inclui também as seguintes variáveis explicativas:

- x_1 , que representa a idade medida em dezenas de anos;
- x_2 , que é uma variável indicadora para sexo (feminino recebe valor 1);
- x_3 , representa anos de escolaridade;
- x_4 , indicadora para etnia (negro ou hispânico recebem valor 1);
- x_5 , representa o número de doenças crônicas;
- x_6 , é uma variável indicadora para presença de seguro.

Utilizamos a variável de interesse Y_1^* , que representa a despesa com serviços médicos, na escala logarítmica, uma vez que ela é altamente assimétrica. A variável Y_2^* , que representa a disposição para gastar, não é observada. Observamos $U = 1\{Y_2^* > 0\}$, a qual representa a decisão de gastar com cuidados médicos.

De acordo com Marchenko e Genton (2012) e Zhelonkin et al. (2016) é natural o ajuste de um modelo de seleção amostral a tais dados, pois a disposição para gastar (Y_2^*) provavelmente está relacionada ao montante de despesa (Y_1^*). Porém, após ajuste do modelo de Heckman clássico e uso dos testes de Wald, da razão de verossimilhanças ou do gradiente para $H_0 : \rho = 0$ contra $H_1 : \rho \neq 0$, a conclusão é que não existem evidências suficientes (p -valor > 0.1) para rejeitar H_0 , ou seja, não há viés de seleção. Cameron e Trivedi (2009) suspeitaram dessa conclusão e Marchenko e Genton (2012) argumentaram que um modelo mais robusto poderia evidenciar o viés de seleção presente nos dados e rejeitaram a hipótese de normalidade dos dados. No entanto, entendemos que o ajuste da dispersão e/ou da correlação pode ser uma alternativa para estudar esses dados sem a necessidade de modificar a suposição de normalidade dos erros, pois como indicam nossas simulações, as estimativas dos parâmetros obtidas pelo ajuste do

modelo de Heckman usuais podem ser mais severamente afetadas pela heterocedasticidade do que pela distribuição incorreta dos termos de erro.

Assim, consideramos o seguinte sistema de equações:

$$Y_{1i}^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \epsilon_{1i}, \quad (2.23)$$

$$Y_{2i}^* = \gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \gamma_3 x_{3i} + \gamma_4 x_{4i} + \gamma_5 x_{5i} + \gamma_6 x_{6i} + \gamma_7 x_{7i} + \epsilon_{2i}, \quad (2.24)$$

em que os erros são definidos de acordo com (2.7), de tal forma que:

$$\begin{aligned} \log(\sigma_i) &= \phi_0 + \phi_1 x_{1i} + \phi_2 x_{2i} + \phi_3 x_{3i} + \phi_4 x_{4i} + \phi_5 x_{5i} + \phi_6 x_{6i}, \text{ e} \\ \arcsen(\rho_i) &= \kappa_0 + \kappa_1 x_{1i} + \kappa_2 x_{2i} + \kappa_3 x_{3i} + \kappa_4 x_{4i} + \kappa_5 x_{5i} + \kappa_6 x_{6i}, i = 1, \dots, 3328. \end{aligned}$$

A regressão de interesse mantém as mesmas covariáveis da equação de seleção com o acréscimo da covariável rendimento (x_7), o que introduz a restrição de exclusão. Na Tabela 2.13 apresentamos os resultados do ajuste dos modelos Heckman clássico e Heckman generalizado.

Tabela 2.13: Estimativas do Modelo de Heckman clássico e os respectivos p -valores juntamente com as estimativas do Modelo de Heckman Generalizado com os respectivos valores de desvio-padrão (DP), valor-z, p -valor e limites inferior e superior para o intervalo de confiança de 95%.

Equação de Seleção								
Parâmetros	Heckman	p -valor	HGen	DP	Valor-z	p -valor	Inferior	Superior
(Intercepto)	-0.676	0.000	-0.651	0.189	-3.448	0.001	-1.022	-0.281
Idade (x_1)	0.088	0.001	0.093	0.027	3.495	0.000	0.041	0.145
Fem (x_2)	0.663	0.000	0.635	0.060	10.574	0.000	0.517	0.753
Educ (x_3)	0.062	0.000	0.060	0.012	5.123	0.000	0.037	0.083
Blhisp (x_4)	-0.364	0.000	-0.330	0.061	-5.436	0.000	-0.449	-0.211
Totcr (x_5)	0.797	0.000	0.720	0.070	10.304	0.000	0.583	0.857
ins (x_6)	0.170	0.007	0.167	0.061	2.747	0.006	0.048	0.287
Renda (x_7)	0.003	0.040	0.002	0.001	1.905	0.057	-0.000	0.005
Equação Primária								
Parâmetros	Heckman	p -valor	HGen	DP	Valor-z	p -valor	Inferior	Superior
(Intercepto)	5.044	0.000	5.730	0.221	25.945	0.000	5.297	6.162
Idade (x_1)	0.212	0.000	0.176	0.027	6.555	0.000	0.123	0.228
Fem (x_2)	0.348	0.000	0.165	0.060	2.739	0.006	0.047	0.283
Educ (x_3)	0.019	0.076	0.008	0.013	0.670	0.503	-0.016	0.033
Blhisp (x_4)	-0.219	0.000	-0.123	0.069	-1.783	0.075	-0.257	0.012
Totcr (x_5)	0.540	0.000	0.417	0.029	14.183	0.000	0.360	0.475
ins (x_6)	-0.030	0.557	-0.101	0.056	-1.822	0.068	-0.210	0.008
Covariáveis do Parâmetro de Dispersão								
Parâmetros	Heckman	p -valor	HGen	DP	Valor-z	p -valor	Inferior	Superior
(Intercepto)	1.271	0.000	0.602	0.111	5.426	0.000	0.385	0.820
Idade (x_1)	-	-	-0.026	0.016	-1.671	0.095	-0.057	0.005
Fem (x_2)	-	-	-0.126	0.033	-3.841	0.000	-0.191	-0.062
Educ (x_3)	-	-	-0.001	0.007	-0.107	0.915	-0.014	0.012
Blhisp (x_4)	-	-	0.040	0.038	1.055	0.291	-0.035	0.115
Totcr (x_5)	-	-	-0.115	0.019	-6.001	0.000	-0.153	-0.078
ins (x_6)	-	-	-0.106	0.033	-3.240	0.001	-0.170	-0.042
Covariáveis do Parâmetro de Correlação								
Parâmetros	Heckman	p -valor	HGen	DP	Valor-z	p -valor	Inferior	Superior
(Intercepto)	-0.131	0.375	-0.161	0.362	-0.446	0.656	-0.871	0.548
Idade (x_1)	-	-	-0.026	0.059	-0.435	0.663	-0.141	0.090
Fem (x_2)	-	-	-0.091	0.101	-0.902	0.367	-0.290	0.107
Educ (x_3)	-	-	-0.040	0.023	-1.758	0.079	-0.085	0.005
Blhisp (x_4)	-	-	0.016	0.116	0.136	0.892	-0.211	0.242
Totcr (x_5)	-	-	-0.260	0.103	-2.524	0.012	-0.462	-0.058
ins (x_6)	-	-	-0.039	0.103	-0.380	0.704	-0.242	0.163

Retiramos as covariáveis correspondentes aos parâmetros não significativos no ajuste de σ_i e ρ_i e ajustamos novamente os modelos Heckman clássico e Heckman generalizado aos dados. A fim de manter as mesmas configurações utilizadas por [Marchenko e Genton \(2012\)](#) e [Zhelonkin et al. \(2016\)](#) na análise destes dados, optamos por não retirar as covariáveis que modelam as equações de regressão e seleção do ajuste, mesmo tendo observado que os parâmetros, acompanhando x_3, x_4 e x_6 na equação primária e x_7 na equação de seleção, foram não significativos no primeiro ajuste. Os novos resultados obtidos são apresentados na Tabela 2.14.

Tabela 2.14: Estimativas do Modelo de Heckman clássico e os respectivos p-valores juntamente com as estimativas do Modelo Heckman Generalizado com os respectivos valores de desvio-padrão (DP), valor-z, p-valor e limites inferior e superior para o intervalo de confiança de 95%.

Equação de Seleção								
Parâmetros	Heckman	<i>p</i> -valor	HGen	DP	Valor-z	<i>p</i> -valor	Inferior	Superior
(Intercepto)	-0.676	0.000	-0.598	0.184	-3.246	0.001	-0.960	-0.237
Idade (x_1)	0.088	0.001	0.095	0.026	3.677	0.000	0.045	0.146
Fem (x_2)	0.663	0.000	0.637	0.059	10.786	0.000	0.521	0.752
Educ (x_3)	0.062	0.000	0.056	0.011	4.901	0.000	0.033	0.078
Blhisp (x_4)	-0.364	0.000	-0.334	0.059	-5.671	0.000	-0.450	-0.219
Totcr (x_5)	0.797	0.000	0.714	0.068	10.497	0.000	0.581	0.847
ins (x_6)	0.170	0.007	0.163	0.060	2.704	0.007	0.045	0.282
Renda (x_7)	0.003	0.040	0.002	0.001	1.866	0.062	-0.000	0.005
Equação Primária								
Parâmetros	Heckman	<i>p</i> -valor	HGen	DP	Valor-z	<i>p</i> -valor	Inferior	Superior
(Intercepto)	5.044	0.000	5.848	0.193	30.377	0.000	5.471	6.226
Idade (x_1)	0.212	0.000	0.174	0.024	7.376	0.000	0.128	0.220
Fem (x_2)	0.348	0.000	0.152	0.055	2.756	0.006	0.044	0.259
Educ (x_3)	0.019	0.076	0.001	0.010	0.052	0.959	-0.019	0.020
Blhisp (x_4)	-0.219	0.000	-0.135	0.058	-2.338	0.019	-0.248	-0.022
Totcr (x_5)	0.540	0.000	0.422	0.030	14.191	0.000	0.363	0.480
ins (x_6)	-0.030	0.557	-0.108	0.051	-2.104	0.035	-0.209	-0.007
Covariáveis do Parâmetro de Dispersão								
Parâmetros	Heckman	<i>p</i> -valor	HGen	DP	Valor-z	<i>p</i> -valor	Inferior	Superior
(Intercepto)	1.271	0.000	0.606	0.063	9.596	0.000	0.483	0.730
Idade (x_1)	—	—	-0.028	0.013	-2.231	0.026	-0.053	-0.003
Fem (x_2)	—	—	-0.130	0.028	-4.656	0.000	-0.185	-0.076
Totcr (x_5)	—	—	-0.109	0.019	-5.650	0.000	-0.146	-0.071
ins (x_6)	—	—	-0.115	0.028	-4.132	0.000	-0.169	-0.060
Covariáveis do Parâmetro de Correlação								
Parâmetros	Heckman	<i>p</i> -valor	HGen	DP	Valor-z	<i>p</i> -valor	Inferior	Superior
(Intercepto)	-0.131	0.375	-0.789	0.075	-10.525	0.000	-0.936	-0.642
Totcr (x_5)	—	—	-0.268	0.086	-3.095	0.002	-0.437	-0.098

O primeiro resultado importante é que as estimativas obtidas com o modelo Heckman generalizado são similares as estimativas do modelo de Heckman clássico. Por outro lado, diferentemente do modelo de Heckman, o modelo Heckman generalizado indica a presença de viés de seleção. Ou seja, ao aplicar o teste de Wald para avaliar $H_0 : \rho = 0$ contra $H_1 : \rho \neq 0$, observamos evidências suficientes (p -valor = 0.000), para rejeitar H_0 ao nível de 5% de significância. Ressaltamos que procedemos também os testes da razão de verossimilhança e do gradiente onde foi observado o mesmo resultado.

Note também que os parâmetros associados as covariáveis *Idade* (x_1), *Fem* (x_2), *Totcr* (x_5) e *ins* (x_6), que modelam a dispersão, são significativos. Tal conclusão indica a presença de heterocedasticidade nos dados e pode justificar os resultados obtidos, pois como observamos em nossas simulações, a presença de heterocedasticidade produz estimativas viesadas e estimadores inconsistentes quando ajustamos o modelo de Heckman clássico. Ressaltamos que o resultado obtido com o ajuste do nosso modelo é similar aos de [Marchenko e Genton \(2012\)](#) e [Zhelonkin et al. \(2016\)](#) que também obtiveram evidências suficientes para rejeitar a ausência de viés de seleção.

2.5 Conclusões

Generalizamos o modelo clássico de Heckman ao considerar covariáveis para os parâmetros de dispersão e correlação. A principal vantagem desta generalização é permitir identificar as fontes de variabilidade dos dados e as covariáveis responsáveis pela presença de viés de seleção. Analisamos o desempenho do modelo proposto comparado aos modelos Heckman clássico, Skew-Normal e Heckman-t via estudo de simulação Monte Carlo. Os resultados apresentados nos Cenários 1, 2 e 3, indicam que para dados com dispersão e/ou correlação variáveis e a imposição da restrição de exclusão, o modelo Heckman generalizado apresenta melhores resultados de estimação, uma vez que suas estimativas são menos viesadas. A REQM converge rapidamente a zero e a variabilidade das estimativas tendem a diminuir com o aumento do tamanho amostral em todos os casos. Observamos também que, na presença de heterocedasticidade, o modelo Heckman generalizado é o único a apresentar valores do nível de significância empírico próximo dos valores nominais para os testes da razão de verossimilhanças e gradiente usados para avaliar $H_0 : \rho = 0$ contra $H_1 : \rho \neq 0$.

Mostramos resultados, conforme Cenários 4 e 5, do ajuste dos modelos para dados simulados sem considerar restrição de exclusão e com alta taxa de censura, respectivamente. Nesses casos, sob ajuste do modelo Heckman generalizado observamos problemas na estimação dos parâmetros de correlação indicando que, assim como, os demais, nosso modelo apresenta problemas de multicolinearidade e problemas na estimação dos parâmetros na presença de altos valores de censura.

Concluimos este capítulo com uma aplicação ao banco de dados *Medical Expenditure Panel Survey (MEPS)* de 2001, obtido no *software* R, para ilustrar a utilidade de nosso modelo e, mais especificamente, a utilidade de considerar um modelo com covariáveis para ajustar os parâmetros de dispersão e correlação.

CAPÍTULO 3

MODELO DE SELEÇÃO AMOSTRAL HECKMAN

BIRNBAUM-SAUNDERS

Resumo do Capítulo

O principal pressuposto para o modelo de Heckman é a suposição de normalidade. Na prática, essa suposição pode não ser realista em muitas situações. Dados como renda e despesa, por exemplo, violam a suposição de normalidade devido a assimetria ou a caudas pesadas. Nesses casos, muitas vezes são utilizadas transformações da variável resposta para diminuir o efeito da assimetria no ajuste do modelo de Heckman. Essa solução tem algumas desvantagens, sendo a mais relevante, a interferência na interpretação dos parâmetros do modelo resultante. Neste capítulo, introduzimos um modelo de seleção amostral baseado na distribuição Birnbaum-Saunders bivariada. Nosso modelo é parcimonioso, possui a mesma quantidade de parâmetros do modelo clássico de Heckman e ajusta-se melhor que os modelos de seleção existentes na literatura a dados assimétricos e positivos, além de manter, nesses casos, a escala original da variável resposta, o que facilita a interpretação dos parâmetros. Estudamos suas propriedades e a performance dos estimadores de máxima verossimilhança em amostras finitas comparada a performance dos estimadores dos modelos de Heckman clássico, Skew-normal e Heckman-t. Os resultados indicam bom ajuste do modelo Heckman-BS e reforçam sua utilidade. Além disso, estudamos e comparamos o poder dos testes da razão de verossimilhança e gradiente sob o ajuste de todos os modelos citados. Apresentamos uma aplicação aos dados de despesas ambulatoriais da pesquisa “Medical Expenditure Panel Survey (MEPS)” de 2001 e comparamos os resultados obtidos com os resultados do modelo clássico de Heckman.

Palavras-chave: Distribuição bivariada, normalidade, assimetria.

3.1 Introdução

A maior crítica ao modelo clássico de Heckman é sua sensibilidade à suposição de normalidade conjunta dos erros. Uma vez que, não muito raro, os dados apresentam observações extremas ou assimetria, o que interfere na estimação dos parâmetros. Para dados com observações extremas, [Marchenko e Genton \(2012\)](#) propuseram, em substituição a distribuição normal no modelo de Heckman, usar a *t* de student bivariada determinando o modelo Heckman-*t*. Para dados com desvios da normalidade devido a assimetria, [Ogundimu e Hutton \(2016\)](#) sugeriram a distribuição Skew-normal bivariada para a distribuição dos erros e introduziram o modelo de seleção amostral Skew-normal.

O problema é que os modelos, Heckman-*t* e Skew-normal, necessitam, respectivamente, da estimação dos parâmetros de grau de liberdade e de assimetria, além dos parâmetros das equações de seleção e regressão e dos parâmetros de dispersão e correlação. Ou seja, tais modelos, apesar de flexíveis, não são parcimoniosos. Além disso, é conhecido na literatura a dificuldade de se estimar o parâmetro de grau de liberdade ([Fernandez e Steel, 1999](#)) e, estimações incorretas deste parâmetro podem levar a estimativas dos demais parâmetros altamente viesadas e conclusões erradas. Adicionalmente, tal como o modelo de Heckman clássico, para o ajuste dos modelos Heckman-*t* e Skew-normal a dados assimétricos e positivos, é necessária a transformação da variável de interesse, de forma, a reduzir a assimetria. Em geral, transformações da variável de interesse nem sempre são apropriadas e podem complicar a interpretação dos parâmetros no modelo resultante, principalmente, quando há problemas de viés de seleção.

O objetivo principal deste capítulo é introduzir um modelo de seleção amostral parcimonioso, flexível, sem maiores problemas de estimação e que tenha fácil interpretação dos seus parâmetros. Assim, propomos mais uma abordagem do modelo de seleção de Heckman ao considerar a distribuição bivariada Birnbaum-Saunders para as variáveis de interesse primária e de seleção. Denominamos esse novo modelo por Heckman-BS. Além de ser uma alternativa para o ajuste de dados assimétricos positivos com problemas de seleção de amostra, nosso modelo tem a vantagem de possuir a mesma quantidade de parâmetros do modelo de Heckman clássico e de não ser necessária a transformação da variável de interesse para seu ajuste. Além disso, é de fácil implementação e envolve o mesmo custo computacional que a abordagem do modelo clássico.

Este capítulo está organizado da seguinte forma: a próxima seção apresenta a formulação do modelo Heckman-BS, explicitando a função de log-verossimilhança e seu vetor escore. Na Seção 3.3, apresentamos estudos de simulação em que comparamos o ajuste do modelo Heckman-BS com o ajuste dos modelos de Heckman clássico, Heckman-*t* e Skew-normal quando os dados são simulados de acordo com o modelo BS. O tamanho e poder dos testes da razão de verossimilhança e gradiente são apresentados na Seção 3.4. Na Seção 3.5 avaliamos o ajuste dos modelos a dados simulados sob falta de especificação utilizando a distribuição gama bivariada. Na Seção 3.6 apresentamos uma aplicação aos dados de despesas ambulatoriais da pesquisa “Medical Expenditure Panel Survey (MEPS)” de 2001 também analisados por [Cameron e Trivedi \(2009\)](#) e [Marchenko e Genton \(2012\)](#), em que comparamos os resultados obtidos com os resultados do modelo clássico de Heckman. As conclusões do capítulo são apresentadas na última seção.

3.2 Formulação do Modelo Heckman-BS

Considere uma amostra aleatória de vetores bivariados $(Y_{11}^*, Y_{21}^*)^\top, \dots, (Y_{1n}^*, Y_{2n}^*)^\top$ com distribuição BS bivariada de parâmetros $\boldsymbol{\mu}_i = (\mu_{1i}, \mu_{2i})^\top \in \mathbb{R}^+ \times \mathbb{R}^+, i = 1, 2, \dots, n$, $\boldsymbol{\phi} = (\phi_1, \phi_2)^\top \in \mathbb{R}^+ \times \mathbb{R}^+$ e $\rho \in (-1, 1)$, isto é, $(Y_{1i}^*, Y_{2i}^*)^\top \stackrel{ind.}{\sim} BS_2(\boldsymbol{\mu}_i, \boldsymbol{\phi}, \rho)$. Suponha que Y_{1i}^* é observado se, e somente se, para $c > 0$, fixo, $Y_{2i}^* > c$. Assim, considere

$$\begin{aligned} U_i &= 1\{Y_{2i}^* > c\}, \\ Y_i &= Y_{1i}^* U_i, \text{ para } i = 1, \dots, n, \end{aligned} \quad (3.1)$$

em que a variável Y_i é a parte observada de Y_{1i}^* e $U_i = 1$, se $Y_{2i}^* > c$ ou $U_i = 0$, se $Y_{2i}^* \leq c$. Fixamos $c = 1$, pois a distribuição Birnbaum-Saunders possui a propriedade de invariância por escala, assim, qualquer outro valor $c \neq 1$, seria absorvido pelo parâmetro de posição. Como ϕ_2 é não identificável, pois não observamos Y_{2i}^* , somente sabemos se $Y_{2i}^* \leq 1$ ou $Y_{2i}^* > 1$, para garantir a identificabilidade do modelo, fixamos $\phi_2 = 1$, no entanto, qualquer outro valor positivo poderia ser considerado. Também no modelo clássico de Heckman, a variância de y_{2i}^* é fixada ($\sigma_2^2 = 1$). Assumimos a seguinte estrutura de regressão:

$$\begin{aligned} g_1(\mu_{1i}) &= \mathbf{x}_i^\top \boldsymbol{\beta} = \eta_{1i}, \\ g_2(\mu_{2i}) &= \mathbf{w}_i^\top \boldsymbol{\gamma} = \eta_{2i}, \end{aligned} \quad (3.2)$$

em que $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ e $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)^\top$ são vetores de parâmetros desconhecidos, $\boldsymbol{\beta} \in \mathbb{R}^p$ e $\boldsymbol{\gamma} \in \mathbb{R}^q$, $p + q < n$. Além disso, $\boldsymbol{\eta}_1 = (\eta_{1i}, \dots, \eta_{1n})^\top$ e $\boldsymbol{\eta}_2 = (\eta_{2i}, \dots, \eta_{2n})^\top$ são preditores lineares e $\mathbf{x}_{1i}, \dots, \mathbf{x}_{pi}, \mathbf{w}_{1i}, \dots, \mathbf{w}_{qi}$ são observações conhecidas, não necessariamente exclusivas. Para garantir a qualidade do ajuste e a positividade dos parâmetros μ_1 e μ_2 , as funções de ligações $g_1 : (0, \infty) \rightarrow \mathbb{R}$ e $g_2 : (0, \infty) \rightarrow \mathbb{R}$ são estritamente monótonas e duas vezes diferenciáveis, com domínio em \mathbb{R}^+ e imagem em \mathbb{R} . Podemos escolher diferentes funções de ligação que garantem a positividade desses parâmetros. Optamos pela mais utilizada nesses casos que é a função logarítmica. Discussões mais aprofundadas sobre funções de ligação podem ser encontradas em [Atkinson \(1985\)](#) e [McCullagh e Nelder \(1989\)](#).

Para a estimação dos parâmetros do modelo por máxima verossimilhança temos que encontrar a densidade da variável aleatória mista Y . Defina o vetor de parâmetros $\boldsymbol{\theta} = (\mu_1, \mu_2, \phi, \rho)$. Temos que,

$$\begin{aligned} P(Y \leq y) &= P(Y \leq y | U = 1)P(U = 1) + P(Y \leq y | U = 0)P(U = 0) \\ &= P(Y_1^* \leq y | Y_2^* > 1)P(Y_2^* > 1) + P(Y_2^* \leq 1) \\ &= 1\{y > 0\}F_{Y_1^* | Y_2^* > 1}(y; \boldsymbol{\theta}) (1 - F_{Y_2^*}(1)) + 1\{y = 0\}F_{Y_2^*}(1) \\ &= \begin{cases} F_{Y_1^* | Y_2^* > 1}(y; \boldsymbol{\theta}) (1 - F_{Y_2^*}(1)), & \text{se } y > 0, \\ F_{Y_2^*}(1), & \text{se } y = 0, \end{cases} \end{aligned}$$

Daí, segue que

$$f_Y(y; \boldsymbol{\theta}) = \begin{cases} f_{Y_1^*|Y_2^* > 1}(y; \boldsymbol{\theta}) (1 - F_{Y_2^*}(1)), & \text{se } y > 0, \\ F_{Y_2^*}(1), & \text{se } y = 0, \end{cases} \quad (3.3)$$

ou seja, a densidade de Y possui um componente discreto e um componente contínuo. Seu componente discreto é definido pelo modelo Probit

$$P(U = u) = (1 - F_{Y_2^*}(1))^u F_{Y_2^*}(1)^{1-u}, \text{ para } u = 0, 1, \quad (3.4)$$

em que $F_{Y_2^*}(1)$ é a fda de $Y_2^* \sim BS(\mu_2, 1)$ avaliada no ponto 1, ou seja,

$$F_{Y_2^*}(1) = \Phi \left[\sqrt{\frac{1}{2}} \left(\sqrt{\frac{2}{\mu_2}} - \sqrt{\frac{\mu_2}{2}} \right) \right]. \quad (3.5)$$

O componente contínuo pode ser obtido a partir do trabalho de [Arellano-Valle et al. \(2006\)](#), onde os autores desenvolvem uma fórmula geral para as distribuições que surgem a partir do mecanismo de seleção $\mathbf{Y}|\mathbf{U} \in C$, em que $\mathbf{Y} \in \mathbb{R}^q$, $\mathbf{U} \in \mathbb{R}^p$ e C é um subconjunto mensurável de \mathbb{R}^p tal que $0 < P(\mathbf{U} \in C) < 1$. No nosso caso, não observamos Y_2^* diretamente, sabemos se $Y_2^* > 1$ ou $Y_2^* \leq 1$. Além disso, observamos $Y = Y_1^*$ somente quando $Y_2^* > 1$. Neste caso, dizemos que Y está sujeito a um truncamento oculto. De acordo com [Arellano-Valle et al. \(2006\)](#), a densidade de $Z = Y_1^*|Y_2^* > 1$, que representa o componente contínuo de Y , é dada por

$$f_{Y_1^*|Y_2^* > 1}(y; \boldsymbol{\theta}) = f_{Y_1^*}(y; \boldsymbol{\theta}) \frac{P(Y_2^* > 1|Y_1^* = y, \boldsymbol{\theta})}{P(Y_2^* > 1)}, \quad (3.6)$$

em que $f_{Y_1^*}(y; \boldsymbol{\theta})$ é a densidade de $Y_1^* \sim BS(\mu_1, \phi_1 \equiv \phi)$ e $P(Y_2^* > 1|Y_1^* = y)$ é obtido a partir de [\(1.29\)](#). A igualdade [\(3.6\)](#) porém, já aparece nos trabalhos de [Arnold e Beaver \(2000\)](#), [Arellano-Valle et al. \(2002\)](#) e [Azzalini \(2005\)](#). Segue, portanto, que no modelo Heckman-BS, a densidade de Y é dada por

$$\begin{aligned} f_Y(y; \boldsymbol{\theta}) &= 1\{y > 0\} \frac{1}{2\sqrt{2\pi}} \exp \left[-\frac{\phi}{4} \left(\sqrt{\frac{y(\phi+1)}{\phi\mu_1}} - \sqrt{\frac{\phi\mu_1}{y(\phi+1)}} \right)^2 \right] \\ &\times \left(\frac{\sqrt{\phi+1}}{\sqrt{\phi\mu_1 y}} + \frac{\sqrt{\phi\mu_1}}{\sqrt{y^3(\phi+1)}} \right) \sqrt{\frac{\phi}{2}} \\ &\times \Phi \left[\frac{\mu_2 - 2}{2\sqrt{\mu_2(1-\rho^2)}} + \frac{\sqrt{\phi}\rho}{\sqrt{2(1-\rho^2)}} \left(\sqrt{\frac{(\phi+1)y}{\phi\mu_1}} - \sqrt{\frac{\phi\mu_1}{y(\phi+1)}} \right) \right] \\ &+ 1\{y = 0\} \Phi \left\{ \sqrt{\frac{1}{2}} \left[\sqrt{\frac{2}{\mu_2}} - \sqrt{\frac{\mu_2}{2}} \right] \right\}, \end{aligned} \quad (3.7)$$

em que $\Phi(\cdot)$ é a função de distribuição acumulada da normal padrão. O modelo [\(3.7\)](#) é identificável no sentido de que para qualquer par de parâmetros $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ temos $f_Y(y; \boldsymbol{\theta}_1) \neq f_Y(y; \boldsymbol{\theta}_2)$.

