UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Matemática, Estatística e Computação Científica

KATHERINE ANDREINA LOOR VALERIANO

Likelihood-based inference for models with censored responses using the Student-t and skew-t distribution

Inferência para modelos com respostas censurados usando a distribuição Student-t e skew-t

Campinas

2024

Katherine Andreina Loor Valeriano

Likelihood-based inference for models with censored responses using the Student-*t* and skew-*t* distribution

Inferência para modelos com respostas censurados usando a distribuição Student-*t* e skew-*t*

Tese apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutora em Estatística.

Thesis presented to the Institute of Mathematics, Statistics and Scientific Computing of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Statistics.

Supervisor: Larissa Avila Matos

Co-supervisor: Christian Eduardo Galarza Morales

Este trabalho corresponde à versão final da Tese defendida pela aluna Katherine Andreina Loor Valeriano e orientada pela Profa. Dra. Larissa Avila Matos.

Campinas

2024

Ficha catalográfica
Universidade Estadual de Campinas (UNICAMP)
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

Loor Valeriano, Katherine Andreina, 1990-
L873L Likelihood-based inference for models with censored responses using the Student-t and skew-t distribution / Katherine Andreina Loor Valeriano. – Campinas, SP : [s.n.], 2024.

Orientador: Larissa Avila Matos.
Coorientador: Christian Eduardo Galarza Morales.
Tese (doutorado) – Universidade Estadual de Campinas (UNICAMP), Instituto de Matemática, Estatística e Computação Científica.

1. Observações censuradas (Estatística). 2. Estimativa de parâmetro. 3. Algoritmos de esperança-maximização. I. Matos, Larissa Avila, 1987-. II. Galarza Morales, Christian Eduardo, 1988-. III. Universidade Estadual de Campinas (UNICAMP). Instituto de Matemática, Estatística e Computação Científica. IV. Título.

Informações Complementares

Título em outro idioma: Inferência para modelos com respostas censuradas usando a distribuição Student-t e skew-t
Palavras-chave em inglês:
Censored observations (Statistics)
Parameter estimation
EM algorithm
Área de concentração: Estatística
Titulação: Doutora em Estatística
Banca examinadora:
Larisa Avila Matos
Filidor Edilfonso Vilca Labra
Mariana Rodrigues Motta
Celso Rômulo Barbosa Cabral
Marcos Oliveira Prates
Data de defesa: 13-08-2024
Programa de Pós-Graduação: Estatística

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0001-6388-4753
- Currículo Lattes do autor: http://lattes.cnpq.br/8992394367424165

Tese de Doutorado defendida em 13 de agosto de 2024 e aprovada

pela banca examinadora composta pelos Profs. Drs.

Prof(a). Dr(a). LARISSA AVILA MATOS

Prof(a). Dr(a). FILIDOR EDILFONSO VILCA LABRA

Prof(a). Dr(a). MARIANA RODRIGUES MOTTA

Prof(a). Dr(a). CELSO RÔMULO BARBOSA CABRAL

Prof(a). Dr(a). MARCOS OLIVEIRA PRATES

A Ata da Defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós-Graduação do Instituto de Matemática, Estatística e Computação Científica.

In loving memory of my grandfather, Nelson Loor Martínez.

Acknowledgements

I would like to express my deepest gratitude to all those who have supported me throughout this journey.

First and foremost, I acknowledge the grace of God for His blessings every day of my life and for providing me the opportunity to pursue my studies.

I sincerely thank my supervisor, Larissa Matos, for her invaluable help, support, guidance, and patience throughout this process. I also extend my gratitude to my co-supervisor, Christian Galarza, whose patience and insightful suggestions have greatly enhanced this work. His constructive feedback has been crucial in shaping the direction and quality of my research. I must also express my appreciation to Fernanda Schumacher and Victor Hugo Lachos for their assistance and support in some of the chapters presented in this thesis.

My heartfelt thanks go to my beloved fiancé, Lourival, for his moral and emotional support during the completion of this work. His encouraging words and practical advice have kept me motivated and focused. I also owe immense gratitude to my grandmother, Elsa, for her support throughout my academic journey, from my first day of school to this moment, including her comforting presence during the most challenging times of my life. Additionally, I must extend a heartfelt thank you to my grandfather, Nelson, for his love, encouragement, and support. Although he is no longer with us, he would be eternally proud of this significant achievement. Lastly, I wish to thank my parents, siblings, and uncles for their constant support and inspiring words of faith and love.

Additionally, I would like to thank my professors, friends, and colleagues who have provided help directly or indirectly. I also appreciate the IMECC women's soccer team, the Lovelaces, who have been a constant source of inspiration and motivation.

Finally, I thank everyone who contributed to the successful completion of my thesis. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Resumo

O estudo de modelos nos quais a variável de resposta está sujeita a limites de detecção, tem sido de interesse em muitas áreas da estatística. Este tipo de dados surgem frequentemente em monitoramento ambiental, medicina, economia, agronomia e biologia. A maioria dos modelos existentes na literatura para lidar com dados censurados assume uma distribuição normal para a variável de resposta, e essa suposição pode ser irrealista na presença de desvios da normalidade ou de outliers. Neste trabalho, propomos uma série de modelos que consideram distribuições assimétricas e de caudas pesadas, como as distribuições Student-t e skew-t, para lidar com observações censuradas e/ou faltantes na variável de resposta.

Os parâmetros dos modelos são estimados utilizando o algoritmo Expectation-Maximization (EM) (Dempster *et al.*, 1977), um método amplamente utilizado para aproximar iterativamente as estimativas de máxima verossimilhança (ML). Este algoritmo exige o cálculo de algumas esperanças condicionais. Em nossos modelos, isso inclui os dois primeiros momentos das distribuições Student-t, skew-t e extended skew-t. Para calcular os momentos da distribuição Student-t, desenvolvemos um método baseado na integração de Monte Carlo, complementado por resultados derivados da esperança condicional (veja, por exemplo, Galarza *et al.*, 2021c). Além disso, quando as esperanças condicionais não podem ser derivadas de forma explícita, empregamos uma versão de aproximação estocástica do algoritmo EM, conhecido como algoritmo SAEM (Delyon *et al.*, 1999), para a estimação dos parâmetros. Para cada modelo, também fornecemos procedimentos para aproximar o erro padrão das estimativas e expressões para prever observações futuras. As propriedades assintóticas e a robustez das estimativas são demonstradas através de estudos de simulação, e aplicações em conjuntos de dados reais são apresentadas para esses modelos.

Palavras-chave: Observações censuradas. Família de distribuições elípticas. Algoritmo EM. Distribuições de caudas pesadas. Distribuições assimétricas. Distribuições truncadas.

Abstract

The study of models where the variable of interest is subjected to threshold values below, above, or both has been the scope of many areas of the statistic. Such data frequently arise in environmental monitoring, medicine, economics, agronomy, and biology. While most of the models to deal with censored data in the literature assume a normal distribution for the response variable, this assumption can be unrealistic in the presence of deviations from normality or outliers. In this work, we propose a series of models considering asymmetric and heavy-tailed distributions, such as the Student-t and skew-t distributions, to handle censored and missing observations in the response variable.

The parameters of the models are estimated using the Expectation-Maximization (EM) algorithm (Dempster *et al.*, 1977), a widely used method for iteratively approximating the maximum likelihood (ML) estimates. This algorithm requires the computation of certain conditional expectations, specifically for our models, the first two moments from the Student-t, skew-t, and extended skew-t distributions. To compute these moments from the Student-t distribution, we develop a method based on Monte Carlo integration and results derived from the conditional distribution (see, for instance, Galarza *et al.*, 2021c). Additionally, when conditional expectations cannot be derived in a closed form, we employ a variation of the EM algorithm, known as the Stochastic Approximation EM (SAEM) algorithm (Delyon *et al.*, 1999), for parameter estimation. For each model, we also provide procedures to approximate the standard error of the estimates and expressions for predicting future observations. The asymptotic properties and robustness of the estimates are demonstrated through simulation studies, and applications to real datasets are presented for these models.

Keywords: Censored observations. Elliptical family of distributions. EM algorithm. Heavy-tailed distributions. Skewed distributions. Truncated distributions.

List of Figures

Figure 1 – Missouri data. Level of TCDD observed on each location, where ∘ represents an observed value and • represents a left-censored value. . . 28
Figure 2 – Ammonia-concentration data. Time series plot of the $NH_3 - N$ (black line) and the detection limit (red dotted line). 29
Figure 3 – Ammonia-nitrogen concentration data. Boxplot for the dissolved oxygen (mg/L), pH, and temperature ($^{\circ}C$) by month. 29
Figure 4 – Phophorus concentration data. Time series plots of P (black line) and the limit of detection (red dotted line). Censored observations are marked by circles. 30
Figure 5 – VDEQ data. Histograms of the original five trace metal concentrations: Cu, Pb, Zn (in μg/L of water), Ca, and Mg (in mg/L of water). The red dashed lines represent the detection limits. 31
Figure 6 – Stellar data. Scatter plot for the logarithm of beryllium (top) and lithium (bottom) abundances against temperature. 32
Figure 7 – A5055 data. Trajectories of $\log_{10}$RNA for 44 HIV-1 infected patients who were randomized in two IDV-RTV regimens. The blue line denotes the mean of the viral load per week, and the red dashed line indicates the detection limit. 33
Figure 8 – Slice sampling algorithm for univariate random variables. 39
Figure 9 – Boxplot based on 100 estimates of the truncated mean. The red line denotes the median of `TT.moment` estimates. 47
Figure 10 – Boxplot based on 100 estimates of the variance-covariance elements. The red line represents the median of the estimates obtained from function `TT.moment`. 48
Figure 11 – Simulation I. Boxplot of the estimates by level of censoring. 59
Figure 12 – Simulation II. MSE of the MCEM (left) and SAEM (right) estimates considering different sample sizes and levels of censoring. 63
Figure 13 – Missouri data. Convergence of the parameter estimates via EM, MCEM, and SAEM algorithm. 64
Figure 14 – Simulation I. Boxplot of the estimates for CARt(2) model by sample size and detection limit. 75
Figure 15 – Simulation I. MSE of the estimates for the CARt(2) model by sample size and detection limit. 76
Figure 16 – Simulation I. Plots of quantile residuals for a sample of size $n = 600$ generated from the CARt(2) model considering different levels of left censoring. 77

Figure 17 – Simulation I. Plots of the residuals for a sample of size $n = 600$ generated from the CAR$t(2)$ model and fitting a model with normal innovations, considering different levels of left censoring. 78
Figure 18 – Simulation II. Boxplot of the estimates obtained from CAR(2) and CAR$t(2)$ model based on 300 MC samples of size $n = 100$. 81
Figure 19 – Square-root of NH_3-N (μg/L). Quantile residuals for the CAR(1) and CAR$t(1)$ models. 83
Figure 20 – Square-root of NH_3-N (μg/L). Observed (black solid line) and imputed values considering Student-t (pink line) and normal (light blue line) innovations. 84
Figure 21 – Square-root of NH_3-N (μg/L). Convergence of the parameter estimates for CAR$t(1)$ model. 84
Figure 22 – Simulation I. MSE of parameter estimates under the ST-CR model, based on 500 Monte Carlo samples simulated from the skew-t distribution considering different sample sizes and types of censoring. 103
Figure 23 – Simulation III. Mean bias ± 1 standard deviation for the estimates of $\boldsymbol{\beta}$ obtained under the SN-CR and ST-CR models, based on 500 MC samples simulated from the SN distribution considering two cases: (a) without censoring and (b) average 15% of right-censored observations. . 106
Figure 24 – Simulation III. Mean bias ± 1 standard deviation for the estimates of $\boldsymbol{\beta}$ obtained under the SN-CR and ST-CR models, based on 500 MC samples simulated from the SSL distribution with $\nu = 1.15$ considering two cases: (a) without censoring and (b) average 15% of right-censored observations. 107
Figure 25 – Simulation IV. MSE of $\boldsymbol{\beta}$ estimated from the SN-CR and ST-CR models, based on 500 MC samples of size $n = 300$ simulated from the bivariate SN distribution considering different perturbation levels and censoring proportions. 108
Figure 26 – Simulation IV. Boxplot of the $\boldsymbol{\beta}$ estimates obtained from the SN-CR and ST-CR model, based on 500 MC samples of size $n = 300$ simulated from the bivariate SN distribution considering two levels of censoring. . 109
Figure 27 – Simulation IV. MSE of the estimates for the variance-covariance elements obtained from the SN-CR and ST-CR model, based on 500 MC samples of size $n = 300$ simulated from the bivariate SN distribution, considering different levels of perturbation. 109

Figure 28 – Stellar data. Scatter plot for the logarithm of beryllium (top) and lithium (bottom) abundances against temperature together with the regression lines fitted to points with no treatment, i.e., considering censored data as observed (dashed line) and under our proposed ST-CR model (solid line). Plots for non-planets-hosting stars are shown in the left panels. . 114
Figure 29 – Simulation I. Boxplot of the estimates for the LMEC-ST model considering different sample sizes and types of censoring: without censoring (No), left-censored (Left), and right-censored (Right) observations. The red line denotes the true value. 128
Figure 30 – Simulation I. MSE of the estimates for the LMEC-ST model considering different sample sizes and types of censoring. 129
Figure 31 – Simulation II. MSE estimated from the LMEC-SN and LMEC-ST models, based on 300 MC samples of size $n = 300$ simulated from the SN distribution considering 1% of perturbed observations and two censoring cases: without censoring and an average of 10% right-censored observations. 131
Figure 32 – Simulation III. Boxplot of the estimates for the regression coefficients obtained after fitting the LMEC-SN (SN) and LMEC-ST (ST) models to data simulated from different distributions incorporating 10% of left-censored observations. 133
Figure 33 – A5055 data. (Left panel) Individual profiles for HIV viral load are displayed on a $\log_{10}$ scale. (Right panel) Boxplot of the random effects b_{1i} predicted from the t-LMEC model. 133
Figure 34 – A5055 data. Observed viral loads on a $\log_{10}$ scale (black solid line) for six subjects, alongside fitted trajectories for the LMEC-ST (dark grey) and LMEC-t (light grey) models, both incorporating the AR(1) structure. The final two observations were reserved to assess the predictive capabilities of the models. 136
Figure 35 – Trace plots of the evolution of the MC estimates for the mean and variance-covariance elements of $\mathbf{X} \mid (\mathbf{X} \in A)$ under scenarios a), b), and c). The red dashed line represents the true estimated value computed using numerical methods 151
Figure 36 – Scatterplot and marginal histograms for the $n = 10^4$ observations sampled for some bivariate truncated elliptical distributions. 157
Figure 37 – Sample autocorrelation plots of X_1 and X_2 sampled from the bivariate truncated elliptical distributions in Figure 36. 158
Figure 38 – Simulation I. Mean bias and ± 1 standard deviation of MCEM estimates based on 300 MC samples. 162

Figure 39 – Simulation II. Boxplot of the MCEM estimates considering different sample sizes and censoring proportions. Red lines represent the true parameter values. 163
Figure 40 – Simulation II. Boxplot of the SAEM estimates considering different sample sizes and censoring proportions. Red lines represent the true parameter values. 163
Figure 41 – Simulation I. Boxplot of the estimates of ν in the CARt(2) model by sample size and detection limits. 169
Figure 42 – Quantile residuals for a model with normally distributed independent errors fitted to the ammonia-nitrogen concentration dataset. 170
Figure 43 – Phosphorus concentration. Plots of residuals for the CAR(1) and CARt(1) models. 172
Figure 44 – Phosphorus concentration. Observed (black solid line) and predicted values considering Student-t (pink line) and normal (light blue line) innovations. Black dashed lines represent the period with missing observations. 172
Figure 45 – Phosphorus concentration. Convergence of the SAEM parameter estimates for the CARt(1) model. 173
Figure 46 – Simulation I. Boxplot of the estimates obtained under the ST-CR model considering different sample sizes and types of censoring based on 500 Monte Carlo samples. The red line denotes the true parameter value. . 178

List of Tables

Table 1 – Median of the CPU time (in seconds) and relative time (R. Time) considering $n = 10^5$ with *thinning* $= 3$ as the reference method, based on 100 simulations. 49

Table 2 – Simulation I. Summary statistics of the MCEM estimates based on 300 samples of size $n = 200$, considering different censoring proportions and number of observations (L) for the MC method. 58

Table 3 – Simulation II. Summary statistics of the MCEM estimates based on 300 samples of sizes $n = 100, 300$, and 600 with different censoring proportions. 61

Table 4 – Simulation II. Summary statistics of the SAEM estimates based on 300 samples of sizes $n = 100, 300$, and 600 with different censoring proportions. 62

Table 5 – Missouri data. ML estimates and information criteria (AIC and BIC) were obtained through MCEM, SAEM, and EM algorithms considering the exponential correlation function. 63

Table 6 – Simulation I. Summary statistics of parameter estimates for the CARt(2) model based on 300 samples of sizes $n = 100, 300, 600$, and different levels of censoring. 74

Table 7 – Simulation II. Mean of the estimates for CAR(2) and CARt(2) model based on 300 MC samples of size $n = 100$ for different levels of perturbation. 80

Table 8 – Square-root of NH_3-N (μg/L). Parameter estimates and their standard errors (SE) for the CAR(p) and CARt(p) model, for $p = 1$ and 2. 83

Table 9 – Simulation I. Summary statistics based on 500 MC samples of size $n = 50, 100, 200$, and 400 without censoring. MC-AV, MC-MD, and MC-SD refer to the mean, median, and standard deviation of the estimates, respectively. IM-SE denotes the average of standard errors obtained as described in Subsection 5.3.5. 101

Table 10 – Simulation I. Results based on 500 MC samples of size $n = 50, 100, 200$, and 400 with an average 15% of left-censored observations. 102

Table 11 – Simulation I. Results based on 500 MC samples of size $n = 50, 100, 200$, and 400 with an average 15% of right-censored observations. 103

Table 12 – Simulation II. Mean of the estimates (MC-AV) and average of the approximated standard errors (IM-SE) based on 500 MC samples of size 300 generated from the bivariate skew-t distribution with skewness parameter $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^\top$ and average 15% of left-censored observations. 104

Table 13 – Simulation III. Results based on 500 MC samples of size $n = 300$ when generating data from the SN distribution considering different censoring levels (c) and estimating the models SN-CR and ST-CR. MC-AV and IM-SE denote to the mean of the estimates and the average of the standard error, respectively. 106
Table 14 – Simulation III. Summary statistics based on 500 MC samples of size $n = 300$ when generating data from the SSL distribution with $\nu = 1.15$ considering different censoring levels (c) and estimating the models SN-CR and ST-CR. MC-AV and IM-SE denote to the mean of the estimates and the average of the standard error, respectively. 107
Table 15 – VDEQ data. ML estimate, approximated standard error (SE), and model comparison criteria (AIC and BIC) from fitting the ST-CR, SN-CR, T-CR, and N-CR to the trace metals concentration dataset. 111
Table 16 – log-VDEQ data. ML estimate, approximated standard error (SE), and model selection criteria (AIC and BIC) from fitting the ST-CR, SN-CR, T-CR, and N-CR to the logarithmic transformation of the trace metals concentration dataset. 111
Table 17 – Stellar data. ML estimate, approximated standard error (SE), and model selection criteria (AIC and BIC) from fitting the ST-CR, SN-CR, T-CR, and N-CR to the natural logarithm of beryllium and lithium. 113
Table 18 – Simulation I. Summary statistics are presented for 300 Monte Carlo (MC) samples of sizes $n = 50, 100, 200$, and 400, each with an average of 10% left-censored (upper table) and right-censored (lower table) observations, simulated from the LMEC-ST model. The statistics MC-AV, MC-MD, and MC-SD refer to the mean, median, and standard deviation of the estimates, respectively. IM-SE denotes the average of the standard errors as detailed in Subsection 6.2.4. The true parameter values are provided in parentheses for reference. 127
Table 19 – Simulation II. Mean of the estimates obtained after fitting the LMEC-SN (SN) and LMEC-ST (ST) models, based on 300 MC samples of size $n = 300$ simulated from the SN distribution considering 10% of right-censored and 1% of perturbed observations. 130
Table 20 – Simulation III. Mean of the estimates obtained after fitting the LMEC-SN (SN) and LMEC-ST (ST) models, based on 300 MC samples of size $n = 300$ simulated from different distributions considering 10% of left-censored observations. 132
Table 21 – A5055 data. Information criteria were obtained after fitting the LMEC-ST model with different correlation structures. 134

Table 22 – A5055 data. Parameter estimates (Est.) and standard error (SE) obtained after fitting some LMEC models with AR(1) structure. 135
Table 23 – Median of the CPU time (in seconds) based on 100 simulations. 155
Table 24 – Phosphorus concentration. Parameter estimates and their standard errors (SE) for the CARt(1) and CAR(1) models. 171
Table 25 – Simulation II. Summary statistics based on 500 MC samples of size 300 generated from the bivariate skew-t distribution with skewness parameter $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^\top$ and average 15% of left-censored observations. MC-AV denotes the mean of the estimates, and IM-SE refers to the average number of standard errors. 177

Contents

Introduction **19**

1 Preliminaries **23**

1.1 The EM, MCEM, and SAEM algorithms 23

1.1.1 The EM algorithm 23

1.1.2 The MCEM algorithm 24

1.1.3 The SAEM algorithm 24

1.2 Standard error approximation 25

1.3 Case studies 27

1.3.1 Missouri dioxin contamination data 27

1.3.2 Ammonia-nitrogen concentration data 28

1.3.3 Phosphorus concentration data 29

1.3.4 Trace metals in freshwater streams data 30

1.3.5 Stellar abundances data 31

1.3.6 A5055 data 32

2 Moments and random number generation for the TE family of distributions **34**

2.1 Introduction 34

2.2 Preliminaries 37

2.2.1 Elliptical family of distributions 37

2.2.2 Truncated elliptical family of distributions 38

2.2.3 Slice sampling algorithm 38

2.3 Sampling from the truncated elliptical family of distributions 39

2.4 Moments of truncated multivariate elliptical distributions 42

2.4.1 Mean and variance for the truncated elliptical distributions 44

2.4.2 Numerical examples 46

2.4.2.1 Simulation study I 46

2.4.2.2 Simulation study II 48

2.5 Remarks 50

3 Spatial censored linear model **51**

3.1 Introduction 51

3.2 Linear spatial model 52

3.3 Model formulation and parameter estimation 53

3.3.1 Spatial censored linear model 53

3.3.2 The likelihood function 53

3.3.3 Parameter estimation 54

3.3.4 Standard error approximation 56

3.4 Simulation study 57

3.4.1 Simulation study I: MCEM estimates . . . 57
3.4.2 Simulation study II: Asymptotic properties of the estimates . . . 59
3.5 Missouri dioxin contamination data . . . 60
3.6 Remarks . . . 65
4 Censored autoregressive regression models with Student-*t* innovations . . . 66
4.1 Introduction . . . 66
4.2 The censored AR regression model of order p . . . 67
4.2.1 The log-likelihood function . . . 68
4.2.2 The SAEM algorithm for ML Estimation . . . 69
4.2.3 Standard error approximation . . . 71
4.2.4 Prediction . . . 72
4.3 Simulation Study . . . 73
4.3.1 Simulation study I . . . 73
4.3.1.1 Asymptotic properties . . . 73
4.3.1.2 Residual Analysis . . . 76
4.3.2 Simulation study II: Robustness of the estimators . . . 79
4.4 The ammonia-nitrogen concentration data . . . 81
4.5 Remarks . . . 84
5 Likelihood-based inference for the multivariate skew-*t* regression . . . 86
5.1 Introduction . . . 86
5.2 Preliminaries . . . 87
5.2.1 The multivariate skew-t distribution . . . 87
5.2.2 The multivariate extended skew-t distribution . . . 88
5.2.3 The truncated extended skew-t distribution . . . 90
5.3 The ST censored regression model for multiple outcomes . . . 91
5.3.1 The likelihood function . . . 91
5.3.2 Parameter estimation via EM-type algorithm . . . 92
5.3.3 Details for the expectations in the EM algorithm . . . 94
5.3.4 Initial values and stopping criterion . . . 98
5.3.5 Standard error approximation . . . 98
5.3.6 Imputation of censored components . . . 99
5.4 Simulation Studies . . . 100
5.4.1 Simulation study I: Asymptotic properties . . . 100
5.4.2 Simulation study II: Effect of skewness on left-censored data . . . 104
5.4.3 Simulation study III: Model misspecification . . . 105
5.4.4 Simulation study IV: Robustness of the estimators . . . 108
5.5 Applications . . . 110
5.5.1 Trace metals in freshwater streams across the Commonwealth of Virginia . . . 110

5.5.2 Stellar abundances data . . . 112
5.6 Remarks . . . 113
6 Likelihood-based inference for the censored linear mixed-effects skew-*t* model 115
6.1 Introduction . . . 115
6.2 The linear mixed-effects skew-t model for censored responses . . . 116
6.2.1 The linear mixed-effects model . . . 116
6.2.2 The likelihood function . . . 118
6.2.3 Parameter estimation via EM-type algorithm . . . 119
6.2.4 Standard error approximation . . . 122
6.2.5 Estimation of random effects . . . 123
6.2.6 Prediction of future observations . . . 124
6.3 Simulation studies . . . 125
6.3.1 Simulation I. Asymptotic properties . . . 125
6.3.2 Simulation II. Robustness of the estimators . . . 129
6.3.3 Simulation III. Model misspecification . . . 131
6.4 Application: A5055 data . . . 133
6.5 Remarks . . . 136
7 Final Considerations . . . 137
7.1 Published papers . . . 137
7.2 R packages . . . 137
7.3 Conclusions . . . 138
7.4 Future research . . . 139

Bibliography . . . 140
APPENDIX A CHAPTER 2 . . . 149
APPENDIX B CHAPTER 3 . . . 160
APPENDIX C CHAPTER 4 . . . 164
APPENDIX D CHAPTER 5 . . . 174
APPENDIX E CHAPTER 6 . . . 179

Introduction

The study of models where the variable of interest is subjected to certain threshold values, whether below, above, or both, has been the scope of many statistical areas. This data type frequently appears in environmental monitoring, medicine, economics, agronomy, and biology, among others. For instance, Lachos *et al.* (2017) investigated the level of contamination by dioxin (2,3,7,8-tetrachlorodibenzo-p-dioxin or TCDD) in a Missouri River, USA, using a spatially censored linear model. This dataset exhibited various detection limits depending on the measurement instruments employed. Similarly, Matos *et al.* (2016) proposed linear and nonlinear censored mixed-effects models to analyze the human immunodeficiency virus (HIV) behavior from UTI data and clinical trial ACTG 315 data, respectively, focusing on the frequently left-censored quantification of HIV-1 RNA viral load.

Most models in the literature dealing with censored data assume a normal distribution for the response variable. However, this assumption can be unrealistic when deviations from normality or outliers exist. The Student-t model, with its robust approach, offers a practical alternative. For instance, Matos *et al.* (2013b) and Lachos *et al.* (2019) demonstrated that the Student-t linear mixed-effects censored model outperforms the normal model in analyzing UTI data. This study pertains to the health of 72 perinatal HIV-infected children.

Moreover, datasets often exhibit skewness alongside censoring. Some authors have typically analyzed such data using a symmetric distribution, often after a transformation. Galarza *et al.* (2022b) proposed a multivariate regression model with a skew-normal distribution for the error term to circumvent this. They showed that, without transforming the response, a model with asymmetrical errors better fits the dissolved trace metals data. Also, using the skew-normal distribution, Mattos *et al.* (2022b) developed a linear mixed-effects model for censored responses, concluding that the aforementioned UTI data exhibit skewness.

Additionally, one of the most popular algorithms for handling missing and partially observed data is the Expectation-Maximization (EM) (Dempster *et al.*, 1977) algorithm, which is frequently employed because of its facility to treat censored data. Typically, this algorithm requires the computation of conditional truncated moments, usually the first two. Consequently, developing more efficient methods for calculating the moments of truncated distributions has been an area of significant interest. For example, recently, Kan & Robotti (2017) proposed a recursive approach for computing arbitrary-order product moments of truncated multivariate normal (TMVN) distributions, which is

implemented in the R library `MomTrunc` (Galarza *et al.*, 2021a). In contrast, Galarza *et al.* (2021c) developed a recurrence approach to compute arbitrary-order product moments of folded and truncated multivariate t (TMVT) distributions. Meanwhile, Galarza *et al.* (2022a) derived the moments for a doubly truncated selection elliptical class of distributions, including some multivariate asymmetric versions of elliptical distributions.

Other algorithms employed to handle partially observed data include the Stochastic Approximation EM (SAEM) (Delyon *et al.*, 1999) and Monte Carlo EM (MCEM) (Wei & Tanner, 1990), which replace conditional expectations with approximations requiring random draws from a truncated distribution. Hence, various methods have been developed to generate random samples from truncated distributions, with rejection sampling being the most common technique. However, this method can be inefficient, especially when the truncation interval is small. To address this, Kotecha & Djuric (1999) proposed an approach based on the Gibbs sampling algorithm for generating vectors from TMVN distributions, and Ho *et al.* (2012) suggested the slice sampling algorithm for generating random observations from the TMVT distributions.

Therefore, firstly, we develop a general method to simulate from any truncated multivariate elliptical distribution with a strictly decreasing density generating function (dgf) as an extension of the algorithm proposed by Ho *et al.* (2012), improving the computational time needed for sampling. Using conditional expectation properties, we also propose an algorithm to approximate the moments of the most common distribution of this class, such as the truncated multivariate normal, Student-t, slash, contaminated normal, and Pearson VII distributions. This method requires less CPU time when compared with the existing ones since it deals with the truncated and non-truncated parts of the vector separately. Both methods will be used in several applications related to a linear spatial model with censored responses, censored regression models considering autoregressive errors with Student-t innovations, multivariate censored regression models with errors following skew-t distribution, and linear mixed-effect models with censored response using skew-t distribution.

This thesis unfolds through chapters that apply various models and techniques to handle censored data. The organization of the thesis is as follows:

Chapter 1: This chapter reviews the background material, including definitions and methodologies. It also describes the datasets used throughout the thesis.

Chapter 2: We propose an algorithm to generate random numbers from any member of the truncated multivariate elliptical family of distributions with a strictly decreasing density generating function based on the slice sampling algorithm (Neal, 2003); this extends the algorithm initially proposed by Ho *et al.* (2012). We also provide a faster approach to approximate the first and the second moments for specific truncated multivariate elliptical distributions using Monte Carlo (MC) integration for the truncated

partition and explicit expressions for the non-truncated part (Galarza *et al.*, 2022a). Methods are accessible in the R library `relliptical` (Valeriano *et al.*, 2022).

Chapter 3: We compare the EM, SAEM, and MCEM estimates for the censored spatial linear model previously discussed by Lachos *et al.* (2017) and Ordoñez *et al.* (2018) with parameters estimated via the SAEM algorithm. Additionally, we provide a methodology to approximate the standard error of the estimates using the squared root of the inverse of the observed information matrix, following the method proposed by Louis (1982). All methods are available in the R package `RcppCensSpatial` (Valeriano *et al.*, 2021a), which utilizes algorithms from Chapter 2 to generate random numbers and approximate moments from the TMVN distribution.

Chapter 4: We develop an SAEM algorithm to estimate the parameters of a censored linear regression model where regression errors are autocorrelated, and the innovations follow a Student-t distribution. Several simulation studies are conducted to examine the asymptotic properties and robustness of the estimates. The methods are illustrated using two datasets with left-censored and missing observations. The codes are available in the package `ARCensReg` (Schumacher *et al.*, 2016).

Chapter 5: We propose a multivariate linear model with censored responses considering the multivariate skew-t distribution for the error term. This work addresses the need for asymmetric and heavy-tailed distributions in regression error. The proposed EM algorithm for maximum likelihood estimation employs closed-form expressions at the E-step based on formulas for the mean and variance of truncated multivariate extended skew-t (EST), skew-t (ST), and Student-t distributions. Moments from the latter distribution are computed using the method from Chapter 2.

Chapter 6: This chapter extends the work of Mattos *et al.* (2022b) by simultaneously considering asymmetric and heavy tails in the random effect distribution. We develop a linear mixed-effects model for censored responses using the skew-t distribution. Several simulation studies are conducted to evaluate the asymptotic properties and robustness of the parameter estimates obtained via the EM-type algorithm. The method is also applied to analyze the RNA viral load in HIV-1 infected patients, with the dataset being left-censored at the threshold $\log_{10}(50)$. The codes will be available in the package `skewlmm` (Schumacher *et al.*, 2023).

Chapter 7: We present final remarks and further research related to this thesis.

The majority of the material in this thesis is based on original publications. Below is a list of the chapters primarily based on those publications:

Chapter 2 and 3.
Valeriano, K. A., Galarza, C. E., and Matos, L. A. (2023). Moments and random number generation for the truncated elliptical family of distributions. *Statistics and Computing*, DOI: 10.1007/s11222-022-10200-4. *Reproduced with permission from Springer Nature.*

Chapter 4.
Valeriano, K. A., Schumacher, F. L., Galarza, C. E., and Matos, L. A. (2024). Censored autoregressive regression models with Student-t innovations. *Canadian Journal of Statistics*, DOI: 10.1002/cjs.11804.

Chapter 5.
Valeriano, K. A., Galarza, C. E., Matos, L. A., and Lachos, V. H. (2023). Likelihood-based inference for the multivariate skew-t regression with censored or missing responses. *Journal of Multivariate Analysis*, DOI: 10.1016/j.jmva.2023.105174.

1 Preliminaries

This chapter provides an overview of the algorithms employed throughout this thesis to derive the maximum likelihood (ML) estimates for problems involving partially observed data. Additionally, we introduce essential notations and basic definitions used in our analyses.

A random variable with a Gamma distribution is denoted by $Gamma(\alpha, \beta)$, where $\alpha > 0$ is the shape parameter and $\beta > 0$ is the rate parameter. A random variable X uniformly distributed over the interval (a, b) is represented as $X \sim U(a, b)$. The notation $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ specifies a p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. Similarly, $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ denotes the p-variate Student-t distribution with location parameter vector $\boldsymbol{\mu} \in \mathbb{R}^p$, a positive definite scale matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, and $\nu > 0$ degrees of freedom. The probability density function (pdf) and cumulative distribution function (cdf) for this distribution are represented by $t_p(.; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ and $T_p(.; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, respectively. When $p = 1$, the index p is omitted. If $\mu = 0$ and $\sigma^2 = 1$ (the standard case), the notations $t(.; \nu)$ and $T(.; \nu)$ are used for the pdf and cdf, respectively.

Additionally, the term *ind* means independent, and *iid* stands for independent and identically distributed. The symbol $\mathbf{I}_p$ denotes a $p \times p$ identity matrix, $\mathbf{A}^\top$ represents the transpose of matrix $\mathbf{A}$, $|\mathbf{A}|$ denotes the determinant of a square matrix $\mathbf{A}$, and $\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}\mathrm{d}x$ is the gamma function evaluated at $a > 0$.

1.1 The EM, MCEM, and SAEM algorithms

1.1.1 The EM algorithm

The expectation-maximization (EM) algorithm was first introduced by Dempster *et al.* (1977) and provides a general approach for iteratively computing maximum likelihood (ML) estimates in the presence of incomplete data. Its name comes from the fact that each algorithm iteration consists of an expectation step followed by a maximization step. The main features of the EM algorithm are the ease of implementation and the stability of monotone convergence.

Let $\boldsymbol{\theta}$ be the parameter vector and $\mathbf{y}_c = (\mathbf{y}_o^\top, \mathbf{y}_m^\top)^\top$ denote the complete data, where $\mathbf{y}_o$ and $\mathbf{y}_m$ are the observed and missing/censored (incomplete) data, respectively. The EM algorithm maximizes the complete-data log-likelihood function $\ell_c(\boldsymbol{\theta}; \mathbf{y}_c)$ at each iteration until convergence. The algorithm proceeds as follows:

- **E-step**: Compute the conditional expectation of the complete log-likelihood function

$Q_k(\boldsymbol{\theta}) = \mathbb{E}\left[\ell_c(\boldsymbol{\theta}; \mathbf{y}_c) \mid \mathbf{y}_o, \widehat{\boldsymbol{\theta}}^{(k)}\right]$, where $\widehat{\boldsymbol{\theta}}^{(k)}$ is the estimate of $\boldsymbol{\theta}$ at the kth iteration.

- **M-step**: Maximize $Q_k(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ to update the estimate to $\widehat{\boldsymbol{\theta}}^{(k+1)}$.

This process is iterated until some distance between two successive evaluations of the actual log-likelihood function becomes small enough.

Although the EM algorithm is a powerful tool when the analytical expressions required by the E-steps have a closed form, it becomes a problem when the analytical expressions cannot be evaluated. The required expectations may be approximated using Monte Carlo (MC) integration techniques in such cases.

1.1.2 The MCEM algorithm

The Monte Carlo EM (MCEM) algorithm is a variation of the standard EM algorithm that incorporates Monte Carlo methods to approximate the expectation in the computation of $Q_k(\boldsymbol{\theta})$. This approach is based on a large number of independent simulations of the missing data (Wei & Tanner, 1990). In the MCEM algorithm, the E-step is replaced by the following two steps:

1. Simulation: Generate L samples of the missing data $\mathbf{y}_m^{(l)}$, $l = 1, \ldots, L$ from the conditional distribution $f(\mathbf{y}_m \mid \widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_o)$.

2. Approximation: Update $Q_k(\boldsymbol{\theta})$ by

$$\widehat{Q}_k(\boldsymbol{\theta}) = \frac{1}{L} \sum_{l=1}^{L} \ell_c(\boldsymbol{\theta}; \mathbf{y}_m^{(l)}, \mathbf{y}_o). \tag{1.1}$$

If $L = 1$, the algorithm reduces to the stochastic EM algorithm (Celeux, 1985). On the other hand, using large values of L leads to more accurate estimates, but the algorithm becomes slow (Booth & Hobert, 1999). Hence, Wei & Tanner (1990) suggested using smaller values of L in the initial iterations when the parameter estimates are relatively far from their true values, and gradually increasing L as the iterations increase. However, the MCEM algorithm may not converge in the same manner as the conventional EM algorithm. Typically, the estimates $\widehat{\boldsymbol{\theta}}$ keep varying around the maximum, and the variability depends on L. One strategy to mitigate this issue is to average the parameter estimates over the final iterations of the algorithm.

1.1.3 The SAEM algorithm

As an alternative to the computationally intensive MCEM algorithm, Delyon *et al.* (1999) proposed the Stochastic Approximation EM (SAEM) algorithm, which modifies the E-step to include both a simulation step and an integration step through stochastic

approximation while leaving the maximization step of the EM algorithm unchanged. This adaptation, noted for its robust theoretical properties, accurately estimates the true parameters and converges to the global maximum under general conditions. The E-step of the SAEM algorithm is structured as follows:

1. Simulation: Draw M samples of the missing data, $\mathbf{y}_m^{(l,k)}, l = 1, \ldots, M$, from the conditional distribution $f(\mathbf{y}_m \,|\, \widehat{\boldsymbol{\theta}}^{(k)}, \mathbf{y}_o)$.

2. Stochastic approximation: Update $Q_k(\boldsymbol{\theta})$ according to

$$\widehat{Q}_k(\boldsymbol{\theta}) = \widehat{Q}_{k-1}(\boldsymbol{\theta}) + \delta_k \left(\frac{1}{M} \sum_{l=1}^{M} \ell_c(\boldsymbol{\theta}; \mathbf{y}_m^{(k,l)}, \mathbf{y}_o) - \widehat{Q}_{k-1}(\boldsymbol{\theta}) \right), \tag{1.2}$$

   where δ_k is a smoothing parameter defined as a decreasing sequence of positive numbers satisfying $\sum_{k=1}^{\infty} \delta_k = \infty$ and $\sum_{k=1}^{\infty} \delta_k^2 < \infty$ (Kuhn & Lavielle, 2005).

Following Galarza *et al.* (2017), we adopt the variable δ_k as

$$\delta_k = \begin{cases} 1, & \text{if } 1 \leqslant k \leqslant cW; \\ \dfrac{1}{k - cW}, & \text{if } cW + 1 \leqslant k \leqslant W, \end{cases} \tag{1.3}$$

where W is the maximum number of iterations, and c is a cutoff point ($0 \leqslant c \leqslant 1$) that determines the percentage of initial iterations with no memory. If $c = 0$, the algorithm will have a memory for all iterations and hence will converge slowly to the ML estimates, and W needs to be large. If $c = 1$, the algorithm will be memory-free, it will converge quickly to a solution neighborhood, and the algorithm will initiate a Markov chain leading to a reasonably well-estimated mean after applying the necessary *burn-in* and *thinning* steps. A number between 0 and 1 ($0 < c < 1$) will assure an initial convergence in distribution to a solution neighborhood for the first cW iterations and an almost sure convergence for the rest of the iterations.

The choice of c and W could impact the convergence speed; therefore, it is recommended to choose c and W such that $50 < cW < 100$ (Kuhn & Lavielle, 2005). A graphical approach can monitor the convergence of the estimates for all parameters and determine the values for these constants, as Lavielle (2014) suggested. An advantage of the SAEM algorithm is that, even though it performs an MCMC E-step, it only requires a small and fixed sample size M (suggested to be $M \leqslant 20$), making it much faster than the MCEM algorithm.

1.2 Standard error approximation

The EM algorithm does not provide the variance-covariance matrix of the ML estimates. As an alternative, we can approximate it by the inverse of the observed

information matrix, which can be derived from the negative of the second derivatives (hessian matrix) of the observed log-likelihood function:

$$\mathbf{I}_o(\boldsymbol{\theta}) = -\frac{\partial^2 \ell(\boldsymbol{\theta}; \mathbf{y}_o)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}. \tag{1.4}$$

Calculating the Hessian matrix of the observed log-likelihood can be challenging. To address this, Louis (1982) introduced a method for extracting the observed information matrix when utilizing the EM algorithm for ML estimation. Define $\mathbf{S}_c(\mathbf{y}_c; \boldsymbol{\theta})$ and $\mathbf{B}_c(\mathbf{y}_c; \boldsymbol{\theta})$ as the first derivative and the negative of the second derivative of the complete-data log-likelihood function with respect to the parameter vector $\boldsymbol{\theta}$, respectively. Similarly, let $\mathbf{S}_o(\mathbf{y}_o; \boldsymbol{\theta})$ and $\mathbf{B}_o(\mathbf{y}_o; \boldsymbol{\theta})$ be the derivatives of the log-likelihood function of the observed data. The observed information matrix is given by

$$\begin{aligned} \mathbf{I}_o(\boldsymbol{\theta}) &= \mathbf{B}_o(\mathbf{y}_o; \boldsymbol{\theta}) \\ &= \mathbb{E}\left[\mathbf{B}_c(\mathbf{y}_c; \boldsymbol{\theta}) \mid \mathbf{y}_o\right] - \mathbb{E}\left[\mathbf{S}_c(\mathbf{y}_c; \boldsymbol{\theta}) \mathbf{S}_c^\top(\mathbf{y}_c; \boldsymbol{\theta}) \mid \mathbf{y}_o\right] + \mathbf{S}_o(\mathbf{y}_o; \boldsymbol{\theta}) \mathbf{S}_o^\top(\mathbf{y}_o; \boldsymbol{\theta}). \end{aligned} \tag{1.5}$$

The first term in (1.5) represents the conditional expected observed information matrix of the full data, and the remaining terms denote the observed information matrix associated with the missing or censored data. The proof of this result can be found in Louis (1982) or Walsh (2006). The authors also demonstrated that the third term in (1.5) is equal to $\mathbb{E}[\mathbf{S}_c(\mathbf{y}_c; \boldsymbol{\theta}) | \mathbf{y}_o]$, and by definition of the EM algorithm, this term is equal to zero at the ML estimates $\widehat{\boldsymbol{\theta}}$, i.e., $\mathbf{S}_o(\mathbf{y}_o; \widehat{\boldsymbol{\theta}}) = \mathbb{E}[\mathbf{S}_c(\mathbf{y}_c; \widehat{\boldsymbol{\theta}}) \,|\, \mathbf{y}_o] = \mathbf{0}$. Then, the observed information matrix at the ML estimate can be expressed as follows

$$\mathbf{I}_o(\widehat{\boldsymbol{\theta}}) = \mathbb{E}\left[\mathbf{B}_c(\mathbf{y}_c; \widehat{\boldsymbol{\theta}}) \mid \mathbf{y}_o\right] - \mathbb{E}\left[\mathbf{S}_c(\mathbf{y}_c; \widehat{\boldsymbol{\theta}})\, \mathbf{S}_c^\top(\mathbf{y}_c; \widehat{\boldsymbol{\theta}}) \mid \mathbf{y}_o\right]. \tag{1.6}$$

The computational complexity increases when these expectations are difficult to calculate analytically, but they can be estimated via the Monte Carlo method as

$$\widehat{\mathbf{I}}_o(\boldsymbol{\theta}) \approx -\frac{1}{L} \sum_{l=1}^{L} \frac{\partial^2 \ell_c(\boldsymbol{\theta}; \mathbf{y}^{(l)})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} - \frac{1}{L} \sum_{l=1}^{L} \frac{\partial \ell_c(\boldsymbol{\theta}; \mathbf{y}^{(l)})}{\partial \boldsymbol{\theta}} \left(\frac{\partial \ell_c(\boldsymbol{\theta}; \mathbf{y}^{(l)})}{\partial \boldsymbol{\theta}}\right)^\top, \tag{1.7}$$

where $\mathbf{y}^{(l)} = (\mathbf{y}_o^\top, \mathbf{y}_m^{(l)\top})^\top$, for $l \in \{1, \ldots, L\}$, represents the complete data composed of the observed data and simulated missing data from the conditional distribution $f(\mathbf{y}_m | \boldsymbol{\theta}, \mathbf{y}_o)$.

Furthermore, when the SAEM algorithm is used to estimate the parameters, Delyon *et al.* (1999) adapted Louis' method to compute the observed information matrix. This adaptation involves calculating additional terms as outlined:

$$\mathbf{H}_k = -\mathbf{G}_k + \boldsymbol{\Delta}_k \boldsymbol{\Delta}_k^\top, \tag{1.8}$$

$$\mathbf{G}_k = \mathbf{G}_{k-1} + \delta_k \left(\frac{1}{M} \sum_{l=1}^{M} \left(\frac{\partial^2 \ell_c(\boldsymbol{\theta}; \mathbf{y}^{(k,l)})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} + \frac{\partial \ell_c(\boldsymbol{\theta}; \mathbf{y}^{(k,l)})}{\partial \boldsymbol{\theta}} \frac{\partial \ell_c(\boldsymbol{\theta}; \mathbf{y}^{(k,l)})}{\partial \boldsymbol{\theta}}^\top \right) - \mathbf{G}_{k-1} \right), \quad \text{and}$$

$$\boldsymbol{\Delta}_k = \boldsymbol{\Delta}_{k-1} + \delta_k \left(\frac{1}{M} \sum_{l=1}^{M} \frac{\partial \ell_c(\boldsymbol{\theta}; \mathbf{y}^{(k,l)})}{\partial \boldsymbol{\theta}} - \boldsymbol{\Delta}_{k-1} \right),$$

with $\mathbf{y}^{(k,l)} = (\mathbf{y}_o^\top, \mathbf{y}_m^{(k,l)\top})^\top$ being the complete data at iteration (k,l) for $k \in \{1,\ldots,W\}$ and $l \in \{1,\ldots,M\}$. The inverse of the limiting value of $\mathbf{H}_k$ can be used to assess the dispersion of the estimators (Delyon *et al.*, 1999).

On the other hand, for independent observations, a consistent estimator for the Fisher information matrix is the empirical information matrix (Meilijson, 1989), which is given by

$$\mathbf{I}_e(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathbf{s}(\mathbf{y}_i;\boldsymbol{\theta})\,\mathbf{s}(\mathbf{y}_i;\boldsymbol{\theta})^\top - \frac{1}{n}\mathbf{S}(\mathbf{y};\boldsymbol{\theta})\,\mathbf{S}(\mathbf{y};\boldsymbol{\theta})^\top, \tag{1.9}$$

where $\mathbf{S}(\mathbf{y};\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathbf{s}(\mathbf{y}_i;\boldsymbol{\theta})$ and $\mathbf{s}(\mathbf{y}_i;\boldsymbol{\theta}) = \mathbf{s}_i(\boldsymbol{\theta})$ is the empirical score function for the ith sample unit (see Meilijson, 1989; Lin, 2010). According to Louis (1982), the individual score can be determined as

$$\mathbf{s}_i(\boldsymbol{\theta}) = \frac{\partial f(\mathbf{y}_i;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbb{E}\left[\frac{\partial \ell_{ic}(\boldsymbol{\theta};\mathbf{y}_i)}{\partial \boldsymbol{\theta}}\middle|\mathbf{y}_i^o, \boldsymbol{\theta}\right],$$

with $\ell_{ic}(\boldsymbol{\theta};\mathbf{y}_i)$ and $\mathbf{y}_i^o$ denoting the complete-data log-likelihood function and the observed data for the ith observation, respectively. Substituting the ML estimates $\widehat{\boldsymbol{\theta}}$ into $\boldsymbol{\theta}$, we get $\mathbf{S}(\mathbf{y};\widehat{\boldsymbol{\theta}}) = \mathbf{0}$, then (1.9) is reduced to

$$\mathbf{I}_e(\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \mathbf{s}_i(\widehat{\boldsymbol{\theta}})\,\mathbf{s}_i(\widehat{\boldsymbol{\theta}})^\top. \tag{1.10}$$

1.3 Case studies

This section presents the datasets that will be analyzed in this thesis.

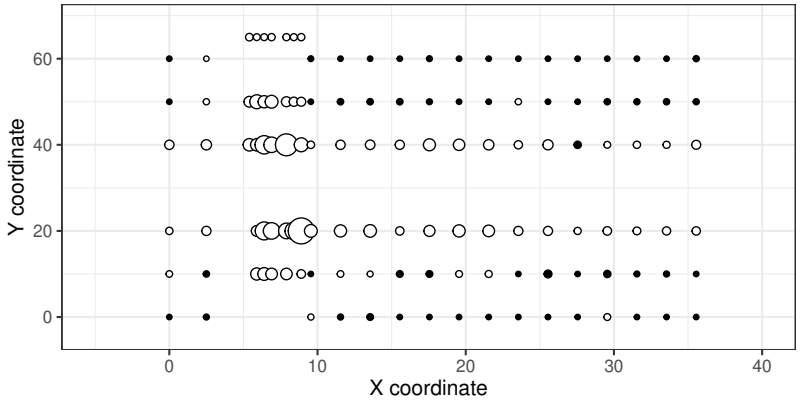
1.3.1 Missouri dioxin contamination data

The Missouri dioxin contamination dataset pertains to a study of contamination level by dioxin (2,3,7,8-tetrachlorodibenzo-p-dioxin or TCDD) at sampled points along the road in Missouri, USA (Zirschky & Harris, 1986). This study aimed to determine contamination levels and identify areas requiring cleanup. The dataset contains 127 observations distributed in an area of $3600 \times 65\ m^2$ on the shoulders of a country road, with 43% of the observations (55 sites) being left-censored, i.e., falling below the detection limit, which ranges from 0.10 to 0.79 mg/kg. For illustration purposes, the reported data is treated as coming from a single sampled location. The spatial directions are the x-direction (measured in 1/100 ft) taken parallel to the road, and the y-direction (in ft) represents the direction perpendicular to the road. Refer to Zirschky & Harris (1986) and Fridley & Dixon (2007) for more details. This data is available in the R package `CensSpatial` (Ordoñez *et al.*, 2020).

Figure 1 illustrates the sampled locations, where each bold point represents a censored observation and each white point an observed value. The size of each point

reflects the magnitude of the observed value, which varies between 0.10 and 48.05 mg/kg. It is worth mentioning that this data was first analyzed by Zirschky & Harris (1986), who concluded that the data appeared to be log-normally distributed. Hence, the logarithms of the observed responses were used in the analysis. The dataset was further examined by Lachos *et al.* (2017) and Ordoñez *et al.* (2018), who employed an SAEM algorithm to estimate model parameters. Using information criteria (AIC and BIC), both studies determined that an exponential correlation function best assessed the data's correlation. Lachos *et al.* (2017) also identified four observations as potentially influential in the estimation of parameters.

Figure 1 – Missouri data. Level of TCDD observed on each location, where ∘ represents an observed value and • represents a left-censored value.



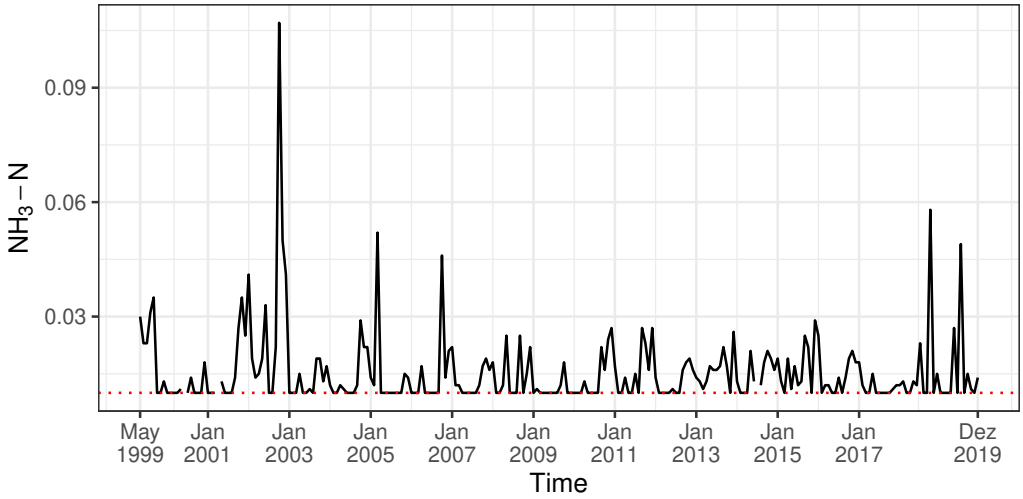
1.3.2 Ammonia-nitrogen concentration data

The ammonia-nitrogen (NH_3-N) measurements were taken in the Samish River in Washington State, USA. The data were collected monthly by the Washington State Department of Ecology, and it is available for free download on the official website (https://ecology.wa.gov).

In this work, we consider a subset of the data that consists of $n = 248$ observations of the NH_3-N concentration in mg/L monthly measured from May 1999 to December 2019. Measurements less than $0.01mg/L$ were labeled as undetected, and such concentrations were listed as censored. Another feature to be considered in this data is that it is not evenly spaced; when this happens, a blank observation is added and treated as missing. Therefore, the dataset contains 102 (41.13%) censored observations and 3 (1.21%) missing observations. Figure 2 shows the ammonia-nitrogen concentration time series, where the red dotted line represents the detection limit.

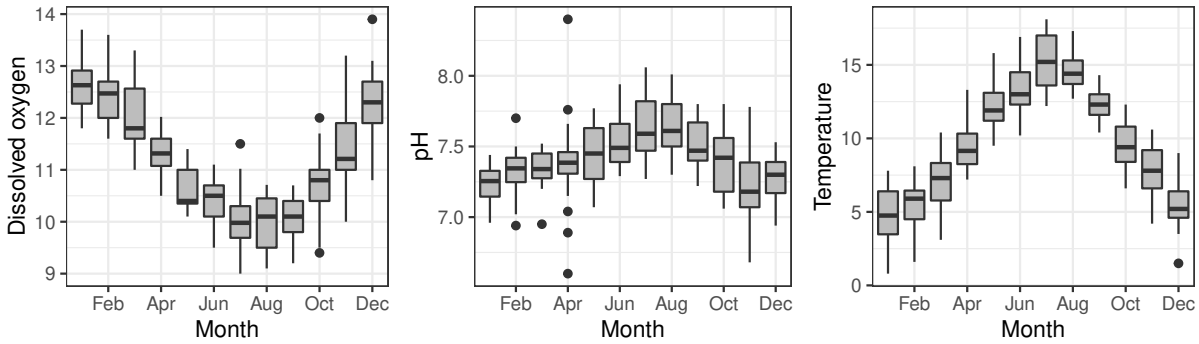
Following De Abreu *et al.* (2022), the predictors dissolved oxygen (DO) in mg/L, pH, and temperature (T) in $^{\circ}C$ could have a statistical effect on the ammonia-nitrogen. An analysis of the temporal behavior of these covariates is therefore conducted. The boxplot

Figure 2 – Ammonia-concentration data. Time series plot of the $NH_3 - N$ (black line) and the detection limit (red dotted line).



of the covariates by month is shown in Figure 3, where we can see that the DO varies from 9.0 to 13.9 mg/L, decreasing from January to July, remaining constant (in median) from July to September and then increasing. Besides, the pH varies from 6.6 to 8.4, slightly increasing from January to July, and then decreasing. On the other hand, the temperature varies from $0.8°C$ to $18.1°C$, with the highest temperatures happening in July. It is worth mentioning that the summer in Washington goes from June to August, while the lowest temperatures were observed from December to February during the winter.

Figure 3 – Ammonia-nitrogen concentration data. Boxplot for the dissolved oxygen (mg/L), pH, and temperature ($°C$) by month.

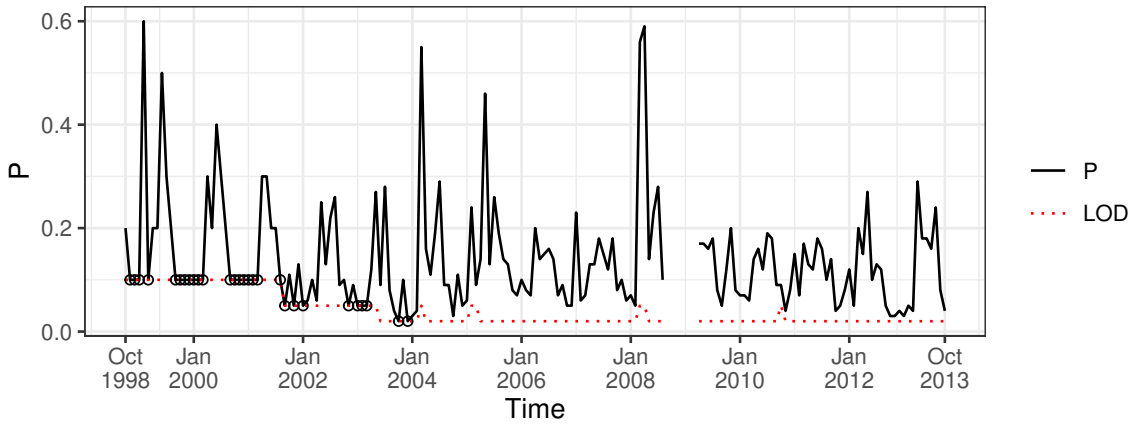


1.3.3 Phosphorus concentration data

The phosphorus concentration levels indicate the river water quality since, for instance, excessive phosphorus in surface water may result in eutrophication. Phosphorus concentration data of West Fork Cedar River at Finchford, Iowa, USA, was collected under the ambient water quality program conducted by the Iowa Department of Natural Resources (Iowa DNR). This data is available in the package `ARCensReg` (Schumacher

et al., 2016). The data consists of $n = 181$ observations of phosphorus concentration (P) in mg/L monthly measured from October 1998 to October 2013. The phosphorus concentration measurement was subject to a detection limit (DL) of 0.02, 0.05, or 0.10 mg/L, depending on the time, and therefore 28 (15.47%) observations are left-censored. Moreover, there are 7 (3.87%) missing observations from September 2008 to March 2009 due to a program suspension caused by a temporary lack of funding. Figure 4 displays the phosphorus concentration (P) time series; the red dotted line represents the DLs.

Figure 4 – Phophorus concentration data. Time series plots of P (black line) and the limit of detection (red dotted line). Censored observations are marked by circles.



This dataset was previously analyzed by Schumacher *et al.* (2017), considering a censored regression model with independent errors under the normal distribution. The analysis indicated that the assumption of independent errors is not valid, and hence, a model with autocorrelated errors was evaluated. Based on information criteria and the mean squared prediction error (MSPE), this last study concluded that the error modeled as an autoregressive (AR) process of order one best fits this data. The authors also found some influential observations. A similar dataset was studied by Wang & Chan (2018), who proposed a quasi-likelihood estimation method for censored autoregressive models with exogenous variables.

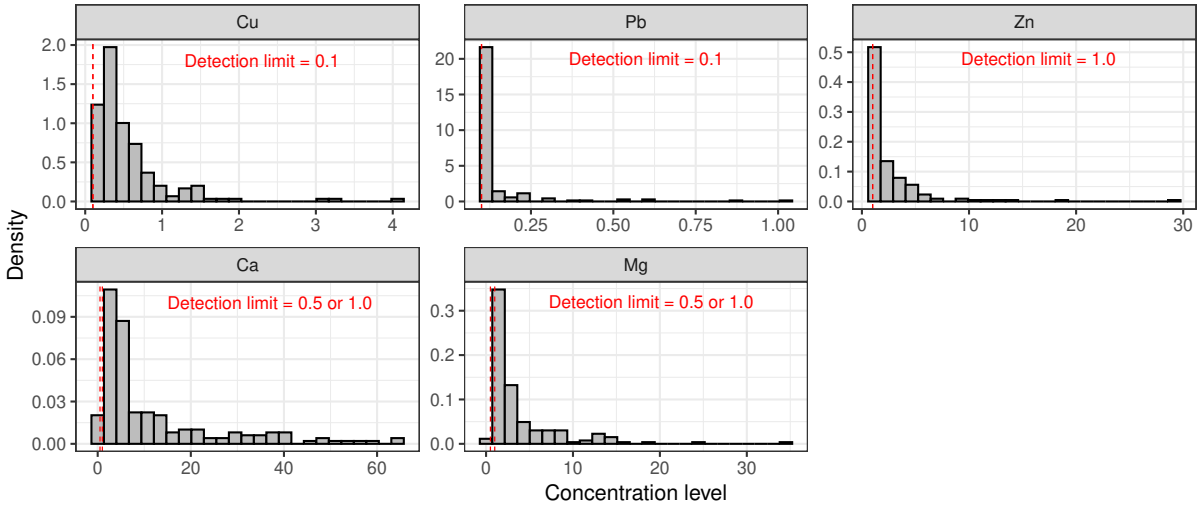
1.3.4 Trace metals in freshwater streams data

This dataset includes data on the concentration levels of five dissolved trace metals –copper (Cu), lead (Pb), zinc (Zn), calcium (Ca), and magnesium (Mg)– measured at 184 randomly selected freshwater sites across Virginia. The concentrations for Cu, Pb, and Zn are expressed in micrograms per liter (μg/L) of water, while the concentrations for Ca and Mg are in milligrams per liter (mg/L). The detection limits are set at 0.1 μg/L for Cu and Pb, 1.0 μg/L for Zn, and vary between 0.5 mg/L and 1.0 mg/L for Ca and Mg, depending on the timing of measurements. The dataset also indicates the proportion

of measurements below the detection limits, known as left-censored values: 2.7% for Ca, 4.9% for Cu, 9.8% for Mg, 38.6% for Zn, and 78.3% for Pb.

Figure 5 shows the histogram for each original metal concentration level with the detection limits and censoring proportion. Here, all the distributions are right-skewed and heavy-tailed.

Figure 5 – VDEQ data. Histograms of the original five trace metal concentrations: Cu, Pb, Zn (in μg/L of water), Ca, and Mg (in mg/L of water). The red dashed lines represent the detection limits.



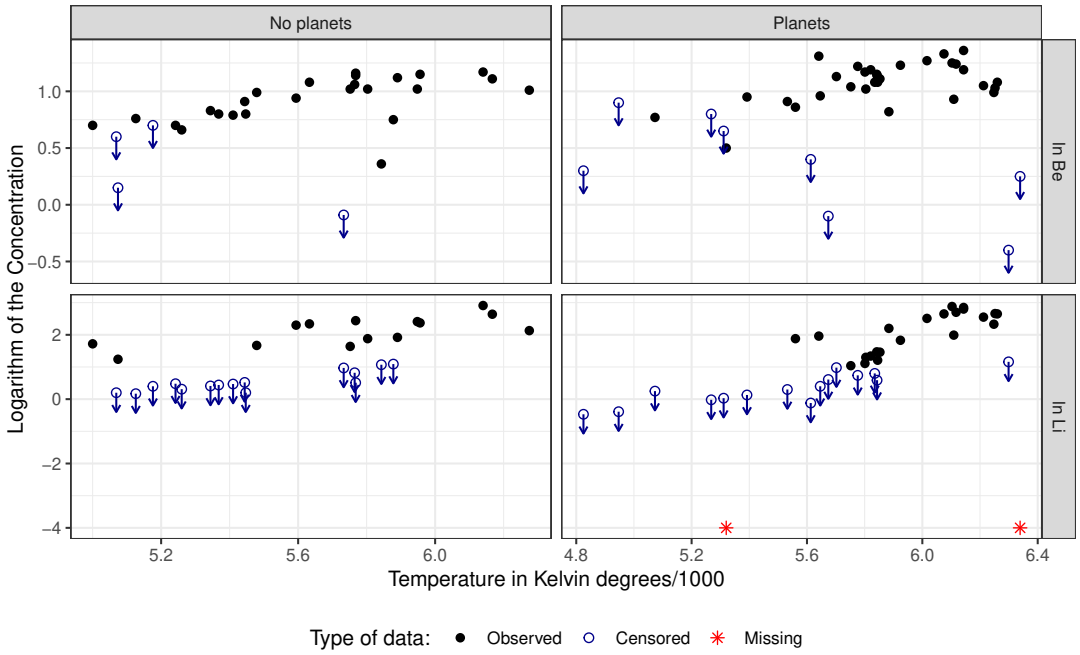
Galarza *et al.* (2022b) recently analyzed this dataset, concluding through a likelihood ratio test (LRT) that the skewness parameter is different from zero for the model that considers the original variables. On the other hand, when a logarithmic transformation is applied, the LRT does not reject the null hypothesis, i.e., $\boldsymbol{\lambda} = \mathbf{0}$.

1.3.5 Stellar abundances data

The Stellar abundances dataset consists of 3 measured variables for $n = 68$ solar-type stars, 39 stars known to host planets, and 29 in a control sample of stars without planets. The proportion of left-censored and missing values is 17.6% and 47.1% for Be and Li, respectively. Moreover, 51.5% of the stars had all measures fully observed, 32.3% had one metal partially detected, and the other 16.2% had both partially detected. This dataset has been analyzed by Santos *et al.* (2002) using a nonparametric approach, and it is available in the R package `astrodatR` (Feigelson, 2014).

Figure 6 shows the scatter plot of the logarithm of beryllium (top) and lithium (bottom) against the temperature in Kelvin degrees/1000 for stars known to host planets (right) and the control sample stars (left). Here, black points represent completely observed values, blue points denote left-censored observations, which are also depicted by the direction of the arrow, and red symbols are representing missing data.

Figure 6 – Stellar data. Scatter plot for the logarithm of beryllium (top) and lithium (bottom) abundances against temperature.

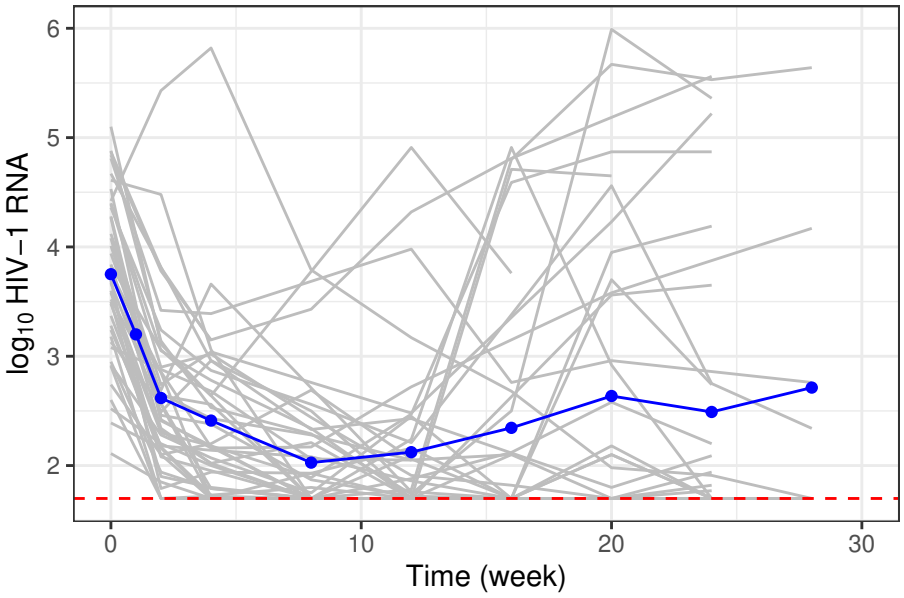


1.3.6 A5055 data

The A5055 dataset comes from the AIDS clinical trial study A5055 (Wang, 2013), which involves 44 infected patients with the human immunodeficiency virus type 1 (HIV-1). These patients were treated with one of the two potent antiretroviral (ARV) therapies: IDV 800mg twice daily (q12h) plus RTV 200mg q12h (treatment 1), or IDV 400mg q12h plus RTV 400mg q12h (treatment 2). In AIDS research, plasma viral load (number of RNA copies) and its trajectories are crucial for diagnosing HIV-1 disease progression post-ARV treatment (Paxton *et al.*, 1997). Another important immunologic marker for monitoring disease progression in AIDS studies is the cluster of differentiation 4 (CD4).

The dataset includes plasma viral load measurements (in copies per milliliter) and CD4 cell counts, collected at approximately days 0, 7, 14, 28, 56, 84, 112, 140, and 168 of follow-up for each patient. This study focuses on the longitudinal trajectories of RNA viral load, converted to a log-base-10 scale and denoted by $\log_{10}$(RNA). The lower detection limit for RNA viral load is 50 copies/milliliter, corresponding to 1.698 on the log-base-10 scale, with 33.5% (106 out of 316) of measurements below this quantification limit (left-censored). Figure 7 shows the trajectories of immunologic responses over time. The red dashed line denotes the detection limit, $\log_{10}(50)$, and the blue points represent the mean of the viral load per week.

Figure 7 – A5055 data. Trajectories of $\log_{10}$RNA for 44 HIV-1 infected patients who were randomized in two IDV-RTV regimens. The blue line denotes the mean of the viral load per week, and the red dashed line indicates the detection limit.



2 Moments and random number generation for the Truncated Elliptical family of distributions

2.1 Introduction

The use of truncated distributions arises in a wide variety of statistical models, such as survival analysis, censored data models, Bayesian models with truncated parameter space, and abound in agronomy, biology, environmental monitoring, medicine, and economics. Algorithms like Expectation-Maximization (EM) (Dempster *et al.*, 1977) are employed frequently in multivariate censored data analysis under a likelihood-based perspective for its facility to deal with missing and partially observed data. This algorithm requires the computation of conditional truncated moments, commonly the first two moments. For example, Matos *et al.* (2016) and Matos *et al.* (2013b) estimated the parameters of a censored mixed-effects model for irregularly repeated measures via the EM algorithm, which needed to compute the first two moments of a truncated multivariate t (TMVT) and a truncated multivariate normal (TMVN) distributions, respectively. Variations of the EM algorithm, such as Stochastic Approximation EM (SAEM) (Delyon *et al.*, 1999) and Monte Carlo EM (MCEM) (Wei & Tanner, 1990), replace the conditional expectations with an approximation that requires drawing independent random observations from a truncated distribution. For instance, Lachos *et al.* (2017) estimated the parameters of a linear spatial model for censored data using the SAEM algorithm, which needed to generate random samples from the TMVN distribution to perform the stochastic approximation step. More recently, using the SAEM algorithm, Lachos *et al.* (2019) proposed a robust multivariate linear mixed model for multiple censored responses based on the scale mixtures of normal (SMN) distributions. Moreover, Gelfand *et al.* (1992) showed how to perform Bayesian analysis for constrained parameters or truncated data problems using Gibbs sampling. Therefore, sampling random observations and computing moments from truncated distributions is a task of considerable interest.

There are several methods to generate random samples from a truncated distribution in the literature, making the rejection sampling (RS) technique the most common. For instance, in the simplest case, when the non-truncated distribution is considered as the "proposal" probability density function (pdf), the RS method draws samples from the latter and retains only the samples inside the support region. However, the procedure may be inefficient for truncated distributions, especially when the truncation

interval is too small or located at a less probable area of the pdf. As an alternative, Robert (1995) proposed an accept-reject algorithm, which dramatically improves the RS method's efficiency for simulating one-dimensional truncated normal distributions. This method is available through the R (R Core Team, 2021) function `rtnorm` (Hadfield, 2022). Besides, the Gibbs sampler algorithm is the most commonly used Markov Chain Monte Carlo (MCMC) method, which generates random observations from a multivariate density by sampling in succession from the full conditional distributions. Nevertheless, its implementation may need methods for sampling from nonstandard univariate distributions. See that an extension to the multivariate framework of the accept-reject algorithm was implemented through a Gibbs sampler (Geman & Geman, 1984; Gelfand & Smith, 1990).

Later, some automatic and self-tuning samplers for univariate distributions emerged, such as the adaptive rejection sampling (ARS) (Gilks & Wild, 1992), the adaptive rejection Metropolis sampling (ARMS) (Gilks *et al.*, 1995), the adaptive rejection Metropolis sampling using Lagrange interpolation polynomials of degree 2 (ARMS2) (Meyer *et al.*, 2008), the independent doubly adaptive rejection Metropolis sampling (IA^{2}RMS) (Martino *et al.*, 2015a), among others. ARS is an efficient random generator for log-concave distributions, which leads to independent samples and ensures that the sequence of proposals converges to the target pdf. This method reduces the number of evaluations of the target in RS by improving the proposal after each rejection, which is piecewise exponential. To deal with non-log-concave distributions, ARMS generalizes the ARS method by incorporating a Metropolis-Hastings step. This algorithm may return correlated samples, and it cannot guarantee the convergence of the sequence of proposals to the target; even so, it has often been preferred over other MCMC techniques due to its good performance. This method is available in the R library `armspp` (Bertolacci, 2019). In contrast, ARMS2 is an extension of ARS and ARMS, whose proposal function is a sequence of two piecewise exponential and $n-2$ truncated normal densities, with n representing the number of elements in the set of support points. On the other hand, IA^{2}RMS is a variation of ARMS, which ensures that the sequence of proposals converges to the target leading to a reduction of the correlation. This algorithm, unlike ARMS, decouples the adaptation mechanism from the proposed construction, allowing one to consider simpler alternatives for the candidate density.

Another alternative to sampling from truncated multivariate distributions is the slice sampler algorithm (Neal, 2003), which turns sampling from a truncated density into sampling repeatedly from uniform distributions instead by introducing an auxiliary variable. This approach is often easier to implement than Gibbs sampling. In general, it is easy to code, fast and does not generate samples out of the truncation region, making it more efficient than the conventional rejection method. The auxiliary variables were also employed by Damien & Walker (2001) to sample from the TMVN distribution. In the same way, Ho *et al.* (2012) used slice sampling to draw random points from the TMVT

distribution.

Regarding calculating moments from truncated distributions, there are a few libraries in R that provide truncated multivariate moments. For instance, the package `tmvtnorm` (Wilhelm, 2015) computes the mean and the variance of the TMVN distribution by deriving its moment generating function, which is an extension of the method described by Tallis (1961). In contrast, the `MomTrunc` library (Galarza *et al.*, 2021a) uses a recursive approach method proposed by Kan & Robotti (2017) to compute arbitrary higher-order moments (Galarza *et al.*, 2022b). For the TMVT distribution, the packages `TTmoment` and `MomTrunc` compute its two first moments. The first library only handles integer degrees of freedom greater than 4, while the latter can compute even high-order moments for any degrees of freedom (Galarza *et al.*, 2021c).

To the best of our knowledge, there are no proposals in the literature to generate samples from other multivariate truncated distributions in the elliptical class other than the TMVN and TMVT distributions (available in the `tmvtnorm` and `TTmoment` packages). Hence, motivated by the slice sampling algorithm, we propose a general method to obtain samples from any truncated multivariate elliptical distribution with a strictly decreasing density generating function (dgf). Using conditional expectation properties, we also construct an efficient algorithm to approximate the moments of the most common distribution of this class: the truncated multivariate normal, Student-t, slash, contaminated normal, and Pearson VII distributions. This method requires less running time than the existing ones since it deals separately with the truncated and non-truncated parts of the vector. Our proposal can be reached through the R package `relliptical`.

Finally, it is worth mentioning that moments of truncated elliptical distributions can be used to compute truncated moments for the selection elliptical family of distributions. This wide family includes complex multivariate asymmetric versions of the elliptical distributions as the extended skew-normal and the unified skew-t distributions, among others. Therefore, our proposal opens the doors for the calculation of truncated moments of complex elliptical asymmetric distributions, which are of particular interest for the development of robust censored models with asymmetry, heavy tails, and missingness (see, for instance, De Alencar *et al.*, 2021; Galarza *et al.*, 2021b).

This chapter is organized as follows. Section 2.2 shows some results related to the elliptical and truncated elliptical family of distributions and a brief description of the slice sampling algorithm. Section 2.3 is devoted to the formulation of the sampling algorithm for the truncated elliptical distributions, whereas Section 2.4 focuses on our proposed method to approximate the first and the second moment. For the last two sections, we present a brief introduction to its respective R function. A simulation study that compares the mean and covariance matrix for the TMVT distribution estimated through different methods in R is presented as well.

2.2 Preliminaries

2.2.1 Elliptical family of distributions

As defined in Muirhead (2009) and Fang (2018), a p-variate random vector $\mathbf{X}$ is said to follow an elliptical distribution with location parameter $\boldsymbol{\mu} \in \mathbb{R}^p$, positive-definite scale matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, and density generating function g, if its pdf is given by

$$f_{\mathbf{X}}(\mathbf{x}) = c_p \left|\boldsymbol{\Sigma}\right|^{-\frac{1}{2}} g\left((\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right), \tag{2.1}$$

for $\mathbf{x} \in \mathbb{R}^p$, where $g(t)$ is a non-negative Lebesgue measurable function on $[0,\infty)$ such that $\int_0^\infty t^{p/2-1} g(t)\,\mathrm{d}t < \infty$.Moreover,

$$c_p = \frac{\Gamma(p/2)}{\pi^{p/2}} \left(\int_0^\infty t^{p/2-1} g(t)\,\mathrm{d}t\right)^{-1}$$

is the normalizing constant, with $\Gamma(\cdot)$ representing the complete gamma function. We will use the notation $\mathbf{X} \sim E\ell_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g)$.

Members of the elliptical family of distributions are characterized by their density-generating function g. Some examples of the elliptical family of distributions are:

- The *multivariate normal* distribution, $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, arises when the dgf takes the form $g(t) = \exp(-t/2)$, $t \geqslant 0$.
- The *multivariate Student-t* distribution, $\mathbf{X} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, with location parameter $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$, and degrees of freedom $\nu > 0$, is obtained when $g(t) = (1 + t/\nu)^{-(\nu+p)/2}$, $t \geqslant 0$.
- The *multivariate power exponential*, $\mathbf{X} \sim \mathrm{PE}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \beta)$, is characterized by a kurtosis parameter $\beta > 0$ and the dgf $g(t) = \exp(-t^\beta/2), t \geqslant 0$. A particular case of the power exponential distribution is the normal distribution, which arises when $\beta = 1$.
- The *multivariate slash*, $\mathbf{X} \sim \mathrm{SL}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, is defined for a random variable $\mathbf{X}$ with dgf $g(t) = \int_0^1 u^{\nu+p/2-1} \exp(-ut/2)\,\mathrm{d}u$, $t \geqslant 0$, $\nu > 0$.
- The *multivariate contaminated normal*, $\mathbf{X} \sim CN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu, \rho)$, $0 \leqslant \nu \leqslant 1, 0 < \rho \leqslant 1$, is characterized by the dgf $g(t) = (1-\nu)\exp(-t/2) + \nu \exp(-t/(2\rho)), t \geqslant 0$. Here, ν represents the proportion of contamination, and ρ scales the variance of the contaminating component relative to the primary distribution.
- The *multivariate Pearson VII* distribution, $\mathbf{X} \sim \mathrm{PVII}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, m, \nu)$, with parameters $\boldsymbol{\mu} \in \mathbb{R}^p$, $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, $m > p/2$, and $\nu > 0$ is obtained when $g(t) = (1 + t/\nu)^{-m}$, $t \geqslant 0$.

For more distributions belonging to this family, please see Fang (2018).

2.2.2 Truncated elliptical family of distributions

Let $A \subseteq \mathbb{R}^p$ be a measurable set. We say that a p-variate random vector $\mathbf{Y}$ has a truncated elliptical distribution with support A, location parameter $\boldsymbol{\mu} \in \mathbb{R}^p$, scale parameter $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, and dgf g, if its pdf is given by

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{g\left((\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)}{\int_A g\left((\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right) \mathrm{d}\mathbf{y}} = \frac{f_{\mathbf{X}}(\mathbf{y})}{\Pr(\mathbf{X} \in A)}, \quad \mathbf{y} \in A, \tag{2.2}$$

where $\mathbf{X} \sim \mathrm{E}\ell_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g)$. We use the notation $\mathbf{Y} \sim \mathrm{TE}\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g, A)$. Notice that the pdf of $\mathbf{Y}$ is written as the ratio between the pdf of $\mathbf{X} \sim \mathrm{E}\ell_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g)$ and $\Pr(\mathbf{X} \in A)$, so the pdf of $\mathbf{Y}$ exists if the pdf of $\mathbf{X}$ does, which occurs if $\boldsymbol{\Sigma}$ is positive-definite (see, Morán-Vásquez & Ferrari, 2021). The variable $\mathbf{Y}$ is also said to be an elliptical distribution truncated on A, being represented by $\mathbf{Y} \overset{d}{=} \mathbf{X} \mid (\mathbf{X} \in A)$, where $\overset{d}{=}$ means "has the same distribution as".

As in the elliptical family of distributions, the dgf g determines any distribution within the truncated elliptical class. For example, if $g(t) = (1 + t/\nu)^{-(\nu+p)/2}$, $t \geqslant 0$, $\nu > 0$, then $\mathbf{Y}$ has TMVT distribution. We will denote the different members of the truncated elliptical family defined in the subsection before as $\mathbf{Y} \sim \mathrm{TN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; A)$ for the TMVN distribution, $\mathbf{Y} \sim \mathrm{T}t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; A)$ for the TMVT distribution, $\mathbf{Y} \sim \mathrm{TPE}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \beta; A)$ for the truncated multivariate power exponential, $\mathbf{Y} \sim \mathrm{TSL}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; A)$ for the truncated multivariate slash distribution, and $\mathbf{Y} \sim \mathrm{TPVII}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, m, \nu; A)$ for the truncated multivariate Pearson VII distribution.

2.2.3 Slice sampling algorithm

A slice sampler is a form of auxiliary variable technique in which one or more variables are introduced to facilitate the construction of an MCMC method. The idea of using auxiliary variables in MCMC methods was established for the Ising model by Swendsen & Wang (1987), and it was brought into the statistical literature by Besag & Green (1993). Following Neal (2003), an MCMC method can be constructed using the principle that it can sample from a given distribution by simulating uniformly from the region under the plot of its density function.

Suppose we are interested in sampling from the distribution of a random variable $\mathbf{X} \in \mathbb{R}^p$, whose pdf is proportional to the function $f(\mathbf{x})$. The slice sampler algorithm simulates uniformly from the $(p+1)$-dimensional region under the plot of $f(\mathbf{x})$ by introducing a real auxiliary variable, Y, such that the joint pdf of $\mathbf{X}$ and Y is uniform over the region $V = \{(\mathbf{x}, y) : 0 < y < f(\mathbf{x})\}$, i.e., $f_{X,Y}(\mathbf{x}, y) \propto \mathbb{I}(0 < y < f(\mathbf{x}))$, with $\mathbb{I}(\cdot)$ being an indicator function. Therefore, we can obtain samples from the distribution of $\mathbf{X}$ by sampling jointly $(\mathbf{x}, y)$ and then ignoring y.

Note that generating independent random points uniformly distributed on V may not be easy. To overcome this problem, Neal (2003) defined a Markov Chain that

converges to this uniform distribution similar to Gibbs sampler or Metropolis-Hastings algorithms. Then, considering Gibbs sampler steps, the slice sampling at iteration k works as follows:

1. Given the current value of $\mathbf{x}_{k-1}$, sample y_k from $Y \mid (\mathbf{X} = \mathbf{x}_{k-1}) \sim U(0, f(\mathbf{x}_{k-1}))$.
2. Draw $\mathbf{x}_k$ from the conditional distribution of $\mathbf{X}$ given y_k, which is uniform over the region $S_k = \{\mathbf{x} : y_k < f(\mathbf{x})\}$, i.e., $\mathbf{X} \mid (Y = y_k) \sim U(\{\mathbf{x} : y_k < f(\mathbf{x})\})$.

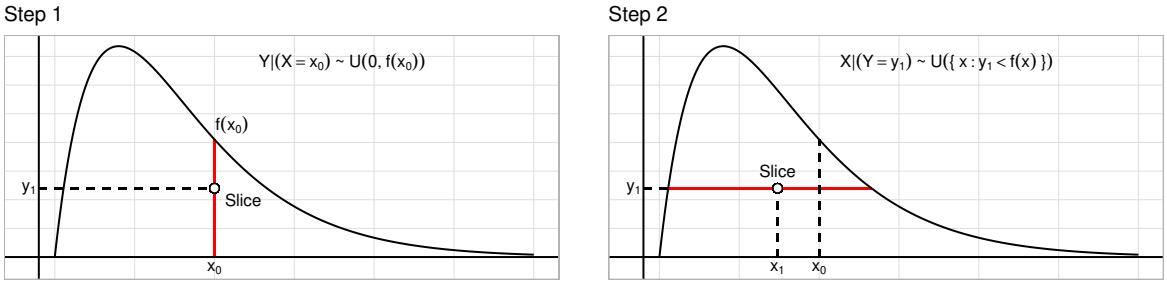
For all $k = 1, 2, \ldots, n$, where n is the desired sample size. We can notice that the slice sampling method is easily implemented for univariate random variables. In contrast, for the multivariate case ($p > 1$), sampling uniformly from the region S_k may be complex, in which case we can employ some update for $\mathbf{x}$ that leaves the uniform distribution invariant over this slice.

