Some extensions in measurement error models

Lorena Yanet Cáceres Tomaya

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA
UFSCar–USP

Lorena Yanet Cáceres Tomaya

Some extensions in measurement error models

Doctoral dissertation submitted to the Department of Statistics – DEs-UFSCar and Institute of Mathematics and Computer Sciences – ICMC-USP – in partial fulfillment of the requirements for the degree of the doctorate in Statistic of the Statistics Interagency Graduate Program UFSCar–USP.

Advisor: Prof. Dr. Mário de Castro Andrade Filho

São Carlos
January 2019

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA
UFSCar–USP

Lorena Yanet Cáceres Tomaya

Algumas extensões em modelos com erros de medição

Tese apresentada ao Departamento de Estatística – DEs-UFSCar e ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Doutora em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística UFSCar–USP.

Orientador: Prof. Dr. Mário de Castro Andrade Filho

São Carlos
Janeiro de 2019

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Tese de Doutorado da candidata Lorena Yanet Cáceres Tomaya, realizada em 14/12/2018:

Prof. Dr. Mário de Castro Andrade Filho
USP

Prof. Dr. Josemar Rodrigues
USP

Prof. Dr. Filidor Edilfonso Vilca Labra
UNICAMP

Prof. Dr. Edwin Moises Marcos Ortega
ESALQ/USP

Prof. Dr. Danilo Lourenço Lopes
UFSCar

To my parents

ACKNOWLEDGEMENTS

Above all I thank God for letting me through this journey of life and for giving me strength and perseverance in difficult times. I will always remain grateful to God.

I would like to express my deepest gratitude to my advisor, Professor Mário de Castro, for his great patience and kindness together with endless enthusiasm for research. With his constant encouragement and insight into the subject matter this work would not have been completed.

I would also like to thank to all the Professors and administrative staff of the Department of Statistics at UFSCar and USP, for helping me during these years of my study academically and intellectually.

A special word of thanks to the committee members for the valuable suggestions.

My thanks also to Miguel and my daughters Sofia and Clarice, for their unlimited love.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

ABSTRACT

TOMAYA, L. Y. C. **Some extensions in measurement error models**. 2019. 89 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística UFSCar–USP) – Departamento de Estatística - DEs-UFSCar e Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

In this dissertation, we approach three different contributions in measurement error model (MEM). Initially, we carry out maximum penalized likelihood inference in MEM's under the normality assumption. The methodology is based on the method proposed by Firth (1993), which can be used to improve some asymptotic properties of the maximum likelihood estimators. In the second contribution, we develop two new estimation methods based on generalized fiducial inference for the precision parameters and the variability product under the Grubbs model considering the two-instrument case. One method is based on a fiducial generalized pivotal quantity and the other one is built on the method of the generalized fiducial distribution. Comparisons with two existing approaches are reported. Finally, we propose to study inference in a heteroscedastic MEM with known error variances. Instead of the normal distribution for the random components, we develop a model that assumes a skew-*t* distribution for the true covariate and a centered Student's *t* distribution for the error terms. The proposed model enables to accommodate skewness and heavy-tailedness in the data, while the degrees of freedom of the distributions can be different. We use the maximum likelihood method to estimate the model parameters and compute them via an EM-type algorithm. All proposed methodologies are assessed numerically through simulation studies and illustrated with real datasets extracted from the literature.

Keywords: Errors-in-variables model, Fiducial inference, Heteroscedastic errors, Penalized likelihood, Skew-*t* distribution.

RESUMO

TOMAYA, L. Y. C. **Algumas extensões em modelos com erros de medição**. 2019. 89 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística UFSCar–USP) – Departamento de Estatística - DEs-UFSCar e Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