A densidade condicional, que corresponde ao componente contínuo da variável mista Y , foi introduzida

na literatura no trabalho recente de [Jamalizadeh e Kundu \(2015\)](#). Eles propõem uma generalização da distribuição Birnbaum-Saunders multivariada baseada na distribuição Skew-Normal multivariada e apresentam as distribuições marginal, recíproca e condicional da distribuição proposta. Assim, seguindo a estrutura obtida a partir da distribuição Birnbaum-Saunders Skew-Normal introduzida por [Vilca-Labra e Leiva-Sánchez \(2006\)](#) e generalizada no trabalho de [Jamalizadeh e Kundu \(2015\)](#), podemos rescrever de forma simplificada

$$f_{Y_1^*|Y_2^*>1}(y; \boldsymbol{\theta}) = \frac{\phi(a(y; \mu_1, \phi)) \Phi(\lambda_0 + \lambda_1 a(y; \mu_1, \phi))}{\Phi\left(\frac{\lambda_0}{\sqrt{1 + \lambda_1^2}}\right)} A(y; \mu_1, \phi), \quad (3.8)$$

em que,

$$a(y; \mu_i, \phi) = \frac{\phi}{2} \left(\sqrt{\frac{y(\phi + 1)}{\phi \mu_1}} - \sqrt{\frac{\phi \mu_1}{y(\phi + 1)}} \right), \quad A(y; \mu_1, \phi) = \frac{1}{2} \sqrt{\frac{\phi}{2}} \left(\sqrt{\frac{\phi + 1}{\phi \mu_1 y}} + \sqrt{\frac{\phi \mu_1}{y^3(\phi + 1)}} \right)$$

$$\lambda_0 = \frac{\mu_2 - 2}{2\sqrt{\mu_2(1 - \rho^2)}} \in \mathbb{R}, \quad \text{e} \quad \lambda_1 = \frac{\rho}{\sqrt{1 - \rho^2}} \in \mathbb{R},$$

Tomando $\lambda_0 = 0$ em (3.8), temos

$$f_{Y_1^*|Y_2^*>1}(y; \boldsymbol{\theta}) = \frac{1}{2} \phi(a(y; \mu_1, \phi)) \Phi(\lambda_1 a(y; \mu_1, \phi)) A(y; \mu_1, \phi), \quad (3.9)$$

que coincide com a distribuição univariada Skew-normal Birnbaum-Saunders (BSSN) com parâmetro de forma ϕ , parâmetro de escala μ_1 e parâmetro de assimetria λ_1 , definida em [Vilca-Labra e Leiva-Sánchez \(2006\)](#). Note que o parâmetro ρ determina λ_1 , porém, a assimetria resulta também de λ_0 , que é um parâmetro de locação, função do parâmetro de seleção μ_2 . O parâmetro ρ determina a correlação de Y_{1i}^* e Y_{2i}^* . Quando $\rho = 0$, temos que Y_1^* é independente de Y_2^* e, nesse caso, não há viés de seleção.

3.2.1 Função de Verossimilhança e Vetor Escore

Nesta seção, apresentamos a estimação dos parâmetros por máxima verossimilhança. A função de verossimilhança para o modelo é obtida a partir da função densidade dada em (3.7) e definida por:

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^n \mathcal{L}_i(\mu_{1i}, \mu_{2i}, \phi, \rho), \quad (3.10)$$

em que

$$\begin{aligned}
\mathcal{L}_i(\mu_{1i}, \mu_{2i}, \phi, \rho) &= u_i \log g(y_i) + (1 - u_i) \log F_{Y_2}^*(1) \\
&= u_i \left\{ -\frac{\phi}{4} \left(\sqrt{\frac{y_i(\phi+1)}{\phi\mu_{1i}}} - \sqrt{\frac{\phi\mu_{1i}}{y_i(\phi+1)}} \right)^2 \right. \\
&\quad + \log \Phi \left[\frac{[\mu_{2i} - 2]}{2\sqrt{\mu_{2i}(1-\rho^2)}} + \frac{\sqrt{\phi}\rho}{\sqrt{2(1-\rho^2)}} \left(\sqrt{\frac{(\phi+1)y_i}{\phi\mu_{1i}}} - \sqrt{\frac{\phi\mu_{1i}}{y_i(\phi+1)}} \right) \right] \\
&\quad + \log \left(\frac{\sqrt{\phi+1}}{\sqrt{\phi\mu_{1i}y_i}} + \frac{\sqrt{\phi\mu_{1i}}}{\sqrt{y_i^3(\phi+1)}} \right) + \frac{1}{4} \sqrt{\frac{\phi}{\pi}} \left. \right\} \\
&\quad + (1 - u_i) \log \Phi \left\{ \sqrt{\frac{1}{2}} \left[\sqrt{\frac{2}{\mu_{2i}}} - \sqrt{\frac{\mu_{2i}}{2}} \right] \right\}, \tag{3.11}
\end{aligned}$$

$\mu_{1i} = \exp(\eta_{1i})$ e $\mu_{2i} = \exp(\eta_{2i})$, como definido anteriormente, são funções dos vetores de parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$, respectivamente. O método da máxima verossimilhança busca o vetor $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\gamma}}^\top, \hat{\boldsymbol{\beta}}^\top, \hat{\phi}, \hat{\rho})$ que maximiza a função de verossimilhança. Tal vetor é solução do sistema $U(\boldsymbol{\theta}) = 0$, em que $H|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ é negativa definida, $U(\cdot)$ é o vetor escore e H é a matriz Hessiana.

O vetor escore $U(\boldsymbol{\theta}) = \partial\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ tem componentes

$$\begin{aligned}
\frac{\partial\mathcal{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}_j} &= \sum_{i=1}^n \frac{\partial\mathcal{L}_i(\boldsymbol{\theta})}{\partial\mu_{2i}} \frac{\partial\mu_{2i}}{\partial\eta_{2i}} \frac{\partial\eta_{2i}}{\partial\boldsymbol{\gamma}_j}, \quad j = 1, \dots, q, \\
\frac{\partial\mathcal{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}_k} &= \sum_{i=1}^n \frac{\partial\mathcal{L}_i(\boldsymbol{\theta})}{\partial\mu_{1i}} \frac{\partial\mu_{1i}}{\partial\eta_{1i}} \frac{\partial\eta_{1i}}{\partial\boldsymbol{\beta}_k}, \quad k = 1, \dots, p, \\
\frac{\partial\mathcal{L}(\boldsymbol{\theta})}{\partial\phi} &= \sum_{i=1}^n \frac{\partial\mathcal{L}_i(\boldsymbol{\theta})}{\partial\phi}, \\
\frac{\partial\mathcal{L}(\boldsymbol{\theta})}{\partial\rho} &= \sum_{i=1}^n \frac{\partial\mathcal{L}_i(\boldsymbol{\theta})}{\partial\rho},
\end{aligned}$$

em que $\frac{\partial\mu_{2i}}{\partial\eta_{2i}} = \mu_{2i}$, $\frac{\partial\mu_{1i}}{\partial\eta_{1i}} = \mu_{1i}$, $\frac{\partial\eta_{2i}}{\partial\boldsymbol{\gamma}_j} = \frac{\partial\boldsymbol{w}_i^\top \boldsymbol{\gamma}}{\partial\boldsymbol{\gamma}_j} = \boldsymbol{w}_{ji}$ e $\frac{\partial\eta_{1i}}{\partial\boldsymbol{\beta}_k} = \frac{\partial\boldsymbol{x}_i^\top \boldsymbol{\beta}}{\partial\boldsymbol{\beta}_k} = \boldsymbol{x}_{ki}$, para $i = 1, \dots, n$. Visando simplificar as notações define:

$$t_{0i}(y_i, \mu_{1i}, \phi) = \left(\sqrt{\frac{y_i(\phi+1)}{\phi\mu_{1i}}} - \sqrt{\frac{\phi\mu_{1i}}{y_i(\phi+1)}} \right), \quad t_{1i}(y_i, \mu_{1i}, \phi) = -\frac{\phi}{4} [t_{0i}(y_i, \mu_{1i}, \phi)]^2,$$

$$t_{2i}(y_i, \mu_{1i}, \phi) = \left(\frac{\sqrt{\phi+1}}{\sqrt{\phi\mu_{1i}y_i}} + \frac{\sqrt{\phi\mu_{1i}}}{\sqrt{y_i^3(\phi+1)}} \right), \quad t_3(\phi) = \frac{1}{4} \sqrt{\frac{\phi}{\pi}},$$

$$t_{4i}(\mu_{2i}, \rho) = \frac{\mu_{2i} - 2}{2\sqrt{\mu_{2i}(1-\rho^2)}}, \quad t_5(\phi, \rho) = \frac{\rho\sqrt{\phi}}{\sqrt{2(1-\rho^2)}}, \quad t_{6i}(\mu_{2i}) = \sqrt{\frac{1}{2}} \left[\sqrt{\frac{2}{\mu_{2i}}} - \sqrt{\frac{\mu_{2i}}{2}} \right],$$

e note que, para $k = 1, \dots, p$, e $j = 1, \dots, q$, temos:

$$\frac{\partial t_{0i}}{\partial \gamma_j} = \frac{\partial t_{1i}}{\partial \gamma_j} = \frac{\partial t_{2i}}{\partial \gamma_j} = \frac{\partial t_3}{\partial \gamma_j} = \frac{\partial t_5}{\partial \gamma_j} = 0, \quad \frac{\partial t_{4i}}{\partial \gamma_j} = \frac{\partial t_{4i}}{\partial \mu_{2i}} \frac{\partial \mu_{2i}}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\partial \gamma_j} = \frac{\mu_{2i} + 2}{4\sqrt{\mu_{2i}(1-\rho^2)}} \mathbf{w}_{ji},$$

$$\frac{\partial t_{6i}}{\partial \gamma_j} = \frac{\partial t_{6i}}{\partial \mu_{2i}} \frac{\partial \mu_{2i}}{\partial \eta_{2i}} \frac{\partial \eta_{2i}}{\partial \gamma_j} = -\frac{1}{2\sqrt{2}} \left(\sqrt{\frac{2}{\mu_{2i}}} + \sqrt{\frac{\mu_{2i}}{2}} \right) \mathbf{w}_{ji}, \quad \frac{\partial t_3}{\partial \beta_k} = \frac{\partial t_{4i}}{\partial \beta_k} = \frac{\partial t_5}{\partial \beta_k} = \frac{\partial t_{6i}}{\partial \beta_k} = 0,$$

$$\frac{\partial t_{0i}}{\partial \beta_k} = \frac{\partial t_{0i}}{\partial \mu_{1i}} \frac{\partial \mu_{1i}}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \beta_k} = -\frac{1}{2} \left(\sqrt{\frac{y_i(\phi+1)}{\phi\mu_{1i}}} + \sqrt{\frac{\phi\mu_{1i}}{y_i(\phi+1)}} \right) \mathbf{x}_{ki}, \quad \frac{\partial t_{4i}}{\partial \phi} = \frac{\partial t_{6i}}{\partial \phi} = 0,$$

$$\frac{\partial t_{1i}}{\partial \beta_k} = \frac{\partial t_{1i}}{\partial \mu_{1i}} \frac{\partial \mu_{1i}}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \beta_k} = -\frac{\phi}{2} t_{0i} \frac{\partial t_{0i}}{\partial \beta_k}, \quad \frac{\partial t_{0i}}{\partial \rho} = \frac{\partial t_{1i}}{\partial \rho} = \frac{\partial t_{2i}}{\partial \rho} = \frac{\partial t_3}{\partial \rho} = \frac{\partial t_{6i}}{\partial \rho} = 0,$$

$$\frac{\partial t_{2i}}{\partial \beta_k} = \frac{\partial t_{2i}}{\partial \mu_{1i}} \frac{\partial \mu_{1i}}{\partial \eta_{1i}} \frac{\partial \eta_{1i}}{\partial \beta_k} = \frac{1}{2} \left(\sqrt{\frac{\phi\mu_{1i}}{y_i^3(\phi+1)}} - \sqrt{\frac{\phi+1}{\phi\mu_{1i}y_i}} \right) \mathbf{x}_{ki}, \quad \frac{\partial t_3}{\partial \phi} = \frac{1}{8\sqrt{\pi\phi}},$$

$$\frac{\partial t_{0i}}{\partial \phi} = -\frac{1}{2} \left(\sqrt{\frac{y_i}{(\phi+1)\phi^3\mu_{1i}}} + \sqrt{\frac{\mu_{1i}}{y_i(\phi+1)^3\phi}} \right), \quad \frac{\partial t_5}{\partial \phi} = \frac{\rho}{2\sqrt{2\phi(1-\rho^2)}},$$

$$\frac{\partial t_{1i}}{\partial \phi} = -\frac{t_{0i}^2}{4} + \frac{t_{0i}}{4} \left(\sqrt{\frac{y_i}{(\phi+1)\phi\mu_{1i}}} + \sqrt{\frac{\phi\mu_{1i}}{y_i(\phi+1)^3}} \right), \quad \frac{\partial t_{4i}}{\partial \rho} = \frac{\rho t_{4i}}{(1-\rho^2)},$$

$$\frac{\partial t_{2i}}{\partial \phi} = \frac{1}{2} \left(\sqrt{\frac{\mu_{1i}}{y_i^3(\phi+1)^3\phi}} - \frac{1}{\sqrt{y_i\mu_{1i}(\phi+1)\phi^3}} \right), \quad \frac{\partial t_5}{\partial \rho} = \frac{t_5}{\rho(1-\rho^2)}.$$

Ainda com o objetivo de simplificar as notações, para $i = 1, \dots, n$ definimos:

$$f_{1i}(y_i, \mu_{1i}, \phi) = t_{1i}(y_i, \mu_{1i}, \phi), \quad f_{2i}(y_i, \mu_{1i}, \phi) = \log [t_{2i}(y_i, \mu_{1i}, \phi)], \quad f_3(\phi) = \log [t_3(\phi)],$$

$$f_{4i}(y_i, \mu_{1i}, \mu_{2i}, \phi, \rho) = \log \Phi(t_{4i} + t_5 t_{0i}), \quad f_{5i}(\mu_{2i}) = \log \Phi[t_{6i}(\mu_{2i})].$$

Segue que:

$$\frac{\partial f_{1i}}{\partial \gamma_j} = \frac{\partial f_{2i}}{\partial \gamma_j} = \frac{\partial f_3}{\partial \gamma_j} = 0, \quad \frac{\partial f_{4i}}{\partial \gamma_j} = \frac{\phi(t_{4i} + t_5 t_{0i})}{\Phi(t_{4i} + t_5 t_{0i})} \frac{\partial t_{4i}}{\partial \gamma_j}, \quad \frac{\partial f_{5i}}{\partial \gamma_j} = \frac{\phi(t_{6i})}{\Phi(t_{6i})} \frac{\partial t_6}{\partial \gamma_j},$$

$$\frac{\partial f_{1i}}{\partial \beta_k} = -\frac{\phi}{2} t_{0i} \frac{\partial t_{0i}}{\partial \beta_k}, \quad \frac{\partial f_{2i}}{\partial \beta_k} = \frac{1}{t_{2i}} \frac{\partial t_{2i}}{\partial \beta_k}, \quad \frac{\partial f_3}{\partial \beta_k} = \frac{\partial f_{5i}}{\partial \beta_k} = 0,$$

$$\frac{\partial f_{4i}}{\partial \beta_k} = \frac{\phi(t_{4i} + t_5 t_{0i})}{\Phi(t_{4i} + t_5 t_{0i})} t_5 \frac{\partial t_{0i}}{\partial \beta_k}, \quad \frac{\partial f_{1i}}{\partial \phi} = \frac{\partial t_{1i}}{\partial \phi}, \quad \frac{\partial f_{2i}}{\partial \phi} = \frac{1}{t_{2i}} \frac{\partial t_{2i}}{\partial \phi},$$

$$\frac{\partial f_3}{\partial \phi} = \frac{1}{2\phi}, \quad \frac{\partial f_{4i}}{\partial \phi} = \frac{\phi(t_{4i} + t_5 t_{0i})}{\Phi(t_{4i} + t_5 t_{0i})} \left(\frac{\partial t_5}{\partial \phi} t_{0i} + t_5 \frac{\partial t_{0i}}{\partial \phi} \right), \quad \frac{\partial f_{5i}}{\partial \phi} = 0,$$

$$\frac{\partial f_{1i}}{\partial \rho} = \frac{\partial f_{2i}}{\partial \rho} = \frac{\partial f_3}{\partial \rho} = \frac{\partial f_{5i}}{\partial \rho} = 0, \quad \frac{\partial f_{4i}}{\partial \rho} = \frac{\phi(t_{4i} + t_5 t_{0i})}{\Phi(t_{4i} + t_5 t_{0i})} \left(\frac{\partial t_{4i}}{\partial \rho} + \frac{\partial t_5}{\partial \rho} t_{0i} \right).$$

Assim, temos que os possíveis estimadores de máxima verossimilhança são obtidos pela solução do sistema:

$$\left\{ \begin{array}{l} \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \gamma_j} = \sum_{i=1}^n u_i \left(\frac{\partial f_{4i}}{\partial \gamma_j} \right) + \sum_{i=1}^n (1 - u_i) \left(\frac{\partial f_{5i}}{\partial \gamma_j} \right) = 0, \quad j = 1, \dots, q, \\ \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \beta_k} = \sum_{i=1}^n u_i \left(\frac{\partial f_{1i}}{\partial \beta_k} + \frac{\partial f_{2i}}{\partial \beta_k} + \frac{\partial f_3}{\partial \beta_k} + \frac{\partial f_{4i}}{\partial \beta_k} \right) = 0, \quad k = 1, \dots, p, \\ \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \phi} = \sum_{i=1}^n u_i \left(\frac{\partial f_{1i}}{\partial \phi} + \frac{\partial f_{2i}}{\partial \phi} + \frac{\partial f_3}{\partial \phi} + \frac{\partial f_{4i}}{\partial \phi} \right) = 0, \\ \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \rho} = \sum_{i=1}^n u_i \frac{\partial f_{4i}}{\partial \rho} = 0. \end{array} \right.$$

Como não há expressões de forma fechada para os EMV's dos parâmetros, as estimativas de máxima verossimilhança são obtidas através de um método de maximização numérico da função de log-verossimilhança usando um algoritmo de otimização não linear, por exemplo, o algoritmo quase-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS). Para detalhes, ver [Press et al. \(1992\)](#) e [Wright e Nocedal \(1999\)](#). O método BFGS é implementado no *software* R pelas funções `optim(.)` e `optimx(.)`. Para inferências intervalares e testes de hipóteses é necessário o conhecimento da matriz Hessiana que é apresentada no Apêndice B.

3.3 Comportamento dos Estimadores em Amostras Finitas

Nesta seção avaliaremos por meio de simulações Monte Carlo o modelo de seleção BS. Comparamos sua performance com os modelos de Heckman Clássico ([Heckman,1976](#)), Skew-normal ([Ogundimu e Hutton,2016](#)) e Heckman-t ([Marchenko e Genton,2012](#)) quando os dados são gerados a partir do modelo de seleção BS. Consideramos dois cenários de simulação distintos, o Cenário 1 com restrição de exclusão, com estrutura das médias dada por:

$$\begin{aligned} g_1(\mu_{1i}) &= \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i}, \\ g_2(\mu_{2i}) &= \gamma_1 + \gamma_2 x_{2i} + \gamma_3 x_{3i} + \gamma_4 x_{4i}, \quad \text{para } i = 1, 2, \dots, n, \end{aligned} \tag{3.12}$$

e o Cenário 2, sem restrição de exclusão, com a estrutura das médias dada por:

$$\begin{aligned} g_1(\mu_{1i}) &= \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i}, \\ g_2(\mu_{2i}) &= \gamma_1 + \gamma_2 x_{2i} + \gamma_3 x_{3i}, \text{ para } i = 1, 2, \dots, n, \end{aligned} \quad (3.13)$$

em que, nos dois casos, os regressores $x_{ji} \stackrel{iid}{\sim} \mathcal{N}(0, 1), j = 2, 3, 4, \forall i$. Além disso, x_{ji} e x_{ki} foram gerados de forma independente para todo $j \neq k$. Em ambos os cenários focamos na estimação do parâmetro ρ , para isso, avaliamos o ajuste dos modelos para sete valores distintos deste parâmetro, $\rho = 0, \rho = \pm 0.2, \rho = \pm 0.5$ e $\rho = \pm 0.7$. Os tamanhos amostrais considerados foram $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo em todos os casos. Consideramos $n \geq 500$ para as simulações, pois frequentemente, banco de dados para o ajuste de modelos de seleção amostral possuem tamanho $n \geq 500$.

Em cada réplica Monte Carlo, geramos n vetores $(Y_1^*, Y_2^*) \stackrel{ind.}{\sim} BS(\mu_{1i}, \mu_{2i}, \phi, 1, \rho)$ em que μ_{1i} e μ_{2i} são dados em (3.12) e (3.13), para os Cenários 1 e 2, respectivamente. Consideramos a função log como função de ligação, dessa forma, utilizamos para o processo de geração dos dados:

$$\mu_{1i} = \exp\left(\beta_1 + \sum_{j=2}^p x_{ij}\beta_j\right) \quad \text{e} \quad \mu_{2i} = \exp\left(\gamma_1 + \sum_{j=2}^q w_{ij}\gamma_j\right), \text{ para } i = 1, \dots, n. \quad (3.14)$$

A partir daí, assumamos $Y_i = Y_{1i}^* U_i$, em que $U_i = 1\{Y_{2i}^* > 1\}$. Ajustamos então os modelos Heckman-BS, Heckman Clássico, Skew-normal e Heckman-t aos dados simulados e para cada cenário considerado, a fim de examinar o desempenho dos EMV's dos modelos ajustados, calculamos a média empírica das estimativas dos parâmetros e a raiz quadrada do erro quadrático médio (REQM). Nos cenários simulados o grau de censura (valores para os quais $Y_{2i}^* \leq 1$) é de aproximadamente 30%.

3.3.1 Cenário 1

Nosso objetivo aqui é usar simulação de Monte Carlo para avaliar o desempenho dos EMV's dos parâmetros do modelo Heckman-BS. Considere $\beta_1 = 1, \beta_2 = 0.7, \beta_3 = 1.1, \gamma_1 = 1.6, \gamma_2 = 0.8, \gamma_3 = 0.2, \gamma_4 = 0.7, \phi = 1.2$ e ρ como descrito acima.

Assim, temos o modelo:

$$Y_i = Y_{1i}^* 1\{Y_{2i}^* > 1\}, \quad (3.15)$$

em que $(Y_{1i}^*, Y_{2i}^*)^\top \stackrel{ind.}{\sim} BS_2(\boldsymbol{\mu}_i, \phi, 1, \rho)$, $\boldsymbol{\mu}_i = (\mu_{1i}, \mu_{2i})^\top \in \mathbb{R}^+ \times \mathbb{R}^+, i = 1, 2, \dots, n$, $\phi \in \mathbb{R}^+, \rho \in (-1, 1)$ e $\boldsymbol{\mu}_i$ é definido conforme (3.14) para $i = 1, \dots, n$.