Figure 8 shows the steps of the slice sampling algorithm for a univariate random variable X. Given an initial value $X = x_0$,

1. Simulate y_1 from $Y \mid (X = x_0) \sim U(0, f(x_0))$.
2. Simulate x_1 from $X \mid (Y = y_1) \sim U(S_1)$, with $S_1 = \{x : y_1 < f(x)\}$.

These two steps are repeated n times by making $x_0 = x_1$ in the next iteration.

Figure 8 – Slice sampling algorithm for univariate random variables.



2.3 Sampling from the truncated elliptical family of distributions

This section is devoted to describing our slice sampling algorithm with Gibbs steps to generate random observations from a multivariate elliptical distribution with strictly decreasing dgf.

We first consider a p-variate truncated elliptical distribution with zero location parameter, positive-definite scale matrix $\mathbf{R} \in \mathbb{R}^{p \times p}$, dgf g, and truncation region $A = \{(x_1, \ldots, x_p)^\top : a_1 < x_1 < b_1, \ldots, a_p < x_p < b_p\} = \{\mathbf{x} : \mathbf{a} < \mathbf{x} < \mathbf{b}\}$, $\mathbf{a}, \mathbf{b} \in \mathbb{R}^p$, in other

words, we will consider $\mathbf{X} \sim \mathrm{TE}\ell_p(\mathbf{0}, \mathbf{R}; g, A)$. Here $\mathbf{R}$ is a correlation matrix, such that the scale matrix can be written as $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{R}\mathbf{\Lambda}$, where $\mathbf{\Lambda} = \mathrm{diag}(\sqrt{\sigma_{11}}, \sqrt{\sigma_{22}}, \ldots, \sqrt{\sigma_{pp}})$. The pdf of $\mathbf{X}$ is given by

$$f_{\mathbf{X}}(\mathbf{x}) \propto g\left(\mathbf{x}^\top \mathbf{R}^{-1}\mathbf{x}\right) \mathbb{I}\left(\mathbf{x} \in A\right). \tag{2.3}$$

Now, in order to sample uniformly from the $(p+1)$-dimensional region under the plot of $f_{\mathbf{X}}(\mathbf{x})$, we introduce an auxiliary variable Y, such that the joint pdf of $\mathbf{X}$ and Y is

$$f_{\mathbf{X},Y}(\mathbf{x}, y) \propto \mathbb{I}\left(0 < y < g\left(\mathbf{x}^\top \mathbf{R}^{-1}\mathbf{x}\right)\right) \mathbb{I}\left(\mathbf{a} < \mathbf{x} < \mathbf{b}\right). \tag{2.4}$$

It is enough to calculate the conditional distributions of $Y \mid \mathbf{X}$ and $\mathbf{X} \mid Y$ to establish our slice sampling algorithm with Gibbs steps to generate random observations from the pdf in (2.4). These are given by:

$$f_{Y|\mathbf{X}}(y \mid \mathbf{x}) \propto \mathbb{I}\big(0 < y < g\big(\mathbf{x}^\top \mathbf{R}^{-1}\mathbf{x}\big)\big) \quad \text{and}$$
$$f_{\mathbf{X}|Y}(\mathbf{x} \mid y) \propto \mathbb{I}\big(\{\mathbf{x} : y < g\big(\mathbf{x}^\top \mathbf{R}^{-1}\mathbf{x}\big) \wedge \mathbf{a} < \mathbf{x} < \mathbf{b}\}\big).$$

Note that sampling y from the distribution of $Y \mid (\mathbf{X} = \mathbf{x})$ is straightforward, but sampling from $\mathbf{X} \mid (Y = y)$ is not trivial. Thus, we use the idea of Damien & Walker (2001), and Ho *et al.* (2012), which consists in sampling each element of $\mathbf{X}$ in succession from the full conditional distributions; in other words, we will apply the Gibbs sampler algorithm within the slice sampling mechanism. Therefore, note that the full conditional distributions are

$$f_{X_j|\mathbf{X}_{-j},Y}(x_j|\mathbf{x}_{-j}, y) \propto \mathbb{I}(x_j \in A_j),$$

where $\mathbf{X}_{-j} = (X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_p)^\top$ and $A_j = \{x_j : y < g\big(\mathbf{x}^\top \mathbf{R}^{-1}\mathbf{x}\big) \wedge a_j < x_j < b_j\}$, for all $j = 1, \ldots, p$. To find the elements of the real set A_j, let y be the value sampled at the current iteration of the algorithm and $\kappa_y = g^{-1}(y)$. Then, we have that

1. Since g is a strictly decreasing function, it follows that $y < g(\mathbf{x}^\top \mathbf{R}^{-1}\mathbf{x})$ is equivalent to $\kappa_y > \mathbf{x}^\top \mathbf{R}^{-1}\mathbf{x}$.

2. Write $\mathbf{x}^\top \mathbf{R}^{-1}\mathbf{x} = \rho^{jj}\left(x_j - \lambda_j\right)^2 - \rho^{jj}\lambda_j^2 + \eta_j$, where ρ^{ij} is the (i,j)th element of the inverse of $\mathbf{R}$, $\eta_j = \sum_{t \neq j}\sum_{r \neq j} x_t x_r \rho^{tr}$ and $\lambda_j = -\dfrac{1}{\rho^{jj}}\sum_{r \neq j} x_r \rho^{jr}$.

3. Combining items 1 and 2, we obtain that $\lambda_j - \tau_j < x_j < \lambda_j + \tau_j$, where $\tau_j = \left(\lambda_j^2 + \dfrac{1}{\rho^{jj}}\left(\kappa_y - \eta_j\right)\right)^{1/2}$.

4. Because $x_j \in (a_j, b_j)$, thereby $a_j^* = \max(a_j, \lambda_j - \tau_j) < x_j < \min(b_j, \lambda_j + \tau_j) = b_j^*$.

It follows that $A_j = (a_j^*, b_j^*)$. The steps to draw n samples from a p-variate truncated elliptical distribution $\mathbf{X} \sim \text{TE}\ell_p(\mathbf{0}, \mathbf{R}; g, A)$ are summarized in Algorithm 1. As seen, only univariate uniform simulations are involved in the algorithm, which is fast to compute. Note also that the assumption that the dgf g is strictly decreasing has been used in step 1. However, when it is not possible to find an analytical expression for $\kappa_y = g^{-1}(y)$, a numerical method is used, leading us to a more computationally expensive algorithm.

Algorithm 1 Slice sampling algorithm

Require: Sample size $n \geqslant 1$, initial value $\mathbf{x}_0 = (x_1^{(0)}, \ldots, x_p^{(0)})^\top \in \mathbb{R}^p$, scale matrix $\mathbf{R} \in \mathbb{R}^{p \times p}$, lower bound $\mathbf{a} \in \mathbb{R}^p$, upper bound $\mathbf{b} \in \mathbb{R}^p$, and strictly decreasing dgf $g(t), t \geqslant 0$.

1: **for** $i \leftarrow 1$ **to** n **do**
2: Simulate y_i from $Y \mid (\mathbf{X} = \mathbf{x}_{i-1}) \sim U(0, g(\mathbf{x}_{i-1}^\top \mathbf{R}^{-1} \mathbf{x}_{i-1}))$
3: $\kappa_y \leftarrow g^{-1}(y_i)$
4: **for** $j \leftarrow 1$ **to** p **do**
5: $\eta_j \leftarrow \sum_{t \neq j} \sum_{r \neq j} x_t x_r \rho^{tr}$
6: $\lambda_j \leftarrow -\frac{1}{\rho^{jj}} \sum_{r \neq j} x_r \rho^{jr}$
7: $\tau_j \leftarrow \left(\lambda_j^2 + \frac{1}{\rho^{jj}} (\kappa_y - \eta_j) \right)^{1/2}$
8: Sample $x_j^{(i)}$ from $X_j \mid (\mathbf{X}_{-j} = \mathbf{x}_{-j}^{(i)}, Y = y) \sim U(\max(a_j, \lambda_j - \tau_j), \min(b_j, \lambda_j + \tau_j))$, where $\mathbf{x}_{-j}^{(i)} = (x_1^{(i)}, \ldots, x_{j-1}^{(i)}, x_{j+1}^{(i-1)}, \ldots, x_p^{(i-1)})^\top$
9: $\mathbf{x}_i[j] \leftarrow x_j^{(i)}$
10: $\mathbf{X}[i, j] \leftarrow x_j^{(i)}$
11: **end for**
12: **end for**

Ensure: $\mathbf{X}$

Additionally, the slice sampler method generates random observations conditioned on previous values, resulting in a sequence of correlated samples. Thus, it is essential to analyze the dependence effect. See, for instance, Robert & Casella (2010), and Section A.3 from Appendix, where samples generated from specific bivariate distributions were observed, and the autocorrelation drops quickly, being negligibly small when lags become large, evidencing well mixing and quickly converging. The sampling method described here can be extended for a general dgf g by constructing an adequate "slice" at each iteration. Please refer to Section 4.1 in Neal (2003).

Moreover, members of the truncated elliptical family of distributions are closed under affine transformations (Fang, 2018). Hence drawing samples from $\mathbf{Y} \sim \text{TE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g, (\mathbf{a}, \mathbf{b}))$ may be readily done by sampling first from $\mathbf{X} \sim \text{TE}\ell_p(\mathbf{0}, \mathbf{R}; g, (\mathbf{a}^*, \mathbf{b}^*))$ and then recovering $\mathbf{Y}$ by the following transformation $\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{X}$, such that $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\mathbf{R}\boldsymbol{\Lambda}$, $\mathbf{a}^* = \boldsymbol{\Lambda}^{-1}(\mathbf{a} - \boldsymbol{\mu})$, and $\mathbf{b}^* = \boldsymbol{\Lambda}^{-1}(\mathbf{b} - \boldsymbol{\mu})$.

2.4 Moments of truncated multivariate elliptical distributions

Here we describe an adaptation of the method proposed by Galarza *et al.* (2022a) to compute the first moment and the variance-covariance matrix of a random vector whose distribution belongs to the truncated elliptical family. Furthermore, this method will be applied to some well-known distributions.

Let $\mathbf{Y}$ be a p-variate random vector that follows a truncated multivariate elliptical distribution with location parameter $\boldsymbol{\mu} \in \mathbb{R}^p$, positive-definite scale matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, dgf g, and support $A \subseteq \mathbb{R}^p$, i.e., $\mathbf{Y} \sim \text{TE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g, A)$. The more straightforward approach for computing the first two moments of $\mathbf{Y}$ is to use Monte Carlo integration. Following this method, the moments $\mathbb{E}(\mathbf{Y}), \mathbb{E}(\mathbf{Y}\mathbf{Y}^\top)$, and $\text{Cov}(\mathbf{Y})$ can be approximated by

$$\widehat{\mathbb{E}(\mathbf{Y})} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{y}_i, \quad \widehat{\mathbb{E}(\mathbf{Y}\mathbf{Y}^\top)} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{y}_i\mathbf{y}_i^\top, \quad \text{and} \quad \widehat{\text{Cov}(\mathbf{Y})} = \widehat{\mathbb{E}(\mathbf{Y}\mathbf{Y}^\top)} - \widehat{\mathbb{E}(\mathbf{Y})}\widehat{\mathbb{E}(\mathbf{Y})}^\top,$$

where $\mathbf{y}_i$ is the ith sample draws independently from $\text{TE}\ell_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g, A)$. However, it is well-known that the execution time needed to perform Monte Carlo integration depends on the algorithm employed to draw samples, the number of random points (n) used in the approximation, and the length of the random vector (p). Then, it depends on some variables that might represent a considerable computational effort. Nevertheless, we can save time when the random vector $\mathbf{Y}$ has non-truncated components, following the idea of Galarza *et al.* (2022a). They proposed to decompose $\mathbf{Y}$ into two vectors, $\mathbf{Y}_1$ and $\mathbf{Y}_2$, in such a way that $\mathbf{Y}_1$ is the random vector of truncated components and $\mathbf{Y}_2$ is the non-truncated part. Then, the first two moments for the truncated variables are computed using any method, and the remaining moments are computed using properties of the conditional expectation. Before showing our algorithm, we state an extremely important result.

Proposition 1 (*Marginal and conditional distribution of the Elliptical family*)**.** *Let $\boldsymbol{X}$ be partitioned into two vectors, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ of dimensions p_1 and p_2, such that $p = p_1 + p_2$, and $\boldsymbol{X} = (\boldsymbol{X}_1^\top, \boldsymbol{X}_2^\top)^\top$ has joint multivariate elliptical distribution as follows*

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \sim E\ell_{p_1+p_2}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g^{(p_1+p_2)}\right),$$

with

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

partitioned such that $\boldsymbol{\mu}_1 \in \mathbb{R}^{p_1}$, $\boldsymbol{\mu}_2 \in \mathbb{R}^{p_2}$ are location vectors, $\boldsymbol{\Sigma}_{11} \in \mathbb{R}^{p_1\times p_1}, \boldsymbol{\Sigma}_{22} \in \mathbb{R}^{p_2\times p_2}, \boldsymbol{\Sigma}_{12} \in \mathbb{R}^{p_1\times p_2}, \boldsymbol{\Sigma}_{21} \in \mathbb{R}^{p_2\times p_1}$ are dispersion matrices, and $g^{(p_1+p_2)}$ is the dgf. The distributions of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2 \mid (\boldsymbol{X}_1 = \boldsymbol{x})$ belong to the elliptical family of distributions, as

follows

$$\begin{aligned}\boldsymbol{X}_1 &\sim E\ell_{p_1}\left(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}; g^{(p_1)}\right)\\ \boldsymbol{X}_2 \mid (\boldsymbol{X}_1 = \boldsymbol{x}) &\sim E\ell_{p_2}\left(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}; g_{\boldsymbol{x}}^{(p_2)}\right).\end{aligned}$$

For additional information, refer to Chapter 2 in Fang (2018).

Building on Proposition 1, if $\mathbf{Y}_1$ denotes the vector of variables truncated within region $A_1 = \{\mathbf{y}_1 : \mathbf{a}_1 < \mathbf{y}_1 < \mathbf{b}_1\}$ and $\mathbf{Y}_2$ represents the vector of non-truncated variables, with the combined truncation region specified as $A = A_1 \times \mathbb{R}^{p_2}$, then it follows that

$$\begin{aligned}\mathbf{Y}_1 &\sim \mathrm{TE}\ell_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}; g^{(p_1)}, A_1)\\ \mathbf{Y}_2 \mid (\mathbf{Y}_1 = \mathbf{y}) &\sim \mathrm{E}\ell_{p_2}\left(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y} - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}; g_{\mathbf{y}}^{(p_2)}\right).\end{aligned}$$

Now let $\boldsymbol{\xi}_1 = \mathbb{E}(\mathbf{Y}_1)$ and $\boldsymbol{\Omega}_{11} = \mathrm{Cov}(\mathbf{Y}_1)$. Then, using the results exposed by Galarza *et al.* (2022a), the first moment of $\mathbf{Y}$ can be computed by

$$\begin{aligned}\mathbb{E}(\mathbf{Y}) &= \mathbb{E}(\mathbb{E}(\mathbf{Y} \mid \mathbf{Y}_1)) = \mathbb{E}\begin{pmatrix}\mathbf{Y}_1\\ \mathbb{E}(\mathbf{Y}_2|\mathbf{Y}_1)\end{pmatrix} = \mathbb{E}\begin{pmatrix}\mathbf{Y}_1\\ \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{Y}_1 - \boldsymbol{\mu}_1)\end{pmatrix}\\ &= \begin{pmatrix}\boldsymbol{\xi}_1\\ \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{\xi}_1 - \boldsymbol{\mu}_1)\end{pmatrix}.\end{aligned} \tag{2.5}$$

On the other hand, the variance-covariance matrix of $\mathbf{Y}$ is given by

$$\mathrm{Cov}(\mathbf{Y}) = \begin{pmatrix}\boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{11}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Omega}_{11} & \boldsymbol{\Psi}_{22}\end{pmatrix}, \tag{2.6}$$

where $\boldsymbol{\Psi}_{22} = \omega_{2.1}\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\left(\omega_{2.1}\mathbf{I}_{p_1} - \boldsymbol{\Omega}_{11}\boldsymbol{\Sigma}_{11}^{-1}\right)\boldsymbol{\Sigma}_{12}$, $\omega_{2.1} = \mathbb{E}\left(h(\mathbf{Y}_1)\right)$, $h(\mathbf{Y}_1) = \mathrm{tr}\left(\mathrm{Cov}(\mathbf{Y}_2|\mathbf{Y}_1)\ \boldsymbol{\Sigma}_{2.1}^{-1}\right)/p_2$, and $\boldsymbol{\Sigma}_{2.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$. Note that $h(\mathbf{Y}_1)$ depends on the conditional distribution of $\mathbf{Y}_2 \,|\, \mathbf{Y}_1$, taking a different expression for each member of the elliptical family.

Thereby, this work proposes to use Monte Carlo integration to approximate the truncated moments $\boldsymbol{\xi}_1$, $\boldsymbol{\Omega}_{11}$, and $\omega_{2.1}$ (when necessary) instead of computing them using recursion-based methods (e.g., Galarza *et al.*, 2021c) which can be computationally more expensive for high dimensions or complex elliptical distributions. For instance, to the best of our knowledge, closed-form expressions to compute the first two moments of truncated elliptical distributions only exist for the TMVN and TMVT distributions. These expressions are not very efficient (as demonstrated in a simulation study in Section 2.4.2) in practice because, due to their recursive nature, they suffer from error propagation and require an intensive calculation of probabilities, which in turn depend on numerical approximation methods. A summary of how our method works is given in Algorithm 2. The first moment and the variance-covariance matrix are approximated by (2.5)–(2.6).

Algorithm 2 Mean and variance approximation

Require: Sample size $n \geqslant 1$, location parameter $\boldsymbol{\mu} \in \mathbb{R}^p$, scale matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, lower bound $\mathbf{a} \in \mathbb{R}^p$, upper bound $\mathbf{b} \in \mathbb{R}^p$, and strictly decreasing dgf $g(t), t \geqslant 0$.

1: Identify: $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{22}, \boldsymbol{\Sigma}_{12}, A_1 = \{\mathbf{y}_1 : \mathbf{a}_1 < \mathbf{y}_1 < \mathbf{b}_1\}$

2: **for** $i \leftarrow 1$ **to** n **do**

3: Draw $\mathbf{y}_{1i}$ from $\mathbf{Y}_1 \sim \mathrm{TE}\ell_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}; g^{(p_1)}, A_1)$ using Algorithm 1

4: **end for**

5: $\widehat{\boldsymbol{\xi}}_1 \leftarrow \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\mathbf{y}_{1i}; \qquad \widehat{\boldsymbol{\Omega}}_{11} \leftarrow \dfrac{1}{n}\sum_{i=1}^{n}\mathbf{y}_{1i}\mathbf{y}_{1i}^\top - \widehat{\boldsymbol{\xi}}_1\widehat{\boldsymbol{\xi}}_1^\top; \qquad \widehat{\omega}_{2.1} \leftarrow \dfrac{1}{n}\sum_{i=1}^{n}h(\mathbf{y}_{1i})$

6: $\widehat{\mathbb{E}(\mathbf{Y})} \leftarrow \begin{pmatrix} \widehat{\boldsymbol{\xi}}_1 \\ \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\widehat{\boldsymbol{\xi}}_1 - \boldsymbol{\mu}_1) \end{pmatrix}$

7: $\widehat{\mathrm{Cov}(\mathbf{Y})} \leftarrow \begin{pmatrix} \widehat{\boldsymbol{\Omega}}_{11} & \widehat{\boldsymbol{\Omega}}_{11}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\widehat{\boldsymbol{\Omega}}_{11} & \widehat{\boldsymbol{\Psi}}_{22} \end{pmatrix}$

8: $\widehat{\mathbb{E}(\mathbf{Y}\mathbf{Y}^\top)} \leftarrow \widehat{\mathrm{Cov}(\mathbf{Y})} + \widehat{\mathbb{E}(\mathbf{Y})}\widehat{\mathbb{E}(\mathbf{Y})}^\top$

Ensure: $\widehat{\mathbb{E}(\mathbf{Y})}, \widehat{\mathbb{E}(\mathbf{Y}\mathbf{Y}^\top)}, \widehat{\mathrm{Cov}(\mathbf{Y})}$

2.4.1 Mean and variance for the truncated elliptical distributions

This subsection is devoted to analyzing how Algorithm 2 works for some specific distributions considering all the conditions used previously. Let $\boldsymbol{\mu}_{2.1}(\mathbf{x}) = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x} - \boldsymbol{\mu}_1)$, $\boldsymbol{\Sigma}_{2.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, $\delta_1(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_1)^\top\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x} - \boldsymbol{\mu}_1)$, and $A = A_1 \times \mathbb{R}^{p_2}$. Here, it is considered that $\mathbf{Y} \overset{d}{=} \mathbf{X}|(\mathbf{X} \in A)$.

- **Normal**: If $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the marginal distribution is $\mathbf{X}_1 \sim N_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and the conditional distribution is $\mathbf{X}_2|(\mathbf{X}_1 = \mathbf{x}) \sim N_{p_2}(\boldsymbol{\mu}_{2.1}(\mathbf{x}), \boldsymbol{\Sigma}_{2.1})$. Then, to compute the moments for $\mathbf{Y} \sim \mathrm{TN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; A)$ with the conditions above, Algorithm 2 firstly samples $\mathbf{Y}_1$ from the TMVN distribution with location parameter $\boldsymbol{\mu}_1$, scale matrix $\boldsymbol{\Sigma}_{11}$, truncation region A_1, and $\omega_{2.1} = 1$.

- **Student-t**: If $\mathbf{X} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, $\nu > 0$, the marginal and conditional distributions are $\mathbf{X}_1 \sim t_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \nu)$ and $\mathbf{X}_2|(\mathbf{X}_1 = \mathbf{x}) \sim t_{p_2}(\boldsymbol{\mu}_{2.1}(\mathbf{x}), \lambda(\mathbf{x})\boldsymbol{\Sigma}_{2.1}, \nu + p_1)$, respectively, such that $\lambda(\mathbf{x}) = (\nu + \delta_1(\mathbf{x}))/(\nu + p_1)$. For this distribution $\mathbb{E}(\mathbf{X})$ exists, if $\nu > 1$ and $\mathrm{Cov}(\mathbf{X})$ exists, if $\nu > 2$. Therefore, the moments for $\mathbf{Y} \sim \mathrm{T}t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; A)$ are computed by sampling $\mathbf{Y}_1$ from the TMVT distribution with location parameter $\boldsymbol{\mu}_1$, scale matrix $\boldsymbol{\Sigma}_{11}$, ν degrees of freedom, truncation region A_1, and $\omega_{2.1}$ given by

$$\omega_{2.1} = \frac{\nu + \mathbb{E}(\delta_1(\mathbf{Y}_1))}{\nu + p_1 - 2},$$

with $\mathbb{E}(\delta_1(\mathbf{Y}_1)) = \mathrm{tr}(\boldsymbol{\Omega}_{11}\boldsymbol{\Sigma}_{11}^{-1}) + (\boldsymbol{\xi}_1 - \boldsymbol{\mu}_1)^\top\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{\xi}_1 - \boldsymbol{\mu}_1)$. It is worth mentioning that for doubly truncated variables, variables which are constrained both below and above, the mean and the variance exist for all $\nu > 0$. Then, if $\mathbf{Y}$ has at least two doubly truncated components, the mean and the variance-covariance matrix exist

for all $\nu > 0$. For more details about the existence of these moments, please refer to Galarza *et al.* (2022a).

- **Pearson VII**: If $\mathbf{X} \sim \text{PVII}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, m, \nu)$, $m > p/2$, $\nu > 0$, then $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$ and $\text{Cov}(\mathbf{X}) = \dfrac{\nu}{2m-p-2}\boldsymbol{\Sigma}$. In this case, $\mathbb{E}(\mathbf{X})$ exists, if $m > (p+1)/2$ and $\text{Cov}(\mathbf{X})$ exists, if $m > (p+2)/2$. The marginal and the conditional distributions are $\mathbf{X}_1 \sim \text{PVII}_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, m - p_2/2, \nu)$ and $\mathbf{X}_2 \mid (\mathbf{X}_1 = \mathbf{x}) \sim \text{PVII}_{p_2}(\boldsymbol{\mu}_{2.1}(\mathbf{x}), \boldsymbol{\Sigma}_{2.1}, m, \delta_1(\mathbf{x}) + \nu)$, respectively. So, the proposed algorithm for $\mathbf{Y} \sim \text{TPVII}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, m, \nu; A)$ was implemented by sampling $\mathbf{Y}_1$ from the truncated multivariate Pearson VII distribution with location parameter $\boldsymbol{\mu}_1$, scale matrix $\boldsymbol{\Sigma}_{11}$, additional parameters $m - p_2/2 > p_1/2$, $\nu > 0$, and truncation region A_1. The constant $\omega_{2.1}$ is

$$\omega_{2.1} = \frac{\nu + \mathbb{E}(\delta_1(\mathbf{Y}_1))}{2m - p_2 - 2},$$

where $\mathbb{E}(\delta_1(\mathbf{Y}_1))$ is given as in the Student-t distribution. For this distribution, first and second moments for doubly truncated variables exist for all $m > p/2$. Then, if $\mathbf{X}$ has at least two doubly truncated variables, the mean and the variance exist for all $m > p/2$. For more details about the existence of the moments, refer to Subsection A.1.1 in the Appendix section.

- **Slash**: If $\mathbf{X} \sim \text{SL}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, $\nu > 0$, then $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$ and $\text{Cov}(\mathbf{X}) = \dfrac{\nu}{\nu-1}\boldsymbol{\Sigma}$. In this case, $\text{Cov}(\mathbf{X})$ exists, if $\nu > 1$. The marginal distribution is $\mathbf{X}_1 \sim \text{SL}_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \nu)$ and the conditional distribution is $\mathbf{X}_2 \mid (\mathbf{X}_1 = \mathbf{x}) \sim \text{E}\ell_{p_2}(\boldsymbol{\mu}_{2.1}(\mathbf{x}), \boldsymbol{\Sigma}_{2.1}; g_x^{(p_2)})$, such that $g_x^{(p_2)}(t) = \displaystyle\int_0^1 u^{\nu+p/2-1} \exp\{-u(t + \delta_1(\mathbf{x}))/2\}\, \mathrm{d}u$. Note that $\mathbf{X}_2 \,|\, \mathbf{X}_1$ does not follow a slash distribution, but its distribution belongs to the elliptical family (see Appendix, Subsection A.1.2). So, the moments for $\mathbf{Y} \sim \text{TSL}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; A)$ are calculated by sampling $\mathbf{Y}_1$ from the truncated multivariate slash distribution with location parameter $\boldsymbol{\mu}_1$, scale matrix $\boldsymbol{\Sigma}_{11}$, ν degrees of freedom, and truncation region A_1. The constant $\omega_{2.1}$ is

$$\omega_{2.1} = \frac{\nu}{\nu-1}\,\mathbb{E}\left(\frac{SL_{p_1}(\mathbf{Y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \nu - 1)}{SL_{p_1}(\mathbf{Y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \nu)}\right),$$

where $SL_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ denotes the pdf of a p-variate slash distribution with parameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}$, and ν. As usual, this constant can also be approximated via Monte Carlo integration.

- **Contaminated Normal**: If $\mathbf{X} \sim \text{CN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu, \rho)$, $0 \leqslant \nu \leqslant 1$, $0 < \rho \leqslant 1$, then the marginal is $\mathbf{X}_1 \sim \text{CN}_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \nu, \rho)$ and the conditional distribution is $\mathbf{X}_2 | (\mathbf{X}_1 = \mathbf{x}) \sim \text{CN}_{p_2}(\boldsymbol{\mu}_{2.1}(\mathbf{x}), \boldsymbol{\Sigma}_{2.1}, \nu_{2.1}(\mathbf{x}), \rho)$, with $\nu_{2.1}(\mathbf{x}) = \nu\phi_{p_1}(\mathbf{x}; \boldsymbol{\mu}_1, \rho^{-1}\boldsymbol{\Sigma}_{11})/\kappa(\mathbf{x})$, $\kappa(\mathbf{x}) = \nu\phi_{p_1}(\mathbf{x}; \boldsymbol{\mu}_1, \rho^{-1}\boldsymbol{\Sigma}_{11}) + (1-\nu)\phi_{p_1}(\mathbf{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$, and $\phi_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denoting the pdf of a p-variate normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$ evaluated at point $\mathbf{x} \in \mathbb{R}^p$.

Thus, our algorithm for $\mathbf{Y} \sim \text{TCN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu, \rho; A)$ samples $\mathbf{Y}_1$ from the truncated contaminated normal distribution with parameters $\boldsymbol{\mu}_1$, $\boldsymbol{\Sigma}_{11}$, ν, and ρ. The constant is $\omega_{2.1} = \nu^*_{2.1}/\rho + 1 - \nu^*_{2.1}$, where $\nu^*_{2.1} = \mathbb{E}(\nu_{2.1}(\mathbf{Y}_1))$. This value is also approximated via Monte Carlo integration.

- **Power exponential**: If $\mathbf{X} \sim \text{PE}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \beta)$, $\beta > 0$, then $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$ and $Cov(\mathbf{X}) = \omega\boldsymbol{\Sigma}$, with $\omega = 2^{1/\beta}\Gamma\left(\frac{n+2}{2\beta}\right) / \left(n\Gamma\left(\frac{n}{2\beta}\right)\right)$. The marginal distribution of $\mathbf{X}_1$ belongs to the elliptical family of distributions with dgf

$$g^{(p_1)}(t) = t^{\frac{p-p_1}{2}} \int_0^1 w^{\frac{p_1-p}{2}} (1-w)^{\frac{p-p_1}{2}-1} \exp\{-\frac{t^\beta}{2w^\beta}\}\, \mathrm{d}w,$$

i.e., $\mathbf{X}_1 \sim \text{E}\ell_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}; g^{(p_1)})$. The conditional distribution is $\mathbf{X}_2 \,|\, (\mathbf{X}_1 = \mathbf{x}) \sim \text{E}\ell_{p_2}(\boldsymbol{\mu}_{2.1}(\mathbf{x}), \boldsymbol{\Sigma}_{2.1}; g^{(p_2)}_x)$ where $g^{(p_2)}_x(t) = \exp\left\{-\frac{1}{2}(t + \delta_1(\mathbf{x}))^\beta\right\}$ (Gómez *et al.*, 1998). Therefore to approximate the moments for $\mathbf{Y} \sim \text{TPE}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \beta; A)$, we will use a different approach that consists of drawing points from the whole random vector of length p and then approximate the moments using Monte Carlo integration, since sampling directly from the marginal distribution of $\mathbf{Y}_1$ could be really complicated, as well as to compute $\omega_{2.1}$.

2.4.2 Numerical examples

2.4.2.1 Simulation study I

We illustrate the application of the method by considering a random vector $\mathbf{X} = (X_1, X_2, X_3, X_4)^\top$ of length 4 with truncated Student-t distribution, $\mathbf{X} \sim \text{T}t_4(\boldsymbol{\mu}, \boldsymbol{\Sigma}; (\mathbf{a}\ , \mathbf{b}))$, characterized by the following parameters:

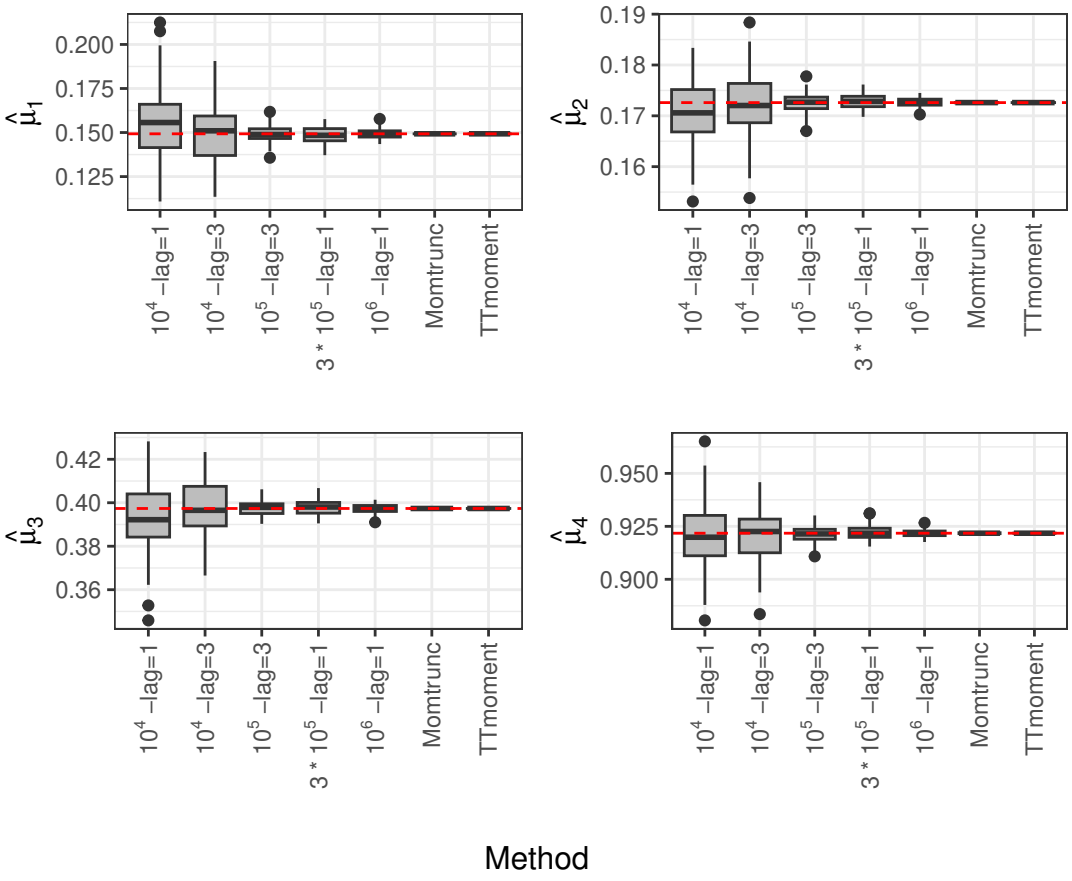
$$\boldsymbol{\mu} = \begin{pmatrix} 0.125 \\ 0.25 \\ 0.375 \\ 0.50 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} 6.321 & -1.780 & -4.596 & -2.630 \\ -1.780 & 4.149 & 1.175 & 0.568 \\ -4.596 & 1.175 & 3.494 & 1.691 \\ -2.630 & 0.568 & 1.691 & 5.738 \end{pmatrix}, \mathbf{a} = \begin{pmatrix} -1.297 \\ -\infty \\ -0.747 \\ -1.269 \end{pmatrix}, \text{ and } \mathbf{b} = \begin{pmatrix} 4.449 \\ \infty \\ 4.976 \\ 3.870 \end{pmatrix}.$$

In this example, X_2 is not truncated, and the other components are doubly truncated. The objective is to study the performance of the estimates for the mean and the variance-covariance elements obtained through Algorithm 2, considering a different number of samples for approximation (n) and thinning, where thinning is a factor for reducing autocorrelation between observations. After that, we compare those results with the ones obtained from the R functions `meanvarTMD` and `TT.moment` available in packages `MomTrunc` and `TTmoment`, respectively.

Figure 9 displays the boxplot for each element of the mean vector based on 100 estimates obtained through our proposal considering $n = 10^4$ with *thinning* = 1 and 3, $n = 10^5$ with *thinning* = 3, $n = 3 \times 10^5$, and $n = 10^6$ with no thinning (*thinning* = 1). It also shows the results came from functions `meanvarTMD` and `TT.moment`. It is worth

noting that even though these two latter functions compute the first two moments using closed-form expressions, there is a notable variability due to the approximation methods used during the calculations. Hence, we will refer to the values obtained through these functions as "estimates" rather than "true values". On the other hand, the red dashed line represents the median of the estimates achieved from function `TT.moment`.

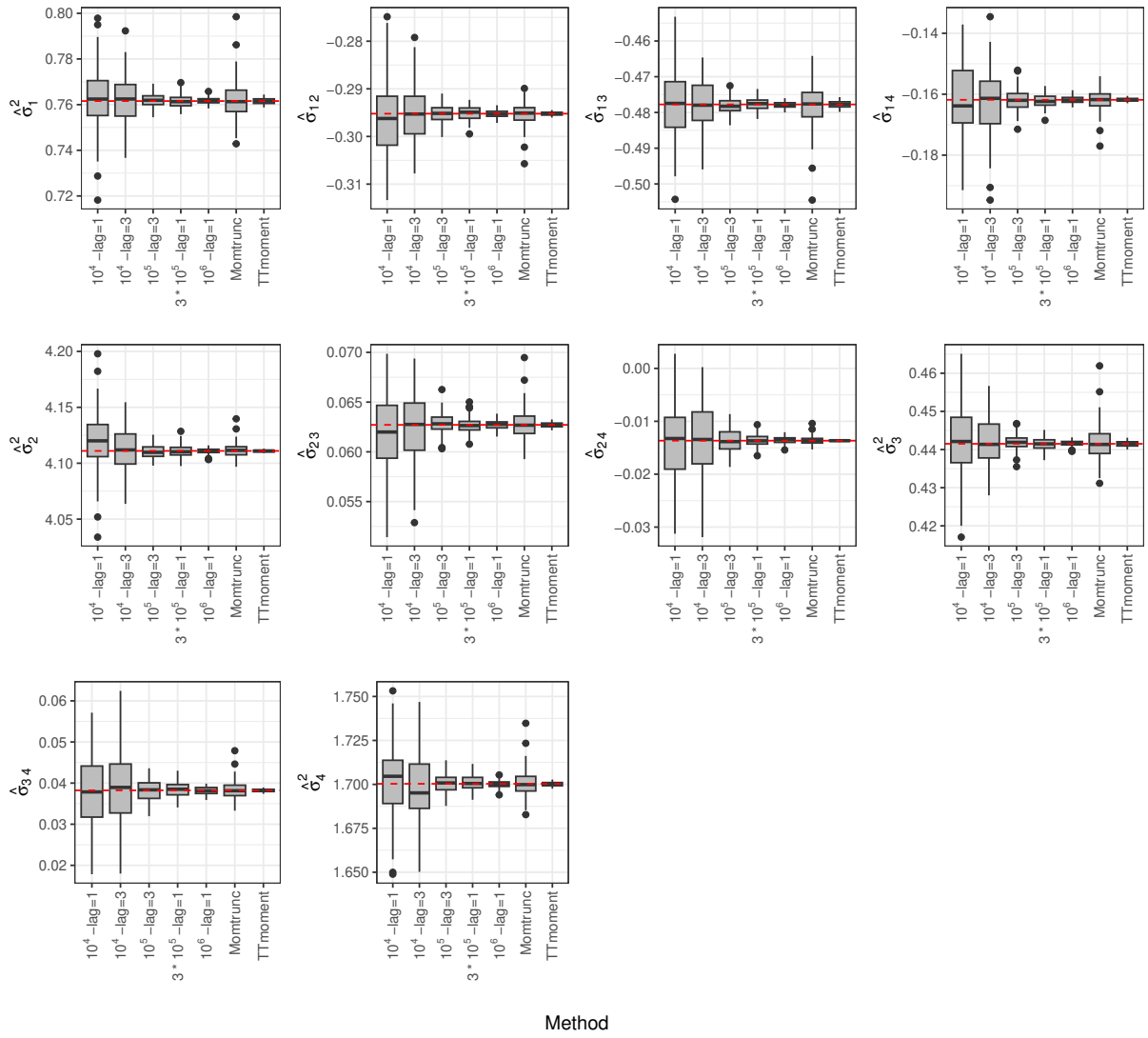
Figure 9 – Boxplot based on 100 estimates of the truncated mean. The red line denotes the median of `TT.moment` estimates.



For the case of $n = 10^4$, the estimates obtained with no thinning have more variability than those with *thinning* $= 3$ (observations with lower autocorrelation). The median of `TT.moment` estimates is closer to the median of our method in most cases, except for $n = 10^4$ with *thinning* $= 1$. As expected, the variability in the estimates was reduced when the number of observations was increased. The estimates from our algorithm with 10^5 samples and *thinning* $= 3$ were similar to those considering $n = 3 \times 10^5$ and no thinning. Recall that both methods need to generate the same number of samples; the only difference here is that the first one (*thinning* $= 3$) will need less memory space than the other one. The best results were obtained throughout `TT.moment` and `meanvarTMD` functions, and those are comparable with the estimates achieved from our proposal with $n = 10^6$ and no thinning.

Figure 10 shows the boxplot for the variance-covariance elements of the trun-

Figure 10 – Boxplot based on 100 estimates of the variance-covariance elements. The red line represents the median of the estimates obtained from function `TT.moment`.



cated random vector considering each method described above. We noticed a slight reduction in the variability of the estimates when considering a thinning equal to 3. Another interesting fact is observed when we set $n = 10^5$ and *thinning* $= 3$; in this case, it returned similar results to estimating the covariances with 3×10^5 samples and no thinning. The estimates achieved through our proposal considering $n = 10^6$ are comparable with those from `TT.moment`. In contrast, the estimates obtained from `meanvarTMD` are similar to those from MC with $n = 10^5$ and *thinning* $= 3$ in most cases, except for σ_{11}, σ_{33}, and σ_{13}. For these parameters, our method showed better performance.

2.4.2.2 Simulation study II

In the previous example, it was observed that the estimates obtained from Algorithm 2 with $n = 10^5$ and *thinning* $= 3$ are good enough to estimate the mean and

variance matrix of a multivariate ($p = 4$) variable with TMVT distribution, even though the best results were obtained through the `TT.moment` function. In this example, our goal is to analyze the computational time required for our method and the functions `meanvarTMD` and `TT.moment` to estimate the first two moments and the variance-covariance matrix of a p-variate random vector with TMVT distribution considering $p = 50, 100$, and 150. In each case, we set 10%, 20%, and 40% of the variables doubly truncated. The methods were run on a Windows 10 machine using R 4.0.3 on an Intel Core i7-7700 Processor with 3.60 GHz and 32 GB of RAM.

Table 1 displays the median of the CPU time (in seconds) required for our algorithm and functions `meanvarTMD` and `TT.moment`. For our proposal, three scenarios were considered $n = 10^4$ with no thinning, $n = 10^4$, and $n = 10^5$ with $thinning = 3$. The results are based on 100 simulations, and they were computed through the R function `microbenchmark`. This table also shows the relative time (R.Time) computed, taking the time used by our method with $n = 10^5$ and $thinning = 3$ as reference. We will refer to this configuration as the "reference method."

Table 1 – Median of the CPU time (in seconds) and relative time (R. Time) considering $n = 10^5$ with $thinning = 3$ as the reference method, based on 100 simulations.

Method	Measure	$p = 50$			$p = 100$			$p = 150$		
		10%	20%	40%	10%	20%	40%	10%	20%	40%
$n = 10^4$	Median	0.011	0.030	0.139	0.030	0.140	0.952	0.071	0.382	2.842
$thinning = 1$	R.Time	0.035	0.035	0.034	0.036	0.035	0.034	0.036	0.034	0.034
$n = 10^4$	Median	0.031	0.084	0.404	0.085	0.405	2.820	0.199	1.118	8.461
$thinning = 3$	R.Time	0.100	0.100	0.100	0.100	0.100	0.100	0.101	0.100	0.100
$n = 10^5$	Median	0.314	0.844	4.042	0.846	4.044	28.217	1.974	11.182	84.619
$thinning = 3$	R.Time	-	-	-	-	-	-	-	-	-
`meanvarTMD`	Median	0.118	4.102	49.189	3.781	48.681	367.243	21.209	157.179	1215.630
	R.Time	0.375	4.861	12.170	4.467	12.037	13.015	10.746	14.056	14.366
`TT.moment`	Median	7.452	24.027	94.408	62.026	202.704	789.641	242.701	800.360	3081.367
	R.Time	23.767	28.0473	23.358	73.279	50.122	27.984	122.974	71.574	36.414

For our algorithm, we observe that the time required to estimate the moments depends on the number of random observations sampled and the number of truncated variables. Note that estimating the moments with $n = 10^4$ took 3.50% of the time required for the reference method, and it is worth mentioning that the number of samples needed for the first method is 3.33% of the number of samples used for the reference one. Our proposal with $n = 10^4$ and $thinning = 3$ already needed 10% of the execution time used by the reference method. Observe that `meanvarTMD` was faster than the reference procedure for vectors with 5 doubly truncated variables. It also seems that the time needed by `meanvarTMD` depends only on the number of truncated variables. In comparison, the `TT.moment` function is much more time-consuming in all scenarios if compared with our proposal and function `meanvarTMD`, e.g., for a random vector of length $p = 100$ and

40 doubly truncated variables, it needed 28 times longer than the reference method. An additional example regarding the computational time required to compute the truncated moments for other elliptical distributions can be found in Appendix A.2.

2.5 Remarks

This chapter described an algorithm to generate random numbers from members of the truncated elliptical family of distributions with a strictly decreasing density generating function through a slice sampling algorithm with Gibbs sampler steps. In addition, we presented an efficient approach to approximate the first and the second moment for these distributions. We briefly introduced the functions available in our R package `relliptical` in order to perform sample generation and estimation of the first two moments. Simulation studies were performed to investigate the properties of estimates and the robustness of our algorithm. Moreover, we compared our approach with others available in the R software (only for the normal and Student-t cases), showing that our approach outperformed others in terms of precision and computational time. In the next chapters, we will illustrate the usefulness of truncated moments in analysing some datasets related to spatial censored data, censored time series, multivariate censored data, and longitudinal censored data. The results presented in this work can be reproduced through the R package `relliptical`, available at CRAN for download.

Future extensions of the work include considering a more general class of density generating functions (not only strictly decreasing) in the sampling method. Other methods could also be explored to sample from the truncated elliptical family, such as IA^{2}RMS (Martino *et al.*, 2015a) or the fast universal self-tuned sampler (FUSS) (Martino *et al.*, 2015b) within Gibbs. The first one is interesting because it returns asymptotically independent samples and tries to maintain the computational cost bounded (as in ARS and ARMS). About the latter one, it was demonstrated through simulation studies to be faster than some well-known MCMC methods for sampling from some specific bivariate distributions, besides the accuracy and generation of virtually independent samples. On the other hand, the method used to approximate moments for the truncated elliptical family can be extended to the context of asymmetric multivariate elliptical distributions, so the fast computation of their truncated moments may lead the way to propose more flexible and robust models relating to censored models for mixed-effects models, longitudinal data, and spatial models, among others.

3 Spatial censored linear model

3.1 Introduction

The algorithms previously proposed for simulating random observations and computing moments from the truncated elliptical family of distributions are applied in various statistical models, including censored data models, survival analysis, and Bayesian models with truncated parameter space, among others. For example, our `relliptical` package has been integrated into the `ARpLMEC` package (Olivari *et al.*, 2022), which estimates parameters for censored mixed-effects models using a symmetric elliptical error distribution. Mattos *et al.* (2022a) have also employed our methods in semiparametric mixed models for longitudinal data with censored responses and heavy tails. This chapter evaluates the performance of our methods and demonstrates their broad applicability. We focus on the Gaussian spatial censored linear (SCL) model as defined by Lachos *et al.* (2017) and Ordoñez *et al.* (2018).

Spatial data are common in fields such as ecology, environmental health, hydrology, and epidemiology. The main challenge in analyzing such data is selecting the correlation model that best fits the observations, such as the Matérn or exponential correlation structures. An added complexity occurs when data are subject to upper or lower detection limits. For instance, environmental (spatial) monitoring often deals with left-censored observations, where values fall below the minimum detectable limits of the instruments. As discussed by Lachos *et al.* (2017) and Ordoñez *et al.* (2018), the SCL model uses the SAEM algorithm for parameter estimation. This implementation is available in the R package `CensSpatial` Ordoñez *et al.* (2020), which utilizes the `rtmvnorm` function (available in the `tmvtnorm` package) to draw random observations from the truncated multivariate normal (TMVN) distribution and `optimx` for optimization.

The main objective of this chapter is to compare the estimates obtained using the Expectation-Maximization (EM, Dempster *et al.*, 1977), Stochastic Approximation EM (SAEM, Delyon *et al.*, 1999), and Monte Carlo EM (MCEM, Wei & Tanner, 1990) algorithms for the SCL model, particularly when random observations and moments are computed using our proposal (detailed in Chapter 2). We also aim to benchmark our methods against others available in R, such as `CensSpatial` and `MomTrunc`, through the analysis of real-world data concerning dioxin contamination (2,3,7,8-tetrachlorodibenzo-p-dioxin or TCDD). All methods discussed in this chapter are accessible via the R library `RcppCensSpatial` (Valeriano *et al.*, 2021a).

The chapter is organized as follows. Section 3.2 briefly introduces the linear

spatial model, and Section 3.3 describes the SCL model and the estimation procedures involving the EM, SAEM, and MCEM algorithms. Section 3.4 examines the asymptotic properties of the MCEM and SAEM estimates through simulation studies. Section 3.5 compares the performance of our method with that of `CensSpatial` by analyzing the Missouri dioxin contamination data.

3.2 Linear spatial model

Consider the real-valued Gaussian stochastic process $\{Y(\mathbf{s}), \mathbf{s} \in \mathbf{D} \subset \mathbb{R}^d\}$, as defined in Gaetan & Guyon (2010), where $\mathbb{R}^d$ is the d-dimensional Euclidean space. For any integer $n \geqslant 1$, it supposes that a realization of this process $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$ is observed at known sites (locations) $\mathbf{s}_i$, for $i \in \{1, \ldots, n\}$, where $\mathbf{s}_i$ is a d-dimensional vector of spatial coordinates, and the data is generated from the model

$$Y(\mathbf{s}_i) = \mu(\mathbf{s}_i) + \xi(\mathbf{s}_i), \tag{3.1}$$

where both the deterministic term $\mu(\mathbf{s}_i)$ and the stochastic term $\xi(\mathbf{s}_i)$ may depend on the spatial location at which $Y(\mathbf{s}_i)$ is observed. We assume that the stochastic errors are normally distributed with zero mean, $\mathbb{E}[\xi(\mathbf{s}_i)] = 0$, and the variation between spatial points is determined by a stationary covariance function $\text{Cov}(\xi(\mathbf{s}_i), \xi(\mathbf{s}_j)) = \text{Cov}(||\mathbf{h}||)$, where $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j \in \mathbb{R}^d$ and $||\cdot||$ denotes the Euclidean distance. Suppose that for some known functions of $\mathbf{s}_i$, say $x_1(\mathbf{s}_i), x_2(\mathbf{s}_i), \ldots, x_q(\mathbf{s}_i)$, the mean of the stochastic process is

$$\mu(\mathbf{s}_i) = \sum_{k=1}^{q} x_k(\mathbf{s}_i)\beta_k.$$

Equivalently, in matrix notation, we have the linear spatial model is given by

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}, \tag{3.2}$$

where $\mathbf{X}$ is the $n \times q$ matrix with $\mathbf{x}_i^\top = (x_1(\mathbf{s}_i), \ldots, x_q(\mathbf{s}_i))$ in the ith row, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)^\top$ is the $q \times 1$ vector of regression parameters, $\boldsymbol{\xi} = (\xi(\mathbf{s}_1), \ldots \xi(\mathbf{s}_n))^\top$ is the $n \times 1$ vector of errors, which has multivariate normal distribution with zero-mean and variance-covariance matrix $\boldsymbol{\Sigma} = [\text{Cov}(\mathbf{s}_i, \mathbf{s}_j)] = \sigma^2\mathbf{R}(\phi) + \tau^2\mathbf{I}_n$. In geostatistical analysis, the parameter $\tau^2 > 0$ is referred to as the nugget effect, while $\sigma^2 > 0$ represents the partial sill. Additionally, $\mathbf{R} = \mathbf{R}(\phi) = [r_{ij}]$ is defined as the $n \times n$ symmetric matrix, where the diagonal elements $r_{ii} = 1$ for all $i \in \{1, \ldots, n\}$, and the off-diagonal elements are calculated using an isotropic spatial correlation function. Various spatial correlation functions were considered, including exponential, Matérn, Gaussian, and power exponential functions. For further reference, see works such as Diggle & Ribeiro (2007), Lachos *et al.* (2017), and Valeriano *et al.* (2021b). It is also assumed that $\boldsymbol{\Sigma}$ is non-singular and that $\mathbf{X}$ has full rank.

3.3 Model formulation and parameter estimation

This section presents the linear spatial model with censored responses previously formulated by Lachos *et al.* (2017) and Ordoñez *et al.* (2018). We explore using the EM, MCEM, and SAEM algorithms to obtain the maximum likelihood (ML) estimates of the model parameters and propose a methodology to approximate the standard error estimates for each method.

3.3.1 Spatial censored linear model

Now, suppose that the linear spatial model defined in (3.2), where $\boldsymbol{\xi} \sim N_n(\mathbf{0}, \boldsymbol{\Sigma})$, has the particularity that the response variable Y_i is not fully observed for all i. Let $R_i \subseteq \mathbb{R}$ denote the censoring region, such that Y_i is not observed if $Y_i \in R_i$. Further, let (V_i, C_i) be the observed data at site $\mathbf{s}_i$, where C_i is the censoring indicator, and V_i is given by

$$V_i = \begin{cases} r_i, & \text{if } Y_i \in R_i, \quad \text{(censored)} \\ Y_i, & \text{otherwise}, \quad \text{(observed)} \end{cases} \tag{3.3}$$

where R_i is an interval of the form $(-\infty, r_i)$, (r_i, ∞), or (r_{i1}, r_{i2}) for left, right, or interval censoring, respectively. The constant $r_i \in \mathbb{R}$ is equal to the detection limit for left and right censoring, and $r_i = (r_{i1} + r_{i2})/2$ for interval censoring. Moreover, missing observations can be handled by setting $R_i = (-\infty, \infty)$ and $r_i = \text{NA}$. Thus, the *spatial censored linear* (SCL) model is defined by (3.2) and (3.3).

Remark: In cases of missing data, the value $r_i = \text{NA}$ is used as a reference and is excluded from the initial parameter estimation calculations. Instead, the associated observation Y_i is imputed in each iteration of the algorithm, as will be described in Subsection 3.3.3. This imputation method is valid under the assumption that the data are missing at random (MAR).

3.3.2 The likelihood function

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \tau^2, \phi)^\top$ be the vector with all parameters in the SCL model. To obtain the likelihood function, given the observed sample $\mathbf{y} = (y_1, y_2, \ldots, y_n)^\top$, we treat the observed and censored components of $\mathbf{y}$ separately. Let $\mathbf{y}_o \in \mathbb{R}^{n_o}$ and $\mathbf{y}_m \in \mathbb{R}^{n_m}$ be the vector of observed and censored variables, respectively, with $n = n_o + n_m$. After reordering the elements of $\mathbf{y}$, $\mathbf{V}$, $\mathbf{X}$, and $\boldsymbol{\Sigma}$, they can be partitioned as

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_o \\ \mathbf{y}_m \end{pmatrix}, \quad \mathbf{V} = \begin{pmatrix} \mathbf{V}_o \\ \mathbf{V}_m \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_o \\ \mathbf{X}_m \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{oo} & \boldsymbol{\Sigma}_{om} \\ \boldsymbol{\Sigma}_{mo} & \boldsymbol{\Sigma}_{mm} \end{pmatrix},$$

where $\mathbf{X}_j \in \mathbb{R}^{n_j \times q}$, and $\boldsymbol{\Sigma}_{jk} \in \mathbb{R}^{n_j \times n_k}$, for $j, k \in \{o, m\}$. Then, we have that $\mathbf{Y}_o \sim N_{n_o}(\mathbf{X}_o\boldsymbol{\beta}, \boldsymbol{\Sigma}_{oo})$ and $\mathbf{Y}_m|\mathbf{Y}_o \sim N_{n_m}(\boldsymbol{\mu}, \mathbf{S})$, with $\boldsymbol{\mu} = \mathbf{X}_m\boldsymbol{\beta} + \boldsymbol{\Sigma}_{mo}\boldsymbol{\Sigma}_{oo}^{-1}(\mathbf{Y}_o - \mathbf{X}_o\boldsymbol{\beta})$ and

$\mathbf{S} = \boldsymbol{\Sigma}_{mm} - \boldsymbol{\Sigma}_{mo}\boldsymbol{\Sigma}_{oo}^{-1}\boldsymbol{\Sigma}_{om}$. Let $\mathbf{V} = (V_1, \ldots, V_n)^\top$ and $\mathbf{C} = (C_1, \ldots, C_n)^\top$ be the observed data, therefore, the likelihood function (using conditional probability arguments, see Vaida & Liu, 2009) is given by

$$\begin{aligned} L(\boldsymbol{\theta}) = L(\boldsymbol{\theta} \mid \mathbf{V}, \mathbf{C}) &= f(\mathbf{y}_o \mid \boldsymbol{\theta}) \Pr(\mathbf{r}_1^m < \mathbf{Y}_m < \mathbf{r}_2^m \mid \mathbf{y}_o, \boldsymbol{\theta}) \\ &= \phi_{n_o}(\mathbf{y}_o; \mathbf{X}_o\boldsymbol{\beta}, \boldsymbol{\Sigma}_{oo}) \Pr(\mathbf{Y}_m \in \mathbf{B} | \mathbf{y}_o), \end{aligned}$$

where $\mathbf{B} = \{\mathbf{y}_m = (y_1^m, \ldots, y_{n_m}^m)^\top | r_{11}^m < y_1^m < r_{12}^m, \ldots, r_{n_m1}^m < y_{n_m}^m < r_{n_m2}^m\}$, $\phi(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the probability density function of $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ computed at vector $\mathbf{x}$, and $\Pr(\mathbf{Y}_m \in \mathbf{B} \mid \mathbf{y}_o)$ is the conditional probability of $\mathbf{Y}_m$ being in the set $\mathbf{B}$ given the observed responses.

3.3.3 Parameter estimation

Due to challenges in handling the observed log-likelihood function directly, Lachos *et al.* (2017) suggested utilizing an EM-type algorithm to derive the ML estimates for the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \tau^2, \phi)^\top$. This estimation approach involves a specific parameterization for $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Psi}$, where $\boldsymbol{\Psi} = \mathbf{R} + \nu^2\mathbf{I}_n$ and $\nu^2 = \tau^2/\sigma^2$. Such parameterization aids in improving the identifiability of the parameters. Additional insights into this methodology can be found in Diggle & Ribeiro (2007).

Excluding constants that do not depend on $\boldsymbol{\theta}$, the complete log-likelihood function for the complete data $\mathbf{y}_c = (\mathbf{V}, \mathbf{C}, \mathbf{y})$ is expressed as follows

$$\ell_c(\boldsymbol{\theta}; \mathbf{y}_c) = -\frac{1}{2}\left\{ n \ln \sigma^2 + \ln|\boldsymbol{\Psi}| + \frac{1}{\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Psi}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\}.$$

Hence, the EM algorithm works as follows:

- **E-step**: Let $\widehat{\boldsymbol{\theta}}^{(k)}$ be the current estimate of $\boldsymbol{\theta}$, then the conditional expectation of the complete-data log-likelihood without the constant is

$$Q_k(\boldsymbol{\theta}) = \mathbb{E}\left[\ell_c(\boldsymbol{\theta}; \mathbf{y}_c) | \mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}\right] = -\frac{1}{2}\left\{ \ln|\boldsymbol{\Psi}| + n \ln \sigma^2 + \frac{1}{\sigma^2}\widehat{A}^{(k)} \right\},$$

  where $\widehat{A}^{(k)} = \operatorname{tr}(\widehat{\mathbf{y}\mathbf{y}^\top}^{(k)}\boldsymbol{\Psi}^{-1}) - 2\widehat{\mathbf{y}}^{(k)\top}\boldsymbol{\Psi}^{-1}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^\top\mathbf{X}^\top\boldsymbol{\Psi}^{-1}\mathbf{X}\boldsymbol{\beta}$. Therefore, the E-step reduces only to the computation of

$$\widehat{\mathbf{y}\mathbf{y}^\top}^{(k)} = \mathbb{E}\left[\mathbf{y}\mathbf{y}^\top | \mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}\right] \quad \text{and} \quad \widehat{\mathbf{y}}^{(k)} = \mathbb{E}\left[\mathbf{y} | \mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}\right]. \tag{3.4}$$

  Note that when $\mathbf{y}$ is not fully observed, the elements of $\widehat{\mathbf{y}}^{(k)}$ and $\widehat{\mathbf{y}\mathbf{y}^\top}^{(k)}$ related to the censored ($C_i = 1$) variables are the first two moments of the TMVN distribution, $\mathrm{TN}_{n_m}(\boldsymbol{\mu}, \mathbf{S}; \mathbf{B})$, respectively, with $\boldsymbol{\mu}, \mathbf{S}$, and $\mathbf{B}$ as defined in the likelihood function. For $C_i = 0$, these components are obtained directly from the observed values.

- **M-step**: The conditional maximization step is carried out, and $\widehat{\boldsymbol{\theta}}^{(k)}$ is updated by maximizing $\widehat{Q}_k(\boldsymbol{\theta})$ over $\boldsymbol{\theta}$ to obtain a new estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$, which leads to the expressions:

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}^{(k+1)} &= \left(\mathbf{X}^\top \widehat{\boldsymbol{\Psi}}^{-1(k)} \mathbf{X}\right)^{-1} \mathbf{X}^\top \widehat{\boldsymbol{\Psi}}^{-1(k)} \widehat{\mathbf{y}}^{(k)}, \\
\widehat{\sigma}^{2(k+1)} &= \frac{1}{n}\left\{ \operatorname{tr}\left(\widehat{\mathbf{y}\mathbf{y}^\top}^{(k)} \widehat{\boldsymbol{\Psi}}^{-1(k)}\right) - 2\widehat{\mathbf{y}^\top}^{(k)} \widehat{\boldsymbol{\Psi}}^{-1(k)} \mathbf{X}\widehat{\boldsymbol{\beta}}^{(k+1)} + \widehat{\boldsymbol{\beta}}^{\top(k+1)} \mathbf{X}^\top \widehat{\boldsymbol{\Psi}}^{-1(k)} \mathbf{X}\widehat{\boldsymbol{\beta}}^{(k+1)} \right\}, \\
\widehat{\boldsymbol{\alpha}}^{(k+1)} &= \underset{\boldsymbol{\alpha}\in\mathbb{R}^+\times\mathbb{R}^+}{\operatorname{argmax}} \left(-\frac{1}{2}\ln|\boldsymbol{\Psi}| - \frac{1}{2\widehat{\sigma}^{2(k+1)}} \left\{ \operatorname{tr}\left(\widehat{\mathbf{y}\mathbf{y}^\top}^{(k)} \boldsymbol{\Psi}^{-1}\right) - 2\widehat{\mathbf{y}^\top}^{(k)} \boldsymbol{\Psi}^{-1} \mathbf{X}\widehat{\boldsymbol{\beta}}^{(k+1)} \right.\right. \\
&\quad \left.\left. + \widehat{\boldsymbol{\beta}}^{\top(k+1)} \mathbf{X}^\top \boldsymbol{\Psi}^{-1} \mathbf{X}\widehat{\boldsymbol{\beta}}^{(k+1)} \right\} \right),
\end{aligned}
$$

  with $\boldsymbol{\alpha} = (\phi, \nu^2)^\top$. Note that $\widehat{\tau}^2$ can be recovered by $\widehat{\tau}^{2(k+1)} = \widehat{\nu}^{2(k+1)}\widehat{\sigma}^{2(k+1)}$. An efficient M-step could be easily accomplished by using, for instance, the `roptim` package (Pan & Pan, 2020).

The algorithm is iterated until some distance between two successive parameter estimations, such as $\sqrt{(\widehat{\boldsymbol{\theta}}^{(k+1)} - \widehat{\boldsymbol{\theta}}^{(k)})^\top (\widehat{\boldsymbol{\theta}}^{(k+1)} - \widehat{\boldsymbol{\theta}}^{(k)})}$, becomes small enough. For comparison purposes, we consider the EM, MCEM, and SAEM algorithms to estimate the parameters of the SCL model; hence, the E-step for each algorithm performs as follows:

- In the *EM algorithm*, the expectations defined in (3.4) are calculated using the R package `MomTrunc`. The ML estimates are determined based on the results from the final iteration of the algorithm. It is important to note that this procedure can be computationally intensive, particularly when there is a significant proportion of censored observations.

- In the *MCEM algorithm*, the expectations are approximated via Monte Carlo integration as outlined in Section 2.4 (refer to Chapter 2). This approach distinctly handles the censored and missing observations by exclusively sampling the censored variables using the proposed slice sampling algorithm, while missing values are imputed based on the properties of the conditional distribution. Given the inherent variability of Monte Carlo (MC) methods, the estimates for $\boldsymbol{\theta}$ might exhibit slight fluctuations around the true solution. Consequently, to stabilize these estimates, the final estimates produced by the MCEM are determined by averaging the estimates from each iteration. This procedure is done after discarding the first half of the iterations, a process known as *burn-in*, and applying a *thinning* factor of 3.

- In the *SAEM algorithm*, the E-step is approximated as follows

$$
\widehat{\mathbf{y}}^{(k)} = \widehat{\mathbf{y}}^{(k-1)} + \delta_k \left(\frac{1}{L}\sum_{l=1}^{L} \mathbf{y}^{(k,l)} - \widehat{\mathbf{y}}^{(k-1)} \right), \tag{3.5}
$$

$$
\widehat{\mathbf{y}\mathbf{y}^\top}^{(k)} = \widehat{\mathbf{y}\mathbf{y}^\top}^{(k-1)} + \delta_k \left(\frac{1}{L}\sum_{l=1}^{L} \mathbf{y}^{(k,l)}\mathbf{y}^{(k,l)\top} - \widehat{\mathbf{y}\mathbf{y}^\top}^{(k-1)} \right), \tag{3.6}
$$

where δ_k is the smoothness parameter selected as described in Subsection 1.1.3, L is the number of samples used in the approximation, and $\mathbf{y}^{(k,l)} = (\mathbf{y}_o^\top, \mathbf{y}_m^{(k,l)\top})^\top$ is the vector of complete data consisting of the observed data $\mathbf{y}_o$ and $\mathbf{y}_m^{(k,l)}$ simulated from the TMVN distribution $\mathbf{Y}_m|\mathbf{Y}_o, \mathbf{V}, \mathbf{C} \sim \text{TN}(\boldsymbol{\mu}, \mathbf{S}; \mathbf{B})$, with the same $\boldsymbol{\mu}$, $\mathbf{S}$, and $\mathbf{B}$ settings considered in the likelihood function. For this algorithm, the samples are obtained through the method proposed in Section 2.3, which is available in the R package `relliptical`.

3.3.4 Standard error approximation

Following the methodology described in Section 1.2, we compute the asymptotic covariance of the ML estimates through the observed information matrix using the Louis method. So, we have that

$$\mathbf{I}_o(\boldsymbol{\theta}) = \mathbb{E}\left[\mathbf{B}_c(\mathbf{y}_c;\boldsymbol{\theta}) \mid \mathbf{y}_o\right] - \mathbb{E}\left[\mathbf{S}_c(\mathbf{y}_c;\boldsymbol{\theta})\mathbf{S}_c^\top(\mathbf{y}_c;\boldsymbol{\theta}) \mid \mathbf{y}_o\right] + \mathbf{S}_o(\mathbf{y}_o;\boldsymbol{\theta})\mathbf{S}_o^\top(\mathbf{y}_o;\boldsymbol{\theta}), \tag{3.7}$$

where

$$\mathbf{B}_c(\mathbf{y}_c;\boldsymbol{\theta}) = -\frac{\partial^2 \ell_c(\boldsymbol{\theta};\mathbf{y}_c)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}, \ \mathbf{S}_c(\mathbf{y}_c;\boldsymbol{\theta}) = \frac{\partial \ell_c(\boldsymbol{\theta};\mathbf{y}_c)}{\partial\boldsymbol{\theta}}, \text{ and } \mathbf{S}_o(\mathbf{y}_o;\boldsymbol{\theta}) = \mathbb{E}\left[\mathbf{S}_c(\mathbf{y}_c;\boldsymbol{\theta})\middle|\mathbf{y}_o\right].$$

The explicit expressions for the elements of $\mathbb{E}\left[\mathbf{B}_c(\mathbf{y}_c;\boldsymbol{\theta}) \mid \mathbf{y}_o\right]$ and $\mathbf{S}_o(\mathbf{y}_o;\boldsymbol{\theta})$ can be found in Appendix B.1, while for the second term $G(\boldsymbol{\theta}) = \mathbb{E}\left[\mathbf{S}_c(\mathbf{y}_c;\boldsymbol{\theta})\mathbf{S}_c^\top(\mathbf{y}_c;\boldsymbol{\theta})|\mathbf{y}_o\right]$, we use an approximation for each proposed method because of the complexity of the expressions, as follows:

- For the EM and MCEM algorithms

$$G(\boldsymbol{\theta}) \approx \frac{1}{L}\sum_{l=1}^{L} \frac{\partial \ell_c(\boldsymbol{\theta};\mathbf{y}_c^{(l)})}{\partial\boldsymbol{\theta}} \left(\frac{\partial \ell_c(\boldsymbol{\theta};\mathbf{y}_c^{(l)})}{\partial\boldsymbol{\theta}}\right)^\top, \tag{3.8}$$

  where $\mathbf{y}_c^{(l)} = (\mathbf{y}_o, \mathbf{y}_m^{(l)})^\top, l \in \{1,\ldots,L\}$ is the vector of complete data consisting of the observed values and the data simulated from the TMVN distribution as described previously.

- For the SAEM algorithm

$$G(\boldsymbol{\theta})^{(k)} \approx G(\boldsymbol{\theta})^{(k-1)} + \delta_k \left(\frac{1}{L}\sum_{l=1}^{L} \frac{\partial \ell_c(\boldsymbol{\theta};\mathbf{y}_c^{(k,l)})}{\partial\boldsymbol{\theta}} \left(\frac{\partial \ell_c(\boldsymbol{\theta};\mathbf{y}_c^{(k,l)})}{\partial\boldsymbol{\theta}}\right)^\top - G(\boldsymbol{\theta})^{(k-1)}\right), \tag{3.9}$$

  with the same δ_k and $\mathbf{y}^{(k,l)} = (\mathbf{y}_o^\top, \mathbf{y}_m^{(k,l)\top})^\top$ settings as in the estimation procedure, for $k \in \{1,\ldots,W\}$ and $l \in \{1,\ldots,L\}$, where W is the maximum number of SAEM iterations and L denotes the number of samples used in the approximation.

3.4 Simulation study

In order to evaluate the performance of the proposed methods, we present two simulation studies focusing on a) the influence of the number of samples used in the MC approximation within the MCEM algorithm and b) the asymptotic behavior of the estimates produced by the MCEM and SAEM algorithms, considering varying sample sizes and levels of censoring. For simplicity, we generated 300 MC datasets from the Gaussian spatial model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}$, where the design matrix $\mathbf{X} = (\mathbf{1}, \mathbf{X}_1, \mathbf{X}_2)$ includes a first column of ones, and the elements of the second and third columns independently simulated from a normal distribution $N(0, 1)$. The vector of regression parameters was set as $\boldsymbol{\beta} = (1, 3, -2)^\top$. The error vector $\boldsymbol{\xi}$ was simulated from the multivariate normal distribution with zero mean and variance-covariance matrix $\boldsymbol{\Sigma} = \sigma^2\mathbf{R} + \tau^2\mathbf{I}_n$, which was calculated using

$$\text{Cov}(\xi(\mathbf{s}_i), \xi(\mathbf{s}_j)) = \sigma^2 \exp\left\{-\left(\frac{||\mathbf{h}||}{\phi}\right)^{\kappa}\right\} + \tau^2\mathbb{I}(\mathbf{s}_i = \mathbf{s}_j),$$

with the partial sill and nugget effect set at $\sigma^2 = 3.50$ and $\tau^2 = 1$, respectively. This configuration suggests that the spatial process accounts for 77.78% of the variability observed in the data. The elements of the matrix $\mathbf{R}$ were computed using the power exponential correlation function, characterized by parameters $\kappa = 1.50$ and $\phi = 4$; these parameters indicate that the correlation drops to less than 0.05 for distances greater than 8.31 (spatial) units. Lastly, the spatial coordinates were randomly selected from within a 20×20 square.

3.4.1 Simulation study I: MCEM estimates

In the first study, we aimed to assess the accuracy of the MCEM estimates with varying numbers of samples (L) used in the MC approximation. It was considered 300 MC samples of size $n = 200$. We examined several scenarios: three with 5% missing and 10%, 20%, and 30% left-censored responses, respectively, and another scenario with 30% censored observations and 10% missing. The censoring procedure followed the method outlined by Schelin & Sjöstedt-de Luna (2014), where α% of responses were left-censored, and then β% of the non-censored observations were randomly selected and treated as missing values. Additionally, the case without censoring (original data) was included for comparative purposes. The estimation procedure was conducted using the function `MCEM.sclm` from `RcppCensSpatial` package. For each sample, the E-step was approximated through MC integration under three configurations: i) $L = 20$ random samples, ii) L increasing linearly from 20 to 5000, and iii) $L = 5000$ random samples.

The findings are summarized in Table 2, where MC-AV and MC-SD represent the mean and the standard deviation of the 300 MC estimates, respectively. The results

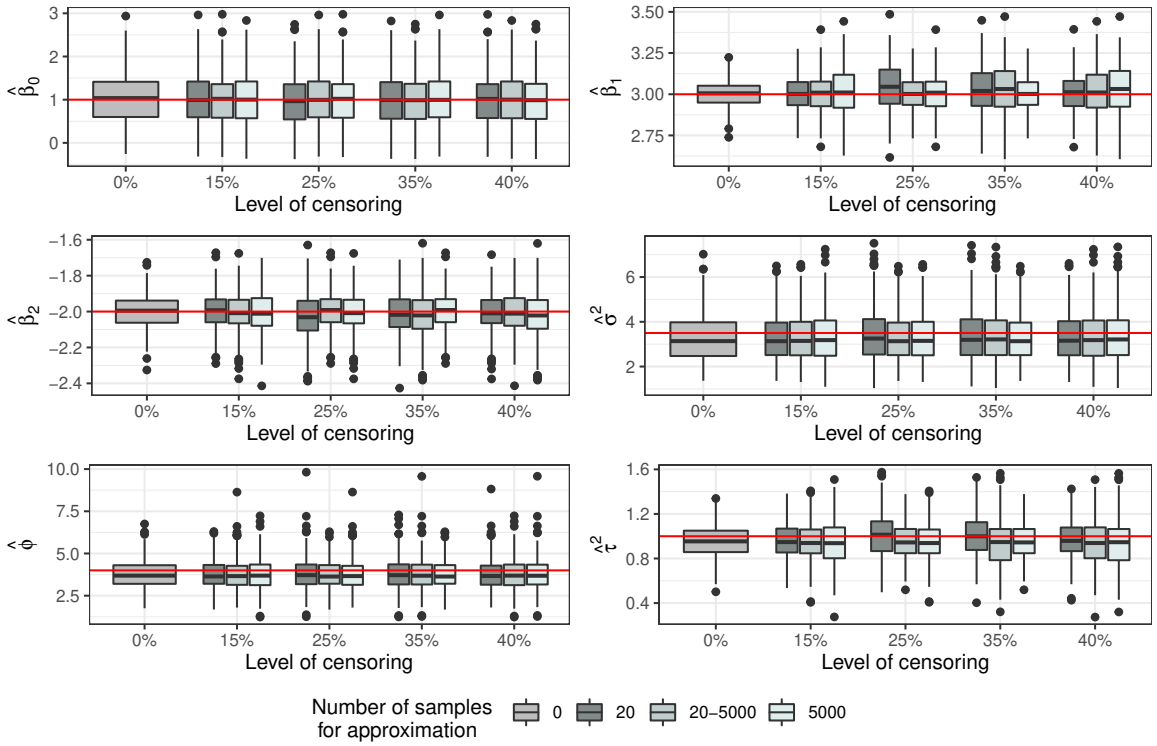
indicate that the mean estimates for β_0, β_1, and β_2 are close to their true values across all scenarios. It is worth mentioning that the standard deviation for β_1 and β_2 increases slightly with the level of censoring, while the variation for β_0 remains negligible. It was observed that the mean of the estimates obtained with $L = 5000$ random points is comparable to that obtained when L increased linearly from 20 to 5000; the standard deviations in these two scenarios also align closely. The mean estimates were lower than the true parameter values for the variance structure parameters σ^2, ϕ, and τ^2. Nonetheless, the ratio $\sigma^2/(\sigma^2 + \tau^2)$ consistently approximated 0.775, suggesting that the spatial process explains approximately 77.5% of the data variability. This finding aligns closely with the true value. For additional details on the consistency of this parameter estimation, please refer to Zhang (2004).

Table 2 – Simulation I. Summary statistics of the MCEM estimates based on 300 samples of size $n = 200$, considering different censoring proportions and number of observations (L) for the MC method.

Censoring	L	Measure	Parameters					
			β_0	β_1	β_2	σ^2	ϕ	τ^2
			1.00	3.00	-2.00	3.50	4.00	1.00
0%	-	MC-AV	1.032	3.001	-1.999	3.295	3.771	0.956
		MC-SD	0.570	0.079	0.092	1.054	0.812	0.146
15%	20	MC-AV	1.027	3.004	-1.996	3.302	3.766	0.958
		MC-SD	0.573	0.101	0.102	1.055	0.842	0.161
	20 - 5000	MC-AV	1.027	3.004	-1.996	3.299	3.762	0.952
		MC-SD	0.573	0.101	0.102	1.054	0.841	0.161
	5000	MC-AV	1.027	3.004	-1.996	3.299	3.762	0.952
		MC-SD	0.573	0.101	0.102	1.054	0.841	0.161
25%	20	MC-AV	1.016	3.011	-2.005	3.330	3.771	0.968
		MC-SD	0.570	0.114	0.107	1.085	0.926	0.175
	20 - 5000	MC-AV	1.018	3.009	-2.003	3.317	3.762	0.949
		MC-SD	0.570	0.114	0.107	1.077	0.922	0.178
	5000	MC-AV	1.018	3.009	-2.003	3.317	3.761	0.949
		MC-SD	0.570	0.114	0.107	1.077	0.922	0.178
35%	20	MC-AV	0.997	3.028	-2.012	3.382	3.801	1.002
		MC-SD	0.582	0.137	0.115	1.155	0.940	0.194
	20 - 5000	MC-AV	1.010	3.018	-2.005	3.350	3.781	0.941
		MC-SD	0.582	0.136	0.115	1.136	0.933	0.205
	5000	MC-AV	1.010	3.018	-2.005	3.350	3.781	0.941
		MC-SD	0.582	0.136	0.115	1.136	0.933	0.205
40%	20	MC-AV	0.963	3.045	-2.026	3.418	3.801	1.007
		MC-SD	0.585	0.144	0.126	1.175	0.995	0.206
	20 - 5000	MC-AV	0.979	3.033	-2.018	3.382	3.778	0.936
		MC-SD	0.585	0.143	0.126	1.149	0.985	0.220
	5000	MC-AV	0.979	3.033	-2.018	3.382	3.778	0.936
		MC-SD	0.585	0.143	0.126	1.150	0.985	0.220

Figure 11 displays boxplots of the estimates derived via the MCEM algorithm, segmented by level of censoring and number of random points used for approximation. In these plots, the solid red line indicates the true parameter value. The plots show that the median of the estimates for β_0, β_1, and β_2 coincides with the actual value, while the medians for σ^2, ϕ, and τ^2 tend to underestimate the true value of the parameters. The boxplots for ϕ and σ^2 exhibit some outliers across all levels of censoring compared to other parameters. Further details can be found in Figure 38 (Appendix B.2), which illustrates the mean bias ± 1 standard deviation for each parameter by censoring proportion and number of samples (L) used for approximation. This analysis reveals that, across different levels of censoring, the bias remains close to zero, and the standard deviation is almost the same independent of L.

Figure 11 – Simulation I. Boxplot of the estimates by level of censoring.



3.4.2 Simulation study II: Asymptotic properties of the estimates

Aiming to provide empirical evidence for the consistency of the ML estimates obtained via the MCEM and SAEM algorithms, we considered sample sizes of $n = 100$, 300, and 600. Additionally, 10% and 30% of the observations were left-censored or missing, following the scheme described in the previous section, with 20% of the desired censored rate randomly selected to be treated as missing. We also evaluated the case without censoring (original data) for comparison purposes. Our proposed MCEM and SAEM algorithms were implemented through the functions `MCEM.sclm` and `SAEM.sclm`, respectively. The number of random observations in the MC approximation increased linearly from 20 to

5000. As shown in Simulation study I, this setting seems to yield results comparable to using $L = 5000$ samples. For the SAEM algorithm, we set $M = 20$ and $c = 0.25$ (refer to Subsection 1.1.3 for more details).

For each combination of sample size, level of censoring, and method, we computed the mean (MC-AV) and standard deviation (MC-SD) of the 300 MC estimates, the mean of the standard error (IM-SE) computed by the inverse of the observed information matrix (detailed in Subsection 3.3.4), and the coverage probability (CP) of a 95% confidence interval.

The results for the MCEM and SAEM estimates, displayed in Tables 3 and 4, respectively, show that the mean of the estimates (MC-AV) is close to the true parameter values across all sample sizes and levels of censoring. As expected, discrepancies decrease as the sample size increases, except for β_0. Notably, the mean of the standard errors (IM-SE) is, in general, close to the standard deviation of the estimates (MC-SD) for all scenarios, indicating the reliability of the proposed method for obtaining standard errors. It is worth noting that the standard error for β_0 appears constant regardless of the sample size or censoring level, while for other parameters, the standard error decreases with increasing sample size and increases with higher levels of censoring. Overall, both methods performed well in fitting the simulated spatial censored data.

Figures 39 and 40 (found in Appendix B.2) present the boxplot of the MCEM and SAEM estimates, respectively, by sample size and level of censoring. The real parameter values are indicated by red lines, showing that the median of the estimates is often close to the true values, with few outliers observed.

Lastly, we analyzed the mean squared error (MSE) of the estimates for all scenarios, defined as:

$$\text{MSE}_i = \frac{1}{n}\sum_{j=1}^{n}(\widehat{\theta}_i^{(j)} - \theta_i)^2,$$

where $\widehat{\theta}_i^{(j)}$ is the estimate of the ith parameter $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \sigma^2, \phi, \tau^2)^\top$ in the jth MC sample. The MSE results for each parameter, sample size, and censoring level are presented in Figure 12 for both MCEM (left panel) and SAEM (right panel) estimates. The MSE tends to zero as the sample size increases, except for β_0 with 30% censoring, where a slight increase is observed for $n = 600$. Thus, while the proposed MCEM and SAEM algorithms provide ML estimates with good asymptotic properties for the spatial censored linear model, careful attention must be given to the estimates of β_0 under high censoring rates.