Neste trabalho abordamos três contribuições diferentes em modelos com erros de medição (MEM). Inicialmente estudamos inferência pelo método de máxima verossimilhança penalizada em MEM sob a suposição de normalidade. A metodologia baseia-se no método proposto por Firth (1993), o qual pode ser usado para melhorar algumas propriedades assintóticas de os estimadores de máxima verossimilhança. Em seguida, propomos construir dois novos métodos de estimação baseados na inferência fiducial generalizada para os parâmetros de precisão e a variabilidade produto no modelo de Grubbs para o caso de dois instrumentos. O primeiro método é baseado em uma quantidade pivotal generalizada fiducial e o outro é baseado no método da distribuição fiducial generalizada. Comparações com duas abordagens existentes são reportadas. Finalmente, propomos estudar inferência em um MEM heterocedástico em que as variâncias dos erros são consideradas conhecidas. Nós desenvolvemos um modelo que assume uma distribuição t-assimétrica para a covariável verdadeira e uma distribuição t de Student centrada para os termos dos erros. O modelo proposto permite acomodar assimetria e cauda pesada nos dados, enquanto os graus de liberdade das distribuições podem ser diferentes. Usamos o método de máxima verossimilhança para estimar os parâmetros do modelo e calculá-los através de um algoritmo tipo EM. Todas as metodologias propostas são avaliadas numericamente em estudos de simulação e são ilustradas com conjuntos de dados reais extraídos da literatura

Palavras-chave: Modelo com erros nas variáveis, Inferência fiducial, Erros heteroscedásticos,Verossimilhança penalizada, Distribuição t-assimétrica.

CONTENTS

CHAPTER 1

INTRODUCTION

In several areas of knowledge, regression models are statistical devices often used to investigate the relationship between one variable of interest (response) and a set of covariates. In practical situations, some covariates associated with the response variable can be error-prone. For these cases, measurement error models, also called errors-in-variables model, are considered. There are a number of specialized books on the measurement error, such as Fuller (1987), Cheng and Ness (1999), Carroll *et al.* (2006) and Buonaccorsi (2010). In the last decades, measurement error models have been receiving considerable attention by many researchers with works being published in several fields of study such as Medicine, Epidemiology, Analytical Chemistry and Astrophysics, among others.

Motivated by a large number of recent works on these models and different statistical tools, in this dissertation we focus on the development of three new contributions for the linear measurement error models that consider one covariate measured with error. First, we study maximum penalized likelihood inference in a normal measurement error model (MEM) considering five identifiable conditions, which can provide some improvement on the maximum likelihood estimators. Second, we construct two estimation methods based on generalized fiducial inference for the variance parameters in the Grubbs model for the two-instrument case. Third, we propose a heteroscedastic MEM based on skew and heavy-tailed distributions with known error variances. The proposed model can accommodate different levels of heaviness in the tails of the unobserved covariate and random error distributions.

Therefore, the purpose of this dissertation is to show three different contributions in measurement error models. We can list the following five main goals: (i) To develop maximum penalized likelihood inference, proposed by Firth (1993), in a normal measurement error model taking into account five identifiability cases and to carry out an extensive simulation study making a comparison with the maximum likelihood estimators, (ii) To construct a fiducial generalized pivotal quantity and the generalized fiducial distribution for the parameters of interest under the Grubbs model for the two-instruments case, (iii) To present a new heteroscedastic MEM by modeling the unobserved covariate by a skew-t (ST) distribution and for the random errors, we assume the centered Student's t (cT) distribution, where the degrees of freedom for both distributions can be different, (iv) To implement an EM-type algorithm for the estimation of the model parameters for the proposed model in (iii) and to present simulation studies to assess the performance of the estimators and the robustness of the model and (v) To illustrate each methodology mentioned above with a real dataset.

The chapters of this dissertation are written independently. Each chapter has its own literature review and theoretical and numerical developments. The chapters are related in the sense that they deal with measurement error models. This dissertation is structured as follows.

In Chapter 1, after a brief introduction of this dissertation we present a summary of some notations and distributions used.

In Chapter 2, we briefly review the Firth's method and after that, we apply it to the normal measurement error models under five identifiability cases. The estimation process is based on the maximum penalized likelihood method. Next, we present some theoretical results and approximate confidence intervals for the slope parameter. Also, we show simulation studies to gauge the performance of the estimators of the parameters of interest and we present two applications with real datasets to illustrate the proposed methodology.

In Chapter 3, we give a short account of two existing estimation approaches for the Grubbs model and then we construct two new estimation procedures based on generalized fiducial inference. We compare these two new procedures with two existing approaches to examine properties of the estimators of parameters in the model via a simulation study. We apply the proposed methodology in the analysis of a real dataset from a methods comparison study. Some remarks are also considered.

In Chapter 4, we formulate the proposed model based on skew and heavy-tailed distributions and develop an EM-type algorithm for the parameter estimation. We investigate the performance of the maximum likelihood estimators through simulation studies. We illustrate the proposed model with a data set from a methods comparison study. Some concluding remarks are also presented.