Para a estimação dos parâmetros utilizamos o método BFGS. O valor inicial para a função $optim(\cdot)$ foi obtido pelo método de dois passos de Heckman. Nossas simulações, conforme resultados apresentados nas Tabelas 3.1 a 3.7, indicam uma boa performance dos EMV's do modelo Heckman-BS sob a correta

especificação do modelo, uma vez que as estimativas da raiz quadrada do erro quadrático médio dos parâmetros estimados são próximas de zero para todos os tamanhos amostrais e tendem a zero quando o tamanho amostral aumenta conforme apresentado nas Tabelas 3.1 a 3.7. Fato que indica consistência dos estimadores de máxima verossimilhança. Comparado aos demais modelos, as estimativas de máxima verossimilhança do modelo Heckman-BS apresentaram menor viés e, em geral, apresentaram menor erro quadrático médio.

Tabela 3.1: Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas após ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 1 e $\rho = -0.7$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.

Parâmetros	n	Heckman BS		Heckman Clássico		Heckman Skew		Heckman t-Student		
		Média	REQM	Média	REQM	Média	REQM	Média	REQM	
γ_1	1.600	500	1.608	0.086	0.719	0.884	1.048	0.572	0.723	0.879
		1000	1.605	0.061	0.718	0.884	1.053	0.563	0.722	0.880
		2000	1.601	0.041	0.713	0.887	1.076	0.529	0.717	0.884
γ_2	0.800	500	0.815	0.096	0.650	0.169	0.572	0.246	0.654	0.166
		1000	0.805	0.064	0.644	0.165	0.571	0.238	0.647	0.162
		2000	0.801	0.044	0.640	0.164	0.568	0.237	0.642	0.162
γ_3	0.200	500	0.205	0.079	0.163	0.074	0.144	0.081	0.164	0.074
		1000	0.202	0.059	0.163	0.060	0.145	0.071	0.164	0.060
		2000	0.200	0.042	0.159	0.054	0.141	0.067	0.159	0.054
γ_4	0.700	500	0.712	0.091	0.573	0.146	0.502	0.218	0.577	0.143
		1000	0.704	0.064	0.565	0.145	0.499	0.212	0.568	0.142
		2000	0.701	0.042	0.564	0.141	0.498	0.207	0.566	0.138
β_1	1.000	500	0.994	0.138	0.367	0.641	-0.453	1.509	0.367	0.641
		1000	0.995	0.092	0.370	0.634	-0.422	1.463	0.370	0.634
		2000	0.997	0.062	0.370	0.631	-0.462	1.476	0.370	0.631
β_2	0.700	500	0.700	0.067	0.717	0.072	0.705	0.072	0.717	0.072
		1000	0.701	0.047	0.715	0.052	0.704	0.051	0.715	0.052
		2000	0.699	0.033	0.715	0.039	0.705	0.037	0.715	0.039
β_3	1.100	500	1.100	0.055	1.104	0.059	1.101	0.060	1.104	0.060
		1000	1.101	0.036	1.106	0.039	1.103	0.039	1.106	0.039
		2000	1.100	0.030	1.105	0.032	1.102	0.032	1.105	0.032
ϕ	1.200	500	1.222	0.140	1.124	0.092	1.502	0.420	1.118	0.097
		1000	1.210	0.101	1.129	0.081	1.465	0.349	1.124	0.085
		2000	1.207	0.066	1.128	0.076	1.461	0.301	1.124	0.080
ρ	-0.700	500	-0.686	0.122	-0.648	0.130	-0.751	0.142	-0.649	0.130
		1000	-0.696	0.080	-0.662	0.090	-0.759	0.113	-0.663	0.089
		2000	-0.697	0.053	-0.660	0.068	-0.761	0.089	-0.661	0.067

Tabela 3.2: Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas após ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 1 e $\rho = -0.5$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.

Parâmetros	n	Heckman BS		Heckman Clássico		Heckman Skew		Heckman t-Student		
		Média	REQM	Média	REQM	Média	REQM	Média	REQM	
γ_1	1.600	500	1.610	0.086	0.720	0.883	0.939	0.679	0.723	0.880
		1000	1.606	0.061	0.718	0.883	0.944	0.668	0.721	0.881
		2000	1.601	0.042	0.713	0.888	0.943	0.663	0.715	0.886
γ_2	0.800	500	0.817	0.097	0.649	0.170	0.614	0.205	0.652	0.167
		1000	0.806	0.064	0.643	0.166	0.613	0.196	0.646	0.163
		2000	0.803	0.044	0.639	0.165	0.614	0.191	0.641	0.163
γ_3	0.200	500	0.204	0.081	0.162	0.075	0.154	0.078	0.163	0.075
		1000	0.201	0.059	0.162	0.061	0.155	0.065	0.163	0.060
		2000	0.201	0.044	0.159	0.054	0.153	0.059	0.160	0.054
γ_4	0.700	500	0.713	0.094	0.570	0.150	0.538	0.184	0.572	0.148
		1000	0.703	0.066	0.560	0.150	0.533	0.178	0.563	0.148
		2000	0.700	0.043	0.559	0.145	0.536	0.169	0.561	0.144
β_1	1.000	500	0.999	0.141	0.375	0.636	-0.291	1.356	0.375	0.636
		1000	0.998	0.099	0.376	0.628	-0.260	1.300	0.376	0.628
		2000	0.998	0.067	0.377	0.625	-0.258	1.281	0.377	0.625
β_2	0.700	500	0.699	0.072	0.711	0.077	0.704	0.079	0.711	0.077
		1000	0.701	0.052	0.710	0.056	0.704	0.057	0.710	0.056
		2000	0.699	0.035	0.710	0.040	0.706	0.040	0.710	0.040
β_3	1.100	500	1.100	0.057	1.103	0.061	1.101	0.062	1.103	0.062
		1000	1.101	0.037	1.105	0.039	1.103	0.039	1.105	0.039
		2000	1.101	0.031	1.103	0.033	1.102	0.033	1.103	0.033
ϕ	1.200	500	1.211	0.124	1.131	0.084	1.393	0.318	1.126	0.088
		1000	1.204	0.092	1.133	0.076	1.354	0.238	1.129	0.080
		2000	1.204	0.061	1.131	0.073	1.332	0.182	1.129	0.075
ρ	-0.500	500	-0.486	0.165	-0.455	0.176	-0.533	0.205	-0.455	0.178
		1000	-0.495	0.121	-0.468	0.126	-0.541	0.154	-0.469	0.127
		2000	-0.496	0.080	-0.468	0.087	-0.535	0.105	-0.468	0.087

Tabela 3.3: Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas após ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 1 e $\rho = -0.2$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.

Parâmetros	n	Heckman BS		Heckman Clássico		Heckman Skew		Heckman t-Student		
		Média	REQM	Média	REQM	Média	REQM	Média	REQM	
γ_1	1.600	500	1.611	0.086	0.720	0.883	0.778	0.833	0.722	0.880
		1000	1.606	0.061	0.718	0.884	0.777	0.828	0.720	0.881
		2000	1.602	0.042	0.713	0.887	0.771	0.831	0.715	0.885
γ_2	0.800	500	0.818	0.098	0.649	0.170	0.642	0.176	0.651	0.168
		1000	0.807	0.065	0.644	0.165	0.640	0.169	0.646	0.163
		2000	0.803	0.046	0.640	0.164	0.638	0.166	0.642	0.163
γ_3	0.200	500	0.203	0.082	0.162	0.076	0.160	0.077	0.162	0.076
		1000	0.201	0.058	0.162	0.060	0.161	0.061	0.163	0.060
		2000	0.201	0.044	0.160	0.054	0.159	0.054	0.160	0.053
γ_4	0.700	500	0.716	0.094	0.569	0.152	0.563	0.158	0.571	0.150
		1000	0.703	0.066	0.558	0.152	0.554	0.156	0.560	0.150
		2000	0.700	0.043	0.557	0.147	0.555	0.149	0.559	0.146
β_1	1.000	500	1.005	0.128	0.390	0.624	-0.045	1.119	0.390	0.624
		1000	1.002	0.093	0.389	0.617	0.025	1.022	0.389	0.618
		2000	1.002	0.063	0.390	0.613	0.036	0.990	0.390	0.613
β_2	0.700	500	0.697	0.075	0.701	0.083	0.701	0.085	0.701	0.084
		1000	0.699	0.056	0.701	0.062	0.700	0.062	0.701	0.062
		2000	0.698	0.038	0.702	0.043	0.702	0.044	0.702	0.044
β_3	1.100	500	1.099	0.057	1.101	0.062	1.100	0.062	1.101	0.062
		1000	1.100	0.038	1.102	0.041	1.102	0.041	1.102	0.041
		2000	1.101	0.031	1.101	0.033	1.101	0.034	1.101	0.034
ϕ	1.200	500	1.200	0.103	1.139	0.074	1.266	0.179	1.135	0.077
		1000	1.198	0.072	1.137	0.069	1.227	0.115	1.133	0.073
		2000	1.200	0.048	1.135	0.068	1.206	0.076	1.132	0.071
ρ	-0.200	500	-0.196	0.196	-0.183	0.218	-0.199	0.246	-0.183	0.221
		1000	-0.202	0.152	-0.192	0.164	-0.207	0.181	-0.192	0.166
		2000	-0.201	0.102	-0.191	0.112	-0.202	0.119	-0.191	0.112

Tabela 3.4: Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas após ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 1 e $\rho = 0$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.

Parâmetros	n	Heckman BS		Heckman Clássico		Heckman Skew		Heckman t-Student		
		Média	REQM	Média	REQM	Média	REQM	Média	REQM	
γ_1	1.600	500	1.612	0.085	0.721	0.882	0.715	0.893	0.723	0.880
		1000	1.606	0.060	0.718	0.883	0.720	0.883	0.721	0.881
		2000	1.602	0.043	0.714	0.887	0.714	0.887	0.716	0.885
γ_2	0.800	500	0.815	0.097	0.647	0.171	0.645	0.174	0.649	0.170
		1000	0.808	0.065	0.644	0.165	0.643	0.166	0.647	0.163
		2000	0.804	0.045	0.641	0.164	0.640	0.164	0.642	0.162
γ_3	0.200	500	0.205	0.081	0.163	0.075	0.162	0.075	0.164	0.075
		1000	0.201	0.059	0.162	0.061	0.162	0.061	0.163	0.061
		2000	0.200	0.044	0.159	0.054	0.159	0.054	0.159	0.054
γ_4	0.700	500	0.714	0.094	0.567	0.153	0.565	0.156	0.569	0.152
		1000	0.703	0.067	0.557	0.153	0.556	0.154	0.559	0.151
		2000	0.702	0.045	0.558	0.147	0.558	0.147	0.560	0.145
β_1	1.000	500	1.005	0.116	0.396	0.618	0.092	0.974	0.396	0.618
		1000	1.002	0.084	0.395	0.612	0.174	0.866	0.395	0.612
		2000	1.003	0.057	0.394	0.609	0.190	0.828	0.394	0.609
β_2	0.700	500	0.698	0.075	0.698	0.084	0.698	0.084	0.698	0.085
		1000	0.698	0.056	0.697	0.063	0.697	0.063	0.697	0.064
		2000	0.698	0.038	0.699	0.045	0.699	0.045	0.699	0.045
β_3	1.100	500	1.099	0.056	1.100	0.062	1.100	0.062	1.100	0.062
		1000	1.100	0.038	1.101	0.041	1.101	0.041	1.101	0.041
		2000	1.100	0.031	1.101	0.033	1.101	0.033	1.101	0.034
ϕ	1.200	500	1.199	0.098	1.140	0.072	1.213	0.128	1.136	0.075
		1000	1.199	0.066	1.137	0.068	1.180	0.079	1.134	0.072
		2000	1.200	0.044	1.135	0.067	1.162	0.059	1.132	0.070
ρ	0.000	500	0.000	0.205	0.000	0.227	0.003	0.243	0.000	0.231
		1000	-0.006	0.159	-0.007	0.178	-0.007	0.185	-0.007	0.180
		2000	-0.003	0.106	-0.001	0.120	-0.000	0.123	-0.001	0.121

Tabela 3.5: Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas após ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 1 e $\rho = 0.2$. Tamanhos amostrais $n = 500$, $n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.

Parâmetros	n	Heckman BS		Heckman Clássico		Heckman Skew		Heckman t-Student		
		Média	REQM	Média	REQM	Média	REQM	Média	REQM	
γ_1	1.600	500	1.611	0.084	0.720	0.883	0.683	0.923	0.722	0.880
		1000	1.605	0.060	0.717	0.884	0.697	0.905	0.720	0.881
		2000	1.602	0.043	0.714	0.887	0.694	0.907	0.716	0.885
γ_2	0.800	500	0.814	0.098	0.646	0.172	0.644	0.174	0.649	0.170
		1000	0.806	0.066	0.642	0.167	0.642	0.167	0.645	0.165
		2000	0.804	0.046	0.640	0.164	0.640	0.164	0.642	0.162
γ_3	0.200	500	0.205	0.081	0.163	0.075	0.163	0.075	0.164	0.075
		1000	0.200	0.059	0.162	0.062	0.162	0.062	0.163	0.061
		2000	0.201	0.043	0.159	0.054	0.159	0.054	0.160	0.054
γ_4	0.700	500	0.713	0.094	0.567	0.154	0.565	0.156	0.569	0.152
		1000	0.703	0.067	0.557	0.153	0.557	0.153	0.559	0.151
		2000	0.702	0.045	0.559	0.146	0.558	0.146	0.560	0.145
β_1	1.000	500	1.007	0.101	0.404	0.610	0.200	0.852	0.404	0.611
		1000	1.004	0.072	0.402	0.605	0.277	0.750	0.402	0.605
		2000	1.002	0.050	0.398	0.605	0.279	0.729	0.398	0.605
β_2	0.700	500	0.696	0.074	0.693	0.083	0.693	0.083	0.693	0.084
		1000	0.697	0.054	0.693	0.062	0.693	0.062	0.693	0.062
		2000	0.698	0.038	0.697	0.044	0.697	0.044	0.697	0.044
β_3	1.100	500	1.099	0.057	1.099	0.062	1.099	0.062	1.099	0.063
		1000	1.100	0.038	1.099	0.041	1.099	0.041	1.099	0.041
		2000	1.100	0.031	1.100	0.033	1.100	0.033	1.100	0.033
ϕ	1.200	500	1.202	0.104	1.139	0.075	1.178	0.099	1.135	0.078
		1000	1.201	0.070	1.136	0.070	1.154	0.069	1.132	0.073
		2000	1.201	0.047	1.135	0.068	1.144	0.062	1.132	0.071
ρ	0.200	500	0.192	0.197	0.182	0.219	0.187	0.225	0.181	0.224
		1000	0.188	0.152	0.178	0.172	0.180	0.174	0.178	0.173
		2000	0.195	0.102	0.191	0.112	0.192	0.113	0.191	0.113

Tabela 3.6: Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas após ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 1 e $\rho = 0.5$. Tamanhos amostrais $n = 500$, $n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.

Parâmetros	n	Heckman		Heckman		Heckman		Heckman		
		BS		Clássico		Skew		t-Student		
		Média	REQM	Média	REQM	Média	REQM	Média	REQM	
γ_1	1.600	500	1.612	0.084	0.721	0.882	0.676	0.929	0.724	0.878
		1000	1.605	0.061	0.717	0.884	0.697	0.905	0.720	0.882
		2000	1.602	0.042	0.713	0.887	0.683	0.918	0.715	0.885
γ_2	0.800	500	0.815	0.096	0.648	0.171	0.646	0.172	0.651	0.168
		1000	0.806	0.066	0.643	0.166	0.642	0.167	0.645	0.164
		2000	0.804	0.045	0.641	0.164	0.640	0.164	0.643	0.162
γ_3	0.200	500	0.208	0.081	0.166	0.073	0.165	0.074	0.166	0.073
		1000	0.199	0.058	0.161	0.061	0.161	0.061	0.162	0.061
		2000	0.200	0.042	0.159	0.053	0.159	0.054	0.159	0.053
γ_4	0.700	500	0.712	0.093	0.569	0.151	0.567	0.153	0.572	0.149
		1000	0.704	0.068	0.560	0.150	0.560	0.150	0.563	0.148
		2000	0.703	0.045	0.561	0.144	0.561	0.144	0.563	0.142
β_1	1.000	500	1.005	0.079	0.416	0.595	0.308	0.720	0.416	0.595
		1000	1.002	0.056	0.413	0.591	0.363	0.648	0.413	0.592
		2000	1.002	0.040	0.410	0.593	0.338	0.666	0.410	0.593
β_2	0.700	500	0.696	0.070	0.686	0.075	0.685	0.075	0.685	0.076
		1000	0.697	0.050	0.686	0.055	0.686	0.055	0.686	0.055
		2000	0.698	0.036	0.690	0.040	0.689	0.040	0.689	0.040
β_3	1.100	500	1.099	0.056	1.097	0.060	1.097	0.060	1.097	0.061
		1000	1.099	0.037	1.097	0.040	1.097	0.040	1.097	0.040
		2000	1.100	0.030	1.097	0.032	1.097	0.032	1.097	0.032
ϕ	1.200	500	1.210	0.130	1.132	0.084	1.146	0.089	1.127	0.088
		1000	1.209	0.090	1.131	0.078	1.135	0.076	1.126	0.082
		2000	1.205	0.059	1.132	0.072	1.135	0.070	1.129	0.075
ρ	0.500	500	0.490	0.158	0.461	0.165	0.464	0.165	0.461	0.166
		1000	0.488	0.123	0.458	0.131	0.460	0.130	0.459	0.131
		2000	0.494	0.080	0.468	0.087	0.469	0.086	0.469	0.087

Tabela 3.7: Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas após ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 1 e $\rho = 0.7$. Tamanhos amostrais $n = 500$, $n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.

Parâmetros	n	Heckman BS		Heckman Clássico		Heckman Skew		Heckman t-Student		
		Média	REQM	Média	REQM	Média	REQM	Média	REQM	
γ_1	1.600	500	1.610	0.084	0.720	0.883	0.674	0.932	0.725	0.878
		1000	1.605	0.060	0.717	0.884	0.703	0.900	0.721	0.880
		2000	1.602	0.042	0.714	0.887	0.673	0.928	0.717	0.884
γ_2	0.800	500	0.814	0.095	0.649	0.169	0.646	0.172	0.653	0.166
		1000	0.806	0.064	0.644	0.165	0.642	0.167	0.647	0.162
		2000	0.805	0.045	0.643	0.161	0.643	0.162	0.646	0.158
γ_3	0.200	500	0.208	0.080	0.166	0.073	0.165	0.073	0.167	0.073
		1000	0.199	0.057	0.161	0.061	0.160	0.061	0.162	0.061
		2000	0.200	0.042	0.160	0.053	0.159	0.053	0.160	0.053
γ_4	0.700	500	0.714	0.089	0.575	0.144	0.573	0.146	0.579	0.141
		1000	0.704	0.065	0.565	0.144	0.564	0.146	0.568	0.142
		2000	0.703	0.043	0.565	0.139	0.565	0.140	0.568	0.137
β_1	1.000	500	1.005	0.068	0.427	0.581	0.350	0.672	0.427	0.581
		1000	1.001	0.046	0.420	0.584	0.398	0.616	0.420	0.584
		2000	1.000	0.033	0.417	0.585	0.348	0.655	0.417	0.585
β_2	0.700	500	0.694	0.067	0.678	0.072	0.678	0.072	0.678	0.073
		1000	0.697	0.046	0.681	0.051	0.682	0.051	0.681	0.052
		2000	0.699	0.033	0.685	0.038	0.685	0.038	0.685	0.038
β_3	1.100	500	1.100	0.056	1.096	0.058	1.096	0.058	1.095	0.058
		1000	1.099	0.036	1.096	0.039	1.096	0.039	1.096	0.039
		2000	1.100	0.029	1.096	0.031	1.096	0.031	1.096	0.031
ϕ	1.200	500	1.224	0.144	1.124	0.092	1.135	0.100	1.118	0.097
		1000	1.216	0.100	1.126	0.083	1.132	0.093	1.120	0.088
		2000	1.207	0.064	1.129	0.075	1.131	0.073	1.124	0.080
ρ	0.700	500	0.687	0.116	0.650	0.125	0.653	0.125	0.651	0.125
		1000	0.690	0.083	0.654	0.096	0.655	0.097	0.655	0.096
		2000	0.696	0.053	0.661	0.067	0.662	0.067	0.662	0.067

Apresentamos também alguns histogramas, Figuras 3.1, 3.2 e 3.3, referentes ao ajuste do modelo Heckman-BS conforme resultados apresentados na Tabela 3.7 e os box-plots, figuras 3.4, 3.5 e 3.6, de todos os modelos da Tabela 3.7. Os histogramas e os box-plots referentes as demais tabelas e demais parâmetros seguem o mesmo padrão e por isso não serão apresentados aqui. Note que em todos os histogramas, conforme aumentamos o tamanho amostral, o intervalo de variação do eixo das abcissas diminui sempre em torno do verdadeiro valor do parâmetro, indicando assim a convergência assintótica dos EMV's. Além disso, ao analisar os box-plots confirmamos a eficiência dos estimadores de máxima verossimilhança do modelo Heckman-BS. Quando comparamos com os demais modelos, observamos que todos os outros sobre-estimam os parâmetros.

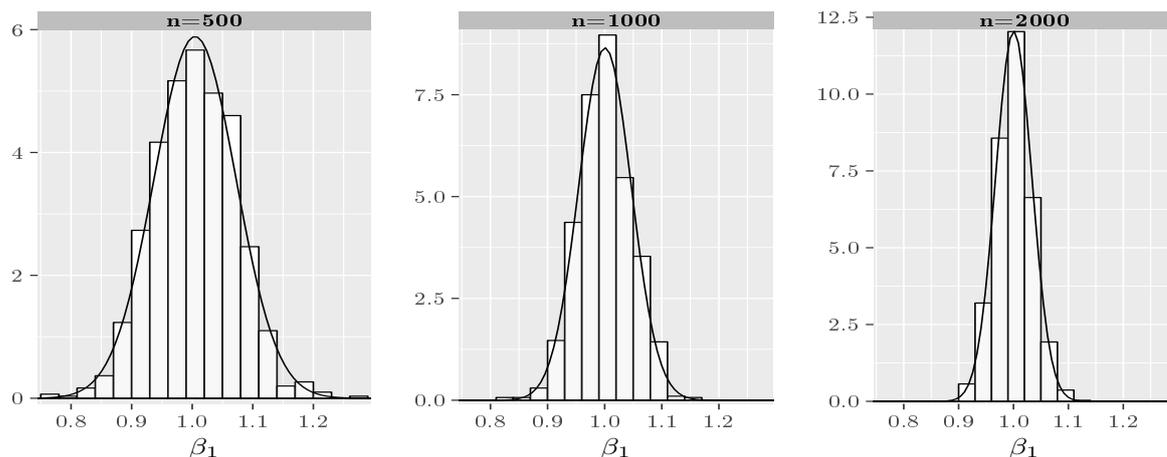


Figura 3.1: Histograma das estimativas de máxima verossimilhança do parâmetro β_1 , conforme Tabela 3.7, para os três tamanhos amostrais utilizados na simulação $n = 500$, $n = 1000$ e $n = 2000$.

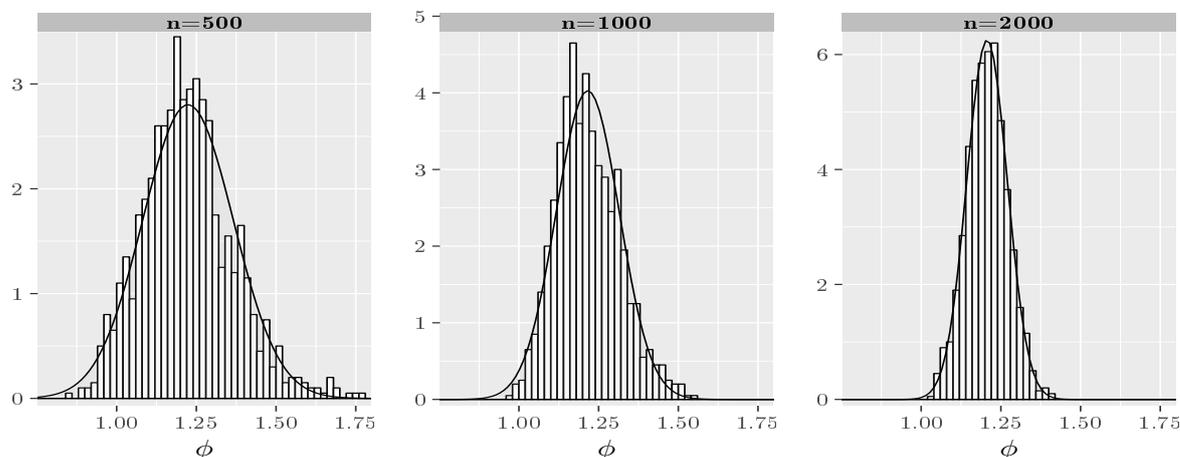


Figura 3.2: Histograma das estimativas de máxima verossimilhança do parâmetro ϕ , conforme Tabela 3.7, para os três tamanhos amostrais utilizados na simulação $n = 500$, $n = 1000$ e $n = 2000$.

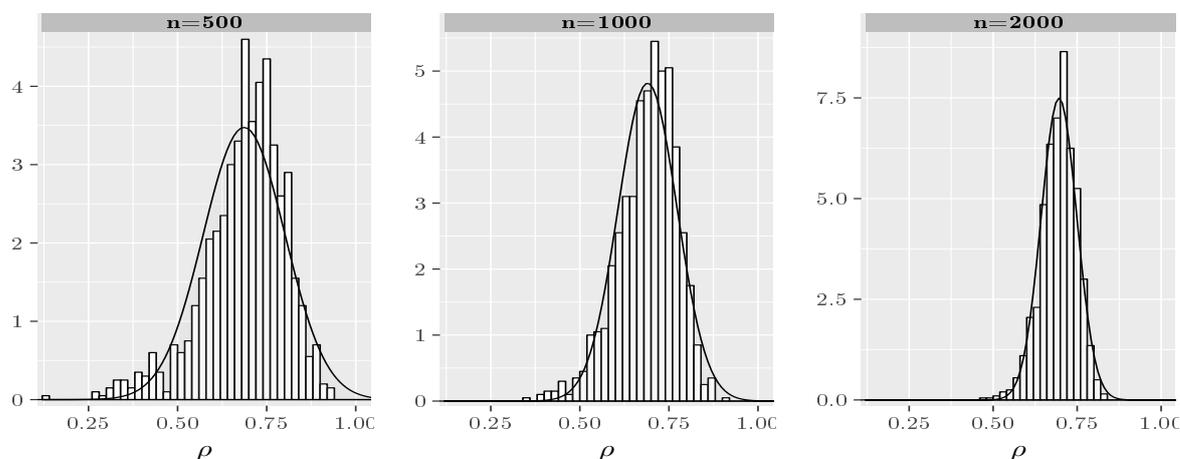


Figura 3.3: Histograma das estimativas de máxima verossimilhança do parâmetro ρ , conforme Tabela 3.7, para os três tamanhos amostrais utilizados na simulação, $n = 500$, $n = 1000$ e $n = 2000$.