3.5 Missouri dioxin contamination data

The methodologies discussed will be applied to analyze the Missouri dioxin contamination dataset described in Subsection 1.3.1. Following the approach suggested by

Table 3 – Simulation II. Summary statistics of the MCEM estimates based on 300 samples of sizes $n = 100, 300$, and 600 with different censoring proportions.

n	Censoring	Measure	Parameters					
			β_0	β_1	β_2	σ^2	ϕ	τ^2
			1.00	3.00	-2.00	3.50	4.00	1.00
100	0%	MC-AV	1.049	2.984	-2.009	3.318	3.796	0.901
		IM-SE	0.560	0.133	0.137	1.169	1.109	0.260
		MC-SD	0.574	0.137	0.152	1.203	1.240	0.267
		CP (%)	91.3	94.3	92.0	-	-	-
	10%	MC-AV	1.047	2.981	-2.009	3.317	3.790	0.898
		IM-SE	0.563	0.143	0.144	1.184	1.143	0.277
		MC-SD	0.582	0.152	0.156	1.213	1.247	0.283
		CP (%)	91.0	93.0	92.7	-	-	-
	30%	MC-AV	1.027	2.987	-2.017	3.344	3.782	0.889
		IM-SE	0.572	0.170	0.164	1.251	1.244	0.333
		MC-SD	0.588	0.183	0.173	1.304	1.361	0.321
		CP (%)	89.3	91.7	94.7	-	-	-
300	0%	MC-AV	0.959	2.997	-2.004	3.303	3.818	0.981
		IM-SE	0.532	0.071	0.069	1.027	0.805	0.120
		MC-SD	0.567	0.074	0.069	1.070	0.777	0.121
		CP (%)	90.3	94.0	95.0	–	–	–
	10%	MC-AV	0.960	2.999	-2.008	3.300	3.823	0.984
		IM-SE	0.535	0.078	0.075	1.035	0.820	0.128
		MC-SD	0.561	0.079	0.075	1.079	0.797	0.129
		CP (%)	90.0	94.3	95.7	-	-	-
	30%	MC-AV	0.946	3.012	-2.016	3.310	3.812	0.992
		IM-SE	0.539	0.096	0.087	1.056	0.858	0.150
		MC-SD	0.564	0.097	0.088	1.088	0.855	0.147
		CP (%)	91.3	96.0	94.3	-	-	-
600	0%	MC-AV	0.925	3.000	-2.005	3.276	3.798	0.995
		IM-SE	0.518	0.045	0.048	0.971	0.709	0.076
		MC-SD	0.554	0.044	0.049	0.950	0.695	0.079
		CP (%)	90.0	95.7	94.0	-	-	-
	10%	MC-AV	0.922	3.003	-2.005	3.278	3.794	0.994
		IM-SE	0.519	0.050	0.052	0.976	0.714	0.081
		MC-SD	0.554	0.049	0.052	0.955	0.701	0.085
		CP (%)	90.0	95.7	95.0	-	-	-
	30%	MC-AV	0.906	3.011	-2.010	3.293	3.771	0.996
		IM-SE	0.521	0.060	0.061	0.991	0.733	0.094
		MC-SD	0.561	0.057	0.061	0.965	0.725	0.096
		CP (%)	89.7	96.3	95.3	-	-	-

Zirschky & Harris (1986), we considered the logarithm transformation of the response and fitted the model:

$$\ln(Y_i) = \beta_0 + \xi_i, \quad i \in \{1, \ldots, 127\}.$$

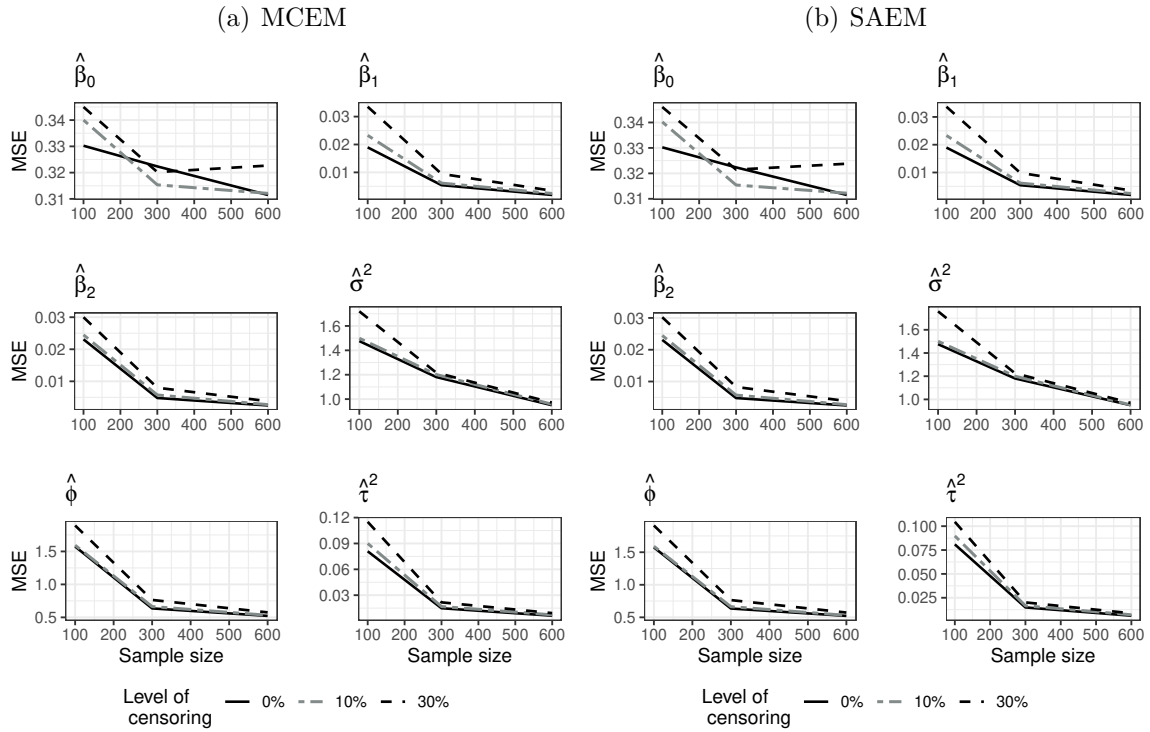
The model parameters were estimated using the MCEM, SAEM, and EM algorithms, each performed over 500 iterations with an exponential correlation function

Table 4 – Simulation II. Summary statistics of the SAEM estimates based on 300 samples of sizes $n = 100, 300$, and 600 with different censoring proportions.

n	Censoring	Measure	Parameters					
			β_0	β_1	β_2	σ^2	ϕ	τ^2
			1.00	3.00	-2.00	3.50	4.00	1.00
100	0%	MC-AV	1.049	2.984	-2.009	3.318	3.796	0.901
		IM-SE	0.560	0.133	0.137	1.169	1.109	0.260
		MC-SD	0.574	0.137	0.152	1.203	1.240	0.267
		CP (%)	91.3	94.3	92.0	-	-	-
	10%	MC-AV	1.047	2.981	-2.009	3.318	3.789	0.899
		IM-SE	0.563	0.143	0.144	1.184	1.143	0.277
		MC-SD	0.582	0.152	0.156	1.214	1.246	0.282
		CP (%)	91.3	93.0	92.7	-	-	-
	30%	MC-AV	1.018	2.994	-2.022	3.368	3.777	0.917
		IM-SE	0.576	0.177	0.169	1.277	1.269	0.362
		MC-SD	0.589	0.184	0.173	1.322	1.366	0.313
		CP (%)	90.0	93.0	94.7	-	-	-
300	0%	MC-AV	0.959	2.997	-2.004	3.302	3.818	0.981
		IM-SE	0.532	0.071	0.069	1.027	0.805	0.120
		MC-SD	0.567	0.074	0.069	1.070	0.777	0.121
		CP (%)	90.3	94.0	95.0	-	-	-
	10%	MC-AV	0.959	3.000	-2.008	3.301	3.823	0.985
		IM-SE	0.535	0.078	0.075	1.035	0.821	0.128
		MC-SD	0.561	0.079	0.075	1.079	0.797	0.128
		CP (%)	90.3	94.3	95.7	-	-	-
	30%	MC-AV	0.936	3.021	-2.022	3.328	3.812	1.021
		IM-SE	0.541	0.100	0.089	1.065	0.864	0.159
		MC-SD	0.564	0.097	0.089	1.095	0.856	0.141
		CP (%)	91.0	96.0	95.0	-	-	-
600	0%	MC-AV	0.925	3.000	-2.005	3.276	3.797	0.995
		IM-SE	0.518	0.045	0.048	0.971	0.709	0.076
		MC-SD	0.554	0.044	0.049	0.950	0.695	0.079
		CP (%)	90.0	95.7	94.0	-	-	-
	10%	MC-AV	0.922	3.002	-2.005	3.278	3.795	0.994
		IM-SE	0.519	0.050	0.052	0.976	0.715	0.081
		MC-SD	0.554	0.049	0.052	0.952	0.700	0.085
		CP (%)	90.0	96.0	95.0	-	-	-
	30%	MC-AV	0.903	3.014	-2.012	3.300	3.771	1.007
		IM-SE	0.522	0.061	0.062	0.994	0.734	0.096
		MC-SD	0.562	0.058	0.061	0.966	0.723	0.093
		CP (%)	90.0	96.7	95.3	-	-	-

to assess the variation between spatial points. For the MCEM algorithm, we explored four scenarios, including one with linearly increasing sample sizes from 100 to 1000. The other scenarios involved constant sample sizes of 20, 5000, and 10^5. Two configurations were used for the SAEM algorithm: one using the `rmvtnorm` function for simulating and `optimx` for optimization, referred to henceforth as SAEM. This method is available in the

Figure 12 – Simulation II. MSE of the MCEM (left) and SAEM (right) estimates considering different sample sizes and levels of censoring.



`CensSpatial` package. The second configuration employed our proposed slice sampler for simulation, with optimization executed by the R function `roptim`, referred to as SAEM-SS. Moments were computed using the `MomTrunc` package for the EM algorithm. Functions for parameter estimation via MCEM, SAEM-SS, and EM are integrated into the `RcppCensSpatial` package.

Table 5 – Missouri data. ML estimates and information criteria (AIC and BIC) were obtained through MCEM, SAEM, and EM algorithms considering the exponential correlation function.

Algorithm	L	c	β_0	σ^2	ϕ	τ^2	$\ell(\widehat{\boldsymbol{\theta}})$	AIC	BIC	Time (min)
MCEM	20		-2.355	6.577	14.702	0.213	-143.128	294.257	305.633	0.936
	10^2 - 10^3	-	-2.402	6.808	15.095	0.207	-143.108	294.216	305.592	1.819
	5000	-	-2.410	6.847	15.076	0.206	-143.095	294.191	305.568	10.206
	10^5	-	-2.408	6.845	15.053	0.205	-143.136	294.272	305.649	185.341
SAEM-SS	20	0.25	-2.311	6.218	14.964	0.220	-143.175	294.350	305.727	0.658
SAEM	20	0.25	-2.014	4.858	14.206	0.245	-143.840	295.681	307.057	6.079
	10^5	1.00	-2.010	4.829	14.136	0.245	-143.865	295.729	307.106	9151.149
EM	-	-	-2.417	6.888	15.092	0.206	-143.122	294.244	305.620	1661.472

Table 5 displays the results of the ML estimates where L represents the number of samples to approximate the conditional mean, and c denotes the percentage of memory-free iterations in the SAEM algorithm (see Subsection 1.1.3). Final estimates for MCEM and EM methods were calculated as the mean of the estimates from each iteration after

applying a burn-in of 250 and thinning of 3 observations. In contrast, estimates from the SAEM and SAEM-SS were taken from the final iteration. The SAEM-SS estimates are similar to those from MCEM at $L = 20$, while the EM algorithm's estimates are similar to MCEM at $L = 5000$. The estimated regression coefficient β_0 from the MCEM and EM was -2.400, whereas the SAEM algorithm estimated this parameter at -2.010. The estimates from MCEM and SAEM for the partial sill σ^2 and the nugget effect τ^2 suggest that the spatial process accounts for 97% and 95% of the data variability, respectively. The spatial scaling parameter ϕ was estimated at approximately 15.05 and 14.10 by MCEM and SAEM, respectively, indicating that correlations between observations fall below 0.05 for distances exceeding 45 and 42 feet. This table presents the maximized log-likelihood values, information criteria (AIC and BIC), and running times. Notably, the information criteria are nearly identical across all methods, which aligns with expectations.

Figure 13 – Missouri data. Convergence of the parameter estimates via EM, MCEM, and SAEM algorithm.

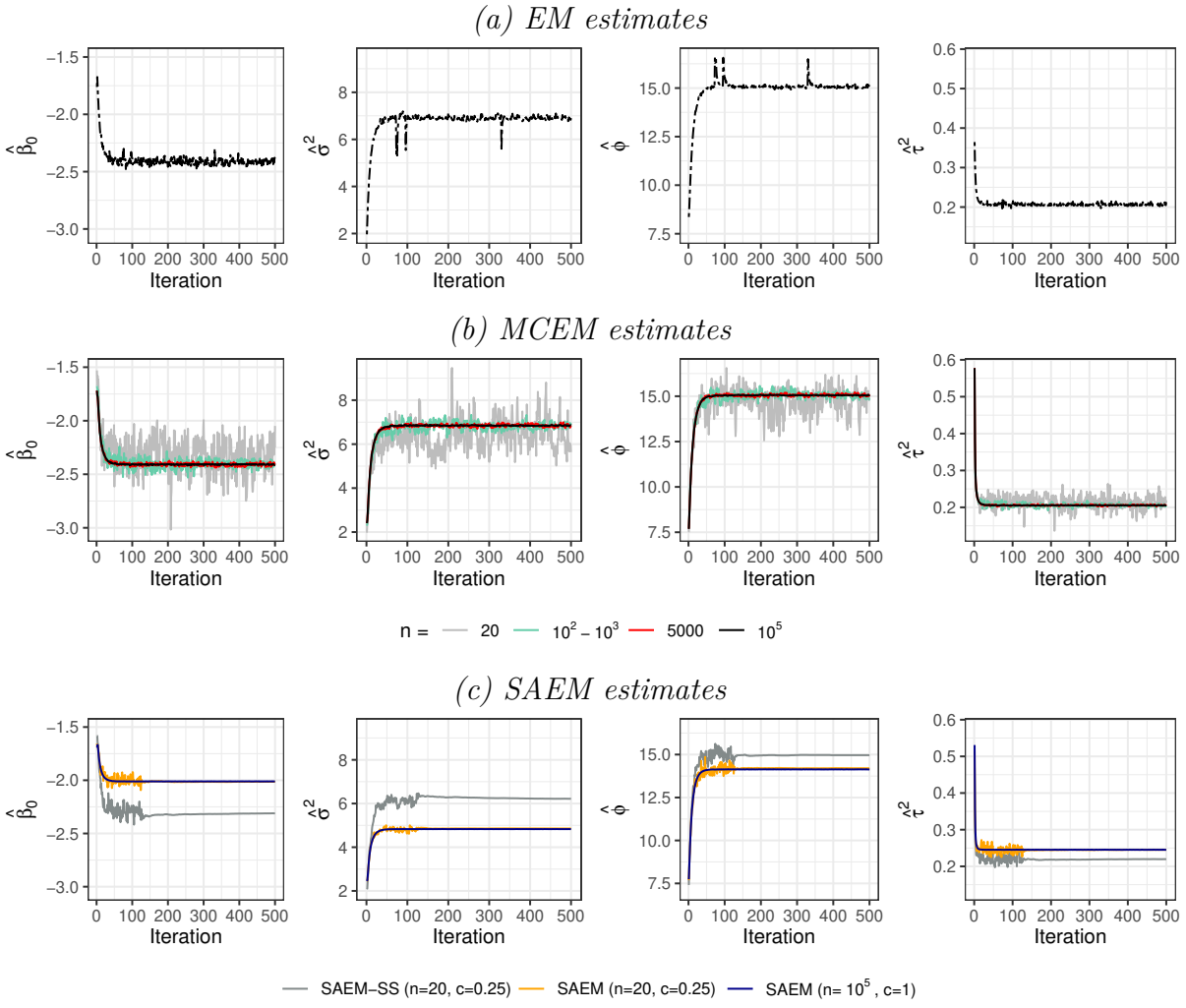


Figure 13 shows the convergence of parameter estimates from MCEM, SAEM-SS, SAEM, and EM algorithms. Notice that the variability in the MCEM estimates

decreases with increasing sample sizes from 100 to 1000 (aquamarine line). As expected, MCEM at $L = 20$ (gray line) presents more variability than larger sample sizes. Conversely, MCEM at $L = 10^5$ (black line) exhibits the lowest variability. Estimates from the EM algorithm show greater variability than those from MCEM at $L = 5000$ (red line), likely due to computational instability in the numerical methods employed by the `MomTrunc` package. this is why we decided to consider a burn-in and thinning procedure to compute the EM final estimates.

3.6 Remarks

In this chapter, we extended the works of Lachos *et al.* (2017) and Ordoñez *et al.* (2018) by employing the MCEM and EM algorithms to estimate the parameters of the spatial censored linear (SCL) model. We also proposed a variation of the SAEM algorithm, which utilizes our slice sampling method, detailed in Chapter 2 and available in the `relliptical` package. This variation of the SAEM algorithm has been demonstrated to be faster than the existing method in the `CensSpatial` package.

Additionally, we developed a method to approximate the standard error of the estimates using the Louis method, as described in Section 1.2. Based on our simulation studies, this method has proven to be reliable. The proposed methodologies have been coded and implemented in the R package `RcppCensSpatial`, readily available to users on the CRAN repository.

4 Censored autoregressive regression models with Student-t innovations

4.1 Introduction

A linear regression model is a commonly used statistical tool to analyze the relationship between a response (dependent) and some explanatory (independent) variables. In these models, the errors are usually considered independent and identically distributed random variables with zero mean and constant variance. However, observations collected over time are often autocorrelated rather than independent. Disregarding correlation in the error term may lead to underestimated standard errors (SE) of the estimated regression coefficients and fitted values (Gareth *et al.*, 2013); therefore, studying the dependence among observations is of considerable practical interest (see Tsay, 2005; Box *et al.*, 2015). Examples of this kind of data abound in economics, medicine, social sciences, and environmental monitoring; for instance, the total phosphorus concentration measured monthly in a river with correlated measures over time.

A stochastic model that has been successfully used in many real-world applications to deal with serial correlation in the error term is the autoregressive (AR) model. In the AR model, the current state of the process is expressed as a finite linear combination of previous states and a stochastic shock of disturbance, called an innovation in the time series literature. In general, it is assumed that the disturbance follows a normal distribution. For example, Alpuim & El-Shaarawi (2008) proposed a linear regression model with the sequence of error terms following a Gaussian autoregressive stationary process, in which the model parameters are estimated using maximum likelihood (ML) and ordinary least squares. However, it is well known that the normality assumption is often unrealistic, especially in the presence of outliers.

In this sense, Genton (2004) provides a broad review of extensions to parametric models using non-normal distributions. Besides, Tiku *et al.* (1999) suggested considering non-normal symmetric innovations in a simple regression model with AR(1) error term. More recently, Nduka (2018) developed an Expectation-Maximization (EM) algorithm to estimate the parameters in autoregressive models of order p with Student-t innovations.

An additional challenge arises when some observations are censored or missing. In the first case, values can occur out of the range of a measuring instrument, and in the second, the value is unknown. Censored time series are frequently encountered in environmental monitoring, medicine, and economics. Furthermore, there are some proposals in the literature related to censored autoregressive linear regression models with normal

innovations. For example, Wang & Chan (2018) suggested a quasi-likelihood method using the complete-incomplete data framework to estimate the parameters of an autoregressive censored linear model with Gaussian innovations. Schumacher *et al.* (2017) estimated the parameters via an analytically tractable and efficient stochastic approximation of the EM (SAEM) algorithm (Delyon *et al.*, 1999). Liu *et al.* (2019) proposed a coupled Monte Carlo Markov Chain (MCMC)-SAEM algorithm to fit an AR(p) regression model with Student-t innovations accounting for missing data.

Nevertheless, to the best of our knowledge, no studies consider the Student-t distribution for the innovations in censored autoregressive models from a likelihood-based perspective. Hence, in this work, we propose an EM-type algorithm to estimate the parameters of a censored regression model with autoregressive errors and innovations following a Student-t distribution. Specifically, the SAEM algorithm is considered to avoid the direct computation of complex expressions on the E-step of the EM algorithm (Dempster *et al.*, 1977). Its computational effort is much smaller in comparison to the Monte Carlo EM algorithm (Wei & Tanner, 1990), as shown by Jank (2006).

The chapter is organized as follows. Section 4.2 formulates the censored autoregressive model of order p with Student-t innovations, providing the log-likelihood function, the estimation procedure, the standard error approximation, and the expression to make predictions. Section 4.3 displays some results of two simulation studies carried out to examine the asymptotic properties of the estimators and demonstrate the robustness of the model. Section 4.4 applies the proposed model to the ammonia-nitrogen concentration dataset, along with an analysis of quantile residuals. Finally, Section 4.5 concludes with a discussion.

4.2 The censored AR regression model of order p

First, ignoring censoring, consider a linear regression model with autocorrelated errors as a discrete-time autoregressive process of order p; thus, the model for observation at time t is given by

$$Y_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \xi_t, \tag{4.1}$$

$$\xi_t = \phi_1 \xi_{t-1} + \ldots + \phi_p \xi_{t-p} + \eta_t, \quad \eta_t \overset{iid}{\sim} F(\cdot), \quad t \in \{1, \ldots, n\}, \tag{4.2}$$

where Y_t represents the response variable, $\mathbf{x}_t = (x_{t1}, \ldots, x_{tq})^\top$ is a $q \times 1$ vector of non-stochastic regressor variables, $\boldsymbol{\beta}$ is a $q \times 1$ vector of unknown regression parameters, and ξ_t is the AR error, with autoregressive coefficients $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^\top$ of length p and shock of disturbance η_t following a distribution $F(\cdot)$. The term η_t is also known as the innovation in the time series literature (see, for instance, Box *et al.*, 2015; Schumacher *et al.*, 2017; Wang & Chan, 2018).

Suppose that the innovation variable η_t in (4.2) follows a Student-t distribution with location parameter 0, scale parameter $\sigma^2 > 0$, and $\nu > 0$ degrees of freedom, denoted by $\eta_t \sim t(0, \sigma^2, \nu)$, whose probability density function can be written as

$$f(\eta; \sigma^2, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\pi\nu\sigma^2)^{\frac{1}{2}}}\left(1 + \frac{\eta^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}, \quad \eta \in \mathbb{R}, \tag{4.3}$$

where $\Gamma(a)$ denotes the gamma function evaluated at $a > 0$. Then, the model defined by (4.1)–(4.3) will be called the *autoregressive regression t model of order p* (AR$t(p)$).

Now assume that the response variable Y_t is not fully observed for all t, i.e., the true response lies within an interval instead of being observed exactly. Hence, let $R_t \subset \mathbb{R}$ denote the censoring region, such that Y_t is not observed if $Y_t \in R_t$. Further, let (V_t, C_t) be the observed data at time t, where C_t is the censoring indicator, and V_t is given by

$$V_t = \begin{cases} c_t, & \text{if } Y_t \in R_t, \quad \text{(censored)} \\ Y_t, & \text{otherwise}, \quad \text{(observed)} \end{cases} \tag{4.4}$$

where R_t is an interval of the form $(-\infty, c_t)$, (c_t, ∞), or (c_{t1}, c_{t2}) for left, right, or interval censoring, respectively. The constant $c_t \in \mathbb{R}$ is equal to the detection limit for left and right censoring, and $c_t = (c_{t1} + c_{t2})/2$ for interval censoring. Moreover, missing observations can be handled by setting $R_t = (-\infty, \infty)$ and $c_t = \text{NA}$. Thereby, the model defined by (4.1)–(4.4) will be referred to the *censored autoregressive regression t model of order p* (CAR$t(p)$) hereinafter. It is worth mentioning that our proposal assumes that the order p of the AR model is known in the parameter estimation process, but it can be chosen using information criteria such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

Next, we discuss an ML approach for estimating the parameters of the CAR$t(p)$ model using the SAEM algorithm. Alternatively, Bayesian parameter estimation can be performed using MCMC algorithms, which also take advantage of the conditional expressions derived in the Appendices but may require more computational effort.

4.2.1 The log-likelihood function

Let $\mathbf{y} = (y_1, \ldots, y_n)^\top$ be an observed sample of $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$. Let $\mathcal{F}_t = \sigma(Y_1, Y_2, \ldots, Y_t)$ denote the σ-field generated by $\{Y_1, Y_2, \ldots, Y_t\}$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \sigma^2, \nu)^\top$ be a generic parameter vector. The likelihood function for the AR$t(p)$ model, i.e., the model with non-censored observations, is given by

$$L(\boldsymbol{\theta}) = f(y_1, \ldots, y_p \mid \boldsymbol{\theta}) \prod_{t=p+1}^{n} f(y_t \mid \boldsymbol{\theta}, \mathcal{F}_{t-1}),$$

where $f(y_t \mid \boldsymbol{\theta}, \mathcal{F}_{t-1}) = f(y_t \mid \boldsymbol{\theta}, y_{t-1}, \ldots, y_{t-p})$ due to the AR(p) representation of the regression errors ξ_t. Note also that the conditional distribution of Y_t given $\mathbf{y}_{(t,p)} =$

$(y_{t-1},\ldots,y_{t-p})^\top$ will be a Student-t with location parameter $\omega_t = \mathbf{x}_t^\top\boldsymbol{\beta}+(\mathbf{y}_{(t,p)}-\mathbf{X}_{(t,p)}\boldsymbol{\beta})^\top\boldsymbol{\phi}$, scale parameter σ^2, and ν degrees of freedom, where $\mathbf{X}_{(t,p)} = [\mathbf{x}_{t-1}\,\mathbf{x}_{t-2}\,\cdots\,\mathbf{x}_{t-p}]^\top$ is a $p\times q$ matrix with the covariates related to the vector $\mathbf{y}_{(t,p)}$. After suppressing contributions from the initial p first values, we obtain a simpler conditional likelihood

$$L_*(\boldsymbol{\theta}) = \prod_{t=p+1}^{n} f(y_t \mid \boldsymbol{\theta}, y_{t-1}, y_{t-2},\ldots,y_{t-p}).$$

Now, to compute the log-likelihood function for the CAR$t(p)$ model, i.e., a model with censored/missing observations in the response vector, we treat the observed and censored components of the outcome $\mathbf{y}$ separately, i.e., as a partition. Let $\mathbf{y}_o \in \mathbb{R}^{n_o}$ be the vector of dimension n_o containing the fully observed elements, and let $\mathbf{y}_m$ denote the vector of dimension n_m of censored or missing components with censoring region $\mathbf{R} = R_{t_1}\times R_{t_2}\times\ldots\times R_{t_{n_m}}$, where $t_1, t_2,\ldots,t_{n_m}$ are the times in which censored observations occur. Thus, the observed (conditional) log-likelihood function can be computed by

$$\ell(\boldsymbol{\theta};\mathbf{y}_o) = \ln\left(\int_{\mathbf{R}}\prod_{t=p+1}^{n} f(y_t \mid \boldsymbol{\theta}, y_{t-1}, y_{t-2},\ldots,y_{t-p})\,\mathrm{d}\mathbf{y}_m\right). \tag{4.5}$$

Then, the ML estimation problem relies on maximizing the expression (4.5) over $\boldsymbol{\theta}$. However, the observed log-likelihood function given above involves expressions too complex to work with directly, which makes it not feasible in practical terms. To overcome this problem, we resort to an EM-type algorithm, which turns the maximization of (4.5) into optimizing a sequence of simpler approximations of this function.

4.2.2 The SAEM algorithm for ML Estimation

This section is devoted to obtaining the ML estimates of the CAR$t(p)$ model using the SAEM algorithm. It is worth mentioning that this approach enables the automatic approximation of the standard errors and the direct imputation of the censored and/or missing observations.

To implement an EM-type method, we require a representation of the model in terms of the missing data. Let's start by considering that if $Y_t \mid \mathbf{y}_{(t,p)} \sim t(\omega_t,\sigma^2,\nu)$, we can use the stochastic representation of the Student-t distribution to express $Y_t \mid \mathbf{y}_{(t,p)}$ as $Y_t \mid \mathbf{y}_{(t,p)} \overset{d}{=} \omega_t + U_t^{-1/2}Z_t$, where Z_t and U_t are two independent random variables. Specifically, Z_t follows a univariate normal distribution with zero-mean and variance σ^2, and U_t follows a gamma distribution with shape and rate parameters equal to $\nu/2$, (see, for instance, Kotz & Nadarajah, 2004). The expression $\overset{d}{=}$ means "equality in distribution". Consequently, it follows that

$$\begin{aligned} Y_t \mid (\mathbf{Y}_{(t,p)} = \mathbf{y}_{(t,p)}, U_t = u_t) &\sim N(\omega_t,\sigma^2/u_t),\\ U_t &\sim \text{Gamma}(\nu/2,\nu/2), \end{aligned}$$

for all $t \in \{p+1, \ldots, n\}$. This relationship is a convenient hierarchical representation and will be helpful in the E-step of the EM algorithm (Dempster *et al.*, 1977), which is discussed next.

To use an EM-type algorithm, we must regard both $\mathbf{y} = (y_1, \ldots, y_n)^\top$ and $\mathbf{u} = (u_{p+1}, \ldots, u_n)^\top$ as hypothetical missing data and $(\mathbf{V}, \mathbf{C})$ as observed data, with $\mathbf{V} = (V_1, \ldots, V_n)^\top$ and $\mathbf{C} = (C_1, \ldots, C_n)^\top$. The resulting complete response is $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{u}^\top, \mathbf{V}^\top, \mathbf{C}^\top)^\top$, and the complete (conditional) log-likelihood function $\ell_c(\boldsymbol{\theta}; \mathbf{y}_c)$ will take the form

$$\ell_c(\boldsymbol{\theta}; \mathbf{y}_c) = \frac{1}{2}\left[g(\nu) - (n-p)\ln\sigma^2 - \frac{1}{\sigma^2}\sum_{t=p+1}^{n} u_t(y_t - \omega_t)^2\right] + cte,$$

where cte represents a constant independent of the vector of parameters $\boldsymbol{\theta}$, $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \sigma^2, \nu)^\top$, and $g(\nu) = (n-p)g_1(\nu) + \nu g_2$ with $g_1(\nu) = \nu\ln(\nu/2) - 2\ln\Gamma(\nu/2)$ and $g_2 = \sum_{t=p+1}^{n}(\ln(u_t) - u_t)$. Therefore, the EM algorithm proceeds as follows:

E-step: Let $\widehat{\boldsymbol{\theta}}^{(k)}$ denote the current estimate of $\boldsymbol{\theta}$. The E-step calculates the conditional expectation of the complete data log-likelihood function, i.e.,

$$Q_k(\boldsymbol{\theta}) = \mathbb{E}\left[\ell_c(\boldsymbol{\theta}; \mathbf{y}_c) \mid \mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}\right].$$

For the proposed CAR$t(p)$ model, this step requires the computation of some conditional expectations provided in detail in Appendix C.1. However, the EM algorithm is not doable because of the difficulties in calculating the conditional expectations, specifically when we have successive censored observations. As an alternative, Wei & Tanner (1990) proposed the MCEM algorithm, which replaces the calculation of the expectations with a Monte Carlo approximation based on a large number of independent simulations of the missing data. However, due to the computational intensity, another option to consider is a stochastic approximation of these expectations, as Delyon *et al.* (1999) suggested with the so-called SAEM algorithm, which requires a much smaller number of observations for their approximation and makes more efficient use of the imputed missing values.

The SAEM algorithm splits the E-step of the EM algorithm into a simulation and an integration step (using a stochastic approximation), while the maximization step remains identical, except for the conditional expectations. Then, at the kth iteration, it proceeds as follows:

Step E-1 (Sampling). Simulate L samples from the conditional distribution of the latent variables via Gibbs sampler according to the following scheme:

- *Step 1*: Sample $\mathbf{y}_m^{(k,\ell)}$ from $f(\mathbf{y}_m \mid \mathbf{V}, \mathbf{y}_o, \mathbf{u}^{(k,\ell-1)}, \widehat{\boldsymbol{\theta}}^{(k)})$, which corresponds to the truncated multivariate normal distribution with location parameter $\mathbf{m}$ and scale matrix $\mathbf{S}$ as

defined at the end of Appendix C.2, and truncation region $\mathbf{R} = R_{t_1} \times R_{t_2} \times \ldots \times R_{t_{n_m}}$. Then update the full vector of observations as $\mathbf{y}^{(k,\ell)} = (\mathbf{y}_o^\top, \mathbf{y}_m^{(k,\ell)\top})^\top$.

- *Step 2*: Sample $\mathbf{u}^{(k,\ell)}$ from $f(\mathbf{u} \mid \mathbf{y}^{(k,\ell)}, \widehat{\boldsymbol{\theta}}^{(k)})$, where each component $u_t^{(k,\ell)}$ is simulated independently from a Gamma(a_t, b_t) distribution, for all $t \in \{p+1, \ldots, n\}$ and $\ell \in \{1, \ldots, L\}$, as demonstrated in Appendix C.2.

Step E-2 (Stochastic Approximation). Update $\widehat{Q}_k(\boldsymbol{\theta})$ as

$$\widehat{Q}_k(\boldsymbol{\theta}) = \widehat{Q}_{k-1}(\boldsymbol{\theta}) + \delta_k \left(\frac{1}{L} \sum_{\ell=1}^{L} \ell_c(\boldsymbol{\theta}; \mathbf{y}_c^{(k,\ell)}) - \widehat{Q}_{k-1}(\boldsymbol{\theta}) \right), \tag{4.6}$$

where $\mathbf{y}_c^{(k,\ell)} = (\mathbf{y}^{(k,\ell)\top}, \mathbf{u}^{(k,\ell)\top}, \mathbf{V}^\top, \mathbf{C}^\top)^\top$ and δ_k is a smoothing parameter, a decreasing sequence of positive numbers such that $\sum_{k=1}^{\infty} \delta_k = \infty$ and $\sum_{k=1}^{\infty} \delta_k^2 < \infty$, which was selected as described in Subsection 1.1.3.

M-step: Update the estimate $\widehat{\boldsymbol{\theta}}^{(k)}$ as $\widehat{\boldsymbol{\theta}}^{(k+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, \widehat{Q}_k(\boldsymbol{\theta})$.
The resulting expressions to estimate the model parameters are given in Appendix C.1, where closed-form expressions were obtained for $\boldsymbol{\beta}$, $\boldsymbol{\phi}$, and σ^2, while ν is estimated through an optimization procedure.

The steps above are iterated until some distance between two successive parameter estimations, such as $\left(\left(\widehat{\boldsymbol{\theta}}^{(k+1)} - \widehat{\boldsymbol{\theta}}^{(k)} \right)^\top \left(\widehat{\boldsymbol{\theta}}^{(k+1)} - \widehat{\boldsymbol{\theta}}^{(k)} \right) \right)^{1/2}$, becomes small enough. The initial values used in the SAEM algorithm were calculated by considering the censored values as observed ones and proceeding as in the Supplementary Material of Schumacher *et al.* (2017). Initial tests suggest that this choice of initial values is expected to give good numerical properties. However, to ensure that the estimates obtained by the EM-type algorithm correspond to a global maximum, it is recommended that users rerun the algorithm using a few different options of initial values and examine the attained log-likelihood value.

4.2.3 Standard error approximation

The Fisher information matrix is a good measure of the amount of information a sample dataset provides about the parameters of a given model, and it can be used to compute the asymptotic variance of the estimators. Louis (1982) developed a procedure for extracting the observed information matrix when the EM algorithm is applied to find the ML estimates in problems with partially observed data. On the other hand, Delyon *et al.* (1999) adapted the Louis method for the SAEM algorithm, resulting in the following stochastic approximation:

$$
\begin{aligned}
\mathbf{H}_k &= -\mathbf{F}_k - \mathbf{G}_k + \boldsymbol{\Delta}_k \boldsymbol{\Delta}_k^\top, \\
\mathbf{F}_k &= \mathbf{F}_{k-1} + \delta_k \left(\frac{1}{L} \sum_{l=1}^{L} \left(\frac{\partial^2 \ell_c(\boldsymbol{\theta}; \mathbf{y}_c^{(k,\ell)})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right) - \mathbf{F}_{k-1} \right), \\
\mathbf{G}_k &= \mathbf{G}_{k-1} + \delta_k \left(\frac{1}{L} \sum_{l=1}^{L} \left(\frac{\partial \ell_c(\boldsymbol{\theta}; \mathbf{y}_c^{(k,\ell)})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \ell_c(\boldsymbol{\theta}; \mathbf{y}_c^{(k,\ell)})}{\partial \boldsymbol{\theta}} \right)^\top - \mathbf{G}_{k-1} \right), \\
\boldsymbol{\Delta}_k &= \boldsymbol{\Delta}_{k-1} + \delta_k \left(\frac{1}{L} \sum_{l=1}^{L} \frac{\partial \ell_c(\boldsymbol{\theta}; \mathbf{y}_c^{(k,\ell)})}{\partial \boldsymbol{\theta}} - \boldsymbol{\Delta}_{k-1} \right),
\end{aligned}
\tag{4.7}
$$

where $\mathbf{y}_c^{(k,\ell)}$ is the complete data sampled at iteration (k, ℓ), for $k \in \{1, \ldots, W\}$ and $\ell \in \{1, \ldots, L\}$, with W denoting the maximum number of iterations and L the number of MC samples for the stochastic approximation. The inverse of the limiting value of $\mathbf{H}_k$ can be used to assess the dispersion of the estimators (Delyon *et al.*, 1999). The analytical expressions for the first and second derivatives of the complete data log-likelihood function are given in Appendix C.3.

4.2.4 Prediction

Interested in predicting values from the CAR$t(p)$ model, we denote by $\mathbf{y}_{\text{obs}}$ the n-vector of random variables corresponding to the given sample and by $\mathbf{y}_{\text{pred}}$ the vector of random variables of length n_{pred} corresponding to the time points that we are interested in predicting.

Let $\mathbf{y}_{\text{obs}} = (\mathbf{y}_o^\top, \mathbf{y}_m^\top)^\top$, where $\mathbf{y}_o$ is the vector of uncensored observations, and $\mathbf{y}_m$ is the vector of censored/missing observations. To deal with the incomplete values existing in $\mathbf{y}_{\text{obs}}$, we use an imputation procedure that consists of replacing the censored values with the values obtained in the last iteration of the SAEM algorithm, i.e., $\widehat{\mathbf{y}}_m = \mathbb{E}[\mathbf{y}_m \mid \mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(W)}] \approx \widehat{\mathbf{y}}_m^{(W)}$, since elements of $\widehat{\mathbf{y}}_m^{(k)}$ can also be updated during Step E-2 of the SAEM algorithm as

$$
\widehat{\mathbf{y}}_m^{(k)} = \widehat{\mathbf{y}}_m^{(k-1)} + \delta_k \left(\frac{1}{L} \sum_{l=1}^{L} \mathbf{y}_m^{(k,\ell)} - \widehat{\mathbf{y}}_m^{(k-1)} \right), \quad k \in \{1, \ldots, W\}, \tag{4.8}
$$

with the same δ_k, L, and W settings considered in the estimation procedure. The new vector of observed values will be denoted by $\mathbf{y}_{\text{obs}^*} = (\mathbf{y}_o^\top, \widehat{\mathbf{y}}_m^{(W)\top})^\top$.

Now, supposing that all values in $\mathbf{y}_{\text{obs}^*}$ are completely observed and that the explanatory variables for $\mathbf{y}_{\text{pred}}$ are available, the forecasting procedure will be performed

recursively (Box *et al.*, 2015) as follows

$$\widehat{y}_{n+k}(\boldsymbol{\theta}) = \begin{cases} \mathbf{x}_{n+k}^{\top}\boldsymbol{\beta} + \sum_{j=k}^{p} \phi_j y_{n+k-j}^{*}, & k = 1, \\ \mathbf{x}_{n+k}^{\top}\boldsymbol{\beta} + \sum_{i=1}^{k-1} \phi_i \widehat{y}_{n+k-i}^{*} + \sum_{j=k}^{p} \phi_j y_{n+k-j}^{*}, & 1 < k \leqslant p, \\ \mathbf{x}_{n+k}^{\top}\boldsymbol{\beta} + \sum_{j=1}^{p} \phi_j \widehat{y}_{n+k-j}^{*}, & p < k \leqslant n_{\text{pred}}, \end{cases}$$

such that $y_i^{*} = y_i - \mathbf{x}_i^{\top}\boldsymbol{\beta}$. In practice, y_{n+k} is obtained by substituting $\widehat{\boldsymbol{\theta}}$ in the last expression, i.e., $\widehat{y}_{n+k} = \widehat{y}_{n+k}(\widehat{\boldsymbol{\theta}})$, where $\widehat{\boldsymbol{\theta}}$ is the SAEM estimate. Therefore, the resulting vector with the predicted values is $\widehat{\mathbf{y}}_{\text{pred}} = (\widehat{y}_{n+1}, \widehat{y}_{n+2}, \ldots, \widehat{y}_{n+n_{\text{pred}}})^{\top}$.

4.3 Simulation Study

In this section, we examine the asymptotic properties of the SAEM estimates through a simulation study considering different sample sizes and levels of censoring. A second simulation study is performed to demonstrate the robustness of the estimates obtained from the proposed model when the data are perturbed.

4.3.1 Simulation study I

This study aims to provide empirical evidence about the consistency of the ML estimates under different scenarios. Therefore, 300 MC samples were generated with different sample sizes: $n = 100, 300$, and 600. The data were generated from the model defined by (4.1)–(4.3). The parameters were set as $\sigma^2 = 2$, $\nu = 4$, $\boldsymbol{\beta} = (5, 0.50, 0.90)^{\top}$, $\boldsymbol{\phi} = (-0.40, 0.12)^{\top}$, and the vector of covariables $\mathbf{x}_t = (1, x_{t1}, x_{t2})^{\top}$, with x_{t1} and x_{t2} simulated from a normal $N(0, 1)$ and a uniform $U(0, 1)$ distribution, respectively, for $t \in \{1, \ldots, n\}$. From this scenario, two analyses were conducted and will be discussed next.

4.3.1.1 Asymptotic properties

Aiming to have scenarios with an average level of censoring/missing of 5%, 20%, and 35%, respectively, we considered the following detection limits: 1.60, 3.45, and 4.30, i.e., values that fall below the detection limits (DL) are replaced with the detection limit value. Furthermore, 20% of the desired censored rate corresponds to observations randomly selected to be treated as missing. Additionally, we considered the case without censoring (original data) for comparison.

For each sample size and censoring level, we computed the mean (MC-Mean) and standard deviation (MC-SD) of the 300 MC estimates, the mean of the standard

Table 6 – Simulation I. Summary statistics of parameter estimates for the CAR$t(2)$ model based on 300 samples of sizes $n = 100, 300, 600$, and different levels of censoring.

n	DL	Average censoring level	Measure	Parameter						
				β_0	β_1	β_2	ϕ_1	ϕ_2	σ^2	ν
				5.00	0.50	0.90	-0.40	0.12	2.00	4.00
100	No	0%	MC-Mean	4.988	0.501	0.937	-0.409	0.093	1.779	4.150
			IM-SE	0.325	0.141	0.565	0.086	0.086	0.531	2.612
			MC-SD	0.334	0.149	0.567	0.090	0.091	0.465	2.089
			CP (%)	94.0	91.3	94.0	-	-	-	-
	1.60	5.10%	MC-Mean	4.984	0.502	0.942	-0.409	0.092	1.781	4.345
			IM-SE	0.319	0.141	0.557	0.091	0.088	0.533	3.326
			MC-SD	0.339	0.149	0.576	0.096	0.094	0.494	2.692
			CP (%)	94.3	92.0	94.3	-	-	-	-
	3.45	19.59%	MC-Mean	4.977	0.505	0.931	-0.418	0.089	1.812	4.559
			IM-SE	0.337	0.150	0.586	0.105	0.099	0.584	3.957
			MC-SD	0.349	0.156	0.588	0.110	0.103	0.516	2.914
			CP (%)	93.3	93.7	93.3	-	-	-	-
	4.30	34.11%	MC-Mean	4.939	0.511	0.928	-0.434	0.077	1.856	4.930
			IM-SE	0.371	0.163	0.644	0.122	0.115	0.721	4.872
			MC-SD	0.372	0.171	0.642	0.126	0.120	0.571	3.517
			CP (%)	94.0	93.7	93.3	-	-	-	-
300	No	0%	MC-Mean	5.015	0.501	0.893	-0.401	0.115	1.937	4.162
			IM-SE	0.167	0.084	0.299	0.048	0.048	0.293	1.172
			MC-SD	0.180	0.090	0.312	0.052	0.050	0.295	1.266
			CP (%)	93.3	93.7	93.7	-	-	-	-
	1.60	5.31%	MC-Mean	5.017	0.501	0.889	-0.399	0.115	1.944	4.244
			IM-SE	0.169	0.085	0.303	0.053	0.051	0.314	1.444
			MC-SD	0.182	0.091	0.319	0.055	0.051	0.311	1.498
			CP (%)	93.0	94.0	93.0	-	-	-	-
	3.45	20.58%	MC-Mean	4.999	0.507	0.891	-0.409	0.110	1.997	4.645
			IM-SE	0.177	0.090	0.317	0.060	0.057	0.365	2.122
			MC-SD	0.191	0.095	0.332	0.061	0.058	0.341	1.935
			CP (%)	93.7	92.3	93.7	-	-	-	-
	4.30	35.44%	MC-Mean	4.958	0.518	0.901	-0.409	0.110	2.080	5.043
			IM-SE	0.194	0.098	0.341	0.068	0.065	0.421	2.716
			MC-SD	0.199	0.106	0.339	0.074	0.069	0.385	2.258
			CP (%)	95.0	93.3	95.3	-	-	-	-
600	No	0%	MC-Mean	5.001	0.507	0.909	-0.400	0.118	1.966	4.072
			IM-SE	0.125	0.063	0.225	0.037	0.036	0.221	0.889
			MC-SD	0.125	0.064	0.226	0.038	0.037	0.212	0.918
			CP (%)	93.0	94.7	92.3	-	-	-	-
	1.60	5.11%	MC-Mean	4.999	0.508	0.911	-0.401	0.117	1.977	4.145
			IM-SE	0.125	0.063	0.225	0.037	0.036	0.221	0.899
			MC-SD	0.124	0.065	0.224	0.038	0.037	0.212	0.937
			CP (%)	92.7	94.3	92.0	-	-	-	-
	3.45	19.99%	MC-Mean	4.981	0.513	0.921	-0.405	0.117	2.034	4.447
			IM-SE	0.132	0.066	0.235	0.042	0.040	0.254	1.252
			MC-SD	0.127	0.069	0.229	0.042	0.041	0.239	1.342
			CP (%)	94.0	93.7	94.3	-	-	-	-
	4.30	34.64%	MC-Mean	4.939	0.521	0.933	-0.410	0.112	2.113	4.826
			IM-SE	0.142	0.072	0.251	0.048	0.045	0.300	1.672
			MC-SD	0.139	0.074	0.251	0.049	0.046	0.284	1.657
			CP (%)	93.7	94.7	94.7	-	-	-	-

errors (IM-SE) computed by the inverse of the observed information matrix given in (4.7), and the coverage probability (CP) of a 95% confidence interval, i.e.,

$$\text{MC-Mean}_i = \bar{\widehat{\theta}}_i = \frac{1}{300}\sum_{j=1}^{300}\widehat{\theta}_i^{(j)}, \quad \text{MC-SD}_i = \sqrt{\frac{1}{299}\sum_{j=1}^{299}\left(\widehat{\theta}_i^{(j)} - \bar{\widehat{\theta}}_i\right)^2}, \quad \text{and}$$

$$\text{IM-SE}_i = \frac{1}{300}\sum_{j=1}^{300} SE\left(\widehat{\theta}_i^{(j)}\right),$$

where $\widehat{\theta}_i^{(j)}$ is the estimate of the ith parameter of $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \phi_1, \phi_2, \sigma^2, \nu)^\top$ in the jth MC sample.

The results are shown in Table 6, where we can observe that the mean of the estimates (MC-Mean) is close to the true parameter value in all combinations of sample size and censoring levels, except for the scale parameter σ^2 for a sample size of $n = 100$. As expected, this difference decreases as the sample size increases. Notice that the mean of the standard errors obtained through the inverse of the observed information matrix (IM-SE) is, in general, close to the standard deviation of the estimates (MC-SD) for all scenarios, indicating that the proposed method to obtain the standard errors is reliable.

Figure 14 – Simulation I. Boxplot of the estimates for CARt(2) model by sample size and detection limit.

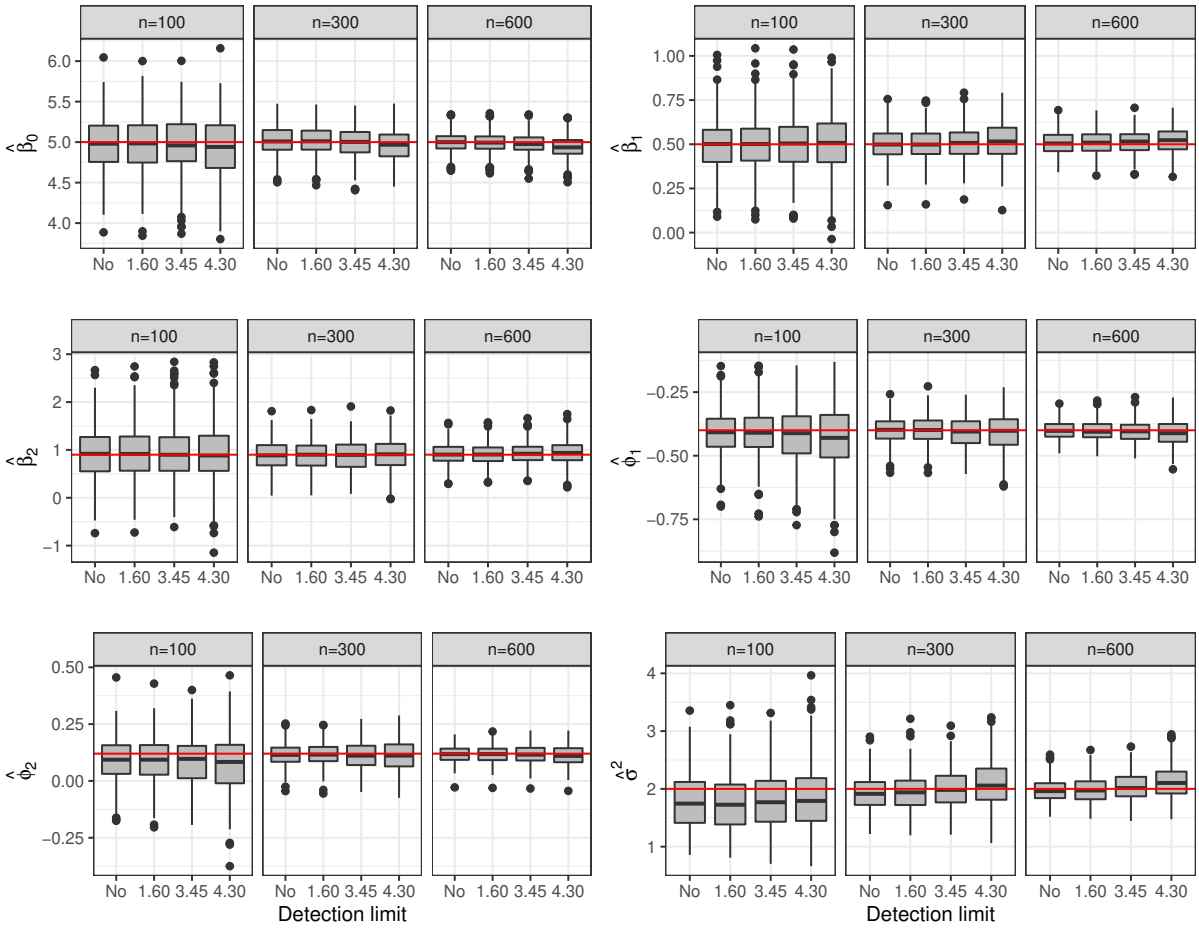


Figure 14 shows the boxplots of the estimates for each parameter by sample size and detection limit. The solid red line represents the true parameter value. In general, the median of the estimates is close to the real value independent of the sample size and

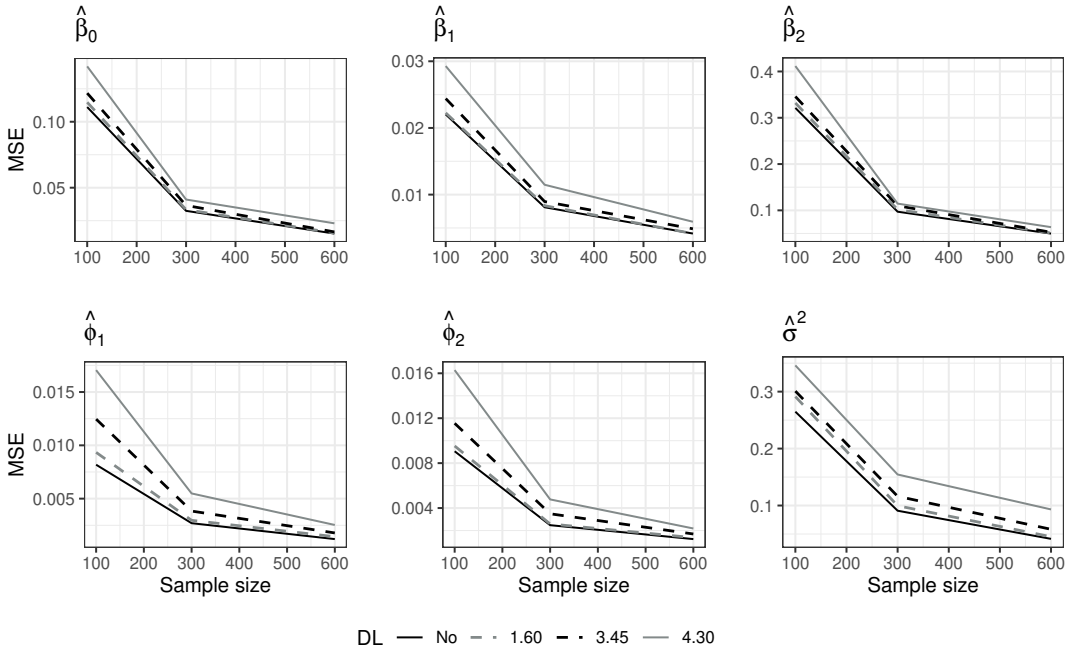
censoring level. However, for ϕ_2 and σ^2, the median underestimates the true value in samples of size $n = 100$, i.e., the smallest sample size in the simulation study. Furthermore, interquartile ranges decrease as sample sizes increase, suggesting the consistency of the estimates. Additionally, boxplots for the estimates of ν are shown in Appendix C.4.1, Figure 41.

To study the asymptotic properties of the estimates, we analyzed the mean squared error (MSE) of the estimates obtained from the proposed algorithm for all scenarios, which can be defined by

$$\text{MSE}_i = \frac{1}{n}\sum_{j=1}^{n}(\widehat{\theta}_i^{(j)} - \theta_i)^2.$$

The results for each parameter, sample size, and detection limits are shown in Figure 15, where we may note that the MSE tends to zero as the sample size increases. Thus, the proposed SAEM algorithm seems to provide ML estimates with good asymptotic properties for our proposed autoregressive censored linear model with Student-t innovations.

Figure 15 – Simulation I. MSE of the estimates for the CARt(2) model by sample size and detection limit.

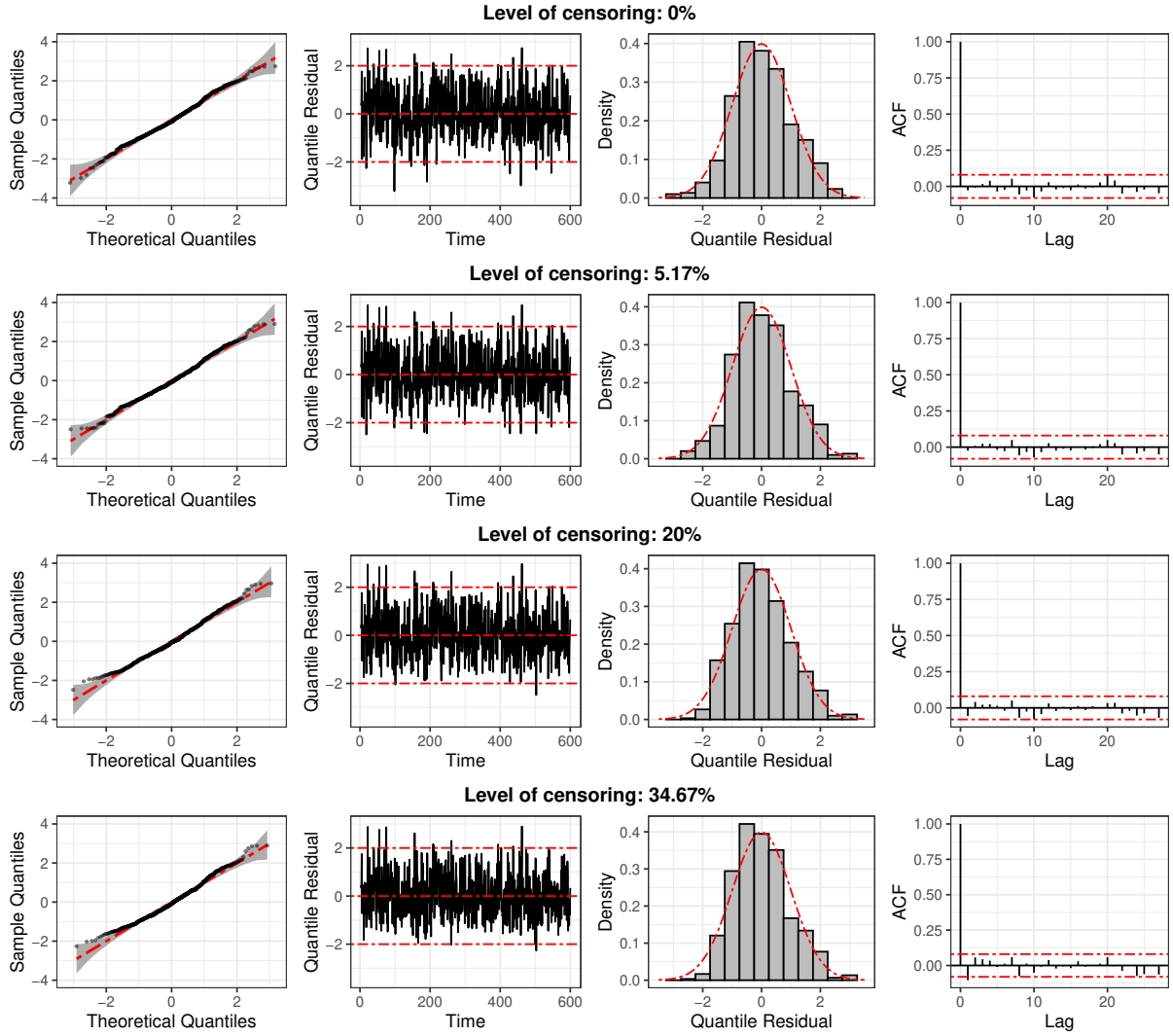


4.3.1.2 Residual Analysis

Checking the specification of a statistical model usually involves statistical tests and graphical methods based on residuals. However, conventional residuals (such as Pearson's residuals) are not appropriate for some models since they may lead to erroneous inference, as Kalliovirta (2012) demonstrated for models based on mixtures of distributions. As an alternative, Dunn & Smyth (1996) developed the quantile residuals method for regression models with independent responses, which produces normally

distributed residuals by inverting the fitted distribution function for each response value and finding the equivalent standard normal quantile. These results assume the model is correctly specified and parameters are consistently estimated. The method can be extended to dependent data by expressing the likelihood as a product of univariate conditional likelihoods.

Figure 16 – Simulation I. Plots of quantile residuals for a sample of size $n = 600$ generated from the CAR$t(2)$ model considering different levels of left censoring.



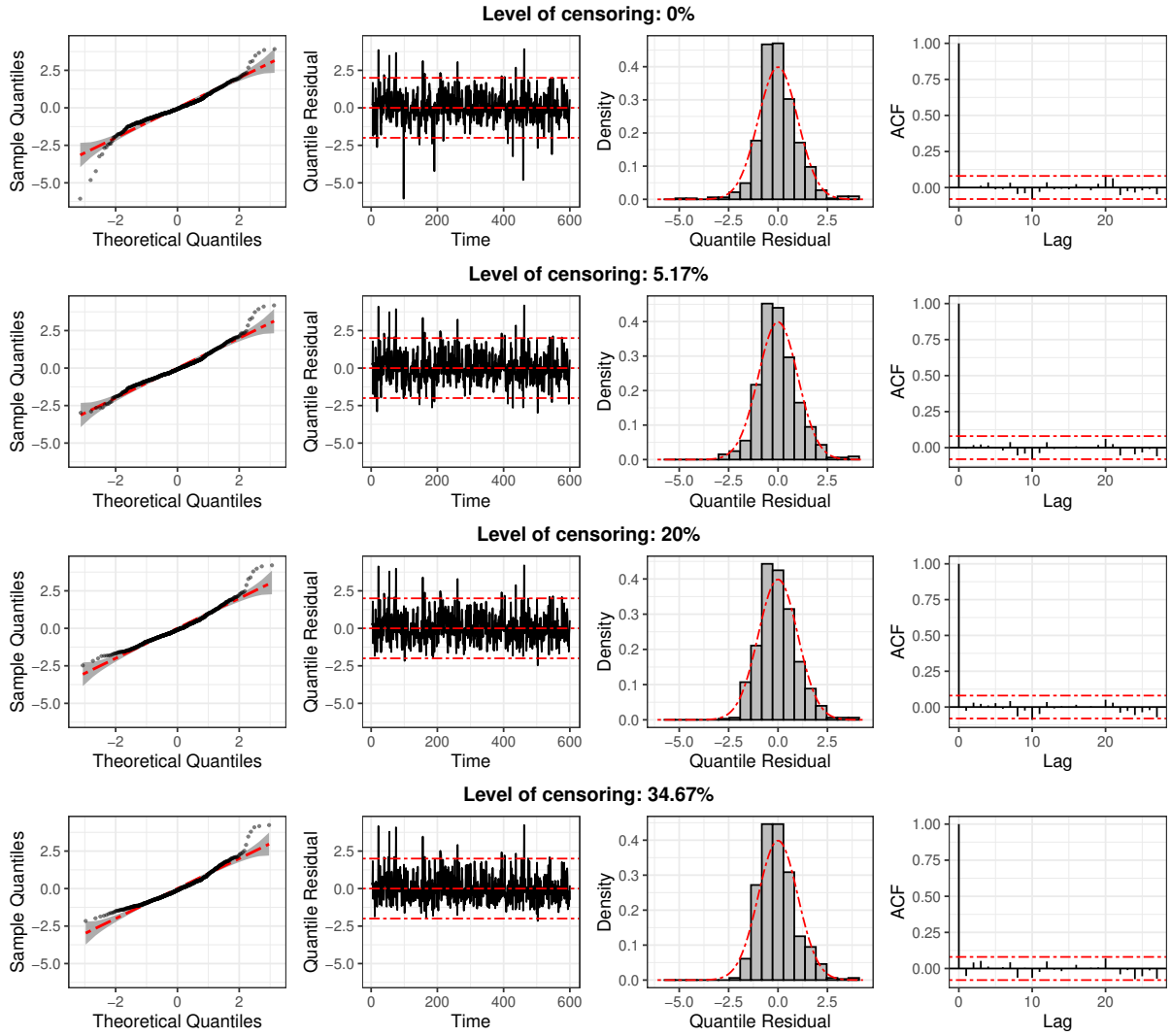
To compute the quantile residuals for the CAR$t(p)$ model, we first impute the censored or missing observations as defined in (4.8) in Subsection 4.2.4. Then, considering all values as completely observed, the residual for the tth observation is computed by

$$\widehat{r}_t = \Phi^{-1}\left(T\left(y_t; \hat{\omega}_t, \hat{\sigma}^2, \hat{\nu}\right)\right), \quad t \in \{p+1, \ldots n\}, \tag{4.9}$$

where $\Phi^{-1}(\cdot)$ denotes the inverse of the cumulative distribution function (cdf) of the standard normal distribution, $T(y_t; \hat{\omega}_t, \hat{\sigma}^2, \hat{\nu})$ is the cdf of the univariate Student-t distribution with location parameter $\widehat{\omega}_t = \mathbf{x}_t^\top \widehat{\boldsymbol{\beta}} + (\mathbf{y}_{(t,p)} - \mathbf{X}_{(t,p)}\widehat{\boldsymbol{\beta}})^\top \widehat{\boldsymbol{\phi}}$, scale parameter $\hat{\sigma}^2$, and $\hat{\nu}$ degrees of freedom evaluated at point y_t. Here $\widehat{\boldsymbol{\theta}}$ refers to the ML estimates of $\boldsymbol{\theta}$ obtained through

the SAEM algorithm. Note that the quantile residual is calculated only for the latest $n-p$ observations.

Figure 17 – Simulation I. Plots of the residuals for a sample of size $n = 600$ generated from the CAR$t(2)$ model and fitting a model with normal innovations, considering different levels of left censoring.



To analyze how the quantile residuals behave for the proposed model, they were computed for a simulated dataset of sample size $n = 600$, and we considered four levels of left censoring: 0%, 5.17%, 20%, and 34.67%. Figure 16 shows the Quantile-Quantile (Q-Q) plot, the quantile residual time series, the histogram, and the residual autocorrelation coefficients. For all levels of censoring, the histogram seems to correspond to a histogram of a normally distributed variable, and the dispersion plot shows independent residuals. We can deduce through the Q-Q plot that the residuals are roughly normally distributed because all points form a roughly straight line inside the confidence band. However, for samples with 20% and 34.67% of censoring, the Q-Q plots present a slight deviation from the center line in the lower tail, which might be due to the high proportion of censored values.

For comparison, we fitted the same dataset assuming the normal distribution (i.e., disregarding the heavy tails) and computed the corresponding quantile residuals. The resulting plots are given in Figure 17, where we can see clear signs of non-normality, such as large residuals and several points outside the confidence band in the Q-Q plots.

This illustration indicates that this method can help check the CAR$t(p)$ model specification. Nevertheless, more caution is needed in analyzing residuals for significant levels of censoring since our proposal imputes the unobserved data by its conditional expectation.

4.3.2 Simulation study II: Robustness of the estimators

This simulation study aims to compare the performance of the estimates for two censored AR models in the presence of outliers on the response variable. In this case, we simulated 300 MC samples of size $n = 100$ under the model defined in (4.1)–(4.2), considering the standard normal distribution for the innovations. The parameters were set as $\boldsymbol{\beta} = (4, 0.50)^\top$, $\boldsymbol{\phi} = (0.48, -0.20)^\top$, and the covariates $\mathbf{x}_t = (1, x_{t1})^\top$, where $x_{t1} \sim \mathcal{N}(0, 1)$, $t \in \{1, \ldots, 100\}$.

After generating the data, each MC sample was perturbed under the following scheme: the maximum value was increased in ϑ times the sample standard deviation, i.e., $y_{\text{pert}} = \max(\mathbf{y}) + \vartheta \text{SD}(\mathbf{y})$, for $\vartheta \in \{0, 1, 2, 3, 4, 5, 7\}$. Furthermore, we considered three different levels of censoring: the first case corresponds to the case without censoring; the second case considered 2.34 as a detection limit, where all values lower or equal than 2.34 were substituted with this particular value. (which implied an average of 10.04% of censoring); and finally the third scenario considered a limit of 3.31 (which yielded an average of 30% of censoring). For each scenario, we fitted two models: the first considers normal innovations, denoted by CAR(2), and the second one is our proposal CAR$t(2)$.

Table 7 displays the mean of the estimates for each parameter by the level of perturbation and censoring rate. To obtain comparable values of σ^2 for the model with Student-t innovations, we reported the estimated variance of the innovation $\widehat{\sigma}^{2*} = \widehat{\nu}\widehat{\sigma}^2/(\widehat{\nu} - 2)$, where $\widehat{\sigma}^2$ is the estimate of the scale parameter under our proposal. Moreover, the last column of Table 7 reports the percentage of times in which our model detected the perturbed observation as influential. This value was computed as the number of times the estimated weight ($\widehat{u}_i$, calculated during the E-step in the SAEM algorithm) of the perturbed observation was the lowest, divided by the number of MC samples.

Table 7a and Figure 18a show the results for the non-censored dataset. For the normal distribution, it is possible to observe that the bias increases as the perturbation increases. However, when Student-t innovations are considered, the bias is much smaller, illustrating the robustness of the model against atypical observations. As expected,

Table 7 – Simulation II. Mean of the estimates for CAR(2) and CARt(2) model based on 300 MC samples of size $n = 100$ for different levels of perturbation.

a. Level of censoring: 0%

Pert.	CAR(2)					CARt(2)						
(ϑ)	β_0	β_1	σ^2	ϕ_1	ϕ_2	β_0	β_1	σ^{2*}	ϕ_1	ϕ_2	ν	DI (%)
0	4.007	0.510	0.963	0.478	-0.215	4.007	0.510	0.979	0.480	-0.216	24.424	17.0
1	4.021	0.526	1.030	0.469	-0.209	4.008	0.518	1.042	0.464	-0.207	17.098	73.3
2	4.036	0.541	1.135	0.447	-0.195	3.999	0.518	1.130	0.433	-0.186	9.568	98.7
3	4.050	0.557	1.275	0.416	-0.175	3.991	0.515	1.229	0.398	-0.162	6.064	99.7
4	4.064	0.572	1.449	0.383	-0.154	3.987	0.513	1.328	0.369	-0.140	4.896	100.0
5	4.079	0.588	1.655	0.349	-0.133	3.986	0.510	1.426	0.346	-0.121	4.324	100.0
7	4.107	0.619	2.160	0.287	-0.100	3.985	0.508	1.628	0.309	-0.090	3.710	100.0

b. Level of censoring: 10.04%

Pert.	CAR(2)					CARt(2)						
(ϑ)	β_0	β_1	σ^2	ϕ_1	ϕ_2	β_0	β_1	σ^{2*}	ϕ_1	ϕ_2	ν	DI (%)
0	4.006	0.511	0.966	0.478	-0.214	4.006	0.511	0.997	0.478	-0.213	20.892	21.0
1	4.018	0.529	1.044	0.468	-0.208	4.007	0.519	1.079	0.460	-0.203	14.290	77.7
2	4.027	0.549	1.165	0.445	-0.194	4.002	0.518	1.207	0.424	-0.180	7.764	98.7
3	4.036	0.570	1.327	0.415	-0.175	3.999	0.514	1.362	0.387	-0.155	5.061	99.7
4	4.044	0.592	1.527	0.383	-0.156	3.999	0.511	1.535	0.357	-0.133	4.105	100.0
5	4.052	0.614	1.764	0.351	-0.138	4.000	0.509	1.704	0.332	-0.115	3.668	100.0
7	4.065	0.659	2.340	0.294	-0.109	4.003	0.505	2.170	0.295	-0.088	3.166	100.0

c. Level of censoring: 30%

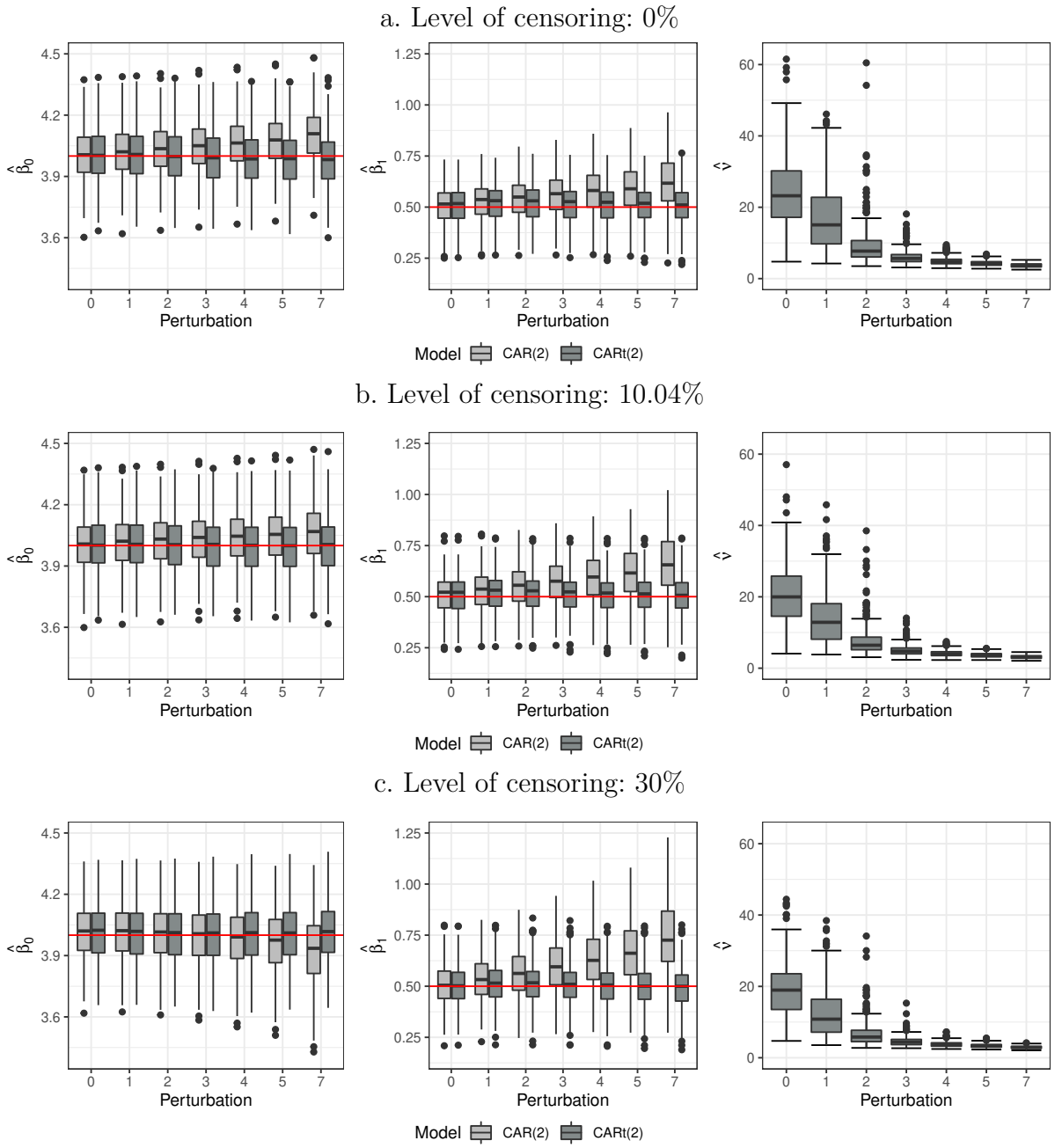
Pert.	CAR(2)					CARt(2)						
(ϑ)	β_0	β_1	σ^2	ϕ_1	ϕ_2	β_0	β_1	σ^{2*}	ϕ_1	ϕ_2	ν	DI (%)
0	4.014	0.505	0.944	0.476	-0.219	4.015	0.503	0.972	0.476	-0.218	19.338	24.7
1	4.014	0.532	1.050	0.464	-0.213	4.011	0.513	1.076	0.454	-0.206	12.553	82.3
2	4.007	0.562	1.213	0.439	-0.199	4.007	0.512	1.230	0.415	-0.181	6.854	98.3
3	3.996	0.594	1.429	0.409	-0.182	4.006	0.507	1.406	0.376	-0.156	4.651	100.0
4	3.982	0.628	1.694	0.378	-0.165	4.008	0.503	1.616	0.344	-0.135	3.803	100.0
5	3.965	0.664	2.005	0.349	-0.151	4.010	0.500	1.837	0.318	-0.118	3.386	100.0
7	3.926	0.737	2.760	0.301	-0.130	4.016	0.495	2.428	0.280	-0.092	2.926	100.0

estimates for ν decrease as the perturbation grows. Note that the observation with the maximum value was detected as influential in only 17% of the non-perturbed samples, but this percentage increases fast as the perturbation increases. The results for samples with an average of 10.04% censoring are displayed in Table 7b and Figure 18b, where the estimates for β_0 have a distribution similar to the non-censored case. On the other hand, for β_1, a more significant difference was observed between the real value and its estimate in the normal model. In contrast, the model with heavy-tailed innovations performs better in recovering the true parameter values, with a mean value of ν smaller than previously observed.

Results for the scenario with an average of 30% left censoring are shown in Table 7c and Figure 18c. For the normal case, the bias for β_1 is larger than was observed in the previous two cases. At the same time, the mean and median of the estimates obtained from the CARt(2) model are not much affected by the perturbations. For β_0, the normal

model returned estimates close to the real value only for levels of perturbation lower than 4 ($\vartheta < 4$), while for larger perturbations, the model tends to underestimate it. These results confirm that the heavy tails of the Student-t distribution allow our model to mitigate the effect of outliers, i.e., a much more robust method against atypical values.

Figure 18 – Simulation II. Boxplot of the estimates obtained from CAR(2) and CARt(2) model based on 300 MC samples of size $n = 100$.



4.4 The ammonia-nitrogen concentration data

This section provides an application of the CAR$t(p)$ model to a real environmental dataset with both missing and censored observations. We consider the ammonia-nitrogen

(NH_3-N) measurements taken in the Samish River in Washington State, USA, and described in Subsection 1.3.2. The data were collected monthly by the Washington State Department of Ecology, available for free download on the official website (https://ecology.wa.gov). Following IPCS (1986) and De Abreu *et al.* (2022), some water parameter factors that could have a statistical effect on the ammonia-nitrogen levels are the dissolved oxygen (DO) in mg/L, pH, and temperature (T) in $^{\circ}C$. In this study, we are interested in tracking the relationship between NH_3-N and those variables over time as an indicator of the river water quality.

Preliminary analysis (reported in Appendix C.4.2) for the linear regression model $y_t = \beta_0 + \beta_{\rm DO}{\rm DO}_t + \beta_{\rm pH}{\rm pH}_t + \beta_{\rm T}{\rm T}_t + \xi_t$, considering the errors as independent and identically normal distributed variables, indicates the presence of serial residual autocorrelation, hence we model ξ_t as an autoregressive process. Then, we fit the following censored regression model:

$$y_t = \beta_0 + \beta_{\rm DO}{\rm DO}_t + \beta_{\rm pH}{\rm pH}_t + \beta_{\rm T}{\rm T}_t + \xi_t, \quad t =\in \{1, \ldots, 248\},$$

where ξ_t is considered as an AR model of order $p = 1$ and 2, with innovations η_t independent and identically distributed as either $\mathcal{N}(0, \sigma^2)$ or $t(0, \sigma^2, \nu)$ (denoted by CAR(p) and CARt(p), respectively).

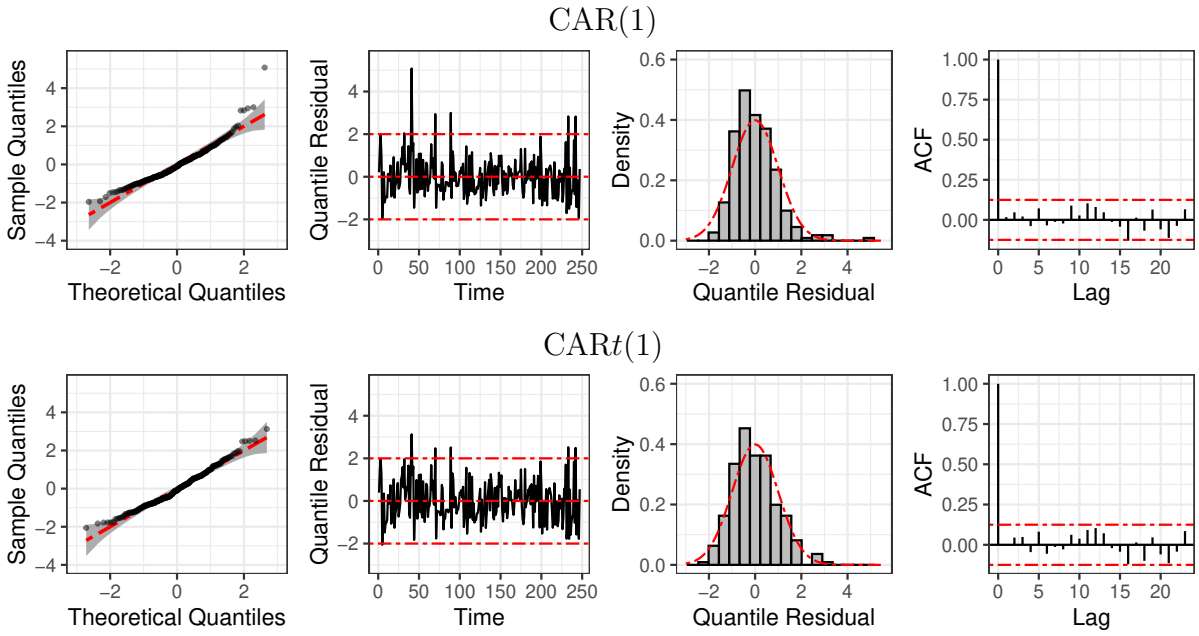
Parameter estimates and their corresponding standard errors (SEs) are displayed in Table 8. For model selection, we computed the observed conditional log-likelihood $\ell(\boldsymbol{\theta}; \mathbf{y}_o)$ defined in (4.5) by a Monte Carlo approximation, which is also shown, along with AIC and BIC values, in Table 8. We note that the estimated intercept (β_0) under the CARt(p) model is lower than the estimate from the CAR(p) model. The estimates for the dissolved oxygen, pH, and temperature regression coefficients are all negative, indicating an inverse relationship between those variables and the ammonia-nitrogen concentration. The ones provided by the CARt(p) model are larger than those obtained based on normality. The nitrification process could explain the inverse relationship between temperature and ammonia levels. According to Shammas (1986), the nitrification process increases as temperature increases (for $T \leqslant 35^oC$), and this process reduces the concentration of ammonia-nitrogen. Besides, the optimal pH for nitrification is between 7.5 and 8.5 (De Abreu *et al.*, 2022). In our study, the pH ranged between 6.6 and 8.4. Then, the increase in pH, within the range analyzed, provides more adequate conditions for the nitrification process; therefore, the concentration of ammonia-nitrogen is reduced. Nevertheless, we do not have evidence that the pH effect is significant.

Regarding the autoregressive coefficient ϕ_1, the normal estimate was smaller than the value estimated through CARt(1) model and ϕ_2 is not statistically significant for both CAR(2) and CARt(2). Therefore, comparing the AR(1) models based on information criteria (AIC and BIC), we conclude that the heavy-tailed Student-t model (with $\widehat{\nu} = 4.091$) provides a better fit to the ammonia-nitrogen concentration data.

Table 8 – Square-root of NH_3-N (μg/L). Parameter estimates and their standard errors (SE) for the CAR(p) and CARt(p) model, for $p = 1$ and 2.

Param.	CAR(1)		CAR(2)		CARt(1)		CARt(2)	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
β_0	17.958	3.533	18.277	3.589	16.441	3.016	16.667	3.140
β_{DO}	-0.904	0.184	-0.888	0.184	-0.689	0.167	-0.661	0.171
β_{pH}	-0.253	0.445	-0.331	0.461	-0.432	0.377	-0.522	0.406
β_{T}	-0.269	0.059	-0.262	0.059	-0.224	0.055	-0.213	0.057
σ^2	1.462	0.161	1.462	0.163	0.745	0.163	0.751	0.167
ϕ_1	0.319	0.067	0.285	0.073	0.341	0.070	0.295	0.076
ϕ_2	-		0.086	0.078	-		0.119	0.072
ν	-		-		4.091	1.384	4.197	1.370
$\ell(\widehat{\boldsymbol{\theta}}; \mathbf{y}_o)$	-302.957		-301.226		-297.503		-295.853	
AIC	617.914		616.452		609.007		**607.705**	
BIC	638.970		640.989		**633.572**		635.748	

Figure 19 – Square-root of NH_3-N (μg/L). Quantile residuals for the CAR(1) and CARt(1) models.

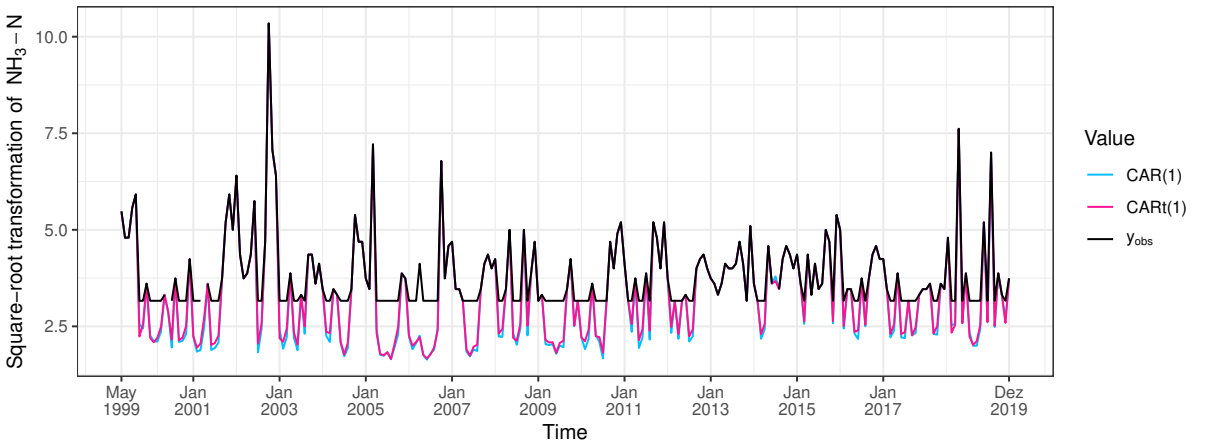


Graphs related to the quantile residuals computed for the CAR(1) and CARt(1) models are shown in Figure 19. The Q-Q plot for the CAR(1) model presents several points outside the confidence bands on the upper and lower tails, indicating that the distribution is heavy-tailed. Additionally, we see larger residual values from the histogram and residual plot. For the CARt(1) model, we see in the Q-Q plot that all points form a roughly straight line and lie within the confidence bands. Further, the histogram seems to correspond to a normally distributed variable, and the autocorrelation coefficients fall within the two standard error bounds. Therefore, the CARt(1) model seems to fit better the ammonia-nitrogen concentration data than the CAR(1) model, which is confirmed

through a likelihood-ratio test whose p-value is less than 0.001.

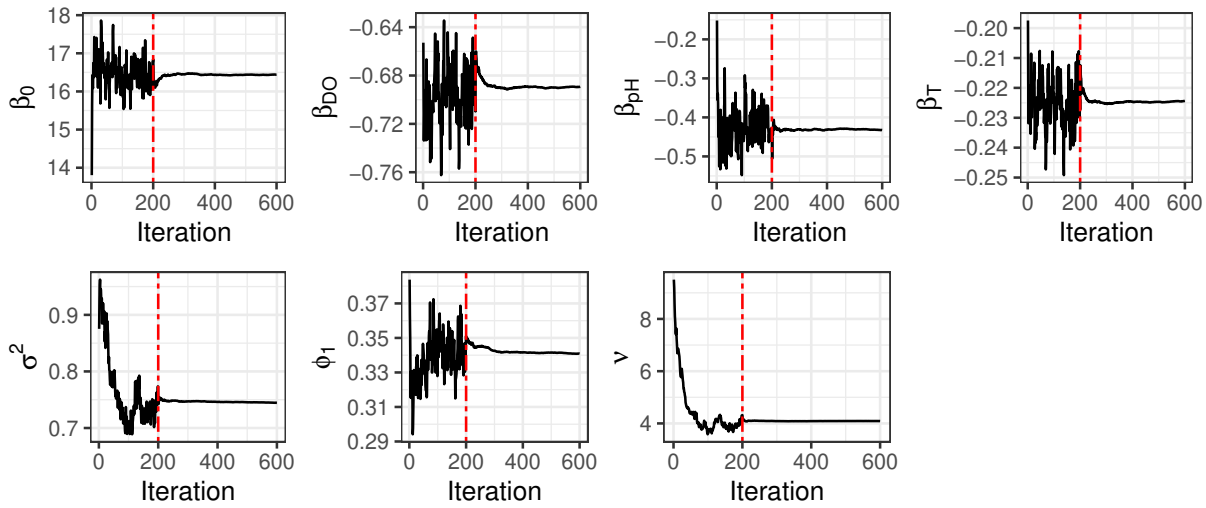
Figure 20 shows the observed values (solid black line) and the imputed values for the censored and missing observations from May 1999 to December 2019 under the CAR(1) and CARt(1) models. Blank spaces in the time series represent the missing values.

Figure 20 – Square-root of NH_3-N (μg/L). Observed (black solid line) and imputed values considering Student-t (pink line) and normal (light blue line) innovations.



In addition, for assessing the convergence of SAEM parameter estimates, convergence plots are displayed in Figure 21. Finally, an additional application related to the total phosphorus concentration can be found in Appendix C.4.3.

Figure 21 – Square-root of NH_3-N (μg/L). Convergence of the parameter estimates for CARt(1) model.



4.5 Remarks

Extending autoregressive regression methods to include censored response variables is a promising area of research. This chapter introduced a novel model that can

handle left, right, or interval-censoring time series while simultaneously modeling heavy tails and missing observations, which can be treated as interval-censored observations. Our approach extends some previous works, such as Schumacher *et al.* (2017) and Liu *et al.* (2019).