In Chapter 5, we describe briefly some problems that can be considered as future work.

1.1 Notation and some distributions

In this section, we describe some general notations used throughout the dissertation. We use bold face letters to indicate vectors and matrices. Let $\mathscr{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$, $\mathscr{SN}_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda})$, $t_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$, $ct_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$ and $\mathscr{S}t_p(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda},\nu)$ denote the p-dimensional normal, skew-normal, Student's t, centered Student's t and skew-t distributions, respectively. Here, $\boldsymbol{\mu} \in \mathbb{R}^p$ is a location vector, $\boldsymbol{\Sigma}$ is a $p \times p$ positive definite scale matrix, $\boldsymbol{\lambda} \in \mathbb{R}^p$ is a vector of skewness parameters and $\nu > 0$ is the degrees of freedom (df). Let $\mathscr{G}(\alpha,\beta)$ denote the gamma distribution with shape parameter $\alpha > 0$ and rate parameter $\beta > 0$. We denote the chi-square distribution with ν degrees of freedom by χ^2_ν. Also, we use $\mathscr{T}\mathscr{N}(\mu,\sigma^2;(a,b))$ to denote the $\mathscr{N}(\mu,\sigma^2)$ distribution truncated to lie in the interval (a,b), $a < b$. The subscript in the univariate case will be omitted.

Let $\mathbf{\Sigma}^{1/2}$ denote the square root of a symmetric and positive definite matrix $\mathbf{\Sigma}$ that satisfies $\mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{1/2^\top} = \mathbf{\Sigma}$, where "$\top$" denotes transposition, $\mathbf{\Sigma}^{-1/2}$ is the inverse of $\mathbf{\Sigma}^{1/2}$ and $\det(\mathbf{\Sigma})$ denotes the determinant of $\mathbf{\Sigma}$. $\boldsymbol{I}_r$ indicates the $r \times r$ unity matrix and $\mathbf{0}_r$ is the $r \times 1$ vector of zeros. Also, "diag(a,b)" denotes a 2×2 matrix whose elements a and b are on the main diagonal. Let $f(\boldsymbol{z};\boldsymbol{\theta})$ denote the probability density function (pdf) of a random vector $\boldsymbol{Z}$ with parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^d$. We also use $\phi_p(\cdot;\boldsymbol{\mu},\mathbf{\Sigma})$ and $\Phi_p(\cdot;\boldsymbol{\mu},\mathbf{\Sigma})$ to indicate the pdf and the cumulate distribution function (cdf) of the $\mathscr{N}_p(\boldsymbol{\mu},\mathbf{\Sigma})$ distribution. In addition, we use $\phi(\cdot)$ and $\Phi(\cdot)$ to denote the pdf and cdf of the standard normal distribution, respectively, while $f_t(\cdot;\boldsymbol{\mu},\mathbf{\Sigma},\nu)$ denotes the pdf of the $t_p(\boldsymbol{\mu},\mathbf{\Sigma},\nu)$ distribution and $F_t(\cdot\,;\nu)$ denotes the cdf of the univariate Student's t distribution with ν df.

We say that a random variable Z follows the gamma distribution, we write $Z \sim \mathscr{G}(\alpha,\beta)$, if its pdf is given by $f(z;\alpha,\beta) = \beta^\alpha z^{\alpha-1}\exp(-\beta z)/\Gamma(\alpha)$, $z > 0$, where $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ is the gamma function evaluated at $x > 0$.

We say that a random vector $\boldsymbol{Z}$ follows the centered Student's t (cT) distribution, that is, $\boldsymbol{Z} \sim ct_p(\boldsymbol{\mu},\mathbf{\Sigma},\nu)$, if its pdf is $f(\boldsymbol{z};\boldsymbol{\mu},\mathbf{\Sigma},\nu) = K(\nu,p)[\det(\mathbf{\Sigma})]^{-1/2}\left[(\nu-2)+\Delta^*\right]^{-(\nu+p)/2}$, $\boldsymbol{z} \in \mathbb{R}^p$, where $\Delta^* = (\boldsymbol{z}-\boldsymbol{\mu})^\top\mathbf{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\mu})$ and $K(\nu,p)$ is the normalizing constant given by $\left[(\nu-2)^{\nu/2}\Gamma(\nu/2+p/2)\right]/[\pi^{p/2}\Gamma(\nu/2)]$, $\nu > 2$. It is a version of centered parameterization of the Student's t distribution (SUTRADHAR, 1993), where $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ correspond to the mean vector and covariance matrix, respectively.