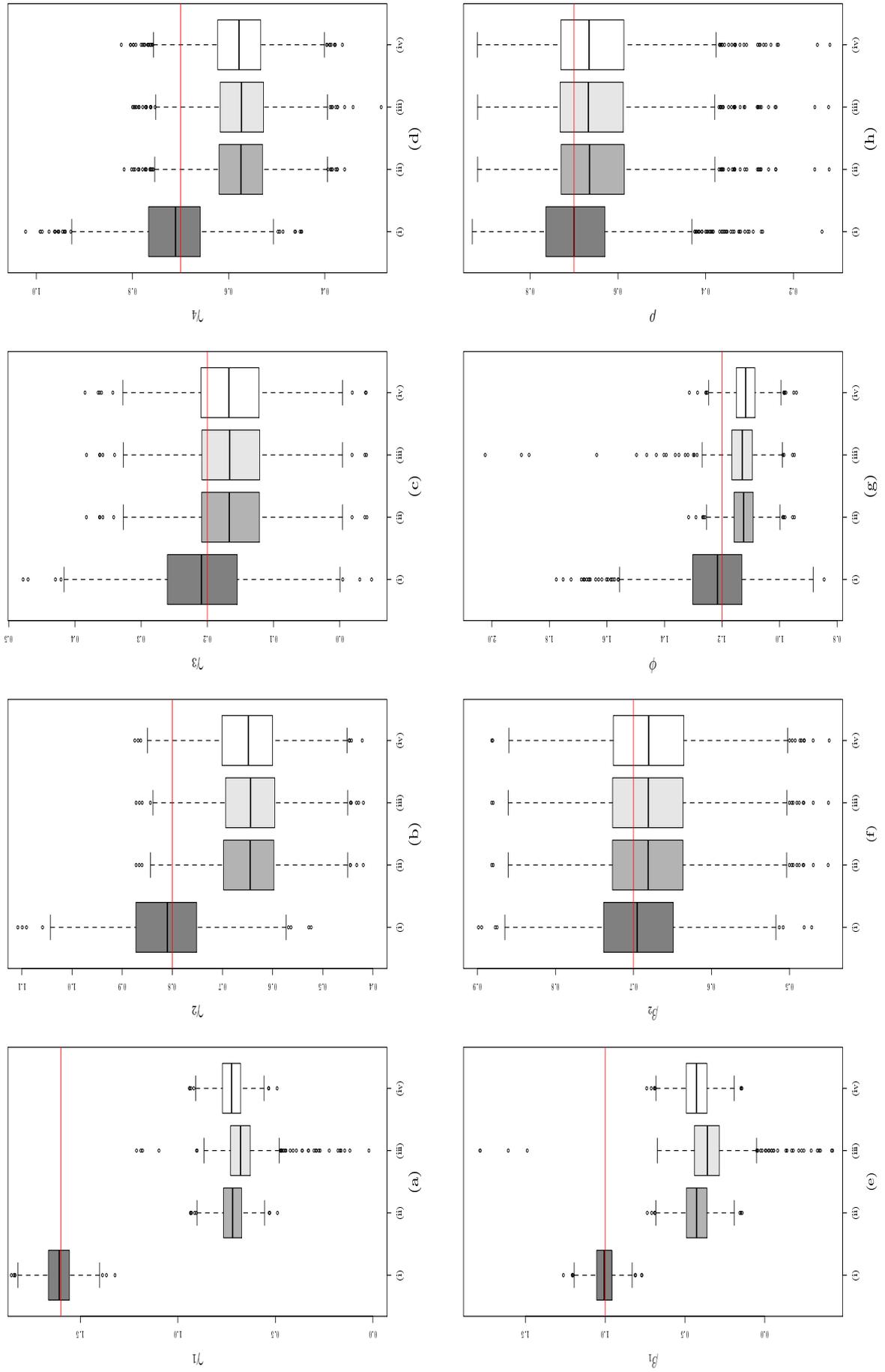


Figura 3.4: Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) ϕ , (f) β_1 , (g) β_2 , (h) d obtidas a partir do ajuste dos modelos (i) Heckman BS, (ii) Heckman Clássico, (iii) Skew-normal e (iv) Heckman-t. Tamanho amostral $n = 500$.

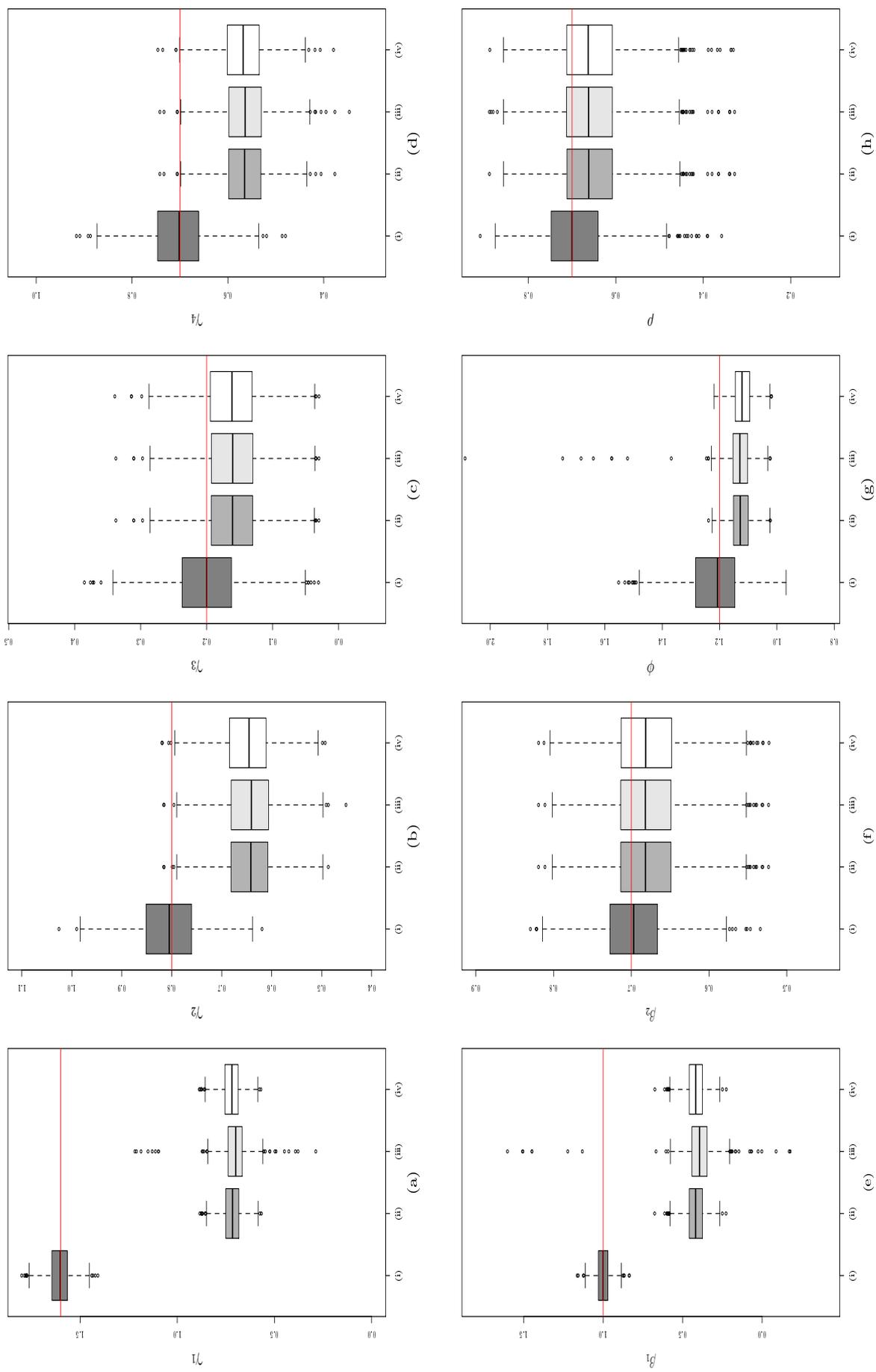


Figura 3.5: Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) ϕ , (f) β_1 , (g) ϕ e (h) ρ obtidas a partir do ajuste dos modelos (i) Heckman BS, (ii) Skew-normal e (iv) Heckman-t. Tamanho amostral $n = 1000$.

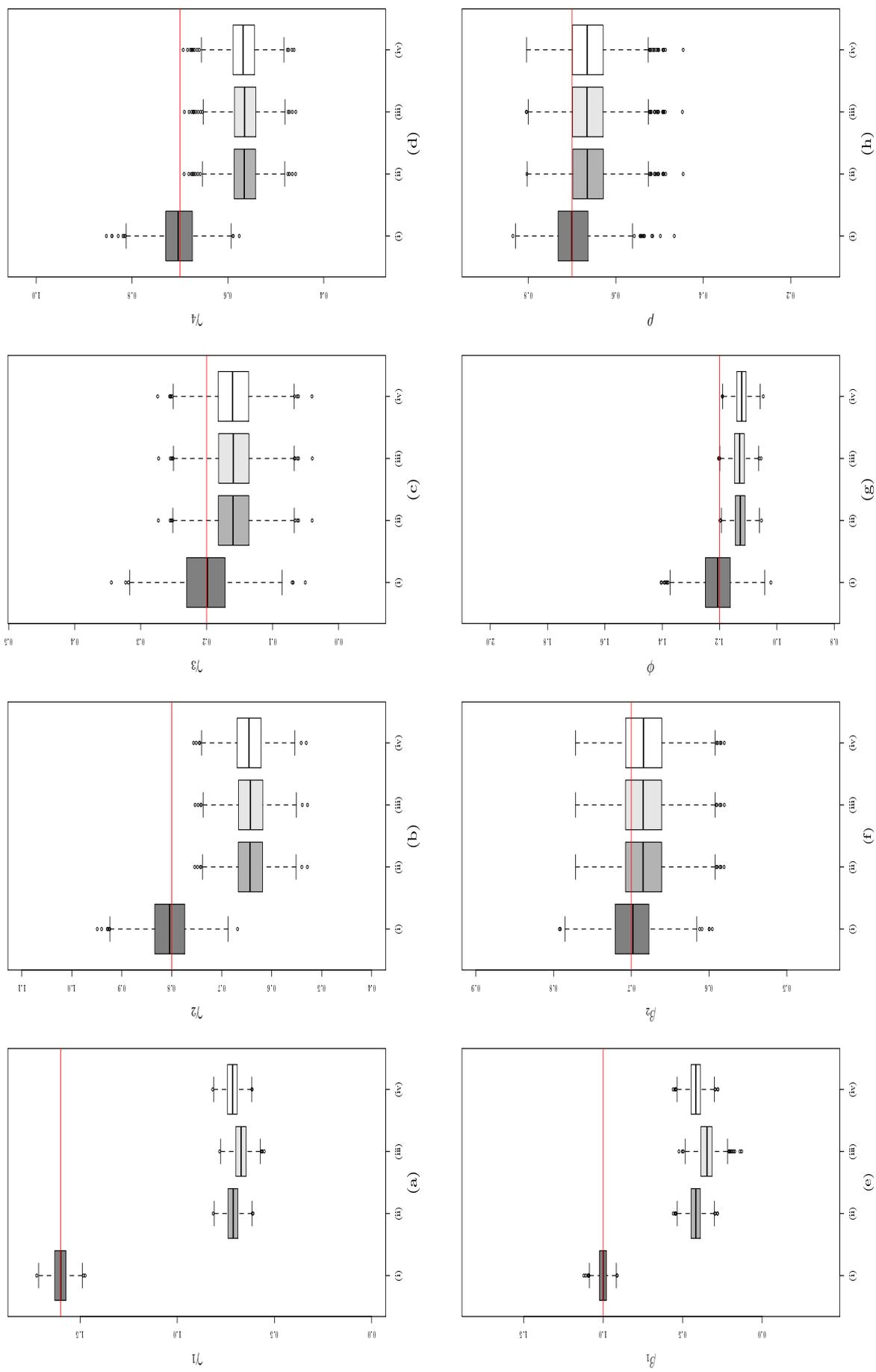


Figura 3.6: Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) γ_4 , (e) β_1 , (f) β_2 , (g) ϕ e (h) ρ obtidas a partir do ajuste dos modelos (i) Heckman BS, (ii) Skew-normal e (iv) Heckman-t. Tamanho amostral $n = 2000$.

3.3.2 Cenário 2

Realizamos simulações Monte Carlo com o objetivo de avaliar e comparar o desempenho dos EMV's dos parâmetros do modelo Heckman-BS, Heckman clássico, Heckman-t e Skew-normal, na ausência de restrição de exclusão, com parâmetros de média descrito em (3.13). Mantendo fixos os coeficientes de regressão que são tomados como $\beta_1 = 1, \beta_2 = 0.7, \beta_3 = 1.1, \gamma_1 = 1.6, \gamma_2 = 0.8, \gamma_3 = 0.2, \phi = 1.2$ e alternando os valores de ρ . Definimos μ_i conforme (3.14) para $i = 1, \dots, n$.

Tabela 3.8: Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 2 com $\rho = -0.7$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.

Parâmetros	n	Heckman BS		Heckman Clássico		Heckman Skew		Heckman t-Student		
		Média	REQM	Média	REQM	Média	REQM	Média	REQM	
γ_1	1.600	500	1.604	0.082	0.708	0.894	0.955	0.683	0.716	0.886
		1000	1.608	0.058	0.708	0.894	0.975	0.656	0.713	0.888
		2000	1.601	0.039	0.705	0.896	1.059	0.554	0.710	0.891
γ_2	0.800	500	0.814	0.097	0.652	0.166	0.579	0.245	0.655	0.165
		1000	0.796	0.067	0.643	0.165	0.573	0.239	0.645	0.163
		2000	0.799	0.045	0.639	0.165	0.557	0.250	0.639	0.165
γ_3	0.200	500	0.206	0.078	0.163	0.073	0.145	0.081	0.164	0.073
		1000	0.200	0.057	0.161	0.059	0.144	0.072	0.162	0.059
		2000	0.200	0.041	0.160	0.053	0.139	0.069	0.160	0.053
β_1	1.000	500	0.818	0.371	0.160	0.896	-0.508	1.562	0.066	1.008
		1000	0.710	0.458	0.218	0.824	-0.456	1.501	0.143	0.917
		2000	0.944	0.203	0.289	0.732	-0.532	1.549	0.143	0.915
β_2	0.700	500	0.794	0.204	0.835	0.225	0.792	0.210	0.889	0.285
		1000	0.862	0.264	0.802	0.182	0.779	0.186	0.846	0.236
		2000	0.729	0.114	0.764	0.120	0.725	0.102	0.847	0.233
β_3	1.100	500	1.124	0.075	1.134	0.079	1.123	0.076	1.147	0.090
		1000	1.142	0.076	1.128	0.060	1.122	0.060	1.139	0.071
		2000	1.108	0.041	1.116	0.042	1.107	0.040	1.137	0.064
ϕ	1.200	500	1.240	0.192	1.113	0.111	1.493	0.494	1.119	0.106
		1000	1.205	0.149	1.111	0.102	1.457	0.408	1.113	0.100
		2000	1.208	0.091	1.110	0.097	1.517	0.393	1.113	0.094
ρ	-0.700	500	-0.359	0.669	-0.265	0.714	-0.431	0.664	-0.089	0.909
		1000	-0.146	0.885	-0.384	0.579	-0.494	0.597	-0.241	0.759
		2000	-0.596	0.375	-0.517	0.373	-0.699	0.324	-0.237	0.759

Tabela 3.9: Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 2 com $\rho = -0.5$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.

Parâmetros	n	Heckman BS		Heckman Clássico		Heckman Skew		Heckman t-Student		
		Média	REQM	Média	REQM	Média	REQM	Média	REQM	
γ_1	1.600	500	1.605	0.080	0.708	0.895	0.823	0.815	0.714	0.888
		1000	1.606	0.059	0.706	0.895	0.848	0.783	0.711	0.891
		2000	1.601	0.039	0.703	0.897	0.884	0.734	0.707	0.893
γ_2	0.800	500	0.812	0.096	0.647	0.171	0.609	0.214	0.648	0.171
		1000	0.801	0.068	0.640	0.169	0.608	0.202	0.641	0.168
		2000	0.802	0.042	0.636	0.168	0.609	0.197	0.636	0.168
γ_3	0.200	500	0.205	0.080	0.162	0.075	0.152	0.079	0.162	0.076
		1000	0.202	0.057	0.161	0.060	0.153	0.065	0.162	0.060
		2000	0.201	0.043	0.159	0.054	0.152	0.059	0.159	0.054
β_1	1.000	500	0.917	0.265	0.177	0.873	-0.401	1.466	0.077	0.985
		1000	0.897	0.250	0.220	0.818	-0.355	1.399	0.120	0.931
		2000	0.992	0.111	0.268	0.758	-0.329	1.351	0.146	0.899
β_2	0.700	500	0.748	0.149	0.823	0.207	0.792	0.194	0.879	0.264
		1000	0.761	0.149	0.798	0.174	0.783	0.173	0.855	0.232
		2000	0.704	0.053	0.773	0.134	0.759	0.137	0.842	0.212
β_3	1.100	500	1.111	0.065	1.130	0.077	1.122	0.075	1.144	0.089
		1000	1.116	0.051	1.126	0.058	1.122	0.058	1.140	0.070
		2000	1.101	0.032	1.119	0.045	1.115	0.046	1.136	0.061
ϕ	1.200	500	1.201	0.161	1.142	0.093	1.409	0.391	1.152	0.090
		1000	1.194	0.140	1.133	0.086	1.361	0.289	1.141	0.082
		2000	1.204	0.084	1.125	0.083	1.340	0.228	1.133	0.080
ρ	-0.500	500	-0.307	0.467	-0.092	0.649	-0.206	0.633	0.085	0.823
		1000	-0.286	0.489	-0.180	0.548	-0.260	0.564	0.003	0.738
		2000	-0.477	0.155	-0.264	0.429	-0.349	0.447	-0.040	0.682

Tabela 3.10: Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 2 com $\rho = -0.2$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.

Parâmetros	n	Heckman BS		Heckman Clássico		Heckman Skew		Heckman t-Student		
		Média	REQM	Média	REQM	Média	REQM	Média	REQM	
γ_1	1.600	500	1.607	0.078	0.708	0.895	0.730	0.895	0.713	0.889
		1000	1.605	0.059	0.706	0.896	0.742	0.872	0.709	0.892
		2000	1.602	0.040	0.703	0.897	0.735	0.872	0.707	0.894
γ_2	0.800	500	0.809	0.095	0.642	0.175	0.625	0.196	0.645	0.173
		1000	0.805	0.064	0.640	0.168	0.630	0.179	0.642	0.166
		2000	0.803	0.043	0.636	0.168	0.632	0.172	0.639	0.165
γ_3	0.200	500	0.204	0.079	0.161	0.075	0.156	0.077	0.161	0.075
		1000	0.204	0.057	0.162	0.060	0.160	0.062	0.162	0.060
		2000	0.201	0.043	0.159	0.054	0.158	0.054	0.159	0.053
β_1	1.000	500	0.993	0.207	0.288	0.754	-0.141	1.226	0.255	0.799
		1000	0.989	0.156	0.315	0.715	-0.064	1.120	0.289	0.749
		2000	1.008	0.105	0.329	0.692	-0.021	1.055	0.302	0.723
β_2	0.700	500	0.715	0.118	0.759	0.157	0.748	0.159	0.778	0.182
		1000	0.713	0.091	0.743	0.126	0.737	0.127	0.758	0.145
		2000	0.698	0.055	0.737	0.104	0.736	0.107	0.752	0.120
β_3	1.100	500	1.103	0.062	1.115	0.070	1.112	0.070	1.119	0.075
		1000	1.104	0.044	1.112	0.051	1.111	0.051	1.116	0.055
		2000	1.100	0.032	1.110	0.040	1.110	0.041	1.113	0.043
ϕ	1.200	500	1.158	0.135	1.167	0.074	1.324	0.284	1.170	0.075
		1000	1.175	0.101	1.156	0.063	1.265	0.171	1.157	0.065
		2000	1.189	0.059	1.147	0.061	1.223	0.093	1.146	0.063
ρ	-0.200	500	-0.131	0.363	-0.000	0.464	-0.035	0.498	0.057	0.533
		1000	-0.153	0.289	-0.055	0.391	-0.082	0.414	-0.008	0.446
		2000	-0.201	0.171	-0.074	0.325	-0.084	0.346	-0.027	0.377

Tabela 3.11: Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 2 com $\rho = 0$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.

Parâmetros	n	Heckman BS		Heckman Clássico		Heckman Skew		Heckman t-Student		
		Média	REQM	Média	REQM	Média	REQM	Média	REQM	
γ_1	1.600	500	1.608	0.077	0.708	0.894	0.727	0.891	0.714	0.889
		1000	1.605	0.060	0.705	0.896	0.721	0.888	0.709	0.893
		2000	1.602	0.040	0.703	0.897	0.706	0.897	0.706	0.894
γ_2	0.800	500	0.810	0.093	0.644	0.174	0.631	0.189	0.647	0.171
		1000	0.806	0.063	0.640	0.169	0.635	0.174	0.643	0.166
		2000	0.803	0.042	0.636	0.167	0.635	0.169	0.639	0.165
γ_3	0.200	500	0.202	0.079	0.160	0.076	0.157	0.077	0.161	0.076
		1000	0.202	0.058	0.161	0.061	0.160	0.062	0.162	0.061
		2000	0.202	0.043	0.160	0.053	0.159	0.053	0.160	0.053
β_1	1.000	500	1.029	0.197	0.388	0.657	0.084	1.005	0.393	0.666
		1000	1.005	0.135	0.393	0.638	0.167	0.890	0.390	0.648
		2000	1.009	0.097	0.394	0.627	0.205	0.828	0.391	0.632
β_2	0.700	500	0.696	0.113	0.703	0.140	0.692	0.147	0.700	0.157
		1000	0.703	0.083	0.699	0.115	0.691	0.117	0.701	0.124
		2000	0.697	0.058	0.700	0.092	0.701	0.094	0.701	0.099
β_3	1.100	500	1.099	0.062	1.101	0.070	1.098	0.070	1.100	0.072
		1000	1.101	0.043	1.101	0.048	1.099	0.048	1.101	0.050
		2000	1.100	0.032	1.101	0.038	1.101	0.039	1.101	0.039
ϕ	1.200	500	1.151	0.132	1.171	0.068	1.278	0.235	1.172	0.069
		1000	1.172	0.090	1.161	0.058	1.217	0.119	1.160	0.060
		2000	1.187	0.052	1.152	0.056	1.180	0.060	1.149	0.058
ρ	0.000	500	-0.007	0.345	0.013	0.401	-0.024	0.443	0.004	0.443
		1000	0.009	0.262	-0.004	0.346	-0.033	0.367	0.002	0.373
		2000	-0.006	0.183	0.001	0.283	0.004	0.293	0.005	0.304

Tabela 3.12: Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 2 com $\rho = 0.2$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.

Parâmetros	n	Heckman BS		Heckman Clássico		Heckman Skew		Heckman t-Student		
		Média	REQM	Média	REQM	Média	REQM	Média	REQM	
γ_1	1.600	500	1.607	0.079	0.708	0.895	0.731	0.883	0.713	0.889
		1000	1.605	0.060	0.706	0.896	0.719	0.886	0.709	0.892
		2000	1.602	0.040	0.703	0.897	0.700	0.902	0.706	0.894
γ_2	0.800	500	0.812	0.095	0.644	0.174	0.634	0.187	0.648	0.171
		1000	0.806	0.063	0.640	0.168	0.636	0.173	0.642	0.166
		2000	0.803	0.043	0.636	0.167	0.636	0.168	0.639	0.165
γ_3	0.200	500	0.205	0.082	0.162	0.077	0.159	0.078	0.162	0.077
		1000	0.202	0.057	0.160	0.061	0.159	0.062	0.161	0.061
		2000	0.202	0.042	0.160	0.053	0.160	0.053	0.160	0.052
β_1	1.000	500	1.053	0.206	0.490	0.571	0.277	0.806	0.526	0.555
		1000	1.019	0.129	0.473	0.565	0.355	0.696	0.500	0.550
		2000	1.012	0.085	0.453	0.571	0.360	0.667	0.476	0.557
β_2	0.700	500	0.681	0.119	0.647	0.158	0.638	0.164	0.627	0.181
		1000	0.694	0.083	0.654	0.127	0.641	0.136	0.639	0.145
		2000	0.694	0.057	0.665	0.102	0.667	0.102	0.653	0.118
β_3	1.100	500	1.096	0.063	1.087	0.072	1.084	0.074	1.082	0.076
		1000	1.099	0.042	1.090	0.050	1.087	0.051	1.086	0.053
		2000	1.100	0.031	1.093	0.039	1.093	0.039	1.089	0.042
ϕ	1.200	500	1.160	0.136	1.170	0.074	1.241	0.214	1.171	0.074
		1000	1.180	0.092	1.155	0.063	1.186	0.099	1.156	0.065
		2000	1.191	0.058	1.147	0.061	1.156	0.061	1.146	0.063
ρ	0.200	500	0.131	0.359	0.023	0.465	-0.005	0.500	-0.038	0.530
		1000	0.171	0.258	0.046	0.389	0.001	0.430	-0.002	0.445
		2000	0.182	0.182	0.086	0.316	0.091	0.318	0.046	0.366

Tabela 3.13: Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 2 com $\rho = 0.5$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.