The proposed methods were applied to several simulation studies, which revealed that our model provides ML estimates with good asymptotic properties since the MSE of the estimates tends to zero as the sample size increases. The model with Student-t innovations (our proposal) was more robust against atypical values than the normal model. In addition, the quantile residuals could help check the CARt(p) model specification; however, for significant levels of censoring, more caution is needed in analyzing residuals. The proposed methods have been coded and implemented in an R package `ARCensReg`.

It is important to remark that we assumed the dropout/censoring mechanism to be missing at random (MAR) (see Diggle *et al.*, 2002, p. 283). However, when MAR with ignorability is not realistic, the relationship between the unobserved measurements and the censoring process should be further investigated. Future directions point to tackling the limitation of assuming that the first p observations are fully observed to fit a CAR$t(p)$ model. Furthermore, a natural and interesting path for future research is to extend this model to a multivariate framework.

5 Likelihood-based inference for the skew-t regression model for multiple outcomes with censored or missing responses

5.1 Introduction

The study of models in which the variable of interest is subject to certain threshold values below or above which the measurements are not quantifiable has been a common topic of interest in the statistical literature in recent years. For example, in environmental research, the concentration levels of the dissolved trace metals in freshwater streams in Virginia are subject to multiple limits of detection values (see, for instance, Galarza *et al.*, 2022b). In AIDS research, the quantification of viral load measurements is typically assessed according to certain upper and lower detection limits. As a result, the viral load responses are either left or right-censored depending on the diagnostic assays used (Wu, 2010). This kind of data can be modeled using censored regression (CR) models, or the Tobit model, and has become quite common in the literature with a wide range of applications in biology, biometrics, genetics, medicine, finance, and marketing, among many others.

CR models usually use the normal distribution for mathematical convenience for continuous data. However, it is well known that the normal distribution (N-CR) is sensitive to outliers. Moreover, the use of N-CR may be unsuitable for a set of data containing observations with heavy tails or asymmetric behavior. It can unduly affect the fit of the CR model. This inconsistency in the N-CR model led to the development of less sensitive estimators to the assumption of normality. Several authors have studied CR models involving response variables with heavier tails than the normal distribution in recent years. For instance, Massuia *et al.* (2015) have studied CR models based on the univariate Student-t distribution (T-CR). In a multivariate setting, Galarza *et al.* (2021c) (see also Matos *et al.*, 2013b; Garay *et al.*, 2017) advocated the use of the multivariate Student-t distribution in the context of CR models, where a simple and efficient EM-type algorithm for iteratively computing maximum likelihood (ML) estimates of the parameters was presented. They demonstrated the robustness aspects of the T-CR model against outliers through extensive simulations using the Expectation-Maximization (EM) algorithm, which is based on the first two moments of the multivariate truncated Student-t distribution. More recently, Galarza *et al.* (2022b) proposed the multivariate skew-normal (SN) distribution to analyze censored or missing data (SN-CR), and a fully likelihood-based

approach is carried out, including the implementation of an EM-type algorithm for ML estimation. However, neither the T-CR model nor the SN-CR model is appropriate when data simultaneously present skewness and heavy-tailed behavior.

In this paper, we attempt to overcome these limitations in the aforementioned CR models by proposing a more adapted and robust CR model that can simultaneously deal with the issues of censored and/or missing data, skewness, and heavy-tailed and atypical data. Our contribution extends the recent works of Galarza *et al.* (2021c) and Galarza *et al.* (2022b) since they considered only the Student-t and the SN distribution, respectively, which are particular cases of the skew-t (ST) family of distributions, including the popular normal one. We show that the E-step reduces to computing the first two moments of a truncated multivariate Student-t, skew-t, and extended skew-t distributions, which are implemented in the `MomTrunc` (Galarza *et al.*, 2021a) and `relliptical` (Valeriano *et al.*, 2022) R packages. The likelihood function is easily computed as a by-product of the E-step and is used for monitoring convergence and model selection. Furthermore, we consider a general information-based method for obtaining the asymptotic covariance matrix of the ML estimates.

The paper is organized as follows. Section 5.2 introduces some notations and outlines the main results of the ST and truncated skew-t (TST) distributions. In Section 5.3, the ST censored regression model (ST-CR) and related likelihood-based inference are presented, including the implementation of an EM-type algorithm called the Expectation/Conditional Maximization Either (ECME) algorithm (Liu & Rubin, 1994) for obtaining ML estimates of the parameters. Section 5.4 presents some simulation studies to illustrate the performance of the proposed method. Section 5.5 discusses two real data applications in environmental and astronomical research. Finally, Section 5.6 concludes with some discussion and possible directions for future research.

5.2 Preliminaries

We start by defining the multivariate ST distribution and multivariate extended ST (EST) distribution and some of their useful properties. Some versions and extensions of the ST family are discussed in works such as Arellano-Valle & Genton (2005, 2010b,a); Azzalini & Capitanio (2003), and Sahu *et al.* (2003).

5.2.1 The multivariate skew-t distribution

We say that a continuous p-dimensional random vector $\mathbf{Y}$ follows a multivariate ST distribution with location vector $\boldsymbol{\mu} \in \mathbb{R}^p$, positive definite dispersion matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, parameter vector $\boldsymbol{\lambda} \in \mathbb{R}^p$, and degrees of freedom $\nu > 0$, denoted by $\mathbf{Y} \sim \mathrm{ST}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$,

if its pdf is given by

$$ST_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu) = 2t_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)\,T\left\{\sqrt{\frac{\nu+p}{\nu+\delta(\mathbf{y})}}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu});\nu+p\right\}, \tag{5.1}$$

where $\delta(\mathbf{y}) \equiv \delta(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma}) = (\mathbf{y}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})$ is the squared Mahalanobis distance, and $\boldsymbol{\Sigma}^{-1/2}$ is the unique square root matrix of $\boldsymbol{\Sigma}^{-1}$ such that $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2}$. It is well known that when $\nu \to \infty$, we retrieve the multivariate skew-normal distribution. Furthermore, if $\boldsymbol{\lambda} = \mathbf{0}$, then (5.1) reduces to the multivariate symmetric $t_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$ pdf. The following propositions of the ST distribution are helpful for our theoretical developments; we start with the cdf of an ST random vector. The proof of the propositions can be found in Arellano-Valle & Genton (2010b).

Proposition 2 (cdf *of the* ST)**.** *If* $\boldsymbol{Y} \sim \mathrm{ST}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)$*, then for any* $\boldsymbol{y} \in \mathbb{R}^p$

$$F_{\boldsymbol{Y}}(\boldsymbol{y}) = Pr(\boldsymbol{Y} \leqslant \boldsymbol{y}) = 2T_{p+1}\big((\boldsymbol{y}^\top,0)^\top;\boldsymbol{\mu}^*,\boldsymbol{\Omega},\nu\big),$$

where $\boldsymbol{\mu}^* = (\boldsymbol{\mu}^\top,0)^\top$, $\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Sigma} & -\boldsymbol{\Delta} \\ -\boldsymbol{\Delta}^\top & 1 \end{pmatrix}$, *and* $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}/(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}$.

Proposition 3 (*Hierarchical representation of the* ST)**.** *The p-variate random vector* $\boldsymbol{Y} \sim \mathrm{ST}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)$ *admits the following hierarchical representation*

$$\boldsymbol{Y} \mid (U=u,T=t) \sim N_p(\boldsymbol{\mu}+\boldsymbol{\Delta}t, u^{-1}\boldsymbol{\Gamma}), \quad T \mid (U=u) \sim \mathrm{TN}(0,u^{-1};(0,\infty)),$$
$$U \sim Gamma(\nu/2,\nu/2),$$

where $\mathrm{TN}(\mu,\sigma^2;(a,b))$ *denotes a univariate truncated normal distribution with location parameter* μ*, scale parameter* σ^2*, and truncation region* (a,b)*, and* $\mathrm{Gamma}(\alpha,\lambda)$ *represents the gamma distribution with shape and rate parameters* α *and* λ*, respectively;* $\boldsymbol{\Delta}$ *as given in Proposition 2, and* $\boldsymbol{\Gamma} = \boldsymbol{\Sigma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top$.

It is worth noting that the multivariate ST distribution is not closed under conditioning, as discussed in Arellano-Valle & Genton (2010b). Consequently, we present its generalization, known as the multivariate extended skew-t (EST) distribution.

5.2.2 The multivariate extended skew-t distribution

We say that a p-dimensional random vector $\mathbf{Y}$ follows an EST distribution with location vector $\boldsymbol{\mu} \in \mathbb{R}^p$, positive definite dispersion matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, skewness parameter vector $\boldsymbol{\lambda} \in \mathbb{R}^p$, shift parameter $\tau \in \mathbb{R}$, and degrees of freedom $\nu > 0$, denoted by $\mathbf{Y} \sim \mathrm{EST}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau,\nu)$, if its pdf is given by

$$EST_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau,\nu) = \frac{t_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)}{T(\tau/\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}};\nu)}\,T\left\{\sqrt{\frac{\nu+p}{\nu+\delta(\mathbf{y})}}\left(\tau+\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})\right);\nu+p\right\}. \tag{5.2}$$

Note that when $\tau = 0$, we retrieve the ST distribution defined in (5.1), that is, $EST_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, 0, \nu) = ST_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$. Here, we employed a variation in the parametrization of the EST distribution compared to that described in Arellano-Valle & Genton (2010b). Specifically, the shape parameter $\boldsymbol{\lambda}^*$ as introduced in Definition 1 of the referenced work corresponds to our parameterization through the relationship $\boldsymbol{\lambda} = \bar{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{\lambda}^*$, where $\bar{\boldsymbol{\Sigma}} = \boldsymbol{\omega}^{-1}\boldsymbol{\Sigma}\boldsymbol{\omega}^{-1}$ and $\boldsymbol{\omega} = diag(\boldsymbol{\Sigma})^{1/2}$. It is evident from the reference that

$$EST_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, \nu) \quad \longrightarrow \quad t_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu), \quad \tau \to +\infty.$$

The following propositions are related to the stochastic representation, cdf, and the marginal and conditional distribution of the EST random vector.

Proposition 4 (*Stochastic representation of the* EST)**.** *Let* $\boldsymbol{X} = (\boldsymbol{X}_1^\top, X_2)^\top \sim t_{p+1}(\boldsymbol{\mu}^*, \boldsymbol{\Omega}, \nu)$. *If*

$$\boldsymbol{Y} \overset{d}{=} (\boldsymbol{X}_1 | X_2 < \tilde{\tau}),$$

it follows that $\boldsymbol{Y} \sim \text{EST}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, \nu)$, *with* $\boldsymbol{\mu}^*$ *and* $\boldsymbol{\Omega}$ *as in Proposition 2, and* $\tilde{\tau} = \tau/(1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}$.

The stochastic representation of an EST is also provided in Arellano-Valle & Genton (2010b, Proposition 1) under a different parametrization. Now letting $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu})$, it follows that $\mathbf{Z} \sim \text{EST}_p(\mathbf{0}, \mathbf{I}, \boldsymbol{\lambda}, \tau, \nu)$. Thence, the mean vector and variance-covariance matrix of $\mathbf{Z}$ can be computed using the stochastic representation given in Proposition 4, which are

$$\mathbb{E}[\mathbf{Z}] = -\eta_1\boldsymbol{\Delta}^*, \quad \text{Cov}[\mathbf{Z}] = \gamma(\mathbf{I}_p - \boldsymbol{\Delta}^*\boldsymbol{\Delta}^{*\top}) + (\eta_2 - \eta_1^2)\boldsymbol{\Delta}^*\boldsymbol{\Delta}^{*\top},$$

with $\boldsymbol{\Delta}^* = \boldsymbol{\lambda}/(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}$, $\gamma = \dfrac{\nu + \eta_2}{\nu - 1}$, $\eta_1 = \mathbb{E}[X_2 | X_2 < \tilde{\tau}] = \dfrac{\nu}{1-\nu}\left(1 + \dfrac{\tilde{\tau}^2}{\nu}\right)\dfrac{t(\tilde{\tau}; \nu)}{T(\tilde{\tau}; \nu)}$, $\nu > 1$, and $\eta_2 = \mathbb{E}[X_2^2 | X_2 < \tilde{\tau}] = \dfrac{\nu(\nu-1)}{\nu-2}\dfrac{T(\sqrt{(\nu-2)/\nu}\,\tilde{\tau}; \nu-2)}{T(\tilde{\tau}; \nu)} - \nu$, $\nu > 2$.

Then, the mean vector and variance-covariance matrix of $\mathbf{Y}$ can be easily computed as

$$\mathbb{E}[\mathbf{Y}] = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbb{E}[\mathbf{Z}], \quad \text{Cov}[\mathbf{Y}] = \boldsymbol{\Sigma}^{1/2}\text{Cov}[\mathbf{Z}]\boldsymbol{\Sigma}^{1/2}. \tag{5.3}$$

Proposition 5 (cdf *of the* EST)**.** *If* $\boldsymbol{Y} \sim \text{EST}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, \nu)$, *then for any* $\boldsymbol{y} \in \mathbb{R}^p$

$$F_{\boldsymbol{Y}}(\boldsymbol{y}) = Pr(\boldsymbol{Y} \leqslant \boldsymbol{y}) = \frac{T_{p+1}\big((\boldsymbol{y}^\top, \tilde{\tau})^\top; \boldsymbol{\mu}^*, \boldsymbol{\Omega}, \nu\big)}{T(\tilde{\tau}; \nu)}.$$

Proof. The proof is a direct consequence of Proposition 4. □

Proposition 6 (*Marginal and conditional distribution of the* EST)**.** *Let* $\boldsymbol{Y} \sim \text{EST}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, \nu)$ *be partitioned as* $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, \boldsymbol{Y}_2^\top)^\top$ *of dimensions* p_1 *and* p_2 $(p_1 + p_2 = p)$, *respectively. Let*

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad \boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top, \quad \textit{and} \quad \boldsymbol{\varphi} = (\boldsymbol{\varphi}_1^\top, \boldsymbol{\varphi}_2^\top)^\top$$

be the corresponding partitions of $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}$, *and* $\boldsymbol{\varphi} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda}$. *Then,*

$$\begin{aligned} \boldsymbol{Y}_1 &\sim \text{EST}_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, c_{12}\boldsymbol{\Sigma}_{11}^{1/2}\tilde{\boldsymbol{\varphi}}_1, c_{12}\tau, \nu), \\ \boldsymbol{Y}_2 | \, \boldsymbol{Y}_1 = \boldsymbol{y}_1 &\sim \text{EST}_{p_2}(\boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{22.1}/\nu^2(\boldsymbol{y}_1), \boldsymbol{\Sigma}_{22.1}^{1/2}\boldsymbol{\varphi}_2, \nu(\boldsymbol{y}_1)\tau_{2.1}, \nu + p_1), \end{aligned}$$

where $c_{12} = (1+\boldsymbol{\varphi}_2^\top \boldsymbol{\Sigma}_{22.1}\boldsymbol{\varphi}_2)^{-1/2}$, $\tilde{\boldsymbol{\varphi}}_1 = \boldsymbol{\varphi}_1 + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\varphi}_2$, $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{y}_1 - \boldsymbol{\mu}_1)$, $\tau_{2.1} = \tau + \tilde{\boldsymbol{\varphi}}_1^\top(\boldsymbol{y}_1 - \boldsymbol{\mu}_1)$, *and* $\nu^2(\boldsymbol{y}_1) = (\nu + p_1)/(\nu + \delta(\boldsymbol{y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}))$.

Proof. See Appendix D.1. □

Now, we introduce a key concept to our theory, namely the truncated extended skew-*t* (TEST) distribution.

5.2.3 The truncated extended skew-*t* distribution

Let $\mathbf{A}$ be a Borel set in $\mathbb{R}^p$. The random vector $\mathbf{Y}$ has a TEST distribution on $\mathbf{A}$ when $\mathbf{Y}$ has the same distribution as $\mathbf{W} \mid (\mathbf{W} \in \mathbf{A})$ such that $\mathbf{W} \sim \text{EST}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, \nu)$. We use the notation $\{\mathbf{Y} \in \mathbf{A}\} = \{\mathbf{a} \leqslant \mathbf{Y} \leqslant \mathbf{b}\}$, where $\mathbf{a} = (a_1, \ldots, a_p)^\top$ and $\mathbf{b} = (b_1, \ldots, b_p)^\top$ are vectors of lower and upper limits. Moreover, if $\mathbf{A}$ has the form

$$\mathbf{A} = \{(y_1, \ldots, y_p)^\top \in \mathbb{R}^p : \, a_1 \leqslant y_1 \leqslant b_1, \ldots, a_p \leqslant y_p \leqslant b_p\} = \{\mathbf{y} \in \mathbb{R}^p : \mathbf{a} \leqslant \mathbf{y} \leqslant \mathbf{b}\},$$

we say that the distribution of $\mathbf{Y}$ is doubly truncated. Analogously, we define $\{\mathbf{Y} \geqslant \mathbf{a}\}$ and $\{\mathbf{Y} \leqslant \mathbf{b}\}$, we say that the distribution of $\mathbf{Y}$ is truncated from below and truncated from above, respectively.

The pdf of $\mathbf{Y}$ is given by

$$f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, \nu; \mathbf{A}) = \frac{EST_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, \nu)}{\Pr(\mathbf{W} \in \mathbf{A})} \mathbb{I}(\mathbf{y} \in \mathbf{A}).$$

We use the notation $\mathbf{Y} \sim \text{TEST}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, \nu; \mathbf{A})$. For convenience, we also use the notation $\mathbf{Y} \sim \text{TEST}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, \nu; [\mathbf{a}, \mathbf{b}])$. For $\boldsymbol{\lambda} = \mathbf{0}$ and $\tau = 0$, $\mathbf{Y}$ follows a truncated Student-*t* distribution on $[\mathbf{a}, \mathbf{b}]$, which can be denoted by $\mathbf{Y} \sim Tt_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu; [\mathbf{a}, \mathbf{b}])$.

We also define the normalizing constant $\mathcal{P}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, \nu) = \Pr(\mathbf{W} \in \mathbf{A})$ as

$$\mathcal{P}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, \nu) = \int_{\mathbf{a}}^{\mathbf{b}} EST_p(\mathbf{w}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, \nu) \, \mathrm{d}\mathbf{w}. \tag{5.4}$$

For the ST distribution ($\tau = 0$), we simply omit the τ parameter, that is, $\mathcal{P}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu) = \mathcal{P}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, 0, \nu)$. If either all $\boldsymbol{\lambda}$ and τ are equal to zero, or $\tau \to +\infty$, we have the normalizing constant of the symmetric Student-*t* distribution as follows

$$\mathcal{P}_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{0}, 0, \nu) \equiv P_p(\mathbf{a}, \mathbf{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \int_{\mathbf{a}}^{\mathbf{b}} t_p(\mathbf{w}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) \, \mathrm{d}\mathbf{w}. \tag{5.5}$$

5.3 The ST censored regression model for multiple outcomes

This section is devoted to formulating the ST-CR model for multiple outcomes, which has asymmetric and heavy-tailed distribution on the error term. We also develop an EM-type algorithm to obtain the ML estimates and propose a method to approximate the standard errors of the estimates based on the empirical information matrix. Hence, let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ip})^\top$ be a $p \times 1$ response vector for the ith sample unit. The ST regression model is defined as

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\xi}_i, \quad i \in \{1, \ldots, n\}, \tag{5.6}$$

where $\mathbf{X}_i = \mathbf{I}_p \otimes \mathbf{x}_i^\top$ is the design matrix for the ith observation, having dimensions $p \times q$, and $\mathbf{x}_i^\top = (1, x_{i1}, \ldots, x_{i,q/p-1})$ is the transposed vector of covariates. Here, $\boldsymbol{\beta}$ represents the q-dimensional vector of population-average regression coefficients, and $\boldsymbol{\xi}_i$ denotes the $p \times 1$ error vector. It is assumed that $\boldsymbol{\xi}_i \overset{ind}{\sim} \mathrm{ST}_p(\kappa\boldsymbol{\Delta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$, where for $\nu > 1$, $\kappa = -(\nu/\pi)^{1/2}\Gamma((\nu-1)/2)/\Gamma(\nu/2)$, and $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}/(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}$. In addition, the chosen location parameter ensures that $\mathbb{E}[\boldsymbol{\xi}_i] = \mathbf{0}$, so that $\mathbb{E}[\mathbf{Y}_i] = \mathbf{X}_i\boldsymbol{\beta}$, for each $i = 1, \ldots, n$, and the regression parameters are all comparable. Thus, the distribution of $\mathbf{Y}_i$ is given by

$$\mathbf{Y}_i \overset{ind}{\sim} \mathrm{ST}_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu), \tag{5.7}$$

with location vector $\boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta} + \kappa\boldsymbol{\Delta}$, dispersion matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\alpha})$ depending on an unknown and reduced parameter vector $\boldsymbol{\alpha}_{\boldsymbol{\Sigma}}$, skewness parameter $\boldsymbol{\lambda}$, and degrees of freedom ν.

However, the response vector $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ip})^\top$ may not be fully observed due to censoring, so we define $(\mathbf{V}_i, \mathbf{C}_i)$ the observed data for the ith sample unit. Let $R_{ij} \subseteq \mathbb{R}$ denote the censoring region for observation (i, j), such that Y_{ij} is censored if $Y_{ij} \in R_{ij}$, for $j \in \{1, \ldots, p\}$. Further, let $\mathbf{V}_i = (V_{i1}, \ldots, V_{ip})^\top$ and $\mathbf{C}_i = (C_{i1}, \ldots, C_{ip})^\top$, where C_{ij} is the censoring indicator, and V_{ij} is given by

$$V_{ij} = \begin{cases} r_{ij}, & \text{if } Y_{ij} \in R_{ij}, \quad \text{(censored)} \\ Y_{ij}, & \text{otherwise}, \quad \text{(observed)} \end{cases} \tag{5.8}$$

where R_{ij} is an interval of the form $(-\infty, r_{ij})$, (r_{ij}, ∞), or (r_{ij1}, r_{ij2}) for left, right, or interval censoring, respectively. The constant $r_{ij} \in \mathbb{R}$ is equal to the detection limit for left and right censoring, and $r_{ij} = (r_{ij1} + r_{ij2})/2$ for interval censoring. Moreover, missing observations can be handled by setting $R_{ij} = (-\infty, \infty)$ and $r_{ij} = \mathrm{NA}$. The NA values serve as references and are excluded from the initial parameter estimation process, as detailed in Subsection 3.3.1. In this case, (5.7) along with (5.8) define the *skew-t interval-censored regression model for multiple outcomes* (hereafter, the ST-CR model).

5.3.1 The likelihood function

Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, where $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})^\top$ is a realization of $\mathbf{Y}_i \sim \mathrm{ST}_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$. To obtain the likelihood function of the ST-CR model, we first treat,

separately, the observed and censored components of $\mathbf{y}_i$, i.e., $\mathbf{y}_i = (\mathbf{y}_i^{o\top}, \mathbf{y}_i^{c\top})^\top$, where $C_{ik} = 0$ for all elements in the p_i^o-dimensional vector $\mathbf{y}_i^o$, and $C_{ik} = 1$ for all elements in the p_i^c-dimensional vector $\mathbf{y}_i^c$. According to that, we write $\mathbf{V}_i = (\mathbf{V}_i^{o\top}, \mathbf{V}_i^{c\top})^\top$, where $\mathbf{V}_i^c = (\mathbf{V}_{1i}^c, \mathbf{V}_{2i}^c)$ with

$$\boldsymbol{\mu}_i = (\boldsymbol{\mu}_i^{o\top}, \boldsymbol{\mu}_i^{c\top})^\top, \quad \boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\alpha}) = \begin{pmatrix} \boldsymbol{\Sigma}_i^{oo} & \boldsymbol{\Sigma}_i^{oc} \\ \boldsymbol{\Sigma}_i^{co} & \boldsymbol{\Sigma}_i^{cc} \end{pmatrix}, \quad \boldsymbol{\varphi}_i = (\boldsymbol{\varphi}_i^{o\top}, \boldsymbol{\varphi}_i^{c\top})^\top,$$

with $\boldsymbol{\varphi}_i = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda}$. To compute the likelihood function, we need to know the marginal and conditional distribution of an ST variable. Then, from Proposition 6, we have that

$$\mathbf{Y}_i^o \sim \mathrm{ST}_{p_i^o}(\boldsymbol{\mu}_i^o, \boldsymbol{\Sigma}_i^{oo}, \tilde{\boldsymbol{\lambda}}_i^o, \nu), \quad \mathbf{Y}_i^c \mid (\mathbf{Y}_i^o = \mathbf{y}_i^o) \sim \mathrm{EST}_{p_i^c}(\boldsymbol{\mu}_i^{co}, \tilde{\boldsymbol{\Sigma}}_i^{cc.o}, \boldsymbol{\lambda}_i^{co}, \tau_i^{co}, \nu_i^{co}),$$

with $\tilde{\boldsymbol{\lambda}}_i^o = c_i^{oc}\boldsymbol{\Sigma}_i^{oo\,1/2}\tilde{\boldsymbol{\varphi}}_i^o$, $c_i^{oc} = (1 + \boldsymbol{\varphi}_i^{c\top}\boldsymbol{\Sigma}_i^{cc.o}\boldsymbol{\varphi}_i^c)^{-1/2}$, $\boldsymbol{\mu}_i^{co} = \boldsymbol{\mu}_i^c + \boldsymbol{\Sigma}_i^{co}\boldsymbol{\Sigma}_i^{oo-1}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o)$, $\tilde{\boldsymbol{\Sigma}}_i^{cc.o} = \boldsymbol{\Sigma}_i^{cc.o}/\nu^2(\mathbf{y}_i^o)$, $\boldsymbol{\Sigma}_i^{cc.o} = \boldsymbol{\Sigma}_i^{cc} - \boldsymbol{\Sigma}_i^{co}(\boldsymbol{\Sigma}_i^{oo})^{-1}\boldsymbol{\Sigma}_i^{oc}$, $\nu^2(\mathbf{y}_i^o) = (\nu + p_i^o)/(\nu + \delta(\mathbf{y}_i^o))$, $\boldsymbol{\lambda}_i^{co} = \boldsymbol{\Sigma}_i^{cc.o\,1/2}\boldsymbol{\varphi}_i^c$, $\tau_i^{co} = \nu(\mathbf{y}_i^o)\tilde{\boldsymbol{\varphi}}_i^{o\top}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o)$, $\tilde{\boldsymbol{\varphi}}_i^o = \boldsymbol{\varphi}_i^o + \boldsymbol{\Sigma}_i^{oo\,-1}\boldsymbol{\Sigma}_i^{oc}\boldsymbol{\varphi}_i^c$, and $\nu_i^{co} = \nu + p_i^o$.

Let $\mathbf{V} = \mathrm{vec}(\mathbf{V}_1, \ldots, \mathbf{V}_n)$ and $\mathbf{C} = \mathrm{vec}(\mathbf{C}_1, \ldots, \mathbf{C}_n)$ denote the observed data, with $\mathrm{vec}(\mathbf{A})$ representing the operator that transforms a matrix into a column vector by vertically stacking the columns of the matrix. Therefore, the log-likelihood function of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}_\Sigma^\top, \boldsymbol{\lambda}^\top, \nu)^\top$, where $\boldsymbol{\alpha}_\Sigma$ denotes a minimal set of parameters such that $\boldsymbol{\Sigma}(\boldsymbol{\alpha})$ is well defined (e.g., the upper triangular elements of $\boldsymbol{\Sigma}$ in the unstructured case), for the observed data $(\mathbf{V}, \mathbf{C})$ is

$$\ell(\boldsymbol{\theta} \mid \mathbf{V}, \mathbf{C}) = \sum_{i=1}^{n} \ln L_i, \tag{5.9}$$

where L_i represents the likelihood function of $\boldsymbol{\theta}$ for the ith sample, given by

$$\begin{aligned} L_i \equiv L_i(\boldsymbol{\theta} \mid \mathbf{V}_i, \mathbf{C}_i) &= \Pr(\mathbf{r}_{1i}^c < \mathbf{y}_i^c < \mathbf{r}_{2i}^c \mid \mathbf{y}_i^o, \boldsymbol{\theta}) f(\mathbf{y}_i^o \mid \boldsymbol{\theta}) \\ &= \mathcal{P}_{p_i^c}(\mathbf{r}_{1i}^c, \mathbf{r}_{2i}^c; \boldsymbol{\mu}_i^{co}, \tilde{\boldsymbol{\Sigma}}_i^{cc.o}, \boldsymbol{\lambda}_i^{co}, \tau_i^{co}, \nu_i^{co})\, ST_{p_i^o}(\mathbf{y}_i^o; \boldsymbol{\mu}_i^o, \boldsymbol{\Sigma}_i^{oo}, \tilde{\boldsymbol{\lambda}}_i^o, \nu), \end{aligned}$$

where $\{\mathbf{y}_i^c \in \mathbb{R}^{p_i^c} : \mathbf{r}_{i1}^c \leqslant \mathbf{y}_i^c \leqslant \mathbf{r}_{i2}^c\} = \{(y_{i1}^c, \ldots, y_{ip_i^c}^c)^\top \in \mathbb{R}^{p_i^c} : r_{i11} \leqslant y_{i1}^c \leqslant r_{i12}^c, \ldots, r_{ip_i^c 1} \leqslant y_{ip_i^c}^c \leqslant r_{ip_i^c 2}^c\}$ denotes the censoring region of all partially observed data for sample unit i and $\mathcal{P}_r(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, \nu)$ denotes the integral defined in (5.4), which can be easily evaluated by using the R package `MomTrunc`.

Then, to estimate the parameters of the multivariate ST-CR model, an alternative is to maximize its log-likelihood function directly. However, this procedure can be quite cumbersome. In the next section, we propose a simple EM-type algorithm (Dempster *et al.*, 1977) to obtain the ML estimates. The EM algorithm is a general iterative method of ML estimation for incomplete data.

5.3.2 Parameter estimation via EM-type algorithm

This section describes how to carry out ML estimation for the ST-CR model. Initially proposed by Dempster *et al.* (1977), the EM algorithm is a popular iterative

optimization strategy commonly used to obtain ML estimates for incomplete data problems. This algorithm has many attractive features, such as numerical stability, implementation simplicity, and quite reasonable memory requirements (McLachlan & Krishnan, 2008). However, ML estimation for the ST-CR model is complicated because of the censoring, while the EM algorithm is less advisable due to the computational difficulty at the M-step. To overcome this problem, we use an extension of the EM called the ECME algorithm (Liu & Rubin, 1994). The ECME replaces the M-step with a sequence of conditional maximization (CM) steps. A key feature of this algorithm is that it preserves the stability of the EM and has a typically faster convergence rate than the original EM.

To propose the ECME algorithm for our ST-CR model, we first consider the marginal stochastic representation of a multivariate ST random vector given in Proposition 3. The model defined in (5.6) and (5.7) can be written hierarchically as

$$\begin{aligned}\mathbf{Y}_i \mid (U_i = u_i, T_i = t_i) &\overset{ind}{\sim} N_p(\mathbf{X}_i\boldsymbol{\beta} + t_i\boldsymbol{\Delta}, u_i^{-1}\boldsymbol{\Gamma}), \quad T_i|U_i = u_i \overset{ind}{\sim} \mathrm{TN}(\kappa, u_i^{-1}; (\kappa, \infty)),\\ U_i &\overset{iid}{\sim} Gamma(\nu/2, \nu/2), \quad i \in \{1, \ldots, n\},\end{aligned}$$

with $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}/\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}$ and $\boldsymbol{\Gamma} = \boldsymbol{\Sigma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top$. In the sequel, we define $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{u} = (u_1, \ldots, u_n)^\top$, and $\mathbf{t} = (t_1, \ldots, t_n)^\top$ such that $(\mathbf{y}, \mathbf{u}, \mathbf{t})$ are hypothetical missing data, and augmenting with the observed data $\mathbf{V}, \mathbf{C}$ corresponding to the censoring mechanism. Consequently, we set the complete data vector as $\mathbf{y}_c = (\mathbf{V}, \mathbf{C}, \mathbf{y}, \mathbf{u}, \mathbf{t})$. Then, fixing the value of ν, the complete data log-likelihood function of an equivalent set of parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\Delta}^\top, \boldsymbol{\alpha}_\Gamma^\top)^\top$, where $\boldsymbol{\alpha}_\Gamma = \mathrm{vech}(\boldsymbol{\Gamma})$ denotes the column vector obtained by vectorizing only the lower triangular part of $\boldsymbol{\Gamma}$, is given by $\ell_c(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_{ic}(\boldsymbol{\theta})$, where the individual complete data log-likelihood is as follows

$$\ell_{ic}(\boldsymbol{\theta}) \;=\; -\frac{1}{2}\Big\{ \ln|\boldsymbol{\Gamma}| + u_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - t_i\boldsymbol{\Delta})^\top\boldsymbol{\Gamma}^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - t_i\boldsymbol{\Delta})\Big\} + K_i(\nu) + c,$$

with c being a constant that does not depend on $\boldsymbol{\theta}$, $K_i(\nu)$ being a function that depends only on ν, and $|\mathbf{A}|$ denoting the determinant of the square matrix $\mathbf{A}$. Subsequently, the EM algorithm for the ST-CR model can be summarized as follows:

E-step: Given the current estimate $\widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\boldsymbol{\beta}}^{(k)\top}, \widehat{\boldsymbol{\Delta}}^{(k)\top}, \widehat{\boldsymbol{\alpha}}_\Gamma^{(k)\top})^\top$ at the kth step of the algorithm, the E-step provides the conditional expectation of the complete data log-likelihood function, i.e.,

$$Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) = \mathbb{E}\Big[\ell_c(\boldsymbol{\theta}) \mid \mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}\Big] = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}),$$

where

$$\begin{aligned}Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) \;\propto\; &-\frac{1}{2}\ln|\boldsymbol{\Gamma}| - \frac{1}{2}\mathrm{tr}\Big[\Big\{\widehat{u\mathbf{y}_i^2}^{(k)} + \widehat{u}_i^{(k)}\mathbf{X}_i\boldsymbol{\beta}\boldsymbol{\beta}^\top\mathbf{X}_i^\top + \widehat{ut_i^2}^{(k)}\boldsymbol{\Delta}\boldsymbol{\Delta}^\top - 2\widehat{u\mathbf{y}}_i^{(k)}\boldsymbol{\beta}^\top\mathbf{X}_i^\top\\ &-2\widehat{ut\mathbf{y}}_i^{(k)}\boldsymbol{\Delta}^\top + 2\widehat{ut}_i^{(k)}\boldsymbol{\Delta}\boldsymbol{\beta}^\top\mathbf{X}_i^\top\Big\}\boldsymbol{\Gamma}^{-1}\Big],\end{aligned}$$

with $\widehat{u\mathbf{y}_i^r}^{(k)} = \mathbb{E}_{U_iT_i\mathbf{Y}_i}[U_i\mathbf{Y}_i^r \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}]$, $\widehat{ut_i^r}^{(k)} = \mathbb{E}_{U_iT_i\mathbf{Y}_i}[U_iT_i^r \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}]$ (for $r \in \{1, 2\}$ with $\mathbf{Y}_i^1 = \mathbf{Y}_i$ and $\mathbf{Y}_i^2 = \mathbf{Y}_i\mathbf{Y}_i^\top$), $\widehat{ut\mathbf{y}_i}^{(k)} = \mathbb{E}_{U_iT_i\mathbf{Y}_i}[U_iT_i\mathbf{Y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}]$, and $\widehat{u}_i^{(k)} = \mathbb{E}_{U_iT_i\mathbf{Y}_i}[U_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}]$.

M-step: Conditionally maximizing $Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)})$ with respect to each entry of $\boldsymbol{\theta}$, we update the estimate $\widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\boldsymbol{\beta}}^{(k)\top}, \widehat{\boldsymbol{\Delta}}^{(k)\top}, \widehat{\boldsymbol{\alpha}}_{\Gamma}^{(k)\top})^\top$ by

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \left\{\sum_{i=1}^{n} \widehat{u}_i^{(k)}\mathbf{X}_i^\top\widehat{\boldsymbol{\Gamma}}^{-1(k)}\mathbf{X}_i\right\}^{-1} \sum_{i=1}^{n} \mathbf{X}_i^\top\widehat{\boldsymbol{\Gamma}}^{-1(k)}\left\{\widehat{u\mathbf{y}_i}^{(k)} - \widehat{ut_i}^{(k)}\widehat{\boldsymbol{\Delta}}^{(k)}\right\}, \tag{5.10}$$

$$\widehat{\boldsymbol{\Delta}}^{(k+1)} = \left\{\sum_{i=1}^{n} \widehat{ut_i^2}^{(k)}\right\}^{-1} \sum_{i=1}^{n}\left\{\widehat{ut\mathbf{y}_i}^{(k)} - \widehat{ut_i}^{(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k+1)}\right\}, \tag{5.11}$$

$$\widehat{\boldsymbol{\Gamma}}^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{u\mathbf{y}_i^2}^{(k)} + \widehat{u}_i^{(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k+1)}\widehat{\boldsymbol{\beta}}^{(k+1)\top}\mathbf{X}_i^\top + \widehat{ut_i^2}^{(k)}\widehat{\boldsymbol{\Delta}}^{(k+1)}\widehat{\boldsymbol{\Delta}}^{(k+1)\top} + \widehat{\mathbf{A}}_i^{(k)}\right\}, \tag{5.12}$$

where $\widehat{\mathbf{A}}_i^{(k)} = \widehat{\mathbf{A}}_{1i}^{(k)} - \widehat{\mathbf{A}}_{2i}^{(k)} - \widehat{\mathbf{A}}_{3i}^{(k)}$, with $\widehat{\mathbf{A}}_{1i}^{(k)} = \widehat{ut_i}^{(k)}(\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k+1)}\widehat{\boldsymbol{\Delta}}^{(k+1)\top} + \widehat{\boldsymbol{\Delta}}^{(k+1)}\widehat{\boldsymbol{\beta}}^{(k+1)\top}\mathbf{X}_i^\top)$, $\widehat{\mathbf{A}}_{2i}^{(k)} = \widehat{u\mathbf{y}_i}^{(k)}\widehat{\boldsymbol{\beta}}^{(k+1)\top}\mathbf{X}_i^\top + \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k+1)}\widehat{u\mathbf{y}_i}^{(k)\top}$, and $\widehat{\mathbf{A}}_{3i}^{(k)} = \widehat{ut\mathbf{y}_i}^{(k)}\widehat{\boldsymbol{\Delta}}^{(k+1)\top} + \widehat{\boldsymbol{\Delta}}^{(k+1)}\widehat{ut\mathbf{y}_i}^{(k)\top}$.

Then we update the parameter ν by maximizing the marginal log-likelihood function with respect to ν, as follows

$$\widehat{\nu}^{(k+1)} = \underset{\nu}{\operatorname{argmax}} \sum_{i=1}^{n} \ln f(\mathbf{V}_i \mid \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k+1)}; \nu).$$

We employ the algorithm proposed by Brent (2013), a combination of golden section search and successive parabolic interpolation, to perform the maximization procedure. It was designed for use with continuous functions of one variable.

The EM algorithm is iterated until a suitable convergence rule is satisfied. Once converged, we can recover $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\lambda}}$ using the expressions

$$\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Gamma}} + \widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{\Delta}}^\top, \quad \widehat{\boldsymbol{\lambda}} = (1 - \widehat{\boldsymbol{\Delta}}^\top\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\Delta}})^{-1/2}\widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\boldsymbol{\Delta}}.$$

It is important to stress that, from (5.10) to (5.12), the E-step reduces to the computation of $\widehat{u}_i^{(k)}$, $\widehat{u\mathbf{y}_i}^{(k)}$, $\widehat{u\mathbf{y}_i^2}^{(k)}$, $\widehat{ut_i}^{(k)}$, $\widehat{ut_i^2}^{(k)}$, and $\widehat{ut\mathbf{y}_i}^{(k)}$. The computations of these conditional expectations are discussed next.

5.3.3 Details for the expectations in the EM algorithm

To compute the expected values needed in the E-step, first note that for any multiplicative separable measurable function of U_i, T_i, and $\mathbf{Y}_i$, such that $g(U_i, T_i, \mathbf{Y}_i) = g_1(\mathbf{Y}_i)g_2(U_i)g_3(T_i)$, we have that

$$\begin{aligned}\mathbb{E}_{U_iT_i\mathbf{Y}_i}[g(U_i, T_i, \mathbf{Y}_i) \mid \mathbf{V}_i, \mathbf{C}_i] &= \mathbb{E}_{\mathbf{Y}_i}[g_1(\mathbf{Y}_i)\,\mathbb{E}_{U_iT_i}[g_2(U_i)g_3(T_i)|\mathbf{Y}_i] \mid \mathbf{V}_i, \mathbf{C}_i] \\ &= \mathbb{E}_{\mathbf{Y}_i}[g_1(\mathbf{Y}_i)\,\mathbb{E}_{U_i}[g_2(U_i)\,\mathbb{E}_{T_i}[g_3(T_i)|U_i, \mathbf{Y}_i] \mid \mathbf{Y}_i] \mid \mathbf{V}_i, \mathbf{C}_i].\end{aligned}$$

Hence,

$$\widehat{u\mathbf{y}_i^r} = \mathbb{E}_{U_iT_i\mathbf{Y}_i}[U_i\mathbf{Y}_i^r|\mathbf{V}_i,\mathbf{C}_i] = \mathbb{E}_{\mathbf{Y}_i}[\mathbf{Y}_i^r\mathbb{E}_{U_i}[U_i|\mathbf{Y}_i] \mid \mathbf{V}_i,\mathbf{C}_i],$$
$$\widehat{ut_i^r} = \mathbb{E}_{U_iT_i\mathbf{Y}_i}[U_iT_i^r|\mathbf{V}_i,\mathbf{C}_i] = \mathbb{E}_{\mathbf{Y}_i}[\mathbb{E}_{U_iT_i}[U_iT_i^r|\mathbf{Y}_i] \mid \mathbf{V}_i,\mathbf{C}_i],$$
$$\widehat{ut\mathbf{y}_i^r} = \mathbb{E}_{U_iT_i\mathbf{Y}_i}[U_iT_i\mathbf{Y}_i^r|\mathbf{V}_i,\mathbf{C}_i] = \mathbb{E}_{\mathbf{Y}_i}[\mathbf{Y}_i^r\mathbb{E}_{U_iT_i}[U_iT_i|\mathbf{Y}_i] \mid \mathbf{V}_i,\mathbf{C}_i],$$

for $r \in \{0,1,2\}$. From Cabral *et al.* (2012), we know that

$$\mathbb{E}_{U_iT_i}[U_iT_i|\mathbf{Y}_i] = (\kappa+\zeta_i)\,\mathbb{E}_{U_i}[U_i|\mathbf{Y}_i] + \varrho\,\phi(\boldsymbol{\theta},\mathbf{Y}_i), \tag{5.13}$$
$$\mathbb{E}_{U_iT_i}[U_iT_i^2|\mathbf{Y}_i] = \varrho^2 + (\kappa+\zeta_i)^2\,\mathbb{E}_{U_i}[U_i|\mathbf{Y}_i] + \varrho(2\kappa+\zeta_i)\,\phi(\boldsymbol{\theta},\mathbf{Y}_i), \tag{5.14}$$

with $\varrho = (1+\boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta})^{-1/2}$, $\zeta_i = \varrho^2\boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}(\mathbf{Y}_i-\boldsymbol{\mu}_i)$, and

$$\phi(\boldsymbol{\theta},\mathbf{Y}_i) = \mathbb{E}_{U_i}\left[U_i^{1/2}\frac{\phi_1(U_i^{1/2}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}_i-\boldsymbol{\mu}_i))}{\Phi_1(U_i^{1/2}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}_i-\boldsymbol{\mu}_i))}\middle|\mathbf{Y}_i\right].$$

As noted, both expectations $\mathbb{E}_{U_iT_i}[U_iT_i|\mathbf{Y}_i]$ and $\mathbb{E}_{U_iT_i}[U_iT_i^2|\mathbf{Y}_i]$ depend on $\mathbb{E}_{U_i}[U_i|\mathbf{Y}_i]$ and $\phi(\boldsymbol{\theta},\mathbf{Y}_i)$. From Lachos *et al.* (2010), we have also that

$$\mathbb{E}_{U_i}[U_i|\mathbf{Y}_i] = \frac{2\nu^2(\mathbf{y}_i)\,t_p(\mathbf{y}_i;\boldsymbol{\mu}_i,\boldsymbol{\Sigma},\nu)}{ST_p(\mathbf{y}_i;\boldsymbol{\mu}_i,\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)}T\left(\sqrt{\frac{\nu+p+2}{\nu+\delta_i}}A_i;\nu+p+2\right)$$

and

$$\phi(\boldsymbol{\theta},\mathbf{y}_i) = \frac{2\,t_p(\mathbf{y}_i;\boldsymbol{\mu}_i,\boldsymbol{\Sigma},\nu)}{ST_p(\mathbf{y}_i;\boldsymbol{\mu}_i,\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)}\frac{\Gamma((\nu+p+1)/2)}{\sqrt{\pi}\Gamma((\nu+p)/2)}\frac{(\nu+\delta_i)^{(\nu+p)/2}}{(\nu+\delta_i+A_i^2)^{(\nu+p+1)/2}},$$

where $\nu^2(\mathbf{y}_i) = (\nu+p)/(\nu+\delta_i)$, $\delta_i = \delta(\mathbf{y}_i;\boldsymbol{\mu}_i,\boldsymbol{\Sigma})$, and $A_i = \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_i-\boldsymbol{\mu}_i)$.

By using the fact that $t_p(\mathbf{y}_i;\boldsymbol{\mu}_i,\boldsymbol{\Sigma},\nu) = t_p(\mathbf{y}_i;\boldsymbol{\mu}_i,\frac{\nu}{\nu+2}\boldsymbol{\Sigma},\nu+2)/\nu^2(\mathbf{y}_i)$, $|\boldsymbol{\Sigma}| = (1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})|\boldsymbol{\Gamma}|$, $\delta_i = \frac{\nu}{\nu+2}\delta(\mathbf{y}_i;\boldsymbol{\mu}_i,\frac{\nu}{\nu+2}\boldsymbol{\Sigma})$, and $\delta_i + A_i^2 = \frac{\nu}{\nu+1}\delta(\mathbf{y}_i;\boldsymbol{\mu}_i,\frac{\nu}{\nu+1}\boldsymbol{\Gamma})$, we can easily propose simplified versions of the equations above after straightforward algebraic manipulations as follows

$$\mathbb{E}_{U_i}[U_i|\mathbf{Y}_i] = \frac{ST_p(\mathbf{y}_i;\boldsymbol{\mu}_i,\frac{\nu}{\nu+2}\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu+2)}{ST_p(\mathbf{y}_i;\boldsymbol{\mu}_i,\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)} \tag{5.15}$$

and

$$\phi(\boldsymbol{\theta},\mathbf{y}_i) = \frac{2}{\sqrt{\pi\nu(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\frac{t_p(\mathbf{y}_i;\boldsymbol{\mu}_i,\frac{\nu}{\nu+1}\boldsymbol{\Gamma},\nu+1)}{ST_p(\mathbf{y}_i;\boldsymbol{\mu}_i,\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)}. \tag{5.16}$$

Now, let us define the expectation of interest $\widehat{\phi\mathbf{y}_i^r} = \mathbb{E}_{\mathbf{Y}_i}[\mathbf{Y}_i^r\phi(\boldsymbol{\theta},\mathbf{Y}_i)|\mathbf{V}_i,\mathbf{C}_i]$, for $r \in \{0,1,2\}$. Next, we present two crucial propositions to compute these expectations, proofs can be found in the Appendix D.1.

Proposition 7. *Let* $\boldsymbol{Y} \sim \mathrm{ST}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)$. *For any measurable function* $g(\boldsymbol{Y})$, *it holds that*

$$\mathbb{E}\left[\phi(\boldsymbol{\theta},\boldsymbol{Y})g(\boldsymbol{Y}) \mid \boldsymbol{\alpha} \leqslant \boldsymbol{Y} \leqslant \boldsymbol{\beta}\right] = \frac{2}{\sqrt{\pi\nu(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\frac{P_p(\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\mu},\frac{\nu}{\nu+1}\boldsymbol{\Gamma},\nu+1)}{\mathcal{P}_p(\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)}\mathbb{E}[g(\boldsymbol{W}_1)], \tag{5.17}$$

and

$$\mathbb{E}[U\, g(\boldsymbol{Y}) \mid \boldsymbol{\alpha} \leqslant \boldsymbol{Y} \leqslant \boldsymbol{\beta}] = \frac{\mathcal{P}_p(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{\mu}, \frac{\nu}{\nu+2}\boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu+2)}{\mathcal{P}_p(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)} \mathbb{E}[g(\boldsymbol{W}_2)], \tag{5.18}$$

where $\boldsymbol{W}_1 \sim Tt_p(\boldsymbol{\mu}, \frac{\nu}{\nu+1}\boldsymbol{\Gamma}, \nu+1; (\boldsymbol{\alpha}, \boldsymbol{\beta}))$, $\boldsymbol{W}_2 \sim \mathrm{TST}_p(\boldsymbol{\mu}, \frac{\nu}{\nu+2}\boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu+2; (\boldsymbol{\alpha}, \boldsymbol{\beta}))$, $\boldsymbol{\Gamma} = \boldsymbol{\Sigma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top$, *and* $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}/(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}$.

Proposition 8. *Let* $\boldsymbol{Y} \sim \mathrm{ST}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$ *be partitioned as* $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, \boldsymbol{Y}_2^\top)^\top$ *of dimensions* p_1 *and* p_2 $(p_1 + p_2 = p)$*, respectively. Let*

$$\boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{\Gamma}_{11} & \boldsymbol{\Gamma}_{12} \\ \boldsymbol{\Gamma}_{21} & \boldsymbol{\Gamma}_{22} \end{pmatrix}, \quad \boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top, \boldsymbol{\alpha}_2^\top)^\top, \quad \boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$$

be the corresponding partitions of $\boldsymbol{\Gamma}$, $\boldsymbol{\alpha}$, *and* $\boldsymbol{\beta}$*. For a multiplicative separable measurable function* g*, it follows that*

$$\begin{aligned}\mathbb{E}[\phi(\boldsymbol{\theta}, \boldsymbol{Y}) g(\boldsymbol{Y}) |\, \boldsymbol{Y}_1, \boldsymbol{\alpha}_2 \leqslant \boldsymbol{Y}_2 \leqslant \boldsymbol{\beta}_2] &= g_1(\boldsymbol{y}_1) \frac{t_{p_1}(\boldsymbol{y}_1; \boldsymbol{\mu}_1, \frac{\nu}{\nu+1}\boldsymbol{\Gamma}_{11}, \nu+1)}{ST_{p_1}(\boldsymbol{y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \tilde{\boldsymbol{\lambda}}_1, \nu)} \frac{2}{\sqrt{\pi\nu(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \\ &\quad \times \frac{P_{p_2}(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2; \boldsymbol{\mu}_{2.1}, \frac{\nu_{2.1}}{\nu_{2.1}+1}\tilde{\boldsymbol{\Gamma}}_{22.1}, \nu_{2.1}+1)}{\mathcal{P}_{p_2}(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2; \boldsymbol{\mu}_{2.1}, \tilde{\boldsymbol{\Sigma}}_{22.1}, \boldsymbol{\lambda}_{2.1}, \tau_{2.1}, \nu_{2.1})} \mathbb{E}[g_2(\boldsymbol{W}_1^*)],\end{aligned} \tag{5.19}$$

and

$$\begin{aligned}\mathbb{E}[U\, g(\boldsymbol{Y}) |\, \boldsymbol{Y}_1, \boldsymbol{\alpha}_2 \leqslant \boldsymbol{Y}_2 \leqslant \boldsymbol{\beta}_2] &= g_1(\boldsymbol{y}_1) \frac{ST_{p_1}(\boldsymbol{y}_1; \boldsymbol{\mu}_1, \frac{\nu}{\nu+2}\boldsymbol{\Sigma}_{11}, \tilde{\boldsymbol{\lambda}}_1, \nu+2)}{ST_{p_1}(\boldsymbol{y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \tilde{\boldsymbol{\lambda}}_1, \nu)} \\ &\quad \times \frac{\mathcal{P}_{p_2}(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2; \boldsymbol{\mu}_{2.1}, \frac{\nu_{2.1}}{\nu_{2.1}+2}\tilde{\boldsymbol{\Sigma}}_{22.1}, \boldsymbol{\lambda}_{2.1}, \sqrt{\frac{\nu_{2.1}+2}{\nu_{2.1}}}\tau_{2.1}, \nu_{2.1}+2)}{\mathcal{P}_{p_2}(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2; \boldsymbol{\mu}_{2.1}, \tilde{\boldsymbol{\Sigma}}_{22.1}, \boldsymbol{\lambda}_{2.1}, \tau_{2.1}, \nu_{2.1})} \\ &\quad \times \mathbb{E}[g_2(\boldsymbol{W}_2^*)],\end{aligned} \tag{5.20}$$

where $g(\boldsymbol{Y}) = g_1(\boldsymbol{Y}_1) g_2(\boldsymbol{Y}_2)$, $\boldsymbol{W}_1^* \sim Tt_{p_2}(\boldsymbol{\mu}_{2.1}, \frac{\nu_{2.1}}{\nu_{2.1}+1}\tilde{\boldsymbol{\Gamma}}_{22.1}, \nu_{2.1}+1; (\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2))$, *and* $\boldsymbol{W}_2^* \sim \mathrm{TEST}_{p_2}(\boldsymbol{\mu}_{2.1}, \frac{\nu_{2.1}}{\nu_{2.1}+2}\tilde{\boldsymbol{\Sigma}}_{22.1}, \boldsymbol{\lambda}_{2.1}, \sqrt{(\nu_{2.1}+2)/\nu_{2.1}}\tau_{2.1}, \nu_{2.1}+2; (\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2))$, *with* $\nu_{2.1} = \nu + p_1$, $\tilde{\boldsymbol{\Gamma}}_{22.1} = (\boldsymbol{\Gamma}_{22} - \boldsymbol{\Gamma}_{21}\boldsymbol{\Gamma}_{11}^{-1}\boldsymbol{\Gamma}_{12})/\upsilon^2(\boldsymbol{y}_1)$, $\upsilon^2(\boldsymbol{y}_1) = (\nu+p_1)/(\nu+\delta(\boldsymbol{y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Gamma}_{11}))$, $\tau_{2.1} = \nu(\boldsymbol{y}_1)(\tilde{\boldsymbol{\varphi}}_1^\top(\boldsymbol{y}_1 - \boldsymbol{\mu}_1))$, *and the remaining parameters as in Proposition 6.*

Subsequently, according to expressions (5.13)–(5.20), we have the implementable expressions to the conditional expectations under the following three possible scenarios:

1. If the ith subject has only non-censored components, $\mathbb{E}_{U_i T_i \mathbf{Y}_i}[\mathbf{Y}_i^r | \mathbf{V}_i, \mathbf{C}_i] = \mathbf{y}_i^r$; then

$$\widehat{u\mathbf{y}_i^r}^{(k)} = \widehat{u}_i^{(k)} \mathbf{y}_i^r, \quad \widehat{u}_i^{(k)} = \mathbb{E}_{U_i}[U_i | \mathbf{Y}_i, \widehat{\boldsymbol{\theta}}^{(k)}], \quad \widehat{ut_i^r}^{(k)} = \mathbb{E}_{U_i T_i}[U_i T_i^r | \mathbf{Y}_i, \widehat{\boldsymbol{\theta}}^{(k)}],$$
$$\widehat{ut\mathbf{y}_i^r}^{(k)} = \mathbf{y}_i^r \mathbb{E}_{U_i T_i}[U_i T_i | \mathbf{Y}_i, \widehat{\boldsymbol{\theta}}^{(k)}],$$

where $\mathbf{y}_i^0 = 1$, $\mathbf{y}_i^1 = \mathbf{y}_i$, and $\mathbf{y}_i^2 = \mathbf{y}_i \mathbf{y}_i^\top$.

2. If the ith subject has only censored components, we have

$$\begin{aligned}
\widehat{u\mathbf{y}_i^r}^{(k)} &= \widehat{u}_i^{(k)}\widehat{\mathbf{w}_{2i}^r}^{(k)},\\
\widehat{u}_i^{(k)} &= \frac{\mathcal{P}_p(\mathbf{v}_{1i},\mathbf{v}_{2i};\hat{\boldsymbol{\mu}}_i^{(k)},\frac{\hat{\nu}^{(k)}}{\hat{\nu}^{(k)}+2}\hat{\boldsymbol{\Sigma}}^{(k)},\hat{\boldsymbol{\lambda}}^{(k)},\hat{\nu}^{(k)}+2)}{\mathcal{P}_p(\mathbf{v}_{1i},\mathbf{v}_{2i};\hat{\boldsymbol{\mu}}_i^{(k)},\hat{\boldsymbol{\Sigma}}^{(k)},\hat{\boldsymbol{\lambda}}^{(k)},\hat{\nu}^{(k)})},\\
\widehat{ut_i}^{(k)} &= \widehat{\kappa}^{(k)}\widehat{u}_i^{(k)} + \hat{\varrho}^{2(k)}\hat{\boldsymbol{\Delta}}^{(k)\top}\hat{\boldsymbol{\Gamma}}^{-1(k)}\left(\widehat{u\mathbf{y}_i}^{(k)} - \widehat{u}_i^{(k)}\hat{\boldsymbol{\mu}}_i^{(k)}\right) + \hat{\varrho}^{(k)}\widehat{\phi\mathbf{y}_i^0}^{(k)},\\
\widehat{ut_i^2}^{(k)} &= \hat{\varrho}^{2(k)} + \widehat{\varrho}^{2(k)}\hat{\boldsymbol{\Delta}}^{(k)\top}\hat{\boldsymbol{\Gamma}}^{-1(k)}\Big[2\widehat{\kappa}^{(k)}(\widehat{u\mathbf{y}_i}^{(k)} - \widehat{u}_i^{(k)}\hat{\boldsymbol{\mu}}_i^{(k)}) + \hat{\varrho}^{(k)}(\widehat{\phi\mathbf{y}_i^1}^{(k)} - \hat{\boldsymbol{\mu}}_i^{(k)}\widehat{\phi\mathbf{y}_i^0}^{(k)})\\
&\quad + \widehat{\varrho}^{2(k)}(\widehat{u\mathbf{y}_i^2}^{(k)} - 2\widehat{u\mathbf{y}_i}^{(k)}\hat{\boldsymbol{\mu}}_i^{\top(k)} + \widehat{u}_i^{(k)}\hat{\boldsymbol{\mu}}_i^{(k)}\hat{\boldsymbol{\mu}}_i^{\top(k)})\hat{\boldsymbol{\Gamma}}^{-1(k)}\hat{\boldsymbol{\Delta}}^{(k)}\Big]\\
&\quad + \widehat{\kappa}^{2(k)}\widehat{u}_i^{(k)} + 2\widehat{\kappa}^{(k)}\widehat{\varrho}^{(k)}\widehat{\phi\mathbf{y}_i^0}^{(k)},\\
\widehat{ut\mathbf{y}_i}^{(k)} &= \widehat{\kappa}^{(k)}\widehat{u\mathbf{y}_i}^{(k)} + \hat{\varrho}^{2(k)}(\widehat{u\mathbf{y}_i^2}^{(k)} - \widehat{u\mathbf{y}_i}^{(k)}\hat{\boldsymbol{\mu}}_i^{(k)\top})\hat{\boldsymbol{\Gamma}}^{-1(k)}\hat{\boldsymbol{\Delta}}^{(k)} + \hat{\varrho}^{(k)}\widehat{\phi\mathbf{y}_i^1}^{(k)},
\end{aligned}$$

with

$$\widehat{\phi\mathbf{y}_i^r}^{(k)} = \frac{2}{\sqrt{\pi\hat{\nu}^{(k)}(1+\hat{\boldsymbol{\lambda}}^{(k)\top}\hat{\boldsymbol{\lambda}}^{(k)})}}\frac{\Gamma(\frac{\hat{\nu}^{(k)}+1}{2})}{\Gamma(\frac{\hat{\nu}^{(k)}}{2})}\frac{P_p(\mathbf{v}_{1i},\mathbf{v}_{2i};\hat{\boldsymbol{\mu}}_i^{(k)},\frac{\hat{\nu}^{(k)}}{\hat{\nu}^{(k)}+1}\hat{\boldsymbol{\Gamma}}^{(k)},\hat{\nu}^{(k)}+1)}{\mathcal{P}_p(\mathbf{v}_{1i},\mathbf{v}_{2i};\hat{\boldsymbol{\mu}}_i^{(k)},\hat{\boldsymbol{\Sigma}}^{(k)},\hat{\boldsymbol{\lambda}}^{(k)},\hat{\nu}^{(k)})}\widehat{\mathbf{w}}_{1i}^{r(k)},$$

where $\widehat{\mathbf{w}}_{si}^{(k)} = \mathbb{E}[\mathbf{W}_{si} \mid \widehat{\boldsymbol{\theta}}^{(k)}]$ and $\widehat{\mathbf{w}}_{si}^{2(k)} = \mathbb{E}[\mathbf{W}_{si}\mathbf{W}_{si}^\top \mid \widehat{\boldsymbol{\theta}}^{(k)}]$ for $s \in \{1,2\}$, with

$$\mathbf{W}_{1i} \sim Tt_{p_i}(\boldsymbol{\mu}_i, \tfrac{\nu}{\nu+1}\boldsymbol{\Gamma}, \nu+1; (\mathbf{v}_{1i},\mathbf{v}_{2i})) \quad \text{and} \quad \mathbf{W}_{2i} \sim \mathrm{TST}_{p_i}(\boldsymbol{\mu}_i, \tfrac{\nu}{\nu+2}\boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu+2; (\mathbf{v}_{1i},\mathbf{v}_{2i})).$$

3. If the ith subject has both censored and uncensored components and given that the following processes $(\mathbf{Y}_i \,|\, \mathbf{V}_i, \mathbf{C}_i)$ and $(\mathbf{Y}_i \,|\, \mathbf{V}_i, \mathbf{C}_i, \mathbf{Y}_i^o)$are equivalent, we have

$$\begin{aligned}
\widehat{u\mathbf{y}_i}^{(k)} &= \mathbb{E}[U_i\mathbf{Y}_i \mid \mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \widehat{u}_i^{(k)}\,\mathrm{vec}(\mathbf{y}_i^o, \widehat{\mathbf{w}}_{2i}^{c(k)}),\\
\widehat{u\mathbf{y}_i^2}^{(k)} &= \mathbb{E}[U_i\mathbf{Y}_i\mathbf{Y}_i^\top \mid \mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \begin{pmatrix} \widehat{u}_i^{(k)}\mathbf{y}_i^o\mathbf{y}_i^{o\top} & \widehat{u}_i^{(k)}\mathbf{y}_i^o\widehat{\mathbf{w}}_{2i}^{c(k)\top} \\ \widehat{u}_i^{(k)}\widehat{\mathbf{w}}_{2i}^{c(k)}\mathbf{y}_i^{o\top} & \widehat{u}_i^{(k)}\widehat{\mathbf{w}}_{2i}^{2c(k)} \end{pmatrix},\\
\widehat{u}_i^{(k)} &= \mathbb{E}[U_i \mid \mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \frac{ST_{p_i^o}(\mathbf{y}_i^o; \hat{\boldsymbol{\mu}}_i^{o(k)}, \frac{\hat{\nu}^{(k)}}{\hat{\nu}^{(k)}+2}\hat{\boldsymbol{\Sigma}}_i^{oo(k)}, \tilde{\boldsymbol{\lambda}}_i^{o(k)}, \hat{\nu}^{(k)}+2)}{ST_{p_i^o}(\mathbf{y}_i^o; \hat{\boldsymbol{\mu}}_i^{o(k)}, \hat{\boldsymbol{\Sigma}}_i^{oo(k)}, \tilde{\boldsymbol{\lambda}}_i^{o(k)}, \hat{\nu}^{(k)})}\\
&\quad \times \frac{\mathcal{P}_{p_i^c}(\mathbf{v}_{1i}^c, \mathbf{v}_{2i}^c; \widehat{\boldsymbol{\mu}}_i^{co(k)}, \frac{\hat{\nu}_i^{co(k)}}{\hat{\nu}_i^{co(k)}+2}\tilde{\boldsymbol{\Sigma}}_i^{cc.o(k)}, \hat{\boldsymbol{\lambda}}_i^{co(k)}, \sqrt{\frac{\hat{\nu}_i^{co(k)}+2}{\hat{\nu}_i^{co(k)}}}\hat{\tau}_i^{co(k)}, \hat{\nu}_i^{co(k)}+2)}{\mathcal{P}_{p_i^c}(\mathbf{v}_{1i}^c, \mathbf{v}_{2i}^c; \widehat{\boldsymbol{\mu}}_i^{co(k)}, \tilde{\boldsymbol{\Sigma}}_i^{cc.o(k)}, \hat{\boldsymbol{\lambda}}_i^{co(k)}, \hat{\tau}_i^{co(k)}, \hat{\nu}_i^{co(k)})},
\end{aligned}$$

with $\widehat{ut_i}^{(k)}$, $\widehat{ut_i^2}^{(k)}$, and $\widehat{ut\mathbf{y}_i}^{(k)}$ as in item 2, and

$$\begin{aligned}
\widehat{\phi\mathbf{y}_i^r}^{(k)} &= \frac{2}{\sqrt{\pi\hat{\nu}^{(k)}(1+\hat{\boldsymbol{\lambda}}^{(k)\top}\hat{\boldsymbol{\lambda}}^{(k)})}}\frac{\Gamma(\frac{\hat{\nu}^{(k)}+1}{2})}{\Gamma(\frac{\hat{\nu}^{(k)}}{2})}\frac{P_{p_i^c}(\mathbf{v}_{1i}^c, \mathbf{v}_{2i}^c; \widehat{\boldsymbol{\mu}}_i^{co(k)}, \frac{\hat{\nu}_i^{co(k)}}{\hat{\nu}_i^{co(k)}+1}\tilde{\boldsymbol{\Gamma}}_i^{cc.o(k)}, \hat{\nu}_i^{co(k)}+1)}{\mathcal{P}_{p_i^c}(\mathbf{v}_{1i}^c, \mathbf{v}_{2i}^c; \widehat{\boldsymbol{\mu}}_i^{co(k)}, \tilde{\boldsymbol{\Sigma}}_i^{cc.o(k)}, \hat{\boldsymbol{\lambda}}_i^{co(k)}, \hat{\tau}_i^{co(k)}, \hat{\nu}_i^{co(k)})}\\
&\quad \times \frac{t_{p_i^o}(\mathbf{y}_i^o; \hat{\boldsymbol{\mu}}_i^{o(k)}, \frac{\hat{\nu}^{(k)}}{\hat{\nu}^{(k)}+1}\hat{\boldsymbol{\Gamma}}_i^{oo(k)}, \hat{\nu}^{(k)}+1)}{ST_{p_i^o}(\mathbf{y}_i^o; \hat{\boldsymbol{\mu}}_i^{o(k)}, \hat{\boldsymbol{\Sigma}}_i^{oo(k)}, \tilde{\boldsymbol{\lambda}}_i^{o(k)}, \hat{\nu}^{(k)})}\widehat{\mathbf{w}}_{1i}^{r(k)},
\end{aligned}$$

where $\widehat{\mathbf{w}}_{si}^{(k)} = \mathbb{E}[\mathbf{W}_{si}^* \mid \widehat{\boldsymbol{\theta}}^{(k)}]$ and $\widehat{\mathbf{w}}_{si}^{2(k)} = \mathbb{E}[\mathbf{W}_{si}^* \mathbf{W}_{si}^{*\top} \mid \widehat{\boldsymbol{\theta}}^{(k)}]$ for $s \in \{1, 2\}$, with

$$\mathbf{W}_{1i}^* \sim Tt_{p_i^c}\left(\boldsymbol{\mu}_i^{co}, \tfrac{\nu_i^{co}}{\nu_i^{co}+1}\tilde{\boldsymbol{\Gamma}}_i^{cc.o}, \nu_i^{co}+1; (\mathbf{v}_{1i}^c, \mathbf{v}_{2i}^c)\right),$$
$$\mathbf{W}_{2i}^* \sim \text{TEST}_{p_i^c}\left(\boldsymbol{\mu}_i^{co}, \tfrac{\nu_i^{co}}{\nu_i^{co}+2}\tilde{\boldsymbol{\Sigma}}_i^{cc.o}, \boldsymbol{\lambda}_i^{co}, \sqrt{\frac{\nu_i^{co}+2}{\nu_i^{co}}}\tau_i^{co}, \nu_i^{co}+2; (\mathbf{v}_{1i}^c, \mathbf{v}_{2i}^c)\right),$$

with $\boldsymbol{\Gamma}_i$ being partitioned like $\boldsymbol{\Sigma}_i$, $\tau_i^{co} = \nu(\mathbf{y}_i^o)(\tilde{\boldsymbol{\varphi}}_i^{o\top}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o))$, $\nu_i^{co} = \nu + p_i^o$, $\tilde{\boldsymbol{\Gamma}}_i^{cc.o} = (\boldsymbol{\Gamma}_i^{cc} - \boldsymbol{\Gamma}_i^{co}\boldsymbol{\Gamma}_i^{oo-1}\boldsymbol{\Gamma}_i^{oc})/\upsilon^2(\mathbf{y}_i^o)$, and the remaining parameters as in Proposition 8. Superscripts (k) have been omitted for simplicity.

The computation of the truncated moments $\widehat{\mathbf{w}}_{si}^{(k)}$ and $\widehat{\mathbf{w}}_{si}^{2(k)}$ (for $s \in \{1, 2\}$), in items 1 to 3, is based on Galarza *et al.* (2022a), which is available through the `MomTrunc` and `relliptical` R packages.

5.3.4 Initial values and stopping criterion

A reasonable convergence is attained using least squares estimates for the initial values of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, that is,

$$\widehat{\boldsymbol{\beta}}^{(0)} = \left(\sum_{i=1}^n \mathbf{X}_i^\top \mathbf{X}_i\right)^{-1} \sum_{i=1}^n \mathbf{X}_i^\top \mathbf{y}_i, \quad \widehat{\boldsymbol{\Sigma}}^{(0)} = \frac{1}{n}\sum_{i=1}^n (\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(0)})(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(0)})^\top,$$

while for the skewness parameter $\boldsymbol{\lambda}$, we use the coefficient of sample skewness of the residuals $\mathbf{r}_i = \mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(0)}$, $i \in \{1, \ldots, n\}$.

We have adopted the stopping criterion $|\ell(\widehat{\boldsymbol{\theta}}^{(k+1)}|\mathbf{V}, \mathbf{C})/\ell(\widehat{\boldsymbol{\theta}}^{(k)}|\mathbf{V}, \mathbf{C}) - 1| < \epsilon$, for example, $\epsilon = 10^{-6}$, i.e., the algorithm stops when the relative distance between two successive evaluations of the log-likelihood defined in (5.9) is less than tolerance.

5.3.5 Standard error approximation

According to large sample theory, the asymptotic covariance of the ML estimates can be approximated by the empirical information matrix, which evaluated at the ML estimates is reduced to

$$\mathbf{I}_e(\widehat{\boldsymbol{\theta}}|\mathbf{y}) = \sum_{i=1}^n \widehat{\mathbf{s}}_i(\boldsymbol{\theta})\,\widehat{\mathbf{s}}_i(\boldsymbol{\theta})^\top.$$

For the ST-CR model, let $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\boldsymbol{\alpha}}_{\boldsymbol{\Sigma}}^\top, \widehat{\boldsymbol{\lambda}}^\top, \widehat{\nu})^\top$ denote the vector of ML estimates obtained at the last iteration of the EM algorithm, with $\widehat{\boldsymbol{\alpha}}_{\boldsymbol{\Sigma}} = (\widehat{\alpha}_1, \ldots, \widehat{\alpha}_{p(p+1)/2})^\top$ representing the $p(p+1)/2$ vector of distinct elements of $\widehat{\boldsymbol{\Sigma}}$, then the vector $\widehat{\mathbf{s}}_i(\boldsymbol{\theta}) = (\widehat{\mathbf{s}}_i(\boldsymbol{\beta})^\top, \widehat{\mathbf{s}}_i(\boldsymbol{\alpha}_{\boldsymbol{\Sigma}})^\top, \widehat{\mathbf{s}}_i(\boldsymbol{\lambda})^\top, \widehat{\mathbf{s}}_i(\nu))^\top$ has elements

$$\widehat{\mathbf{s}}_i(\boldsymbol{\beta}) = \mathbf{X}_i^\top \widehat{\boldsymbol{\Gamma}}^{-1}\left(\widehat{u\mathbf{y}}_i - \widehat{ut}_i\widehat{\boldsymbol{\Delta}} - \widehat{u}_i\mathbf{X}_i\widehat{\boldsymbol{\beta}}\right),$$

$$\widehat{\mathbf{s}}_i(\boldsymbol{\alpha}_{\boldsymbol{\Sigma}}) = \left(\widehat{\mathbf{s}}_i(\alpha_1), \widehat{\mathbf{s}}_i(\alpha_2), \ldots, \widehat{\mathbf{s}}_i(\alpha_{p(p+1)/2})\right)^\top, \quad \widehat{\mathbf{s}}_i(\boldsymbol{\lambda}) = \left(\widehat{\mathbf{s}}_i(\lambda_1), \widehat{\mathbf{s}}_i(\lambda_2), \ldots, \widehat{\mathbf{s}}_i(\lambda_p)\right)^\top,$$

$$\widehat{\mathbf{s}}_i(\nu) = \frac{1}{2}\left\{\ln\left(\frac{\widehat{\nu}}{2}\right) + 1 - \psi\left(\frac{\widehat{\nu}}{2}\right) + \mathbb{E}\left[\ln U_i | \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\right] - \widehat{u}_i\right\},$$

with $\psi(x) = \Gamma'(x)/\Gamma(x)$ representing the digamma function,

$$\begin{aligned}
\widehat{\mathbf{s}}_i(\alpha_l) &= -\frac{1}{2}\mathrm{tr}\left\{\widehat{\boldsymbol{\Gamma}}^{-1}\dot{\boldsymbol{\Gamma}}_{\alpha_l} + \widehat{\mathbf{A}}_i\dot{\boldsymbol{\Gamma}}^{-1}_{\alpha_l} + \left[\widehat{ut_i^2}\left(\dot{\boldsymbol{\Delta}}_{\alpha_l}\widehat{\boldsymbol{\Delta}}^\top + \widehat{\boldsymbol{\Delta}}\dot{\boldsymbol{\Delta}}^\top_{\alpha_l}\right) - 2\dot{\boldsymbol{\Delta}}_{\alpha_l}(\widehat{ut\mathbf{y}}_i - \widehat{ut}_i\mathbf{X}_i\widehat{\boldsymbol{\beta}})^\top\right]\widehat{\boldsymbol{\Gamma}}^{-1}\right\}, \\
\widehat{\mathbf{s}}_i(\lambda_j) &= -\frac{1}{2}\mathrm{tr}\left\{\widehat{\boldsymbol{\Gamma}}^{-1}\dot{\boldsymbol{\Gamma}}_{\lambda_j} + \widehat{\mathbf{A}}_i\dot{\boldsymbol{\Gamma}}^{-1}_{\lambda_j} + \left[\widehat{ut_i^2}\left(\dot{\boldsymbol{\Delta}}_{\lambda_j}\widehat{\boldsymbol{\Delta}}^\top + \widehat{\boldsymbol{\Delta}}\dot{\boldsymbol{\Delta}}^\top_{\lambda_j}\right) - 2\dot{\boldsymbol{\Delta}}_{\lambda_j}(\widehat{ut\mathbf{y}}_i - \widehat{ut}_i\mathbf{X}_i\widehat{\boldsymbol{\beta}})^\top\right]\widehat{\boldsymbol{\Gamma}}^{-1}\right\},
\end{aligned}$$

where

$$\widehat{\mathbf{A}}_i = \widehat{u\mathbf{y}_i^2} + (\widehat{u}_i\mathbf{X}_i\widehat{\boldsymbol{\beta}} - 2\widehat{u\mathbf{y}}_i + 2\widehat{ut}_i\widehat{\boldsymbol{\Delta}})\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_i^\top + \widehat{\boldsymbol{\Delta}}(\widehat{ut_i^2}\widehat{\boldsymbol{\Delta}} - 2\widehat{ut\mathbf{y}}_i)^\top, \quad \dot{\boldsymbol{\Gamma}}^{-1}_{\alpha_l} = -\widehat{\boldsymbol{\Gamma}}^{-1}\dot{\boldsymbol{\Gamma}}_{\alpha_l}\widehat{\boldsymbol{\Gamma}}^{-1},$$

$$\dot{\boldsymbol{\Gamma}}_{\alpha_l} = \frac{\partial\boldsymbol{\Gamma}}{\partial\alpha_l}\Big|_{\boldsymbol{\alpha}=\widehat{\boldsymbol{\alpha}}} = \dot{\boldsymbol{\Sigma}}_{\alpha_l} - \dot{\boldsymbol{\Delta}}_{\alpha_l}\widehat{\boldsymbol{\Delta}}^\top - \widehat{\boldsymbol{\Delta}}\dot{\boldsymbol{\Delta}}^\top_{\alpha_l}, \quad \dot{\boldsymbol{\Gamma}}^{-1}_{\lambda_j} = -\widehat{\boldsymbol{\Gamma}}^{-1}\dot{\boldsymbol{\Gamma}}_{\lambda_j}\widehat{\boldsymbol{\Gamma}}^{-1}, \quad \dot{\boldsymbol{\Gamma}}_{\lambda_j} = -\dot{\boldsymbol{\Delta}}_{\lambda_j}\widehat{\boldsymbol{\Delta}}^\top - \widehat{\boldsymbol{\Delta}}\dot{\boldsymbol{\Delta}}^\top_{\lambda_j},$$

$$\dot{\boldsymbol{\Delta}}_{\alpha_l} = \frac{\partial\boldsymbol{\Delta}}{\partial\alpha_l}\Big|_{\boldsymbol{\alpha}=\widehat{\boldsymbol{\alpha}}} = \frac{\dot{\boldsymbol{\Sigma}}^{1/2}_{\alpha_l}\widehat{\boldsymbol{\lambda}}}{(1+\widehat{\boldsymbol{\lambda}}^\top\widehat{\boldsymbol{\lambda}})^{1/2}}, \quad \dot{\boldsymbol{\Delta}}_{\lambda_j} = \frac{\partial\boldsymbol{\Delta}}{\partial\lambda_j}\Big|_{\boldsymbol{\lambda}=\widehat{\boldsymbol{\lambda}}} = \widehat{\boldsymbol{\Sigma}}^{1/2}\left(\frac{\dot{\boldsymbol{\lambda}}_j}{(1+\widehat{\boldsymbol{\lambda}}^\top\widehat{\boldsymbol{\lambda}})^{1/2}} - \frac{\widehat{\lambda}_j\widehat{\boldsymbol{\lambda}}}{(1+\widehat{\boldsymbol{\lambda}}^\top\widehat{\boldsymbol{\lambda}})^{3/2}}\right),$$

for $l \in \{1, \ldots, p(p+1)/2\}$, $j \in \{1, \ldots, p\}$, $i \in \{1, \ldots, n\}$. Additionally, $\dot{\boldsymbol{\lambda}}_j = \dfrac{\partial\boldsymbol{\lambda}}{\partial\lambda_j}$ denotes the $p \times 1$ vector with 1 in the jth entry and 0 otherwise, $\dot{\boldsymbol{\Sigma}}_{\alpha_l} = \dfrac{\partial\boldsymbol{\Sigma}}{\partial\alpha_l}$ represents the $p \times p$ matrix with 1 in the position of the element α_l and the remaining elements equal to 0, and $\dot{\boldsymbol{\Sigma}}^{1/2}_{\alpha_l} = \dfrac{\partial\boldsymbol{\Sigma}^{1/2}}{\partial\alpha_l}\Big|_{\boldsymbol{\alpha}=\widehat{\boldsymbol{\alpha}}}$ is computed by

$$\mathrm{vec}\left(\dot{\boldsymbol{\Sigma}}^{1/2}_{\alpha_l}\right) = \left(\boldsymbol{\Sigma}^{1/2} \oplus \boldsymbol{\Sigma}^{1/2}\right)^{-1}\mathrm{vec}\left(\dot{\boldsymbol{\Sigma}}_{\alpha_l}\right),$$

where $\boldsymbol{\Sigma}^{1/2} \oplus \boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma}^{1/2} \otimes \mathbf{I}_n + \mathbf{I}_n \otimes \boldsymbol{\Sigma}^{1/2}$ represents the Kronecker sum with $\otimes$ denoting the Kronecker product. Since $\boldsymbol{\Sigma}$ is positive definite, $\boldsymbol{\Sigma}^{1/2}$ is positive definite, and hence the Kronecker sum is positive definite.

It is important to stress that the standard error (SE) of ν depends on the calculation of $\mathbb{E}[\ln U_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}]$, which would rely on computationally intensive Monte Carlo integration or other numerical methods. Therefore, we focus solely on studying the SEs of the parameters $\boldsymbol{\beta}$, $\boldsymbol{\alpha}_{\boldsymbol{\Sigma}}$, and $\boldsymbol{\lambda}$.

5.3.6 Imputation of censored components

Let $\mathbf{y}_i^c$ be the true unobserved response vector for the censored components of the ith subject. Now, as a by-product of the EM algorithm, we can obtain the predictor of the censored components, denoted by $\widehat{\mathbf{y}}_i^c$, as follows

$$\widehat{\mathbf{y}}_i^c = \mathbb{E}\left[\mathbf{Y}_i^c \,|\, \mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\right], \tag{5.21}$$

which is obtained considering two possible cases:

1. If subject i has only censored components

$$\widehat{\mathbf{y}}_i^c = \mathbb{E}\left[\mathbf{Y}_i \,|\, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\right],$$

where $\mathbf{Y}_i \,|\, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}} \sim \text{TST}_p(\widehat{\boldsymbol{\mu}}_i, \widehat{\boldsymbol{\Sigma}}, \widehat{\boldsymbol{\lambda}}, \widehat{\nu}; \mathbf{A}_i)$, with $\widehat{\boldsymbol{\mu}}_i = \mathbf{X}_i\widehat{\boldsymbol{\beta}} + \kappa\widehat{\boldsymbol{\Delta}}$, $\mathbf{A}_i = \{\mathbf{y}_i \in \mathbb{R}^p : \mathbf{v}_{1i} < \mathbf{y}_i < \mathbf{v}_{2i}\}$, $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})^\top$, $\mathbf{v}_{1i} = (v_{1i1}, \ldots, v_{1ip})^\top$, and $\mathbf{v}_{2i} = (v_{2i1}, \ldots, v_{2ip})^\top$.

2. If subject i has both censored and uncensored components, then

$$\widehat{\mathbf{y}}_i^c = \mathbb{E}\left[\mathbf{Y}_i^c \,|\, \mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}\right],$$

where $\mathbf{Y}_i^c \,|\, \mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}} \sim \text{TEST}_{p_i^c}(\widehat{\boldsymbol{\mu}}_i^{co}, \widetilde{\boldsymbol{\Sigma}}_i^{cc.o}, \widehat{\boldsymbol{\lambda}}_i^{co}, \widehat{\tau}_i^{co}, \widehat{\nu}_i^{co}; \mathbf{A}_i)$, with $\mathbf{A}_i = \{\mathbf{y}_i^c \in \mathbb{R}^{p_i^c} : \mathbf{v}_{1i}^c < \mathbf{y}_i^c < \mathbf{v}_{2i}^c\}$, $\mathbf{y}_i^c = (y_{i1}, \ldots, y_{ip_i^c})^\top$, $\mathbf{v}_{1i}^c = (v_{1i1}^c, \ldots, v_{1ip_i^c}^c)^\top$, and $\mathbf{v}_{2i}^c = (v_{2i1}^c, \ldots, v_{2ip_i^c}^c)^\top$. The remaining parameters of the distribution are given as in Proposition 6.

5.4 Simulation Studies

This section is dedicated to examining the performance of the proposed methods. We present four simulation studies to investigate: i) the asymptotic behavior of the ML estimates for our proposal under different scenarios, ii) the effect of the skewness parameter $\boldsymbol{\lambda}$ in the estimation procedure for left-censored responses, iii) the impact on the parameter estimates of model misspecification, and iv) the robustness of the estimates when the data is perturbed.

In all simulation studies, we considered 500 Monte Carlo (MC) samples generated from the model $\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\xi}_i$, for $i \in \{1, \ldots, n\}$, where the errors follow a bivariate skew distribution. We set the parameters $\boldsymbol{\beta} = (1, -2, 2, -1)^\top$, the design matrix $\mathbf{X}_i = \mathbf{I}_2 \otimes \mathbf{x}_i^\top$, with $\mathbf{I}_2$ being the identity matrix of dimension 2×2, $\mathbf{x}_i^\top = (1, x_{i1})$, and x_{i1} being independent simulated from the standard normal distribution. The elements of $\boldsymbol{\Sigma}$ equal to $\sigma_{11} = 4$, $\sigma_{12} = -0.50$, and $\sigma_{22} = 1.50$. Moreover, we consider the skewness parameter vector $\boldsymbol{\lambda} = (2, -3)^\top$ for simulations I, III, and IV.

5.4.1 Simulation study I: Asymptotic properties

Aiming to provide empirical evidence about the consistency of the ML estimates obtained through the proposed method, we simulated samples of sizes $n = 50, 100, 200$, and 400 from the bivariate skew-t distribution, $\mathbf{Y}_i \overset{ind}{\sim} \text{ST}_2(\mathbf{X}_i\boldsymbol{\beta} + \kappa\boldsymbol{\Delta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$, with $\nu = 4$ degrees of freedom, κ, and $\boldsymbol{\Delta}$ as in Section 5.3. In addition, we study the effect of censoring on the parameter estimates. Thus, three scenarios were evaluated: without censoring, an average of 15% left-censored, and an average of 15% right-censored observations. The detection limits were set $\mathbf{v}_2^c = (-2, 0.50)^\top$ and $\mathbf{v}_1^c = (4.14, 3.69)^\top$ for the left and right-censored datasets, respectively. These values ensure almost the same censoring proportion

for each component of the response vector $\mathbf{Y} \in \mathbb{R}^{n\times 2}$. For each sample size and type of censoring, we computed the mean (MC-AV), median (MC-MD), and standard deviation (MC-SD) of the 500 MC estimates. To examine the consistency of the approximated method to get standard errors, we calculated the average of the standard error estimates (denoted by IM-SE). We compared it with the empirical MC-SD for each scenario.