A random vector $\boldsymbol{Z}$ is said to follow the skew-normal (SN) distribution, that is, $\boldsymbol{Z} \sim \mathscr{S}\mathscr{N}_p(\boldsymbol{\mu},\mathbf{\Sigma},\boldsymbol{\lambda})$, if its pdf is $f(\boldsymbol{z};\boldsymbol{\mu},\mathbf{\Sigma},\boldsymbol{\lambda}) = 2\phi_p(\boldsymbol{z};\boldsymbol{\mu},\mathbf{\Sigma})\Phi(\boldsymbol{\lambda}^\top\boldsymbol{z}^*)$, $\boldsymbol{z} \in \mathbb{R}^p$, where $\boldsymbol{z}^* = \mathbf{\Sigma}^{-1/2}(\boldsymbol{z}-\boldsymbol{\mu})$.

Finally, we say that a random vector $\boldsymbol{Z}$ follows the skew-t (ST) distribution, $\boldsymbol{Z} \sim \mathscr{S}t_p(\boldsymbol{\mu},\mathbf{\Sigma},\boldsymbol{\lambda},\nu)$, if its pdf is $f(\boldsymbol{z};\boldsymbol{\mu},\mathbf{\Sigma},\boldsymbol{\lambda},\nu) = 2f_t(\boldsymbol{z};\boldsymbol{\mu},\mathbf{\Sigma},\nu)F_t(\boldsymbol{\lambda}^\top\boldsymbol{z}^*\{(\nu+p)/(\nu+\Delta^*)\}^{1/2};\nu+p)$, $\boldsymbol{z} \in \mathbb{R}^p$.

For additional properties related to the skew-normal and the skew-t distributions, the interested reader is referred to Azzalini and Capitanio (1999) and Azzalini and Capitanio (2003), respectively. Moreover, for developments on skewness distributions in the context of MEM, we refer to Arellano-Valle *et al.* (2005) and Lachos *et al.* (2010).

CHAPTER

2

MAXIMUM PENALIZED LIKELIHOOD INFERENCE IN MEASUREMENT ERROR MODELS

In this chapter, we study maximum penalized likelihood inference in a normal measurement error model under five identifiability cases. We briefly review some recent works on Firth's method and our motivation for the development of this chapter is also presented. After a brief description of the model, we present a summary of Firth's method, and then we apply it to the model in Section 2.2.1. Also, we propose different approximate confidence intervals for the slope parameter. Moreover, we conduct several simulation studies to assess the performance of some properties of the estimators for the parameter of interest and illustrate with two datasets the proposed methodology. Lastly, concluding remarks are presented.

2.1 Introduction

Due to the great importance of the maximum likelihood (ML) estimators, many techniques have been developed to attempt to improve the properties of the ML estimates in small or moderate samples. For instance, analytical expressions for the second-order bias of the ML estimators in several statistical models have been widely reported; see, Efron (1975), Cox and Snell (1968), Schaefer (1983) and Cordeiro and McCullagh (1991), among others. Additionally, these different methodologies are of bias-correction type and may only be applied if the ML estimate is finite.

On the other hand, another alternative approach that has been suggested to remove of the second-order bias of the ML estimate, the method proposed by Firth (1993) has received considerable recent attention, which can be viewed as a preventive method. A noteworthy advantage of Firth's method is that it produces a family of estimators that are not computed directly from the ML estimates. Since then, the latter characteristic has motivated research in some common situations where the ML estimate can be infinite. For instance, logistic regression models and Cox regression, see Heinze and Schemper (2002), Heinze and Schemper (2001); for censored data with exponential lifetimes, see Pettitt, Kelly and Gao (1998), models based on the asymmetric distributions, see Sartori (2006), Lagos Álvarez and Jiménez Gamero (2012), Arrué, Arellano-Valle and Gómez (2016). Also, for the skew-normal and skew-t distributions, a more general extension of Firth's method was studied by Azzalini and Arellano-Valle (2013).

Furthermore, the bias reduction for the parameters of interest in a class of multivariate generalized nonlinear models was studied in Kosmidis and Firth (2009); in multinomial logistic models by using the equivalent Poisson log-linear model was developed in Kosmidis and Firth (2011) and in cumulative link models for ordinal data was investigated in Kosmidis (2014). Also, Ospina, Cribari-Neto and Vasconcellos (2006), Hefley and Hooten (2015), Meyvisch (2016) and Siino, Fasola and Muggeo (2016) are other recent applications of Firth's method.