Parâmetros	n	Heckman BS		Heckman Clássico		Heckman Skew		Heckman t-Student		
		Média	REQM	Média	REQM	Média	REQM	Média	REQM	
γ_1	1.600	500	1.606	0.078	0.708	0.894	0.755	0.861	0.715	0.887
		1000	1.603	0.059	0.705	0.896	0.752	0.859	0.710	0.892
		2000	1.601	0.041	0.704	0.897	0.707	0.896	0.708	0.893
γ_2	0.800	500	0.814	0.095	0.648	0.170	0.633	0.190	0.650	0.169
		1000	0.807	0.061	0.641	0.167	0.628	0.183	0.641	0.167
		2000	0.803	0.043	0.637	0.166	0.636	0.168	0.638	0.166
γ_3	0.200	500	0.205	0.080	0.162	0.075	0.158	0.077	0.163	0.075
		1000	0.200	0.057	0.159	0.062	0.156	0.064	0.159	0.062
		2000	0.202	0.041	0.159	0.053	0.159	0.053	0.160	0.053
β_1	1.000	500	1.103	0.262	0.608	0.487	0.449	0.641	0.705	0.448
		1000	1.027	0.124	0.559	0.502	0.504	0.562	0.665	0.452
		2000	1.022	0.112	0.512	0.522	0.460	0.566	0.634	0.459
β_2	0.700	500	0.644	0.150	0.577	0.208	0.565	0.224	0.523	0.262
		1000	0.684	0.083	0.604	0.169	0.560	0.216	0.544	0.233
		2000	0.686	0.069	0.631	0.129	0.632	0.129	0.562	0.211
β_3	1.100	500	1.087	0.065	1.069	0.078	1.065	0.081	1.055	0.089
		1000	1.096	0.041	1.077	0.055	1.066	0.064	1.063	0.068
		2000	1.097	0.032	1.084	0.044	1.084	0.044	1.066	0.061
ϕ	1.200	500	1.205	0.160	1.142	0.093	1.218	0.224	1.150	0.090
		1000	1.209	0.122	1.130	0.087	1.185	0.164	1.140	0.084
		2000	1.206	0.087	1.124	0.084	1.134	0.095	1.132	0.080
ρ	0.500	500	0.300	0.469	0.103	0.640	0.063	0.697	-0.069	0.810
		1000	0.435	0.259	0.188	0.530	0.042	0.689	-0.004	0.736
		2000	0.452	0.216	0.280	0.405	0.283	0.407	0.054	0.670

Tabela 3.14: Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) obtidas a partir do ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados gerados conforme Cenário 2 com $\rho = 0.7$. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.

Parâmetros	n	Heckman BS		Heckman Clássico		Heckman Skew		Heckman t-Student		
		Média	REQM	Média	REQM	Média	REQM	Média	REQM	
γ_1	1.600	500	1.606	0.077	0.710	0.893	0.791	0.834	0.718	0.885
		1000	1.602	0.057	0.706	0.895	0.830	0.796	0.711	0.890
		2000	1.601	0.041	0.704	0.896	0.711	0.894	0.710	0.891
γ_2	0.800	500	0.811	0.091	0.650	0.167	0.621	0.205	0.652	0.166
		1000	0.807	0.062	0.644	0.164	0.602	0.215	0.646	0.162
		2000	0.801	0.043	0.640	0.163	0.635	0.171	0.641	0.163
γ_3	0.200	500	0.205	0.080	0.163	0.075	0.155	0.078	0.163	0.076
		1000	0.200	0.055	0.159	0.061	0.149	0.069	0.159	0.061
		2000	0.201	0.041	0.160	0.052	0.159	0.053	0.161	0.052
β_1	1.000	500	1.178	0.371	0.621	0.487	0.415	0.660	0.725	0.464
		1000	1.033	0.150	0.552	0.509	0.421	0.641	0.650	0.479
		2000	1.091	0.267	0.487	0.536	0.427	0.587	0.619	0.491
β_2	0.700	500	0.596	0.212	0.566	0.224	0.538	0.260	0.507	0.287
		1000	0.677	0.094	0.605	0.170	0.496	0.294	0.549	0.240
		2000	0.648	0.148	0.643	0.108	0.637	0.123	0.568	0.219
β_3	1.100	500	1.074	0.075	1.066	0.079	1.059	0.086	1.051	0.093
		1000	1.095	0.042	1.078	0.054	1.050	0.082	1.065	0.067
		2000	1.087	0.046	1.086	0.039	1.084	0.042	1.067	0.061
ϕ	1.200	500	1.239	0.188	1.115	0.110	1.247	0.296	1.121	0.104
		1000	1.232	0.140	1.108	0.105	1.273	0.293	1.111	0.102
		2000	1.210	0.096	1.109	0.097	1.134	0.141	1.111	0.095
ρ	0.700	500	0.348	0.678	0.280	0.701	0.184	0.828	0.085	0.913
		1000	0.618	0.294	0.412	0.533	0.043	0.950	0.227	0.768
		2000	0.526	0.488	0.541	0.327	0.523	0.382	0.290	0.707

Nossas simulações indicam que a ausência da restrição de exclusão e a presença de correlação moderada ou alta, afetam a estimação dos parâmetros para todos os modelos. Para correlação fraca ou nula, a simulação indica consistência somente do modelo Heckman-BS na estimação de todos os parâmetros. Os demais modelos apresentaram inconsistências na estimação, pois conforme aumentamos o tamanho amostral a raiz quadrada do REQM não necessariamente diminuiu. Tais resultados indicam a necessidade e a importância da restrição de exclusão no ajuste dos modelos de seleção amostral.

3.4 Tamanho e Poder dos Testes da Razão de Verossimilhança e Gradiente

Com o objetivo de avaliar e comparar o desempenho dos testes da razão de verossimilhanças e gradiente para testar as hipóteses $H_0 : \rho = 0$ contra $H_1 : \rho \neq 0$, sob o ajuste dos modelos Heckman-BS, Heckman clássico e Skew-normal, todos ajustados aos dados simulados de acordo com o Cenário 1, apresentamos os resultados do estudo de simulação do poder e do nível de significância empírico destes testes. Não mostraremos os resultados dos testes sob o ajuste do modelo Heckman-t, uma vez que observamos

resultados que contrariam a nossa expectativa teórica e que devem ser melhor investigados. Considere as hipóteses definidas anteriormente, os níveis de significância nominais $\alpha = 1\%$, $\alpha = 5\%$ e $\alpha = 10\%$, tamanhos amostrais $n = 500$, $n = 1000$ e $n = 2000$ e $N = 1000$ réplicas Monte Carlo.

Para cada amostra simulada ajustamos os modelos Heckman-BS, Heckman clássico e Skew-normal, e encontramos as estimativas de máxima verossimilhança sob a restrição imposta pela hipótese nula e irrestrita. Encontramos então as estatísticas dos testes e comparamos com os respectivos quantis da distribuição qui-quadrado com 1 grau de liberdade. Estimamos a proporção de vezes em que a hipótese H_0 foi rejeitada nas $N = 1000$ réplicas Monte Carlo. Fixado um cenário específico e um tamanho amostral n , os passos seguidos para o cálculo do nível de significância empírico dos testes e poder são detalhados a seguir:

1. Geramos uma amostra de tamanho n de acordo com o Cenário 1 com o parâmetro ρ fixo;
2. Calculamos os estimadores de máxima verossimilhança para a amostra dada sob a hipótese nula, e depois sem a restrição da hipótese nula;
3. Construimos as estatísticas S_{RV} e S_G ;
4. Comparamos S_{RV} e S_G com os quantis da distribuição qui-quadrado com 1 grau de liberdade e rejeitamos a hipótese nula para cada teste respectivo se $S_{RV} > \chi_1^2(1 - \alpha)$ e $S_G > \chi_1^2(1 - \alpha)$;
5. Repetimos os Passos de 1 a 4 N vezes (aqui consideramos $N=1.000$);
6. Calculamos a proporção de rejeições obtidas sob H_0 (**Nível de significância empírico do teste**) e o poder do teste (probabilidade de rejeitar sob H_1);

A Tabela 3.15 mostra o nível de significância empírico e a taxa de rejeição nula dos testes da razão de verossimilhança e do gradiente, isto é, a taxa de rejeição (em porcentagem) dos testes baseados nas estatísticas S_{RV} e S_G para avaliar a hipótese nula $H_0 : \rho = 0$ contra $H_1 : \rho \neq 0$. Notamos que ambos os testes mantém, aproximadamente, o nível nominal correto quando os dados são simulados de acordo com o Cenário 1, descrito na seção anterior, com os modelos Heckman-BS, Heckman clássico e Skew-normal ajustados aos dados.

Tabela 3.15: Nível de significância empírico dos testes da razão de verossimilhança e gradiente para $H_0 : \rho = 0$ contra $H_1 : \rho \neq 0$ e censura média de 30%.

Valor nominal do teste	Tamanho da Amostra	Heckman BS		Heckman Clássico		Heckman Skew	
		S_{RV}	S_G	S_{RV}	S_G	S_{RV}	S_G
1%	$n = 500$	0.7	0.3	1.1	0.5	1.2	0.4
	$n = 1000$	1.3	0.8	1.3	0.5	1.3	0.5
	$n = 2000$	1.3	1.3	1.1	0.9	1.1	0.9
5%	$n = 500$	4.7	4.1	4.4	3.2	4.4	3
	$n = 1000$	6.7	6.1	5.3	4.6	5.5	4.5
	$n = 2000$	5	4.9	5	4.6	5	4.7
10%	$n = 500$	9.8	8.4	8.9	7.4	9.1	7.3
	$n = 1000$	11.2	10.6	10.7	9.6	10.8	9.7
	$n = 2000$	9.9	9.5	9.9	9.8	9.8	9.7

Na Tabela 3.16 apresentamos o resultado das simulações de poder dos testes RV e gradiente. Observamos que a medida que ρ se afasta de 0, e aumentamos o tamanho amostral, o poder dos testes aumentam. Notamos ainda que, em geral, o teste gradiente é menos poderoso que o teste da razão de verossimilhança sob o ajuste de todos os modelos.

Tabela 3.16: Poder (em porcentagem) dos testes da razão de verossimilhança e gradiente para $\alpha = 1\%$, $\alpha = 5\%$ e $\alpha = 10\%$.

Parâmetros	Tamanho da Amostra	Heckman BS		Heckman Clássico		Heckman Skew	
		S_{RV}	S_G	S_{RV}	S_G	S_{RV}	S_G
$\alpha = 1\%$							
$\rho = -0.7$	$n = 500$	90.6	84.9	84.1	73.4	86	75.8
	$n = 1000$	99.5	99.4	99.3	97.6	99.5	98.8
	$n = 2000$	100	100	100	100	100	100
$\rho = -0.3$	$n = 500$	11.4	7.8	8.5	4.1	9.1	4.6
	$n = 1000$	29.7	25.9	20.7	14.8	21.1	15.6
	$n = 2000$	62	60.1	46.8	42.9	47	43.5
$\rho = 0.3$	$n = 500$	14.2	8	8	4.8	8	4.8
	$n = 1000$	26.9	23.6	18.8	13.2	18.8	12.9
	$n = 2000$	61.8	60.4	47.2	44.3	47.1	44.3
$\rho = 0.7$	$n = 500$	90.3	85.4	84.9	73.6	84.9	72.4
	$n = 1000$	99.2	98.8	98.8	97.5	98.8	97.5
	$n = 2000$	100	100	100	100	100	100
$\alpha = 5\%$							
$\rho = -0.7$	$n = 500$	96.6	95.8	94.6	91.5	95.4	92.7
	$n = 1000$	99.9	99.9	100	99.9	100	99.9
	$n = 2000$	100	100	100	100	100	100
$\rho = -0.3$	$n = 500$	31.8	28.6	23	19	23.7	19.2
	$n = 1000$	54.5	52.9	42.7	40.2	43.1	40.8
	$n = 2000$	83	82.4	70.4	69.4	70.8	69.4
$\rho = 0.3$	$n = 500$	30.7	28.1	24.6	19.3	24.4	19.3
	$n = 1000$	50.2	48.4	39.3	36.8	39.3	36.5
	$n = 2000$	81.9	81.5	71.4	69.6	71.3	69.8
$\rho = 0.7$	$n = 500$	97.2	96.4	94.9	92.1	94.9	91.9
	$n = 1000$	99.8	99.8	99.6	99.5	99.6	99.5
	$n = 2000$	100	100	100	100	100	100
$\alpha = 10\%$							
$\rho = -0.7$	$n = 500$	98.6	97.9	97.2	95.9	97.5	96.7
	$n = 1000$	100	100	100	100	100	100
	$n = 2000$	100	100	100	100	100	100
$\rho = -0.3$	$n = 500$	45.5	42.8	34.8	31.9	35	31.9
	$n = 1000$	65.2	64	55.1	53.6	54.7	53.4
	$n = 2000$	90.8	90.1	81.1	79.3	80.9	79.3
$\rho = 0.3$	$n = 500$	43.6	41.9	37.7	33.8	37.4	33.6
	$n = 1000$	63.5	62.5	52	49.5	51.9	49
	$n = 2000$	89.8	89.7	81.5	80.7	81.4	80.7
$\rho = 0.7$	$n = 500$	98.2	98	97.8	96.7	97.8	96.6
	$n = 1000$	100	99.9	99.9	99.8	99.9	99.8
	$n = 2000$	100	100	100	100	100	100

Como esperado o poder dos testes RV e gradiente sob o ajuste do modelo Heckman-BS apresentaram poder superior. Nossas simulações também indicaram que mesmo na presença de assimetria o modelo Skew-normal não produziu testes com poder superior ao modelo de Heckman clássico.

3.5 Ajuste do Modelo Heckman-BS sob Falta de Especificação

Com o objetivo de comparar o ajuste dos modelos Heckman-BS, Heckman clássico, Skew-normal e Heckman-t sob a especificação dos dados com distribuição assimétrica e positiva, diferente da distribuição BS, simulamos dados com distribuição gama bivariada usando um algoritmo descrito no trabalho de [Schmeiser e Lal \(1982\)](#). O algoritmo é resultado de uma técnica conhecida como redução trivariada, que é utilizada para a construção de distribuições bivariadas a partir da combinação de três ou mais variáveis.

Em síntese, consideramos X_1, X_2 e X_3 variáveis aleatórias, não necessariamente independentes, com função de distribuição $F_i(x_i; \lambda_i), i = 1, \dots, 3$, e λ_i um parâmetro qualquer. Tomando

$$\begin{aligned} X &= T_1(X_1, X_2, X_3) \\ Y &= T_2(X_1, X_2, X_3), \end{aligned}$$

com $T_j : \mathbb{R}^3 \rightarrow \mathbb{R}, j = 1, 2$, dizemos que o vetor (X, Y) têm distribuição bivariada gerada por uma redução trivariada. As funções de distribuição de X_i são, em geral, da mesma família, porém isso não é condição necessária. Uma definição restrita é

$$\begin{aligned} X &= X_1 + X_3 \\ Y &= X_2 + X_3, \end{aligned}$$

em que $X_i, i = 1, \dots, 3$, são independentes. Nesse caso, X e Y são positivamente correlacionadas.

Antes de construir o algoritmo, considere que uma variável aleatória Y segue uma distribuição gama com parâmetros $\alpha, \beta > 0$, denotada por $Y \sim \Gamma(\alpha, \beta)$, se a sua densidade é dada por

$$f(y; \alpha, \beta) = \frac{y^{(\alpha-1)}}{\beta^\alpha \Gamma(\alpha)} \exp\left(-\frac{y}{\beta}\right), \text{ para } y > 0, \quad (3.16)$$

em que α e β são os parâmetros de forma e de escala, respectivamente. A média e a variância de Y são, respectivamente, dadas por

$$E(Y) = \alpha\beta \quad \text{e} \quad Var(Y) = \alpha\beta^2. \quad (3.17)$$

Reparametrizamos a distribuição (3.16) ao considerar α e $\beta = \frac{\mu}{\alpha}$, e denotamos $Y \sim \Gamma\left(\alpha, \frac{\mu}{\alpha}\right)$. Segue

que, sob esta nova parametrização

$$E(Y) = \mu \quad \text{e} \quad \text{Var}(Y) = \frac{\mu^2}{\alpha}. \quad (3.18)$$

Considerando então a distribuição gama univariada e a técnica de redução trivariada, podemos obter a distribuição gama bivariada de Cheria, que é uma das diferentes formas de parametrização da distribuição gama bivariada, detalhada no livro de [Balakrishnan e Lai \(2009\)](#).

Um vetor aleatório (Y_1, Y_2) tem distribuição gama bivariada de Cheria, se sua fdp conjunta é dada por

$$f(y_1, y_2) = \frac{\exp(-y_1 - y_2)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} \int_0^{\min(y_1, y_2)} (y_1 - z)^{\theta_1 - 1} (y_2 - z)^{\theta_2 - 1} z^{\theta_3 - 1} \exp(z) dz, \quad (3.19)$$

em que $y_1, y_2 > 0$ e $\theta_1, \theta_2, \theta_3 > 0$. As variáveis Y_1 e Y_2 são correlacionadas e tem distribuições marginais gama com parâmetros de forma $\alpha_1 = \theta_1 + \theta_3$ e $\alpha_2 = \theta_2 + \theta_3$, respectivamente.

Para obter essa distribuição, via método de redução trivariada, consideramos $X_i \stackrel{iid}{\sim} \Gamma(\theta_i, 1)$ para $i = 1, \dots, 3$. Segue que, como os X_i são independentes, a função de densidade conjunta é dada pelo produto das marginais, ou seja:

$$\begin{aligned} f(x_1, x_2, x_3) &= f(x_1) \times f(x_2) \times f(x_3) \\ &= \frac{\exp[-(x_1 + x_2 + x_3)]}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} x_1^{\theta_1 - 1} x_2^{\theta_2 - 1} x_3^{\theta_3 - 1}. \end{aligned} \quad (3.20)$$

Se considerarmos as variáveis X e Y definidas por

$$\begin{aligned} X &= X_1 + X_3 \Rightarrow X_1 = X - X_3 \\ Y &= X_2 + X_3 \Rightarrow X_2 = Y - X_3, \end{aligned}$$

a partir de (3.20) a fdp conjunta de (X, Y, X_3) fica dada por:

$$\begin{aligned} f(x, y, x_3) &= \frac{\exp[-(x - x_3 + y - x_3 + x_3)]}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} (x - x_3)^{\theta_1 - 1} (y - x_3)^{\theta_2 - 1} x_3^{\theta_3 - 1} \\ &= \frac{\exp(x_3 - x - y)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} (x - x_3)^{\theta_1 - 1} (y - x_3)^{\theta_2 - 1} x_3^{\theta_3 - 1}, \end{aligned} \quad (3.21)$$

com $0 \leq x_3 \leq x$ e $0 \leq x_3 \leq y$. Integrando $f(x, y, x_3)$ em relação a variável X_3 e considerando $\delta = \min(x, y)$ temos:

$$\begin{aligned} f(x, y) &= \int_0^\delta f(x, y, x_3) dx_3 \\ &= \frac{\exp[-(x + y)]}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} \int_0^\delta (x - x_3)^{\theta_1 - 1} (y - x_3)^{\theta_2 - 1} x_3^{\theta_3 - 1} dx_3, \end{aligned} \quad (3.22)$$

em que $f(x, y)$ corresponde à fdp conjunta de (X, Y) com distribuição gama bivariada de Cheria.

Enfim, para gerarmos uma amostra aleatória de vetores bivariados $(Y_{11}, Y_{21})^\top, \dots, (Y_{1n}, Y_{2n})^\top$ com distribuição gama bivariada de Cherian, de tal forma, que as variáveis marginais possuam distribuição gama e sejam correlacionadas, consideramos o algoritmo de redução trivariada descrito abaixo com os seguintes parâmetros $\alpha_1, \alpha_2, \mu_1, \mu_2 > 0$ e $0 < \rho < \frac{\min(\alpha_1, \alpha_2)}{\sqrt{\alpha_1 \alpha_2}}$.

• Algoritmo:

1. Geramos $X_1 \sim \Gamma(\alpha_1 - \rho\sqrt{(\alpha_1 \alpha_2)}, 1)$, $X_2 \sim \Gamma(\alpha_2 - \rho\sqrt{(\alpha_1 \alpha_2)}, 1)$ e $X_3 \sim \Gamma(\rho\sqrt{(\alpha_1 \alpha_2)}, 1)$;
2. Fazemos $Y_1 = (X_1 + X_3)\frac{\mu_1}{\alpha_1}$ e $Y_2 = (X_2 + X_3)\frac{\mu_2}{\alpha_2}$;

O vetor (Y_1, Y_2) tem distribuição bivariada gama de Cherian, o mais importante é que $Y_i \sim \Gamma\left(\alpha_i, \frac{\mu_i}{\alpha_i}\right)$ para $i = 1, 2$, com $\rho = \text{corr}(Y_1, Y_2)$. Assim, podemos simular dados assimétricos, positivos e correlacionados, com distribuição gama e avaliar os modelos Heckman-BS, Heckman clássico, Skew-normal e Heckman-t sob falta de especificação. Para isso, consideramos a seguinte estrutura de regressão,

$$\mu_{1i} = \exp(0.5 + 1.5x_{2i}), \quad \mu_{2i} = \exp(-0.5 + 1.2x_{2i} + 2x_{3i}), \text{ para } i = 1, \dots, n,$$

com parâmetros de forma $\alpha_1 = 2$ e $\alpha_2 = 1$, e em que, os regressores $x_{ji} \stackrel{iid}{\sim} \mathcal{U}[0, 1], j = 2, 3, \forall i$. Logo, a partir do algoritmo anterior, geramos n variáveis aleatórias correlacionadas para cada réplica de Monte Carlo definidas por:

$$Y_{1i} \sim \Gamma\left(2, \frac{\mu_{1i}}{2}\right) \quad \text{e} \quad Y_{2i} \sim \Gamma(1, \mu_{2i}), i = 1, \dots, n.$$

Consideramos então $Y_i = Y_{1i}U_i$ em que $U_i = 1\{Y_{2i} > 1\}$, $i = 1, \dots, n$, e ajustamos os modelos Heckman-BS, Heckman Clássico, Skew-normal e Heckman-t aos dados simulados. Os tamanhos amostrais considerados foram $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo em todos os casos. A fim de examinar o desempenho dos EMV's dos modelos ajustados, calculamos a média empírica das estimativas dos parâmetros e a raiz quadrada do erro quadrático médio (REQM). O grau de censura (valores para os quais $Y_2^* \leq 1$) é de aproximadamente 30%. Os resultados do ajuste dos quatro modelos são apresentados na Tabela 3.17. O parâmetro de forma α_2 não é estimado, tal como os parâmetros ϕ_2 do modelo Heckman-BS e σ_2 dos modelos Heckman clássico, Skew-normal e Heckman-t, assumimos $\alpha_2 = 1$. Apresentamos também, nas Figuras 3.7, 3.8 e 3.9, o box-plot das estimativas para o ajuste dos quatro modelos.

De acordo com os resultados apresentados na Tabela 3.17, o modelo Heckman-BS não estima bem o intercepto da equação de seleção e o parâmetro de forma (α_1). Porém, estima melhor, todos os demais parâmetros, inclusive o parâmetro de correlação. Além disso, é o único modelo cujo os EMV's são consistentes para todos os parâmetros. Ressaltamos ainda que, diferentemente dos demais modelos, não foi necessário a transformação da variável de interesse para o ajuste do modelo Heckman-BS. Adicionalmente, as Figuras 3.7, 3.8 e 3.9 indicam que o modelo Heckman-BS é o mais eficiente na estimação do parâmetro de correlação, principal interesse na aplicação de um modelo de seleção amostral.

Tabela 3.17: Média empírica das estimativas de máxima verossimilhança com a respectiva raiz quadrada do erro quadrático médio (REQM) do ajuste dos modelos de Heckman BS, Heckman clássico, Skew-normal e Heckman-t aos dados simulados com distribuição gama bivariada de Cherian. Tamanhos amostrais $n = 500, n = 1000$ e $n = 2000$ com $N = 1000$ réplicas de Monte Carlo.