Table 9 displays results obtained for data without censoring, where we can see that the mean (MC-AV) and median (MC-MD) of the estimates for the regression coefficients ($\beta_{10}, \beta_{11}, \beta_{20}$, and β_{21}) are close to the true parameter value for all sample sizes. Additionally, there is a bias in the estimates of $\sigma_{11}, \sigma_{12}, \sigma_{22}, \boldsymbol{\lambda}$, and ν, but it gets smaller as n increases. One can also observe that MC-AV overestimates $\boldsymbol{\lambda}$ and ν for samples of size $n = 50$, but the median was a better estimator. On the other hand, the estimation methods of the SE for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_{\boldsymbol{\Sigma}}$ provide results close to the empirical ones, and the closeness improves as the sample size increases. Besides, the inverse of the empirical information matrix was unreliable in estimating the SE of the skewness parameter $\boldsymbol{\lambda}$ for the two smallest sample sizes $n = 50$ and 100; however, very decent results were obtained for samples of size $n = 200$ and 400, as expected.

Table 9 – Simulation I. Summary statistics based on 500 MC samples of size $n = 50, 100, 200$, and 400 without censoring. MC-AV, MC-MD, and MC-SD refer to the mean, median, and standard deviation of the estimates, respectively. IM-SE denotes the average of standard errors obtained as described in Subsection 5.3.5.

n	Measure	β_{10}	β_{11}	β_{20}	β_{21}	σ_{11}	σ_{12}	σ_{22}	λ_1	λ_2	ν
		1.00	-2.00	2.00	-1.00	4.00	-0.50	1.50	2.00	-3.00	4.00
50	MC-AV	1.036	-2.033	2.009	-1.013	4.592	-0.410	1.613	18.354	-25.286	13.355
	MC-MD	1.013	-2.026	2.021	-1.018	4.123	-0.391	1.533	2.867	-3.952	4.728
	MC-SD	0.379	0.329	0.192	0.163	2.090	0.842	0.683	34.010	41.342	
	IM-SE	0.491	0.381	0.287	0.207	1.753	0.762	0.641	864.88	1147.8	
100	MC-AV	1.009	-2.003	2.014	-1.002	4.169	-0.438	1.507	3.432	-4.807	5.589
	MC-MD	0.989	-2.005	2.024	-1.002	3.942	-0.429	1.427	2.113	-3.064	4.182
	MC-SD	0.248	0.196	0.127	0.097	1.312	0.449	0.446	7.224	8.753	
	IM-SE	0.244	0.196	0.129	0.096	1.060	0.428	0.393	22.924	26.016	
200	MC-AV	0.996	-1.992	2.008	-0.994	4.033	-0.442	1.485	2.065	-3.069	4.345
	MC-MD	0.981	-1.996	2.004	-0.994	3.944	-0.417	1.455	1.937	-2.935	4.130
	MC-SD	0.174	0.135	0.095	0.066	0.790	0.301	0.303	0.783	0.923	
	IM-SE	0.167	0.142	0.089	0.068	0.727	0.289	0.276	0.788	0.970	
400	MC-AV	0.997	-2.011	2.007	-1.001	3.983	-0.452	1.474	1.930	-2.901	4.160
	MC-MD	0.995	-2.014	2.008	-1.002	3.947	-0.447	1.468	1.903	-2.861	4.046
	MC-SD	0.120	0.095	0.067	0.046	0.531	0.196	0.209	0.469	0.562	
	IM-SE	0.117	0.096	0.062	0.045	0.510	0.201	0.194	0.493	0.604	

To evaluate how the type of censoring affects the parameter estimates, we considered samples with an average of 15% of censored observations. Results are shown in Tables 10 and 11 for the left-censored and right-censored datasets, respectively. From these tables, we can see that the mean (MC-AV) and median (MC-MD) of the estimates for the regression coefficients $\boldsymbol{\beta}$ are close to the true parameter value for each type of censoring,

and this difference decreases as the sample size increases. Regarding the scale matrix elements for left-censored data, the method returned reasonable estimates for σ_{11}, but it had problems recovering σ_{12} and σ_{22}. It could happen since left-censored observations affect the lower tail of the distribution and then the retrieval of λ_2, the skewness parameter with the negative signal. On the other hand, for right-censored datasets, the proposed EM algorithm achieved MC-AV close to the true parameter value for σ_{22}. However, as expected, we got biased estimates for σ_{11} and σ_{12}. It happens since right-censored data affects the upper tail of the distribution and then the recovery of λ_1, the positive skewness parameter. Even though there is a bias in the estimates of σ_{11}, σ_{12}, and σ_{22}, it gets smaller as n increases, indicating the consistency of the EM estimates.

In both cases, MC-AV for ν is close to the true parameter value ($\nu = 4$) for samples greater than 100, while the median (MC-MD) was approximately 4 for all sample sizes. Additionally, notice that the mean of the SE of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_{\boldsymbol{\Sigma}}$, obtained through the inverse of the empirical information matrix, is close to the empirical standard deviation of the estimates (MC-SD) for all scenarios, indicating that the proposed method to obtain the standard error of these parameters is reliable. Once again, the variability of $\boldsymbol{\lambda}$ is more complex to be estimated for small sample sizes (see e.g., $n = 50$) compared to the location and scale parameter. Therefore, more caution is necessary when performing inference for $\boldsymbol{\lambda}$ under small sample sizes.

Table 10 – Simulation I. Results based on 500 MC samples of size $n = 50, 100, 200$, and 400 with an average 15% of left-censored observations.

n	Censoring Level	Measure	β_{10}	β_{11}	β_{20}	β_{21}	σ_{11}	σ_{12}	σ_{22}	λ_1	λ_2	ν
			1.00	-2.00	2.00	-1.00	4.00	-0.50	1.50	2.00	-3.00	4.00
50	14.4%	MC-AV	1.119	-2.030	2.054	-1.015	4.757	-0.147	1.160	7.241	-9.125	13.754
		MC-MD	1.091	-2.012	2.065	-1.016	4.190	-0.168	1.015	2.210	-1.763	4.143
		MC-SD	0.385	0.364	0.183	0.172	2.336	0.651	0.599	55.990	89.321	
		IM-SE	0.400	0.373	0.204	0.192	2.084	0.655	0.481	14.199	19.711	
100	15.5%	MC-AV	1.074	-1.996	2.057	-1.004	4.527	-0.309	1.155	2.395	-2.267	6.691
		MC-MD	1.046	-2.011	2.061	-1.004	4.238	-0.294	1.071	2.156	-1.900	4.321
		MC-SD	0.262	0.212	0.128	0.099	1.589	0.394	0.397	1.221	1.634	
		IM-SE	0.257	0.213	0.128	0.102	1.327	0.390	0.328	1.435	1.416	
200	16.3%	MC-AV	1.082	-1.972	2.070	-0.996	4.543	-0.315	1.068	2.124	-1.719	4.695
		MC-MD	1.065	-1.982	2.067	-0.999	4.386	-0.302	1.026	2.073	-1.613	4.233
		MC-SD	0.188	0.157	0.101	0.070	1.040	0.272	0.258	0.614	0.731	
		IM-SE	0.180	0.153	0.089	0.076	0.953	0.251	0.212	0.759	0.654	
400	14.9%	MC-AV	1.068	-1.998	2.063	-1.002	4.527	-0.392	1.128	2.060	-1.816	4.332
		MC-MD	1.064	-1.999	2.059	-1.002	4.472	-0.397	1.109	2.053	-1.789	4.166
		MC-SD	0.126	0.101	0.069	0.047	0.721	0.184	0.196	0.415	0.507	
		IM-SE	0.127	0.101	0.062	0.050	0.655	0.176	0.158	0.476	0.425	

Figure 46 (of Appendix D.2.1) presents the boxplot of the estimates obtained through the ST-CR model by sample size and censoring, where the red line indicates the real parameter value. In most cases, we observe that the median is close to the true parameter value, and there are some outliers for σ_{11}, σ_{12}, and σ_{22}.

Finally, we analyzed the mean square error (MSE) of the regression coefficients $\boldsymbol{\beta}$ and the scale matrix elements $\boldsymbol{\Sigma}$ estimated from the ST-CR model for sample sizes

Table 11 – Simulation I. Results based on 500 MC samples of size $n = 50, 100, 200$, and 400 with an average 15% of right-censored observations.

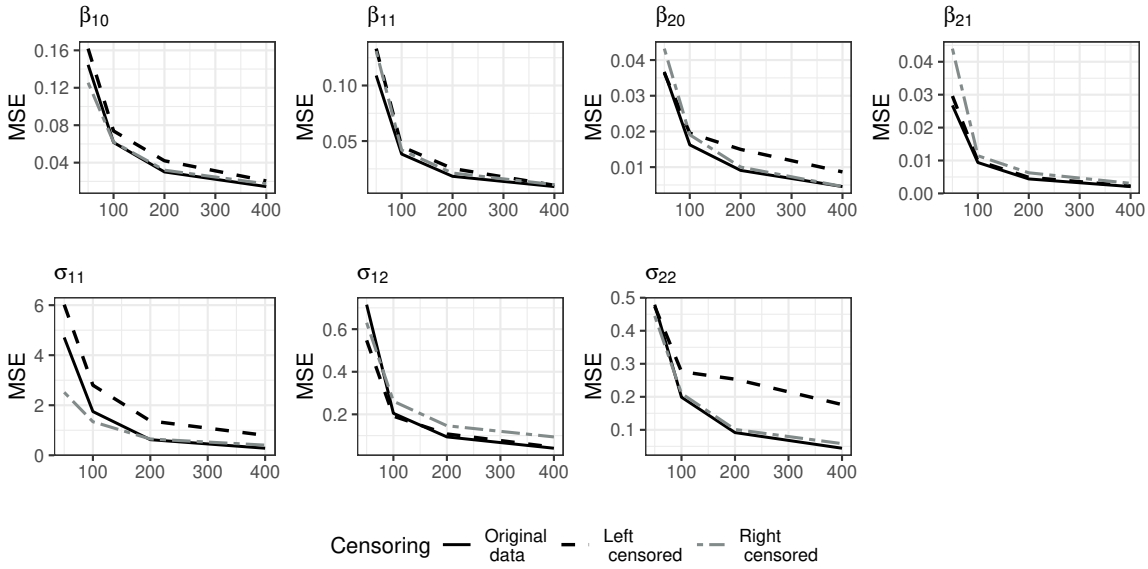
n	Censoring Level	Measure	β_{10}	β_{11}	β_{20}	β_{21}	σ_{11}	σ_{12}	σ_{22}	λ_1	λ_2	ν
			1.00	-2.00	2.00	-1.00	4.00	-0.50	1.50	2.00	-3.00	4.00
50	12.0%	MC-AV	0.989	-2.025	1.989	-1.024	3.800	-0.141	1.510	2.915	-5.466	13.993
		MC-MD	0.991	-2.012	2.008	-1.007	3.512	-0.163	1.409	1.581	-3.306	4.138
		MC-SD	0.354	0.361	0.208	0.209	1.575	0.709	0.668	5.150	7.335	
		IM-SE	0.388	0.411	0.204	0.210	1.600	0.703	0.601	7.354	10.394	
100	15.1%	MC-AV	0.965	-2.007	1.998	-1.001	3.635	-0.215	1.484	1.541	-3.109	6.126
		MC-MD	0.949	-1.997	2.013	-0.999	3.492	-0.210	1.402	1.337	-2.792	4.105
		MC-SD	0.247	0.205	0.138	0.107	1.100	0.425	0.458	1.262	1.820	
		IM-SE	0.249	0.210	0.133	0.107	0.962	0.437	0.420	1.199	1.828	
200	12.1%	MC-AV	0.947	-1.995	1.994	-0.992	3.550	-0.246	1.494	1.404	-2.922	4.367
		MC-MD	0.948	-1.996	1.993	-0.991	3.459	-0.238	1.463	1.316	-2.771	4.017
		MC-SD	0.171	0.146	0.100	0.079	0.670	0.286	0.318	0.655	0.905	
		IM-SE	0.171	0.158	0.093	0.077	0.622	0.290	0.294	0.655	1.009	
400	15.5%	MC-AV	0.943	-2.014	1.995	-1.001	3.546	-0.264	1.511	1.313	-2.756	4.181
		MC-MD	0.942	-2.006	1.996	-0.999	3.525	-0.275	1.503	1.288	-2.727	4.065
		MC-SD	0.119	0.103	0.068	0.055	0.447	0.195	0.241	0.408	0.527	
		IM-SE	0.120	0.105	0.065	0.053	0.447	0.206	0.212	0.401	0.619	

$n = 50, 100, 200$, and 400. The idea is to provide empirical evidence about the consistency of the ML estimates. The MSE measure is defined by

$$\text{MSE}_i = \frac{1}{n}\sum_{j=1}^{n}\left(\widehat{\theta}_i^{(j)} - \theta_i\right)^2, \tag{5.22}$$

where $\widehat{\theta}_i^{(j)}$ is the ML estimate of parameter θ_i in the jth sample for $\boldsymbol{\theta} = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}, \sigma_{11}, \sigma_{12}, \sigma_{22})^\top$. Figure 22 shows that the MSE of the parameter estimates tends to zero as the sample size increases, providing empirical evidence about the consistency of the ML estimates of the ST-CR model for the three evaluated scenarios.

Figure 22 – Simulation I. MSE of parameter estimates under the ST-CR model, based on 500 Monte Carlo samples simulated from the skew-t distribution considering different sample sizes and types of censoring.



5.4.2 Simulation study II: Effect of skewness on left-censored data

This simulation study seeks to assess the effect of the skewness parameter $\boldsymbol{\lambda}$ in the estimation process for datasets with an average of 15% of left-censored responses. Hence, we simulated 500 MC datasets of size $n = 300$ from the bivariate ST-CR model with the same parameter values described above, and the following four scenarios for $\boldsymbol{\lambda}$ and the detection limit $\mathbf{v}_2^c$:

i) $\boldsymbol{\lambda} = (2, -3)^\top$, $\mathbf{v}_2^c = (-2.090, 0.490)^\top$;

ii) $\boldsymbol{\lambda} = (2, 3)^\top$, $\mathbf{v}_2^c = (-2.230, 0.290)^\top$;

iii) $\boldsymbol{\lambda} = (-2, 3)^\top$, $\mathbf{v}_2^c = (-1.960, 0.455)^\top$; and

iv) $\boldsymbol{\lambda} = (-2, -3)^\top$, $\mathbf{v}_2^c = (-2.220, 0.315)^\top$.

The values of $\mathbf{v}_2^c$ assure almost the same number of censored observations on each component of the bivariate response vector. For each case, we fitted the ST-CR model. The summary statistics are shown in Table 12, where MC-AV denotes the mean of the 500 MC estimates, and IM-SE represents the average of the 500 standard errors approximated from the empirical information matrix. From this table, we can see that the mean (MC-AV) of the estimates for the regression coefficients ($\beta_{10}, \beta_{11}, \beta_{20}$, and β_{21}) are close to the true parameter value independent of the skewness parameter.

Table 12 – Simulation II. Mean of the estimates (MC-AV) and average of the approximated standard errors (IM-SE) based on 500 MC samples of size 300 generated from the bivariate skew-t distribution with skewness parameter $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^\top$ and average 15% of left-censored observations.

Par.	$\lambda_1 = 2,\ \lambda_2 = -3$		$\lambda_1 = 2,\ \lambda_2 = 3$		$\lambda_1 = -2,\ \lambda_2 = 3$		$\lambda_1 = -2,\ \lambda_2 = -3$	
	MC-AV	IM-SE	MC-AV	IM-SE	MC-AV	IM-SE	MC-AV	IM-SE
β_{10}	1.064	0.145	1.200	0.203	1.070	0.139	1.081	0.155
β_{11}	-1.996	0.119	-1.956	0.115	-1.993	0.128	-2.016	0.129
β_{20}	2.065	0.071	2.103	0.111	2.007	0.076	2.005	0.085
β_{21}	-1.004	0.058	-0.968	0.061	-1.001	0.064	-1.015	0.065
σ_{11}	4.544	0.760	3.269	0.534	3.528	0.509	3.864	0.563
σ_{12}	-0.381	0.205	-0.478	0.176	-0.240	0.239	-0.842	0.252
σ_{22}	1.122	0.182	1.098	0.186	1.510	0.246	1.563	0.288
λ_1	2.122	0.580	2.761	0.815	-1.284	0.481	-1.030	0.511
λ_2	-1.870	0.521	3.620	0.934	2.798	0.763	-2.333	0.653
ν	4.404		2.473		4.239		4.723	

As demonstrated in Simulation study I for left-censored datasets, the algorithm has some difficulties recovering the skewness component of the negative signal since left-censored observations affect the lower tail of the distribution. This is also observed for scenarios i), iii), and iv), i.e., cases with at least one negative $\lambda_j, j \in \{1, 2\}$. For scenario ii),

where all components of the skewness parameter are positive, our method overestimated the true value of $\boldsymbol{\lambda}$. For this case, it is worth noting that the mean of the ν estimates is equal to 2.473, whereas, for the other three cases, it is close to the true value ($\nu = 4$).

Moreover, this simulation study was also performed considering different detection limits to generate left-censored datasets. The new limits guarantee that the first component of the response vector has approximately 70% of the desirable censoring rate. The results are displayed in Table 25 (see Appendix D.2.2), where we can notice that this configuration returned similar results as shown above.

5.4.3 Simulation study III: Model misspecification

To evaluate the performance of the proposed model and the impact of estimating with the wrong distribution, we simulated 500 MC datasets from the model $\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\xi}_i$, for $i \in \{1, \ldots, 300\}$, considering the same parameter values as above. For data generation, two scenarios were considered:

i) $\mathbf{Y}_i \overset{ind}{\sim} \mathrm{SN}_2(\mathbf{X}_i\boldsymbol{\beta} + \kappa_1\boldsymbol{\Delta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ and

ii) $\mathbf{Y}_i \overset{ind}{\sim} \mathrm{SSL}_2(\mathbf{X}_i\boldsymbol{\beta} + \kappa_2\boldsymbol{\Delta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$, with $\nu = 1.15$,

where $\mathrm{SN}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ and $\mathrm{SSL}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$ denote the bivariate skew-normal and skew-slash distribution, respectively, with $\kappa_1 = -\sqrt{\frac{2}{\pi}}$ and $\kappa_2 = -\sqrt{\frac{2}{\pi}}\left(\frac{\nu}{\nu - 1/2}\right)$.

For comparison, we considered two scenarios: one without censoring (original data) and another with an average of 15% right-censored observations. The detection limits were set $\mathbf{v}_1^c = (3.81, 3.40)^\top$ for the skew-normal and $\mathbf{v}_1^c = (4.46, 3.96)^\top$ for the skew-slash distribution, i.e., $y_{ij} = v_{1j}^c$ if $y_{ij} \geqslant v_{1j}^c$ and it keeps unchangeable otherwise, for $i \in \{1, \ldots, 300\}$ and $j \in \{1, 2\}$. We fitted the SN-CR and ST-CR models for each simulated dataset. The SN-CR model was previously studied by Galarza *et al.* (2022b). The model selection criteria AIC and BIC and the estimates of the model parameters were recorded at each simulation. Summary statistics are reported, such as the MC mean estimate (MC-AV), and the standard error approximated through the empirical information matrix (IM-SE).

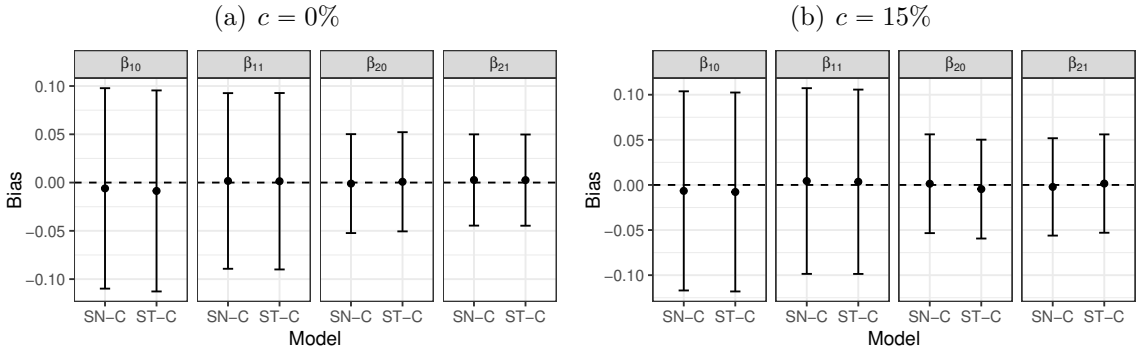
Table 13 shows the results for data simulated from the SN distribution by censoring proportion and model fitting. Notice that both models estimate the regression coefficients well (i.e., MC-AV is close to the true value) with approximately the same standard error. Note that the mean of the ν estimates are 80.610 and 65.383 for the original and censored data, respectively. It is worth mentioning that the scale matrix components are not comparable directly. Moreover, both models returned close values for the information criteria AIC and BIC. Figure 23 displays the bias of $\widehat{\beta}_{10}, \widehat{\beta}_{11}, \widehat{\beta}_{20}$, and $\widehat{\beta}_{21}$

for both censoring proportions when estimating with SN-CR and ST-CR models, where we can see that both distributions seem to fit the data equally well.

Table 13 – Simulation III. Results based on 500 MC samples of size $n = 300$ when generating data from the SN distribution considering different censoring levels (c) and estimating the models SN-CR and ST-CR. MC-AV and IM-SE denote to the mean of the estimates and the average of the standard error, respectively.

Par.	$c = 0\%$				$c = 15\%$			
	SN-CR		ST-CR		SN-CR		ST-CR	
	MC-AV	IM-SE	MC-AV	IM-SE	MC-AV	IM-SE	MC-AV	IM-SE
β_{10}	0.994	0.103	0.991	0.103	0.993	0.107	0.992	0.108
β_{11}	-1.998	0.097	-1.999	0.097	-1.996	0.107	-1.996	0.109
β_{20}	1.999	0.051	2.001	0.051	2.001	0.054	1.995	0.054
β_{21}	-0.997	0.047	-0.997	0.046	-1.002	0.054	-0.998	0.054
σ_{11}	3.942	0.728	3.805	0.576	3.920	0.834	3.353	0.538
σ_{12}	-0.374	0.322	-0.395	0.228	-0.271	0.344	-0.150	0.267
σ_{22}	1.416	0.267	1.381	0.212	1.371	0.270	1.356	0.230
λ_1	1.771	1.426	1.835	0.652	1.476	0.951	1.080	0.599
λ_2	-2.728	1.387	-2.769	0.765	-2.482	1.070	-2.409	0.767
ν	-		80.610		-		65.383	
AIC	1937.2		1937.7		1792.4		1795.3	
BIC	1976.8		1981.6		1831.9		1839.3	

Figure 23 – Simulation III. Mean bias ± 1 standard deviation for the estimates of $\boldsymbol{\beta}$ obtained under the SN-CR and ST-CR models, based on 500 MC samples simulated from the SN distribution considering two cases: (a) without censoring and (b) average 15% of right-censored observations.



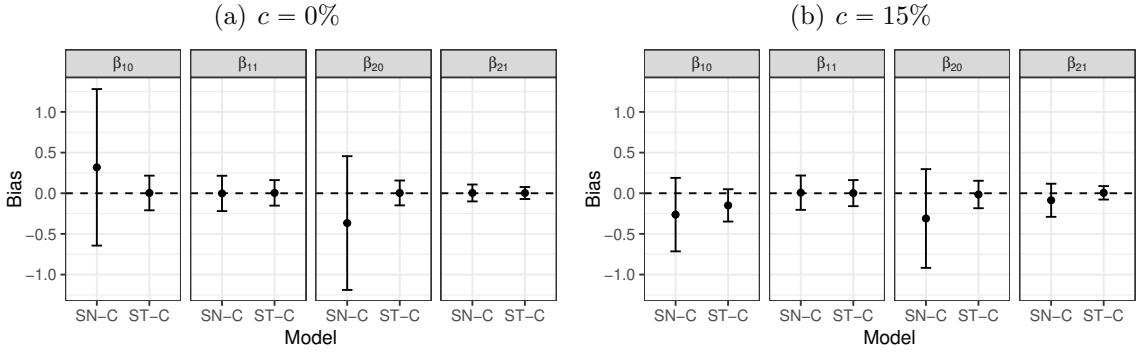
Finally, results for data generated from the skew-slash (SSL) distribution with $\nu = 1.15$ are displayed in Table 14. Notice that the mean (MC-AV) for the regression coefficients under the ST-CR model is close to the true value, except for β_{10} with an average of 15% left-censored observations, where the mean underestimates the true parameter value. On the other hand, the SN-CR model returned bias estimates for β_{10} and β_{20}, the intercepts. Observe that the mean of the standard errors (IM-SE) under the ST-CR model is smaller than the estimate under the SN-CR model. Based on information criteria AIC

and BIC, we conclude that the ST-CR model fits better with the simulated data. It is worth noting that both models recover the correct signal for the skewness parameters. For the ST-CR model, the mean of λ_1 and λ_2 are close to the true values in data without censoring (original), but the model underestimates λ_1 for censored datasets, as expected. It happens since right-censored observations affect the upper tail of the distribution and then the retrieval of the positive skewness parameter. Moreover, Figure 24 shows the bias of $\widehat{\beta}_{10}$, $\widehat{\beta}_{11}$, $\widehat{\beta}_{20}$, and $\widehat{\beta}_{21}$ for both censoring rates, where we see that estimating disregarding the heavy tail increases the variance of the estimator.

Table 14 – Simulation III. Summary statistics based on 500 MC samples of size $n = 300$ when generating data from the SSL distribution with $\nu = 1.15$ considering different censoring levels (c) and estimating the models SN-CR and ST-CR. MC-AV and IM-SE denote to the mean of the estimates and the average of the standard error, respectively.

Par.	$c = 0\%$				$c = 15\%$			
	SN-CR		ST-CR		SN-CR		ST-CR	
	MC-AV	IM-SE	MC-AV	IM-SE	MC-AV	IM-SE	MC-AV	IM-SE
β_{10}	1.319	0.300	1.003	0.197	0.737	0.205	0.851	0.196
β_{11}	-2.002	0.285	-1.995	0.152	-1.993	0.208	-1.998	0.161
β_{20}	1.633	0.165	2.004	0.108	1.690	0.132	1.985	0.111
β_{21}	-0.996	0.152	-0.997	0.071	-1.086	0.132	-0.994	0.078
σ_{11}	26.919	1.587	6.609	0.998	12.909	1.112	5.693	0.806
σ_{12}	-2.474	0.754	-0.792	0.398	1.356	0.933	-0.342	0.410
σ_{22}	11.297	0.627	2.503	0.390	12.838	0.632	2.643	0.432
λ_1	4.283	3.500	2.049	0.637	1.331	5.772	1.200	0.466
λ_2	-6.704	5.148	-3.137	0.816	-7.412	9.436	-3.051	0.840
ν	-		3.258		-		3.409	
AIC	2888.4		2617.8		2534.8		2380.3	
BIC	2927.9		2661.8		2574.4		2424.2	

Figure 24 – Simulation III. Mean bias ±1 standard deviation for the estimates of $\boldsymbol{\beta}$ obtained under the SN-CR and ST-CR models, based on 500 MC samples simulated from the SSL distribution with $\nu = 1.15$ considering two cases: (a) without censoring and (b) average 15% of right-censored observations.



5.4.4 Simulation study IV: Robustness of the estimators

This simulation aims to compare the performance of the estimates in the presence of outliers on the response variable. In this case, 500 MC samples of size $n = 300$ were simulated from a bivariate skew-normal distribution, $\mathbf{Y}_i \overset{ind}{\sim} \text{SN}_2(\mathbf{X}_i\boldsymbol{\beta} + \kappa\boldsymbol{\Delta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, with $\kappa = -\sqrt{2/\pi}$ and the remaining parameters described in the introduction of this section. After generating the data, each MC sample was perturbed considering the following scheme: the maximum value of each column of $\mathbf{Y} \in \mathbb{R}^{n\times 2}$ was increased in ϑ times the sample standard deviation, i.e., $y_{j,\text{pert}} = \max(\mathbf{y}_{\cdot j}) + \vartheta SD(\mathbf{y})$, for $\mathbf{y}_{\cdot j} = (y_{1j}, y_{2j}, \ldots, y_{300j})^\top$, $\vartheta \in \{0, 1, 2, 3, 4, 5, 6, 7\}$, and $j \in \{1, 2\}$. Furthermore, we considered two different levels of censoring: the first case corresponds to the case without censoring, and the second case considered $\mathbf{v}_2^c = (-1.75,\ 0.61)^\top$ as detection limits, i.e., $y_{ij} = v_{2j}^c$ if $y_{ij} \leqslant v_{2j}^c$ and it keeps unchangeable otherwise. The latter scenario implies that each dataset has around 15% of left-censored observations.

Figure 25 shows the MSE of the estimates of $\beta_{10}, \beta_{11}, \beta_{20}$, and β_{21} obtained after fitting the SN-CR and ST-CR models by levels of perturbation. These plots reveal, in most cases, that the MSE computed from the ST-CR is lower than the obtained from the SN-CR model for levels of perturbation greater and equal to two sample standard deviations (except for β_{10} in datasets with left-censored observations for $\vartheta \leqslant 4$).

Figure 25 – Simulation IV. MSE of $\boldsymbol{\beta}$ estimated from the SN-CR and ST-CR models, based on 500 MC samples of size $n = 300$ simulated from the bivariate SN distribution considering different perturbation levels and censoring proportions.

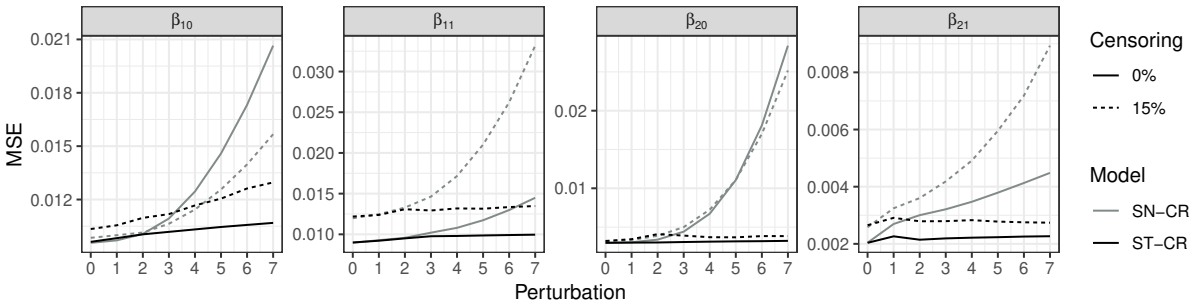
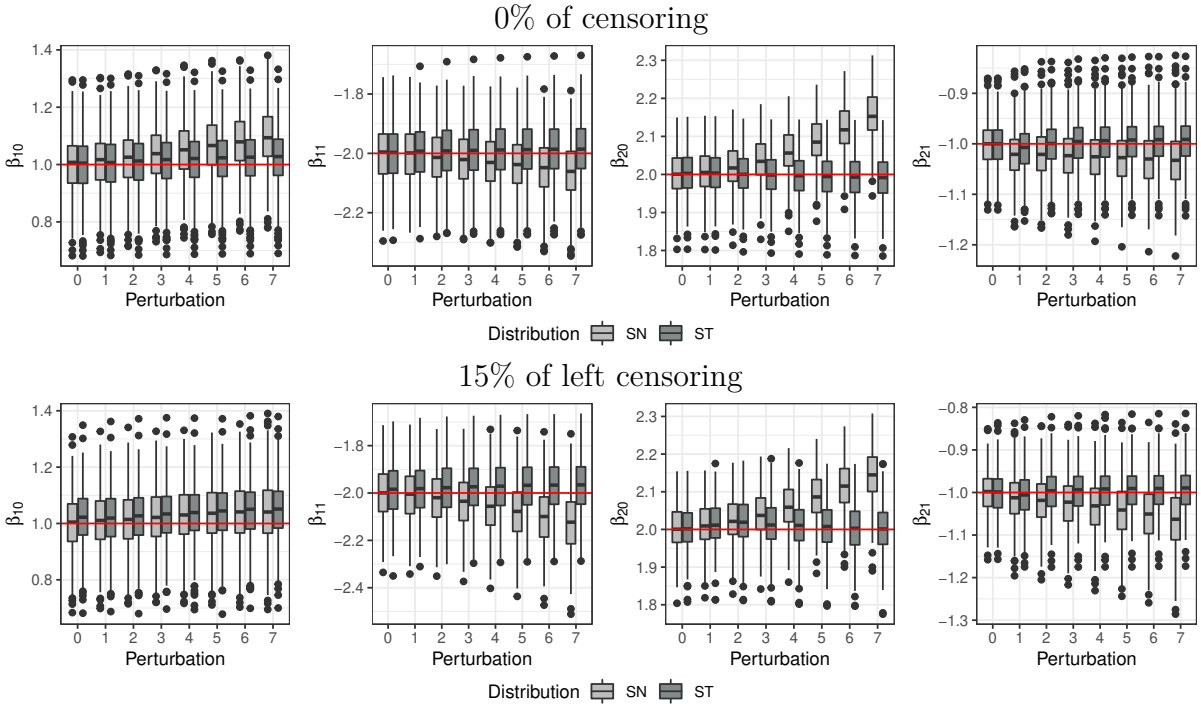


Figure 26 displays the boxplot for the estimates of $\beta_{10}, \beta_{11}, \beta_{20}$, and β_{21} obtained from the SN-CR and ST-CR model by level of perturbation and censoring type. Here we can observe that the median of the ST estimates is close to the true value (red line), while the difference between the median of the SN estimates and the true value increases along the perturbation level. Note also, for the SN, that the interquartile range increases with the level of perturbation for all scenarios.

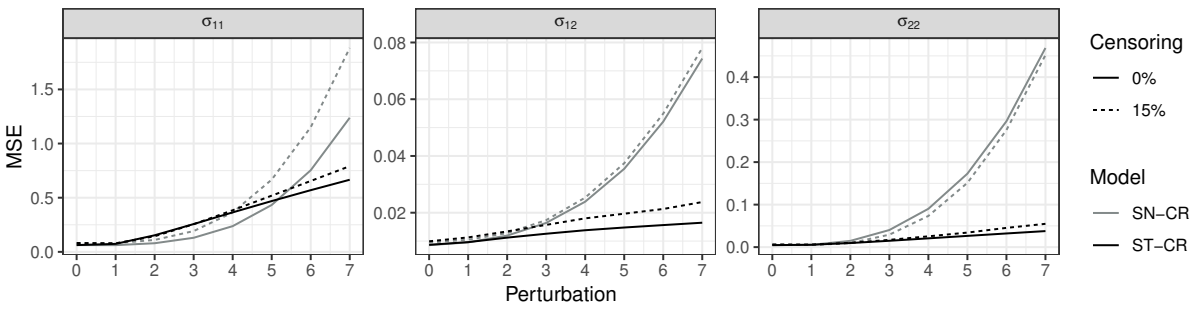
Yet, Figure 27 shows the MSE for the variance-covariance elements σ_{11}, σ_{12}, and σ_{22}, computed using expression (4) in Ferreira *et al.* (2016) for the SN-CR, and (5.3) for the ST-CR model. Here, we can note that for σ_{12} and σ_{22}, the MSE for the ST distribution is quite smaller than the SN one for $\vartheta \geqslant 2$. Regarding σ_{11}, we have a better

Figure 26 – Simulation IV. Boxplot of the $\boldsymbol{\beta}$ estimates obtained from the SN-CR and ST-CR model, based on 500 MC samples of size $n = 300$ simulated from the bivariate SN distribution considering two levels of censoring.



performance of the ST model for $\vartheta \geqslant 4$ when data is left-censored and for $\vartheta \geqslant 5$ under no censoring. It is worth mentioning that the variance-covariance elements depend on the skewness parameter, whose estimation relies on the censoring type (left, right, or interval). For instance, under right-censoring, σ_{11} and σ_{22} behavior would be the opposite. These results confirm that the heavy tails of the skew-t distribution allow our model to mitigate the effect of outliers, i.e., a much more robust method against atypical values.

Figure 27 – Simulation IV. MSE of the estimates for the variance-covariance elements obtained from the SN-CR and ST-CR model, based on 500 MC samples of size $n = 300$ simulated from the bivariate SN distribution, considering different levels of perturbation.



5.5 Applications

In order to show the usefulness of the method developed in this work, we analyze two datasets: a) the trace metals in freshwater streams across the commonwealth of Virginia and b) the stellar abundances dataset. The first has interval-censored observations, and the second has left-censored and missing observations.

5.5.1 Trace metals in freshwater streams across the Commonwealth of Virginia

Aiming to illustrate the performance of our proposal, we consider the trace metals concentration dataset described in Subsection 1.3.4, which was previously analyzed by Hoffman & Johnson (2015). They proposed a pseudo-likelihood approach for estimating parameters considering the multivariate normal and log-normal models, while recently, Galarza *et al.* (2022b) analyzed this data by fitting the multivariate SN-CR model.

It is important to note that the concentration levels of trace metals are strictly positive measures; to guarantee this, we consider an interval censoring approach. Specifically, we set the lower detection limits to zero for all trace metals, while the upper detection limits are variable-specific as detailed in Subsection 1.3.4. The analysis then proceeds with the assessment of the following model:

$$\mathbf{y}_i = \boldsymbol{\mu} + \boldsymbol{\xi}_i, \quad i \in \{1, \ldots, 184\}, \tag{5.23}$$

with $\mathbf{y}_i = (y_{i1}, y_{i2} \ldots, y_{i5})^\top$ representing the ith 5×1 response vector containing the concentration level of Cu, Pb, Zn, Ca, and Mg, respectively, $\boldsymbol{\mu}$ denoting the population-average, and $\boldsymbol{\xi}_i$ the vector of errors. Moreover, censored responses and the asymmetric behavior of the data led us to evaluate the multivariate skew-t (ST-CR) and skew-normal (SN-CR) censored models. We also fit the multivariate Student-t (T-CR) and normal (N-CR) censored models for comparison.

The ML estimates and the standard error (SE) for the population-average $\widehat{\boldsymbol{\mu}}$ are displayed in Table 15, as well as the skewness parameter $\widehat{\boldsymbol{\lambda}}$, the degrees of freedom $\widehat{\nu}$, the log-likelihood, and the information criteria AIC and BIC. Based on information criteria AIC and BIC, our ST-CR model fits this data better since it shows the lower values for these statistics.In other words, the result indicates that the asymmetric model is necessary for modeling the VDEQ data.

For the sake of comparison, we fitted the logarithm of the response using the ST-CR, SN-CR, T-CR, and N-CR models, obtaining that the T-CR has the best performance based on information criteria (see Table 16). Then, the log transformation does its job for this dataset by symmetrizing the data enough to make asymmetric models unnecessary. This can be evidenced in Table 16, where the estimates for $\boldsymbol{\lambda}$'s obtained through the ST-CR are not significantly different from zero. However, the AIC of the

Table 15 – VDEQ data. ML estimate, approximated standard error (SE), and model comparison criteria (AIC and BIC) from fitting the ST-CR, SN-CR, T-CR, and N-CR to the trace metals concentration dataset.

Par.	ST-CR		SN-CR		T-CR		N-CR	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
μ_1	0.650	0.075	0.559	0.081	0.396	0.023	0.556	0.078
μ_2	0.129	0.017	0.098	0.034	0.072	0.010	0.099	0.030
μ_3	2.233	0.538	2.320	0.751	1.261	0.139	2.314	0.639
μ_4	9.890	2.361	12.821	1.896	6.041	0.957	12.084	2.298
μ_5	3.564	0.614	4.020	1.232	2.190	0.249	3.814	0.734
λ_1	21.461	16.978	-1.730	6.346	-		-	
λ_2	21.632	15.455	0.230	2.100	-		-	
λ_3	30.568	26.840	3.001	11.954	-		-	
λ_4	28.895	17.177	26.956	70.259	-		-	
λ_5	21.697	11.935	8.438	27.712	-		-	
ν	2.186		-		2.019		-	
$\ell(\widehat{\boldsymbol{\theta}})$	-1647.309		-1936.688		-1673.541		-2007.606	
AIC	**3346.617**		3923.376		3389.081		4055.212	
BIC	**3472.051**		4043.985		3490.393		4151.699	

SN-CR is close to the AIC of the T-CR model, indicating a lack of adequacy of the symmetry distribution. Note also that the estimates for $\boldsymbol{\mu}$'s are roughly the same for the SN-CR, T-CR, and N-CR models. As seen, the logarithm transformation performs exceptionally well on this dataset due to the strong positive skewness and a large number of observations near zero; however, we will see in the following application a case where more than transformation is needed.

Table 16 – log-VDEQ data. ML estimate, approximated standard error (SE), and model selection criteria (AIC and BIC) from fitting the ST-CR, SN-CR, T-CR, and N-CR to the logarithmic transformation of the trace metals concentration dataset.

Par.	ST-CR		SN-CR		T-CR		N-CR	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
μ_1	-0.753	0.119	-0.875	0.066	-0.822	0.060	-0.873	0.067
μ_2	-2.201	0.126	-3.309	0.301	-3.069	0.204	-3.249	0.259
μ_3	0.459	0.219	0.294	0.094	0.263	0.085	0.281	0.098
μ_4	1.808	0.228	1.931	0.087	1.846	0.087	1.914	0.094
μ_5	0.817	0.193	0.905	0.071	0.868	0.078	0.874	0.088
λ_1	-1.101	2.046	-0.934	0.581	-		-	
λ_2	10.953	8.067	-2.781	2.028	-		-	
λ_3	3.147	3.243	1.773	1.263	-		-	
λ_4	1.400	3.149	2.309	1.289	-		-	
λ_5	-0.111	2.839	3.795	1.936	-		-	
ν	2.010		-		5.725		-	
$\ell(\widehat{\boldsymbol{\theta}})$	-899.975		-857.227		-861.032		-875.996	
AIC	1851.950		1764.455		**1764.065**		1791.992	
BIC	1977.384		1885.064		**1865.376**		1888.480	

5.5.2 Stellar abundances data

In astronomical research, a previously identified sample of objects (stars, galaxies, quasars, x-ray sources, etc.) is observed at some new wavebands. Due to limited sensitivities, some object features may be undetected or partially detected, leading to missing and censored data. In fact, astronomical data are typically left-censored. The probability of finding a new planet is related to the star's metal content; however, it is unclear whether this arises from the metallicity at birth or a later accretion of planetary bodies (Feigelson & Babu, 2012).

Here, we study whether the presence or absence of a giant planet is correlated with the level of lithium (Li) and beryllium (Be) in the photosphere of the host star. The dataset consists of $n = 68$ solar-type stars, 39 stars known to host planets, and 29 stars without planets, see Subsection 1.3.5 for more details. The proportion of left-censored and missing values is 17.6% and 47.1% for Be and Li, respectively. Therefore, we propose the following model:

$$\mathbf{y}_i = \left[\mathbf{I}_2 \otimes \mathbf{x}_i^\top\right] \boldsymbol{\beta} + \boldsymbol{\xi}_i, \qquad i \in \{1, \ldots, 68\}, \tag{5.24}$$

where for the ith object, $\mathbf{y}_i = (\ln \text{Be}, \ln \text{Li})_i^\top$ is the bivariate response of interest containing the natural logarithm of beryllium and lithium, and $\boldsymbol{\beta} = (\boldsymbol{\beta}_b^\top, \boldsymbol{\beta}_l^\top)^\top$ is the 6×1 vector, with $\boldsymbol{\beta}_b$ and $\boldsymbol{\beta}_l$ being 3×1 regression coefficients vectors for $\ln$ Be and $\ln$ Li. The vector of covariates is $\mathbf{x}_i = (1, \text{Type}_i, \text{Temp}_i - \overline{\text{Temp}})^\top$, where Type_i indicates planet-hosting stars (=1), Temp_i is the effective stellar surface temperature (in Kelvin degrees/1000), and $\overline{\text{Temp}} = 5.708$ represents the average temperature. The error term, denoted by $\boldsymbol{\xi}_i = (\xi_{bi},\ \xi_{li})^\top$, is considered independent and identically distributed. Note that model in (5.24) is equivalent to fit simultaneously

$$\begin{cases} \ln \text{Be}_i &= \beta_{b0} + \beta_{b1}\text{Type}_i + \beta_{b2}(\text{Temp}_i - \overline{\text{Temp}}) + \xi_{bi}. \\ \ln \text{Li}_i &= \beta_{l0} + \beta_{l1}\text{Type}_i + \beta_{l2}(\text{Temp}_i - \overline{\text{Temp}}) + \xi_{li}. \end{cases}$$

This model considers a correlation structure between the $\ln$ Be and $\ln$ Li, since $\text{cov}(\xi_b, \xi_l) \neq 0$.

Due to censored responses and asymmetric behavior of the Stellar abundance data, we propose to fit the skew-t censored (ST-CR) model. The multivariate skew-normal (SN-CR), Student-t (T-CR), and normal (N-CR) models are also fitted for comparison. Table 17 shows the estimated values for the regression coefficients $\widehat{\boldsymbol{\beta}}$, scale matrix $\widehat{\boldsymbol{\Sigma}}$, and skewness parameter $\widehat{\boldsymbol{\lambda}}$ obtained using the EM algorithm. The log-likelihood $\ell(\widehat{\boldsymbol{\theta}})$, AIC, and BIC information criteria are considered for model selection. Based on AIC and BIC, the ST-CR model is suitable to fit this data,indicating that the asymmetric model with heavy tails is necessary.

Considering the results obtained from the ST-CR model, we observe that at temperatures of 5708 Kelvin degrees, the expected logarithm concentration of beryllium

Table 17 – Stellar data. ML estimate, approximated standard error (SE), and model selection criteria (AIC and BIC) from fitting the ST-CR, SN-CR, T-CR, and N-CR to the natural logarithm of beryllium and lithium.

Par.	ST-CR		SN-CR		T-CR		N-CR	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_{b0}	0.797	0.051	0.736	0.071	0.957	0.047	0.859	0.126
β_{b1}	0.064	0.046	0.086	0.094	0.055	0.058	-0.022	0.123
β_{b2}	0.486	0.088	0.428	0.176	0.473	0.069	0.445	0.174
β_{l0}	0.947	0.226	1.053	0.253	1.184	0.161	1.187	0.252
β_{l1}	-0.453	0.235	-0.392	0.371	-0.411	0.247	-0.511	0.344
β_{l2}	4.077	0.556	2.983	0.722	4.077	0.547	2.970	0.735
σ_{bb}	0.050	0.015	0.285	0.043	0.021	0.005	0.148	0.032
σ_{bl}	0.093	0.047	0.346	0.147	0.047	0.017	0.202	0.089
σ_{ll}	0.436	0.201	1.195	0.546	0.389	0.127	1.046	0.445
λ_b	-3.681	1.865	-9.485	11.441	-		-	
λ_l	-1.957	1.236	-3.823	4.191	-		-	
ν	2.026		-		2.213		-	
$\ell(\widehat{\boldsymbol{\theta}})$	-61.384		-80.526		-68.178		-101.989	
AIC	**146.768**		183.053		156.357		221.978	
BIC	**181.720**		215.092		185.484		248.192	

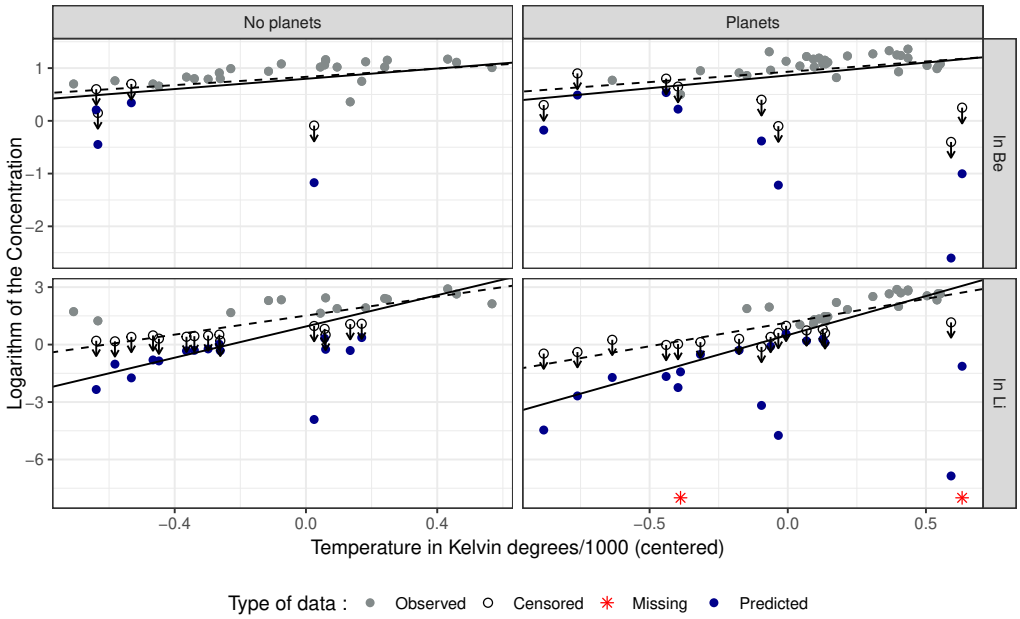
is 0.797 for stars without planets and 0.861 for planet-hosting stars, evidencing slight differences between these two types of stars. On the other hand, the expected level of the logarithm of lithium is 0.494 for planet-hosting stars and 0.947 otherwise. This last result shows a correlation between the lithium level and the presence or absence of a planet.

Finally, Figure 28 displays scatter plots and fitted regression lines for the logarithm of beryllium (top panel) and lithium (bottom panel) abundances against centered temperature in Kelvin degrees/1000. The dashed line denotes the regression line fitted under the skew-t model for the error term to points with no treatment, i.e., considering censored data as observed. In contrast, the solid line results from our proposed ST-CR model with blue points denoting the predicted values for the censored and missing observations. Here, we can see some differences in the intercepts and slopes for the lithium concentration, evidencing the importance of considering a proper model to deal with missing and/or censored data.

5.6 Remarks

This chapter introduced a robust multivariate regression model for censoring and/or missing data using the multivariate skew-t distribution, extending the recent work by Galarza *et al.* (2022b); Garay *et al.* (2017); Matos *et al.* (2019), among many others. The main advantage of the proposed ST-CR model is that it can reduce the negative impact of distributional misspecification and outliers on the estimation of the parameters. Moreover, the ST class has a convenient framework for implementing the EM algorithm.

Figure 28 – Stellar data. Scatter plot for the logarithm of beryllium (top) and lithium (bottom) abundances against temperature together with the regression lines fitted to points with no treatment, i.e., considering censored data as observed (dashed line) and under our proposed ST-CR model (solid line). Plots for non-planets-hosting stars are shown in the left panels.



The experimental results and the analysis of two real datasets support the usefulness and effectiveness of our proposal for handling multiple censoring and/or missing in the presence of asymmetric heavy-tailed data.

In a recent paper, Mattos *et al.* (2018) considered the problem of censored linear regression models using the univariate class of scale mixtures of skew-normal (SMSN) distributions. Therefore, it would be worthwhile to investigate the applicability of a likelihood-based treatment in the context of multivariate SMSN censored regression models. Other extensions of the current work include, for example, a generalization of the ST-CR model to the nonlinear setting (Matos *et al.*, 2019), censored mixed-effects models with skew-t random effects as will be proposed in the next chapter, or a finite mixture of censored data using the multivariate skew-t distribution (De Alencar *et al.*, 2021).

6 Likelihood-based inference for the censored linear mixed-effects skew-t model

6.1 Introduction

Linear mixed-effects (LME) models have been extensively used to analyze longitudinal, hierarchical, or grouped data. These models are particularly useful in fields such as biology, medicine, and agriculture, where data often exhibit correlation and non-constant variability within groups. LME models address these complexities by incorporating both fixed and random effects. Fixed effects estimate the overall impact of predictor variables of interest across all units, while random effects capture variations at the group level.

Traditionally, the most popular LME models assume that random effects and error terms follow normal distributions (LME-N). This assumption can be seen in Verbeke *et al.* (1997) and Pinheiro & Bates (2000). However, these restrictive assumptions often lead to robustness issues against the presence of deviations from normality and can result in invalid statistical inferences, particularly when the data exhibit heavy tails or outliers. Outliers can be particularly problematic in LME models compared to fixed effects models because they may occur in the random effects, the within-subject errors, or both, complicating their detection and treatment. Consequently, several models have been developed to accommodate heavy-tailed, asymmetric, or mixed distributions in both the random effects and error terms, as discussed in works by Verbeke & Lesaffre (1996), Pinheiro *et al.* (2001), and Schumacher *et al.* (2021).

Although conventional LME models assume that the response variable is fully observed, some observations may be censored in many practical scenarios. This censorship can occur due to limitations in measurement instruments or other methodological constraints. Addressing this, researchers such as Vaida & Liu (2009) and Matos *et al.* (2013a) have extended LME models to effectively handle censored data, assuming that the random effects and errors follow a normal distribution. Further advancements have been made to account for heavy tails in the data, as explored in studies by Matos *et al.* (2013b), Wang *et al.* (2018), and Lachos *et al.* (2019).

Moreover, employing asymmetric distributions has proven beneficial for data types like virologic responses, which often exhibit censored observations and non-normal features such as skewness with heavy tails. This approach avoids unrealistic normality assumptions and eliminates the need for data transformations, as proposed in a Bayesian context by Bandyopadhyay *et al.* (2012) and further extended in a likelihood framework by

Mattos *et al.* (2022b). The latter work considered a damped exponential correlation (DEC) for the within-subject dependence structure, replacing the independence assumption used in previous research.

No previous work has investigated LMEC models based on skew-t (ST) distributions from a likelihood-based perspective. Thus, this paper explores the linear mixed-effects censored skew-t (LMEC-ST) model, which incorporates the skewness and heavy-tailed nature of the random effects. The model assumes that the random effects for each subject follow an ST distribution (Arellano-Valle & Genton, 2005), while the within-subject errors follow a multivariate t distribution. We employ the Expectation/Conditional Maximization Either (ECME) algorithm (Liu & Rubin, 1994) for parameter estimation, chosen for its facility in computing the E-step, which involves calculating moments from the truncated multivariate extended skew-t (EST), ST, and t distributions (Galarza *et al.*, 2021c). The log-likelihood function is easily computed as a byproduct of the E-step and is used for monitoring convergence and model selection. All code used in this study will be available in the R library `skewlmm`.

The structure of the chapter is as follows: Section 6.2 details the LMEC-ST model, the computation of the log-likelihood function, and the parameter estimation process via the ECME algorithm. It also outlines methods for approximating the standard error of estimates, estimating random effects, and predicting future observations. Section 6.3 presents simulation study results that provide evidence about the asymptotic properties and robustness of the estimates obtained. The application of the model to a real dataset is described in Section 6.4, where we demonstrate that a model accounting for asymmetry and heavy tails better fits the data based on information criteria. We conclude with a discussion in Section 6.5.

6.2 The linear mixed-effects skew-t model for censored responses

This section introduces a linear mixed-effects model for censored observations in the response variable. It is assumed that the random effects are characterized by an asymmetric and heavy-tailed distribution, whereas the error term is modeled with a symmetric heavy-tailed distribution. Furthermore, an EM-type algorithm is developed to derive the maximum likelihood (ML) estimates. We also propose a method for approximating the standard errors of these estimates based on the empirical information matrix.

6.2.1 The linear mixed-effects model

Consider a scenario where a variable of interest, along with several covariates, is repeatedly measured across a set of n subjects at specific occasions during a study period. Let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^\top$ be the $n_i \times 1$ vector of observed continuous responses

related to the ith sample individual. The linear mixed-effects regression model is given by

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\xi}_i, \quad i \in \{1, \ldots, n\}, \tag{6.1}$$

where $\mathbf{X}_i$ is the $n_i \times \ell$ design matrix corresponding to the fixed effects, $\boldsymbol{\beta}$ is the $\ell \times 1$ vector of fixed effects, $\mathbf{Z}_i$ is the $n_i \times q$ design matrix corresponding to the $q \times 1$ random effects vector $\mathbf{b}_i$, and $\boldsymbol{\xi}_i$ is the $n_i \times 1$ vector of random errors. It is assumed that the random effects and the error components are independently distributed as $\mathbf{b}_i \overset{iid}{\sim} ST_q(\kappa\boldsymbol{\Delta}, \mathbf{D}, \boldsymbol{\lambda}, \nu)$ and $\boldsymbol{\xi}_i \overset{ind}{\sim} t_{n_i}(\mathbf{0}, \boldsymbol{\Omega}_i, \nu)$. The $q \times q$ random effects covariance matrix $\mathbf{D}$ may be unstructured or structured. Here, it is considered an unstructured matrix, which has at most $q(q+1)/2$ different elements. On the other hand, the $n_i \times n_i$ error covariance matrix $\boldsymbol{\Omega}_i$ is commonly written as $\boldsymbol{\Omega}_i = \sigma^2\mathbf{R}_i$, where $\mathbf{R}_i = \mathbf{R}_i(\boldsymbol{\phi})$ can be a known matrix or a structured matrix depending on a vector of parameter, say $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_\ell)^\top$. For $\nu > 1$, the parameters $\kappa = -(\nu/\pi)^{1/2}\dfrac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}$ and $\boldsymbol{\Delta} = \mathbf{D}^{1/2}\boldsymbol{\lambda}/(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}$ assure that $\mathbb{E}[\mathbf{b}_i] = \mathbf{0}$, consequently $\mathbb{E}[\mathbf{Y}_i] = \mathbf{X}_i\boldsymbol{\beta}$, for each $i \in \{1, \ldots, n\}$. Then, the marginal distribution of $\mathbf{Y}_i$ is given by

$$\mathbf{Y}_i \overset{ind}{\sim} ST_{n_i}(\mathbf{X}_i\boldsymbol{\beta} + \kappa\mathbf{Z}_i\boldsymbol{\Delta}, \boldsymbol{\Sigma}_i, \bar{\boldsymbol{\lambda}}_i, \nu), \tag{6.2}$$

where $\boldsymbol{\Sigma}_i = \boldsymbol{\Omega}_i + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top$ and $\bar{\boldsymbol{\lambda}}_i = \boldsymbol{\Sigma}_i^{-1/2}\mathbf{Z}_i\mathbf{D}\boldsymbol{\zeta}/(1+\boldsymbol{\zeta}^\top\boldsymbol{\Lambda}_i\boldsymbol{\zeta})^{1/2}$, with $\boldsymbol{\zeta} = \mathbf{D}^{-1/2}\boldsymbol{\lambda}$ and $\boldsymbol{\Lambda}_i = (\mathbf{D}^{-1} + \mathbf{Z}_i^\top\boldsymbol{\Omega}_i^{-1}\mathbf{Z}_i)^{-1}$. This result is demonstrated using Proposition 5 from Arellano-Valle & Genton (2010b) in Appendix E.1. We refer to the model defined by (6.1) and (6.2) as the *linear mixed-effects skew-t model* (LME-ST).

The LME-ST model can be represented hierarchically as follows:

$$\begin{aligned}
\mathbf{Y}_i \,|\, \mathbf{b}_i, U_i = u_i &\overset{ind}{\sim} N_{n_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, u_i^{-1}\sigma^2\mathbf{R}_i) \\
\mathbf{b}_i \,|\, T_i = t_i, U_i = u_i &\overset{ind}{\sim} N_q(t_i\boldsymbol{\Delta}, u_i^{-1}\boldsymbol{\Gamma}) \\
T_i \,|\, U_i = u_i &\overset{ind}{\sim} TN(\kappa, u_i^{-1}, (\kappa, \infty)) \\
U_i &\overset{iid}{\sim} Gamma(\nu/2, \nu/2),
\end{aligned} \tag{6.3}$$

with $\boldsymbol{\Gamma} = \mathbf{D} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top$, $\boldsymbol{\Delta} = \mathbf{D}^{1/2}\boldsymbol{\lambda}/\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}$, $\mathbf{D}^{1/2}$ denoting the square root of $\mathbf{D}$, and $TN(\mu, \sigma^2; (a, b))$ the univariate normal distribution with parameters μ and σ^2, truncated on the interval (a, b). This representation will be useful in Section 6.2.3 to implement the EM-type algorithm.

Additionally, in order to introduce flexibility in modeling the within-subject dependence structure, we consider three correlation structures: uncorrelated (UNC), autoregressive dependence of order p ($AR(p)$), and damped exponential correlation (DEC). The UNC model assumes that the error terms are conditionally uncorrelated, i.e., $\mathbf{R}_i = \mathbf{I}_{n_i}$, where $\mathbf{I}_{n_i}$ denotes the identity matrix of dimensions $n_i \times n_i$. However, measurements taken over time are generally correlated, invalidating the use of a UNC model. Therefore, we consider other structures that account for the within-subject serial correlation.

The $AR(p)$ structure assumes that the observations are measured at regularly discrete time points, then the correlation matrix is given by $\mathbf{R}_i = 1/(1 - \phi_1\rho_1 - \ldots - \phi_p\rho_p)[\rho_{|r-s|}]$, where $r, s \in \{1, \ldots, n_i\}$ and $\rho_1, \ldots, \rho_p$ are the theoretical autocorrelations of the process, and thereby they are functions of autoregressive parameters $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^\top$, and satisfy the Yule-Walker equations $\rho_k = \phi_1\rho_{k-1} + \ldots + \phi_p\rho_{k-p}, \rho_0 = 1, k \in \{1, \ldots, p\}$ (see for instance Lin & Lee, 2007; Box *et al.*, 2015).

On the other hand, the DEC model computes the entry (j, k) of the correlation matrix $\mathbf{R}_i$ by $R_i^{(j,k)} = \phi_1^{|t_{ij}-t_{ik}|^{\phi_2}}$, where t_{ij} and t_{ik} denote the times at which the jth and kth observation were measured for the ith subject. In this model, $0 \leqslant \phi_1 < 1$ and $\phi_2 \geqslant 0$. However, there are certain combinations of ϕ_1 and ϕ_2 that yield matrices which are not positive definite. Some particular cases of the DEC model are the compound symmetry correlation structure (CS), which occurs when $\phi_2 = 0$, the CAR(1) correlation is obtained when $\phi_2 = 1$, and the moving-average of order 1 (MA(1)) is attained when ϕ_2 tends to $+\infty$. For further details on these correlations, please refer to Schumacher *et al.* (2021).

6.2.2 The likelihood function

From now on, it is assumed that the response vector $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^\top$ is not fully observed due to censoring, i.e., the true response lies within a region instead of being observed exactly. Let $R_{ij} \subseteq \mathbb{R}$ denote the censoring region, such that Y_{ij} is not observed if $Y_{ij} \in R_{ij}$, for $j \in \{1, \ldots, n_i\}$. Further, let (V_{ij}, C_{ij}) be the observed data at time t_j for subject i, where C_{ij} is the censoring indicator, and V_{ij} is given by

$$V_{ij} = \begin{cases} r_{ij}, & \text{if } Y_{ij} \in R_{ij}, \quad \text{(censored)} \\ Y_{ij}, & \text{otherwise}, \quad \text{(observed)} \end{cases} \tag{6.4}$$

where R_{ij} is an interval of the form $(-\infty, r_{ij})$, (r_{ij}, ∞), or (r_{ij1}, r_{ij2}) for left, right, or interval censoring, respectively. The constant $r_{ij} \in \mathbb{R}$ is equal to the detection limit for left and right censoring, and $r_{ij} = (r_{ij1} + r_{ij2})/2$ for interval censoring. Moreover, missing observations can be handled by setting $R_{ij} = (-\infty, \infty)$ and $r_{ij} = \text{NA}$. Thereby, the model defined by (6.1)–(6.4) will be referred to as the *linear mixed-effects censored skew-t model* (LMEC-ST).

Therefore, let $\mathbf{y} = (\mathbf{y}_1^\top, \mathbf{y}_2^\top, \ldots, \mathbf{y}_n^\top)^\top$ be the vector of all responses, where $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$ is a realization of $\mathbf{Y}_i \sim ST_{n_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \bar{\boldsymbol{\lambda}}_i, \nu)$, with $\boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta} + \kappa\mathbf{Z}_i\boldsymbol{\Delta}$. In order to obtain the likelihood function of the LMEC-ST model, we treat, separately, the observed and censored components of $\mathbf{y}_i$, i.e., $\mathbf{y}_i = (\mathbf{y}_i^{o\top}, \mathbf{y}_i^{c\top})^\top$, where $C_{ij} = 0$ for all elements in the n_i^o-dimensional vector $\mathbf{y}_i^o$, and $C_{ij} = 1$ for all elements in the n_i^c-dimensional vector $\mathbf{y}_i^c$ of censored components. According to this partition, we write

$$\mathbf{V}_i = (\mathbf{V}_i^{o\top}, \mathbf{V}_i^{c\top})^\top, \quad \boldsymbol{\mu}_i = (\boldsymbol{\mu}_i^{o\top}, \boldsymbol{\mu}_i^{c\top})^\top, \quad \boldsymbol{\Sigma}_i = \begin{pmatrix} \boldsymbol{\Sigma}_i^{oo} & \boldsymbol{\Sigma}_i^{oc} \\ \boldsymbol{\Sigma}_i^{co} & \boldsymbol{\Sigma}_i^{cc} \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\varphi}_i = (\boldsymbol{\varphi}_i^{o\top}, \boldsymbol{\varphi}_i^{c\top})^\top,$$

where $\boldsymbol{\varphi}_i = \boldsymbol{\Sigma}_i^{-1/2}\bar{\boldsymbol{\lambda}}_i$. To compute the likelihood function, we need to know the marginal and conditional distribution of an ST variable. Then, from Proposition 6, we have that

$$\begin{aligned}\mathbf{Y}_i^o &\sim ST_{n_i^o}(\boldsymbol{\mu}_i^o, \boldsymbol{\Sigma}_i^{oo}, \tilde{\boldsymbol{\lambda}}_i^o, \nu),\\ \mathbf{Y}_i^c \mid \mathbf{Y}_i^o = \mathbf{y}_i^o &\sim EST_{n_i^c}(\boldsymbol{\mu}_i^{co}, \tilde{\boldsymbol{\Sigma}}_i^{cc.o}, \boldsymbol{\lambda}_i^{co}, \tau_i^{co}, \nu_i^{co}),\end{aligned}$$

with $\tilde{\boldsymbol{\lambda}}_i^o = c_i^{oc}\boldsymbol{\Sigma}_i^{oo\,1/2}\tilde{\boldsymbol{\varphi}}_i^o$, $c_i^{oc} = (1+\boldsymbol{\varphi}_i^{c\top}\boldsymbol{\Sigma}_i^{cc.o}\boldsymbol{\varphi}_i^c)^{-1/2}$, $\boldsymbol{\mu}_i^{co} = \boldsymbol{\mu}_i^c + \boldsymbol{\Sigma}_i^{co}\boldsymbol{\Sigma}_i^{oo-1}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o)$, $\tilde{\boldsymbol{\Sigma}}_i^{cc.o} = \boldsymbol{\Sigma}_i^{cc.o}/\nu^2(\mathbf{y}_i^o)$, $\boldsymbol{\Sigma}_i^{cc.o} = \boldsymbol{\Sigma}_i^{cc} - \boldsymbol{\Sigma}_i^{co}(\boldsymbol{\Sigma}_i^{oo})^{-1}\boldsymbol{\Sigma}_i^{oc}$, $\nu^2(\mathbf{y}_i^o) = (\nu + n_i^o)/(\nu + \delta(\mathbf{y}_i^o))$, $\boldsymbol{\lambda}_i^{co} = \boldsymbol{\Sigma}_i^{cc.o\,1/2}\boldsymbol{\varphi}_i^c$, $\tau_i^{co} = \nu(\mathbf{y}_i^o)\tilde{\boldsymbol{\varphi}}_i^{o\top}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o)$, $\tilde{\boldsymbol{\varphi}}_i^o = \boldsymbol{\varphi}_i^o + \boldsymbol{\Sigma}_i^{oo\,-1}\boldsymbol{\Sigma}_i^{oc}\boldsymbol{\varphi}_i^c$, and $\nu_i^{co} = \nu + n_i^o$.

Let $\mathbf{V} = (\mathbf{V}_1^\top, \ldots, \mathbf{V}_n^\top)^\top$ and $\mathbf{C} = (\mathbf{C}_1^\top, \ldots, \mathbf{C}_n^\top)^\top$ denote the observed data. Therefore, the log-likelihood function of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\alpha}_{\mathbf{D}}^\top, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\lambda}^\top, \nu)^\top$, where $\boldsymbol{\alpha}_{\mathbf{D}}$ denotes a minimal set of parameters such that $\mathbf{D}$ is well defined (e.g., the upper triangular elements of $\mathbf{D}$ in the unstructured case), for the observed data $(\mathbf{V}, \mathbf{C})$ is

$$\ell(\boldsymbol{\theta} \mid \mathbf{V}, \mathbf{C}) = \sum_{i=1}^{n} \ln L_i, \tag{6.5}$$

where L_i represents the likelihood function of $\boldsymbol{\theta}$ for the ith subject, given by

$$\begin{aligned}L_i \equiv L_i(\boldsymbol{\theta} \mid \mathbf{V}_i, \mathbf{C}_i) &= \Pr(\mathbf{r}_{i1}^c \leqslant \mathbf{y}_i^c \leqslant \mathbf{r}_{i2}^c \mid \mathbf{y}_i^o, \boldsymbol{\theta}) f(\mathbf{y}_i^o \mid \boldsymbol{\theta})\\ &= \mathcal{P}_{n_i^c}(\mathbf{r}_{i1}^c, \mathbf{r}_{i2}^c; \boldsymbol{\mu}_i^{co}, \tilde{\boldsymbol{\Sigma}}_i^{cc.o}, \boldsymbol{\lambda}_i^{co}, \tau_i^{co}, \nu_i^{co})\, ST_{n_i^o}(\mathbf{y}_i^o; \boldsymbol{\mu}_i^o, \boldsymbol{\Sigma}_i^{oo}, \tilde{\boldsymbol{\lambda}}_i^o, \nu),\end{aligned}$$

with $\{\mathbf{y}_i^c \in \mathbb{R}^{n_i^c} : \mathbf{r}_{i1}^c \leqslant \mathbf{y}_i^c \leqslant \mathbf{r}_{i2}^c\} = \{(y_{i1}^c, \ldots, y_{in_i^c}^c)^\top \in \mathbb{R}^{n_i^c} : r_{i11} \leqslant y_{i1}^c \leqslant r_{i12}^c, \ldots, r_{in_i^c 1} \leqslant y_{in_i^c}^c \leqslant r_{in_i^c 2}^c\}$ denoting the censoring region of all partially observed data for subject i, and $\mathcal{P}_r(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, \nu)$ denotes the integral defined in (5.4), which can be easily evaluated by using the R package `MomTrunc` (Galarza *et al.*, 2021a).

The estimation procedure for the model parameters $\boldsymbol{\theta}$ is performed by maximizing the log-likelihood function given in (6.5). This process is not straightforward, as the expression involves intractable integrals required to compute the probabilities of an EST distribution for censored cases. Consequently, an EM-type algorithm is the most widely used method to address this complexity. This algorithm optimizes a sequence of simpler approximations of the observed log-likelihood function. Further details on the implementation of the algorithm will be discussed in the subsequent section.

6.2.3 Parameter estimation via EM-type algorithm

The EM algorithm, as introduced by Dempster *et al.* (1977), is a widely used method for deriving ML estimates in scenarios involving partially observed data, such as censored or missing data. This algorithm iteratively estimates parameters by maximizing the conditional expectation of the complete-data log-likelihood function. The EM algorithm is noted for its numerical stability, simplicity of implementation, and modest memory

requirements (McLachlan & Krishnan, 2008). However, ML estimation for the LMEC-ST model presents challenges due to censoring issues, rendering the EM algorithm less suitable due to computational complexities during the M-step. To address these challenges, we employ an extension of the EM algorithm known as the Expectation-Conditional Maximization Either (ECME) algorithm (Liu & Rubin, 1994). The ECME algorithm modifies the M-step with a series of conditional maximization (CM) steps, maintaining the stability characteristic of the EM algorithm and typically achieving a faster convergence rate than the original EM.

In order to develop the ECEM algorithm for parameter estimation in the LMEC-ST model, we adopt the hierarchical representation of the model as expressed in (6.3). Define $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)^\top$, $\mathbf{u} = (u_1, \ldots, u_n)^\top$ and $\mathbf{t} = (t_1, \ldots, t_n)^\top$ as the hypothetical missing data, complemented by the observed data $\mathbf{V}$ and $\mathbf{C}$, which correspond to the censoring mechanism. The complete dataset, therefore, comprises $\mathbf{y}_c = (\mathbf{V}, \mathbf{C}, \mathbf{y}, \mathbf{u}, \mathbf{t})$. Holding the value of ν constant, we define the complete-data log-likelihood function for a corresponding set of parameters $\boldsymbol{\theta}^* = (\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\Delta}, \boldsymbol{\alpha}_{\boldsymbol{\Gamma}})^\top$ by

$$\ell_c(\boldsymbol{\theta}^* \mid \mathbf{y}_c) = \sum_{i=1}^{n} \ell_{ic}(\boldsymbol{\theta}^* \mid \mathbf{y}_{ic}),$$

where

$$\begin{aligned}\ell_{ic}(\boldsymbol{\theta}^* \mid \mathbf{y}_{ic}) \;=\; & -\frac{1}{2}\Big[\ln|\mathbf{R}_i| + \ln|\boldsymbol{\Gamma}| + \frac{u_i}{\sigma^2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i)^\top\mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i) \\ & + n_i\ln\sigma^2 + u_i(\mathbf{b}_i - \boldsymbol{\Delta}t_i)^\top\boldsymbol{\Gamma}^{-1}(\mathbf{b}_i - \boldsymbol{\Delta}t_i)\Big] + K(\nu) + c,\end{aligned}$$

with c denoting a constant that does not depend on the vector of parameters $\boldsymbol{\theta}^*$, $K(\nu)$ is a function that depends only on ν, and $|\mathbf{A}|$ represents the determinant of the square matrix $\mathbf{A}$.

Therefore, the EM-type algorithm for the LMEC-ST model can be summarized in two steps:

E-step: Given the estimates of $\widehat{\boldsymbol{\theta}}^{*(k)} = (\widehat{\boldsymbol{\beta}}^{(k)}, \widehat{\sigma}^{2(k)}, \widehat{\boldsymbol{\phi}}^{(k)}, \widehat{\boldsymbol{\Delta}}^{(k)}, \widehat{\boldsymbol{\alpha}}_{\boldsymbol{\Gamma}}^{(k)})^\top$ at the kth iteration of the algorithm, the E-step provides the conditional expectation of the complete-data log-likelihood function:

$$Q(\boldsymbol{\theta}^*|\widehat{\boldsymbol{\theta}}^{*(k)}) = \mathbb{E}\left[\ell_c(\boldsymbol{\theta}^*|\mathbf{y}_c)|\mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{*(k)}\right] = \sum_{i=1}^{n}\mathbb{E}\left[\ell_c(\boldsymbol{\theta}^*|\mathbf{y}_{ic})|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{*(k)}\right] = \sum_{i=1}^{n}\widehat{Q}_i^{(k)}(\boldsymbol{\theta}^*).$$

It is possible to demonstrate that $\widehat{Q}_i^{(k)}(\boldsymbol{\theta}^*)$ can be decomposed as follows

$$\widehat{Q}_i^{(k)}(\boldsymbol{\theta}) = \widehat{Q}_{1i}^{(k)}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}) + \widehat{Q}_{2i}^{(k)}(\boldsymbol{\alpha}, \boldsymbol{\lambda}) + \widehat{Q}_{3i}(\nu),$$

where

$$\begin{aligned}\widehat{Q}_{1i}^{(k)}(\boldsymbol{\beta},\sigma^2,\boldsymbol{\phi}) &= -\frac{1}{2}\ln|\mathbf{R}_i| - \frac{1}{2\sigma^2}\mathrm{tr}\left(\mathbf{R}_i^{-1}\left[\widehat{\mathbf{uy}_i^2}^{(k)} + \mathbf{Z}_i\widehat{\mathbf{ub}_i^2}^{(k)}\mathbf{Z}_i^\top - 2\mathbf{Z}_i\widehat{\mathbf{uby}_i^\top}^{(k)}\right]\right)\\ &\quad -\frac{n_i}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\left(\widehat{u}_i^{(k)}\mathbf{X}_i\boldsymbol{\beta} - 2\widehat{\mathbf{uy}_i}^{(k)} + 2\mathbf{Z}_i\widehat{\mathbf{ub}_i}^{(k)}\right)^\top\mathbf{R}_i^{-1}\mathbf{X}_i\boldsymbol{\beta}.\\ \widehat{Q}_{2i}^{(k)}(\boldsymbol{\alpha},\boldsymbol{\lambda}) &= -\frac{1}{2}\ln|\boldsymbol{\Gamma}| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Gamma}^{-1}\widehat{\mathbf{ub}_i^2}^{(k)}\right) + \widehat{\mathbf{utb}_i^\top}^{(k)}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta} - \frac{\widehat{ut_i^2}^{(k)}}{2}\boldsymbol{\Delta}^\top\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta},\end{aligned}$$

such that $\widehat{u}_i^{(k)} = \mathbb{E}[U_i|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}]$, $\widehat{ut_i^2}^{(k)} = \mathbb{E}[U_iT_i^2|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}]$, $\widehat{\mathbf{ub}_i^r}^{(k)} = \mathbb{E}[U_i\mathbf{b}_i^r|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}]$, $\widehat{\mathbf{utb}_i}^{(k)} = \mathbb{E}[U_iT_i\mathbf{b}_i|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}]$, $\widehat{\mathbf{uby}_i^\top}^{(k)} = \mathbb{E}[U_i\mathbf{b}_i\mathbf{Y}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}]$, and $\widehat{\mathbf{uy}_i^r}^{(k)} = \mathbb{E}[U_i\mathbf{Y}_i^r|\mathbf{V}_i,\ \mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}]$, for $r \in \{1,2\}$, such that $\mathbf{b}_i^1 = \mathbf{b}_i$, $\mathbf{b}_i^2 = \mathbf{b}_i\mathbf{b}_i^\top$, $\mathbf{Y}_i^1 = \mathbf{Y}_i$, and $\mathbf{Y}_i^2 = \mathbf{Y}_i\mathbf{Y}_i^\top$. The computation of these conditional expectations is detailed in Appendix E.2, where expressions for each scenario are provided (cases where the response vector is uncensored, cases where all components are censored, and cases where only some components are censored). These moments can be approximated using the R libraries `MomTrunc` (Galarza *et al.*, 2021a) and `relliptical` (Chapter 2).

M-step: Conditionally maximizing $Q(\boldsymbol{\theta}^*|\widehat{\boldsymbol{\theta}}^{*(k)})$ with respect to each entry of $\boldsymbol{\theta}^*$, we update the estimated $\widehat{\boldsymbol{\theta}}^{*(k)}$ by:

$$\begin{aligned}\widehat{\boldsymbol{\beta}}^{(k+1)} &= \left(\sum_{i=1}^n \widehat{u}_i^{(k)}\mathbf{X}_i^\top\widehat{\boldsymbol{\Omega}}_i^{-1(k)}\mathbf{X}_i\right)^{-1}\sum_{i=1}^n\mathbf{X}_i^\top\widehat{\boldsymbol{\Omega}}_i^{-1(k)}\left(\widehat{\mathbf{uy}_i}^{(k)} - \mathbf{Z}_i\widehat{\mathbf{ub}_i}^{(k)}\right),\\ \widehat{\sigma}^{2(k+1)} &= \frac{1}{N}\sum_{i=1}^n\left\{\mathrm{tr}\left(\widehat{\mathbf{R}}_i^{-1(k)}\left[\widehat{\mathbf{uy}_i^2}^{(k)} + \mathbf{Z}_i\widehat{\mathbf{ub}_i^2}^{(k)}\mathbf{Z}_i^\top - 2\mathbf{Z}_i\widehat{\mathbf{uby}_i^\top}^{(k)}\right]\right)\right.\\ &\quad \left. + \left(\widehat{u}_i^{(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k+1)} - 2\widehat{\mathbf{uy}_i}^{(k)} + 2\mathbf{Z}_i\widehat{\mathbf{ub}_i}^{(k)}\right)^\top\widehat{\mathbf{R}}_i^{-1(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k+1)}\right\},\\ \widehat{\boldsymbol{\phi}}^{(k+1)} &= \underset{\boldsymbol{\phi}}{\mathrm{argmax}}\sum_{i=1}^n\left\{-\frac{1}{2\widehat{\sigma}^{2(k+1)}}\mathrm{tr}\left(\mathbf{R}_i^{-1}\left[\widehat{\mathbf{uy}_i^2}^{(k)} + \mathbf{Z}_i\widehat{\mathbf{ub}_i^2}^{(k)}\mathbf{Z}_i^\top - 2\mathbf{Z}_i\widehat{\mathbf{uby}_i^\top}^{(k)}\right]\right)\right.\\ &\quad \left. -\frac{1}{2}\ln|\mathbf{R}_i| - \frac{1}{2\widehat{\sigma}^{2(k+1)}}\left(\widehat{u}_i^{(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k+1)} - 2\widehat{\mathbf{uy}_i}^{(k)} + 2\mathbf{Z}_i\widehat{\mathbf{ub}_i}^{(k)}\right)^\top\mathbf{R}_i^{-1}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k+1)}\right\},\\ \widehat{\boldsymbol{\Delta}}^{(k+1)} &= \frac{\sum_{i=1}^n\widehat{\mathbf{utb}_i}^{(k)}}{\sum_{i=1}^n\widehat{ut_i^2}^{(k)}},\\ \widehat{\boldsymbol{\Gamma}}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^n\left\{\widehat{\mathbf{ub}_i^2}^{(k)} - \widehat{\mathbf{utb}_i}^{(k)}\widehat{\boldsymbol{\Delta}}^{(k+1)\top} - \widehat{\boldsymbol{\Delta}}^{(k+1)}\widehat{\mathbf{utb}_i}^{(k)\top} + \widehat{ut_i^2}^{(k)}\widehat{\boldsymbol{\Delta}}^{(k+1)}\widehat{\boldsymbol{\Delta}}^{(k+1)\top}\right\},\end{aligned}$$

where $N = \sum_{i=1}^n n_i$. Additionally, estimates for certain parameters associated with the random effects, such as the unstructured scale matrix $\mathbf{D}$ and the skewness parameter $\boldsymbol{\lambda}$, can be obtained using the following expressions

$$\widehat{\mathbf{D}}^{(k+1)} = \widehat{\boldsymbol{\Gamma}}^{(k+1)} + \widehat{\boldsymbol{\Delta}}^{(k+1)}\widehat{\boldsymbol{\Delta}}^{(k+1)\top} \qquad \text{and} \qquad \widehat{\boldsymbol{\lambda}}^{(k+1)} = \frac{\widehat{\mathbf{D}}^{-1/2(k+1)}\widehat{\boldsymbol{\Delta}}^{(k+1)}}{\left(1 - \widehat{\boldsymbol{\Delta}}^{(k+1)\top}\widehat{\mathbf{D}}^{-1(k+1)}\widehat{\boldsymbol{\Delta}}^{(k+1)}\right)^{1/2}}.$$

To update the degrees of freedom ν, we maximize the marginal log-likelihood function, given in (6.5), with respect to ν as follows

$$\widehat{\nu}^{(k+1)} = \underset{\nu}{\operatorname{argmax}} \left\{ \sum_{i=1}^{n} \ln \left(L_i(\nu | \widehat{\boldsymbol{\theta}}^{(k+1)}, \mathbf{V}_i, \mathbf{C}_i) \right) \right\}.$$

To perform the maximization procedure, we employ the algorithm proposed by Brent (2013), which is a combination of golden section search and successive parabolic interpolation, designed for use with continuous functions of a single variable. The EM algorithm iterations continue until the difference between two successive values of the observed log-likelihood function is less than a specified tolerance level, denoted as `tol` (e.g. `tol` $< 10^{-5}$).