In exponential family models with canonical parameterization, Firth (1993) showed that the prevention scheme consists in maximizing a penalized likelihood, so that the Firth's estimator can be computed by numerically maximizing that modified likelihood function. Furthermore, as pointed out by him, the reduction in bias may sometimes be accompanied by inflation of variance, i.e., the resulting estimator may have a mean squared error larger than that of the uncorrected one. Despite this fact, it is clear from published empirical studies such as those mentioned earlier, that in some models, bias prevention by Firth's method may improve the uncorrected ML estimator, especially when the sample size is not large.

Additionally, for interval estimation in a penalized likelihood context, Heinze and Schemper (2002) studied approximate confidence intervals based on the profile penalized likelihood ratio test. Numerical findings reported by them showed an improvement on the properties of ML estimators in logistic regression models. For that model, Siino, Fasola and Muggeo (2016) also studied approximate confidence intervals based on the Wald-type statistic with a sandwich formula for the asymptotic variance of the Firth's estimator. Moreover, Heinze and Schemper (2002) indicated that the confidence intervals based on the profile penalized likelihood ratio test can perform better when the intervals present an asymmetric shape, while for symmetric shape the Wald-type intervals can be considered. Then, a better estimator here means not only in terms of bias, but also in terms of other properties such as finiteness, mean squared error and the coverage probability of approximate confidence intervals. The foregoing remarks stress that such improvements can not be attained simultaneously.

Motivated by a large number of recent papers on Firth's method, which can be used to improve some asymptotic properties of the ML estimators, we carry out a study applying Firth's method to linear measurement error models. We also construct the approximate confidence intervals for the slope parameter using the maximum penalized likelihood method and we compare with that of the ML method.

2.2 Model and parameter estimation

In this section, we describe the normal MEM, a review on the Firth's method and its application to the model.

2.2.1 Model

The linear MEM with one covariate assumes that the unobserved variables x_i and y_i are related by $y_i = \beta_0 + \beta_1 x_i$ and wherein

$$Y_i = y_i + \varepsilon_i \quad \text{and} \quad X_i = x_i + u_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where β_0 and β_1 are the intercept and the slope parameters, respectively; x_i and y_i are the true covariate and the true response variable, respectively; X_i and Y_i denote the observed variables corresponding to x_i and y_i; ε_i and u_i are the additive error terms and n is the sample size. An important characteristic of (2.1) is that if the unobserved covariates x_i's are random variables, then it is called a structural MEM and if the x_i's are unknown constants, then we have a functional MEM.

Assume in (2.1) that x_i, u_i and ε_i are distributed as random variables from the $\mathcal{N}(\mu_x, \sigma_x^2)$, $\mathcal{N}(0, \sigma_u^2)$ and $\mathcal{N}(0, \sigma_\varepsilon^2)$ distributions, respectively, and x_i, u_j and ε_k are independent for all $i, j, k = 1, \ldots, n$. From these assumptions, it follows that $\mathbf{Z}_i = (X_i, Y_i)^\top$, $i = 1, \ldots, n$, are independent and distributed as

$$\mathbf{Z}_i \sim \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{2.2}$$

where the mean vector is $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ are given by

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_x \\ \beta_0 + \beta_1 \mu_x \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_x^2 + \sigma_u^2 & \beta_1 \sigma_x^2 \\ \beta_1 \sigma_x^2 & \beta_1^2 \sigma_x^2 + \sigma_\varepsilon^2 \end{bmatrix}.$$

Since the true covariate x_i is treated as a random variable, we are dealing with a structural formulation. When normality is assumed, the model in (2.2) poses identifiability problems (CHENG; NESS, 1999, Appendix A). Thus, to deal with it is necessary to impose additional assumptions to make the model identifiable (CHENG; NESS, 1999, Section 1.2). In this chapter, we deal with the following additional assumptions often used:

Case 1: σ_u^2 is known.

Case 2: σ_ε^2 is known.

Case 3: The ratio of the error variances, $\lambda = \sigma_\varepsilon^2 / \sigma_u^2$, is known.

Case 4: The reliability ratio, $k_x = \sigma_x^2 / (\sigma_x^2 + \sigma_u^2)$, is known.