Parâmetros	n	Heckman BS		Heckman Clássico		Heckman Skew		Heckman t-Student	
		Média	REQM	Média	REQM	Média	REQM	Média	REQM
γ_1 -0.500	500	-0.153	0.404	-0.667	0.230	-0.539	0.284	-0.662	0.256
	1000	-0.157	0.373	-0.672	0.205	-0.458	0.271	-0.683	0.231
	2000	-0.161	0.354	-0.672	0.190	-0.529	0.269	-0.697	0.217
γ_2 1.200	500	1.119	0.293	0.883	0.386	0.861	0.405	0.924	0.362
	1000	1.138	0.205	0.899	0.337	0.856	0.377	0.936	0.308
	2000	1.144	0.151	0.902	0.318	0.874	0.348	0.944	0.282
γ_3 2.00	500	1.924	0.278	1.520	0.525	1.400	0.698	1.521	0.573
	1000	1.918	0.206	1.519	0.505	1.349	0.724	1.554	0.497
	2000	1.914	0.159	1.512	0.499	1.448	0.579	1.572	0.449
β_1 0.500	500	0.470	0.135	0.310	0.223	0.533	0.391	0.463	0.247
	1000	0.463	0.094	0.298	0.214	0.656	0.453	0.417	0.189
	2000	0.461	0.070	0.299	0.207	0.517	0.381	0.401	0.145
β_2 1.500	500	1.479	0.184	1.435	0.155	1.389	0.231	1.362	0.229
	1000	1.483	0.145	1.449	0.107	1.397	0.210	1.402	0.157
	2000	1.491	0.098	1.451	0.084	1.452	0.104	1.412	0.121
α_1 2.000	500	2.910	1.024	0.737	1.264	0.841	1.176	0.685	1.316
	1000	2.862	0.935	0.737	1.263	0.909	1.113	0.676	1.324
	2000	2.847	0.886	0.736	1.264	0.867	1.157	0.669	1.332
ρ 0.500	500	0.487	0.139	0.345	0.224	0.247	0.514	0.103	0.582
	1000	0.499	0.059	0.359	0.166	0.239	0.569	0.176	0.444
	2000	0.492	0.040	0.355	0.157	0.401	0.247	0.209	0.345

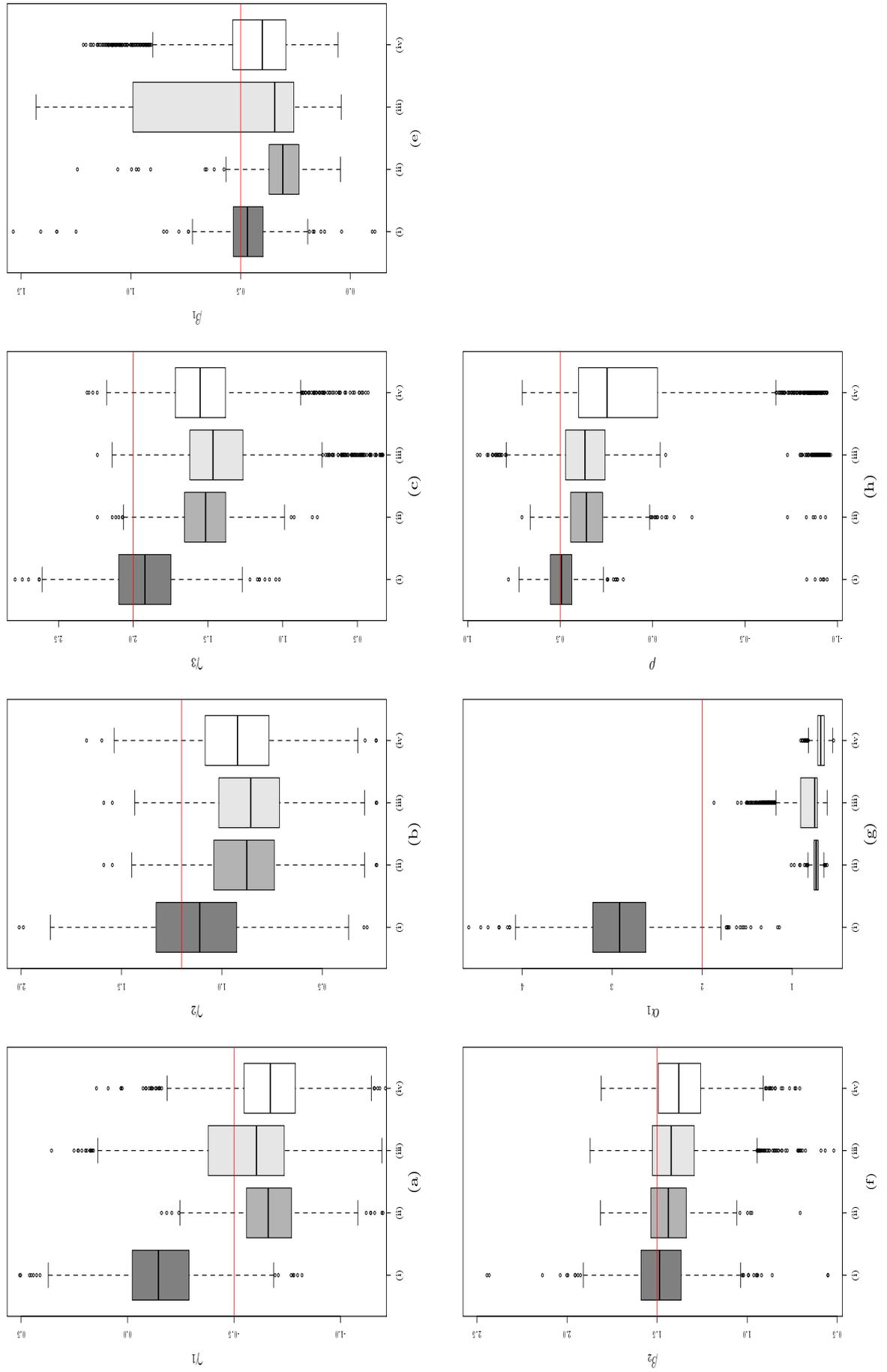


Figura 3.7: Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) β_1 , (e) β_2 , (f) α_1 e (g) ρ obtidas a partir do ajuste dos modelos (i) Heckman BS, (ii) Heckman Clássico, (iii) Skew-normal e (iv) Heckman-t. Tamanho amostral $n = 500$.

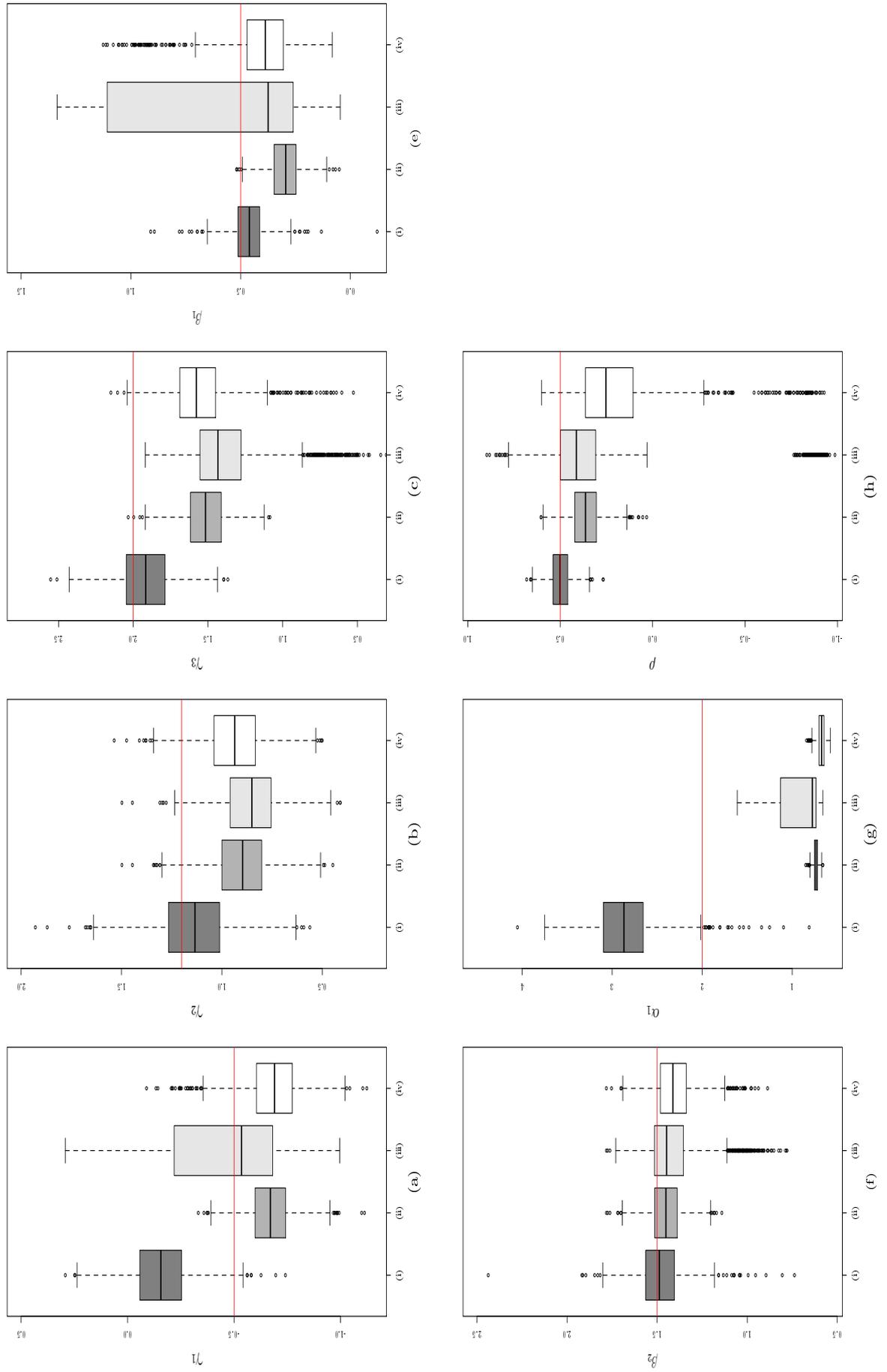


Figura 3-8: Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) β_1 , (e) β_2 , (f) α_1 e (g) ρ obtidas a partir do ajuste dos modelos (i) Heckman BS, (ii) Heckman Clássico, (iii) Skew-normal e (iv) Heckman-t. Tamanho amostral $n = 1000$.

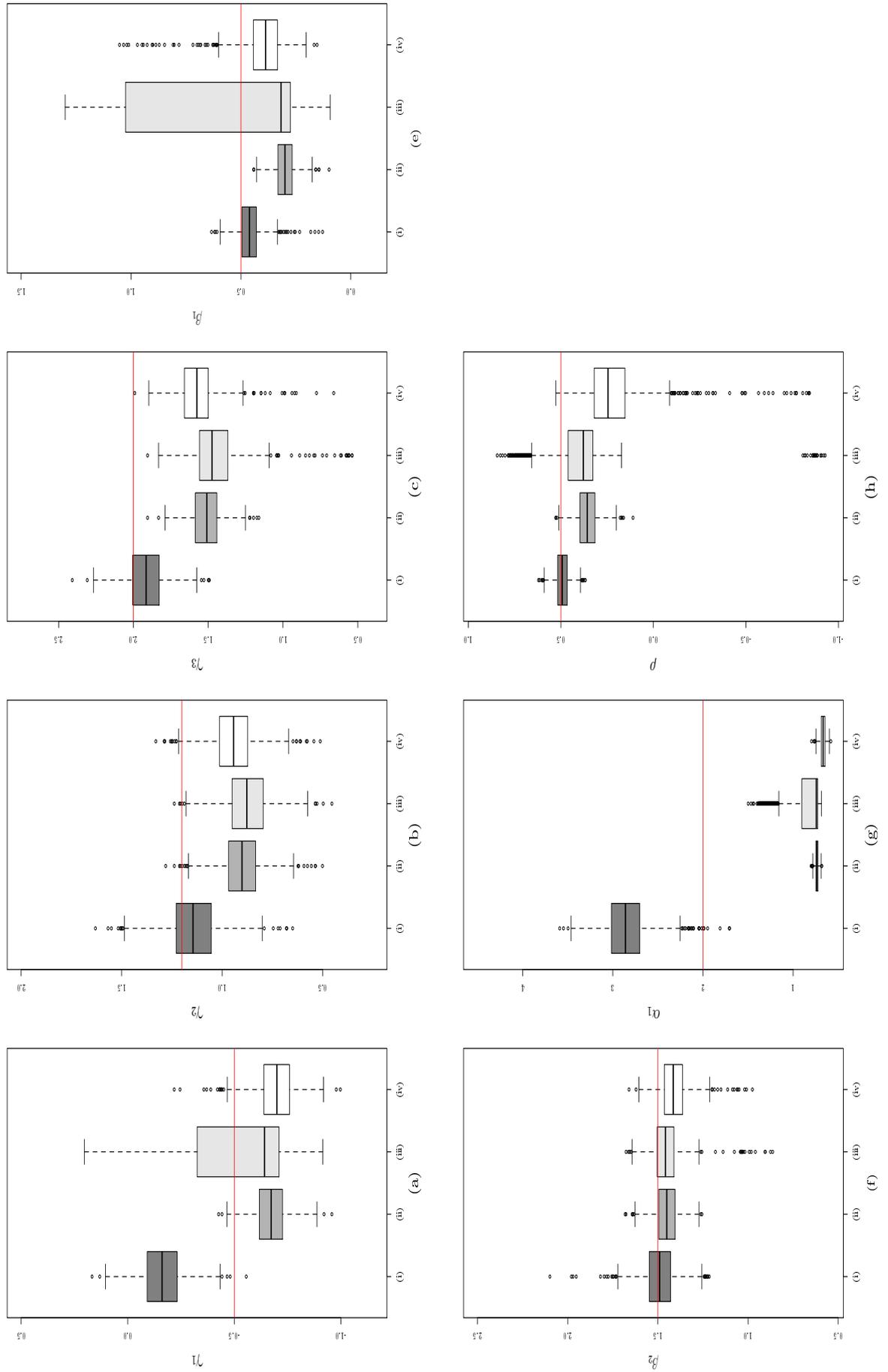


Figura 3-9: Comparação das estimativas de Máxima Verossimilhança dos parâmetros (a) γ_1 , (b) γ_2 , (c) γ_3 , (d) β_1 , (e) β_2 , (f) α_1 e (g) ρ obtidas a partir do ajuste dos modelos (i) Heckman BS, (ii) Heckman Clássico, (iii) Skew-normal e (iv) Heckman-t. Tamanho amostral $n = 2000$.

Para uso do algoritmo de redução trivariada temos que simular dados com parâmetro $\rho > 0$, logo, não calculamos o nível empírico dos testes da razão de verossimilhança e gradiente. Estimamos porém o poder de ambos os testes sob o ajuste dos modelos Heckman-BS, Heckman clássico e Skew-normal. Mais uma vez observamos resultados que contrariam nossa expectativa teórica com o ajuste do modelo Heckman-t e preferimos não apresentá-los. De acordo com a Tabela 3.18, ambos os testes apresentaram resultados de poder superior sob o ajuste do modelo Heckman-BS, com destaque para o teste da razão de verossimilhança. É interessante notar que mais uma vez os testes apresentaram poder semelhante sob o ajuste dos modelos Heckman clássico e Skew-normal, indicando não haver diferença de poder sob o ajuste desses modelos.

Tabela 3.18: Poder (em porcentagem) dos testes da razão de verossimilhança e gradiente para $\alpha = 1\%$, $\alpha = 5\%$ e $\alpha = 10\%$.

ρ	Heckman BS				Heckman Clássico				Heckman Skew			
	$n = 500$		$n = 1000$		$n = 500$		$n = 1000$		$n = 500$		$n = 1000$	
	S_{RV}	S_G	S_{RV}	S_G	S_{RV}	S_G	S_{RV}	S_G	S_{RV}	S_G	S_{RV}	S_G
$\alpha = 1\%$												
0.1	50.6	34.2	87.9	82.4	14.4	1.1	12.0	2.4	19.8	1.1	13.0	1.7
0.2	56.8	42.0	94.2	90.3	8.9	0.5	6.6	1.5	12.5	0.5	7.8	1.8
0.3	63.7	48.8	95.9	94.1	5.6	0.9	9.0	4.7	11.2	0.9	10.3	5.1
0.4	74.0	59.6	98.7	97.2	8.2	2.4	25.4	14.1	15.0	2.5	25.8	13.9
0.5	80.0	67.5	99.7	99.5	20.9	8.2	55.9	40.9	26.7	8.9	55.1	41.0
0.6	88.0	76.2	99.9	99.8	45.1	24.2	89.0	79.3	49.2	25.9	88.3	79.3
0.7	96.2	90.5	100	100	86.2	66.1	100	99.1	86.6	66.3	100	99.0
$\alpha = 5\%$												
0.1	75.7	63.8	97.5	96.0	19.5	6.1	16.6	10.5	24.5	6.0	17.4	8.9
0.2	80.4	70.6	99.0	98.1	14.0	5.7	16.3	10.4	17.3	6.1	17.3	10.5
0.3	86.7	80.5	98.8	98.0	17.1	8.6	29.4	22.2	22.5	9.1	30.2	22.7
0.4	92.4	86.8	99.9	99.9	25.9	15.2	55.1	44.7	32.3	16.9	55.3	45.0
0.5	94.7	91.9	100	100	47.7	33.3	83.4	76.5	52.7	35.4	82.2	77.1
0.6	97.1	95.6	100	100	73.1	60.8	97.8	96.2	75.4	62.3	97.1	96.3
0.7	99.5	99.2	100	100	97.1	94.1	100	100	97.3	94.3	100	100
$\alpha = 10\%$												
0.1	84.1	74.7	98.8	98.4	24.2	12.3	21.6	16.9	28.9	12.8	22.4	14.9
0.2	87.9	80.3	99.6	99.4	20.0	12.0	24.4	19.5	23.2	12.4	25.2	19.3
0.3	93.5	89.1	99.9	99.6	27.0	18.8	43.3	37.9	31.9	20.0	43.9	38.0
0.4	96.0	93.8	100	100	40.0	30.6	68.6	62.3	45.2	32.2	68.6	62.4
0.5	97.8	97.0	100	100	60.7	52.3	90.8	88.3	64.4	54.1	89.4	88.4
0.6	98.8	98.5	100	100	84.2	77.5	99.1	98.7	85.7	78.8	98.4	98.7
0.7	99.8	99.8	100	100	98.6	97.6	100	100	98.7	97.9	100	100

3.6 Aplicação a Dados Reais

Analizamos novamente os dados do *Medical Expenditure Panel Survey* (MEPS) de 2001 utilizados na aplicação do modelo Heckman generalizado no Capítulo 2. A amostra é restrita a pessoas com idades entre 21 e 64 anos e contém uma variável resposta, denotada aqui por Y_1^* , com 3328 observações de valores de gastos ambulatoriais, das quais 526 (15,8%) correspondem aos valores de despesas zero. Esse conjunto de dados contém também as seguintes variáveis explicativas:

- x_1 , que representa a idade, medida em dezenas de anos;
- x_2 , que é uma variável indicadora para sexo, sexo feminino recebe valor 1;
- x_3 , representa anos de escolaridade;
- x_4 , indicadora para etnia, negro ou hispânico recebem valor 1;
- x_5 , representa o número de doenças crônicas ;
- e, x_6 , é uma variável indicadora para presença de seguro.

A variável Y_2^* não é observada, temos $U_i = 1\{Y_{2i}^* > 0\}$, para $i = 1, \dots, 3328$, que representa a decisão de gastar ou não com cuidados médicos. Os dados MEPS também foram utilizados por [Cameron e Trivedi \(2009\)](#) e [Marchenko e Genton \(2012\)](#). De acordo com esses autores, a decisão de gastar provavelmente esta relacionada ao valor da despesa, e portanto, é natural considerar um modelo de seleção amostral para a análise. Esses autores rejeitam a hipótese de normalidade conjunta de Y_1^* e Y_2^* . Além disso, a variável despesas ambulatoriais (Y_1^*) é assimétrica e positiva. Nesse sentido, propomos analisar tais dados com o modelo Heckman-BS. Ressaltamos que para o ajuste deste modelo, não é necessária a transformação logarítmica da variável de despesas ambulatoriais. Isso representa uma grande vantagem para a interpretação dos parâmetros do modelo.

Consideramos, então, a seguinte estrutura de regressão para os parâmetros relacionados a média

$$\log \mu_{1i} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i}, \quad (3.23)$$

$$\log \mu_{2i} = \gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \gamma_3 x_{3i} + \gamma_4 x_{4i} + \gamma_5 x_{5i} + \gamma_6 x_{6i} + \gamma_7 x_{7i}, \text{ para } i = 1, 2, \dots, n.$$

A equação de seleção inclui a variável renda (x_7), dessa forma, é imposto a restrição de exclusão. Na Tabela 3.19 apresentamos os resultados do ajuste do modelo de Heckman clássico e do modelo Heckman-BS. Nosso maior interesse é testar $H_0 : \rho = 0$ contra $H_1 : \rho \neq 0$ sob o ajuste destes modelos. Apresentamos os resultados do teste de Wald que possui o mesmo resultado dos testes da razão de verossimilhança e gradiente também realizados, porém não apresentados aqui. Na aplicação de todos os testes, sob o ajuste do modelo de Heckman clássico, não rejeitamos H_0 (p -valor > 0.1). Contrariamente, sob o ajuste do modelo Heckman-BS, todos os testes indicam evidências suficientes (p -valor < 0.001) para rejeitar H_0 . Nosso resultado concorda com os resultados obtidos por [Marchenko e Genton \(2012\)](#), por [Zhelonkin et al. \(2016\)](#) e também com os resultados do ajuste do modelo Heckman generalizado no Capítulo 2. Ressaltamos ainda que as estimativas dos modelos de Heckman clássico e Heckman-BS são similares, com exceção do intercepto da equação de seleção. Tal resultado indica que o ajuste com o modelo Heckman-BS não modifica as conclusões e inferências realizadas para os parâmetros.

Tabela 3.19: Estimativas do Modelo Heckman-BS com os respectivos valores de desvio-padrão (DP), valor-z, p -valor, limite inferior e limite superior para o intervalo de confiança de 95%, juntamente com as estimativas do Modelo Heckman clássico e os respectivos p -valores.

Equação de Seleção								
Parâmetros	Heckman	p -valor	Heckman-BS	DP	Valor-z	p -valor	Inferior	Superior
(Intercepto)	-0.676	0.000	0.117	0.235	0.498	0.618	-0.344	0.578
Idade (x_1)	0.088	0.001	0.090	0.032	2.835	0.005	0.028	0.151
Fem (x_2)	0.663	0.000	0.722	0.069	10.456	0.000	0.587	0.857
Educ (x_3)	0.062	0.000	0.068	0.015	4.651	0.000	0.040	0.097
Blhisp (x_4)	-0.364	0.000	-0.399	0.074	-5.367	0.000	-0.545	-0.253
Totcr (x_5)	0.797	0.000	0.806	0.068	11.812	0.000	0.672	0.939
ins (x_6)	0.170	0.007	0.179	0.073	2.454	0.014	0.036	0.323
Renda (x_7)	0.003	0.040	0.003	0.001	2.395	0.017	0.001	0.006
Equação Primária								
Parâmetros	Heckman	p -valor	Heckman-BS	DP	Valor-z	p -valor	Inferior	Superior
(Intercepto)	5.044	0.000	5.779	0.173	33.469	0.000	5.440	6.117
Idade (x_1)	0.212	0.000	0.246	0.022	11.117	0.000	0.203	0.290
Fem (x_2)	0.348	0.000	0.411	0.048	8.511	0.000	0.316	0.505
Educ (x_3)	0.019	0.076	-0.007	0.010	-0.671	0.503	-0.026	0.013
Blhisp (x_4)	-0.219	0.000	-0.215	0.053	-4.047	0.000	-0.319	-0.111
Totcr (x_5)	0.540	0.000	0.589	0.034	17.165	0.000	0.522	0.656
ins (x_6)	-0.030	0.557	-0.061	0.050	-1.236	0.216	-0.159	0.036
ϕ	1.271	0.000	0.703	0.021	33.917	0.000	0.662	0.743
ρ	-0.131	0.375	0.273	0.057	4.759	0.000	0.161	0.385

3.7 Conclusões

Introduzimos um modelo de seleção amostral baseado na distribuição Birnbaum-Saunders bivariada. Nosso modelo fornece uma maior flexibilidade para a modelagem de dados assimétricos positivos quando comparado ao modelo clássico de Heckman, mantendo a mesma quantidade de parâmetros do modelo clássico, o que é uma vantagem quando comparado também aos modelos Skew-normal e Heckman-t.

Realizamos simulações de Monte Carlo para avaliar nosso modelo com relação aos existentes na literatura. Nossas simulações indicam que, para dados assimétricos e positivos gerados conforme o modelo de seleção BS, os EMV's dos modelos de Heckman clássico, Skew-normal e Heckman-t são tendenciosos e não são eficientes tanto na presença quanto ausência de restrição de exclusão para todos os valores de correlação. Nosso modelo, contrariamente, apresentou excelentes resultados na estimação dos parâmetros para todos os valores de correlação na presença de restrição de exclusão. Na ausência de restrição de exclusão e para altos valores de correlação, nosso modelo também apresentou estimativas dos parâmetros viesadas, apesar de, ainda apresentar estimativas com menor viés quando comparado com as estimativas obtidas pelos demais modelos.

Sob a especificação do modelo de Heckman-BS, nossas simulações de poder e nível de significância empírico, para testar $H_0 : \rho = 0$ contra $H_1 : \rho \neq 0$, utilizando os testes da razão de verossimilhança e gradiente, indicaram que todos os modelos explorados mantiveram o nível nominal correto. Nosso modelo, porém, apresentou maior poder para todos os níveis de significância avaliados com destaque para o teste da RV que foi superior ao teste gradiente em relação ao poder de rejeição da hipótese nula

quando nos afastamos da mesma. É importante destacar que os modelos Skew-normal e Heckman clássico tiveram resultados semelhantes para o poder de ambos os testes. Esperávamos resultados superiores para o modelo Skew-normal, já que os dados são assimétricos. Os resultados de poder e nível de significância empírico sob o ajuste do modelo Heckman-t não foram apresentados.

Avaliamos nosso modelo também sob falta de especificação para dados assimétricos e positivos com distribuição gama, novamente, o modelo Heckman-BS se mostrou melhor, pois as estimativas obtidas sob seu ajuste apresentaram menor viés, maior precisão e menor REQM quando comparadas com o ajuste dos demais modelos.

Na aplicação do modelo Heckman-BS utilizamos dados que foram anteriormente avaliados para o ajuste dos modelos Heckman-t, desenvolvido no trabalho de [Marchenko e Genton \(2012\)](#), e pelo método de dois passos robusto, desenvolvido no trabalho de [Zhelonkin et al. \(2016\)](#). Nossos resultados, assim como aqueles obtidos por esses autores, indicam que há fortes evidências da presença de viés de seleção nos dados. É importante ressaltar que nosso modelo quando comparado ao Heckman-t é mais parcimonioso e possui maior facilidade de interpretação dos parâmetros, uma vez que, não há a necessidade da transformação da variável de interesse para seu ajuste.

Como sugerido por [Ogundimu e Hutton \(2016\)](#), na presença de desvios da normalidade devido a assimetria, o modelo de Heckman clássico, pode atribuir, erroneamente, correlação nula à relação das equações de regressão e seleção. Nesse caso, modelos para dados assimétricos, como o proposto nessa tese, são mais realistas. Logo, nosso modelo é indicado para substituir os modelos de seleção existentes quando os dados são assimétricos e positivos, pois indica ganho de eficiência na estimação dos parâmetros, não necessita da transformação da variável resposta e tem o mesmo número de parâmetros do modelo clássico.