6.2.4 Standard error approximation

According to the large sample theory, the standard error of the ML estimates can be approximated by the inverse of the observed information matrix. However, in the presence of censored observations, this matrix cannot be expressed in closed form. Consequently, we adopt the strategies detailed by Lin (2010) and Meilijson (1989), which utilize the empirical information matrix instead of the observed one. Therefore, the empirical information matrix, evaluated at the ML estimates, is defined as follows:

$$\mathbf{I}_e(\widehat{\boldsymbol{\theta}}|\mathbf{y}) = \sum_{i=1}^{n} \widehat{\mathbf{s}}_i(\boldsymbol{\theta}) \widehat{\mathbf{s}}_i(\boldsymbol{\theta})^\top.$$

For the LMEC-ST model, let $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\sigma}^2, \widehat{\boldsymbol{\phi}}^\top, \widehat{\boldsymbol{\alpha}}_{\mathbf{D}}^\top, \widehat{\boldsymbol{\lambda}}^\top, \nu)^\top$ denote the vector of ML estimates obtained at the last iteration of the EM-type algorithm, with $\widehat{\boldsymbol{\alpha}}_{\mathbf{D}} = (\widehat{\alpha}_1, \ldots, \widehat{\alpha}_{q(q+1)/2})^\top$ representing the $q(q+1)/2$ vector of distinct elements of $\widehat{\mathbf{D}}$, then the vector $\widehat{\mathbf{s}}_i(\boldsymbol{\theta}) = (\widehat{\mathbf{s}}_i(\boldsymbol{\beta})^\top, \widehat{\mathbf{s}}_i(\sigma^2), \widehat{\mathbf{s}}_i(\boldsymbol{\phi})^\top, \widehat{\mathbf{s}}_i(\boldsymbol{\alpha}_{\mathbf{D}})^\top, \widehat{\mathbf{s}}_i(\boldsymbol{\lambda})^\top, \widehat{\mathbf{s}}_i(\nu))^\top$ has elements

$$\begin{aligned}
\widehat{\mathbf{s}}_i(\boldsymbol{\beta}) &= \mathbf{X}_i^\top \widehat{\boldsymbol{\Omega}}_i^{-1} \left(\widehat{\mathbf{uy}}_i - \mathbf{Z}_i \widehat{\mathbf{ub}}_i - \widehat{u}_i \mathbf{X}_i \widehat{\boldsymbol{\beta}} \right), \\
\widehat{\mathbf{s}}_i(\sigma^2) &= -\frac{n_i}{2\widehat{\sigma}^2} + \frac{1}{2\widehat{\sigma}^4} \left[\widehat{a}_i - 2\widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i^\top \widehat{\mathbf{R}}_i^{-1} \left(\widehat{\mathbf{uy}}_i - \mathbf{Z}_i \widehat{\mathbf{ub}}_i \right) + \widehat{u}_i \widehat{\boldsymbol{\beta}}^\top \mathbf{X}_i^\top \widehat{\mathbf{R}}_i^{-1} \mathbf{X}_i \widehat{\boldsymbol{\beta}} \right], \\
\widehat{\mathbf{s}}_i(\boldsymbol{\phi}) &= (\widehat{\mathbf{s}}_i(\phi_1), \widehat{\mathbf{s}}_i(\phi_2), \ldots, \widehat{\mathbf{s}}_i(\phi_r))^\top, \\
\widehat{\mathbf{s}}_i(\boldsymbol{\alpha}_{\mathbf{D}}) &= \left(\widehat{\mathbf{s}}_i(\alpha_1), \widehat{\mathbf{s}}_i(\alpha_2), \ldots, \widehat{\mathbf{s}}_i(\alpha_{q(q+1)/2}) \right)^\top, \\
\widehat{\mathbf{s}}_i(\boldsymbol{\lambda}) &= (\widehat{\mathbf{s}}_i(\lambda_1), \widehat{\mathbf{s}}_i(\lambda_2), \ldots, \widehat{\mathbf{s}}_i(\lambda_q))^\top, \quad \text{and} \\
\widehat{\mathbf{s}}_i(\nu) &= \frac{1}{2} \left\{ \ln \left(\frac{\widehat{\nu}}{2} \right) + 1 - \psi \left(\frac{\widehat{\nu}}{2} \right) + \mathbb{E} \left[\ln U_i | \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}} \right] - \widehat{u}_i \right\},
\end{aligned}$$

with $\widehat{a}_i = \operatorname{tr}(\mathbf{R}_i^{-1} \widehat{\mathbf{A}}_i)$ such that $\widehat{\mathbf{A}}_i = \widehat{\mathbf{uy}_i^2} + \mathbf{Z}_i \widehat{\mathbf{ub}_i^2} \mathbf{Z}_i^\top - 2\mathbf{Z}_i \widehat{\mathbf{uby}_i^\top}$, and $\psi(x) = \Gamma'(x)/\Gamma(x)$ represents the digamma function. Note that $\widehat{\mathbf{s}}_i(\nu)$ depends on the calculation of $\mathbb{E}[\ln U_i | \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}]$, which relies on computationally intensive Monte Carlo integration, then we focus on computing the remaining elements of $\widehat{\mathbf{s}}_i(\boldsymbol{\theta})$. The rest of the elements are given by

$$\begin{aligned}
\widehat{\mathbf{s}}_i(\phi_k) &= \frac{1}{2\widehat{\sigma}^2}\mathrm{tr}\left\{\left[\widehat{\mathbf{A}}_i + \widehat{u}_i\mathbf{X}_i\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_i^\top - 2(\widehat{\mathbf{uy}}_i - \mathbf{Z}_i\widehat{\mathbf{ub}}_i)\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_i^\top\right]\widehat{\mathbf{R}}_i^{-1}\dot{\mathbf{R}}_{i\phi_k}\widehat{\mathbf{R}}_i^{-1}\right\} - \frac{1}{2}\mathrm{tr}\left\{\widehat{\mathbf{R}}_i^{-1}\dot{\mathbf{R}}_{i\phi_k}\right\},\\
\widehat{\mathbf{s}}_i(\alpha_l) &= -\frac{1}{2}\mathrm{tr}\left\{\widehat{\boldsymbol{\Gamma}}^{-1}\dot{\boldsymbol{\Gamma}}_{\alpha_l} - \widehat{\mathbf{E}}_i\widehat{\boldsymbol{\Gamma}}^{-1}\dot{\boldsymbol{\Gamma}}_{\alpha_l}\widehat{\boldsymbol{\Gamma}}^{-1} - [\widehat{\mathbf{utb}}_i\dot{\boldsymbol{\Delta}}_{\alpha_l}^\top + \dot{\boldsymbol{\Delta}}_{\alpha_l}\widehat{\mathbf{utb}}_i^\top]\widehat{\boldsymbol{\Gamma}}^{-1} + \widehat{ut}_i^2[\dot{\boldsymbol{\Delta}}_{\alpha_l}\widehat{\boldsymbol{\Delta}}^\top + \widehat{\boldsymbol{\Delta}}\dot{\boldsymbol{\Delta}}_{\alpha_l}^\top]\widehat{\boldsymbol{\Gamma}}^{-1}\right\},\\
\widehat{\mathbf{s}}_i(\lambda_j) &= -\frac{1}{2}\mathrm{tr}\left\{\widehat{\boldsymbol{\Gamma}}^{-1}\dot{\boldsymbol{\Gamma}}_{\lambda_j} - \widehat{\mathbf{E}}_i\widehat{\boldsymbol{\Gamma}}^{-1}\dot{\boldsymbol{\Gamma}}_{\lambda_j}\widehat{\boldsymbol{\Gamma}}^{-1} - [\widehat{\mathbf{utb}}_i\dot{\boldsymbol{\Delta}}_{\lambda_j}^\top + \dot{\boldsymbol{\Delta}}_{\lambda_j}\widehat{\mathbf{utb}}_i^\top]\widehat{\boldsymbol{\Gamma}}^{-1} + \widehat{ut}_i^2[\dot{\boldsymbol{\Delta}}_{\lambda_j}\widehat{\boldsymbol{\Delta}}^\top + \widehat{\boldsymbol{\Delta}}\dot{\boldsymbol{\Delta}}_{\lambda_j}^\top]\widehat{\boldsymbol{\Gamma}}^{-1}\right\},
\end{aligned}$$

where

$$\widehat{\mathbf{E}}_i = \widehat{\mathbf{ub}_i^2} - \widehat{\mathbf{utb}}_i\widehat{\boldsymbol{\Delta}}^\top - \widehat{\boldsymbol{\Delta}}\widehat{\mathbf{utb}}_i^\top + \widehat{ut}_i^2\widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{\Delta}}^\top,$$

$$\dot{\boldsymbol{\Gamma}}_{\alpha_l} = \frac{\partial\boldsymbol{\Gamma}}{\partial\alpha_l}\Big|_{\boldsymbol{\alpha}=\widehat{\boldsymbol{\alpha}}} = \dot{\mathbf{D}}_{\alpha_l} - \dot{\boldsymbol{\Delta}}_{\alpha_l}\widehat{\boldsymbol{\Delta}}^\top - \widehat{\boldsymbol{\Delta}}\dot{\boldsymbol{\Delta}}_{\alpha_l}^\top, \quad \text{with} \quad \dot{\boldsymbol{\Delta}}_{\alpha_l} = \frac{\dot{\mathbf{D}}_{\alpha_l}^{1/2}\widehat{\boldsymbol{\lambda}}}{(1+\widehat{\boldsymbol{\lambda}}^\top\widehat{\boldsymbol{\lambda}})^{1/2}}, \quad \text{and}$$

$$\dot{\boldsymbol{\Gamma}}_{\lambda_j} = \frac{\partial\boldsymbol{\Gamma}}{\partial\lambda_j}\Big|_{\boldsymbol{\lambda}=\widehat{\boldsymbol{\lambda}}} = -\dot{\boldsymbol{\Delta}}_{\lambda_j}\widehat{\boldsymbol{\Delta}}^\top - \widehat{\boldsymbol{\Delta}}\dot{\boldsymbol{\Delta}}_{\lambda_j}^\top, \text{ with } \dot{\boldsymbol{\Delta}}_{\lambda_j} = \widehat{\mathbf{D}}^{1/2}\left(\frac{\dot{\boldsymbol{\lambda}}_j}{(1+\widehat{\boldsymbol{\lambda}}^\top\widehat{\boldsymbol{\lambda}})^{1/2}} - \frac{\widehat{\lambda}_j\widehat{\boldsymbol{\lambda}}}{(1+\widehat{\boldsymbol{\lambda}}^\top\widehat{\boldsymbol{\lambda}})^{3/2}}\right),$$

for $k \in \{1,\ldots,r\}$, $l \in \{1,\ldots,q(q+1)/2\}$, $j \in \{1,\ldots,q\}$, and $i \in \{1,\ldots,n\}$. Additionally, $\dot{\boldsymbol{\lambda}}_j = \dfrac{\partial\boldsymbol{\lambda}}{\partial\lambda_j}$ denotes the $q\times 1$ vector with 1 in the jth entry and 0 otherwise, $\dot{\mathbf{D}}_{\alpha_l} = \dfrac{\partial\mathbf{D}}{\partial\alpha_l}$ represents the $q\times q$ matrix with 1 in the position of the element α_l and the remaining elements equal to 0, and $\dot{\mathbf{D}}_{\alpha_l}^{1/2} = \dfrac{\partial\mathbf{D}^{1/2}}{\partial\alpha_l}\Big|_{\boldsymbol{\alpha}=\widehat{\boldsymbol{\alpha}}}$ is computed by

$$\mathrm{vec}\left(\dot{\mathbf{D}}_{\alpha_l}^{1/2}\right) = \left(\mathbf{D}^{1/2}\otimes\mathbf{I}_n + \mathbf{I}_n\otimes\mathbf{D}^{1/2}\right)^{-1}\mathrm{vec}\left(\dot{\mathbf{D}}_{\alpha_l}\right),$$

where $\mathrm{vec}(\mathbf{D})$ represents the operator that transforms a matrix into a column vector by vertically stacking its columns. Given that $\mathbf{D}$ is positive definite, the matrix square root $\mathbf{D}^{1/2}$ is unique and positive definite. Consequently, the associated Kronecker sum is also positive definite and non-singular. These results are similar to those presented in Chapter 5 for the multivariate censored skew-t distribution.

6.2.5 Estimation of random effects

To approximate the random effects for each individual, we consider the result presented in (6.3), which corresponds to the hierarchical representation of the ST model. From this foundation, it follows that $\mathbf{Y}_i \,|\, \mathbf{b}_i = \mathbf{b}_i, U_i = u_i \sim N_{n_i}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i, u_i^{-1}\boldsymbol{\Omega}_i)$ and $\mathbf{b}_i \,|\, (U_i = u_i) \sim \mathrm{SN}_q(\kappa\boldsymbol{\Delta}, u_i^{-1}\mathbf{D}, \boldsymbol{\lambda})$. Consequently, given $(\mathbf{Y}_i, U_i)$, the distribution of $\mathbf{b}_i$ falls within the extended skew-normal (ESN) family, as follows:

$$\mathbf{b}_i \,|\, (\mathbf{Y}_i = \mathbf{y}_i, U_i = u_i) \sim \mathrm{ESN}_q(\boldsymbol{\mu}_{bi}, u_i^{-1}\boldsymbol{\Lambda}_i, \boldsymbol{\Lambda}_i^{1/2}\boldsymbol{\zeta}, \tau_{bi}), \tag{6.6}$$

where $\boldsymbol{\mu}_{bi} = \kappa\boldsymbol{\Delta} + \mathbf{D}\mathbf{Z}_i^\top\boldsymbol{\Sigma}_i^{-1/2}\mathbf{y}_{0i}$ and $\tau_{bi} = u_i^{1/2}\boldsymbol{\zeta}^\top\mathbf{D}\mathbf{Z}_i^\top\boldsymbol{\Sigma}_i^{-1/2}\mathbf{y}_{0i}$, with $\mathbf{y}_{0i} = \boldsymbol{\Sigma}_i^{-1/2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \kappa\mathbf{Z}_i\boldsymbol{\Delta})$ and $\boldsymbol{\zeta} = \mathbf{D}^{-1/2}\boldsymbol{\lambda}$. Therefore, the conditional mean of $\mathbf{b}_i$ given $\mathbf{Y}_i = \mathbf{y}_i$ can be computed by

$$\begin{aligned}
\mathbb{E}[\mathbf{b}_i|\mathbf{Y}_i = \mathbf{y}_i, \boldsymbol{\theta}] &= \mathbb{E}[\,\mathbb{E}[\mathbf{b}_i|U_i, \mathbf{Y}_i = \mathbf{y}_i, \boldsymbol{\theta}]\,|\,\mathbf{Y}_i = \mathbf{y}_i, \boldsymbol{\theta}]\\
&= \boldsymbol{\mu}_{bi} + \frac{\mathbb{E}[U_i^{-1/2}W_\Phi(U_i^{1/2}\bar{\boldsymbol{\lambda}}_i^\top\mathbf{y}_{0i})\,|\,\mathbf{Y}_i = \mathbf{y}_i, \boldsymbol{\theta}]}{\sqrt{1+\boldsymbol{\zeta}^\top\boldsymbol{\Lambda}_i\boldsymbol{\zeta}}}\boldsymbol{\Lambda}_i\boldsymbol{\zeta},
\end{aligned}$$

with $W_\Phi(a) = \phi(a)/\Phi(a)$. The minimum mean-squared error (MSE) estimator of $\mathbf{b}_i$ is obtained by the conditional mean of $\mathbf{b}_i$ given the observed values $(\mathbf{V}_i, \mathbf{C}_i)$ (Lachos *et al.*, 2010), which is given by

$$\begin{aligned}\widehat{\mathbf{b}}_i(\boldsymbol{\theta}) &= \mathbb{E}[\,\mathbf{b}_i|\mathbf{V}_i, \mathbf{C}_i, \boldsymbol{\theta}] = \mathbb{E}[\,\mathbb{E}[\mathbf{b}_i|\mathbf{Y}_i, \boldsymbol{\theta}]\,|\,\mathbf{V}_i, \mathbf{C}_i, \boldsymbol{\theta}] \\ &= c\boldsymbol{\Delta} + \mathbf{D}\mathbf{Z}_i^\top \boldsymbol{\Sigma}_i^{-1}(\widehat{\mathbf{y}}_i - \mathbf{X}_i\boldsymbol{\beta} - \kappa\mathbf{Z}_i\boldsymbol{\Delta}) + \frac{\widehat{\xi}_i \boldsymbol{\Lambda}_i\boldsymbol{\zeta}}{\sqrt{1+\boldsymbol{\zeta}^\top\boldsymbol{\Lambda}_i\boldsymbol{\zeta}}},\end{aligned}$$

where $\widehat{\mathbf{y}}_i = \mathbb{E}[\mathbf{Y}_i|\mathbf{V}_i, \mathbf{C}_i, \boldsymbol{\theta}]$ and $\widehat{\xi}_i = \mathbb{E}[\,\mathbb{E}[U_i^{-1/2} W_\Phi(U_i^{1/2}\bar{\boldsymbol{\lambda}}_i^\top \mathbf{y}_{0i})|\mathbf{Y}_i = \mathbf{y}_i, \boldsymbol{\theta}]\,|\,\mathbf{V}_i, \mathbf{C}_i, \boldsymbol{\theta}]$ depend on the censoring scheme of the ith subject.

6.2.6 Prediction of future observations

Interested in predicting values from the LMEC-ST model, we denote by $\mathbf{y}_{i,\text{obs}}$ the $n_{i,\text{obs}}$-vector of random variables corresponding to the given sample for subject i and by $\mathbf{y}_{i,\text{pred}}$ the vector of random variables of length $n_{i,\text{pred}}$ corresponding to the time points that we are interested in predicting. Let $\mathbf{X}_i^* = (\mathbf{X}_{i,\text{obs}}, \mathbf{X}_{i,\text{pred}})$ and $\mathbf{Z}_i^* = (\mathbf{Z}_{i,\text{obs}}, \mathbf{Z}_{i,\text{pred}})$ be the $n_i^* \times p$ and $n_i^* \times q$ design matrices corresponding to $\bar{\mathbf{y}}_i = (\mathbf{y}_{i,\text{obs}}^\top, \mathbf{y}_{i,\text{pred}}^\top)^\top$, with $n_i^* = n_{i,\text{obs}} + n_{i,\text{pred}}$.

Let $\mathbf{y}_{i,\text{obs}} = (\mathbf{y}_i^{o\top}, \mathbf{y}_i^{c\top})^\top$, where $\mathbf{y}_i^o$ and $\mathbf{y}_i^c$ denote the uncensored and censored components of $\mathbf{y}_{i,\text{obs}}$, respectively. To deal with the incomplete values existing in $\mathbf{y}_{i,\text{obs}}$, we use an imputation procedure that consists of replacing the censored values by $\widehat{\mathbf{y}}_i^c = \mathbb{E}[\mathbf{Y}_i^c \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}]$ the conditional mean obtained from the EM algorithm. The new vector of complete dataset will be denoted by $\mathbf{y}_{i,\text{obs}}^* = (\mathbf{y}_i^{o\top}, \widehat{\mathbf{y}}_i^{c\top})^\top$. The reason to use the imputation procedure is that it avoids computing truncated conditional expectations of the skew-t multivariate distribution originated by the censoring scheme. Hence, we have that

$$\bar{\mathbf{Y}}_i^* = (\mathbf{Y}_{i,\text{obs}}^{*\top}, \mathbf{Y}_{i,\text{pred}}^\top)^\top \sim \text{ST}_{n_i^*}(\mathbf{X}_i^*\boldsymbol{\beta} + \kappa\mathbf{Z}_i^*\boldsymbol{\Delta}, \boldsymbol{\Sigma}_i^*, \bar{\boldsymbol{\lambda}}_i^*, \nu), \tag{6.7}$$

such that $\boldsymbol{\Sigma}_i^* = \boldsymbol{\Omega}_i^* + \mathbf{Z}_i^*\mathbf{D}\mathbf{Z}_i^{*\top} = \begin{pmatrix} \boldsymbol{\Sigma}_{i11}^* & \boldsymbol{\Sigma}_{i12}^* \\ \boldsymbol{\Sigma}_{i21}^* & \boldsymbol{\Sigma}_{i22}^* \end{pmatrix}$, $\bar{\boldsymbol{\lambda}}_i^* = \boldsymbol{\Sigma}_i^{*-1/2}\mathbf{Z}_i^*\mathbf{D}\boldsymbol{\zeta}/(1+\boldsymbol{\zeta}^\top\boldsymbol{\Lambda}_i^*\boldsymbol{\zeta})^{1/2}$, with $\boldsymbol{\zeta} = \mathbf{D}^{-1/2}\boldsymbol{\lambda}$ and $\boldsymbol{\Lambda}_i = (\mathbf{D}^{-1} + \mathbf{Z}_i^{*\top}\boldsymbol{\Omega}_i^{*-1}\mathbf{Z}_i^*)^{-1}$. Using Proposition 6, we obtain that the distribution of $\mathbf{Y}_{i,\text{pred}}$ given $\mathbf{Y}_{i,\text{obs}}^* = \mathbf{y}_i$ belongs to the extended skew-t family, as follows

$$\mathbf{Y}_{i,\text{pred}} \mid (\mathbf{Y}_{i,\text{obs}}^* = \mathbf{y}_i) \sim \text{EST}_{n_{i,\text{pred}}}(\boldsymbol{\mu}_i^*, \boldsymbol{\Sigma}_{i22.1}^*/\nu^2(\mathbf{y}_i), \boldsymbol{\Sigma}_{i22.1}^{*1/2}\boldsymbol{\varphi}_i^{\text{pred}}, \tau_i^*, \nu + n_{i,\text{obs}}),$$

where $\boldsymbol{\mu}_i^* = \mathbf{X}_{i,\text{pred}}\boldsymbol{\beta} + \kappa\mathbf{Z}_{i,\text{pred}}\boldsymbol{\Delta} + \boldsymbol{\Sigma}_{i21}^*\boldsymbol{\Sigma}_{i11}^{*-1}(\mathbf{y}_i - \boldsymbol{\mu}_{i,\text{obs}})$, $\tau_i^* = \nu(\mathbf{y}_i)\tilde{\boldsymbol{\varphi}}_i^\top(\mathbf{y}_i - \boldsymbol{\mu}_{i,\text{obs}})$, $\boldsymbol{\mu}_{i,\text{obs}} = \mathbf{X}_{i,\text{obs}}\boldsymbol{\beta} + c\mathbf{Z}_{i,\text{obs}}\boldsymbol{\Delta}$, $\boldsymbol{\Sigma}_{i22.1}^* = \boldsymbol{\Sigma}_{i22}^* - \boldsymbol{\Sigma}_{i21}^*\boldsymbol{\Sigma}_{i11}^{*-1}\boldsymbol{\Sigma}_{i12}^*$, $\nu^2(\mathbf{y}_i) = (\nu + n_{i,\text{obs}})/(\nu + \delta(\mathbf{y}_i))$, $\tilde{\boldsymbol{\varphi}}_i = \boldsymbol{\varphi}_i^{\text{obs}} + \boldsymbol{\Sigma}_{i11}^{*-1}\boldsymbol{\Sigma}_{i12}^*\boldsymbol{\varphi}_i^{\text{pred}}$, and $\boldsymbol{\varphi}_i^* = (\boldsymbol{\varphi}_i^{\text{obs}\top}, \boldsymbol{\varphi}_i^{\text{pred}\top})^\top = \boldsymbol{\Sigma}_i^{*-1/2}\bar{\boldsymbol{\lambda}}_i^*$. Following Lachos *et al.* (2010), the minimum MSE predictor of future measurements of $\mathbf{Y}_i$ is the conditional

mean of $\mathbf{Y}_{i,\text{pred}}$ given $\mathbf{Y}^*_{i,\text{obs}} = \mathbf{y}_i$, which can be easily computed from (5.3), i.e.,

$$\widehat{\mathbf{y}}_{i,\text{pred}}(\boldsymbol{\theta}) = \mathbb{E}[\mathbf{Y}_{i,\text{pred}} \mid \mathbf{Y}^*_{i,\text{obs}} = \mathbf{y}_i, \boldsymbol{\theta}] = \boldsymbol{\mu}^*_i + \frac{\tilde{\eta}\boldsymbol{\Sigma}^*_{i22.1}\boldsymbol{\varphi}^{\text{pred}}_i}{\sqrt{1 + \boldsymbol{\varphi}^{\text{pred}\top}_i \boldsymbol{\Sigma}^*_{i22.1}\boldsymbol{\varphi}^{\text{pred}}_i}}, \tag{6.8}$$

with

$$\tilde{\eta} = \left(\frac{\nu + n_{i,\text{obs}}}{\nu + n_{i,\text{obs}} - 1}\right)\left(\frac{\nu + \delta(\mathbf{y}_i)}{\nu + n_{i,\text{obs}}}\right)^{1/2}\left(1 + \frac{\tilde{\tau}_i^2}{\nu + n_{i,\text{obs}}}\right)\frac{t(\tilde{\tau}_i; \nu + n_{i,\text{obs}})}{T(\tilde{\tau}_i; \nu + n_{i,\text{obs}})},$$

$$\delta(\mathbf{y}_i) = \delta(\mathbf{y}_i; \boldsymbol{\mu}_{i,\text{obs}}, \boldsymbol{\Sigma}^*_{i11}) \quad \text{and} \quad \tilde{\tau}_i = \tau^*_i/(1 + \boldsymbol{\varphi}^{\text{pred}\top}_i \boldsymbol{\Sigma}^*_{i22.1}\boldsymbol{\varphi}^{\text{pred}}_i)^{1/2}.$$

The prediction of $\mathbf{y}_{i,\text{pred}}$ can be obtained by substituting the ML estimates $\widehat{\boldsymbol{\theta}}$ into (6.8), leading to $\widehat{\mathbf{y}_{i,\text{pred}}} = \widehat{\mathbf{y}}_{i,\text{pred}}(\widehat{\boldsymbol{\theta}})$.

6.3 Simulation studies

This section presents the results from three simulation studies. The first study evaluates the asymptotic properties of parameter estimates obtained through the proposed method. The second simulation study assesses the robustness of our approach, specifically its capacity to handle perturbed and censored data simulated from the linear mixed-effect censored skew-normal (LMEC-SN) model. The final study examines the impact of model misspecification, where the data originate from a population with a distribution that differs from the one assumed for parameter estimation.

For the three simulation studies, we considered 300 Monte Carlo (MC) samples simulated from the model $\mathbf{Y}_i = (\beta_0 + b_{0i})\mathbf{1}_{10} + \beta_1\mathbf{x}_i + \boldsymbol{\xi}_i, i \in \{1, \ldots, n\}$, with n being the number of subjects (sample units). Here, $\mathbf{1}_{10}$ denotes a vector with all elements equal to one of length 10, $\mathbf{x}_i = (1, 2, \ldots, 10)^\top$ represents the vector of times at which the measurements were taken, and $\boldsymbol{\xi}_i$ is the error term, which was simulated from a multivariate distribution with zero mean and variance-covariance matrix $\boldsymbol{\Omega}_i = \sigma^2\mathbf{R}_i$, where $\sigma^2 = 0.25$ and $\mathbf{R}_i = \mathbf{R}_i(\boldsymbol{\phi})$ computed from the correlation function of an AR(2) process with parameters $\phi_1 = 0.60$ and $\phi_2 = -0.20$. The regression parameters were set to $\beta_0 = 1$ and $\beta_1 = 2$. The random effect b_{0i} was simulated from a distribution with scale parameter $D = 2$ and skewness parameter $\lambda = 3$.

6.3.1 Simulation I. Asymptotic properties

In this simulation study, we want to evaluate the asymptotic properties of the ML estimates obtained from the proposed EM-type algorithm. Therefore, we simulated samples consisting of n subjects with 10 repeated observations for each sample unit, for $n \in \{50, 100, 200, 400\}$. The observations were simulated considering that the error term

follows a multivariate Student-t distribution, $\boldsymbol{\xi}_i \overset{ind}{\sim} t_{10}(\mathbf{0}, \boldsymbol{\Omega}_i, \nu)$, and the random effect $b_{0i} \overset{iid}{\sim} ST(\kappa\Delta, D, \lambda, \nu)$, with κ selected such that $\mathbb{E}(b_{0i}) = 0$, the degrees of freedom $\nu = 6$, and the remaining parameters were set as mentioned above.

Additionally, we investigated the effect of censoring on the parameter estimates. Two scenarios were evaluated: one with an average of 10% left-censored observations and another with an average of 10% right-censored observations. The scenario without censoring was previously evaluated by Schumacher *et al.* (2021). An observation was considered censored if it fell within the interval $R_{ij} = (-\infty, 3.92)$ for left-censored cases, and $R_{ij} = (20, +\infty)$ for right-censored cases, where $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, 10\}$.

Table 18 presents the mean (MC-AV), median (MC-MD), and standard deviation (MC-SD) calculated from 300 MC samples. To assess the consistency of the method in approximating standard errors, we computed the average of the standard error estimates (denoted by IM-SE) and compared it with the MC-SD. The results for the left-censored observations are detailed in the upper table, while those for the right-censored observations are shown in the lower table. These findings indicate that the estimates for the regression coefficients (β_0, β_1) and the parameters related to the error term (σ^2, ϕ_1, ϕ) closely match the true parameter values, which are denoted in parentheses. Furthermore, a lower increase in the bias of estimates for β_0 was observed for the right-censored cases. Notably, the standard error estimates decrease as the sample size increases, aligning consistently with those derived via MC-SD. This consistency indicates that the inverse of the empirical information matrix, as detailed in Subsection 6.2.4, is reliable for approximating the SE of these parameters.

For the degrees of freedom ν, the estimates obtained through the mean and median approach the true value ($\nu = 6$) in samples of size $n \geqslant 100$, while for samples of size $n = 50$, the median (MC-MD) proved to be a more accurate estimator. Regarding the parameters associated with the random effect (D, λ), the estimates show some bias, which diminishes as the sample size increases. It is also worth noting that IM-SE provides a reliable estimate of the standard error for D, though it is less precise for λ. However, as expected, the discrepancy between IM-SE and MC-SD narrows with increasing sample size. Note also that the IM-SE for λ is greater for right-censored observations than for left-censored observations across all sample sizes; this could occur because the chosen values for the skewness parameter of the response vector vary from 0.80 to 0.87 for each individual, slightly more affecting right-censored observations than those that are left-censored.

The results previously discussed are summarized in the boxplots shown in Figure 29. We consider cases without censoring (No), with an average of 10% left-censored (Left) and 10% right-censored (Right) observations. It can be observed that the median of the estimates is close to the true parameter value (indicated by the red solid line) for the regression coefficients (β_0, β_1) and the parameters related to the error term $(\sigma^2, \phi_1, \phi_2, \nu)$.

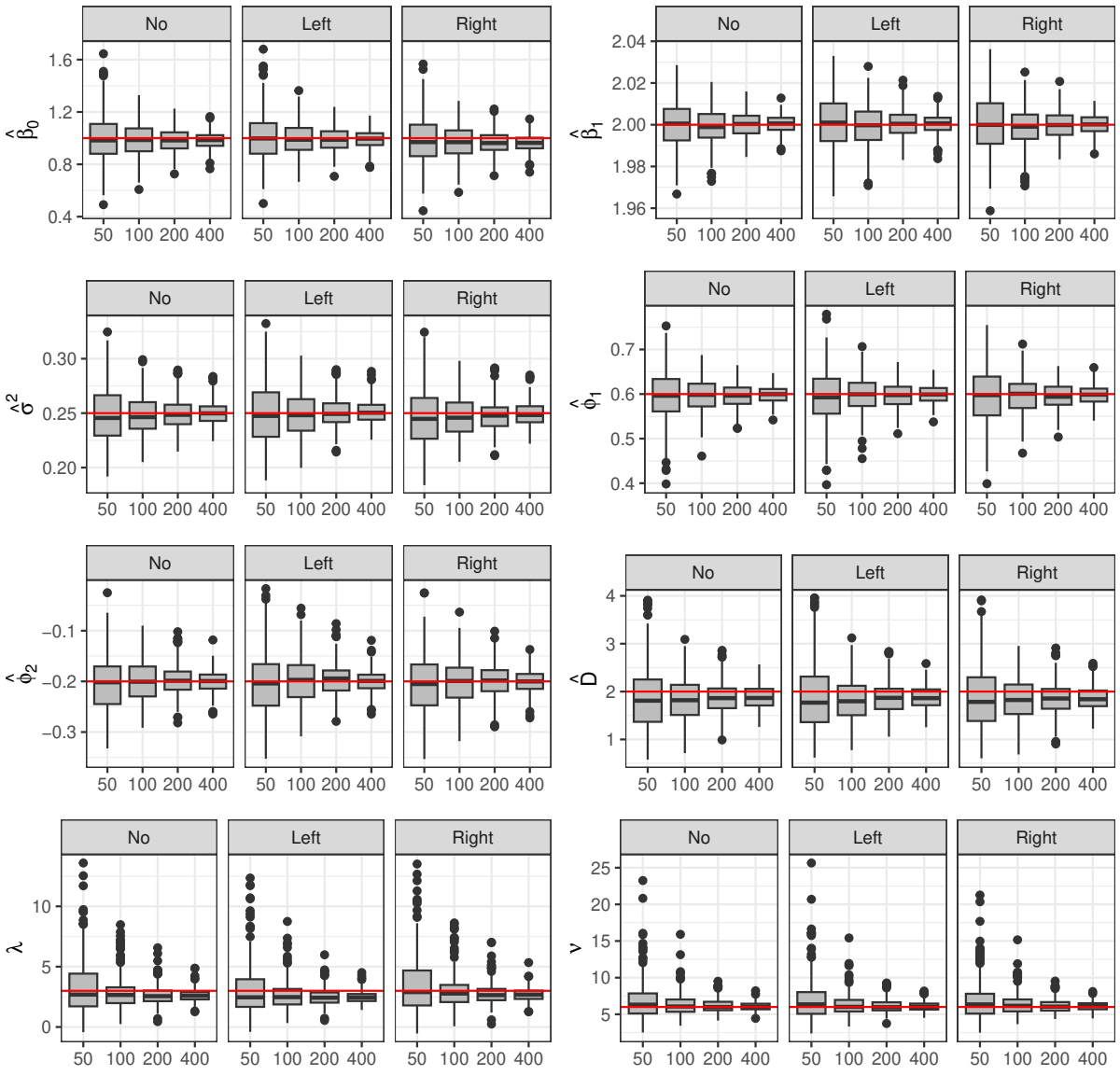
Table 18 – Simulation I. Summary statistics are presented for 300 Monte Carlo (MC) samples of sizes $n = 50, 100, 200$, and 400, each with an average of 10% left-censored (upper table) and right-censored (lower table) observations, simulated from the LMEC-ST model. The statistics MC-AV, MC-MD, and MC-SD refer to the mean, median, and standard deviation of the estimates, respectively. IM-SE denotes the average of the standard errors as detailed in Subsection 6.2.4. The true parameter values are provided in parentheses for reference.

n	cens. (%)	Measure	β_0 (1.00)	β_1 (2.00)	σ^2 (0.25)	ϕ_1 (0.60)	ϕ_2 (-0.20)	D (2.00)	λ (3.00)	ν (6.00)
Left-censored observations										
50	9.96	MC-AV	1.002	2.001	0.250	0.595	-0.204	1.886	3.039	6.973
		MC-MD	0.999	2.001	0.248	0.593	-0.204	1.773	2.466	6.371
		MC-SD	0.186	0.013	0.029	0.059	0.061	0.708	2.104	-
		IM-SE	0.199	0.015	0.030	0.064	0.066	0.747	187.722	-
100	9.96	MC-AV	0.999	1.999	0.249	0.598	-0.197	1.820	2.612	6.290
		MC-MD	0.987	2.000	0.248	0.600	-0.197	1.797	2.477	6.018
		MC-SD	0.131	0.010	0.019	0.040	0.044	0.460	1.148	-
		IM-SE	0.134	0.010	0.021	0.043	0.045	0.495	27.026	-
200	10.00	MC-AV	0.993	2.000	0.251	0.597	-0.197	1.858	2.487	6.156
		MC-MD	0.987	2.000	0.249	0.598	-0.194	1.869	2.429	5.975
		MC-SD	0.093	0.007	0.014	0.029	0.032	0.331	0.713	-
		IM-SE	0.093	0.007	0.015	0.030	0.032	0.346	9.348	-
400	10.01	MC-AV	0.994	2.000	0.251	0.598	-0.200	1.872	2.471	6.089
		MC-MD	0.996	2.001	0.250	0.599	-0.199	1.863	2.449	6.043
		MC-SD	0.069	0.005	0.011	0.020	0.022	0.237	0.459	-
		IM-SE	0.065	0.005	0.010	0.021	0.022	0.241	4.905	-
Right-censored observations										
50	10.16	MC-AV	0.978	2.000	0.247	0.596	-0.206	1.889	3.483	6.928
		MC-MD	0.970	2.000	0.245	0.598	-0.205	1.784	2.894	6.352
		MC-SD	0.178	0.014	0.028	0.064	0.062	0.692	2.381	-
		IM-SE	0.192	0.015	0.030	0.064	0.066	0.718	240.409	-
100	10.07	MC-AV	0.973	1.999	0.246	0.597	-0.201	1.823	2.928	6.353
		MC-MD	0.969	1.999	0.246	0.601	-0.199	1.824	2.775	6.160
		MC-SD	0.128	0.010	0.019	0.041	0.045	0.463	1.406	-
		IM-SE	0.129	0.010	0.021	0.043	0.045	0.479	41.722	-
200	10.06	MC-AV	0.967	2.000	0.248	0.594	-0.198	1.855	2.751	6.200
		MC-MD	0.963	2.000	0.248	0.594	-0.198	1.854	2.642	6.047
		MC-SD	0.086	0.007	0.014	0.029	0.031	0.320	0.839	-
		IM-SE	0.089	0.007	0.014	0.030	0.031	0.336	12.584	-
400	10.09	MC-AV	0.966	2.000	0.249	0.598	-0.201	1.857	2.686	6.134
		MC-MD	0.964	2.000	0.248	0.599	-0.200	1.840	2.686	6.120
		MC-SD	0.064	0.005	0.011	0.020	0.023	0.243	0.523	-
		IM-SE	0.062	0.005	0.010	0.021	0.022	0.234	6.233	-

However, for the parameters associated with the random effects, D and λ, the median underestimates the true parameter value across all sample sizes and types of censoring. Notably, the interquartile range decreases as the sample size increases for all parameters. The boxplots also reveal some outliers in the estimates for most parameters, with a notably higher number of outliers in the estimates of λ and ν.

Finally, we analyzed the mean square error (MSE) of the parameters estimated

Figure 29 – Simulation I. Boxplot of the estimates for the LMEC-ST model considering different sample sizes and types of censoring: without censoring (No), left-censored (Left), and right-censored (Right) observations. The red line denotes the true value.



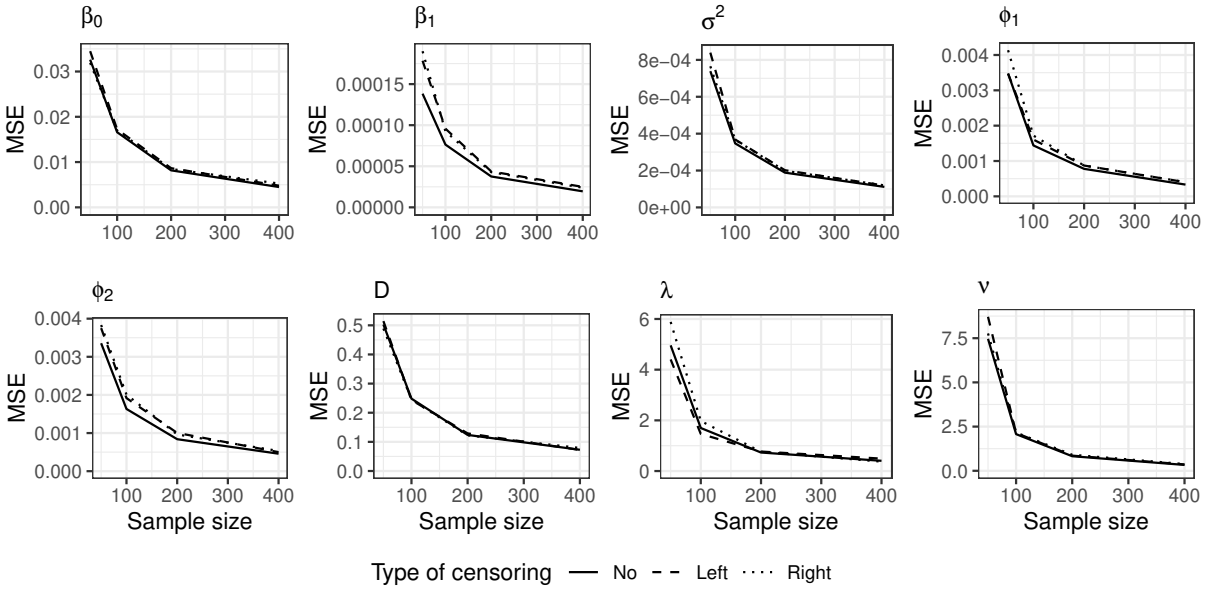
from the LMEC-ST model across samples of size $n = 50, 100, 200, 400$, with each subject i having $n_i = 10$ observations. This analysis aimed to provide empirical evidence regarding the consistency of the ML estimates. The MSE for each parameter is defined as follows

$$MSE_i = \frac{1}{m}\sum_{j=1}^{m}\left(\widehat{\theta}_i^{(j)} - \theta_i\right)^2,$$

where m is the number of MC samples, in this case, $m = 300$; $\widehat{\theta}_i^{(j)}$ denotes the estimate of the ith parameter obtained by the proposed EM-type algorithm from the jth MC sample; and θ_i represents the true value of the ith parameter of $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma^2, \phi_1, \phi_2, D, \lambda, \nu)^\top$.

The MSE results are displayed in Figure 30 for each sample size, with datasets without censoring represented by a solid line, those with left censoring by a dashed line, and right censoring scenarios by dotted lines. It is observable that the MSE tends to decrease toward zero as the sample size increases, substantiating the consistency of the ML estimates of the LMEC-ST model across the three evaluated scenarios.

Figure 30 – Simulation I. MSE of the estimates for the LMEC-ST model considering different sample sizes and types of censoring.



6.3.2 Simulation II. Robustness of the estimators

Aiming to provide evidence about the robustness of the estimates obtained by our proposal (LMEC-ST model), we simulated 300 MC samples, each containing $n = 300$ subjects with $n_i = 10$ observations per subject. The dataset was generated from the skew-normal distribution, assuming that the error term $\boldsymbol{\xi}_i \overset{ind}{\sim} N_{10}(\mathbf{0}, \Omega_i)$ and the random effects $b_{0i} \overset{iid}{\sim} SN(\kappa\Delta, D, \lambda)$, with $\kappa = -\sqrt{2/\pi}$. This model was developed by Mattos *et al.* (2022b), and the values for the remaining parameters were set as described in the introduction of this section.

After generating the data, each MC sample was perturbed by replacing the lowest 20 observations (1%) with the value $\rho = \min(\mathbf{y}) - \varphi SD(\mathbf{y})$, where $\min(\mathbf{y})$ and $SD(\mathbf{y})$ denote the minimum value and the standard deviation of the simulated values, and φ varied among $\{1, 2, 3, 4\}$. The case without perturbation, i.e., $\rho = y_{ij}$, was also considered for comparison. Subsequently, we censored an average of 10% of observations, considering as censored those that fell within the interval $R_{ij} = (20, +\infty)$, constituting the right-censored cases.

We fitted both the LMEC-SN and our proposed LMEC-ST model for comparison purposes. Table 19 displays the means of the estimates obtained by the LMEC-SN (SN)

and LMEC-ST (ST) models for different values of φ, with the case without perturbation denoted by $\varphi = 0$. In the unperturbed scenario, both models performed similarly, as the means of the estimates were close to the true parameter values for each model parameter. It was also noted that the estimate for ν was approximately equal to 95, suggesting that our algorithm tends to approximate the SN distribution.

For the other values of φ ($\{1, 2, 3, 4\}$), the estimates obtained by the LMEC-SN model showed a bias for all parameters, with the bias increasing as the perturbation level increased. Conversely, the mean of the estimates for the regression coefficients by the LMEC-ST model proved to be a better estimator than that obtained by the LMEC-SN model, exhibiting a smaller increase in bias for β_0; this could be related to the presence of a random effect in the intercept and the difficulty for the model to estimate the skewness parameter, λ, a challenge also observed in the right-censored datasets in Simulation Study I. Additionally, the mean of the estimates for the degrees of freedom ν tended to decrease as the perturbation level increased, demonstrating the method's mechanism for dealing with outliers.

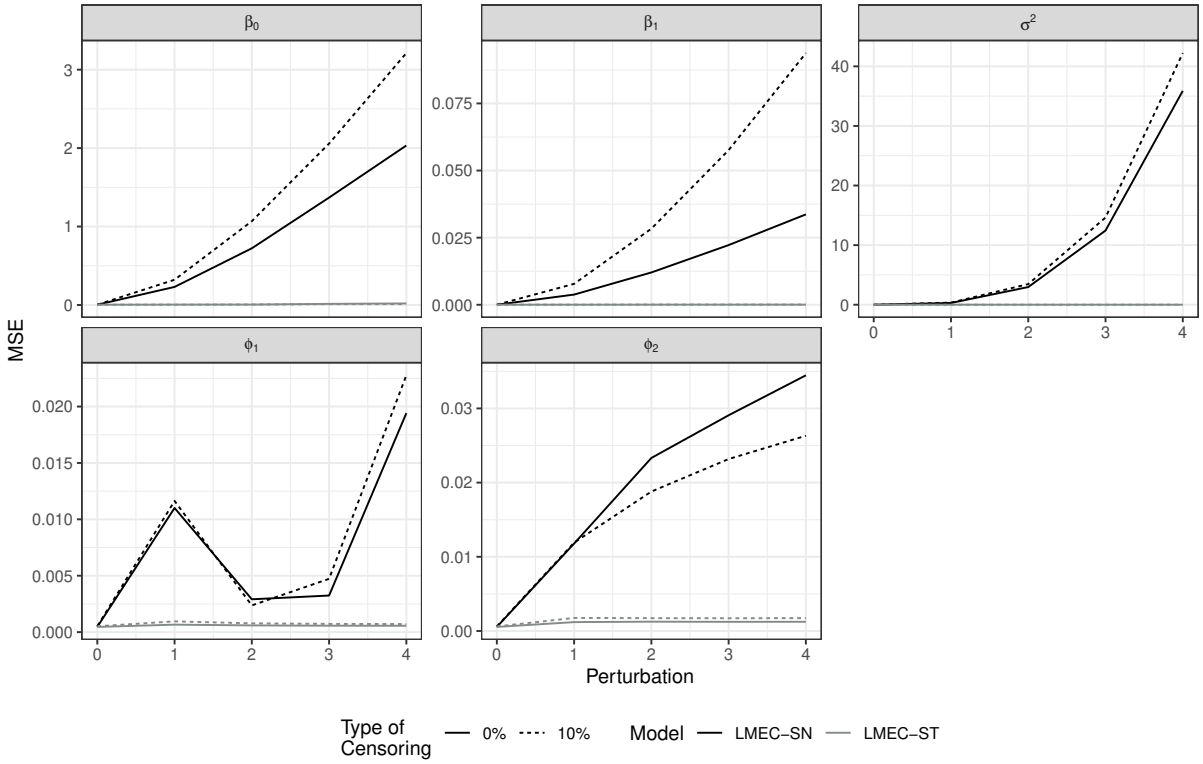
Table 19 – Simulation II. Mean of the estimates obtained after fitting the LMEC-SN (SN) and LMEC-ST (ST) models, based on 300 MC samples of size $n = 300$ simulated from the SN distribution considering 10% of right-censored and 1% of perturbed observations.

Param.	$\varphi = 0$		$\varphi = 1$		$\varphi = 2$		$\varphi = 3$		$\varphi = 4$	
	SN	ST	SN	ST	SN	ST	SN	ST	SN	ST
$\beta_0(1)$	0.999	0.997	0.437	1.003	-0.032	0.958	-0.433	0.921	-0.791	0.915
$\beta_1(2)$	2.000	2.000	2.088	1.998	2.168	1.996	2.240	1.995	2.306	1.994
$\sigma^2(0.25)$	0.249	0.245	0.840	0.255	2.108	0.248	4.073	0.245	6.746	0.245
$\phi_1(0.6)$	0.597	0.597	0.704	0.619	0.638	0.613	0.538	0.611	0.452	0.610
$\phi_2(-0.2)$	-0.202	-0.202	-0.095	-0.168	-0.066	-0.168	-0.050	-0.169	-0.039	-0.168
$D(2)$	1.826	1.808	1.103	0.665	2.869	0.642	3.917	0.636	4.525	0.633
$\lambda(3)$	2.571	2.652	-0.124	-0.332	-3.303	-0.341	-4.212	-0.349	-4.377	-0.344
ν	-	97.165	-	3.407	-	2.433	-	2.073	-	2.002

Figure 31 displays the MSE for the regression coefficients (β_0, β_1) and the parameters related to the error term $(\sigma^2, \phi_1, \phi_2)$ by perturbation level and by the model fitted. The black line represents the LMEC-SN model, and the grey line represents the LMEC-ST model. Results for datasets without censoring are shown with solid lines, while dashed lines indicate right-censored datasets. These plots reveal that the MSE computed from the LMEC-ST model is lower than those from the LMEC-SN model across all perturbation levels. It is also observable that the MSE increases with the level of perturbation for all models and levels of censoring, except for the estimates of ϕ_1 through the LMEC-SN model, which shows a decrease from $\varphi = 1$ to $\varphi = 2$, and then increase for $\varphi \geqslant 2$. Additionally, the censored datasets generally exhibit higher MSE values than the

uncensored datasets, except for the estimates of ϕ_2 through the LMEC-SN model, which displayed a different behavior for $\varphi \geqslant 2$. These results confirm that the heavy tails of the skew-t distribution enable our model to mitigate the effects of outliers, thereby providing a more robust method against atypical values.

Figure 31 – Simulation II. MSE estimated from the LMEC-SN and LMEC-ST models, based on 300 MC samples of size $n = 300$ simulated from the SN distribution considering 1% of perturbed observations and two censoring cases: without censoring and an average of 10% right-censored observations.



6.3.3 Simulation III. Model misspecification

To evaluate the performance of the proposed model and the impact of estimating parameters with the wrong distribution, we simulated 300 MC samples, each containing 300 subjects with $n_i = 10$ observations per subject. In this study, we explored various scenarios in which data were generated from the following distributions:

1. $\boldsymbol{\xi}_i \overset{ind}{\sim} N_{10}(\mathbf{0}, \boldsymbol{\Omega}_i)$, $b_{0i} \overset{iid}{\sim} SN(\kappa_1\Delta, D)$ and $\kappa_1 = -\sqrt{2/\pi}$;
2. $\boldsymbol{\xi}_i \overset{ind}{\sim} t_{10}(\mathbf{0}, \boldsymbol{\Omega}_i, \nu)$, $b_{0i} \overset{iid}{\sim} ST(\kappa_2\Delta, D, \nu)$, $\kappa_2 = -(\nu/\pi)^{1/2}\dfrac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}$ and $\nu = 6$; and,
3. $\boldsymbol{\xi}_i \overset{ind}{\sim} SL_{10}(\mathbf{0}, \boldsymbol{\Omega}_i, \nu_2)$, $b_{0i} \overset{iid}{\sim} SSL(\kappa_3\Delta, D, \nu_2)$, $\kappa_3 = -\sqrt{2/\pi}\dfrac{\nu}{\nu - 1/2}$ and $\nu = 1.15$.

Each scenario corresponds to data simulated from the skew-normal, skew-t, and skew-slash distribution. After generating the datasets, approximately 10% of the observations were censored. An observation was considered censored if it fell within the region $R_{ij} = (-\infty, r_{ij})$, with the threshold r_{ij} varying according to the simulated data distribution. Specifically, the detection limits were set at $r_{ij} = 4$ for the skew-normal distribution, $r_{ij} = 3.81$ for the skew-t distribution, and $r_{ij} = 3.57$ for the skew-slash distribution.

Table 20 displays the mean of the 300 estimates obtained after fitting the LMEC-ST (our proposal). For comparative analysis, we also fitted the LMEC-SN model across the three scenarios and the LMEC-t model specifically for the skew-t distribution data. For data simulated from the skew-normal distribution, the means of the estimates computed from both models are close to the true parameter values, suggesting that both distributions seem to fit the data equally well.

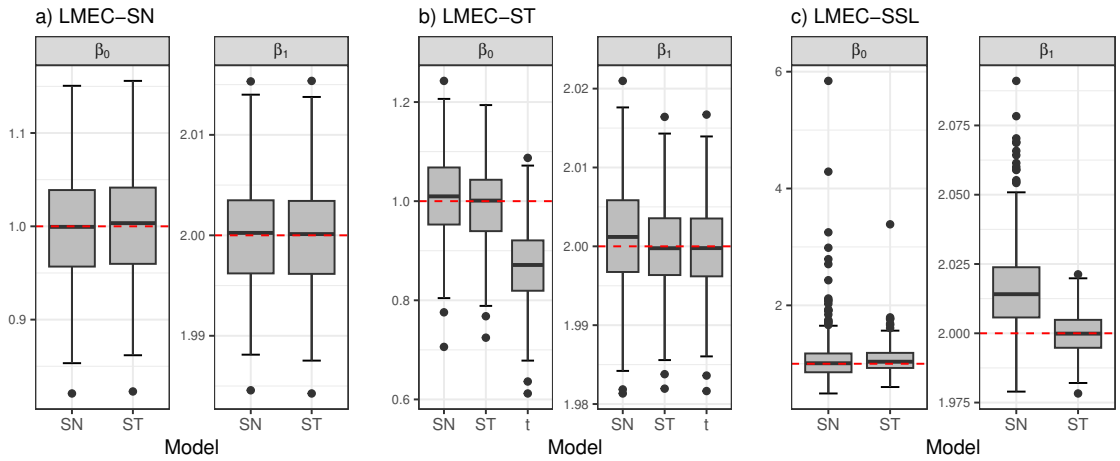
Table 20 – Simulation III. Mean of the estimates obtained after fitting the LMEC-SN (SN) and LMEC-ST (ST) models, based on 300 MC samples of size $n = 300$ simulated from different distributions considering 10% of left-censored observations.

	a) LMEC-SN		b) LMEC-ST			c) LMEC-SSL	
	SN	ST	SN	ST	t	SN	ST
$\beta_0(1)$	0.998	1.002	1.005	0.993	0.866	1.112	1.070
$\beta_1(2)$	2.000	2.000	2.001	2.000	2.000	2.017	2.000
$\sigma^2(0.25)$	0.250	0.245	0.376	0.251	0.253	1.414	0.454
$\phi_1(0.6)$	0.598	0.598	0.598	0.599	0.601	0.594	0.601
ϕ_2(-0.2)	-0.202	-0.202	-0.197	-0.198	-0.195	-0.210	-0.200
$D(2)$	1.897	1.748	3.359	1.857	0.924	13.106	3.791
$\lambda(3)$	2.695	2.352	4.216	2.467	-	7.561	2.841
ν	-	97.610	-	6.147	6.204	-	3.605

For the skew-t distribution data, the LMEC-t model tends to underestimate the true value of β_0 and compensates for the skewness in the distribution of the random effects by underestimating D. Conversely, the means of the estimates for β_0 and β_1 via the LMEC-SN model are close to the true values, although the interquartile range, as seen in the boxplot (Figure 32), is wider for the LMEC-SN estimates compared to those from the LMEC-ST.

Finally, when analyzing data simulated from the heavy-tailed skew-slash distribution (with $\nu = 1.15$), the LMEC-ST model outperforms the LMEC-SN model, which tends to overestimate the regression coefficients. In contrast, the LMEC-ST model's estimates are much closer to the true parameter values. Figure 32 further illustrates that estimates of β_0 and β_1 from the LMEC-SN model show significantly more outliers than those from the LMEC-ST model.

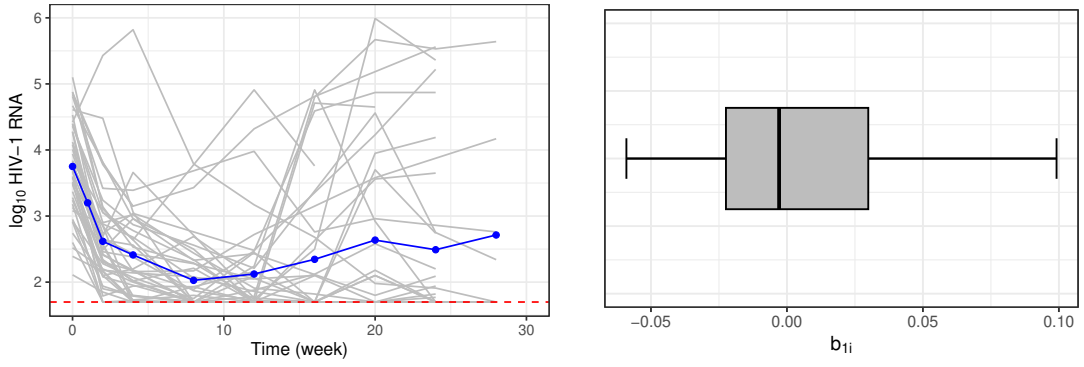
Figure 32 – Simulation III. Boxplot of the estimates for the regression coefficients obtained after fitting the LMEC-SN (SN) and LMEC-ST (ST) models to data simulated from different distributions incorporating 10% of left-censored observations.



6.4 Application: A5055 data

To illustrate the proposed model, we analyze the dataset A5055 from the AIDS clinical trial study by Wang (2013), which involved 44 HIV-1 infected patients. The dataset includes measurements of plasma viral load (in copies per milliliter) and CD4 cell counts, collected at approximately days 0, 7, 14, 28, 56, 84, 112, 140, and 168 of follow-up for each patient. This study focuses on the longitudinal trajectories of RNA viral load, converted to a log-base-10 scale and denoted by $\log_{10}$(RNA). The lower detection limit for RNA viral load is 50 copies/milliliter, corresponding to 1.698 on the log-base-10 scale, with 33.5% (106 out of 316) of measurements below this quantification limit (left-censored). Figure 33 (left panel) shows the trajectories of immunologic responses over time. For assessing the model's predictive performance, we partitioned the dataset into training and test sets, with the former comprising 304 observations and the latter 12 observations, which represent the last two data points from six randomly selected patients.

Figure 33 – A5055 data. (Left panel) Individual profiles for HIV viral load are displayed on a $\log_{10}$ scale. (Right panel) Boxplot of the random effects b_{1i} predicted from the t-LMEC model.



It is worth mentioning that this dataset was previously analyzed by Wang *et al.* (2018) using a multivariate t linear mixed-effects model for censored observations. Additionally, Lachos *et al.* (2019) fitted a multivariate linear mixed-effects model incorporating scale mixtures of normal distribution, while Mattos *et al.* (2022b) employed the LMEC-SN model. In this study, we propose to analyze the dataset using the following model

$$y_{ij} = \beta_0 + (\beta_1 + b_{1i})t_{ij} + \beta_2\sqrt{t_{ij}} + \beta_3 \text{CD4}^+ + \epsilon_{ij},$$

where y_{ij} represents $\log_{10}$(RNA) for subject i measured approximately at day_{ij}, $t_{ij} = \text{day}_{ij}/7$ (week), CD4^+ is the scaled CD4 variable, b_{1i} is the random effect for the ith subject, and ϵ_{ij} denotes the error term. To evaluate the necessity of considering a skewed distribution, we initially fitted an LMEC-t model with different correlation structures for the error term. Based on the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), the model that best fitted this dataset included an AR(1) correlation process. Figure 33 (right panel) displays the boxplot of the random effects predicted from the LMEC-t model with an AR(1) correlation, which appears to correspond to a right-skewed distribution.

Given the results obtained earlier, we fitted a model with skew-t distributed random effects and t distributed error terms, which suggests that the response variable itself follows a skew-t distribution. For comparative analysis, we applied the LMEC-ST model using five different correlation structures for the error terms. These structures included the unstructured (UNC) model, the damped exponential correlation (DEC), the compound symmetry correlation (CS), and the autoregressive models of order 1 and 2 (AR(1) and AR(2)). The log-likelihood functions evaluated at the maximum likelihood (ML) estimates and the information criteria (AIC and BIC) are displayed in Table 21. Here, we observe that the lowest values for these statistics were obtained with the AR(1) correlation, mirroring the results found for the LMEC-t model.

Table 21 – A5055 data. Information criteria were obtained after fitting the LMEC-ST model with different correlation structures.

Information criterion	Correlation structure				
	UNC	DEC	CS	AR(1)	AR(2)
$\ell(\widehat{\boldsymbol{\theta}})$	-312.406	-279.012	-292.648	-278.755	-279.407
AIC	640.811	578.023	603.296	**575.510**	578.813
BIC	670.548	615.194	636.749	**608.963**	615.984

Additionally, we fitted the LMEC-SN and LMEC-N models, considering the AR(1) correlation for the error term. Table 22 presents the parameter estimates and their standard errors (SE), approximated through the inverse of the empirical information matrix as described in Subsection 6.2.4. This table also includes the information criteria (AIC and BIC) along with the log-likelihood evaluated at the ML estimates. Note that the

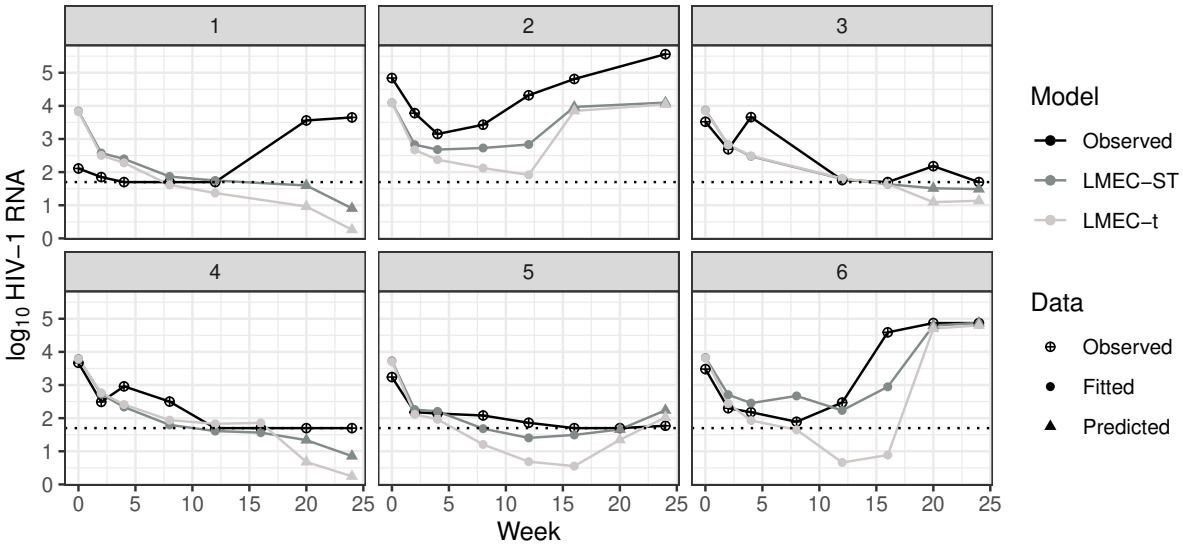
estimates for β_0 and β_3 are similar between the LMEC-ST and LMEC-t models. However, for the skew-normal and normal models, these estimates are lower compared to those obtained through heavy-tailed distributions. The estimated SEs are nearly identical across all models. The estimates for β_1 are around 0.120, while β_2 is roughly estimated at -0.95, with the greater value estimated in the LMEC-N model. In the LMEC-ST and LMEC-t models, the estimated variance of the error term, σ^2, was 0.10. This estimate is notably lower, specifically half, compared to the variance of 0.20 estimated in both the skew-normal and normal models. The estimate for the skewness parameter λ in the LMEC-ST model was 1.668, corroborating the findings shown in Figure 33. Additionally, the AIC favors our proposal, as it shows the lowest values for this statistic, while the BIC is nearly identical for both heavy-tailed models, confirming the adequacy of the LMEC-ST model.

Table 22 – A5055 data. Parameter estimates (Est.) and standard error (SE) obtained after fitting some LMEC models with AR(1) structure.

Parameter	LMEC-ST		LMEC-SN		LMEC-t		LMEC-N	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
β_0	3.734	0.131	3.619	0.131	3.709	0.129	3.582	0.139
β_1	0.129	0.027	0.119	0.027	0.110	0.023	0.115	0.023
β_2	-0.969	0.077	-0.923	0.077	-0.953	0.086	-0.910	0.091
β_3	-0.344	0.098	-0.474	0.098	-0.365	0.118	-0.482	0.102
σ^2	0.102	0.022	0.209	0.022	0.113		0.214	
ϕ_1	0.866	0.030	0.860	0.030	0.863		0.857	
D	0.002	0.002	0.004	0.002	0.002		0.003	
λ	1.668	0.001	1.009	0.001	-		-	
ν	2.946		-		3.633		-	
$\ell(\widehat{\boldsymbol{\theta}})$	-278.755		-290.640		-281.161		-291.201	
AIC	**575.510**		597.280		578.322		596.401	
BIC	608.963		627.016		**608.058**		622.421	

Figure 34 displays the observed values of $\log_{10}$(RNA) against the week of observation for six patients, depicted by black solid lines. The fitted and predicted values are represented by solid grey lines for the LMEC-ST model (dark grey) and the LMEC-t model (light grey), with the last two points indicating predictions. The dotted line marks the detection limit ($\log_{10}(50)$) for censored observations. It is evident that both models strive to match the dynamics of the observed data, with the LMEC-ST model demonstrating superior performance. To quantitatively assess the predictive capabilities of our model, the mean squared prediction error (MSPE) was calculated, excluding censored observations. The results show that the MSPE for the LMEC-ST model is significantly lower at 1.863 compared to 2.844 for the LMEC-t model. This suggests that the LMEC-ST model, which accounts for skewness and heavier tails, is better suited for fitting this dataset.

Figure 34 – A5055 data. Observed viral loads on a $\log_{10}$ scale (black solid line) for six subjects, alongside fitted trajectories for the LMEC-ST (dark grey) and LMEC-t (light grey) models, both incorporating the AR(1) structure. The final two observations were reserved to assess the predictive capabilities of the models.



6.5 Remarks

This chapter introduced a robust linear mixed-effects regression model for handling censored and/or missing data using the skew-t distribution. This model extends the works of Schumacher *et al.* (2021), Lachos *et al.* (2019), and Mattos *et al.* (2022b), among many others. A key advantage of our proposed LMEC-ST model is its ability to mitigate the negative impacts of distributional misspecification and outliers on parameter estimation, as evidenced by our simulation studies. The skew-t class of distributions also offers a convenient framework for implementing the EM-type algorithm since its moments and probabilities can be calculated from the expressions derived by Galarza *et al.* (2021c) and implemented in the `MomTrunc` package. Furthermore, the effectiveness of our method was demonstrated through the analysis of a real dataset, particularly in managing multiple left-censored observations in the presence of asymmetric heavy-tailed data.

Future extensions of this research could include several promising directions. One possibility is developing a multiple-outcome skew-normal and skew-t linear mixed-effects model, extending the works of Wang *et al.* (2018) and Lachos *et al.* (2019). Another direction could involve generalizing the LMEC-ST model to a nonlinear framework, extending the methodologies introduced by Matos *et al.* (2013b). Additionally, exploring a multiple-outcome skew-t nonlinear mixed-effects model could further advance the findings of Lin & Wang (2017).

7 Final Considerations

This section describes the technical production developed in this thesis.

7.1 Published papers

- **Valeriano, K. A.**, Galarza, C. E., and Matos, L. A. (2023). Moments and random number generation for the truncated elliptical family of distributions. *Statistics and Computing*, DOI: 10.1007/s11222-022-10200-4.
- **Valeriano, K. A.**, Galarza, C. E., Matos, L. A., and Lachos, V. H. (2023). Likelihood-based inference for the multivariate skew-t regression with censored or missing responses. *Journal of Multivariate Analysis*, DOI: 10.1016/j.jmva.2023.105174.
- **Valeriano, K. A.**, Schumacher, F. L., Galarza, C. E., and Matos, L. A. (2024). Censored autoregressive regression models with Student-t innovations. *Canadian Journal of Statistics*, DOI: 10.1002/cjs.11804.

7.2 R packages

The research conducted in this thesis resulted in the development of two new `R` packages, `relliptical` and `RcppCensSpatial`, while in other instances, the proposed methodologies were incorporated into existing packages such as `ARCensReg` and `skewlmm` (currently in progress). A brief description of each package is provided below:

- `relliptical`: The Truncated Elliptical Family of Distributions

  It offers random numbers generation from members of the truncated multivariate elliptical family of distribution, such as the truncated versions of the Normal, Student-t, Laplace, Pearson VII, Slash, and Logistic, among others. Particular distributions can be provided by specifying the density generating function. It also computes the first two moments (covariance matrix as well) for some particular distributions.

- `RcppCensSpatial`: Spatial Estimation and Prediction for Censored/Missing Responses

  It provides functions to estimate parameters in linear spatial models with censored/missing responses via the Expectation-Maximization (EM), the Stochastic Approximation EM (SAEM), or the Monte Carlo EM (MCEM) algorithm. These algorithms are widely used to compute the maximum likelihood (ML) estimates in

problems with incomplete data. This package also approximates the standard error of the estimates using the Louis method. Moreover, it has a function that performs spatial prediction in new locations.

- `ARCensReg`: Fitting Univariate Censored Linear Regression Model with Autoregressive Errors

  It fits a univariate left, right, or interval-censored linear regression model with autoregressive errors, considering the normal or the Student-t distribution for the innovations. It provides estimates and standard errors of the parameters, predicts future observations, and supports missing values on the dependent variable.

- `skewlmm`: Scale Mixture of Skew-Normal Linear Mixed Models

  It fits a scale mixture of skew-normal linear mixed models using an expectation–maximization (EM) type algorithm, including some possibilities for modeling the within-subject dependence and censoring or missing observations in the response variable.

7.3 Conclusions

This thesis explored a frequentist approach to modeling censored and missing responses using the Student-t and skew-t distributions. These models extend the work of Schumacher *et al.* (2017) within the context of time series and Mattos *et al.* (2022b) in linear mixed-effects models. For multivariate regression, we proposed a model featuring skew-t errors, building upon the skew-normal model developed by Galarza *et al.* (2022b). These distributions are particularly attractive due to their ability to accommodate heavy tails and skewness. Moreover, this research accounts for the missingness mechanism, assuming it to be missing at random (MAR).

Additionally, we developed a method for generating random numbers from the family of truncated elliptical distributions, extending the work of Ho *et al.* (2012). This method was driven by the necessity to simulate from truncated distributions, which facilitate the estimation procedures for the models discussed. For maximum likelihood (ML) estimation, the EM and the Stochastic Approximation of the EM (SAEM) algorithms were employed, accommodating both censored and missing responses.

Several simulation studies demonstrated the asymptotic properties of the ML estimates, validated the reliability of standard error calculated by the Louis method or the empirical information matrix, and confirmed the robustness of our proposed models in scenarios involving perturbed observations. We also assessed model performance under misspecification, particularly when data arise from heavy-tailed distributions. The practical applicability of these models was further illustrated using real datasets. Each model has

been implemented in an R package, except the multivariate skew-t regression model (Chapter 5).

Moreover, this thesis conducted a comparative analysis of the EM, MCEM, and SAEM algorithms for estimating the parameters of a Gaussian spatial censored linear (SCL) model (Lachos *et al.*, 2017; Ordoñez *et al.*, 2018). It also proposed a new package, `RcppCensSpatial`, which estimates parameters and approximates the standard errors more efficiently than the previously established R package `CensSpatial` (Ordoñez *et al.*, 2020). This new package offers enhanced computational speed, improving its predecessor.

7.4 Future research

Future extensions in Chapter 2 may consider a broader class of density generating functions beyond the strictly decreasing ones currently used in the sampling method. Additionally, alternative sampling methods for the truncated elliptical family could be explored, such as IA^{2}RMS (Martino *et al.*, 2015a) and the fast universal self-tuned sampler (FUSS) (Martino *et al.*, 2015b) within a Gibbs framework. IA^{2}RMS is noteworthy for generating asymptotically independent samples while keeping computational costs comparable to ARS and ARMS. FUSS, demonstrated through simulation studies, offers faster sampling than some well-known MCMC methods for specific bivariate distributions, with high accuracy and virtually independent sample generation.

In Chapter 4, future research directions include relaxing the assumption that the first p observations are fully observed in order to fit a CAR$t(p)$ model. Additionally, a natural and interesting path for further research is extending this model to a multivariate framework.

Further work in Chapter 5 might include generalizing the skew-t censored regression model to a nonlinear context or developing a finite mixture of censored data models using the multivariate skew-t distribution. This would be an extension of the skew-normal finite mixture model proposed by De Alencar *et al.* (2021).

Extensions of Chapter 6 could involve developing a multiple-outcome skew-normal and skew-t linear mixed-effects model, building on the contributions of Wang *et al.* (2018) and Lachos *et al.* (2019). Another direction may involve adapting the LMEC-ST model to a nonlinear framework, extending methodologies like those introduced by Matos *et al.* (2013b). Moreover, exploring a multiple-outcome skew-t nonlinear mixed-effects model could advance the findings of Lin & Wang (2017).

Bibliography

Alpuim, T. & El-Shaarawi, A. (2008). On the efficiency of regression analysis with AR(p) errors. *Journal of Applied Statistics*, **35**(7), 717–737.

Andersen, M., Goedman, R., Grothendieck, G., Højsgaard, S., Pinkus, A. & Mazur, G. (2020). Ryacas: R interface to the YACAS Computer Algebra System.

Arellano-Valle, R. B. & Genton, M. G. (2005). On fundamental skew distributions. *Journal of Multivariate Analysis*, **96**(1), 93–116.

Arellano-Valle, R. B. & Genton, M. G. (2010a). Multivariate unified skew-elliptical distributions. *Chilean Journal of Statistics*, **1**(1), 17–33.

Arellano-Valle, R. B. & Genton, M. G. (2010b). Multivariate extended skew-t distributions and related families. *Metron*, **68**(3), 201–234.

Azzalini, A. & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**(2), 367–389.

Bandyopadhyay, D., Lachos, V. H., Castro, L. M. & Dey, D. K. (2012). Skew-normal/independent linear mixed models for censored responses with applications to HIV viral loads. *Biometrical Journal*, **54**(3), 405–425.

Barroso, F. J. C., Garcia-Perez, C. & Prieto-Alaiz, M. (2020). Modelling income distribution using the log student'st distribution: New evidence for european union countries. *Economic Modelling*, **89**, 512–522.

Bertolacci, M. (2019). armspp: Adaptive rejection metropolis sampling (ARMS) via 'Rcpp'. *R package version*. 0.0.2.

Besag, J. & Green, P. J. (1993). Spatial statistics and bayesian computation. *Journal of the Royal Statistical Society: Series B (Methodological)*, **55**(1), 25–37.

Booth, J. G. & Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **61**(1), 265–285.

Box, G. E., Jenkins, G. M., Reinsel, G. C. & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Brent, R. P. (2013). *Algorithms for minimization without derivatives*. Courier Corporation.

Cabral, C. R. B., Lachos, V. H. & Prates, M. O. (2012). Multivariate mixture modeling using skew-normal independent distributions. *Computational Statistics & Data Analysis*, **56**(1), 126–142.

Celeux, G. (1985). The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational statistics quarterly*, **2**, 73–82.

Damien, P. & Walker, S. G. (2001). Sampling truncated normal, beta, and gamma densities. *Journal of Computational and Graphical Statistics*, **10**(2), 206–215.

De Abreu, D., Martins, A. C., da Silva, K. M., Nunes, A. C., Miranda, Y. C., Battistelli, A. A., de Oliveira, R. P., Camilo, R. & Achcar, J. A. (2022). Identifying water factors that are related to ammonia nitrogen concentrations in columbia river using a reversed hazard exponential model. *River Research and Applications*, **38**(2), 256–266.

De Alencar, F. H., Galarza, C. E., Matos, L. A. & Lachos, V. H. (2021). Finite mixture modeling of censored and missing data using the multivariate skew-normal distribution. *Advances in Data Analysis and Classification*, pages 1–37.

Delyon, B., Lavielle, M., Moulines, E. *et al.* (1999). Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Statistics*, **27**(1), 94–128.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, **39**(1), 1–22.

Diggle, P. & Ribeiro, P. (2007). *Model-based Geostatistics*. Springer.

Diggle, P., Diggle, P. J., Heagerty, P., Liang, K.-Y., Zeger, S. *et al.* (2002). *Analysis of longitudinal data*. Oxford university press.

Dunn, P. K. & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, **5**(3), 236–244.

Fang, K. W. (2018). *Symmetric multivariate and related distributions*. CRC Press.

Feigelson, E. (2014). astrodatR: Astronomical Data. *R Package Version 0.1 URL http://CRAN. R-project.org/package=astrodatR*.

Feigelson, E. D. & Babu, G. J. (2012). *Modern statistical methods for astronomy: with R applications*. Cambridge University Press.

Ferreira, C. S., Lachos, V. H. & Bolfarine, H. (2016). Likelihood-based inference for multivariate skew scale mixtures of normal distributions. *AStA Advances in Statistical Analysis*, **100**, 421–441.

Fridley, B. L. & Dixon, P. (2007). Data augmentation for a bayesian spatial model involving censored observations. *Environmetrics: The official journal of the International Environmetrics Society*, **18**(2), 107–123.

Gaetan, C. & Guyon, X. (2010). *Spatial statistics and modeling*, volume 90. Springer.

Galarza, C. E., Lachos, V. H. & Bandyopadhyay, D. (2017). Quantile regression in linear mixed models: a stochastic approximation EM approach. *Statistics and its Interface*, **10**(3), 471.

Galarza, C. E., Kan, R. & Lachos, V. H. (2021a). Package 'MomTrunc'. *R package version*.

Galarza, C. E., Lachos, V. H. & Bourguignon, M. (2021b). A skew-t quantile regression for censored and missing data. *Stat*, **10**(1), e379.

Galarza, C. E., Lin, T.-I., Wang, W.-L. & Lachos, V. H. (2021c). On moments of folded and truncated multivariate Student-t distributions based on recurrence relations. *Metrika*, **84**, 825–850.

Galarza, C. E., Matos, L. A., Castro, L. M. & Lachos, V. H. (2022a). Moments of the doubly truncated selection elliptical distributions with emphasis on the unified multivariate skew-t distribution. *Journal of Multivariate Analysis*, **189**, 104944.

Galarza, C. E., Matos, L. A. & Lachos, V. H. (2022b). An EM algorithm for estimating the parameters of the multivariate skew-normal distribution with censored responses. *METRON*, **80**, 231–253.

Garay, A. M., Castro, L. M., Leskow, J. & Lachos, V. H. (2017). Censored linear regression models for irregularly observed longitudinal data using the multivariate-t distribution. *Statistical Methods in Medical Research*, **26**(2), 542–566.

Gareth, J., Daniela, W., Trevor, H. & Robert, T. (2013). *An introduction to statistical learning: with applications in R*. Spinger.

Gelfand, A. E. & Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**(410), 398–409.

Gelfand, A. E., Smith, A. F. & Lee, T.-M. (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association*, **87**(418), 523–532.

Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-6**(6), 721–741.

Genton, M. G. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality (1st ed.)*. Chapman and Hall/CRC.

Gilks, W. R. & Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **41**(2), 337–348.

Gilks, W. R., Best, N. G. & Tan, K. K. (1995). Adaptive rejection metropolis sampling within Gibbs sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **44**(4), 455–472.

Gómez, E., Gómez-Villegas, M. & Marín, J. M. (1998). A multivariate generalization of the power exponential family of distributions. *Communications in Statistics-Theory and Methods*, **27**(3), 589–600.

Hadfield, J. (2022). MCMCglmm: MCMC Generalised Linear Mixed Models. *R package version*. 2.34.

Ho, H. J., Lin, T.-I., Chen, H.-Y. & Wang, W.-L. (2012). Some results on the truncated multivariate t distribution. *Journal of Statistical Planning and Inference*, **142**(1), 25–40.

Hoffman, H. J. & Johnson, R. E. (2015). Pseudo-likelihood estimation of multivariate normal parameters in the presence of left-censored data. *Journal of Agricultural, Biological, and Environmental Statistics*, **20**(1), 156–171.

IPCS (1986). (International Programme on Chemical Safety). Environmental Health Criteria 54: Ammonia. *United National Environment Programme, International Labour Organisation, World Health Organization*.

Jank, W. (2006). Implementing and diagnosing the stochastic approximation EM algorithm. *Journal of Computational and Graphical Statistics*, **15**(4), 803–829.

Kalliovirta, L. (2012). Misspecification tests based on quantile residuals. *The Econometrics Journal*, **15**(2), 358–393.

Kan, R. & Robotti, C. (2017). On moments of folded and truncated multivariate normal distributions. *Journal of Computational and Graphical Statistics*, **26**(4), 930–934.

Kotecha, J. H. & Djuric, P. M. (1999). Gibbs sampling approach for generation of truncated multivariate gaussian random variables. In *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258)*, volume 3, pages 1757–1760. IEEE.

Kotz, S. & Nadarajah, S. (2004). *Multivariate t-distributions and their applications*. Cambridge University Press.

Kuhn, E. & Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Computational statistics & data analysis*, **49**(4), 1020–1038.

Lachos, V. H., Ghosh, P. & Arellano-Valle, R. B. (2010). Likelihood-based inference for skew-normal independent linear mixed models. *Statistica Sinica*, **20**(1), 303.

Lachos, V. H., Matos, L. A., Barbosa, T. S., Garay, A. M. & Dey, D. K. (2017). Influence diagnostics in spatial models with censored response. *Environmetrics*, **28**(7).

Lachos, V. H., A. Matos, L., Castro, L. M. & Chen, M.-H. (2019). Flexible longitudinal linear mixed models for multiple censored responses data. *Statistics in medicine*, **38**(6), 1074–1102.

Lavielle, M. (2014). *Mixed effects models for the population approach: models, tasks, methods and tools*. CRC press.

Lin, T.-I. (2010). Robust mixture modeling using multivariate skew t distributions. *Statistics and Computing*, **20**(3), 343–356.

Lin, T. I. & Lee, J. C. (2007). Bayesian analysis of hierarchical linear mixed modeling using the multivariate t distribution. *Journal of Statistical Planning and Inference*, **137**(2), 484–495.

Lin, T.-I. & Wang, W.-L. (2017). Multivariate-nonlinear mixed models with application to censored multi-outcome aids studies. *Biostatistics*, **18**(4), 666–681.

Liu, C. & Rubin, D. B. (1994). The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika*, **81**(4), 633–648.

Liu, J., Kumar, S. & Palomar, D. P. (2019). Parameter estimation of heavy-tailed AR model with missing data via stochastic EM. *IEEE Transactions on Signal Processing*, **67**(8), 2159–2172.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, **44**(2), 226–233.

Marchenko, Y. V. & Genton, M. G. (2010). Multivariate log-skew-elliptical distributions with applications to precipitation data. *Environmetrics: The official journal of the International Environmetrics Society*, **21**(3-4), 318–340.

Martino, L., Read, J. & Luengo, D. (2015a). Independent doubly adaptive rejection Metropolis sampling within Gibbs sampling. *IEEE Transactions on Signal Processing*, **63**(12), 3123–3138.

Martino, L., Yang, H., Luengo, D., Kanniainen, J. & Corander, J. (2015b). A fast universal self-tuned sampler within Gibbs sampling. *Digital Signal Processing*, **47**, 68–83.

Massuia, M. B., Cabral, C. R. B., Matos, L. A. & Lachos, V. H. (2015). Influence diagnostics for Student-t censored linear regression models. *Statistics*, **49**, 1074–1094.

Matos, L. A., Lachos, V. H., Balakrishnan, N. & Labra, F. V. (2013a). Influence diagnostics in linear and nonlinear mixed-effects models with censored data. *Computational Statistics & Data Analysis*, **57**(1), 450–464.

Matos, L. A., Prates, M. O., Chen, M.-H. & Lachos, V. H. (2013b). Likelihood-based inference for mixed-effects models with censored response using the multivariate-t distribution. *Statistica Sinica*, **23**(3), 1323–1345.

Matos, L. A., Castro, L. M. & Lachos, V. H. (2016). Censored mixed-effects models for irregularly observed repeated measures with applications to HIV viral loads. *Test*, **25**(4), 627–653.

Matos, L. A., Lachos, V. H., Lin, T.-I. & Castro, L. M. (2019). Heavy-tailed longitudinal regression models for censored data: a robust parametric approach. *Test*, **28**(3), 844–878.

Mattos, T. B., Lachos, V. H., Castro, L. M. & Matos, L. A. (2022a). Extending multivariate Student's-t semiparametric mixed models for longitudinal data with censored responses and heavy tails. *Statistics in Medicine*, **41**(19), 3696–3719.

Mattos, T. B., Matos, L. A. & Lachos, V. H. (2022b). Likelihood-based inference for linear mixed-effects models with censored response using skew-normal distribution. In *Innovations in Multivariate Statistical Modeling: Navigating Theoretical and Multidisciplinary Domains*, pages 23–43. Springer.

Mattos, T. d. B., Garay, A. M. & Lachos, V. H. (2018). Likelihood-based inference for censored linear regression models with scale mixtures of skew-normal distributions. *Journal of Applied Statistics*, **45**(11), 2039–2066.

McLachlan, G. J. & Krishnan, T. (2008). *The EM algorithm and extensions*. John Wiley & Sons.

Meilijson, I. (1989). A fast improvement to the em algorithm on its own terms. *Journal of the Royal Statistical Society: Series B (Methodological)*, **51**(1), 127–138.

Meyer, R., Cai, B. & Perron, F. (2008). Adaptive rejection Metropolis sampling using Lagrange interpolation polynomials of degree 2. *Computational Statistics & Data Analysis*, **52**(7), 3408–3423.

Morán-Vásquez, R. A. & Ferrari, S. L. (2021). New results on truncated elliptical distributions. *Communications in Mathematics and Statistics*, **9**, 299–313.

Muirhead, R. J. (2009). *Aspects of multivariate statistical theory*, volume 197. John Wiley & Sons.

Nduka, U. C. (2018). EM-based algorithms for autoregressive models with t-distributed innovations. *Communications in Statistics-Simulation and Computation*, **47**(1), 206–228.

Neal, R. M. (2003). Slice sampling. *The Annals of Statistics*, **31**(3), 705–767.

Olivari, R. C., Zhong, K., Garay, A. M. & Lachos, V. H. (2022). ARpLMEC: Censored mixed-effects models with different correlation structures. *R package version*. 2.4.1.

Ordoñez, A., Galarza, C. E. & Lachos, V. H. (2020). Package 'CensSpatial'. *R package version*.

Ordoñez, J. A., Bandyopadhyay, D., Lachos, V. H. & Cabral, C. R. (2018). Geostatistical estimation and prediction for censored responses. *Spatial statistics*, **23**, 109–123.

Pan, Y. & Pan, J. (2020). roptim: An R Package for General Purpose Optimization with C++. *R package version*.

Paxton, W. B., Coombs, R. W., McElrath, M. J., Keefer, M. C., Hughes, J., Sinangil, F., Chernoff, D., Demeter, L., Williams, B., Corey, L. *et al.* (1997). Longitudinal analysis of quantitative virologic measures in human immunodeficiency virus-infected subjects with $\geqslant$ 400 CD4 lymphocytes: implications for applying measurements to individual patients. *Journal of Infectious Diseases*, **175**(2), 247–254.

Pinheiro, J. C. & Bates, D. M. (2000). Linear mixed-effects models: basic concepts and examples. *Mixed-effects models in S and S-Plus*, pages 3–56.

Pinheiro, J. C., Liu, C. & Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics*, **10**(2), 249–276.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Robert, C. P. (1995). Simulation of truncated normal variables. *Statistics and computing*, **5**(2), 121–125.

Robert, C. P. & Casella, G. (2010). *Introducing Monte Carlo Methods with R*, volume 18. Springer.