Case 5: The intercept is null, $\beta_0 = 0$, with $\mu_x \neq 0$.

There are several studies available that addressed for parameter estimation of the model in (2.2) for each identifiability case, where the common methods are: maximum likelihood, weighted least square, orthogonal regression and moment methods, see, e.g., Fuller (1987, Section 1.2–1.3), Hood, Nix and Iles (1999), Thompson and Carter (2007) and Gillard (2010). Some classical references considering the maximum likelihood method are, for instance, for Cases 1 and 2, see Birch (1964); for Case 3, see Madansky (1959); for Case 4, see Cheng and Ness (1999, p. 17) and for Case 5, see Chan and Mak (1979).

2.2.2 Firth's maximum penalized likelihood method

In this section, a brief review of the method proposed by Firth (1993) is presented. Let $L(\boldsymbol{\theta};\boldsymbol{z})$ be the likelihood function of a regular parametric model, where $\boldsymbol{\theta}$ is an unknown parameter vector and $\boldsymbol{z}$ is the vector of the observed data. The maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_p)^\top$ is obtained as a solution of the equation $\partial\log(L(\boldsymbol{\theta};\boldsymbol{z}))/\partial\boldsymbol{\theta} \equiv \boldsymbol{U}(\boldsymbol{\theta};\boldsymbol{z}) = \boldsymbol{0}$, where $\boldsymbol{U}(\cdot;\cdot)$ denotes the score functions. Under certain conditions, Firth (1993) suggested penalizing the likelihood function by the factor $[\det(\boldsymbol{I_\theta})]^{1/2}$, that is, $L_p(\boldsymbol{\theta};\boldsymbol{z}) = L(\boldsymbol{\theta};\boldsymbol{z})\left[\det(\boldsymbol{I_\theta})\right]^{1/2}$, where $L_p(\boldsymbol{\theta};\cdot)$ denotes the penalized likelihood function and $\boldsymbol{I_\theta}$ is the expected information matrix, which is computed as minus the expected value of the second derivatives of the log-likelihood function. Notice that $[\det(\boldsymbol{I_\theta})]^{1/2}$ corresponds to the Jeffreys' prior in a Bayesian context. By using this modification, Firth (1993) showed that second-order bias of the maximum likelihood estimates $\widehat{\boldsymbol{\theta}}$ is removed.

The penalized log-likelihood function takes the form

$$\ell_p(\boldsymbol{\theta}) = \log(L_p(\boldsymbol{\theta};\boldsymbol{z})) = \log(L(\boldsymbol{\theta};\boldsymbol{z})) + \frac{1}{2}\log(\det(\boldsymbol{I_\theta})). \tag{2.3}$$

Thus, the Firth's estimate $\tilde{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ is obtained as a solution of the modified score equations, that is,

$$U_r^*(\boldsymbol{\theta};\boldsymbol{z}) \equiv U_r(\boldsymbol{\theta};\boldsymbol{z}) + \frac{1}{2}\,\mathrm{tr}\left(\boldsymbol{I}_{\boldsymbol{\theta}}^{-1}\frac{\partial \boldsymbol{I_\theta}}{\partial\theta_r}\right) = 0,\ r = 1,\ldots,p,$$

where $U_r(\boldsymbol{\theta};\boldsymbol{z})$ and θ_r are the r-th component of $\boldsymbol{U}(\boldsymbol{\theta};\boldsymbol{z})$ and $\boldsymbol{\theta}$, respectively, while "tr" denotes the trace of a matrix. Hereafter, we refer to the Firth's estimator as the maximum penalized likelihood (MPL) estimator.

2.2.3 Applying Firth's method to measurement error models

It follows from (2.2) that the log-likelihood function of $\boldsymbol{\theta}$ given the observed data $\boldsymbol{z} = (\boldsymbol{z}_1^\top,\ldots,\boldsymbol{z}_n^\top)^\top$ can be expressed as

$$\ell(\boldsymbol{\theta}) = \log(L(\boldsymbol{\theta};\boldsymbol{z})) = -n\log(2\pi) - \frac{n}{2}\left[\frac{A}{B} + \frac{C}{B} + \log(B)\right], \tag{2.4}$$