4.1 Conclusões

Para tratar dados com problemas de viés de seleção, existem na literatura, no âmbito paramétrico, os modelos Heckman clássico, Heckman-Skew e Heckman-t. Nesta tese, abordamos aspectos de modelagem e inferência de dois novos modelos de seleção amostral. Inicialmente, no capítulo 1, apresentamos o modelo de Heckman clássico com o métodos de estimação por máxima verossimilhança e o método de dois passos. Mostramos também as distribuições Birnbaum-Saunders univariada e bivariada, com alguns de seus resultados importantes. Em seguida, as contribuições desta tese são apresentadas de forma independente, nos Capítulos 2 e 3. Nossas principais contribuições e resultados são:

Capítulo 2: Propusemos o modelo Heckman generalizado que estende o modelo de seleção amostral de Heckman ao considerar covariáveis para os parâmetros de dispersão e de correlação. Sua principal vantagem é permitir identificar as fontes de variabilidade dos dados e as covariáveis responsáveis pela presença de viés de seleção. Analisamos o desempenho do modelo proposto comparado aos modelos Heckman clássico, Heckman-Skew e Heckman-t via simulação Monte Carlo para dados simulados de acordo com cinco cenários distintos. Os resultados indicam que nosso modelo é flexível e se ajusta melhor aos dados com problemas de viés de seleção devido a dispersão e/ou correlação variável. Comparamos também o poder dos testes da razão de verossimilhança e gradiente sob ajuste dos quatro modelos e observamos que ambos os testes tem comportamentos similares com uma vantagem para o teste da razão de verossimilhanças. Aplicamos o modelo Heckman generalizado para analisar os dados de despesas ambulatoriais do *Medical Expenditure Panel Survey (MEPS)* de 2001 e, assim como na aplicação do modelo Heckman-t, observamos evidências estatísticas da existência de viés de seleção amostral nesses dados que não foram detectadas pelo modelo de Heckman clássico.

Capítulo 3: Desenvolvemos uma alternativa ao modelo de Heckman para dados positivos e assimétricos com o uso da distribuição Birnbaum-Saunders, denominado Heckman-BS. O desempenho desse

modelo foi avaliado e comparado aos dos modelos Heckman clássico, Heckman-Skew e Heckman-t, via simulação Monte Carlo. Os resultados indicam que nosso modelo se ajusta melhor a dados assimétricos e positivos, pois, para esses dados, os EMV dos modelos de Heckman clássico, Heckman-Skew e Heckman-t são tendenciosos e os testes, da razão de verossimilhança e gradiente, não mantêm o nível nominal correto. Analisamos os dados de despesas ambulatoriais do *Medical Expenditure Panel Survey (MEPS)* de 2001 com nosso modelo e, assim como na aplicação do modelo Heckman generalizado, (Capítulo 2), e o modelo Heckman-t, observamos evidências estatísticas da existência de viés de seleção amostral que não foram detectados pelo modelo de Heckman clássico.

4.2 Trabalhos Futuros

Para o futuro, nosso interesse é introduzir uma análise de diagnóstico, juntamente com análise de resíduos e influência dos modelos Heckman-BS e Heckman generalizado. Nosso objetivo é também estudar novas possibilidades de extensão do modelo de Heckman com novas distribuições e escrever um pacote para o *software* R com os modelos propostos nesta tese e os modelos Heckman-t e Heckman-Skew.

REFERÊNCIAS BIBLIOGRÁFICAS

- H. Ahn e J. L. Powell. Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics*, 58:3–29, 1993.
- A. Arabmazar e P. Schmidt. Further evidence on the robustness of the tobit estimator to heteroskedasticity. *Journal of Econometrics*, 17:253–258, 1981.
- R. Arellano-Valle, G. del Pino, e E. S. Martin. Definition and probabilistic properties of skew-distributions. *Statistics & Probability Letters*, 58:111–121, 2002.
- R. B. Arellano-Valle, M. D. Branco, e M. G. Genton. A unified view on skewed distributions arising from selections. *Canadian Journal of Statistics*, 34:581–601, 2006.
- B. C. Arnold e R. J. Beaver. Hidden truncation models. *Sankhyā: The Indian Journal of Statistics, Series A*, 62:23–35, 2000.
- A. Atkinson. *Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*. Oxford Statistical Science Series. Clarendon Press, 1985.
- A. Azzalini. The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, 32:159–188, 2005.
- N. Balakrishnan e C.-D. Lai. *Continuous Bivariate Distributions*. Springer, Verlag, New York, 2 edition, 2009.
- Z. Birnbaum e S. C. Saunders. A new family of life distributions. *Journal of Applied Probability*, 6: 319–327, 1969a.
- Z. Birnbaum e S. C. Saunders. Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability*, 6:328–347, 1969b.
- C. A. Cameron e P. K. Trivedi. *Microeconometrics Using Stata*. TX: Stata press, College Station, revised edition, 2009.

- D. S. Chang e L. C. Tang. Reliability bounds and critical time for the birnbaum-saunders distribution. *IEEE Transactions on Reliability*, 42:464–469, 1993.
- D. S. Chang e L. C. Tang. Percentile bounds and tolerance limits for the birnbaum-saunders distribution. *Communications in Statistics - Theory and Methods*, 23:2853–2863, 1994.
- M. Das, W. K. Newey, e F. Vella. Nonparametric estimation of sample selection models. *The Review of Economic Studies*, 70:33–58, 2003.
- P. Ding. Bayesian robust inference of sample selection using selection-t models. *Journal of Multivariate Analysis*, 124:451–464, 2014.
- S. G. Donald. Two-step estimation of heteroskedastic sample selection models. *Journal of Econometrics*, 65:347–380, 1995.
- D. J. Dupuis e J. E. Mills. Robust estimation of the birnbaum-saunders distribution. *IEEE Transactions on Reliability*, 47:88–95, 1998.
- C. K. Enders. *Applied Missing Data Analysis*. Guilford Press, New York, NY, 1 edition, 2010.
- C. Fernandez e M. F. Steel. Multivariate student-t regression models: Pitfalls and inference. *Biometrika*, 86:153–167, 1999.
- R. V. Fonseca e F. Cribari-Neto. Inference in a bimodal birnbaum-saunders model. *Mathematics and Computers in Simulation*, 146:134 – 159, 2018.
- S. G. From e L. Li. Estimation of the parameters of the birnbaum-saunders distribution. *Communications in Statistics - Theory and Methods*, 35:2157–2169, 2006.
- R. S. Hartman. A monte carlo analysis of alternative estimators in models involving selectivity. *Journal of Business & Economic Statistics*, 9:41–49, 1991.
- J. J. Heckman. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, 5:475–492, 1976.
- J. J. Heckman. Sample selection bias as a specification error. *Econometrica*, 47:153–161, 1979.
- M. Hurd. Estimation in truncated samples when there is heteroscedasticity. *Journal of Econometrics*, 11:247–258, 1979.
- A. Jamalizadeh e D. Kundu. A multivariate birnbaum-saunders distribution based on the multivariate skew normal distribution. *Journal of the Japan Statistical Society*, 45:1–20, 2015.
- N. L. Johnson, S. Kotz, e N. Balakrishnan. *Continuous Univariate Distributions*. Wiley, New York, 2 edition, 1995.
- D. Kundu, N. Balakrishnan, e A. Jamalizadeh. Bivariate birnbaum-saunders distribution and associated inference. *Journal of Multivariate Analysis*, 101:113–125, 2010.

- L.-F. Lee. Generalized econometric models with selectivity. *Econometrica*, 51:507–512, 1983.
- V. Leiva. *The Birnbaum-Saunders Distribution*. Academic Press, 1 edition, 2015.
- V. Leiva, M. Riquelme, N. Balakrishnan, e A. Sanhueza. Lifetime analysis based on the generalized birnbaum–saunders distribution. *Computational Statistics & Data Analysis*, 52:2079 – 2097, 2008.
- A. J. Lemonte, F. Cribari-Neto, e K. L. Vasconcellos. Improved statistical inference for the two-parameter birnbaum–saunders distribution. *Computational Statistics & Data Analysis*, 51:4656 – 4681, 2007.
- A. J. Lemonte, A. B. Simas, e F. Cribari-Neto. Bootstrap-based improved estimators for the two-parameter birnbaum–saunders distribution. *Journal of Statistical Computation and Simulation*, 78: 37–49, 2008.
- S. F. Leung e S. Yu. On the choice between sample selection and twopart models. *Journal of Econometrics*, 72:197–229, 1996.
- S. F. Leung e S. Yu. Collinearity and two-step estimation of sample selection models: Problems, origins, and remedies. *Computational Economics*, 15:173–199, 2000.
- W. Manning, N. Duan, e W. Rogers. Monte carlo evidence on the choice between sample selection and two-part models. *Journal of Econometrics*, 35:59 – 82, 1987.
- Y. V. Marchenko e M. G. Genton. A heckman selection-t model. *Journal of the American Statistical Association*, 107:304–317, 2012.
- P. McCullagh e J. Nelder. *Generalized Linear Models*. Chapman & Hall, London, 1989.
- K. Nawata. A note on the estimation of models with sample-selection biases. *Economics Letters*, 42:15 – 24, 1993.
- K. Nawata. Estimation of sample selection bias models by the maximum likelihood estimator and heckman’s two-step estimator. *Economics Letters*, 45:33 – 40, 1994.
- F. D. Nelson. Efficiency of the two-step estimator for models with endogenous sample selection. *Journal of Econometrics*, 24:181 – 196, 1984.
- W. K. Newey. Two-step series estimation of sample selection models. *Econometrics Journal*, 12:217–229, 2009.
- H. Ng, D. Kundu, e N. Balakrishnan. Modified moment estimation for the two-parameter birnbaum–saunders distribution. *Computational Statistics & Data Analysis*, 43:283 – 298, 2003.
- H. Ng, D. Kundu, e N. Balakrishnan. Point and interval estimation for the two-parameter birnbaum–saunders distribution based on type-ii censored samples. *Computational Statistics & Data Analysis*, 50:3222 – 3242, 2006.
- E. O. Ogundimu e J. L. Hutton. A sample selection model with skew-normal distribution. *Scandinavian Journal of Statistics*, 43:172–190, 2016.

- H. J. Paarsch. A monte carlo comparison of estimators for censored regression models. *Journal of Econometrics*, 24:197–213, 1984.
- J. L. Powell. Symmetrically trimmed least squares estimation for tobit models. *Econometrica*, pages 1435–1460, 1986.
- J. L. Powell. Estimation of semiparametric models. *Handbook of Econometrics*, 4:2443–2521, 1994.
- W. H. Press, B. P. Flannery, S. A. Teukolsky, e W. T. Vetterling. *Numerical Recipes in C: The Art of Scientific Programming*. Cambridge University Press, 2 edition, 1992.
- H. Saulo, N. Balakrishnan, X. Zhu, J. F. B. Gonzales, e J. Leão. Estimation in generalized bivariate birnbaum-saunders models. *Metrika*, 80:427–453, 2017.
- S. C. Saunders. A family of random variables closed under reciprocation. *Journal of the American Statistical Association*, 69:533–539, 1974.
- B. W. Schmeiser e R. Lal. Bivariate gamma random vectors. *Operations Research*, 30:355–374, 1982.
- M. C. Seiler e F. A. Seiler. Numerical recipes in c: The art of scientific computing. *Risk Analysis*, 9: 415–416, 1989.
- R. M. Stolzenberg e D. A. Relles. Theory testing in a world of constrained research design: The significance of heckman’s censored sampling bias correction for nonexperimental research. *Sociological Methods & Research*, 18:395–415, 1990.
- R. Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA, 2016a. URL <http://www.rstudio.com/>.
- R. C. Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016b. URL <https://www.R-project.org/>.
- F. Vilca-Labra e V. Leiva-Sánchez. A new fatigue life model based on the family of skew-elliptical distributions. *Communications in Statistics - Theory and Methods*, 35:229–244, 2006.
- S. Wright e J. Nocedal. Numerical optimization. *Springer Science*, 35:67–68, 1999.
- M. Zhelonkin, M. G. Genton, e E. Ronchetti. *R Package ssmrob: Robust Estimation and Inference in Sample Selection Models*. CRAN, 2014. URL <https://CRAN.R-project.org/package=ssmrob>. R package version 0.4.
- M. Zhelonkin, M. G. Genton, e E. Ronchetti. Robust inference in sample selection models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78:805–827, 2016.

Apêndice

APÊNDICE A

MATRIZ HESSIANA DO MODELO HECKMAN GENERALIZADO

Seja $\boldsymbol{\theta} = (\boldsymbol{\gamma}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \boldsymbol{\kappa}^\top)$. A função log-verossimilhança para o modelo Heckman-BS baseado em um par de observações (y, u) é

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) = u \left\{ \log \Phi \left[\frac{\boldsymbol{\mu}_2 + \boldsymbol{\rho}(y - \boldsymbol{\mu}_1)/\boldsymbol{\sigma}}{\sqrt{1 - \boldsymbol{\rho}^2}} \right] + \log \phi \left(\frac{y - \boldsymbol{\mu}_1}{\boldsymbol{\sigma}} \right) - \log \boldsymbol{\sigma} \right\} \\ + (1 - u) \log \Phi(-\boldsymbol{\mu}_2). \end{aligned} \quad (\text{A.1})$$

em que $\boldsymbol{\mu}_1 = \mathbf{x}^\top \boldsymbol{\beta} = \boldsymbol{\eta}_1$, $\boldsymbol{\mu}_2 = \mathbf{w}^\top \boldsymbol{\gamma} = \boldsymbol{\eta}_2$, $\boldsymbol{\sigma} = \exp(\boldsymbol{\zeta}^\top \boldsymbol{\phi}) = \boldsymbol{\eta}_3$ e $\boldsymbol{\rho} = \text{sen}(\boldsymbol{\nu}^\top \boldsymbol{\kappa}) = \boldsymbol{\eta}_4$. Se y é observado, $u = 1$ e, se y é não observado, $u = 0$. Para simplificar o cálculo da matriz Hessiana, consideramos

$$z = \frac{y - \boldsymbol{\mu}_1}{\boldsymbol{\sigma}}, \quad A_{\boldsymbol{\rho}} = \frac{1}{\sqrt{1 - \boldsymbol{\rho}^2}}, \quad A_{\boldsymbol{\rho}\boldsymbol{\rho}} = \frac{\boldsymbol{\rho}}{\sqrt{1 - \boldsymbol{\rho}^2}} = \boldsymbol{\rho}A_{\boldsymbol{\rho}}, \quad \eta_\theta = A_{\boldsymbol{\rho}}\boldsymbol{\mu}_2 + A_{\boldsymbol{\rho}\boldsymbol{\rho}}z \text{ e } M(\eta_\theta) = \frac{\phi(\eta_\theta)}{\Phi(\eta_\theta)},$$

assim,

$$\mathcal{L}(\boldsymbol{\theta}) = u \left(\log \Phi(\eta_\theta) + \log \phi(z) - \log \boldsymbol{\sigma} \right) + (1 - u) \log \Phi(-\boldsymbol{\mu}_2).$$

Consideramos $S_{ab} = \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial a \partial b}$, com $a, b \in (\boldsymbol{\gamma}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \boldsymbol{\kappa}^\top)$ a derivada parcial de segunda ordem de $\mathcal{L}(\boldsymbol{\theta})$ com respeito a a e b . Logo, para $i, j = 1, \dots, p$, para $k, l = 1, \dots, q$, para $m, n = 1, \dots, r$, e para $u, v = 1, \dots, s$, temos que as entradas da matrix Hessiana são

$$\begin{aligned}
S_{\gamma_i \gamma_j} &= -u \left[A_{\rho}^2 M(\eta_{\theta}) \left(\eta_{\theta} + M(\eta_{\theta}) \right) \right] w_i w_j - (1-u) M(-\mu_2) \left(\mu_2 + M(-\mu_2) \right) w_i w_j, \\
S_{\gamma_i \beta_l} &= -u \frac{A_{\rho} A_{\rho\rho}}{\sigma} M(\eta_{\theta}) \left(\eta_{\theta} + M(\eta_{\theta}) \right) x_l w_i, \\
S_{\gamma_i \phi_n} &= -u \left[A_{\rho} A_{\rho\rho} z M(\eta_{\theta}) \left(\eta_{\theta} + M(\eta_{\theta}) \right) \right] w_i \zeta_n, \\
S_{\gamma_i \kappa_v} &= u \left[\rho - \left(\mu_2 A_{\rho\rho} + z A_{\rho} \right) \left(\eta_{\theta} + M(\eta_{\theta}) \right) \right] M(\eta_{\theta}) A_{\rho}^3 (\cos \eta_4) \nu_v w_i, \\
S_{\beta_k \gamma_j} &= u \left[\frac{A_{\rho} A_{\rho\rho}}{\sigma} M(\eta_{\theta}) \left(\eta_{\theta} + M(\eta_{\theta}) \right) \right] x_k w_j, \\
S_{\beta_k \beta_l} &= -u \left[1 - \frac{A_{\rho\rho}^2}{\sigma} M(\eta_{\theta}) \left(\eta_{\theta} + M(\eta_{\theta}) \right) \right] \frac{x_k x_l}{\sigma}, \\
S_{\beta_k \phi_n} &= u \left[A_{\rho\rho}^2 z M(\eta_{\theta}) \left(\eta_{\theta} + M(\eta_{\theta}) \right) + M(\eta_{\theta}) A_{\rho\rho} - z \right] \frac{x_k \zeta_n}{\sigma}, \\
S_{\beta_k \kappa_v} &= u \left[\left(\mu_2 A_{\rho\rho}^2 A_{\rho}^2 + z A_{\rho}^3 A_{\rho\rho} \right) \left(\eta_{\theta} + M(\eta_{\theta}) \right) - A_{\rho} (1 + A_{\rho\rho}^2) \right] M(\eta_{\theta}) \frac{(\cos \eta_4) \nu_v x_k}{\sigma}, \\
S_{\phi_m \gamma_j} &= u \left[A_{\rho} A_{\rho\rho} z M(\eta_{\theta}) \left(\eta_{\theta} + M(\eta_{\theta}) \right) \right] \zeta_m w_j, \\
S_{\phi_m \beta_l} &= -u \left[\frac{2z}{\sigma} - \frac{M(\eta_{\theta}) A_{\rho\rho}}{\sigma} - \frac{A_{\rho\rho}^2 z}{\sigma} M(\eta_{\theta}) \left(\eta_{\theta} + M(\eta_{\theta}) \right) \right] x_l \zeta_m, \\
S_{\phi_m \phi_n} &= u \left[A_{\rho\rho}^2 z^2 M(\eta_{\theta}) \left(\eta_{\theta} + M(\eta_{\theta}) \right) - M(\eta_{\theta}) A_{\rho\rho} z + 2z^2 \right] \zeta_m \zeta_n, \\
S_{\phi_m \kappa_v} &= u \left[\left(\mu_2 A_{\rho\rho}^2 A_{\rho}^2 + z A_{\rho}^3 A_{\rho\rho} \right) \left(\eta_{\theta} + M(\eta_{\theta}) \right) - A_{\rho} (1 + A_{\rho\rho}^2) \right] z M(\eta_{\theta}) (\cos \eta_4) \nu_v \zeta_m, \\
S_{\kappa_u \gamma_j} &= u \left[1 - \mu_2 A_{\rho} \left(\eta_{\theta} + M(\eta_{\theta}) \right) \right] A_{\rho}^2 A_{\rho\rho} M(\eta_{\theta}) (\cos \eta_4) \nu_u w_j, \\
S_{\kappa_u \beta_l} &= -u \left[\frac{A_{\rho}^3}{\sigma} + \frac{A_{\rho\rho}^2 A_{\rho}^2 \mu_2}{\sigma} M(\eta_{\theta}) \left(\eta_{\theta} + M(\eta_{\theta}) \right) \right] \nu_u x_l \cos \eta_4, \\
S_{\kappa_u \phi_n} &= -u \left[A_{\rho}^2 A_{\rho\rho}^2 z \mu_2 M(\eta_{\theta}) \left(\eta_{\theta} + M(\eta_{\theta}) \right) - z A_{\rho}^3 \right] (\cos \eta_4) \nu_u \zeta_n, \\
S_{\kappa_u \kappa_v} &= -u \left[\left(\mu_2 A_{\rho\rho} A_{\rho}^2 + z A_{\rho}^3 \right)^2 \left(\eta_{\theta} + M(\eta_{\theta}) \right) - A_{\rho}^3 \left(\mu_2 (1 + 3A_{\rho\rho}^2) + 3z A_{\rho\rho} A_{\rho} \right) \right] \\
&\quad \times M(\eta_{\theta}) (\cos \eta_4) \nu_u \nu_v.
\end{aligned}$$

APÊNDICE B

MATRIZ HESSIANA DO MODELO HECKMAN-BS

Seja $\boldsymbol{\theta} = (\boldsymbol{\gamma}^\top, \boldsymbol{\beta}^\top, \phi, \rho)$. A função log-verossimilhança para o modelo Heckman-BS baseado em um par de observações (y, u) é

$$\begin{aligned}
 \mathcal{L}(\boldsymbol{\theta}) = u & \left\{ -\frac{\phi}{4} \left(\sqrt{\frac{y(\phi+1)}{\phi\boldsymbol{\mu}_1}} - \sqrt{\frac{\phi\boldsymbol{\mu}_1}{y(\phi+1)}} \right)^2 \right. \\
 & + \log \Phi \left[\frac{[\boldsymbol{\mu}_2 - 2]}{2\sqrt{\boldsymbol{\mu}_2(1-\rho^2)}} + \frac{\sqrt{\phi}\rho}{\sqrt{2(1-\rho^2)}} \left(\sqrt{\frac{(\phi+1)y}{\phi\boldsymbol{\mu}_1}} - \sqrt{\frac{\phi\boldsymbol{\mu}_1}{y(\phi+1)}} \right) \right] \\
 & + \log \left(\frac{\sqrt{\phi+1}}{\sqrt{\phi\boldsymbol{\mu}_1 y}} + \frac{\sqrt{\phi\boldsymbol{\mu}_1}}{\sqrt{y^3(\phi+1)}} \right) + \frac{1}{4} \sqrt{\frac{\phi}{\pi}} \left. \right\} \\
 & + (1-u) \log \Phi \left\{ \sqrt{\frac{1}{2}} \left[\sqrt{\frac{2}{\boldsymbol{\mu}_2}} - \sqrt{\frac{\boldsymbol{\mu}_2}{2}} \right] \right\}, \tag{B.1}
 \end{aligned}$$

em que $\boldsymbol{\mu}_1 = \exp(\mathbf{x}^\top \boldsymbol{\beta}) = \exp(\eta_1)$ e $\boldsymbol{\mu}_2 = \exp(\mathbf{w}^\top \boldsymbol{\gamma}) = \exp(\eta_2)$. Se y é observado, $u = 1$, se y é não observado $u = 0$. Para simplificar o cálculo da matriz Hessiana, consideramos:

$$\begin{aligned}
 t_0(y, \boldsymbol{\mu}_1, \phi) &= \left(\sqrt{\frac{y(\phi+1)}{\phi\boldsymbol{\mu}_1}} - \sqrt{\frac{\phi\boldsymbol{\mu}_1}{y(\phi+1)}} \right), & t_1(y, \boldsymbol{\mu}_1, \phi) &= -\frac{\phi}{4} [t_0(y, \boldsymbol{\mu}_1, \phi)]^2, \\
 t_2(y, \boldsymbol{\mu}_1, \phi) &= \left(\frac{\sqrt{\phi+1}}{\sqrt{\phi\boldsymbol{\mu}_1 y}} + \frac{\sqrt{\phi\boldsymbol{\mu}_1}}{\sqrt{y^3(\phi+1)}} \right), & t_3(\phi) &= \frac{1}{4} \sqrt{\frac{\phi}{\pi}}, \\
 t_4(\boldsymbol{\mu}_2, \rho) &= \frac{\boldsymbol{\mu}_2 - 2}{2\sqrt{\boldsymbol{\mu}_2(1-\rho^2)}}, & t_5(\phi, \rho) &= \frac{\rho\sqrt{\phi}}{\sqrt{2(1-\rho^2)}}, & t_6(\boldsymbol{\mu}_2) &= \sqrt{\frac{1}{2}} \left[\sqrt{\frac{2}{\boldsymbol{\mu}_2}} - \sqrt{\frac{\boldsymbol{\mu}_2}{2}} \right],
 \end{aligned}$$

e para $k = 1, \dots, p$, e $i = 1, \dots, q$,

$$\frac{\partial t_0}{\partial \gamma_i} = \frac{\partial t_1}{\partial \gamma_i} = \frac{\partial t_2}{\partial \gamma_i} = \frac{\partial t_3}{\partial \gamma_i} = \frac{\partial t_5}{\partial \gamma_i} = 0, \quad \frac{\partial t_4}{\partial \gamma_i} = \frac{\partial t_4}{\partial \mu_2} \frac{\partial \mu_2}{\partial \eta_2} \frac{\partial \eta_2}{\partial \gamma_i} = \frac{\mu_2 + 2}{4\sqrt{\mu_2(1-\rho^2)}} \mathbf{w}_i,$$

$$\frac{\partial t_6}{\partial \gamma_i} = \frac{\partial t_6}{\partial \mu_2} \frac{\partial \mu_2}{\partial \eta_2} \frac{\partial \eta_2}{\partial \gamma_i} = -\frac{1}{2\sqrt{2}} \left(\sqrt{\frac{2}{\mu_2}} + \sqrt{\frac{\mu_2}{2}} \right) \mathbf{w}_i, \quad \frac{\partial t_3}{\partial \beta_k} = \frac{\partial t_4}{\partial \beta_k} = \frac{\partial t_5}{\partial \beta_k} = \frac{\partial t_6}{\partial \beta_k} = 0,$$

$$\frac{\partial t_0}{\partial \beta_k} = \frac{\partial t_0}{\partial \mu_1} \frac{\partial \mu_1}{\partial \eta_1} \frac{\partial \eta_1}{\partial \beta_k} = -\frac{1}{2} \left(\sqrt{\frac{y(\phi+1)}{\phi \mu_1}} + \sqrt{\frac{\phi \mu_1}{y(\phi+1)}} \right) \mathbf{x}_k, \quad \frac{\partial t_4}{\partial \phi} = \frac{\partial t_6}{\partial \phi} = 0,$$

$$\frac{\partial t_1}{\partial \beta_k} = \frac{\partial t_1}{\partial \mu_1} \frac{\partial \mu_1}{\partial \eta_1} \frac{\partial \eta_1}{\partial \beta_k} = -\frac{\phi}{2} t_0 \frac{\partial t_0}{\partial \beta_k}, \quad \frac{\partial t_0}{\partial \rho} = \frac{\partial t_1}{\partial \rho} = \frac{\partial t_2}{\partial \rho} = \frac{\partial t_3}{\partial \rho} = \frac{\partial t_6}{\partial \rho} = 0,$$

$$\frac{\partial t_2}{\partial \beta_k} = \frac{\partial t_2}{\partial \mu_1} \frac{\partial \mu_1}{\partial \eta_1} \frac{\partial \eta_1}{\partial \beta_k} = \frac{1}{2} \left(\sqrt{\frac{\phi \mu_1}{y^3(\phi+1)}} - \sqrt{\frac{\phi+1}{\phi \mu_1 y}} \right) \mathbf{x}_k, \quad \frac{\partial t_3}{\partial \phi} = \frac{1}{8\sqrt{\pi \phi}},$$

$$\frac{\partial t_0}{\partial \phi} = -\frac{1}{2} \left(\sqrt{\frac{y}{(\phi+1)\phi^3 \mu_1}} + \sqrt{\frac{\mu_1}{y(\phi+1)^3 \phi}} \right), \quad \frac{\partial t_5}{\partial \phi} = \frac{\rho}{2\sqrt{2\phi(1-\rho^2)}},$$

$$\frac{\partial t_1}{\partial \phi} = -\frac{t_0^2}{4} + \frac{t_0}{4} \left(\sqrt{\frac{y}{(\phi+1)\phi \mu_1}} + \sqrt{\frac{\phi \mu_1}{y(\phi+1)^3}} \right), \quad \frac{\partial t_4}{\partial \rho} = \frac{\rho t_4}{(1-\rho^2)},$$

$$\frac{\partial t_2}{\partial \phi} = \frac{1}{2} \left(\sqrt{\frac{\mu_1}{y^3(\phi+1)^3 \phi}} - \frac{1}{\sqrt{y \mu_1 (\phi+1) \phi^3}} \right), \quad \frac{\partial t_5}{\partial \rho} = \frac{t_5}{\rho(1-\rho^2)}.$$