Sahu, S. K., Dey, D. K. & Branco, M. D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *Canadian Journal of Statistics*, **31**(2), 129–150.

Santos, N., López, R. G., Israelian, G., Mayor, M., Rebolo, R., García-Gil, A. & Randich, S. (2002). Beryllium abundances in stars hosting giant planets. *Astronomy & Astrophysics*, **386**(3), 1028–1038.

Schelin, L. & Sjöstedt-de Luna, S. (2014). Spatial prediction in the presence of left-censoring. *Computational Statistics & Data Analysis*, **74**, 125–141.

Schumacher, F., Matos, L., Lachos, VH, V. K., Henderson, N. & Varadhan, R. (2023). Package 'skewlmm'. *R package version*.

Schumacher, F. L., Lachos, V. H. & Galarza, C. E. (2016). Package 'ARCensReg'. *R package version*.

Schumacher, F. L., Lachos, V. H. & Dey, D. K. (2017). Censored regression models with autoregressive errors: A likelihood-based perspective. *Canadian Journal of Statistics*, **45**(4), 375–392.

Schumacher, F. L., Lachos, V. H. & Matos, L. A. (2021). Scale mixture of skew-normal linear mixed models with within-subject serial dependence. *Statistics in medicine*, **40**(7), 1790–1810.

Shammas, N. K. (1986). Interactions of temperature, ph, and biomass on the nitrification process. *Journal (Water Pollution Control Federation)*, pages 52–59.

Swendsen, R. H. & Wang, J.-S. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Physical review letters*, **58**(2), 86.

Tallis, G. M. (1961). The moment generating function of the truncated multi-normal distribution. *Journal of the Royal Statistical Society: Series B (Methodological)*, **23**(1), 223–229.

Tiku, M. L., Wong, W.-K. & Bian, G. (1999). Estimating parameters in autoregressive models in non-normal situations: Symmetric innovations. *Communications in Statistics-Theory and Methods*, **28**(2), 315–341.

Tsay, R. S. (2005). *Analysis of financial time series*, volume 543. John Wiley & Sons.

Vaida, F. & Liu, L. (2009). Fast implementation for normal mixed effects models with censored response. *Journal of Computational and Graphical Statistics*, **18**(4), 797–817.

Valeriano, K., Ordoñez, A., Matos, L. & Galarza, C. (2021a). Package 'RcppCensSpatial'. *R package version*.

Valeriano, K., Matos, L. & Galarza, C. (2022). Package 'relliptical'. *R package version*.

Valeriano, K. A., Lachos, V. H., Prates, M. O. & Matos, L. A. (2021b). Likelihood-based inference for spatiotemporal data with censored and missing responses. *Environmetrics*, **32**(3), e2663.

Verbeke, G. & Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, **91**(433), 217–221.

Verbeke, G., Molenberghs, G. & Verbeke, G. (1997). *Linear mixed models for longitudinal data*. Springer.

Viglione, A. (2010). On the sampling distribution of the coefficient of l-variation for hydrological applications. *Hydrology and Earth System Sciences Discussions*, **7**(4), 5467–5496.

Walsh, M. J. (2006). Computing the observed information matrix for dynamic mixture models. Technical report, 11, 768, Naval Undersea Warfare Center Division, Newport, RI.

Wang, C. & Chan, K.-S. (2018). Quasi-likelihood estimation of a censored autoregressive model with exogenous variables. *Journal of the American Statistical Association*, **113**(523), 1135–1145.

Wang, W.-L. (2013). Multivariate t linear mixed models for irregularly observed multiple repeated measures with missing outcomes. *Biometrical Journal*, **55**(4), 554–571.

Wang, W.-L., Lin, T.-I. & Lachos, V. H. (2018). Extending multivariate-t linear mixed models for multiple longitudinal data with censored responses and heavy tails. *Statistical methods in medical research*, **27**(1), 48–64.

Wei, G. C. & Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, **85**(411), 699–704.

Wilhelm, S. (2015). Package 'tmvtnorm'. *R journal*.

Wu, L. (2010). *Mixed Effects Models for Complex Data*. Chapman & Hall/CRC, Boca Raton, FL.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, **99**(465), 250–261.

Zhou, R., Liu, J., Kumar, S. & Palomar, D. P. (2020). Student's t VAR modeling with missing data via stochastic EM and Gibbs Sampling. *IEEE Transactions on Signal Processing*, **68**, 6198–6211.

Zirschky, J. H. & Harris, D. J. (1986). Geostatistical analysis of hazardous waste site data. *Journal of Environmental Engineering*, **112**(4), 770–784.

APPENDIX A – CHAPTER 2

A.1 Further results for some multivariate elliptical distributions

A.1.1 The multivariate Pearson VII distribution

A p-dimensional random vector $\mathbf{X}$ is said to have a multivariate Pearson VII distribution with location parameter $\boldsymbol{\mu} \in \mathbb{R}^p$, positive-definite scale matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, extra parameters $m > p/2$ and $\nu > 0$, if its pdf is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\Gamma(m)}{(\pi\nu)^{p/2}\Gamma(m-p/2)}\,|\boldsymbol{\Sigma}|^{-1/2}\left(1+\frac{1}{\nu}(\mathbf{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)^{-m},$$

with $\mathbf{x} \in \mathbb{R}^p$. The random vector $\mathbf{X}$ can also be represented as a scale mixture of normal (SMN) distributions, i.e., $\mathbf{X} = \boldsymbol{\mu} + U^{-1/2}\mathbf{Z}$, where $\mathbf{Z}$ has a p-variate normal distribution with mean $\mathbf{0} \in \mathbb{R}^p$ and variance-covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$. Here, U follows Gamma distribution with scale parameter $m - p/2$ and rate parameter $\nu/2$, and $\mathbf{Z}$ is independent of U. This implies that $\mathbf{X} \mid (U = u) \sim N_p(\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})$ and $U \sim Gamma(m - p/2, \nu/2)$.

Therefore, the mean and the variance-covariance matrix of $\mathbf{X}$ are

$$\mathbb{E}(\mathbf{X}) = \mathbb{E}(\mathbb{E}(\mathbf{X} \mid U)) = \boldsymbol{\mu}, \quad m > \frac{p+1}{2}.$$

$$\mathrm{Cov}(\mathbf{X}) = \mathrm{Cov}(\mathbb{E}(\mathbf{X} \mid U)) + \mathbb{E}(\mathrm{Cov}(\mathbf{X} \mid U)) = \mathbb{E}(U^{-1})\boldsymbol{\Sigma} = \frac{\nu\boldsymbol{\Sigma}}{2m-p-2}, \quad m > \frac{p+2}{2}.$$

A.1.1.1 Marginal and conditional distribution

Now suppose that the vector $\mathbf{X}$ is partitioned into two random vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ of dimensions p_1 and p_2, such that $p = p_1 + p_2$, and consider the partition of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ used in Proposition 1, i.e.,

$$\mathbf{X} = \begin{pmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{pmatrix},\ \boldsymbol{\mu} = \begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix},\ \text{and}\ \boldsymbol{\Sigma} = \begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}.$$

First, notice that $(\mathbf{X}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) = \delta_1(\mathbf{X}_1) + \delta_{2.1}(\mathbf{X}_{2.1})$, where $\delta_1(\mathbf{X}_1) = (\mathbf{X}_1 - \boldsymbol{\mu}_1)^\top\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{X}_1-\boldsymbol{\mu}_1)$, $\delta_{2.1}(\mathbf{X}_{2.1}) = (\mathbf{X}_2-\boldsymbol{\mu}_{2.1})^\top\boldsymbol{\Sigma}_{2.1}^{-1}(\mathbf{X}_2-\boldsymbol{\mu}_{2.1})$, $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2+\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{X}_1-\boldsymbol{\mu}_1)$, and $\boldsymbol{\Sigma}_{2.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$. By the results above, the marginal pdf of $\mathbf{X}_1$ is given by

$$\begin{aligned}
f_{\mathbf{X}_1}(\mathbf{x}_1) &= \int_{\mathbb{R}^{p_2}} f_{\mathbf{X}}(\mathbf{x})\mathrm{d}\mathbf{x}_2\\
&= \frac{\Gamma(m)}{(\pi\nu)^{p/2}\Gamma(m-p/2)}\,|\boldsymbol{\Sigma}|^{-1/2}\int_{\mathbb{R}^{p_2}}\left(1+\frac{\delta_1(\mathbf{x}_1)}{\nu}+\frac{\delta_{2.1}(\mathbf{x}_{2.1})}{\nu}\right)^{-m}\mathrm{d}\mathbf{x}_2\\
&= \frac{\Gamma(m)}{(\pi\nu)^{p/2}\Gamma(m-p/2)}\,|\boldsymbol{\Sigma}|^{-1/2}\left(1+\frac{\delta_1(\mathbf{x}_1)}{\nu}\right)^{-m}\int_{\mathbb{R}^{p_2}}\left(1+\frac{\delta_{2.1}(\mathbf{x}_{2.1})}{\nu+\delta_1(\mathbf{x}_1)}\right)^{-m}\mathrm{d}\mathbf{x}_2
\end{aligned}$$

$$f_{\mathbf{X}_1}(\mathbf{x}_1) = \frac{\Gamma(m-p_2/2)}{(\pi\nu)^{p_1/2}\Gamma(m-p/2)}\,|\boldsymbol{\Sigma}_{11}|^{-1/2}\left(1+\frac{\delta_1(\mathbf{x}_1)}{\nu}\right)^{-(m-p_2/2)}, \quad \mathbf{x}_1 \in \mathbb{R}^{p_1}.$$

Hence, the marginal distribution of $\mathbf{X}_1$ is also Pearson VII with parameters $\boldsymbol{\mu}_1$, $\boldsymbol{\Sigma}_{11}$, $m-p_2/2$ and ν, i.e., $\mathbf{X}_1 \sim \text{PVII}_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, m-p_2/2, \nu)$. On the other hand, the conditional pdf of $\mathbf{X}_2 \mid (\mathbf{X}_1 = \mathbf{x}_1)$ is given by

$$\begin{aligned} f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2 \mid \mathbf{x}_1) &= \frac{f_{\mathbf{X}}(\mathbf{x}_1, x_2)}{f_{\mathbf{X}_1}(\mathbf{x}_1)} \\ &= \frac{\Gamma(m)\,|\boldsymbol{\Sigma}_{2.1}|^{-1/2}}{(\pi(\nu+\delta_1(\mathbf{x}_1)))^{p_2/2}\Gamma(m-p_2/2)}\left(1+\frac{\delta_{2.1}(\mathbf{x}_{2.1})}{\nu+\delta_1(\mathbf{x}_1)}\right)^{-m}, \end{aligned}$$

$\mathbf{x}_1 \in \mathbb{R}^{p_1}$, $\mathbf{x}_2 \in \mathbb{R}^{p_2}$. Therefore, the conditional distribution has also a Pearson VII distribution with parameters $\boldsymbol{\mu}_{2.1}$, $\boldsymbol{\Sigma}_{2.1}$, m, and $\nu + \delta_1(\mathbf{x}_1)$, i.e., $\mathbf{X}_2 \mid (\mathbf{X}_1 = \mathbf{x}_1) \sim \text{PVII}_{p_2}(\boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{2.1}, m, \nu + \delta_1(\mathbf{x}_1))$.

A.1.1.2 Existence of its truncated moments

Let $\mathbf{X} \sim \text{PVII}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, m, \nu), m > p/2, \nu > 0$, and let $A \subseteq \mathbb{R}^p$ be a truncation region of interest. Then, the expectation and the variance-covariance matrix of $\mathbf{X}$ given $\mathbf{X} \in A$ exist in the following cases:

- If $A = \mathbb{R}^p$ or A is unbounded (at most one finite limit in each dimension), so the expectation exists for $m > (p+1)/2$ and the covariance matrix exists for $m > (p+2)/2$, as usual.
- If A is bounded (all truncation points are finite), then $\mathbb{E}(\mathbf{X} \mid \mathbf{X} \in A)$ and $\text{Cov}(\mathbf{X} \mid \mathbf{X} \in A)$ exist for all $m > p/2$, since the distribution is bounded.
- If $\mathbf{X}$ can be partitioned into two random variables $\mathbf{X}_1$ and $\mathbf{X}_2$ of dimensions p_1 and p_2, such that the truncation region associated to $\mathbf{X}_1$ (say, A_1) is bounded, from the last item we have $\mathbb{E}(\mathbf{X}_1 \mid \mathbf{X} \in A)$ and $\text{Cov}(\mathbf{X}_1 \mid \mathbf{X} \in A)$ exist for all $m > p/2$ and $\nu > 0$. On the other hand, it follows from Fubini's theorem that $\mathbb{E}(\mathbf{X}_2 \mid \mathbf{X} \in A)$ will exist if and only if $\mathbb{E}(\mathbf{X}_2 \mid \mathbf{X}_1)$ exists; this occurs for all $m > (p_2+1)/2$. Note that the existence of $\mathbb{E}(\mathbf{X}_2 \mid \mathbf{X}_1)$ also implies that $\text{Cov}(\mathbf{X}_1, \mathbf{X}_2 \mid \mathbf{X} \in A)$ exists. Additionally, $\text{Cov}(\mathbf{X}_2 \mid \mathbf{X} \in A)$ exists if and only if $\text{Cov}(\mathbf{X}_2 \mid \mathbf{X}_1)$ exists, which holds for all $m > (p_2+2)/2$.

Remark: It is equivalent to saying that $\mathbb{E}(\mathbf{X} \mid \mathbf{X} \in A)$ exists for all m if at least one dimension containing a finite limit exists. Besides, if at least two dimensions containing finite limits exist, we have that $\text{Cov}(\mathbf{X} \mid \mathbf{X} \in A)$ exists for all $m > p/2$.

In order to illustrate the result, consider $\mathbf{X} \sim \text{PVII}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}, m, \nu)$, with $\nu = 1$, $\boldsymbol{\mu} = \mathbf{0}$, and $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.20 \\ 0.20 & 1 \end{pmatrix}$. We are interested in observing what happens with the elements of $\mathbb{E}(\mathbf{X} \,|\, \mathbf{X} \in A)$ and $\text{Cov}(\mathbf{X} \,|\, \mathbf{X} \in A)$ for $A = \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{a} < \mathbf{x} < \mathbf{b}\}$ in the following three scenarios:

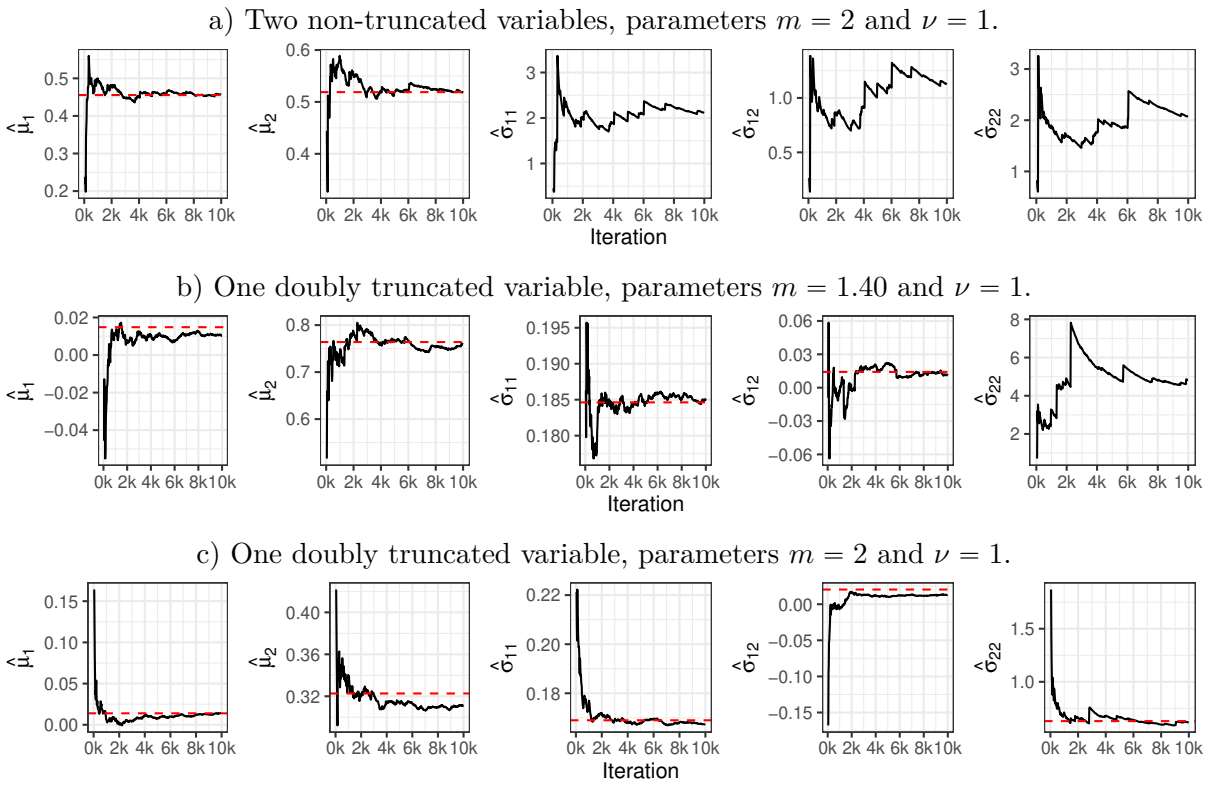
a) $m = 2$, $\mathbf{b} = (\infty, \infty)^\top$;

b) $m = 1.40$, $\mathbf{b} = (0.80, \infty)^\top$;

c) $m = 2$, $\mathbf{b} = (0.80, \infty)^\top$;

and lower limit $\mathbf{a} = (-0.80, -0.60)^\top$ for all scenarios. Figure 35 displays the trace evolution of the MC estimates for the mean and variance-covariance elements μ_1, μ_2, σ_{11}, σ_{12} and σ_{22} for each case. The red dashed line represents the value for the parameter estimated via MC with 10^6 samples, and we refer to this value as the "true value".

Figure 35 – Trace plots of the evolution of the MC estimates for the mean and variance-covariance elements of $\mathbf{X} \mid (\mathbf{X} \in A)$ under scenarios a), b), and c). The red dashed line represents the true estimated value computed using numerical methods



For the first case, we have that $(p + 1)/2 = 3/2 < 2 = m$, then only the first moment exists. Therefore, we observe in the first row of Figure 35 that only the estimates of μ_1 and μ_2 converge to their true values as the sample size increases. In the

second scenario (middle row), we have that all elements converge except σ_{22}. This happens because the truncation limits for the first variable are finite and $m > (p_2 + 1)/2 = 1$. In the last case, scenario c), convergence is attained for all parameters, since the condition $m > (p_2 + 2)/2 = 3/2$ holds. Note that even with 2000 MC simulations there exists a significant variability in the chains.

A.1.2 The multivariate Slash distribution

A random vector $\mathbf{X}$ of length p has multivariate slash distribution with location parameter $\boldsymbol{\mu} \in \mathbb{R}^p$, positive-definite scale matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, and $\nu > 0$ degrees of freedom, denoted by $\mathbf{X} \sim \mathrm{SL}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, if its pdf is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \nu \int_0^1 u^{\nu-1} \phi_p\left(\mathbf{x}; \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}\right) \mathrm{d}u, \quad \mathbf{x} \in \mathbb{R}^p,$$

where $\phi_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the pdf of a p-variate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. We denote its pdf by $SL_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ which can be evaluated through numerical methods, e.g., using the R function `integrate`. The random vector $\mathbf{X}$ can also be represented in the family of the SMN distributions, this is, $\mathbf{X} = \boldsymbol{\mu} + U^{-1/2}\mathbf{Z}$, where the random variables U and $\mathbf{Z}$ are both independent and have $\mathrm{Beta}(\nu, 1)$ and $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ distributions, respectively. Therefore, the mean and variance-covariance matrix of the random vector $\mathbf{X}$ are given by

$$\begin{aligned}
\mathbb{E}(\mathbf{X}) &= \mathbb{E}\left(\mathbb{E}(\mathbf{X} \mid U)\right) = \mathbb{E}(\boldsymbol{\mu}) = \boldsymbol{\mu}. \\
\mathrm{Cov}(\mathbf{X}) &= \mathrm{Cov}(\mathbb{E}(\mathbf{X} \mid U)) + \mathbb{E}(\mathrm{Cov}(\mathbf{X} \mid U)) = \mathbb{E}(U^{-1})\boldsymbol{\Sigma} = \frac{\nu}{\nu - 1}\boldsymbol{\Sigma}, \quad \nu > 1.
\end{aligned}$$

A.1.2.1 Marginal and conditional distribution

Considering a partition in the same manner as used for the Pearson VII distribution, the marginal pdf of $\mathbf{X}_1$ is given by

$$\begin{aligned}
f_{\mathbf{X}_1}(\mathbf{x}_1) &= \int_{\mathbb{R}^{p_2}} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}_2 = \int_{\mathbb{R}^{p_2}} \nu \int_0^1 u^{\nu-1} \phi_p\left(\mathbf{x}; \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}\right) \mathrm{d}u \, d\mathbf{x}_2 \\
&= \nu \int_{\mathbb{R}^{p_2}} \int_0^1 u^{\nu-1} \phi_{p_1}\left(\mathbf{x}_1; \boldsymbol{\mu}_1, u^{-1}\boldsymbol{\Sigma}_{11}\right) \phi_{p_2}\left(\mathbf{x}_2; \boldsymbol{\mu}_{2.1}, u^{-1}\boldsymbol{\Sigma}_{2.1}\right) \mathrm{d}u \, \mathrm{d}\mathbf{x}_2 \\
&= \nu \int_0^1 u^{\nu-1} \phi_{p_1}\left(\mathbf{x}_1; \boldsymbol{\mu}_1, u^{-1}\boldsymbol{\Sigma}_{11}\right) \int_{\mathbb{R}^{p_2}} \phi_{p_2}\left(\mathbf{x}_2; \boldsymbol{\mu}_{2.1}, u^{-1}\boldsymbol{\Sigma}_{2.1}\right) \mathrm{d}\mathbf{x}_2 \, \mathrm{d}u \\
&= \nu \int_0^1 u^{\nu-1} \phi_{p_1}\left(\mathbf{x}_1; \boldsymbol{\mu}_1, u^{-1}\boldsymbol{\Sigma}_{11}\right) \mathrm{d}u.
\end{aligned}$$

Thus, $\mathbf{X}_1$ follows a slash distribution with location parameter $\boldsymbol{\mu}_1 \in \mathbb{R}^{p_1}$, scale matrix $\boldsymbol{\Sigma}_{11} \in \mathbb{R}^{p_1\times p_1}$, and $\nu > 0$ degrees of freedom. On the other hand, the conditional pdf of

$\mathbf{X}_2 \mid (\mathbf{X}_1 = \mathbf{x}_1)$ is

$$\begin{aligned} f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2 \mid \mathbf{x}_1) &= \frac{f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2)}{f_{\mathbf{X}_1}(\mathbf{x}_1)} \\ &= \frac{\nu}{f_{\mathbf{X}_1}(\mathbf{x}_1)} \int_0^1 u^{\nu-1} \phi_p\left(\mathbf{x}; \boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}\right) \mathrm{d}u \\ &= \frac{\nu}{f_{\mathbf{X}_1}(\mathbf{x}_1)} \int_0^1 u^{\nu-1} \phi_{p_1}\left(\mathbf{x}_1; \boldsymbol{\mu}_1, u^{-1}\boldsymbol{\Sigma}_{11}\right) \phi_{p_2}\left(\mathbf{x}_2; \boldsymbol{\mu}_{2.1}, u^{-1}\boldsymbol{\Sigma}_{2.1}\right) \mathrm{d}u. \end{aligned}$$

Then, it is possible to notice that the slash distribution is not closed under conditioning. Furthermore, the pdf of $\mathbf{X}_2 \,|\, (\mathbf{X}_1 = \mathbf{x}_1)$ belongs to the elliptical family of distributions with dgf $g_{x_1}^{(p_2)}(t) = \displaystyle\int_0^1 u^{\nu+p/2-1} \exp\{-u(t+\delta_1(\mathbf{x}_1))/2\} \,\mathrm{d}u$, i.e., $\mathbf{X}_2 \,|\, (\mathbf{X}_1 = \mathbf{x}_1) \sim \mathrm{E}\ell(\boldsymbol{\mu}_{2.1}, \boldsymbol{\Sigma}_{2.1}, \nu; g_{x_1}^{(p_2)})$. To determine the mean of the random vector $\mathbf{X}_2 \,|\, (\mathbf{X}_1 = \mathbf{x}_1)$, we compute the conditional expected value of the ith element of $\mathbf{X}_2$ as follows

$$\begin{aligned} \mathbb{E}(\mathbf{X}_{2i} \mid \mathbf{X}_1 = \mathbf{x}_1) &= \int_{\mathbb{R}^{p_2}} x_{2i} f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2 \mid \mathbf{x}_1) \mathrm{d}\mathbf{x}_2 \\ &= \frac{\nu}{f_{\mathbf{X}_1}(\mathbf{x}_1)} \int_{\mathbb{R}^{p_2}} x_{2i} \int_0^1 u^{\nu-1} \phi_{p_1}\left(\mathbf{x}_1; \boldsymbol{\mu}_1, u^{-1}\boldsymbol{\Sigma}_{11}\right) \phi_{p_2}\left(\mathbf{x}_2; \boldsymbol{\mu}_{2.1}, u^{-1}\boldsymbol{\Sigma}_{2.1}\right) \mathrm{d}u \,\mathrm{d}\mathbf{x}_2 \\ &= \frac{\nu}{f_{\mathbf{X}_1}(\mathbf{x}_1)} \int_0^1 u^{\nu-1} \phi_{p_1}\left(\mathbf{x}_1; \boldsymbol{\mu}_1, u^{-1}\boldsymbol{\Sigma}_{11}\right) \int_{\mathbb{R}^{p_2}} x_{2i} \phi_{p_2}\left(\mathbf{x}_2; \boldsymbol{\mu}_{2.1}, u^{-1}\boldsymbol{\Sigma}_{2.1}\right) d\mathbf{x}_2 \,\mathrm{d}u \\ &= \frac{\mu_{2.1}^{(i)} \nu}{f_{\mathbf{X}_1}(\mathbf{x}_1)} \int_0^1 u^{\nu-1} \phi_{p_1}\left(\mathbf{x}_1; \boldsymbol{\mu}_1, u^{-1}\boldsymbol{\Sigma}_{11}\right) \mathrm{d}u = \mu_{2.1}^{(i)}, \quad \forall i, \nu > 0, \end{aligned}$$

where $\mu_{2.1}^{(i)}$ represents the ith element of the vector $\boldsymbol{\mu}_{2.1}$, and $\mathbb{E}(\mathbf{X}_2 \,|\, \mathbf{X}_1 = \mathbf{x}_1) = \boldsymbol{\mu}_{2.1}$. Now, to compute the elements of the variance-covariance matrix of the conditional random vector, we first determine $\mathbb{E}(X_{2i}X_{2j} \,|\, \mathbf{X}_1 = \mathbf{x}_1)$ for all $i, j = 1, \ldots, p_2$, as

$$\begin{aligned} \mathbb{E}(X_{2i}X_{2j} \mid \mathbf{X}_1 = \mathbf{x}_1) &= \int_{\mathbb{R}^{p_2}} x_{2i}x_{2j} f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2 \mid \mathbf{x}_1) d\mathbf{x}_2 \\ &= \frac{\nu}{f_{\mathbf{X}_1}(\mathbf{x}_1)} \int_{\mathbb{R}^{p_2}} x_{2i}x_{2j} \int_0^1 u^{\nu-1} \phi_{p_1}\left(\mathbf{x}_1; \boldsymbol{\mu}_1, u^{-1}\boldsymbol{\Sigma}_{11}\right) \phi_{p_2}\left(\mathbf{x}_2; \boldsymbol{\mu}_{2.1}, u^{-1}\boldsymbol{\Sigma}_{2.1}\right) \mathrm{d}u \,\mathrm{d}\mathbf{x}_2 \\ &= \frac{\nu}{f_{\mathbf{X}_1}(\mathbf{x}_1)} \int_0^1 u^{\nu-1} \phi_{p_1}\left(\mathbf{x}_1; \boldsymbol{\mu}_1, u^{-1}\boldsymbol{\Sigma}_{11}\right) \int_{\mathbb{R}^{p_2}} x_{2i}x_{2j} \phi_{p_2}\left(\mathbf{x}_2; \boldsymbol{\mu}_{2.1}, u^{-1}\boldsymbol{\Sigma}_{2.1}\right) \mathrm{d}\mathbf{x}_2 \,\mathrm{d}u \\ &= \frac{\nu}{f_{\mathbf{X}_1}(\mathbf{x}_1)} \int_0^1 u^{\nu-1} \phi_{p_1}\left(\mathbf{x}_1; \boldsymbol{\mu}_1, u^{-1}\boldsymbol{\Sigma}_{11}\right) \left(u^{-1}\sigma_{2.1}^{(ij)} + \mu_{2.1}^{(i)}\mu_{2.1}^{(j)}\right) \mathrm{d}u \\ &= \frac{\sigma_{2.1}^{(ij)} \nu}{f_{\mathbf{X}_1}(\mathbf{x}_1)} \int_0^1 u^{\nu-2} \phi_{p_1}\left(\mathbf{x}_1; \boldsymbol{\mu}_1, u^{-1}\boldsymbol{\Sigma}_{11}\right) \mathrm{d}u + \mu_{2.1}^{(i)}\mu_{2.1}^{(j)} \\ &= \frac{\nu}{\nu-1} \left(\frac{SL_{p_1}(\mathbf{x}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \nu-1)}{SL_{p_1}(\mathbf{x}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \nu)} \right) \sigma_{2.1}^{(ij)} + \mu_{2.1}^{(i)}\mu_{2.1}^{(j)}, \quad \nu > 1, \end{aligned}$$

where $\sigma_{2.1}^{(ij)}$ is the (i,j)th element of the matrix $\boldsymbol{\Sigma}_{2.1}$. From these results, we have that

$$\mathrm{Cov}(X_{2i}, X_{2j} \mid \mathbf{X}_1 = \mathbf{x}_1) = \frac{\nu}{\nu-1} \left(\frac{SL_{p_1}(\mathbf{x}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \nu-1)}{SL_{p_1}(\mathbf{x}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \nu)} \right) \sigma_{2.1}^{(ij)},$$

$\nu > 1$. Therefore, the covariance matrix of the random vector $\mathbf{X}_2 \mid (\mathbf{X}_1 = \mathbf{x}_1)$ will be given by

$$\text{Cov}(\mathbf{X}_2 \mid \mathbf{X}_1 = \mathbf{x}_1) = \frac{\nu}{\nu - 1}\left(\frac{SL_{p_1}(\mathbf{x}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \nu - 1)}{SL_{p_1}(\mathbf{x}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \nu)}\right)\boldsymbol{\Sigma}_{2.1}.$$

A.2 CPU time to compute moments from truncated distributions

A complementary study of Simulation study II (Subsection 2.4.2) was conducted to examine the computational time required for our method in order to estimate the first two moments and the variance-covariance matrix of a p-variate random vector considering different distributions in the truncated elliptical family, with $p = 50$ and 100. As in Simulation study II, we consider 10%, 20%, and 40% of doubly truncated variables for each case.

Table 23 shows the median of the CPU time (in seconds) needed for the function `mvtelliptical` to compute the first two moments and the covariance matrix. We considered a TMVN, a truncated contaminated normal with $\nu = 1/2$ and $\rho = 1/5$, a truncated Pearson VII with parameters $m = 55$ and $\nu = 3$, a truncated slash with $\nu = 2$ degrees of freedom, and a truncated power exponential distribution with kurtosis $\beta = 1/2$. For each case, our method was applied setting $n = 10^4$ and 10^5 with $thinning = 3$. Notice that the time needed by the algorithm for TMVN, TMVT, and truncated Pearson VII distributions are similar and depend only on the number of truncated variables and samples used in the approximation. Our method requires more time to compute moments from the truncated contaminated normal distribution when compared to the latter results. This is because the algorithm uses a numerical method to calculate the inverse of the dgf. Besides, it is interesting to note that there is no time difference between computing the moments for a truncated slash distribution with five or ten doubly truncated variables. This occurs since the function used to approximate the integral on the dgf is more time-consuming when $\nu + p/2 - 1$ is not an integer. Finally, the computation of the moments for the truncated power exponential distribution required approximately the same time for random vectors of equal length regardless of the number of doubly truncated variables. For this case, the method samples values for the whole vector, leading to no time difference.

A.3 The `relliptical` R package

The `relliptical` package offers random numbers generation from members of the truncated multivariate elliptical family of distribution such as the truncated versions of the normal, Student-t, Pearson VII, slash, logistic, Kotz-type, among others. Particular distributions can be provided by specifying the density-generating function. It also computes

Table 23 – Median of the CPU time (in seconds) based on 100 simulations.

Distribution (ν)	Sample size	$p = 50$			$p = 100$		
		10%	20%	40%	10%	20%	40%
Normal	10^4	0.028	0.083	0.399	0.081	0.400	2.888
	10^5	0.285	0.840	3.999	0.805	4.003	28.892
Contaminated Normal (1/2, 1/5)	10^4	0.071	0.118	0.440	0.120	0.442	2.928
	10^5	0.706	1.180	4.405	1.192	4.415	29.286
Pearson VII (55, 3)	10^4	0.031	0.083	0.403	0.084	0.403	2.891
	10^5	0.309	0.838	4.030	0.839	4.036	28.944
Slash (2)	10^4	0.202	0.202	0.548	0.200	0.549	3.113
	10^5	2.020	2.026	5.481	1.997	5.489	31.160
Power Exponential (1/2)	10^4	5.101	5.095	5.096	41.870	41.858	41.864
	10^5	51.038	51.013	50.999	418.675	418.604	418.651

the first two moments (covariance matrix as well) for some particular distributions. Next, we will show the functions available.

A.3.1 Random number generator

Its main function for random number generation is called `rtelliptical`, which is based on the methods described in Section 2.3, and whose signature is the following.

```
rtelliptical(n=1e4, mu=rep(0,length(lower)), Sigma=diag(length(lower)),
  lower, upper=rep(Inf,length(lower)), dist="Normal", nu=NULL, expr=NULL,
  gFun=NULL, ginvFun=NULL, burn.in=0, thinning=1)
```

In this function, $n \geqslant 1$ is the number of observations to be sampled, `nu` is the additional parameter or vector of parameters depending on the distribution of $\mathbf{X}$, `mu` is the location parameter, `Sigma` is the positive-definite scale matrix, and `lower` and `upper` are the lower and upper truncation points, respectively. The truncated normal, Student-t, power exponential, Pearson VII, slash, and contaminated normal distributions can be specified through the argument `dist`.

The following examples illustrate the function `rtelliptical`, for drawing samples from truncated bivariate distributions with location parameter $\boldsymbol{\mu} = (0, 0)^\top$, scale matrix elements $\sigma_{11} = \sigma_{22} = 1$, and $\sigma_{12} = \sigma_{21} = 0.70$, and truncation region $A = \{\mathbf{x} : \mathbf{a} < \mathbf{x} < \mathbf{b}\}$, with $\mathbf{a} = (-2, -2)^\top$ and $\mathbf{b} = (3, 2)^\top$. The distributions considered are the predefined ones in the package.

- Truncated normal

```
rtelliptical(n=1e4, mu=c(0,0), Sigma=matrix(c(1,0.7,0.7,1),2,2),
  lower=c(-2,-2), upper=c(3,2), dist="Normal")
```

- Truncated Student-t with $\nu = 3$ degrees of freedom

```
rtelliptical(n=1e4, mu=c(0,0), Sigma=matrix(c(1,0.7,0.7,1),2,2),
  lower=c(-2,-2), upper=c(3,2), dist="t", nu=3)
```

- Truncated power exponential with $\beta = 2$

```
rtelliptical(n=1e4, mu=c(0,0), Sigma=matrix(c(1,0.7,0.7,1),2,2),
  lower=c(-2,-2), upper=c(3,2), dist="PE", nu=2)
```

- Truncated Pearson VII with parameters $m = 5/2$ and $\nu = 3$

```
rtelliptical(n=1e4, mu=c(0,0), Sigma=matrix(c(1,0.7,0.7,1),2,2),
  lower=c(-2,-2), upper=c(3,2), dist="PVII", nu=c(2.50,3))
```

- Truncated slash with $\nu = 1.5$ degrees of freedom

```
rtelliptical(n=1e4, mu=c(0,0), Sigma=matrix(c(1,0.7,0.7,1),2,2),
  lower=c(-2,-2), upper=c(3,2), dist="Slash", nu=1.50)
```

- Truncated contaminated normal with $\nu = 0.70$ and $\rho = 0.20$

```
rtelliptical(n=1e4, mu=c(0,0), Sigma=matrix(c(1,0.7,0.7,1),2,2),
  lower=c(-2,-2), upper=c(3,2), dist="CN", nu=c(0.70,0.20))
```

Note that no additional arguments are passed for the TMVN distribution. On the opposite way, for the truncated contaminated normal and Pearson VII distributions, `nu` is a vector of length two, and for the remaining distributions, this parameter is a non-negative scalar. An important remark is that there exist closed-form expressions to compute $\kappa_y = g^{-1}(y)$ for the normal, Student-t, power exponential, and Pearson VII distributions; however, the contaminated normal and slash distributions require numerical methods for this purpose. This value is calculated as the root of the function $g(t) - y = 0, t \geqslant 0$, through the Newton-Raphson algorithm for the contaminated normal, and using Brent's method (Brent, 2013), for the slash distribution, a mixture of linear interpolation, inverse quadratic interpolation, and the bisection method.

This function also allows generating random numbers from other truncated elliptical distributions not specified in the `dist` argument, by supplying the dgf through arguments either `expr` or `gFun`. The easiest way is to provide the dgf expression to argument `expr` as a character. The notation used in `expr` needs to be understood by the `Ryacas0` package (Andersen *et al.*, 2020), and the R environment. For instance, for the dgf $g(t) = e^{-t}$, the user must provide `expr = "exp(1)^(-t)"`. For this case, when a character expression is provided to `expr`, the algorithm tries to compute a closed-form expression for the inverse function of $g(t)$; however, this is not always possible (a warning message is returned). On the other hand, if it is not possible to pass an expression to `expr`, due to the complexity of the expression, the user may provide a custom R function to the

`gFun` argument. By default, its inverse function is approximated numerically; however, the user may also provide its inverse to the `ginvFun` argument to gain some computational time. When `gFun` is provided, arguments `dist` and `expr` are ignored.

For example, to generate samples from the bivariate truncated logistic distribution with the same parameters as before, and which has dgf $g(t) = e^{-t}/(1+e^{-t})^2, t \geqslant 0$, we can run the following code.

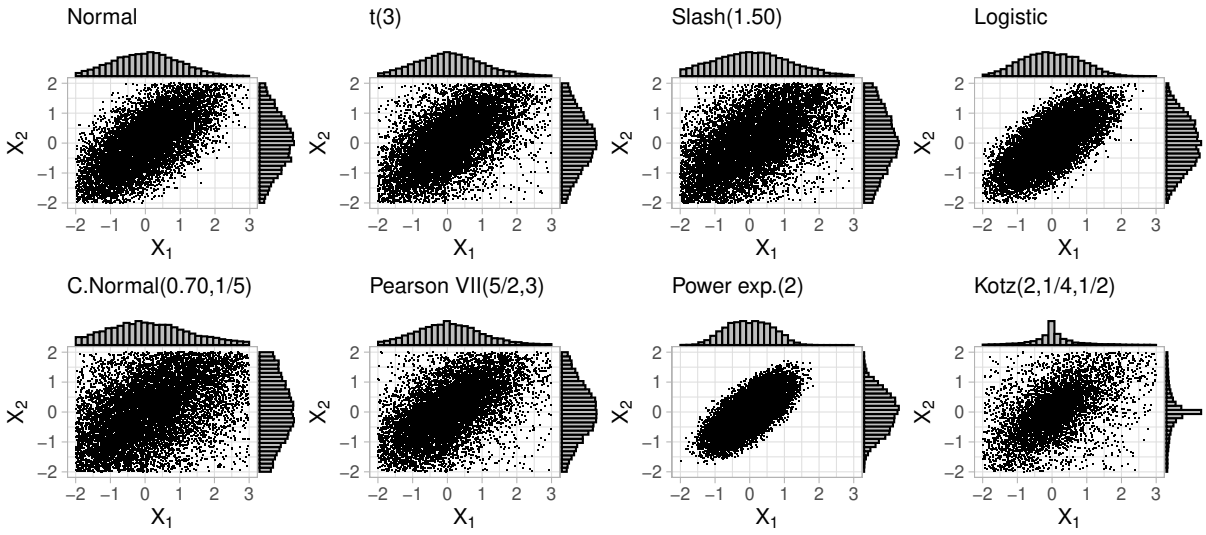
```
rtelliptical(n=1e4, mu=c(0,0), Sigma=matrix(c(1,0.7,0.7,1),2,2),
  lower=c(-2,-2), upper=c(3,2), expr="exp(1)^(-t)/(1+exp(1)^(-t))^2")
```

Another distribution that belongs to the elliptical family is the Kotz-type distribution with parameters $r > 0, s > 0$, and $2N + p > 2$, whose dgf is $g(t) = t^{N-1}e^{-rt^s}, t \geqslant 0$ (Fang, 2018). For this distribution, $g(t)$ is not strictly decreasing for all parameter values, however, for $(2-p)/2 < N \leqslant 1$, it holds. Hence, our proposal works for $r > 0$, $s > 0$, and $(2-p)/2 < N \leqslant 1$. For this type of more complex dgf, it is advisable to pass it through the `gFun` argument as an R function (with other parameters as fixed values). In the following example, we draw samples from a bivariate Kotz-type distribution with settings as before, and extra parameters $r = 2, s = 1/4$, and $N = 1/2$.

```
rtelliptical(n=1e4, mu=c(0,0), Sigma=matrix(c(1,0.7,0.7,1),2,2),
  lower=c(-2,-2), upper=c(3,2), gFun=function(t){t^(-1/2)*exp(-2*t^(1/4))})
```

Figure 36 shows the scatterplot and marginal histograms for the $n = 10^4$ observations sampled from each of the truncated bivariate distributions referred to above.

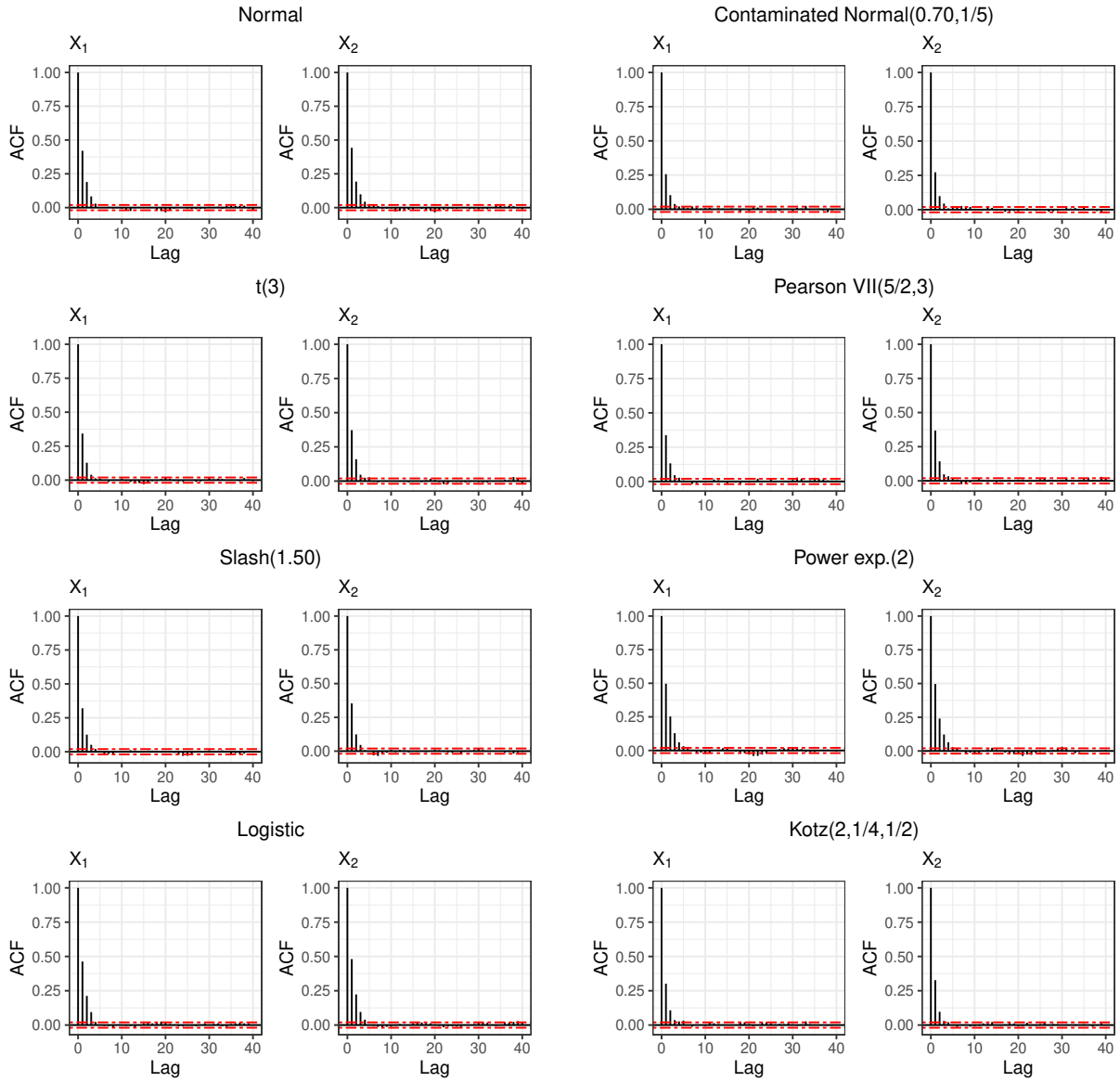
Figure 36 – Scatterplot and marginal histograms for the $n = 10^4$ observations sampled for some bivariate truncated elliptical distributions.



As mentioned by Robert & Casella (2010) and Ho *et al.* (2012), the slice sampling algorithm with Gibbs steps generates random samples conditioned on previous values, resulting in a sequence of correlated samples. Thus, it is essential to analyze

the dependence effect of the proposed algorithm. Figure 37 displays the autocorrelation plots for each one of the distributions, where we notice that the autocorrelation drops quickly and becomes negligibly small when lags become large, evidencing well mixing and quickly converging for these examples. If necessary, initial observations can be discarded by means of the `burn.in` argument. Finally, autocorrelation can be decimated by setting the `thinning` argument. Thinning consists of picking separated points from the sample at each kth step. The thinning factor reduces the autocorrelation of the random points in the Gibbs sampling process. As natural, this value must be an integer greater than or equal to 1.

Figure 37 – Sample autocorrelation plots of X_1 and X_2 sampled from the bivariate truncated elliptical distributions in Figure 36.



A.3.2 Mean and variance-covariance matrix computation

Algorithm 2 for the distributions detailed in Subsection 2.4.1 is available through the function `mvtelliptical`, whose signature, together with default values, is the following.

```
mvtelliptical(lower, upper=rep(Inf,length(lower)), mu=rep(0,length(lower)),
  Sigma=diag(length(lower)), dist="Normal", nu=NULL, n=1e4, burn.in=0,
  thinning=3)
```

The arguments `lower` and `upper` are the lower and upper truncation points of length p, respectively, `mu` is the location parameter of length p, `Sigma` is the $p \times p$ positive-definite scale matrix, `nu` is the additional parameter or vector of parameters depending on the dgf g. The argument `dist` indicates the distribution to be used. The parameters `n`, `burn.in`, and `thinning` are related to the Monte Carlo approximation, where `n` is the number of samples to be generated, `burn.in` is the number of samples to be discarded as burn-in phase, and `thinning` is a factor for reducing autocorrelation between observations.

APPENDIX B – CHAPTER 3

B.1 Elements of the observed information matrix

In Subsection 3.3.4, we defined the observed information matrix for the spatial censored linear (SCL) model based on Louis' method, which is given by

$$\mathbf{I}_o(\boldsymbol{\theta}) = \mathbb{E}\left[\mathbf{B}_c(\mathbf{y}_c;\boldsymbol{\theta}) \mid \mathbf{y}_o\right] - \mathbb{E}\left[\mathbf{S}_c(\mathbf{y}_c;\boldsymbol{\theta})\mathbf{S}_c^\top(\mathbf{y}_c;\boldsymbol{\theta}) \mid \mathbf{y}_o\right] + \mathbf{S}_o(\mathbf{y}_o;\boldsymbol{\theta})\mathbf{S}_o^\top(\mathbf{y}_o;\boldsymbol{\theta}), \tag{B.1}$$

where

$$\mathbf{B}_c(\mathbf{y}_c;\boldsymbol{\theta}) = -\frac{\partial^2 \ell_c(\boldsymbol{\theta};\mathbf{y}_c)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top},\ \mathbf{S}_c(\mathbf{y}_c;\boldsymbol{\theta}) = \frac{\partial \ell_c(\boldsymbol{\theta};\mathbf{y}_c)}{\partial\boldsymbol{\theta}},\ \text{and}\ \mathbf{S}_o(\mathbf{y}_o;\boldsymbol{\theta}) = \mathbb{E}\left[\mathbf{S}_c(\mathbf{y}_c;\boldsymbol{\theta})\middle|\mathbf{y}_o\right].$$

The elements of matrix $\mathbf{H}(\boldsymbol{\theta}) = \mathbb{E}[\mathbf{B}_c(\mathbf{y}_c;\boldsymbol{\theta})|\mathbf{y}_o]$ are the following:

$\mathbf{H}_{\beta\beta^\top} = -\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\mathbf{X}$,

$\mathbf{H}_{\beta\sigma^2} = \mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\mathbf{R}\boldsymbol{\Sigma}^{-1}(\mathbf{X}\boldsymbol{\beta} - \widehat{\mathbf{y}})$,

$\mathbf{H}_{\beta\phi} = \sigma^2\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\dot{\mathbf{R}}_\phi\boldsymbol{\Sigma}^{-1}(\mathbf{X}\boldsymbol{\beta} - \widehat{\mathbf{y}})$,

$\mathbf{H}_{\beta\tau^2} = \mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}(\mathbf{X}\boldsymbol{\beta} - \widehat{\mathbf{y}})$,

$\mathbf{H}_{(\sigma^2)^2} = -\frac{1}{2}\left[tr\left(-\boldsymbol{\Sigma}^{-1}\mathbf{R}\boldsymbol{\Sigma}^{-1}\mathbf{R}\right) + tr\left(\widehat{\mathbf{y}\mathbf{y}^\top}\ddot{\boldsymbol{\Sigma}}^{-1}_{(\sigma^2)^2}\right) - \boldsymbol{\beta}^\top\mathbf{X}^\top\ddot{\boldsymbol{\Sigma}}^{-1}_{(\sigma^2)^2}\left(2\widehat{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta}\right)\right]$,

where $\ddot{\boldsymbol{\Sigma}}^{-1}_{(\sigma^2)^2} = 2\boldsymbol{\Sigma}^{-1}\mathbf{R}\boldsymbol{\Sigma}^{-1}\mathbf{R}\boldsymbol{\Sigma}^{-1}$,

$\mathbf{H}_{\phi\sigma^2} = -\frac{1}{2}\left[tr\left(\boldsymbol{\Sigma}^{-1}\dot{\mathbf{R}}_\phi\left(\mathbf{I} - \sigma^2\boldsymbol{\Sigma}^{-1}\mathbf{R}\right)\right) + tr\left(\widehat{\mathbf{y}\mathbf{y}^\top}\ddot{\boldsymbol{\Sigma}}^{-1}_{\phi\sigma^2}\right) - \boldsymbol{\beta}^\top\mathbf{X}^\top\ddot{\boldsymbol{\Sigma}}^{-1}_{\phi\sigma^2}\left(2\widehat{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta}\right)\right]$,

where $\ddot{\boldsymbol{\Sigma}}^{-1}_{\phi\sigma^2} = \boldsymbol{\Sigma}^{-1}\left(\sigma^2\dot{\mathbf{R}}_\phi\boldsymbol{\Sigma}^{-1}\mathbf{R} - \dot{\mathbf{R}}_\phi + \sigma^2\mathbf{R}\boldsymbol{\Sigma}^{-1}\dot{\mathbf{R}}_\phi\right)\boldsymbol{\Sigma}^{-1}$,

$\mathbf{H}_{\tau^2\sigma^2} = -\frac{1}{2}\left[tr\left(-\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}\mathbf{R}\right) + tr\left(\widehat{\mathbf{y}\mathbf{y}^\top}\ddot{\boldsymbol{\Sigma}}^{-1}_{\tau^2\sigma^2}\right) - \boldsymbol{\beta}^\top\mathbf{X}^\top\ddot{\boldsymbol{\Sigma}}^{-1}_{\tau^2\sigma^2}\left(2\widehat{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta}\right)\right]$,

where $\ddot{\boldsymbol{\Sigma}}^{-1}_{\tau^2\sigma^2} = \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma}^{-1}\mathbf{R} + \mathbf{R}\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{\Sigma}^{-1}$,

$\mathbf{H}_{\phi^2} = -\frac{1}{2}\left[tr\left(\sigma^2\boldsymbol{\Sigma}^{-1}\left(\ddot{\mathbf{R}}_{\phi^2} - \sigma^2\dot{\mathbf{R}}_\phi\boldsymbol{\Sigma}^{-1}\dot{\mathbf{R}}_\phi\right)\right) + tr\left(\widehat{\mathbf{y}\mathbf{y}^\top}\ddot{\boldsymbol{\Sigma}}^{-1}_{\phi^2}\right) - \boldsymbol{\beta}^\top\mathbf{X}^\top\ddot{\boldsymbol{\Sigma}}^{-1}_{\phi^2}\left(2\widehat{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta}\right)\right]$,

where $\ddot{\boldsymbol{\Sigma}}^{-1}_{\phi^2} = \sigma^2\boldsymbol{\Sigma}^{-1}\left(2\sigma^2\dot{\mathbf{R}}_\phi\boldsymbol{\Sigma}^{-1}\dot{\mathbf{R}}_\phi - \ddot{\mathbf{R}}_{\phi^2}\right)\boldsymbol{\Sigma}^{-1}$,

$\mathbf{H}_{\phi\tau^2} = -\frac{1}{2}\left[tr\left(-\sigma^2\boldsymbol{\Sigma}^{-1}\dot{\mathbf{R}}_\phi\boldsymbol{\Sigma}^{-1}\right) + tr\left(\widehat{\mathbf{y}\mathbf{y}^\top}\ddot{\boldsymbol{\Sigma}}^{-1}_{\phi\tau^2}\right) - \boldsymbol{\beta}^\top\mathbf{X}^\top\ddot{\boldsymbol{\Sigma}}^{-1}_{\phi\tau^2}\left(2\widehat{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta}\right)\right]$,

where $\ddot{\boldsymbol{\Sigma}}^{-1}_{\phi\tau^2} = \sigma^2\boldsymbol{\Sigma}^{-1}\left(\dot{\mathbf{R}}_\phi\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\dot{\mathbf{R}}_\phi\right)\boldsymbol{\Sigma}^{-1}$,

$\mathbf{H}_{(\tau^2)^2} = -\frac{1}{2}\left[tr\left(-\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}\right) + tr\left(\widehat{\mathbf{y}\mathbf{y}^\top}\ddot{\boldsymbol{\Sigma}}^{-1}_{(\tau^2)^2}\right) - \boldsymbol{\beta}^\top\mathbf{X}^\top\ddot{\boldsymbol{\Sigma}}^{-1}_{(\tau^2)^2}\left(2\widehat{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta}\right)\right]$,

where $\ddot{\boldsymbol{\Sigma}}^{-1}_{(\tau^2)^2} = 2\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}$,

with $\mathbf{H}_{\phi\sigma^2} = \mathbb{E}\left[\frac{\partial^2 \ell_c(\boldsymbol{\theta};\mathbf{y}_c)}{\partial\phi\partial\sigma^2}\Big|\mathbf{y}_o\right]$, for any parameter in $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2, \phi, \tau^2)^\top$, $\dot{\mathbf{R}}_\phi = \frac{\partial \mathbf{R}}{\partial \phi}$, and $\ddot{\mathbf{R}}_{\phi^2} = \frac{\partial^2 \mathbf{R}}{\partial \phi^2}$.

Moreover, the elements of the vector $\mathbf{S}(\boldsymbol{\theta}) = \mathbf{S}_o(\mathbf{y}_o;\boldsymbol{\theta}) = \mathbb{E}[\mathbf{S}_c(\mathbf{y}_c;\boldsymbol{\theta})|\mathbf{y}_o]$ are given by

$$
\begin{aligned}
\mathbf{S}_{\boldsymbol{\beta}} &= \mathbf{X}^\top\boldsymbol{\Sigma}^{-1}(\widehat{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta}),\\
\mathbf{S}_{\sigma^2} &= -\frac{1}{2}\left[tr\left(\boldsymbol{\Sigma}^{-1}\mathbf{R}\right) - tr\left(\widehat{\mathbf{y}\mathbf{y}^\top}\boldsymbol{\Sigma}^{-1}\mathbf{R}\boldsymbol{\Sigma}^{-1}\right) + \boldsymbol{\beta}^\top\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\mathbf{R}\boldsymbol{\Sigma}^{-1}\left(2\widehat{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta}\right)\right],\\
\mathbf{S}_{\phi} &= -\frac{\sigma^2}{2}\left[tr\left(\boldsymbol{\Sigma}^{-1}\dot{\mathbf{R}}_\phi\right) - tr\left(\widehat{\mathbf{y}\mathbf{y}^\top}\boldsymbol{\Sigma}^{-1}\dot{\mathbf{R}}_\phi\boldsymbol{\Sigma}^{-1}\right) + \boldsymbol{\beta}^\top\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\dot{\mathbf{R}}_\phi\boldsymbol{\Sigma}^{-1}\left(2\widehat{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta}\right)\right],\\
\mathbf{S}_{\tau^2} &= -\frac{1}{2}\left[tr\left(\boldsymbol{\Sigma}^{-1}\right) - tr\left(\widehat{\mathbf{y}\mathbf{y}^\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}\right) + \boldsymbol{\beta}^\top\mathbf{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}\left(2\widehat{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta}\right)\right],
\end{aligned}
$$

where $\mathbf{S}_\phi = \mathbb{E}\left[\frac{\partial \ell_c(\boldsymbol{\theta};\mathbf{y}_c)}{\partial\phi}\Big|\mathbf{y}_o\right]$ and $\dot{\mathbf{R}}_\phi = \frac{\partial \mathbf{R}}{\partial \phi}$. For more details about the functions considered to compute the correlation matrix $\mathbf{R} = \mathbf{R}(\phi)$ and their derivatives, please see Valeriano *et al.* (2021b).

B.2 Extra simulation results

Simulation study I. This simulation aimed to evaluate the impact in the MCEM estimates of the number of random observations used in the Monte Carlo (MC) approximation. Thus, Figure 38 shows the mean bias (black point) of the MCEM estimates ± 1 standard deviation for each parameter by censoring proportion and number of random vectors (L) employed. We observe that the mean bias (for β_0, β_1, and β_2) is close to zero. However, the standard deviations of β_0 seem to be the same independent of L and the censoring rate, while for β_1 and β_2, the deviations increase with the censoring level. Moreover, the mean bias of σ^2, ϕ, and τ^2 is lower than zero, in most cases, except for τ^2 estimated with $L = 20$.

Simulation study II. Figures 39 and 40 display the boxplot for each parameter, sample size, and level of censoring estimated under the MCEM and SAEM algorithms, respectively. Here, the red line denotes the true parameter value. It can be seen that in most cases, the median is close to the real value, and there are few outliers. It is worth noting that the interquartile range increases with the censoring proportion and decreases with the sample size, suggesting the consistency of the estimates for all parameters, except for β_0, which seems to remain constant.

Figure 38 – Simulation I. Mean bias and ±1 standard deviation of MCEM estimates based on 300 MC samples.

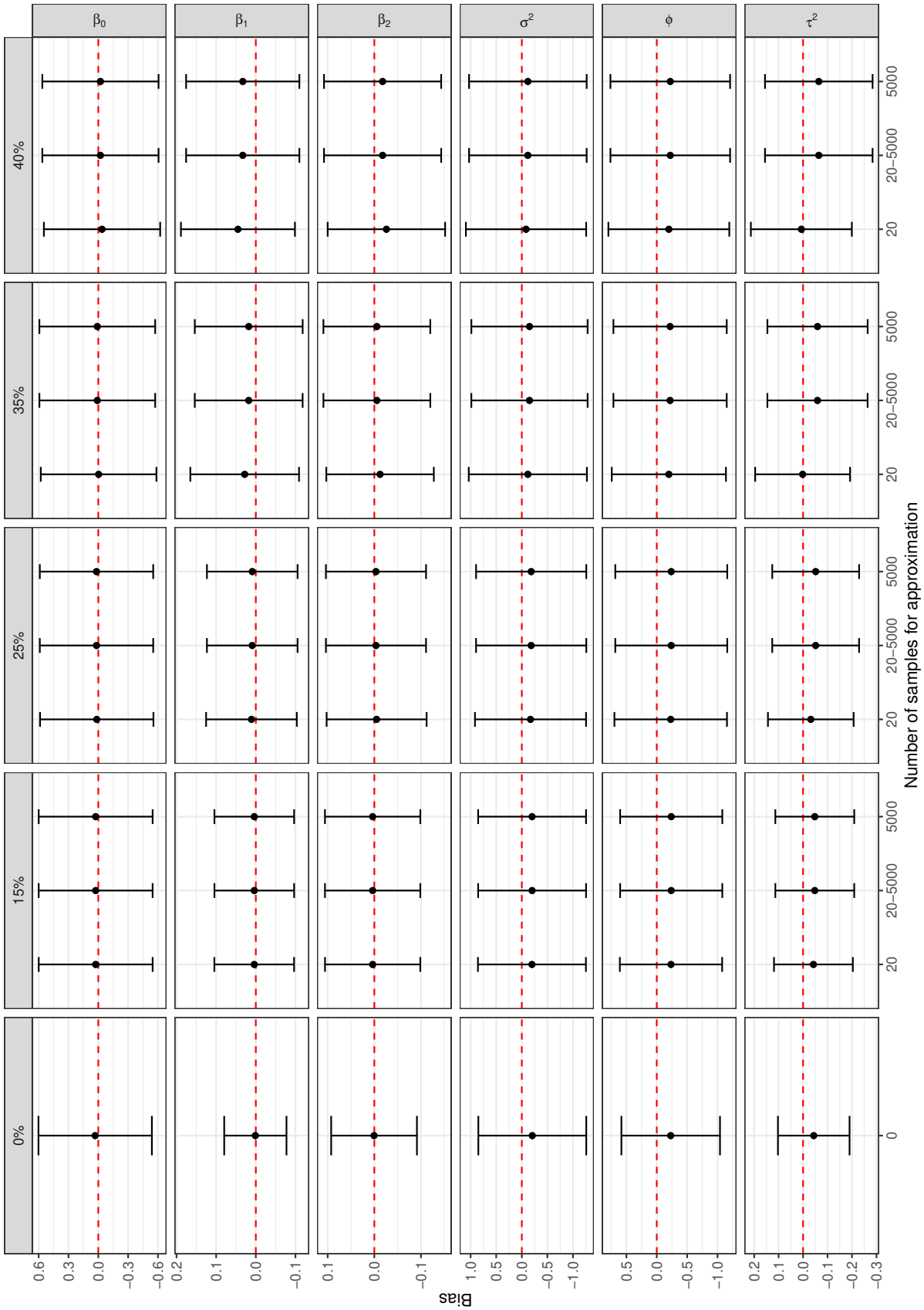


Figure 39 – Simulation II. Boxplot of the MCEM estimates considering different sample sizes and censoring proportions. Red lines represent the true parameter values.

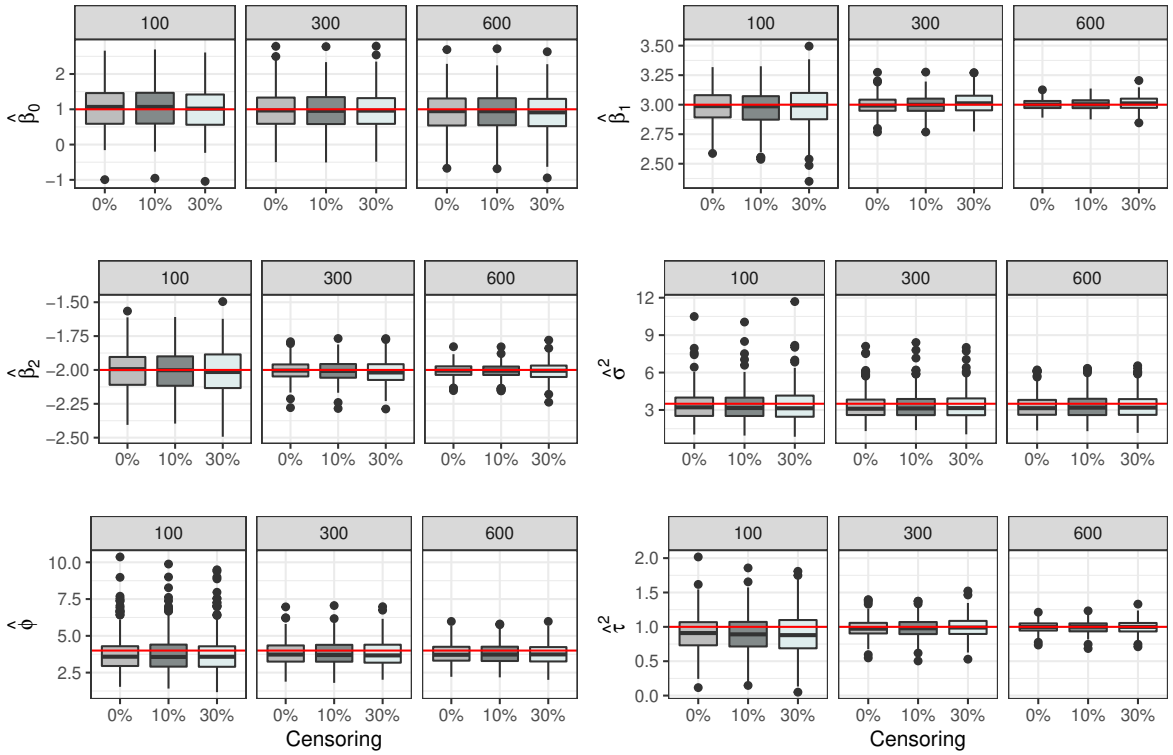
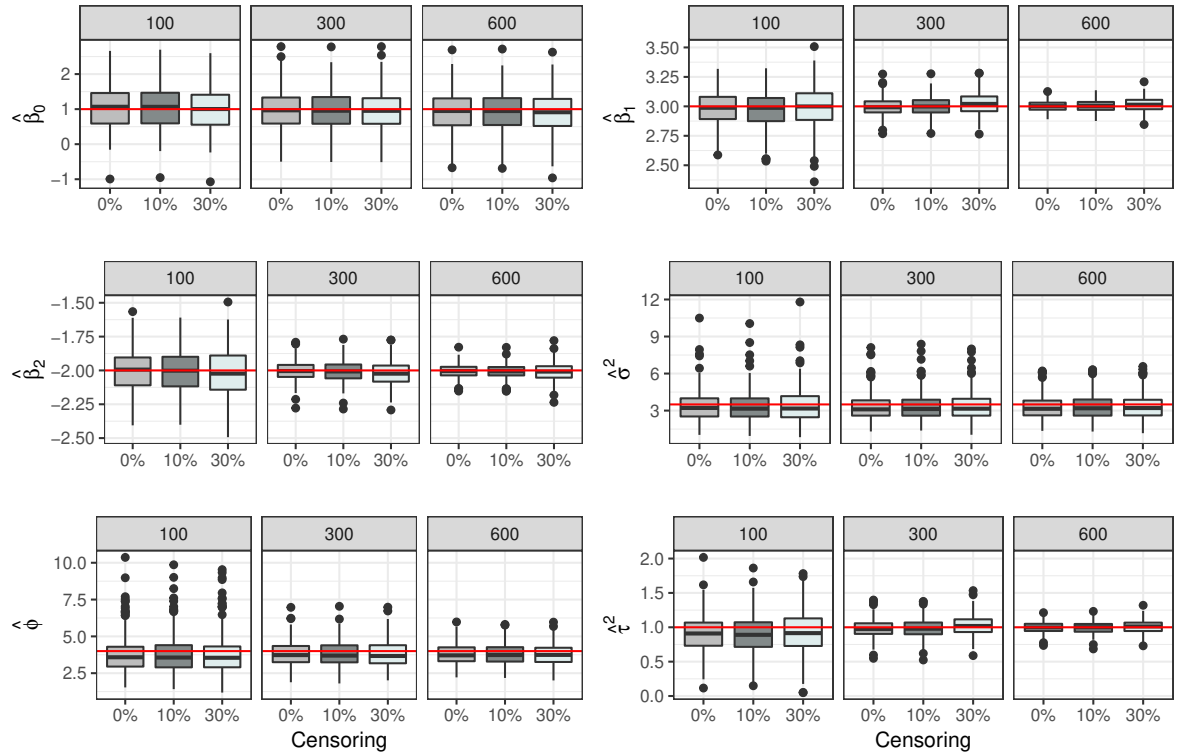


Figure 40 – Simulation II. Boxplot of the SAEM estimates considering different sample sizes and censoring proportions. Red lines represent the true parameter values.



APPENDIX C – CHAPTER 4

C.1 Model estimation details

Consider the censored autoregressive regression model of order p with Student-t innovations defined in Section 4.2, with a complete log-likelihood function given by

$$\ell_c(\boldsymbol{\theta}; \mathbf{y}_c) = \frac{1}{2}\left[g(\nu) - (n-p)\ln\sigma^2 - \frac{1}{\sigma^2}\sum_{t=p+1}^{n} u_t(y_t - \omega_t)^2 \right] + cte, \tag{C.1}$$

where cte represents a constant independent of the vector of parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \sigma^2, \nu)^\top$, $\omega_t = \mathbf{x}_t^\top\boldsymbol{\beta} + (\mathbf{y}_{(t,p)} - \mathbf{X}_{(t,p)}\boldsymbol{\beta})^\top\boldsymbol{\phi}$, $g(\nu) = (n-p)g_1(\nu) + \nu g_2$ with $g_1(\nu) = \nu\ln(\nu/2) - 2\ln\Gamma(\nu/2)$ and $g_2 = \sum_{t=p+1}^{n}(\ln u_t - u_t)$.

Hence, to obtain the ML estimates of $\boldsymbol{\theta}$, we now apply the EM algorithm (Dempster *et al.*, 1977), which at the E-step calculates the conditional expectation of the complete data log-likelihood function, given by

$$\begin{aligned} Q_k(\boldsymbol{\theta}) &= \mathbb{E}\left[\ell_c(\boldsymbol{\theta}; \mathbf{y}_c) \mid \mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}\right] \\ &= \frac{1}{2}\left[\widehat{g(\nu)}^{(k)} - (n-p)\ln\sigma^2 - \frac{1}{\sigma^2}\left(\widehat{uy_*^2}^{(k)} - 2\boldsymbol{\phi}^\top\widehat{u\mathbf{y}_*}^{(k)} + \boldsymbol{\phi}^\top\widehat{u\mathbf{y}_*^2}^{(k)}\boldsymbol{\phi}\right)\right], \end{aligned}$$

where

$$\widehat{g(\nu)}^{(k)} = (n-p)g_1(\nu) + \nu\sum_{t=p+1}^{n}\left(\widehat{\ln(u_t)}^{(k)} - \widehat{u}_t^{(k)}\right),$$

$$\widehat{uy_*^2}^{(k)} = \sum_{t=p+1}^{n}\left(\widehat{uy_t^2}^{(k)} - 2\widehat{uy_t^1}^{(k)}\mu_{(t)} + \widehat{u}_t^{(k)}\mu_{(t)}^2\right),$$

$$\widehat{u\mathbf{y}_*}^{(k)} = \sum_{t=p+1}^{n}\left(\widehat{u\mathbf{y}_t}^{(k)} - \widehat{uy_t^1}^{(k)}\boldsymbol{\mu}_{(t,p)} - \mu_{(t)}\widehat{u\mathbf{y}_t^1}^{(k)} + \widehat{u}_t^{(k)}\mu_{(t)}\boldsymbol{\mu}_{(t,p)}\right),$$

$$\widehat{u\mathbf{y}_*^2}^{(k)} = \sum_{t=p+1}^{n}\left(\widehat{u\mathbf{y}_t^2}^{(k)} - \widehat{u\mathbf{y}_t^1}^{(k)}\boldsymbol{\mu}_{(t,p)}^\top - \boldsymbol{\mu}_{(t,p)}\widehat{u\mathbf{y}_t^1}^{(k)\top} + \widehat{u}_t^{(k)}\boldsymbol{\mu}_{(t,p)}\boldsymbol{\mu}_{(t,p)}^\top\right),$$

with $\mu_{(t)} = \mathbf{x}_t^\top\boldsymbol{\beta}$ and $\boldsymbol{\mu}_{(t,p)} = \mathbf{X}_{(t,p)}\boldsymbol{\beta}$. The remaining expressions are the following conditional expectations: $\widehat{\ln(u_t)}^{(k)} = \mathbb{E}[\ln(U_t) \mid \mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}]$, $\widehat{u}_t^{(k)} = \mathbb{E}[U_t \mid \mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}]$, $\widehat{uy\mathbf{y}_t}^{(k)} = \mathbb{E}[U_t Y_t \mathbf{Y}_{(t,p)} \mid \mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}]$, $\widehat{uy_t^i}^{(k)} = \mathbb{E}[U_t Y_t^i \mid \mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}]$, $\widehat{u\mathbf{y}_t^i}^{(k)} = \mathbb{E}[U_t \mathbf{Y}_{(t,p)}^i \mid \mathbf{V}, \mathbf{C}, \widehat{\boldsymbol{\theta}}^{(k)}]$, for $i \in \{1, 2\}$ and $t \in \{p+1, \ldots, n\}$, such that $\mathbf{Y}_{(t,p)}^1 = \mathbf{Y}_{(t,p)}$ and $\mathbf{Y}_{(t,p)}^2 = \mathbf{Y}_{(t,p)}\mathbf{Y}_{(t,p)}^\top$.

Note that the E-step reduces to the computation of $\widehat{\ln(u_t)}^{(k)}$, $\widehat{u}_t^{(k)}$, $\widehat{uy\mathbf{y}_t}^{(k)}$, $\widehat{uy_t^1}^{(k)}$, $\widehat{uy_t^2}^{(k)}$, $\widehat{u\mathbf{y}_t^1}^{(k)}$, and $\widehat{u\mathbf{y}_t^2}^{(k)}$, for $t \in \{p+1, \ldots, n\}$. However, it is worth mentioning that calculating those expectations becomes a challenge when we have successive censored observations. Therefore, we consider a variation of the EM algorithm called the SAEM algorithm (Delyon *et al.*, 1999), which at the kth iteration proceeds as follows:

E-step:

- Simulation: $\mathbf{y}_m^{(k,\ell)}$ from $f(\mathbf{y}_m \mid \mathbf{V}, \mathbf{y}_o, \widehat{\mathbf{u}}^{(k,\ell-1)}, \widehat{\boldsymbol{\theta}}^{(k)})$ and $u_t^{(k,\ell)}$ from $f(u_t \mid \mathbf{y}^{(k,\ell)}, \widehat{\boldsymbol{\theta}}^{(k)})$, for $t \in \{p+1, \ldots, n\}$ and $\ell \in \{1, \ldots, L\}$, which are the truncated multivariate normal (after conditioning on the observed data $\mathbf{V}$) and gamma distribution, respectively, as will be demonstrated in the next section. Then, construct the vectors $\mathbf{y}^{(k,\ell)} = (\mathbf{y}_o^\top, \mathbf{y}_m^{(k,\ell)\top})^\top$ and $\mathbf{u}^{(k,\ell)} = (u_{p+1}^{(k,\ell)}, \ldots, u_n^{(k,\ell)})^\top$.
- Stochastic Approximation: given the sequence $(\mathbf{y}^{(k,\ell)}, \mathbf{u}^{(k,\ell)})$, the conditional expectations in $Q_k(\boldsymbol{\theta})$ are replaced by the following stochastic approximations:

$$
\begin{aligned}
\widehat{u}_t^{(k)} &= \widehat{u}_t^{(k-1)} + \delta_k \left(\frac{1}{L} \sum_{l=1}^{L} u_t^{(k,\ell)} - \widehat{u}_t^{(k-1)} \right), \\
\widehat{\ln(u_t)}^{(k)} &= \widehat{\ln(u_t)}^{(k-1)} + \delta_k \left(\frac{1}{L} \sum_{l=1}^{L} \ln u_t^{(k,\ell)} - \widehat{\ln(u_t)}^{(k-1)} \right), \\
\widehat{u\mathbf{y}_t}^{(k)} &= \widehat{u\mathbf{y}_t}^{(k-1)} + \delta_k \left(\frac{1}{L} \sum_{l=1}^{L} u_t^{(k,l)} y_t^{(k,\ell)} \mathbf{y}_{(t,p)}^{(k,\ell)} - \widehat{u\mathbf{y}_t}^{(k-1)} \right), \\
\widehat{uy_t^i}^{(k)} &= \widehat{uy_t^i}^{(k-1)} + \delta_k \left(\frac{1}{L} \sum_{l=1}^{L} u_t^{(k,\ell)} y_t^{i(k,\ell)} - \widehat{uy_t^i}^{(k-1)} \right), \\
\widehat{u\mathbf{y}_t^i}^{(k)} &= \widehat{u\mathbf{y}_t^i}^{(k-1)} + \delta_k \left(\frac{1}{L} \sum_{l=1}^{L} u_t^{(k,\ell)} \mathbf{y}_{(t,p)}^{i(k,\ell)} - \widehat{u\mathbf{y}_t^i}^{(k-1)} \right),
\end{aligned}
$$

  for $i \in \{1, 2\}$, such that $\mathbf{y}_{(t,p)}^{1(k,\ell)} = \mathbf{y}_{(t,p)}^{(k,\ell)} = (y_{t-1}^{(k,\ell)}, \ldots, y_{t-p}^{(k,l)})^\top$ and $\mathbf{y}_{(t,p)}^{2(k,\ell)} = \mathbf{y}_{(t,p)}^{(k,\ell)} \mathbf{y}_{(t,p)}^{(k,\ell)\top}$.

M-step:

- Maximization: update $\widehat{\boldsymbol{\theta}}^{(k)}$ by maximizing $\widehat{Q}_k(\boldsymbol{\theta})$ over $\boldsymbol{\theta}$ to obtain a new estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$, which leads to the expressions:

$$
\begin{aligned}
\widehat{\boldsymbol{\phi}}^{(k+1)} &= \left(\widehat{u\mathbf{y}_*^2}^{(k)} \right)^{-1} \widehat{u y \mathbf{y}_*}^{(k)}, \\
\widehat{\sigma}^{2(k+1)} &= \frac{1}{n-p} \left(\widehat{uy_*^2}^{(k)} - 2\widehat{\boldsymbol{\phi}}^{(k+1)\top} \widehat{u y \mathbf{y}_*}^{(k)} + \widehat{\boldsymbol{\phi}}^{(k+1)\top} \widehat{u\mathbf{y}_*^2}^{(k)} \widehat{\boldsymbol{\phi}}^{(k+1)} \right), \\
\widehat{\boldsymbol{\beta}}^{(k+1)} &= \left(\sum_{t=p+1}^{n} \widehat{u}_t^{(k)} \widehat{\boldsymbol{\alpha}}_t^{(k+1)} \widehat{\boldsymbol{\alpha}}_t^{(k+1)\top} \right)^{-1} \sum_{t=p+1}^{n} \left(\widehat{uy_t}^{(k)} - \widehat{\boldsymbol{\phi}}^{(k+1)\top} \widehat{u\mathbf{y}_t}^{(k)} \right) \widehat{\boldsymbol{\alpha}}_t^{(k+1)}, \\
\widehat{\nu}^{(k+1)} &= \underset{\nu}{\operatorname{argmax}} \quad \widehat{g(\nu)}^{(k)},
\end{aligned}
$$

  with $\widehat{\boldsymbol{\alpha}}_t^{(k+1)} = \mathbf{x}_t - \mathbf{X}_{(t,p)}^\top \widehat{\boldsymbol{\phi}}^{(k+1)}$, for $t \in \{p+1, \ldots, n\}$.

C.2 Full conditional distributions

Here we derive the full conditional distributions needed to perform the E-step of the SAEM algorithm, i.e., $f(\mathbf{u} \mid \mathbf{y}, \boldsymbol{\theta})$ and $f(\mathbf{y}_m \mid \mathbf{u}, \mathbf{y}_o, \boldsymbol{\theta})$, where $\mathbf{u} = (u_{p+1}, \ldots, u_n)^\top$

are the mixture weights, and $\mathbf{y}_o$ and $\mathbf{y}_m$ are the observed and the censored/missing part of $\mathbf{y}$, respectively. We first compute the conditional probability density function (pdf) of the mixture weights $\mathbf{u}$, which is given by

$$\begin{aligned} f(\mathbf{u} \mid \mathbf{y}, \boldsymbol{\theta}) &= \frac{f(\mathbf{u}, \mathbf{y}_o, \mathbf{y}_m \mid \boldsymbol{\theta})}{f(\mathbf{y}_o, \mathbf{y}_m \mid \boldsymbol{\theta})} \propto f(\mathbf{u}, \mathbf{y}_o, \mathbf{y}_m \mid \boldsymbol{\theta}) \\ &\propto \prod_{t=p+1}^{n} u_t^{(\nu-1)/2} \exp\left\{ -\frac{u_t}{2} \left(\nu + \frac{(y_t - \mathbf{x}_t^\top \boldsymbol{\beta} - \mathbf{y}_{(t,p)}^\top \boldsymbol{\phi} + \boldsymbol{\beta}^\top \mathbf{X}_{(t,p)}^\top \boldsymbol{\phi})^2}{\sigma^2} \right) \right\}, \end{aligned}$$

which implies that $\{u_t\}$ are independent from each other with $f(u_t \mid \mathbf{y}, \boldsymbol{\theta}) \propto u_t^{a_t - 1} \exp\{-b_t u_t\}$, for $a_t = (\nu+1)/2$ and $b_t = (\nu + \varrho_t^2/\sigma^2)/2$, where $\varrho_t = y_t - \mathbf{x}_t^\top \boldsymbol{\beta} - (\mathbf{y}_{(t,p)} - \mathbf{X}_{(t,p)}\boldsymbol{\beta})^\top \boldsymbol{\phi}$. Notice that $f(u_t \mid \mathbf{y}, \boldsymbol{\theta})$ corresponds to the pdf of a gamma distribution with shape parameter $(\nu+1)/2$ and rate parameter $(\nu + \varrho_t^2/\sigma^2)/2$, i.e.,

$$u_t \overset{ind}{\sim} Gamma\left(\frac{\nu+1}{2}, \frac{\nu + \varrho_t^2/\sigma^2}{2} \right), \quad t \in \{p+1, \ldots, n\}.$$

We now focus on computing the conditional distribution of the censored part $\mathbf{y}_m$. We start by expressing the model defined in Section 4.2 by (4.1)–(4.2) as a function of the first p observations $\mathbf{y}_{(p+1,p)} = (y_p, y_{p-1}, \ldots, y_1)^\top$. To obtain this, we use the VAR(1) model representation, as Zhou *et al.* (2020) suggested, as follows

$$\mathbf{w}_t = \boldsymbol{\Phi} \mathbf{w}_{t-1} + \boldsymbol{\alpha}_t,$$

where $\boldsymbol{\alpha}_t = (\eta_t, 0, \ldots, 0)^\top$ is a vector of dimension p, $\boldsymbol{\Phi} = [\boldsymbol{\phi} \ \mathbf{A}^\top]^\top$ is a $p \times p$ matrix in which $\boldsymbol{\phi}$ is the vector of autoregressive coefficients and $\mathbf{A}$ is a $(p-1) \times p$ matrix with the identity matrix in its first $p-1$ columns and 0s in the last one, and $\mathbf{w}_t = (\tilde{y}_t, \tilde{y}_{t-1}, \ldots, \tilde{y}_{t-p+1})^\top$ is a vector of dimension p with $\tilde{y}_t = y_t - \mathbf{x}_t^\top \boldsymbol{\beta}$, for $t \in \{p, p+1, \ldots, n\}$. Through a recursive process based on VAR(1) form, we have

$$\begin{aligned} t = p+1 \ &\Rightarrow \ \mathbf{w}_{p+1} = \boldsymbol{\Phi} \mathbf{w}_p + \boldsymbol{\alpha}_{p+1}. \\ t = p+2 \ &\Rightarrow \ \mathbf{w}_{p+2} = \boldsymbol{\Phi} \mathbf{w}_{p+1} + \boldsymbol{\alpha}_{p+2} = \boldsymbol{\Phi}(\boldsymbol{\Phi} \mathbf{w}_p + \boldsymbol{\alpha}_{p+1}) + \boldsymbol{\alpha}_{p+2} \\ &\qquad = \boldsymbol{\Phi}^2 \mathbf{w}_p + \boldsymbol{\Phi} \boldsymbol{\alpha}_{p+1} + \boldsymbol{\alpha}_{p+2}. \\ t = p+3 \ &\Rightarrow \ \mathbf{w}_{p+3} = \boldsymbol{\Phi} \mathbf{w}_{p+2} + \boldsymbol{\alpha}_{p+3} = \boldsymbol{\Phi}(\boldsymbol{\Phi}^2 \mathbf{w}_p + \boldsymbol{\Phi} \boldsymbol{\alpha}_{p+1} + \boldsymbol{\alpha}_{p+2}) + \boldsymbol{\alpha}_{p+3} \\ &\qquad = \boldsymbol{\Phi}^3 \mathbf{w}_p + \boldsymbol{\Phi}^2 \boldsymbol{\alpha}_{p+1} + \boldsymbol{\Phi} \boldsymbol{\alpha}_{p+2} + \boldsymbol{\alpha}_{p+3}. \\ &\vdots \\ t = p+k \ &\Rightarrow \ \mathbf{w}_{p+k} = \boldsymbol{\Phi} \mathbf{w}_{p+k-1} + \boldsymbol{\alpha}_{p+k} = \boldsymbol{\Phi}^k \mathbf{w}_p + \sum_{j=0}^{k-1} \boldsymbol{\Phi}^j \boldsymbol{\alpha}_{p+k-j}. \end{aligned}$$

Note that our autoregressive model can be recovered through the first element of the preceding vectors as follows

$$y_{p+k} = \mathbf{x}_{p+k}^\top \boldsymbol{\beta} + \left(\boldsymbol{\Phi}^k\right)_{1.}^\top \left(\mathbf{y}_{(p+1,p)} - \mathbf{X}_{(p+1,p)} \boldsymbol{\beta}\right) + \sum_{j=0}^{k-1} \left(\boldsymbol{\Phi}^j\right)_{11} \eta_{p+k-j}, \tag{C.2}$$

where $\mathbf{\Phi}^k$ represents the matrix $\mathbf{\Phi}$ multiplied by itself k times, $\left(\mathbf{\Phi}^k\right)_{1.}$ is a $p \times 1$ vector whose elements correspond to the first row of $\mathbf{\Phi}^k$, and $\left(\mathbf{\Phi}^k\right)_{11}$ is the element (1,1) of $\mathbf{\Phi}^k$.