Daí, fizemos

$$\begin{aligned} f_1(y, \mu_1, \phi) &= t_1(y, \mu_1, \phi), & f_2(y, \mu_1, \phi) &= \log [t_2(y, \mu_1, \phi)], & f_3(\phi) &= \log [t_3(\phi)], \\ f_4(y, \mu_1, \mu_2, \phi, \rho) &= \log \Phi(t_4 + t_5 t_0), & f_5(\mu_2) &= \log \Phi[t_6(\mu_2)]. \end{aligned}$$

Derivamos parcialmente as funções acima e obtemos

$$\frac{\partial f_1}{\partial \gamma_i} = \frac{\partial f_2}{\partial \gamma_i} = \frac{\partial f_3}{\partial \gamma_i} = 0, \quad \frac{\partial f_4}{\partial \gamma_i} = \frac{\phi(t_4 + t_5 t_0)}{\Phi(t_4 + t_5 t_0)} \frac{\partial t_4}{\partial \gamma_i}, \quad \frac{\partial f_5}{\partial \gamma_i} = \frac{\phi(t_6)}{\Phi(t_6)} \frac{\partial t_6}{\partial \gamma_i},$$

$$\frac{\partial f_1}{\partial \beta_k} = -\frac{\phi}{2} t_0 \frac{\partial t_0}{\partial \beta_k}, \quad \frac{\partial f_2}{\partial \beta_k} = \frac{1}{t_2} \frac{\partial t_2}{\partial \beta_k}, \quad \frac{\partial f_3}{\partial \beta_k} = \frac{\partial f_5}{\partial \beta_k} = 0,$$

$$\frac{\partial f_4}{\partial \beta_k} = \frac{\phi(t_4 + t_5 t_0)}{\Phi(t_4 + t_5 t_0)} t_5 \frac{\partial t_0}{\partial \beta_k}, \quad \frac{\partial f_1}{\partial \phi} = \frac{\partial t_1}{\partial \phi}, \quad \frac{\partial f_2}{\partial \phi} = \frac{1}{t_2} \frac{\partial t_2}{\partial \phi},$$

$$\frac{\partial f_3}{\partial \phi} = \frac{1}{2\phi}, \quad \frac{\partial f_4}{\partial \phi} = \frac{\phi(t_4 + t_5 t_0)}{\Phi(t_4 + t_5 t_0)} \left(\frac{\partial t_5}{\partial \phi} t_0 + t_5 \frac{\partial t_0}{\partial \phi} \right), \quad \frac{\partial f_5}{\partial \phi} = 0,$$

$$\frac{\partial f_1}{\partial \rho} = \frac{\partial f_2}{\partial \rho} = \frac{\partial f_3}{\partial \rho} = \frac{\partial f_5}{\partial \rho} = 0, \quad \frac{\partial f_4}{\partial \rho} = \frac{\phi(t_4 + t_5 t_0)}{\Phi(t_4 + t_5 t_0)} \left(\frac{\partial t_4}{\partial \rho} + \frac{\partial t_5}{\partial \rho} t_0 \right).$$

Seja $M(x) = \frac{\phi(x)}{\Phi(x)}$, $x \in \mathbb{R}$, segue que $\frac{\partial M(x)}{\partial a} = -\frac{\partial x}{\partial a} M(x) \left(x + M(x) \right)$ e $a \in (\boldsymbol{\gamma}^\top, \boldsymbol{\beta}^\top, \phi, \rho)$. As derivadas parciais de segunda ordem das funções acima são:

$$\frac{\partial^2 f_4}{\partial \gamma_i \partial \gamma_j} = \frac{\partial}{\partial \gamma_j} M(t_4 + t_5 t_0) \frac{\partial t_4}{\partial \gamma_i} + M(t_4 + t_5 t_0) \frac{\partial^2 t_4}{\partial \gamma_i \partial \gamma_j},$$

$$\frac{\partial^2 t_4}{\partial \gamma_i \partial \gamma_j} = \frac{t_4}{4} w_i w_j, \quad \frac{\partial^2 f_5}{\partial \gamma_i \partial \gamma_j} = \frac{\partial}{\partial \gamma_j} M(t_6) \frac{\partial t_6}{\partial \gamma_i} + M(t_6) \frac{\partial^2 t_6}{\partial \gamma_i \partial \gamma_j}, \quad \frac{\partial^2 t_6}{\partial \gamma_i \partial \gamma_j} = \frac{1}{4} t_6 w_i w_j,$$

$$\frac{\partial^2 f_5}{\partial \gamma_i \partial \phi} = \frac{\partial}{\partial \phi} M(t_6) \frac{\partial t_6}{\partial \gamma_i}, \quad \frac{\partial^2 f_4}{\partial \gamma_i \partial \beta_l} = \frac{\partial}{\partial \beta_l} M(t_4 + t_5 t_0) \frac{\partial t_4}{\partial \gamma_i}, \quad \frac{\partial^2 f_5}{\partial \gamma_i \partial \beta_l} = \frac{\partial^2 f_5}{\partial \gamma_i \partial \rho} = 0,$$

$$\frac{\partial^2 f_4}{\partial \gamma_i \partial \phi} = \frac{\partial}{\partial \phi} M(t_4 + t_5 t_0) \frac{\partial t_4}{\partial \gamma_i}, \quad \frac{\partial^2 f_4}{\partial \beta_k \partial \gamma_j} = \frac{\partial}{\partial \gamma_j} M(t_4 + t_5 t_0) t_5 \frac{\partial t_0}{\partial \beta_k}, \quad \frac{\partial^2 t_4}{\partial \gamma_i \partial \rho} = \frac{\rho}{(1 - \rho^2)} \frac{\partial t_4}{\partial \gamma_i},$$

$$\frac{\partial^2 t_0}{\partial \beta_k \partial \beta_l} = \frac{1}{4} t_0 x_k x_l, \quad \frac{\partial t_2^{-1}}{\partial \beta_l} = -\frac{1}{t_2^2} \frac{\partial t_2}{\partial \beta_l}, \quad \frac{\partial^2 f_2}{\partial \beta_k \partial \beta_l} = \frac{\partial t_2^{-1}}{\partial \beta_l} \frac{\partial t_2}{\partial \beta_k} + \frac{1}{t_2} \frac{\partial^2 t_2}{\partial \beta_k \partial \beta_l},$$

$$\frac{\partial^2 t_2}{\partial \beta_k \partial \beta_l} = \frac{1}{4} t_2 x_k x_l, \quad \frac{\partial^2 f_3}{\partial \beta_k \partial \beta_l} = 0, \quad \frac{\partial^2 t_0}{\partial \phi \partial \beta_l} = \frac{1}{4} \left(\sqrt{\frac{\boldsymbol{\mu}_1}{y(\phi + 1)^3 \phi}} - \sqrt{\frac{y}{(\phi + 1)\phi^3 \boldsymbol{\mu}_1}} \right) x_l,$$

$$\frac{\partial^2 f_3}{\partial \beta_k \partial \phi} = 0, \quad \frac{\partial^2 f_1}{\partial \beta_k \partial \rho} = 0, \quad \frac{\partial^2 f_1}{\partial \phi \partial \gamma_j} = \frac{\partial^2 f_2}{\partial \phi \partial \gamma_j} = \frac{\partial^2 f_3}{\partial \phi \partial \gamma_j} = 0,$$

$$\frac{\partial^2 f_3}{\partial \beta_k \partial \rho} = 0, \quad \frac{\partial t_2^{-1}}{\partial \phi} = -\frac{1}{t_2^2} \frac{\partial t_2}{\partial \phi}, \quad \frac{\partial^2 f_4}{\partial \phi \partial \gamma_j} = \frac{\partial}{\partial \gamma_j} M(t_4 + t_5 t_0) \left(\frac{\partial t_5}{\partial \phi} t_0 + t_5 \frac{\partial t_0}{\partial \phi} \right),$$

$$\frac{\partial^2 f_2}{\partial \beta_k \partial \rho} = 0, \quad \frac{\partial^2 t_5}{\partial \phi \partial \rho} = \frac{1}{2\sqrt{2\phi(1 - \rho^2)^3}}, \quad \frac{\partial^2 f_2}{\partial \phi^2} = -\frac{1}{t_2^2} \left(\frac{\partial t_2}{\partial \phi} \right)^2 + \frac{1}{t_2} \frac{\partial^2 t_2}{\partial \phi^2},$$

$$\frac{\partial^2 f_3}{\partial \phi^2} = -\frac{1}{2\phi^2}, \quad \frac{\partial^2 t_5}{\partial \phi^2} = -\frac{\rho}{4\sqrt{2\phi^3(1-\rho^2)}}, \quad \frac{\partial^2 f_2}{\partial \phi \partial \beta_l} = -\frac{1}{t_2^2} \frac{\partial t_2}{\partial \beta_l} + \frac{1}{t_2} \frac{\partial^2 t_2}{\partial \phi \partial \beta_l},$$

$$\frac{\partial^2 f_4}{\partial \gamma_i \partial \rho} = \frac{\partial}{\partial \rho} M(t_4 + t_5 t_0) \frac{\partial t_4}{\partial \gamma_i} + M(t_4 + t_5 t_0) \frac{\partial^2 t_4}{\partial \gamma_i \partial \rho}, \quad \frac{\partial^2 f_1}{\partial \beta_k \partial \gamma_j} = \frac{\partial^2 f_2}{\partial \beta_k \partial \gamma_j} = \frac{\partial^2 f_3}{\partial \beta_k \partial \gamma_j} = 0,$$

$$\frac{\partial^2 f_1}{\partial \beta_k \partial \beta_l} = -\frac{\phi}{2} \left(\frac{\partial t_0}{\partial \beta_l} \frac{\partial t_0}{\partial \beta_k} + t_0 \frac{\partial^2 t_0}{\partial \beta_k \partial \beta_l} \right), \quad \frac{\partial^2 f_2}{\partial \beta_k \partial \phi} = \frac{\partial t_2^{-1}}{\partial \phi} \frac{\partial t_2}{\partial \beta_k} + \frac{1}{t_2} \frac{\partial^2 t_2}{\partial \beta_k \partial \phi},$$

$$\frac{\partial^2 f_1}{\partial \beta_k \partial \phi} = -\frac{1}{2} t_0 \frac{\partial t_0}{\partial \beta_k} - \frac{\phi}{2} \left(\frac{\partial t_0}{\partial \phi} \frac{\partial t_0}{\partial \beta_k} + t_0 \frac{\partial^2 t_0}{\partial \beta_k \partial \phi} \right), \quad \frac{\partial^2 f_1}{\partial \phi \partial \rho} = \frac{\partial^2 f_2}{\partial \phi \partial \rho} = \frac{\partial^2 f_3}{\partial \phi \partial \rho} = 0,$$

$$\frac{\partial^2 f_4}{\partial \beta_k \partial \beta_l} = \frac{\partial}{\partial \beta_l} M(t_4 + t_5 t_0) t_5 \frac{\partial t_0}{\partial \beta_k} + M(t_4 + t_5 t_0) \left(t_5 \frac{\partial^2 t_0}{\partial \beta_k \partial \beta_l} \right),$$

$$\frac{\partial^2 t_0}{\partial \beta_k \partial \phi} = \frac{1}{4} \left(\sqrt{\frac{y}{(\phi+1)\phi^3 \mu_1}} + \sqrt{\frac{\mu_1}{y(\phi+1)^3 \phi}} \right) x_k,$$

$$\frac{\partial^2 t_2}{\partial \beta_k \partial \phi} = \frac{1}{4} \left(\sqrt{\frac{\mu_1}{y^3 \phi (\phi+1)^3}} + \frac{1}{\sqrt{\mu_1 y \phi^3 (\phi+1)}} \right) x_k,$$

$$\frac{\partial^2 f_4}{\partial \beta_k \partial \rho} = \frac{\partial}{\partial \rho} M(t_4 + t_5 t_0) t_5 \frac{\partial t_0}{\partial \beta_k} + M(t_4 + t_5 t_0) \frac{\partial t_5}{\partial \rho} \frac{\partial t_0}{\partial \beta_k},$$

$$\frac{\partial^2 t_2}{\partial \phi \partial \beta_l} = \frac{1}{4} \left(\sqrt{\frac{\mu_1}{y^3 \phi (\phi+1)^3}} + \frac{1}{\sqrt{\mu_1 y \phi^3 (\phi+1)}} \right) x_l,$$

$$\frac{\partial^2 f_4}{\partial \phi \partial \beta_l} = \frac{\partial}{\partial \beta_l} M(t_4 + t_5 t_0) \left(\frac{\partial t_5}{\partial \phi} t_0 + t_5 \frac{\partial t_0}{\partial \phi} \right) + M(t_4 + t_5 t_0) \left(\frac{\partial t_5}{\partial \phi} \frac{\partial t_0}{\partial \beta_l} + t_5 \frac{\partial^2 t_0}{\partial \phi \partial \beta_l} \right),$$

$$\frac{\partial^2 t_2}{\partial \phi^2} = \frac{1}{4} \left(\frac{4\phi+3}{\phi^2 \sqrt{\mu_1 y \phi (\phi+1)^3}} - \frac{4\phi+1}{(\phi+1)^2} \sqrt{\frac{\mu_1}{y^3 \phi^3 (\phi+1)}} \right),$$

$$\frac{\partial^2 t_0}{\partial \phi^2} = \frac{1}{4} \left[\frac{4\phi+3}{\phi^4} \sqrt{\frac{y}{\mu_1 \phi (\phi+1)^3}} - \frac{4\phi+1}{(\phi+1)^4} \sqrt{\frac{\mu_1}{y \phi^3 (\phi+1)}} \right],$$

$$\frac{\partial^2 f_4}{\partial \beta_k \partial \phi} = \frac{\partial}{\partial \phi} M(t_4 + t_5 t_0) t_5 \frac{\partial t_0}{\partial \beta_k} + M(t_4 + t_5 t_0) \left(\frac{\partial t_5}{\partial \phi} \frac{\partial t_0}{\partial \beta_k} + t_5 \frac{\partial^2 t_0}{\partial \beta_k \partial \phi} \right),$$

$$\frac{\partial^2 f_4}{\partial \phi^2} = \frac{\partial}{\partial \phi} M(t_4 + t_5 t_0) \left(\frac{\partial t_5}{\partial \phi} t_0 + t_5 \frac{\partial t_0}{\partial \phi} \right) + M(t_4 + t_5 t_0) \left(\frac{\partial^2 t_5}{\partial \phi^2} t_0 + 2 \frac{\partial t_5}{\partial \phi} \frac{\partial t_0}{\partial \phi} + t_5 \frac{\partial^2 t_0}{\partial \phi^2} \right),$$

$$\frac{\partial^2 f_4}{\partial \phi \partial \rho} = \frac{\partial}{\partial \rho} M(t_4 + t_5 t_0) \left(\frac{\partial t_5}{\partial \phi} t_0 + t_5 \frac{\partial t_0}{\partial \phi} \right) + M(t_4 + t_5 t_0) \left(\frac{\partial^2 t_5}{\partial \phi \partial \rho} t_0 + \frac{\partial t_5}{\partial \rho} \frac{\partial t_0}{\partial \phi} \right),$$

$$\frac{\partial^2 f_4}{\partial \rho \partial \gamma_j} = \frac{\partial}{\partial \gamma_j} M(t_4 + t_5 t_0) \left(\frac{\partial t_4}{\partial \rho} + \frac{\partial t_5}{\partial \rho} t_0 \right) + M(t_4 + t_5 t_0) \frac{\partial^2 t_4}{\partial \rho \partial \gamma_j},$$

$$\frac{\partial^2 f_4}{\partial \rho \partial \phi} = \frac{\partial}{\partial \phi} M(t_4 + t_5 t_0) \left(\frac{\partial t_4}{\partial \rho} + \frac{\partial t_5}{\partial \rho} t_0 \right) + M(t_4 + t_5 t_0) \left(\frac{\partial^2 t_5}{\partial \rho \partial \phi} t_0 + \frac{\partial t_5}{\partial \rho} \frac{\partial t_0}{\partial \phi} \right),$$

$$\frac{\partial^2 f_4}{\partial \rho^2} = \frac{\partial}{\partial \rho} M(t_4 + t_5 t_0) \left(\frac{\partial t_4}{\partial \rho} + \frac{\partial t_5}{\partial \rho} t_0 \right) + M(t_4 + t_5 t_0) \left(\frac{\partial^2 t_4}{\partial \rho^2} + \frac{\partial^2 t_5}{\partial \rho^2} t_0 + \frac{\partial t_5}{\partial \rho} \frac{\partial t_0}{\partial \rho} \right),$$

$$\begin{aligned} \frac{\partial^2 f_1}{\partial \phi \partial \beta_i} &= \frac{\partial^2 t_1}{\partial \phi \partial \beta_i} \\ &= -\frac{1}{2} t_0 \frac{\partial t_0}{\partial \beta_i} + \frac{1}{4} \frac{\partial t_0}{\partial \beta_i} \left(\sqrt{\frac{y}{(\phi+1)\phi\mu_1}} + \sqrt{\frac{\phi\mu_1}{y(\phi+1)^3}} \right) \\ &\quad + \frac{t_0}{8} \left(\sqrt{\frac{\phi\mu_1}{y(\phi+1)^3}} - \sqrt{\frac{y}{(\phi+1)\phi\mu_1}} \right) x_i, \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 f_1}{\partial \phi^2} &= \frac{\partial^2 t_1}{\partial \phi^2} \\ &= -\frac{t_0}{2} \frac{\partial t_0}{\partial \phi} + \frac{1}{4} \frac{\partial t_0}{\partial \phi} \left(\sqrt{\frac{y}{(\phi+1)\phi\mu_1}} + \sqrt{\frac{\phi\mu_1}{y(\phi+1)^3}} \right) \\ &\quad - \frac{t_0}{8} \left((2\phi-1) \sqrt{\frac{\mu_1}{y(\phi+1)^5\phi}} + (2\phi+1) \sqrt{\frac{y}{(\phi+1)^3\phi\mu_1}} \right), \end{aligned}$$

$$\frac{\partial^2 t_4}{\partial \rho \partial \gamma_j} = \frac{\rho}{(1-\rho^2)} \frac{\partial t_4}{\partial \gamma_j}, \quad \frac{\partial^2 f_4}{\partial \rho \partial \beta_l} = \frac{\partial}{\partial \beta_l} M(t_4 + t_5 t_0) \left(\frac{\partial t_4}{\partial \rho} + \frac{\partial t_5}{\partial \rho} t_0 \right), \quad \frac{\partial^2 t_5}{\partial \rho^2} = \frac{3t_5}{(1-\rho^2)^2},$$

$$\frac{\partial^2 t_5}{\partial \rho \partial \phi} = \frac{1}{2\sqrt{2\phi(1-\rho^2)^3}}, \quad \frac{\partial^2 t_4}{\partial \rho^2} = \frac{1+\rho^2}{(1-\rho^2)^2} t_4 + \frac{\rho}{(1-\rho^2)} \frac{\partial t_4}{\partial \rho}.$$

Consideramos $S_{ab} = \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial a \partial b}$, com $a, b \in (\boldsymbol{\gamma}^\top, \boldsymbol{\beta}^\top, \phi, \rho)$ a derivada parcial de segunda ordem de $\mathcal{L}(\boldsymbol{\theta})$ com respeito a a e b . Logo, para $i, j = 1, \dots, q$ e $k, l = 1, \dots, p$, temos que as entradas da matrix Hessiana são:

$$\begin{aligned}
S_{\gamma_i \gamma_j} &= u \frac{\partial^2 f_4}{\partial \gamma_i \partial \gamma_j} + (1-u) \frac{\partial^2 f_5}{\partial \gamma_i \partial \gamma_j}, \\
S_{\gamma_i \beta_l} &= u \frac{\partial^2 f_4}{\partial \gamma_i \partial \beta_l} + (1-u) \frac{\partial^2 f_5}{\partial \gamma_i \partial \beta_l}, \\
S_{\gamma_i \phi} &= u \frac{\partial^2 f_4}{\partial \gamma_i \partial \phi} + (1-u) \frac{\partial^2 f_5}{\partial \gamma_i \partial \phi}, \\
S_{\gamma_i \rho} &= u \frac{\partial^2 f_4}{\partial \gamma_i \partial \rho} + (1-u) \frac{\partial^2 f_5}{\partial \gamma_i \partial \rho}, \\
S_{\beta_k \gamma_j} &= u \left(\frac{\partial^2 f_1}{\partial \beta_k \partial \gamma_j} + \frac{\partial^2 f_2}{\partial \beta_k \partial \gamma_j} + \frac{\partial^2 f_3}{\partial \beta_k \partial \gamma_j} + \frac{\partial^2 f_4}{\partial \beta_k \partial \gamma_j} \right), \\
S_{\beta_k \beta_l} &= u \left(\frac{\partial^2 f_1}{\partial \beta_k \partial \beta_l} + \frac{\partial^2 f_2}{\partial \beta_k \partial \beta_l} + \frac{\partial^2 f_3}{\partial \beta_k \partial \beta_l} + \frac{\partial^2 f_4}{\partial \beta_k \partial \beta_l} \right), \\
S_{\beta_k \phi} &= u \left(\frac{\partial^2 f_1}{\partial \beta_k \partial \phi} + \frac{\partial^2 f_2}{\partial \beta_k \partial \phi} + \frac{\partial^2 f_3}{\partial \beta_k \partial \phi} + \frac{\partial^2 f_4}{\partial \beta_k \partial \phi} \right), \\
S_{\beta_k \rho} &= u \left(\frac{\partial^2 f_1}{\partial \beta_k \partial \rho} + \frac{\partial^2 f_2}{\partial \beta_k \partial \rho} + \frac{\partial^2 f_3}{\partial \beta_k \partial \rho} + \frac{\partial^2 f_4}{\partial \beta_k \partial \rho} \right), \\
S_{\phi \gamma_j} &= u \left(\frac{\partial^2 f_1}{\partial \phi \partial \gamma_j} + \frac{\partial^2 f_2}{\partial \phi \partial \gamma_j} + \frac{\partial^2 f_3}{\partial \phi \partial \gamma_j} + \frac{\partial^2 f_4}{\partial \phi \partial \gamma_j} \right), \\
S_{\phi \beta_l} &= u \left(\frac{\partial^2 f_1}{\partial \phi \partial \beta_l} + \frac{\partial^2 f_2}{\partial \phi \partial \beta_l} + \frac{\partial^2 f_3}{\partial \phi \partial \beta_l} + \frac{\partial^2 f_4}{\partial \phi \partial \beta_l} \right), \\
S_{\phi \phi} &= u \left(\frac{\partial^2 f_1}{\partial \phi^2} + \frac{\partial^2 f_2}{\partial \phi^2} + \frac{\partial^2 f_3}{\partial \phi^2} + \frac{\partial^2 f_4}{\partial \phi^2} \right), \\
S_{\phi \rho} &= u \left(\frac{\partial^2 f_1}{\partial \phi \partial \rho} + \frac{\partial^2 f_2}{\partial \phi \partial \rho} + \frac{\partial^2 f_3}{\partial \phi \partial \rho} + \frac{\partial^2 f_4}{\partial \phi \partial \rho} \right), \\
S_{\rho \gamma_j} &= u \frac{\partial^2 f_4}{\partial \rho \partial \gamma_j}, \\
S_{\rho \beta_l} &= u \frac{\partial^2 f_4}{\partial \rho \partial \beta_l}, \\
S_{\rho \phi} &= u \frac{\partial^2 f_4}{\partial \rho \partial \phi}, \\
S_{\rho \rho} &= u \frac{\partial^2 f_4}{\partial \rho^2}.
\end{aligned}$$