From (C.2), it is possible to deduce that the conditional distribution of y_{p+k} given the first p observations $\mathbf{y}_{(p+1,p)}$ and the mixture weights $\mathbf{u}$ is normal, for all $k \in \{1, \ldots, n-p\}$. Hence, our interest is to compute the parameters that characterize the normal distribution, i.e., the mean and the variance-covariance matrix. The conditional expectation is given by

$$\tilde{\mu}_{(k)} = \mathbb{E}[Y_{p+k} \mid \mathbf{u}, \mathbf{y}_{(p+1,p)}, \boldsymbol{\theta}] = \mathbf{x}_{p+k}^\top \boldsymbol{\beta} + \left(\mathbf{\Phi}^k\right)_{1.}^\top (\mathbf{y}_{(p+1,p)} - \mathbf{X}_{(p+1,p)}\boldsymbol{\beta}). \tag{C.3}$$

On the other hand, the conditional variance-covariance can be computed by

$$\begin{aligned} \tilde{\sigma}_{(kl)} &= \operatorname{Cov}(Y_{p+k}, Y_{p+l} \mid \mathbf{u}, \mathbf{y}_{(p+1,p)}, \boldsymbol{\theta}) = \sum_{j=0}^{r-1} \frac{\sigma^2}{u_{p+r-j}} \left(\mathbf{\Phi}^j\right)_{11} \left(\mathbf{\Phi}^{|k-l|+j}\right)_{11} \\ &= \sum_{j=1}^{r} \frac{\sigma^2}{u_{p+j}} \left(\mathbf{\Phi}^{k-j}\right)_{11} \left(\mathbf{\Phi}^{l-j}\right)_{11}, \qquad \text{for } r = \min(k,l). \end{aligned} \tag{C.4}$$

Assuming that the first p observations of the response vector $\mathbf{y}$ are completely observed, i.e., they are not censored or missing. Let $\mathbf{y}$ be partitioned into two vectors, $\mathbf{y}_{1:p} = (y_1, \ldots, y_p)^\top \in \mathbb{R}^p$ containing the first p observed values and $\mathbf{y}_{-1:p} = (y_{p+1}, \ldots, y_n)^\top \in \mathbb{R}^{n-p}$ containing the remaining observations, such that $\mathbf{y} = (\mathbf{y}_{1:p}^\top, \mathbf{y}_{-1:p}^\top)^\top$. Let $\mathbf{y}_o \in \mathbb{R}^{n_o}$ and $\mathbf{y}_m \in \mathbb{R}^{n_m}$ be the observed and the censored/missing part of $\mathbf{y}_{-1:p}$, respectively. Then, using the fact that $y_{p+k} \mid \mathbf{u}, \mathbf{y}_{(p+1,p)}, \boldsymbol{\theta}$ is normally distributed for all $k = 1, \ldots, n-p$, we have that $\mathbf{y}_{-1:p} \mid \mathbf{u}, \mathbf{y}_{1:p}, \boldsymbol{\theta} \overset{d}{=} \mathbf{y}_o, \mathbf{y}_m \mid \mathbf{u}, \mathbf{y}_{1:p}, \boldsymbol{\theta} \overset{d}{=} \mathbf{y}_o, \mathbf{y}_m \mid \mathbf{u}, \mathbf{y}_{(p+1,p)}, \boldsymbol{\theta} \sim N_{n-p}(\tilde{\boldsymbol{\mu}}, \tilde{\mathbf{\Sigma}})$, where the ith element of $\tilde{\boldsymbol{\mu}}$ is $\tilde{\mu}_{(i)}$ and the element (i,j) of $\tilde{\mathbf{\Sigma}}$ is equal to $\tilde{\sigma}_{(ij)}$, which can be computed from (C.3) and (C.4), respectively, for all $i, j = 1, \ldots, n-p$. The expression $\overset{d}{=}$ means "has the same distribution as".

Finally, to compute the conditional distribution of $\mathbf{y}_m \mid \mathbf{u}, \mathbf{y}_o, \mathbf{y}_{(p+1,p)}, \boldsymbol{\theta}$, we rearrange the elements of $\mathbf{y}_{-1:p}, \tilde{\boldsymbol{\mu}}$, and $\tilde{\mathbf{\Sigma}}$ as follows

$$\mathbf{y}_{-1:p} = \begin{pmatrix} \mathbf{y}_o \\ \mathbf{y}_m \end{pmatrix}, \quad \tilde{\boldsymbol{\mu}} = \begin{pmatrix} \tilde{\boldsymbol{\mu}}_o \\ \tilde{\boldsymbol{\mu}}_m \end{pmatrix}, \quad \text{and} \quad \tilde{\mathbf{\Sigma}} = \begin{pmatrix} \tilde{\mathbf{\Sigma}}_{oo} & \tilde{\mathbf{\Sigma}}_{om} \\ \tilde{\mathbf{\Sigma}}_{mo} & \tilde{\mathbf{\Sigma}}_{mm} \end{pmatrix}.$$

Then, using the results for the conditional distribution of a normal distribution, we obtain

$$\mathbf{y}_m \mid \mathbf{u}, \mathbf{y}_o, \mathbf{y}_{(p+1,p)}, \boldsymbol{\theta} \sim N_{n_m}(\mathbf{m}, \mathbf{S}),$$

with $\mathbf{m} = \tilde{\boldsymbol{\mu}}_m + \tilde{\mathbf{\Sigma}}_{mo}\tilde{\mathbf{\Sigma}}_{oo}^{-1}(\mathbf{y}_o - \tilde{\boldsymbol{\mu}}_o)$ and $\mathbf{S} = \tilde{\mathbf{\Sigma}}_{mm} - \tilde{\mathbf{\Sigma}}_{mo}\tilde{\mathbf{\Sigma}}_{oo}^{-1}\tilde{\mathbf{\Sigma}}_{om}$.

C.3 Gradient and hessian matrix of the complete-data log-likelihood function

Here, we derive the calculations required to obtain the observed Fisher information matrix by the stochastic approximation procedure described in Section 1.2. Let

$\ell_c(\boldsymbol{\theta}; \mathbf{y}_c)$ be the complete log-likelihood function defined by (C.1). Then, the elements of the gradient vector are

$$
\begin{aligned}
\frac{\partial \ell_c(\boldsymbol{\theta}; \mathbf{y}_c)}{\partial \nu} &= \frac{n-p}{2}\left(\ln\left(\frac{\nu}{2}\right) + 1 - \psi\left(\frac{\nu}{2}\right)\right) + \frac{1}{2}\sum_{t=p+1}^{n}\left(\ln u_t - u_t\right), \\
\frac{\partial \ell_c(\boldsymbol{\theta}; \mathbf{y}_c)}{\partial \sigma^2} &= -\frac{n-p}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=p+1}^{n} u_t\left(y_t - \omega_t\right)^2, \\
\frac{\partial \ell_c(\boldsymbol{\theta}; \mathbf{y}_c)}{\partial \boldsymbol{\phi}} &= \frac{1}{\sigma^2}\sum_{t=p+1}^{n} u_t\left[\mathbf{z}_{(t,p)}(y_t - \mathbf{x}_t^\top \boldsymbol{\beta}) - \mathbf{z}_{(t,p)}\mathbf{z}_{(t,p)}^\top \boldsymbol{\phi}\right], \\
\frac{\partial \ell_c(\boldsymbol{\theta}; \mathbf{y}_c)}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma^2}\sum_{t=p+1}^{n} u_t\left[\boldsymbol{\alpha}_t(y_t - \mathbf{y}_{(t,p)}^\top \boldsymbol{\phi}) - \boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^\top \boldsymbol{\beta}\right],
\end{aligned}
$$

with $\mathbf{z}_{(t,p)} = \mathbf{y}_{(t,p)} - \mathbf{X}_{(t,p)}\boldsymbol{\beta}$, $\boldsymbol{\alpha}_t = \mathbf{x}_t - \mathbf{X}_{(t,p)}^\top \boldsymbol{\phi}$, and $\psi(x) = \dfrac{\mathrm{d}}{\mathrm{d}x}\ln\Gamma(x) = \dfrac{\Gamma'(x)}{\Gamma(x)}$.

Besides, the elements of the Hessian matrix are given by:

$$
\begin{aligned}
\frac{\partial^2 \ell_c(\boldsymbol{\theta}; \mathbf{y}_c)}{\partial \nu^2} &= \frac{n-p}{2}\left(\frac{1}{\nu} - \frac{1}{2}\psi_1\left(\frac{\nu}{2}\right)\right), \\
\frac{\partial^2 \ell_c(\boldsymbol{\theta}; \mathbf{y}_c)}{\partial \sigma^2 \partial \nu} &= 0, \quad \frac{\partial^2 \ell_c(\boldsymbol{\theta}; \mathbf{y}_c)}{\partial \boldsymbol{\phi} \partial \nu} = \mathbf{0}, \quad \frac{\partial^2 \ell_c(\boldsymbol{\theta}; \mathbf{y}_c)}{\partial \boldsymbol{\beta} \partial \nu} = \mathbf{0}, \\
\frac{\partial^2 \ell_c(\boldsymbol{\theta}; \mathbf{y}_c)}{\partial (\sigma^2)^2} &= \frac{n-p}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{t=p+1}^{n} u_t\left(y_t - \omega_t\right)^2, \\
\frac{\partial^2 \ell_c(\boldsymbol{\theta}; \mathbf{y}_c)}{\partial \boldsymbol{\phi} \partial \sigma^2} &= -\frac{1}{\sigma^4}\sum_{t=p+1}^{n} u_t\left[\mathbf{z}_{(t,p)}\left(y_t - \mathbf{x}_t^\top \boldsymbol{\beta}\right) - \mathbf{z}_{(t,p)}\mathbf{z}_{(t,p)}^\top \boldsymbol{\phi}\right], \\
\frac{\partial^2 \ell_c(\boldsymbol{\theta}; \mathbf{y}_c)}{\partial \boldsymbol{\beta} \partial \sigma^2} &= -\frac{1}{\sigma^4}\sum_{t=p+1}^{n} u_t\left[\boldsymbol{\alpha}_t\left(y_t - \mathbf{y}_{(t,p)}^\top \boldsymbol{\phi}\right) - \boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^\top \boldsymbol{\beta}\right], \\
\frac{\partial^2 \ell_c(\boldsymbol{\theta}; \mathbf{y}_c)}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}^\top} &= -\frac{1}{\sigma^2}\sum_{t=p+1}^{n} u_t\, \mathbf{z}_{(t,p)}\mathbf{z}_{(t,p)}^\top, \\
\frac{\partial^2 \ell_c(\boldsymbol{\theta}; \mathbf{y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\phi}^\top} &= \frac{1}{\sigma^2}\sum_{t=p+1}^{n} u_t\left[\left(\mathbf{z}_{(t,p)}^\top \boldsymbol{\phi} + \mathbf{x}_t^\top \boldsymbol{\beta} - y_t\right)\mathbf{X}_{(t,p)}^\top - \boldsymbol{\alpha}_t\mathbf{z}_{(t,p)}^\top\right], \\
\frac{\partial^2 \ell_c(\boldsymbol{\theta}; \mathbf{y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= -\frac{1}{\sigma^2}\sum_{t=p+1}^{n} u_t\, \boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^\top,
\end{aligned}
$$

with $\psi_1(x) = \dfrac{\mathrm{d}^2}{\mathrm{d}x^2}\ln\Gamma(x)$.

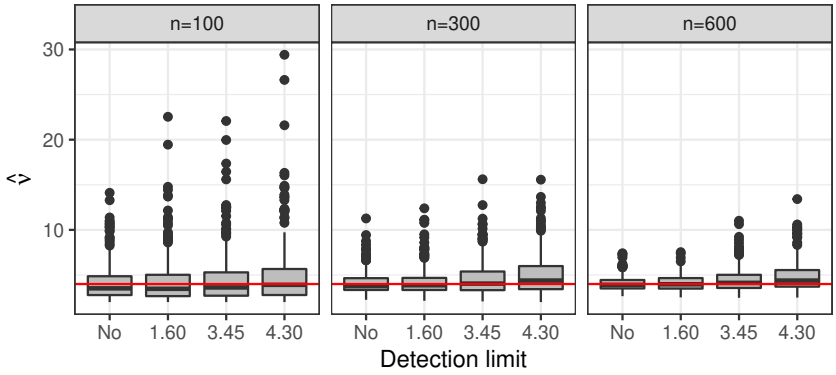
C.4 Additional numerical results

This section displays additional results for the simulation study I, the Ammonia-nitrogen application in Section 4.4, and a new application related to total phosphorus concentration analysis.

C.4.1 Simulation study I: Asymptotic properties for the degrees of freedom

This simulation study aimed to provide empirical evidence about the consistency of the ML estimates for the degrees of freedom ν under different scenarios. Boxplots for the degrees of freedom ν estimates under different settings of sample sizes and detection limits are shown in Figure 41. The red line denotes the true parameter value. Here, it is possible to observe that the median of the estimates is close to the true value ($\nu = 4$), independent of the sample size and detection limit. Furthermore, interquartile ranges decrease as sample sizes increase, suggesting the consistency of the estimates.

Figure 41 – Simulation I. Boxplot of the estimates of ν in the CARt(2) model by sample size and detection limits.



C.4.2 Application: Ammonia-nitrogen concentration

For the sake of comparison, we fit a censored normal regression model given by

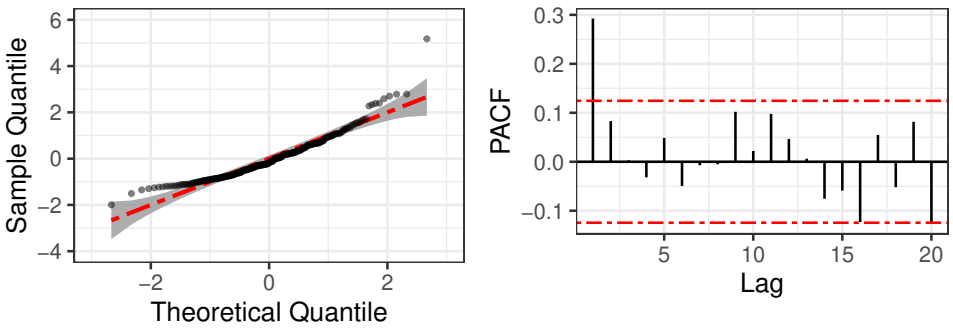
$$y_t = \beta_0 + \beta_{\text{DO}}\text{DO}_t + \beta_{\text{pH}}\text{pH}_t + \beta_{\text{T}}\text{T}_t + \xi_t,$$

where y_t denotes the square root of the ammonia-nitrogen levels (in μg/L) with 42.34% of the observations censored/missing, and ξ_t represents the independent and normally distributed error term. Figure 42 displays the quantile-quantile plot and the partial autocorrelation coefficients of the standardized residuals computed for the fitted model with the censored and missing observations imputed by the conditional mean. Notice that the assumption of independent errors is invalid; hence, a model with autocorrelated errors should be considered.

C.4.3 Application: Total phosphorus concentration

In this analysis, we are interested in tracking the phosphorus concentration levels over time as an indicator of the river water quality since, for instance, excessive phosphorus in surface water may result in eutrophication. This dataset is available in the R package `ARCensReg`, and it was previously described in Subsection 1.3.3.

Figure 42 – Quantile residuals for a model with normally distributed independent errors fitted to the ammonia-nitrogen concentration dataset.



As mentioned in Wang & Chan (2018), P levels are generally correlated with the water discharge (Q), measured in cubic feet per second; then, our objective is to explore the relationship between P and Q when the response contains censored and missing observations. The dataset was train-test split to evaluate the prediction accuracy. The training dataset consists of 169 observations, where 20.71% are left-censored or missing, while the testing dataset contains 12 fully observed values. Following Wang & Chan (2018), the logarithmic transformation of P and Q must be considered to make the P-Q relationship linear. Then, we fitted the following model:

$$\ln(P_t) = \sum_{j=1}^{4} [\beta_{0j} S_{jt} + \beta_{1j} S_{jt} \ln(Q_t)] + \xi_t, \quad t = p+1, \ldots, n, \tag{C.5}$$

where ξ_t follows an autoregressive model and S_j is a dummy seasonal variable for quarters $j = 1, 2, 3$, and 4. The first quarter corresponds from January to March, the second from April to June, and so on. In this model, β_{0j} and β_{1j} are respectively the intercept and slope for quarter j, for $j = 1, 2, 3$, and 4.

This dataset was previously analyzed by Schumacher *et al.* (2017). The authors concluded that a censored autoregressive model of order 1 was the best to fit this data based on information criteria and mean squared prediction error (MSPE). The authors also pointed out that the data has some influential observations, so it seems reasonable to consider a model with innovations following a heavy-tailed distribution. However, it is worth noting that considering Student-t innovations for the model defined in (C.5) implies that the conditional distribution of the phosphorus concentration itself is log-Student t (or log-t for short). The log-t distribution has applications in, for instance, finance (Barroso *et al.*, 2020) and hydrology (Viglione, 2010), and it is extremely heavy-tailed with an infinite mean. Additionally, following Marchenko & Genton (2010), the log-t model may be preferable to the normal model, especially for estimating extreme events, due to its robustness.

For comparison purposes, the model in (C.5) was fitted considering an AR(1) model for the regression error with innovations η_t being independent and identically

distributed as $N(0, \sigma^2)$ and $t(0, \sigma^2, \nu)$ (denoted by CAR(1) and CARt(1), respectively). Parameter estimates and their corresponding standard errors (SEs) are displayed in Table 24. The MSPE and the mean absolute prediction error (MAPE) were also computed considering one-step-ahead predictions for the testing dataset. These criteria indicate that the heavy-tailed Student-t ($\widehat{\nu} = 5.394$) model provides better predictions than the normal model for the phosphorus concentration data. We can also note that the estimates for $\beta_{0j}, j = 1, 2, 3, 4$ under the CARt(1) model are negative and greater than the estimates obtained from the CAR(1) model. On the other hand, estimates for slopes β_{1j} are all positive real numbers and, except for the second quarter, they are significantly different from zero, evidencing the correlation between the water discharge and the phosphorus concentration. Regarding the autoregressive coefficient ϕ_1, both models estimated similar values.

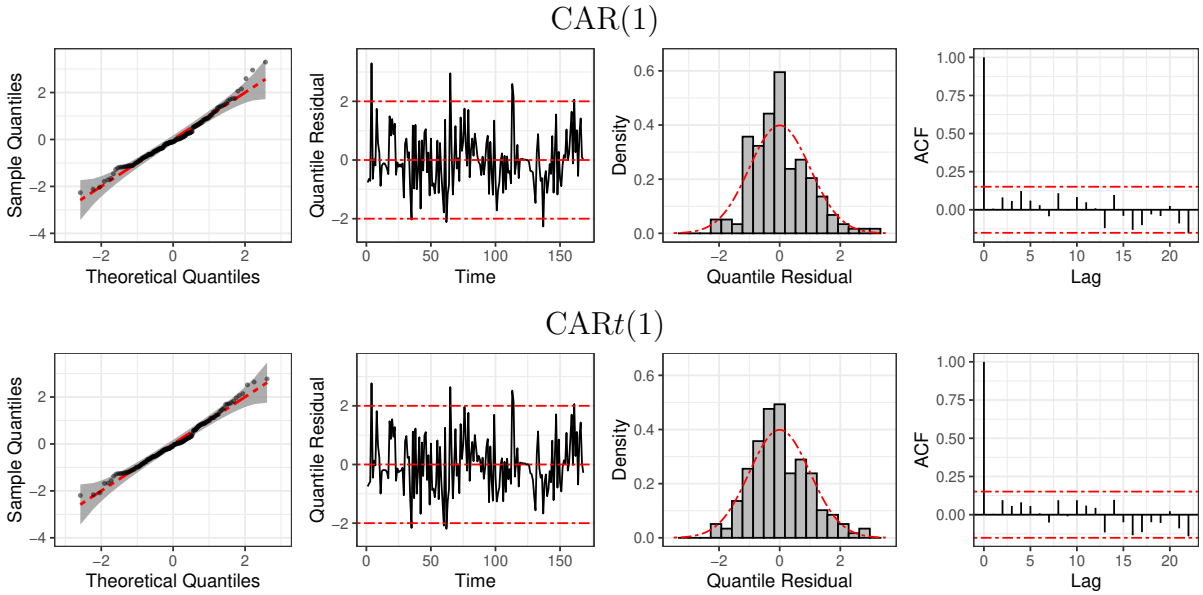
Table 24 – Phosphorus concentration. Parameter estimates and their standard errors (SE) for the CARt(1) and CAR(1) models.

Parameters	CARt(1)			CAR(1)		
	Estimate	SE	95% CI	Estimate	SE	95% CI
β_{01}	-4.345	0.623	(-5.565 , -3.124)	-4.691	0.544	(-5.758 , -3.624)
β_{02}	-2.748	0.767	(-4.252 , -1.244)	-3.029	0.772	(-4.542 , -1.516)
β_{03}	-4.142	0.422	(-4.970 , -3.315)	-4.177	0.457	(-5.072 , -3.283)
β_{04}	-4.657	0.543	(-5.722 , -3.593)	-5.020	0.573	(-6.143 , -3.896)
β_{11}	0.296	0.109	(0.083 , 0.509)	0.376	0.091	(0.197 , 0.555)
β_{12}	0.142	0.108	(-0.069 , 0.353)	0.186	0.108	(-0.026 , 0.398)
β_{13}	0.363	0.070	(0.226 , 0.500)	0.365	0.077	(0.214 , 0.515)
β_{14}	0.352	0.097	(0.162 , 0.543)	0.413	0.103	(0.212 , 0.614)
ϕ_1	-0.086	0.088		-0.075	0.090	
σ^2	0.181	0.040		0.268	0.033	
ν	5.394	2.661		-	-	
MSPE	0.102			0.126		
MAPE	0.241			0.254		

Figure 43 presents from left to right: a Q-Q plot, a residual vs. time plot, a histogram, and the sample autocorrelation plot for residual analysis for both models. The Q-Q plot for the CAR(1) model (top) presents some points outside the confidence bands on the upper tail, and we also see a few larger residual values from the scatterplot and histogram. For the CARt(1) model (bottom), we see in the Q-Q plot that all points form a roughly straight line and lie within the confidence bands; further, the histogram seems to correspond to a normally distributed variable, and the dispersion plot seems to be related to independent residuals. Therefore, the CARt(1) model fits the phosphorus concentration data better than the CAR(1) model.

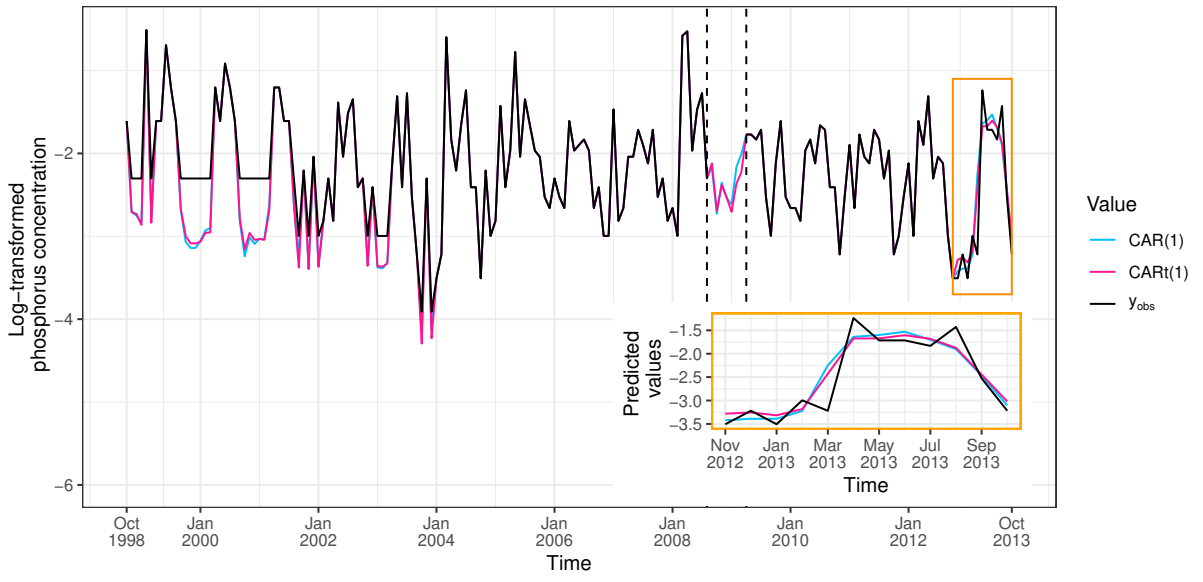
Figure 44 shows the observed values (solid black line) and the imputed values for the censored and missing observations from October 1998 to October 2012 (training

Figure 43 – Phosphorus concentration. Plots of residuals for the CAR(1) and CARt(1) models.



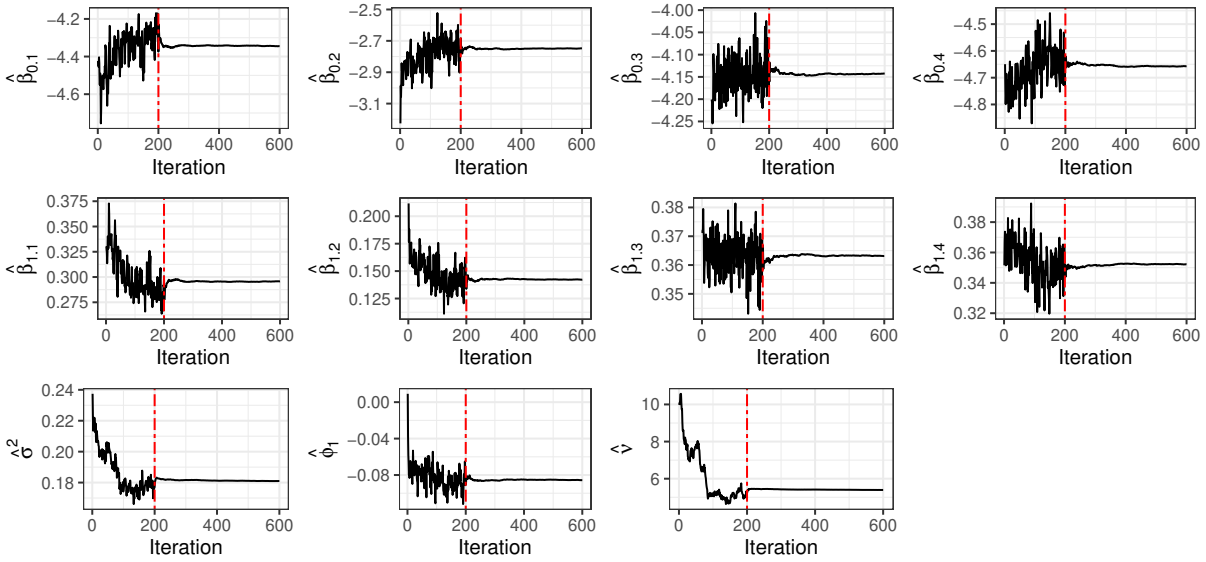
dataset). The vertical dashed black lines denote the period with missing values. The predicted values for the testing data are displayed in the yellow box, where the light blue line represents the values predicted through the CAR(1) model and the pink line predicted from the CARt(1) model. We see slight differences between the predicted values obtained under both fitted models. Besides, the general behavior of the imputed values for the missing period seems to capture the seasonal behavior of the time series well and is similar for both models.

Figure 44 – Phosphorus concentration. Observed (black solid line) and predicted values considering Student-t (pink line) and normal (light blue line) innovations. Black dashed lines represent the period with missing observations.



In addition, for assessing the convergence of SAEM parameter estimates, convergence plots for our proposal are displayed in Figure 45.

Figure 45 – Phosphorus concentration. Convergence of the SAEM parameter estimates for the CAR$t(1)$ model.



APPENDIX D – CHAPTER 5

D.1 Proof of Propositions

Proof of Proposition 6. Consider the partition $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ and the corresponding partitions of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\varphi}$. We based our proof on the factorization of $f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{Y}_1,\mathbf{Y}_2}(\mathbf{y}_1, \mathbf{y}_2)$ as $f_{\mathbf{Y}_1,\mathbf{Y}_2}(\mathbf{y}_1, \mathbf{y}_2) = f_{\mathbf{Y}_1}(\mathbf{y}_1) f_{\mathbf{Y}_2|\mathbf{Y}_1=\mathbf{y}_1}(\mathbf{y}_2)$. First, for the symmetric part, we have that

$$t_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = t_{p_1}(\mathbf{y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \nu) t_{p_2}(\mathbf{y}_2; \boldsymbol{\mu}_{2.1}, \tilde{\boldsymbol{\Sigma}}_{22.1}, \nu + p_1), \tag{D.1}$$

with $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$, $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, $\tilde{\boldsymbol{\Sigma}}_{22.1} = \boldsymbol{\Sigma}_{22.1}/\nu^2(\mathbf{y}_1)$, and $\nu^2(\mathbf{y}_1) = (\nu + p_1)/(\nu + \delta(\mathbf{y}_1))$.

Let now $c_{12} = (1 + \boldsymbol{\varphi}_2^\top \boldsymbol{\Sigma}_{22.1}\boldsymbol{\varphi}_2)^{-1/2}$, $\tilde{\boldsymbol{\varphi}}_1 = \boldsymbol{\varphi}_1 + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\varphi}_2$, $\tau_{2.1} = \nu(\mathbf{y}_1)(\tau + \tilde{\boldsymbol{\varphi}}_1^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))$, and $\nu_{2.1} = \nu + p_1$. By noting after some straightforward algebra that $\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu}) = \boldsymbol{\varphi}^\top(\mathbf{y} - \boldsymbol{\mu}) = \tilde{\boldsymbol{\varphi}}_1^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1) + \boldsymbol{\varphi}_2^\top(\mathbf{y}_2 - \boldsymbol{\mu}_{2.1})$ and $\boldsymbol{\lambda}^\top\boldsymbol{\lambda} = \boldsymbol{\varphi}^\top\boldsymbol{\Sigma}\boldsymbol{\varphi} = \tilde{\boldsymbol{\varphi}}_1^\top\boldsymbol{\Sigma}_{11}\tilde{\boldsymbol{\varphi}}_1 + \boldsymbol{\varphi}_2^\top\boldsymbol{\Sigma}_{22.1}\boldsymbol{\varphi}_2$, we obtain

$$T\left((\tau_1 + \tilde{\boldsymbol{\lambda}}_1^\top\boldsymbol{\Sigma}_{11}^{-1/2}(\mathbf{y}_1 - \boldsymbol{\mu}_1))\nu(\mathbf{y}_1); \nu + p_1\right) = T\left(\frac{\tau_{2.1}}{(1 + \boldsymbol{\lambda}_{2.1}^\top\boldsymbol{\lambda}_{2.1})^{1/2}}; \nu_{2.1}\right), \tag{D.2}$$

and

$$T\left(\frac{\tau}{(1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}}; \nu\right) = T\left(\frac{\tau_1}{(1 + \tilde{\boldsymbol{\lambda}}_1^\top\tilde{\boldsymbol{\lambda}}_1)^{1/2}}; \nu\right), \tag{D.3}$$

where $\tilde{\boldsymbol{\lambda}}_1 = c_{12}\boldsymbol{\Sigma}_{11}^{1/2}\tilde{\boldsymbol{\varphi}}_1$, $\tau_1 = c_{12}\tau$ and $\boldsymbol{\lambda}_{2.1} = \boldsymbol{\Sigma}_{22.1}^{1/2}\boldsymbol{\varphi}_2$. Additionally, it is easy to see that

$$\nu^2(\mathbf{y}) = \frac{\nu + p}{\nu + \delta(\mathbf{y})} = \frac{\nu + p_1}{\nu + \delta(\mathbf{y}_1)}\left(\frac{\nu_{2.1} + p_2}{\nu_{2.1} + \delta(\mathbf{y}_2; \boldsymbol{\mu}_{2.1}, \tilde{\boldsymbol{\Sigma}}_{22.1})}\right) = \nu^2(\mathbf{y}_1)\nu^2_{\mathbf{Y}_{2.1}}(\mathbf{y}_2).$$

From this last equation, it holds that

$$T\big((A + \tau)\nu(\mathbf{y}); \nu + p\big) = T\big((\tau_{2.1} + \boldsymbol{\lambda}_{2.1}^\top\tilde{\boldsymbol{\Sigma}}_{22.1}^{-1/2}(\mathbf{y}_2 - \boldsymbol{\mu}_{2.1}))\nu_{\mathbf{Y}_{2.1}}(\mathbf{y}_2); \nu_{2.1} + p_2\big), \tag{D.4}$$

with $A = \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu})$, Hence, using (D.1), (D.2) and (D.3), we can rewrite the density of $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ as

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= t_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)\frac{T((\tau + \boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu}))\nu(\mathbf{y}); \nu + p)}{T(\tau/(1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})^{1/2}; \nu)} \\ &= t_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)\frac{T\big((\tau_{2.1} + \boldsymbol{\lambda}_{2.1}^\top\tilde{\boldsymbol{\Sigma}}_{22.1}^{-1/2}(\mathbf{y}_2 - \boldsymbol{\mu}_{2.1}))\nu_{\mathbf{Y}_{2.1}}(\mathbf{y}_2); \nu_{2.1} + p_2\big)}{T\left(\tau_1/(1 + \tilde{\boldsymbol{\lambda}}_1^\top\tilde{\boldsymbol{\lambda}}_1)^{1/2}; \nu\right)} \end{aligned}$$

$$
\begin{aligned}
f_{\mathbf{Y}}(\mathbf{y}) &= t_{p_1}(\mathbf{y}_1;\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11},\nu)\frac{T\big((\tau_1+\tilde{\boldsymbol{\lambda}}_1^\top\boldsymbol{\Sigma}_{11}^{-1/2}(\mathbf{y}_1-\boldsymbol{\mu}_1))\nu(\mathbf{y}_1);\nu+p_1\big)}{T\left(\tau_1/(1+\tilde{\boldsymbol{\lambda}}_1^\top\tilde{\boldsymbol{\lambda}}_1)^{1/2};\nu\right)}\\
&\quad\times t_{p_2}(\mathbf{y}_2;\boldsymbol{\mu}_{2.1},\tilde{\boldsymbol{\Sigma}}_{22.1},\nu_{2.1})\frac{T\big((\tau_{2.1}+\boldsymbol{\lambda}_{2.1}^\top\tilde{\boldsymbol{\Sigma}}_{22.1}^{-1/2}(\mathbf{y}_2-\boldsymbol{\mu}_{2.1}))\nu_{\mathbf{Y}_{2.1}}(\mathbf{y}_2);\nu_{2.1}+p_2\big)}{T\left(\tau_{2.1}/(1+\boldsymbol{\lambda}_{2.1}^\top\boldsymbol{\lambda}_{2.1})^{1/2};\nu_{2.1}\right)}\\
&= EST_{p_1}(\mathbf{y}_1;\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11},\tilde{\boldsymbol{\lambda}}_1,\tau_1,\nu)\times EST_{p_2}(\mathbf{y}_2;\boldsymbol{\mu}_{2.1},\tilde{\boldsymbol{\Sigma}}_{22.1},\boldsymbol{\lambda}_{2.1},\tau_{2.1},\nu+p_1).
\end{aligned}
$$

□

Proof of Proposition 7. First note that $\mathbf{Y}\mid(\boldsymbol{\alpha}\leqslant\mathbf{Y}\leqslant\boldsymbol{\beta})\sim\mathrm{TST}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu;(\boldsymbol{\alpha},\boldsymbol{\beta}))$. By direct integration of the simplified expressions (5.15) and (5.16), it is readily that

$$
\begin{aligned}
\mathbb{E}\left[\phi(\boldsymbol{\theta},\mathbf{Y})\,g(\mathbf{Y})\mid\boldsymbol{\alpha}\leqslant\mathbf{Y}\leqslant\boldsymbol{\beta}\right] &= \frac{2\Gamma(\frac{\nu+1}{2})/\Gamma(\frac{\nu}{2})}{\sqrt{\pi\nu(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}}\int_\alpha^\beta\frac{t_p(\mathbf{y};\boldsymbol{\mu},\frac{\nu}{\nu+1}\boldsymbol{\Gamma},\nu+1)}{ST_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)}\frac{ST_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)}{\mathcal{P}_p(\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)}g(\mathbf{y})\mathrm{d}\mathbf{y}\\
&= \frac{2\Gamma(\frac{\nu+1}{2})/\Gamma(\frac{\nu}{2})}{\sqrt{\pi\nu(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}}\frac{1}{\mathcal{P}_p(\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)}\int_\alpha^\beta g(\mathbf{y})t_p(\mathbf{y};\boldsymbol{\mu},\tfrac{\nu}{\nu+1}\boldsymbol{\Gamma},\nu+1)\mathrm{d}\mathbf{y}\\
&= \frac{2}{\sqrt{\pi\nu(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\frac{P_p(\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\mu},\frac{\nu}{\nu+1}\boldsymbol{\Gamma},\nu+1)}{\mathcal{P}_p(\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)}\mathbb{E}[g(\mathbf{W}_1)]
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}_{UT\mathbf{Y}}[U\,g(\mathbf{Y})\mid\boldsymbol{\alpha}\leqslant\mathbf{Y}\leqslant\boldsymbol{\beta}] &= \int_\alpha^\beta\frac{ST_p(\mathbf{y};\boldsymbol{\mu},\frac{\nu}{\nu+2}\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu+2)}{ST_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)}\frac{ST_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)}{\mathcal{P}_p(\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)}g(\mathbf{y})\mathrm{d}\mathbf{y}\\
&= \frac{1}{\mathcal{P}_p(\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)}\int_\alpha^\beta g(\mathbf{y})ST_p(\mathbf{y};\boldsymbol{\mu},\tfrac{\nu}{\nu+2}\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu+2)\mathrm{d}\mathbf{y}\\
&= \frac{\mathcal{P}_p(\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\mu},\frac{\nu}{\nu+2}\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu+2)}{\mathcal{P}_p(\boldsymbol{\alpha},\boldsymbol{\beta};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)}\mathbb{E}[g(\mathbf{W}_2)],
\end{aligned}
$$

$\mathbf{W}_1\sim Tt_p(\boldsymbol{\mu},\frac{\nu}{\nu+1}\boldsymbol{\Gamma},\nu+1;(\boldsymbol{\alpha},\boldsymbol{\beta}))$ and $\mathbf{W}_2\sim\mathrm{TST}_p(\boldsymbol{\mu},\frac{\nu}{\nu+2}\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu+2;(\boldsymbol{\alpha},\boldsymbol{\beta}))$. □

Proof of Proposition 8. It follows from the conditional distribution of an ST distribution that $\mathbf{Y}_2\mid(\mathbf{Y}_1,\boldsymbol{\alpha}_2\leqslant\mathbf{Y}_2\leqslant\boldsymbol{\beta}_2)\sim\mathrm{TEST}_{p_2}(\boldsymbol{\mu}_{2.1},\tilde{\boldsymbol{\Sigma}}_{22.1},\boldsymbol{\lambda}_{2.1},\tau_{2.1},\nu_{2.1};(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2))$, with conditional parameters as in Proposition 6. It is straightforward that

$$
\begin{aligned}
&\mathbb{E}\big[\phi(\boldsymbol{\theta},\mathbf{Y})\,g(\mathbf{Y})\mid\mathbf{Y}_1,\boldsymbol{\alpha}_2\leqslant\mathbf{Y}_2\leqslant\boldsymbol{\beta}_2\big]\\
&=\frac{2}{\sqrt{\pi\nu\,(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\int_{\alpha_2}^{\beta_2}\frac{t_p(\mathbf{y};\boldsymbol{\mu},\frac{\nu}{\nu+1}\boldsymbol{\Gamma},\nu+1)}{ST_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)}\frac{EST_{p_2}(\mathbf{y}_2;\boldsymbol{\mu}_{2.1},\tilde{\boldsymbol{\Sigma}}_{22.1},\boldsymbol{\lambda}_{2.1},\tau_{2.1},\nu_{2.1})}{\mathcal{P}_{p_2}(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2;\boldsymbol{\mu}_{2.1},\tilde{\boldsymbol{\Sigma}}_{22.1},\boldsymbol{\lambda}_{2.1},\tau_{2.1},\nu_{2.1})}g_2(\mathbf{y}_2)\mathrm{d}\mathbf{y}_2\\
&=\frac{2}{\sqrt{\pi\nu\,(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\frac{t_{p_1}(\mathbf{y}_1;\boldsymbol{\mu}_1,\frac{\nu}{\nu+1}\boldsymbol{\Gamma}_{11},\nu+1)}{ST_{p_1}(\mathbf{y}_1;\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11},\tilde{\boldsymbol{\lambda}}_1,\nu)}g_1(\mathbf{Y}_1)\\
&\quad\times\int_{\alpha_2}^{\beta_2}\frac{t_{p_2}(\mathbf{y}_2;\boldsymbol{\mu}_{2.1},\frac{\nu_{2.1}}{\nu_{2.1}+1}\tilde{\boldsymbol{\Gamma}}_{22.1},\nu_{2.1}+1)}{EST_{p_2}(\mathbf{y}_2;\boldsymbol{\mu}_{2.1},\tilde{\boldsymbol{\Sigma}}_{22.1},\boldsymbol{\lambda}_{2.1},\tau_{2.1},\nu_{2.1})}\frac{EST_{p_2}(\mathbf{y}_2;\boldsymbol{\mu}_{2.1},\tilde{\boldsymbol{\Sigma}}_{22.1},\boldsymbol{\lambda}_{2.1},\tau_{2.1},\nu_{2.1})}{\mathcal{P}_{p_2}(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2;\boldsymbol{\mu}_{2.1},\tilde{\boldsymbol{\Sigma}}_{22.1},\boldsymbol{\lambda}_{2.1},\tau_{2.1},\nu_{2.1})}g(\mathbf{y})\mathrm{d}\mathbf{y}_2\\
&=\frac{2}{\sqrt{\pi\nu\,(1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda})}}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\frac{t_{p_1}(\mathbf{y}_1;\boldsymbol{\mu}_1,\frac{\nu}{\nu+1}\boldsymbol{\Gamma}_{11},\nu+1)}{ST_{p_1}(\mathbf{y}_1;\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11},\tilde{\boldsymbol{\lambda}}_1,\nu)}\frac{P_{p_2}(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2;\boldsymbol{\mu}_{2.1},\frac{\nu_{2.1}}{\nu_{2.1}+1}\tilde{\boldsymbol{\Gamma}}_{22.1},\nu_{2.1}+1)}{\mathcal{P}_{p_2}(\boldsymbol{\alpha}_2,\boldsymbol{\beta}_2;\boldsymbol{\mu}_{2.1},\tilde{\boldsymbol{\Sigma}}_{22.1},\boldsymbol{\lambda}_{2.1},\tau_{2.1},\nu_{2.1})}\\
&\quad\times g_1(\mathbf{Y}_1)\mathbb{E}[g_2(\mathbf{W}_1^*)]
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}_{UT\mathbf{Y}}[U\, g(\mathbf{Y}) \mid \boldsymbol{\alpha} \leqslant \mathbf{Y} \leqslant \boldsymbol{\beta}] \\
&\quad= \int_{\alpha_2}^{\beta_2} \frac{ST_p(\mathbf{y}; \boldsymbol{\mu}, \frac{\nu}{\nu+2}\boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu+2)}{ST_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)} \frac{EST_{p_2}(\mathbf{y}_2; \boldsymbol{\mu}_{2.1}, \tilde{\boldsymbol{\Sigma}}_{22.1}, \boldsymbol{\lambda}_{2.1}, \tau_{2.1}, \nu_{2.1})}{\mathcal{P}_{p_2}(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2; \boldsymbol{\mu}_{2.1}, \tilde{\boldsymbol{\Sigma}}_{22.1}, \boldsymbol{\lambda}_{2.1}, \tau_{2.1}, \nu_{2.1})} g(\mathbf{y}) \mathrm{d}\mathbf{y}_2 \\
&\quad= \frac{ST_{p_1}(\mathbf{y}_1; \boldsymbol{\mu}_1, \frac{\nu}{\nu+2}\boldsymbol{\Sigma}_{11}, \tilde{\boldsymbol{\lambda}}_1, \nu+2)}{ST_{p_1}(\mathbf{y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \tilde{\boldsymbol{\lambda}}_1, \nu)} g_1(\mathbf{Y}_1) \\
&\qquad\times \int_{\alpha_2}^{\beta_2} \frac{EST_{p_2}(\mathbf{y}_2; \boldsymbol{\mu}_{2.1}, \frac{\nu_{2.1}}{\nu_{2.1}+2}\tilde{\boldsymbol{\Sigma}}_{22.1}, \boldsymbol{\lambda}_{2.1}, \sqrt{\frac{\nu_{2.1}+2}{\nu_{2.1}}}\tau_{2.1}, \nu_{2.1}+2)}{\mathcal{P}_{p_2}(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2; \boldsymbol{\mu}_{2.1}, \tilde{\boldsymbol{\Sigma}}_{22.1}, \boldsymbol{\lambda}_{2.1}, \tau_{2.1}, \nu_{2.1})} g_2(\mathbf{y}_2) \mathrm{d}\mathbf{y}_2 \\
&\quad= \frac{ST_{p_1}(\mathbf{y}_1; \boldsymbol{\mu}_1, \frac{\nu}{\nu+2}\boldsymbol{\Sigma}_{11}, \tilde{\boldsymbol{\lambda}}_1, \nu+2)}{ST_{p_1}(\mathbf{y}_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \tilde{\boldsymbol{\lambda}}_1, \nu)} \frac{\mathcal{P}_{p_2}(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2; \boldsymbol{\mu}_{2.1}, \frac{\nu_{2.1}}{\nu_{2.1}+2}\tilde{\boldsymbol{\Sigma}}_{22.1}, \boldsymbol{\lambda}_{2.1}, \sqrt{\frac{\nu_{2.1}+2}{\nu_{2.1}}}\tau_{2.1}, \nu_{2.1}+2)}{\mathcal{P}_{p_2}(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2; \boldsymbol{\mu}_{2.1}, \tilde{\boldsymbol{\Sigma}}_{22.1}, \boldsymbol{\lambda}_{2.1}, \tau_{2.1}, \nu_{2.1})} \\
&\qquad\times g_1(\mathbf{Y}_1)\mathbb{E}[g_2(\mathbf{W}_2^*)],
\end{aligned}
$$

where $g(\mathbf{Y}) = g_1(\mathbf{Y}_1)g_2(\mathbf{Y}_2)$, $\mathbf{W}_1^* \sim Tt_{p_2}(\boldsymbol{\mu}_{2.1}, \frac{\nu_{2.1}}{\nu_{2.1}+1}\tilde{\boldsymbol{\Gamma}}_{22.1}, \nu_{2.1}+1; (\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2))$, and $\mathbf{W}_2^* \sim \mathrm{TEST}_{p_2}(\boldsymbol{\mu}_{2.1}, \frac{\nu_{2.1}}{\nu_{2.1}+2}\tilde{\boldsymbol{\Sigma}}_{22.1}, \boldsymbol{\lambda}_{2.1}, \sqrt{(\nu_{2.1}+2)/\nu_{2.1}}\,\tau_{2.1}, \nu_{2.1}+2; (\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2))$ □

D.2 Additional results from Section 5.4

D.2.1 Simulation Study I

This simulation study aimed to provide evidence about the consistency of the ML estimates obtained through the proposed ST-CR model considering left and right-censored datasets. It was also considered no-censored data for comparison. Figure 46 displays the boxplot of the estimates by sample size and type of censoring. The red line denotes the true parameter value. In general, the median of the estimates is close to the true value for all scenarios. However, for right-censored datasets, the median underestimates σ_{11} and σ_{22}, and overestimates σ_{12}. Furthermore, the interquartile range decreases as the sample size increases, suggesting the consistency of the estimates.

D.2.2 Simulation Study II

To assess the effect of the skewness parameter $\boldsymbol{\lambda}$ in the estimation process for left-censored datasets, we simulated 500 MC samples of size $n = 300$ from the bivariate ST-CR model with the same parameter values described in Section 4, and the following four scenarios for $\boldsymbol{\lambda}$ and the detection limit $\mathbf{v}_{2i}^c$:

i) $\boldsymbol{\lambda} = (2, -3)^\top$, $\mathbf{v}_2^c = (-1.426, -0.110)^\top$;

ii) $\boldsymbol{\lambda} = (2, 3)^\top$, $\mathbf{v}_2^c = (-1.553, -0.160)^\top$;

iii) $\boldsymbol{\lambda} = (-2, 3)^\top$, $\mathbf{v}_2^c = (-1.223, 0.037)^\top$; and

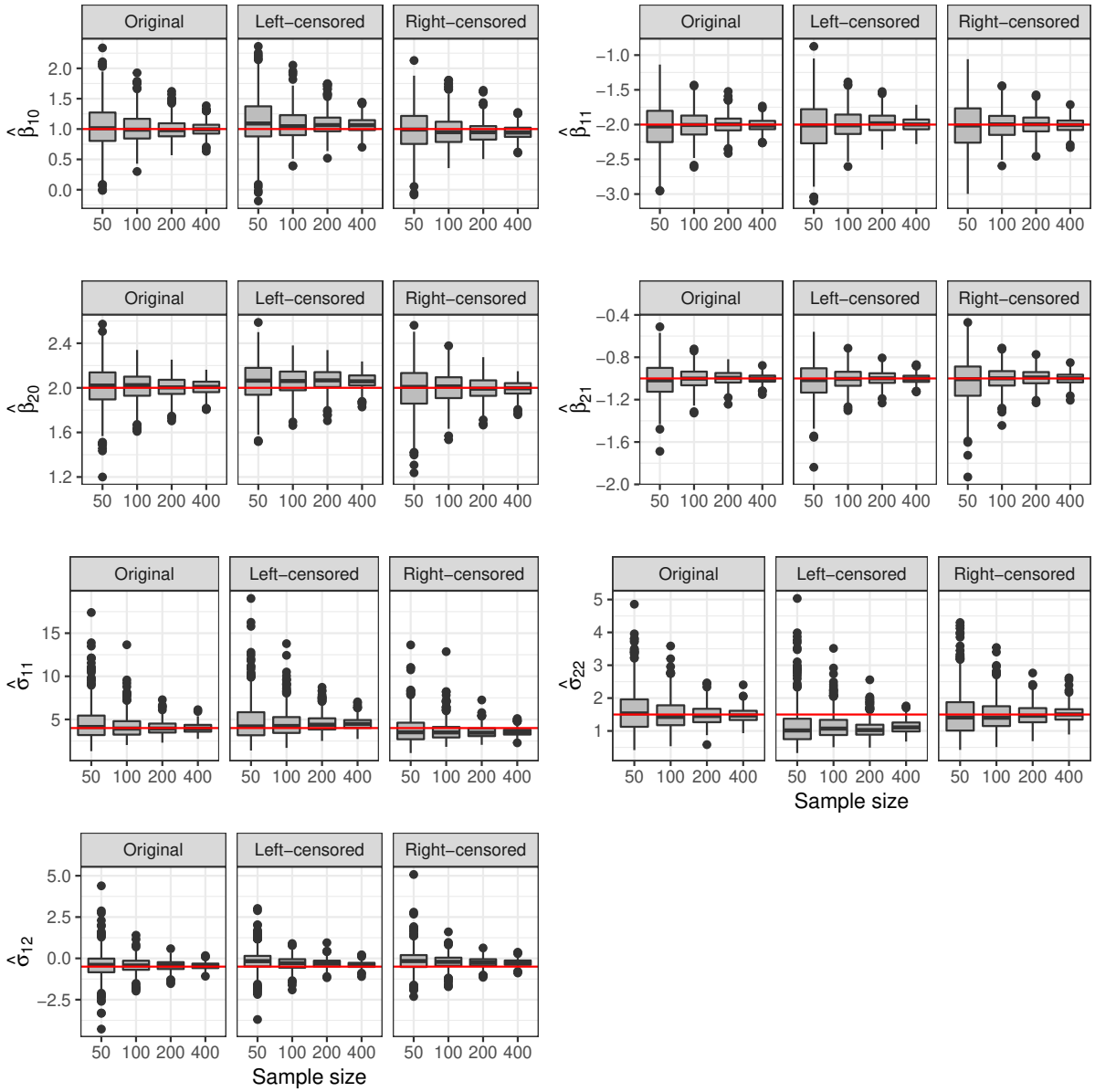
iv) $\boldsymbol{\lambda} = (-2, -3)^\top$, $\mathbf{v}_2^c = (-1.455, -0.300)^\top$.

The values of $\mathbf{v}_2^c$ assure almost 70% of the desirable censoring proportion for the first component of the response vector and 30% for the second component. We fitted the ST-CR model for each case. The summary statistics are shown in Table 25, where MC-AV denotes the mean of the 500 MC estimates, and IM-SE represents the average of the 500 standard errors approximated from the empirical information matrix. This table shows that the mean (MC-AV) of the estimates for the regression coefficients are close to the true parameter value independent of the skewness parameter. In general, the results are similar to those obtained in simulation II.

Table 25 – Simulation II. Summary statistics based on 500 MC samples of size 300 generated from the bivariate skew-t distribution with skewness parameter $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^\top$ and average 15% of left-censored observations. MC-AV denotes the mean of the estimates, and IM-SE refers to the average number of standard errors.

Par.	$\lambda_1 = 2,\ \lambda_2 = -3$		$\lambda_1 = 2,\ \lambda_2 = 3$		$\lambda_1 = -2,\ \lambda_2 = 3$		$\lambda_1 = -2,\ \lambda_2 = -3$	
	MC-AV	IM-SE	MC-AV	IM-SE	MC-AV	IM-SE	MC-AV	IM-SE
β_{10}	1.110	0.154	1.253	0.198	1.063	0.138	1.127	0.159
β_{11}	-1.981	0.122	-1.952	0.116	-1.988	0.131	-2.014	0.133
β_{20}	2.070	0.074	2.045	0.107	1.985	0.073	1.986	0.084
β_{21}	-1.006	0.059	-0.978	0.062	-1.011	0.061	-1.008	0.063
σ_{11}	4.387	0.776	3.549	0.609	3.780	0.585	3.620	0.481
σ_{12}	-0.296	0.199	-0.466	0.175	-0.340	0.246	-0.927	0.277
σ_{22}	1.024	0.158	1.060	0.171	1.586	0.247	1.664	0.298
λ_1	1.993	0.567	3.140	0.915	-1.644	0.545	-0.741	0.499
λ_2	-1.514	0.459	3.195	0.874	3.094	0.811	-2.497	0.647
ν	3.987		2.630		4.697		4.602	

Figure 46 – Simulation I. Boxplot of the estimates obtained under the ST-CR model considering different sample sizes and types of censoring based on 500 Monte Carlo samples. The red line denotes the true parameter value.



APPENDIX E – CHAPTER 6

E.1 Marginal distribution for the LME-ST model

Following Arellano-Valle & Genton (2010b), Proposition 5, if $\mathbf{Y} \sim ST_p(\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\lambda}, \nu)$, then $\mathbf{AY} + \mathbf{b} \sim ST_r(\boldsymbol{\mu}_A, \boldsymbol{\Omega}_A, \boldsymbol{\lambda}_A, \nu)$ for any $r \times p$ matrix $\mathbf{A}$ of rank $r \leqslant p$ and $r \times 1$ vector $\mathbf{b}$, where $\boldsymbol{\mu}_A = \mathbf{A}\boldsymbol{\mu} + \mathbf{b}$, $\boldsymbol{\Omega}_A = \mathbf{A}\boldsymbol{\Omega}\mathbf{A}^\top$, and $\boldsymbol{\lambda}_A = \gamma_A^{-1}\boldsymbol{\Omega}_A^{-1/2}\mathbf{A}\boldsymbol{\Omega}^{1/2}\boldsymbol{\lambda}$, with $\gamma_A = 1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda} - \boldsymbol{\lambda}^\top\boldsymbol{\Omega}^{1/2}\mathbf{A}^\top\boldsymbol{\Omega}_A^{-1}\mathbf{A}\boldsymbol{\Omega}^{1/2}\boldsymbol{\lambda}$.

For the ith subject in the LME-ST model, the joint distribution of $(\mathbf{b}_i, \boldsymbol{\xi}_i)^\top$ is given by

$$\begin{pmatrix} \mathbf{b}_i \\ \boldsymbol{\xi}_i \end{pmatrix} \sim ST_{q+n_i}\left(\begin{pmatrix} \kappa\boldsymbol{\Delta} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_i \end{pmatrix}, \begin{pmatrix} \boldsymbol{\lambda} \\ \mathbf{0} \end{pmatrix}, \nu\right), \quad i \in \{1, \ldots, n\}.$$

Therefore, to compute the marginal distribution of $\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\xi}_i$, we set the $n_i \times 1$ vector $\mathbf{b} = \mathbf{X}_i\boldsymbol{\beta}$, and the $n_i \times (q + n_i)$ matrix $\mathbf{A} = [\mathbf{Z}_i \quad \mathbf{I}_{n_i}]$, with $\mathbf{I}_{n_i}$ denoting the diagonal matrix of dimensions $n_i \times n_i$. Then $\mathbf{Y}_i \sim ST_{n_i}(\boldsymbol{\mu}_A, \boldsymbol{\Omega}_A, \boldsymbol{\lambda}_A, \nu)$, with

$$\begin{aligned}
\boldsymbol{\mu}_A &= \mathbf{X}_i\boldsymbol{\beta} + \begin{pmatrix} \mathbf{Z}_i & \mathbf{I}_{n_i} \end{pmatrix}\begin{pmatrix} \kappa\boldsymbol{\Delta} \\ \mathbf{0} \end{pmatrix} = \mathbf{X}_i\boldsymbol{\beta} + \kappa\mathbf{Z}_i\boldsymbol{\Delta}, \\
\boldsymbol{\Omega}_A &= \begin{pmatrix} \mathbf{Z}_i & \mathbf{I}_{n_i} \end{pmatrix}\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_i \end{pmatrix}\begin{pmatrix} \mathbf{Z}_i^\top \\ \mathbf{I}_{n_i} \end{pmatrix} = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top + \boldsymbol{\Omega}_i, \\
\boldsymbol{\lambda}_A &= \gamma_A^{-1}\boldsymbol{\Omega}_A^{-1/2}\begin{pmatrix} \mathbf{Z}_i & \mathbf{I}_{n_i} \end{pmatrix}\begin{pmatrix} \mathbf{D}^{1/2} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_i^{1/2} \end{pmatrix}\begin{pmatrix} \boldsymbol{\lambda} \\ \mathbf{0} \end{pmatrix} = \gamma_A^{-1}\boldsymbol{\Omega}_A^{-1/2}\mathbf{Z}_i\mathbf{D}^{1/2}\boldsymbol{\lambda}, \text{ with} \\
\gamma_A &= 1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda} - \boldsymbol{\lambda}^\top\mathbf{D}^{1/2}\mathbf{Z}_i^\top\boldsymbol{\Omega}_A^{-1}\mathbf{Z}_i\mathbf{D}^{1/2}\boldsymbol{\lambda} \\
&= 1 + \boldsymbol{\lambda}^\top\mathbf{D}^{-1/2}\left(\mathbf{D} - \mathbf{D}\mathbf{Z}_i^\top(\boldsymbol{\Omega}_i + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top)^{-1}\mathbf{Z}_i\mathbf{D}\right)\mathbf{D}^{-1/2}\boldsymbol{\lambda} \\
&= 1 + \boldsymbol{\lambda}^\top\mathbf{D}^{-1/2}(\mathbf{D}^{-1} + \mathbf{Z}_i^\top\boldsymbol{\Omega}_i^{-1}\mathbf{Z}_i)^{-1}\mathbf{D}^{-1/2}\boldsymbol{\lambda}.
\end{aligned}$$

E.2 Details for the expectations needed in the EM algorithm of the LMEC-ST model

This section derives the conditional expectations required to perform the E-step in the linear mixed-effects censored skew-t (LMEC-ST) model.

Note that:

$$\begin{aligned}
\mathbf{b}_i \mid \mathbf{y}_i, t_i, u_i, \boldsymbol{\theta} &\sim N_p(\mathbf{s}_i t_i + \mathbf{r}_i, u_i^{-1}\mathbf{B}_i) \\
t_i|\mathbf{y}_i, u_i, \boldsymbol{\theta} &\sim TN(\kappa + \mu_i, u_i^{-1}M_i^2; (\kappa, \infty)) \\
\mathbf{Y}_i \mid \boldsymbol{\theta} &\sim ST_{n_i}(\mathbf{X}_i\boldsymbol{\beta} + \kappa\mathbf{Z}_i\boldsymbol{\Delta}, \boldsymbol{\Sigma}_i, \bar{\boldsymbol{\lambda}}_i, \nu),
\end{aligned}$$

where $\mathbf{s}_i = (\mathbf{I}_q - \mathbf{B}_i\mathbf{Z}_i^\top\boldsymbol{\Omega}_i^{-1}\mathbf{Z}_i)\boldsymbol{\Delta}$, $\mathbf{r}_i = \mathbf{B}_i\mathbf{Z}_i^\top\boldsymbol{\Omega}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$, $\mathbf{B}_i = (\boldsymbol{\Gamma}^{-1} + \mathbf{Z}_i^\top\boldsymbol{\Omega}_i^{-1}\mathbf{Z}_i)^{-1}$, $\mu_i = M_i^2\boldsymbol{\Delta}^\top\mathbf{Z}_i^\top\boldsymbol{\Psi}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \kappa\mathbf{Z}_i\boldsymbol{\Delta})$, and $M_i = (1 + \boldsymbol{\Delta}^\top\mathbf{Z}_i^\top\boldsymbol{\Psi}_i^{-1}\mathbf{Z}_i\boldsymbol{\Delta})^{-1/2}$, with $\boldsymbol{\Psi}_i = \boldsymbol{\Omega}_i + \mathbf{Z}_i\boldsymbol{\Gamma}\mathbf{Z}_i^\top$.

To compute the expected values needed in the E-step, first note that for any multiplicative separable measurable function of U_i, T_i, $\mathbf{b}_i$, and $\mathbf{Y}_i$, such that $g(U_i, T_i, \mathbf{b}_i, \mathbf{Y}_i) = g_1(\mathbf{Y}_i)g_2(U_i)g_3(T_i)g_4(\mathbf{b}_i)$, we have that

$$\begin{aligned}\mathbb{E}[g(U_i, T_i, \mathbf{b}_i, \mathbf{Y}_i)|\mathbf{V}_i, \mathbf{C}_i] &= \mathbb{E}[g_1(\mathbf{Y}_i)\,\mathbb{E}[g_2(U_i)g_3(T_i)g_4(\mathbf{b}_i) \mid \mathbf{Y}_i] \mid \mathbf{V}_i, \mathbf{C}_i] \\ &= \mathbb{E}[g_1(\mathbf{Y}_i)\,\mathbb{E}[g_2(U_i)\,\mathbb{E}[g_3(T_i)\,\mathbb{E}[g_4(\mathbf{b}_i) \mid U_i, T_i, \mathbf{Y}_i] \mid U_i, \mathbf{Y}_i] \mid \mathbf{Y}_i] \mid \mathbf{V}_i, \mathbf{C}_i].\end{aligned}$$

Hence,

$$\begin{aligned}
&\widehat{u}_i^{(k)} = \mathbb{E}[U_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \mathbb{E}[\mathbb{E}[U_i|\mathbf{Y}_i] \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}], \\
&\widehat{\mathbf{uy}_i^r}^{(k)} = \mathbb{E}[U_i\mathbf{Y}_i^r|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \mathbb{E}[\mathbf{Y}_i^r\,\mathbb{E}[U_i|\mathbf{Y}_i] \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}], \\
&\widehat{ut_i}^{(k)} = \mathbb{E}[U_iT_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \widehat{\kappa}^{(k)}\widehat{u}_i^{(k)} + \widehat{M}_i^{2(k)}\widehat{\boldsymbol{\Delta}}^{(k)\top}\mathbf{Z}_i^\top\widehat{\boldsymbol{\Psi}}_i^{-1(k)}\left(\widehat{\mathbf{uy}_i}^{(k)} - \widehat{u}_i^{(k)}\widehat{\boldsymbol{\mu}}_i^{(k)}\right) \\
&\qquad + \widehat{M}_i^{(k)}\widehat{\phi\mathbf{y}_i^0}^{(k)}, \\
&\widehat{ut_i^2}^{(k)} = \mathbb{E}[U_iT_i^2|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \widehat{\kappa}^{2(k)}\widehat{u}_i^{(k)} + 2\widehat{\kappa}^{(k)}\widehat{M}_i^{2(k)}\widehat{\boldsymbol{\Delta}}^{(k)\top}\mathbf{Z}_i^\top\widehat{\boldsymbol{\Psi}}_i^{-1(k)}(\widehat{\mathbf{uy}_i}^{(k)} - \widehat{u}_i^{(k)}\widehat{\boldsymbol{\mu}}_i^{(k)}) \\
&\qquad + \widehat{M}_i^{2(k)} + \widehat{M}_i^{3(k)}\widehat{\boldsymbol{\Delta}}^{(k)\top}\mathbf{Z}_i^\top\widehat{\boldsymbol{\Psi}}_i^{-1(k)}\left[\widehat{\phi\mathbf{y}_i^1}^{(k)} - \widehat{\boldsymbol{\mu}}_i^{(k)}\widehat{\phi\mathbf{y}_i^0}^{(k)} + \widehat{M}_i^{(k)}\widehat{\mathbf{F}}_{1i}^{(k)}\widehat{\boldsymbol{\Psi}}_i^{-1(k)}\mathbf{Z}_i\widehat{\boldsymbol{\Delta}}^{(k)}\right] \\
&\qquad + 2\widehat{\kappa}^{(k)}\widehat{M}_i^{(k)}\widehat{\phi\mathbf{y}_i^0}^{(k)}, \\
&\widehat{\mathbf{uty}_i}^{(k)} = \mathbb{E}[U_iT_i\mathbf{Y}_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \widehat{\kappa}^{(k)}\widehat{\mathbf{uy}_i}^{(k)} + \widehat{M}_i^{2(k)}(\widehat{\mathbf{uy}_i^2}^{(k)} - \widehat{\mathbf{uy}_i}^{(k)}\widehat{\boldsymbol{\mu}}_i^{(k)\top})\widehat{\boldsymbol{\Psi}}_i^{-1(k)}\mathbf{Z}_i\widehat{\boldsymbol{\Delta}}^{(k)} \\
&\qquad + \widehat{M}_i^{(k)}\widehat{\phi\mathbf{y}_i^1}^{(k)}, \\
&\widehat{\mathbf{utb}_i} = \mathbb{E}[U_iT_i\mathbf{b}_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \widehat{ut_i^2}^{(k)}\widehat{\mathbf{s}}_i^{(k)} + \widehat{\mathbf{B}}_i^{(k)}\mathbf{Z}_i^\top\widehat{\boldsymbol{\Omega}}_i^{-1(k)}(\widehat{\mathbf{uty}_i}^{(k)} - \widehat{ut_i}^{(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}), \\
&\widehat{\mathbf{ub}_i} = \mathbb{E}[U_i\mathbf{b}_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \widehat{ut_i}^{(k)}\widehat{\mathbf{s}}_i^{(k)} + \widehat{\mathbf{B}}_i^{(k)}\mathbf{Z}_i^\top\widehat{\boldsymbol{\Omega}}_i^{-1(k)}\left(\widehat{\mathbf{uy}_i}^{(k)} - \widehat{u}_i^{(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\right), \\
&\widehat{\mathbf{ub}_i^2} = \mathbb{E}[U_i\mathbf{b}_i^2|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \widehat{ut_i^2}^{(k)}\widehat{\mathbf{s}}_i^{(k)}\widehat{\mathbf{s}}_i^{(k)\top} + \widehat{\mathbf{s}}_i^{(k)}(\widehat{\mathbf{uty}_i}^{(k)} - \widehat{ut_i}^{(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)})^\top\widehat{\boldsymbol{\Omega}}_i^{-1(k)}\mathbf{Z}_i\widehat{\mathbf{B}}_i^{(k)} \\
&\qquad + \widehat{\mathbf{B}}_i^{(k)} + \widehat{\mathbf{B}}_i^{(k)}\mathbf{Z}_i^\top\widehat{\boldsymbol{\Omega}}_i^{-1(k)}(\widehat{\mathbf{uty}_i}^{(k)} - \widehat{ut_i}^{(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)})\widehat{\mathbf{s}}_i^{(k)\top} + \widehat{\mathbf{B}}_i^{(k)}\mathbf{Z}_i^\top\widehat{\boldsymbol{\Omega}}_i^{-1(k)}\widehat{\mathbf{F}}_{2i}^{(k)}\widehat{\boldsymbol{\Omega}}_i^{-1(k)}\mathbf{Z}_i\widehat{\mathbf{B}}_i^{(k)}, \\
&\widehat{\mathbf{uby}_i^\top} = \mathbb{E}[U_i\mathbf{b}_i\mathbf{Y}_i^\top|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \widehat{\mathbf{s}}_i^{(k)}\widehat{\mathbf{uty}_i}^{(k)\top} + \widehat{\mathbf{B}}_i^{(k)}\mathbf{Z}_i^\top\widehat{\boldsymbol{\Omega}}_i^{-1(k)}(\widehat{\mathbf{uy}_i^2}^{(k)} - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\widehat{\mathbf{uy}_i}^{(k)\top}),
\end{aligned}$$

such that $\boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta} + c\mathbf{Z}_i\boldsymbol{\Delta}$, $\widehat{\mathbf{F}}_{1i}^{(k)} = \widehat{\mathbf{uy}_i^2}^{(k)} - \widehat{\mathbf{uy}_i}^{(k)}\widehat{\boldsymbol{\mu}}_i^{\top(k)} - \widehat{\boldsymbol{\mu}}_i^{(k)}\widehat{\mathbf{uy}_i}^{\top(k)} + \widehat{u}_i^{(k)}\widehat{\boldsymbol{\mu}}_i^{(k)}\widehat{\boldsymbol{\mu}}_i^{\top(k)}$, $\widehat{\mathbf{F}}_{2i}^{(k)} = \widehat{\mathbf{uy}_i^2}^{(k)} - \widehat{\mathbf{uy}_i}^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\widehat{\mathbf{uy}_i}^{(k)\top} + \widehat{u}_i^{(k)}\mathbf{X}_i\widehat{\boldsymbol{\beta}}^{(k)}\widehat{\boldsymbol{\beta}}^{(k)\top}\mathbf{X}_i^\top$, and

$$\begin{aligned}\widehat{\phi\mathbf{y}_i^r} &= \mathbb{E}[\mathbf{Y}_i^r\phi(\boldsymbol{\theta}, \mathbf{Y}_i)|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] \\ &= \mathbb{E}\left[\mathbf{Y}_i^r U_i^{1/2}\frac{\phi_1\left(U_i^{1/2}\bar{\boldsymbol{\lambda}}_i^\top\boldsymbol{\Sigma}_i^{-1/2}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - c\mathbf{Z}_i\boldsymbol{\Delta})\right)}{\Phi_1\left(U_i^{1/2}\bar{\boldsymbol{\lambda}}_i^\top\boldsymbol{\Sigma}_i^{-1/2}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - c\mathbf{Z}_i\boldsymbol{\Delta})\right)}\middle| \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right]\end{aligned}$$

for $r \in \{0, 1, 2\}$. Subsequently, we have the implementable expressions to the conditional expectations under the following three possible scenarios:

1. If the ith subject has only non-censored components, $\mathbb{E}[\mathbf{Y}_i^r|\mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \mathbf{y}_i^r$; then

$$\widehat{u}_i^{(k)} = \frac{ST_{n_i}(\mathbf{y}_i; \widehat{\boldsymbol{\mu}}_i^{(k)}, \frac{\widehat{\nu}^{(k)}}{\widehat{\nu}^{(k)}+2}\widehat{\boldsymbol{\Sigma}}_i^{(k)}, \widehat{\tilde{\boldsymbol{\lambda}}}_i^{(k)}, \widehat{\nu}^{(k)}+2)}{ST_{n_i}(\mathbf{y}_i; \widehat{\boldsymbol{\mu}}_i^{(k)}, \widehat{\boldsymbol{\Sigma}}_i^{(k)}, \widehat{\tilde{\boldsymbol{\lambda}}}_i^{(k)}, \widehat{\nu}^{(k)})},$$

$$\widehat{\mathbf{uy}_i^r}^{(k)} = \widehat{u}_i^{(k)}\mathbf{y}_i^r,$$

$$\widehat{\phi\mathbf{y}_i^r}^{(k)} = \frac{2}{\sqrt{\pi\widehat{\nu}^{(k)}\widehat{\varrho}^{(k)}}}\frac{\Gamma((\widehat{\nu}^{(k)}+1)/2)}{\Gamma(\widehat{\nu}^{(k)}/2)}\frac{t_{n_i}(\mathbf{y}_i; \widehat{\boldsymbol{\mu}}_i^{(k)}, \frac{\widehat{\nu}^{(k)}}{\widehat{\nu}^{(k)}+1}\widehat{\boldsymbol{\Psi}}_i^{(k)}, \widehat{\nu}^{(k)}+1)}{ST_{n_i}(\mathbf{y}_i; \widehat{\boldsymbol{\mu}}_i^{(k)}, \widehat{\boldsymbol{\Sigma}}_i^{(k)}, \widehat{\tilde{\boldsymbol{\lambda}}}_i^{(k)}, \widehat{\nu}^{(k)})}\mathbf{y}_i^r,$$

where $\widehat{\varrho}^{(k)} = 1 + \widehat{\boldsymbol{\Delta}}^{(k)\top}\mathbf{Z}_i^\top\widehat{\boldsymbol{\Psi}}_i^{-1(k)}\mathbf{Z}_i\widehat{\boldsymbol{\Delta}}^{(k)}$, $\mathbf{y}_i^0 = 1$, $\mathbf{y}_i^1 = \mathbf{y}_i$, and $\mathbf{y}_i^2 = \mathbf{y}_i\mathbf{y}_i^\top$.

2. If the ith subject has only censored components, we have

$$\widehat{u}_i^{(k)} = \frac{\mathcal{P}_{n_i}(\mathbf{r}_{1i}, \mathbf{r}_{2i}; \widehat{\boldsymbol{\mu}}_i^{(k)}, \frac{\widehat{\nu}^{(k)}}{\widehat{\nu}^{(k)}+2}\widehat{\boldsymbol{\Sigma}}_i^{(k)}, \widehat{\tilde{\boldsymbol{\lambda}}}_i^{(k)}, \widehat{\nu}^{(k)}+2)}{\mathcal{P}_{n_i}(\mathbf{r}_{1i}, \mathbf{r}_{2i}; \widehat{\boldsymbol{\mu}}_i^{(k)}, \widehat{\boldsymbol{\Sigma}}_i^{(k)}, \widehat{\tilde{\boldsymbol{\lambda}}}_i^{(k)}, \widehat{\nu}^{(k)})},$$

$$\widehat{\mathbf{uy}_i^r}^{(k)} = \widehat{u_i}^{(k)}\widehat{\mathbf{w}_{2i}^r}^{(k)},$$

$$\widehat{\phi\mathbf{y}_i^r}^{(k)} = \frac{2}{\sqrt{\pi\widehat{\nu}^{(k)}\widehat{\varrho}^{(k)}}}\frac{\Gamma((\widehat{\nu}^{(k)}+1)/2)}{\Gamma(\widehat{\nu}^{(k)}/2)}\frac{P_{n_i}(\mathbf{r}_{1i}, \mathbf{r}_{2i}; \widehat{\boldsymbol{\mu}}_i^{(k)}, \frac{\widehat{\nu}^{(k)}}{\widehat{\nu}^{(k)}+1}\widehat{\boldsymbol{\Psi}}_i^{(k)}, \widehat{\nu}^{(k)}+1)}{\mathcal{P}_{n_i}(\mathbf{r}_{1i}, \mathbf{r}_{2i}; \widehat{\boldsymbol{\mu}}_i^{(k)}, \widehat{\boldsymbol{\Sigma}}_i^{(k)}, \widehat{\tilde{\boldsymbol{\lambda}}}_i^{(k)}, \widehat{\nu}^{(k)})}\widehat{\mathbf{w}}_{1i}^{r(k)},$$

where $\widehat{\mathbf{w}}_{si}^{(k)} = \mathbb{E}[\mathbf{W}_{si} \mid \widehat{\boldsymbol{\theta}}^{(k)}]$ and $\widehat{\mathbf{w}}_{si}^{2(k)} = \mathbb{E}[\mathbf{W}_{si}\mathbf{W}_{si}^\top \mid \widehat{\boldsymbol{\theta}}^{(k)}]$ for $s \in \{1, 2\}$, with

$$\mathbf{W}_{1i} \sim Tt_{n_i}(\boldsymbol{\mu}_i, \tfrac{\nu\boldsymbol{\Psi}_i}{\nu+1}, \nu+1; (\mathbf{r}_{1i}, \mathbf{r}_{2i})) \text{ and } \mathbf{W}_{2i} \sim \mathrm{TST}_{n_i}(\boldsymbol{\mu}_i, \tfrac{\nu\boldsymbol{\Sigma}_i}{\nu+2}, \bar{\boldsymbol{\lambda}}_i, \nu+2; (\mathbf{r}_{1i}, \mathbf{r}_{2i})).$$

3. If the ith subject has both censored and uncensored components and given that the following processes $(\mathbf{Y}_i \,|\, \mathbf{V}_i, \mathbf{C}_i)$ and $(\mathbf{Y}_i \,|\, \mathbf{V}_i, \mathbf{C}_i, \mathbf{Y}_i^o)$ are equivalent, we have

$$\widehat{u}_i^{(k)} = \mathbb{E}[U_i \mid \mathbf{y}_i^o, \mathbf{V}_i, \mathbf{C}_i, \widehat{\boldsymbol{\theta}}^{(k)}] = \frac{ST_{n_i^o}(\mathbf{y}_i^o; \widehat{\boldsymbol{\mu}}_i^{o(k)}, \frac{\widehat{\nu}^{(k)}}{\widehat{\nu}^{(k)}+2}\widehat{\boldsymbol{\Sigma}}_i^{oo(k)}, \tilde{\boldsymbol{\lambda}}_i^{o(k)}, \widehat{\nu}^{(k)}+2)}{ST_{n_i^o}(\mathbf{y}_i^o; \widehat{\boldsymbol{\mu}}_i^{o(k)}, \widehat{\boldsymbol{\Sigma}}_i^{oo(k)}, \tilde{\boldsymbol{\lambda}}_i^{o(k)}, \widehat{\nu}^{(k)})}$$
$$\times \frac{\mathcal{P}_{n_i^c}(\mathbf{r}_{1i}^c, \mathbf{r}_{2i}^c; \widehat{\boldsymbol{\mu}}_i^{co(k)}, \frac{\widehat{\nu}_i^{co(k)}}{\widehat{\nu}_i^{co(k)}+2}\tilde{\boldsymbol{\Sigma}}_i^{cc.o(k)}, \widehat{\boldsymbol{\lambda}}_i^{co(k)}, \sqrt{\frac{\widehat{\nu}_i^{co(k)}+2}{\widehat{\nu}_i^{co(k)}}}\,\widehat{\tau}_i^{co(k)}, \widehat{\nu}_i^{co(k)}+2)}{\mathcal{P}_{n_i^c}(\mathbf{r}_{1i}^c, \mathbf{r}_{2i}^c; \widehat{\boldsymbol{\mu}}_i^{co(k)}, \tilde{\boldsymbol{\Sigma}}_i^{cc.o(k)}, \widehat{\boldsymbol{\lambda}}_i^{co(k)}, \widehat{\tau}_i^{co(k)}, \widehat{\nu}_i^{co(k)})},$$

$$\widehat{u\mathbf{y}_i^r}^{(k)} = \widehat{u}_i^{(k)}\widehat{\mathbf{w}}_{2i}^{r(k)},$$

$$\widehat{\phi\mathbf{y}_i^r}^{(k)} = \frac{2}{\sqrt{\pi\widehat{\nu}^{(k)}\widehat{\varrho}^{(k)}}}\frac{\Gamma(\frac{\widehat{\nu}^{(k)}+1}{2})}{\Gamma(\frac{\widehat{\nu}^{(k)}}{2})}\frac{P_{n_i^c}(\mathbf{r}_{1i}^c, \mathbf{r}_{2i}^c; \widehat{\boldsymbol{\mu}}_i^{co(k)}, \frac{\widehat{\nu}_i^{co(k)}}{\widehat{\nu}_i^{co(k)}+1}\tilde{\boldsymbol{\Psi}}_i^{cc.o(k)}, \widehat{\nu}_i^{co(k)}+1)}{\mathcal{P}_{n_i^c}(\mathbf{r}_{1i}^c, \mathbf{r}_{2i}^c; \widehat{\boldsymbol{\mu}}_i^{co(k)}, \tilde{\boldsymbol{\Sigma}}_i^{cc.o(k)}, \widehat{\boldsymbol{\lambda}}_i^{co(k)}, \widehat{\tau}_i^{co(k)}, \widehat{\nu}_i^{co(k)})}$$
$$\times \frac{t_{n_i^o}(\mathbf{y}_i^o; \widehat{\boldsymbol{\mu}}_i^{o(k)}, \frac{\widehat{\nu}^{(k)}}{\widehat{\nu}^{(k)}+1}\widehat{\boldsymbol{\Psi}}_i^{oo(k)}, \widehat{\nu}^{(k)}+1)}{ST_{n_i^o}(\mathbf{y}_i^o; \widehat{\boldsymbol{\mu}}_i^{o(k)}, \widehat{\boldsymbol{\Sigma}}_i^{oo(k)}, \tilde{\boldsymbol{\lambda}}_i^{o(k)}, \widehat{\nu}^{(k)})}\widehat{\mathbf{w}}_{1i}^{r(k)},$$

where $\widehat{\mathbf{w}}_{si}^{(k)} = \text{vec}(\mathbf{y}_i^o, \widehat{\mathbf{w}}_{si}^{c(k)})$, $\widehat{\mathbf{w}}_{si}^{2(k)} = \begin{pmatrix} \mathbf{y}_i^o \mathbf{y}_i^{o\top} & \mathbf{y}_i^o \widehat{\mathbf{w}}_{si}^{c(k)\top} \\ \widehat{\mathbf{w}}_{si}^{c(k)} \mathbf{y}_i^{o\top} & \widehat{\mathbf{w}_{si}^c}^{2(k)} \end{pmatrix}$, $\widehat{\mathbf{w}}_{si}^{c(k)} = \mathbb{E}[\mathbf{W}_{si}^c \mid \widehat{\boldsymbol{\theta}}^{(k)}]$, and $\widehat{\mathbf{w}_{si}^c}^{2(k)} = \mathbb{E}[\mathbf{W}_{si}^c \mathbf{W}_{si}^{c\top} \mid \widehat{\boldsymbol{\theta}}^{(k)}]$ for $s \in \{1, 2\}$, such that

$$\mathbf{W}_{1i}^c \sim Tt_{n_i^c}\left(\boldsymbol{\mu}_i^{co(k)}, \tfrac{\nu_i^{co}}{\nu_i^{co}+1}\tilde{\boldsymbol{\Psi}}_i^{cc.o}, \nu_i^{co}+1; (\mathbf{r}_{1i}^c, \mathbf{r}_{2i}^c)\right),$$

$$\mathbf{W}_{2i}^c \sim \text{TEST}_{n_i^c}\left(\boldsymbol{\mu}_i^{co}, \tfrac{\nu_i^{co}}{\nu_i^{co}+2}\tilde{\boldsymbol{\Sigma}}_i^{cc.o}, \boldsymbol{\lambda}_i^{co}, \sqrt{\frac{\nu_i^{co}+2}{\nu_i^{co}}}\tau_i^{co}, \nu_i^{co}+2; (\mathbf{r}_{1i}^c, \mathbf{r}_{2i}^c)\right),$$

with $\boldsymbol{\Psi}_i$ being partitioned like $\boldsymbol{\Sigma}_i$, $\nu_i^{co} = \nu + n_i^o$, $\tilde{\boldsymbol{\Psi}}_i^{cc.o} = (\boldsymbol{\Psi}_i^{cc} - \boldsymbol{\Psi}_i^{co}\boldsymbol{\Psi}_i^{oo-1}\boldsymbol{\Psi}_i^{oc})/\upsilon^2(\mathbf{y}_i^o)$, with $\upsilon^2(\mathbf{y}_i^o) = (\nu + n_i^o)/(\nu + \delta(\mathbf{y}_i^o; \boldsymbol{\mu}_i^o, \boldsymbol{\Psi}^{oo}))$, $\tau_i^{co} = \nu(\mathbf{y}_i^o)(\tilde{\boldsymbol{\varphi}}_i^{o\top}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o))$, and $\nu^2(\mathbf{y}_i^o) = (\nu + n_i^o)/(\nu + \delta(\mathbf{y}_i^o; \boldsymbol{\mu}_i^o, \boldsymbol{\Sigma}_i^{oo}))$.