where $A = S_X'^2(\beta_1^2\sigma_x^2 + \sigma_\varepsilon^2) - 2\beta_1\sigma_x^2 S_{XY}'^2 + S_Y'^2(\sigma_x^2 + \sigma_u^2)$, $B = \beta_1^2\sigma_x^2\sigma_u^2 + \sigma_u^2\sigma_\varepsilon^2 + \sigma_x^2\sigma_\varepsilon^2$ and $C = (\overline{X} - \mu_x)^2(\beta_1^2\sigma_x^2 + \sigma_\varepsilon^2) - 2\beta_1\sigma_x^2(\overline{X} - \mu_x)(\overline{Y} - \beta_0 - \beta_1\mu_x) + (\overline{Y} - \beta_0 - \beta_1\mu_x)^2(\sigma_x^2 + \sigma_u^2)$, with

$\overline{X} = n^{-1}\sum_{i=1}^{n} X_i$, $\overline{Y} = n^{-1}\sum_{i=1}^{n} Y_i$ $S_X'^2 = n^{-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$, $S_Y'^2 = n^{-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$ and $S'_{XY} = n^{-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$. Notice that B is the determinant of the covariance matrix of $\boldsymbol{Z}_i$ given in (2.2) and $\boldsymbol{\theta}$ is the unknown parameter vector with components β_0, β_1, μ_x, σ_x^2, σ_u^2 and σ_ε^2.

The ML estimators for the identifiability Cases 1–5 can also be found in Hood, Nix and Iles (1999) and Cheng and Ness (1999, p. 17), as well as the expressions of the expected information matrix $\boldsymbol{I_\theta}$. Additionally, Wang (2004) calculated the determinant of $\boldsymbol{I_\theta}$ for Case 1 and 3. Algebraic expressions not available in the literature will be provided in this chapter. Next, we briefly review the expressions of the ML estimators and then we apply Firth's method for each case of model identifiable considered here.

Case 1: σ_u^2 is known. In this case, the components of the vector $\boldsymbol{\theta}$ are β_0, β_1, μ_x, σ_x^2 and σ_ε^2 and with ML estimators given by

$$\widehat{\mu}_x = \overline{X}, \quad \widehat{\beta}_1 = \frac{S_{XY}}{S_X'^2 - \sigma_u^2}, \quad \widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1\widehat{\mu}_x, \quad \widehat{\sigma}_x^2 = S_X'^2 - \sigma_u^2 \quad \text{and } \widehat{\sigma}_e^2 = \frac{S_X'^2 S_Y'^2 - \sigma_u^2 S_Y'^2 - S_{XY}'^2}{S_X'^2 - \sigma_u^2},$$

with the constraints $S_X'^2 > \sigma_u^2$ and $S_Y'^2 > S_{XY}'^2/(S_X'^2 - \sigma_u^2)$.

On the other hand, the penalized log-likelihood function associated to this case follows after substituting the log-likelihood function into (2.4) and $\det(\boldsymbol{I_\theta}) = n^5\sigma_x^4/4B^4$ into the function in (2.3), so that we get

$$\ell_p(\boldsymbol{\theta}) = c_1 - \frac{n}{2}\left[\frac{A}{B} + \frac{C}{B} + \left(1 + \frac{4}{n}\right)\log(B)\right] + \log(\sigma_x^2), \tag{2.5}$$

where c_1 is free of parameters, A, B and C are given as in (2.4). The MPL estimators for β_0, β_1, σ_x^2 and σ_u^2 are not available in closed form and estimates need to be computed numerically from the maximization of the function in (2.5), except for μ_x. It can be seen that the penalty term does not depend on μ_x, so that the ML and MPL estimators for μ_x are the same.

Case 2: σ_ε^2 is known. In this case, the components of the vector $\boldsymbol{\theta}$ are β_0, β_1, μ_x, σ_x^2 and σ_u^2 with ML estimators given by

$$\widehat{\mu}_x = \overline{X}, \quad \widehat{\beta}_1 = \frac{S_Y'^2 - \sigma_\varepsilon^2}{S'_{XY}}, \quad \widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1\widehat{\mu}_x, \quad \widehat{\sigma}_x^2 = \frac{S_{XY}'^2}{S_Y'^2 - \sigma_\varepsilon^2} \quad \text{and} \quad \widehat{\sigma}_u^2 = \frac{S_X'^2 S_Y'^2 - \sigma_\varepsilon^2 S_X'^2 - S_{XY}'^2}{S_Y'^2 - \sigma_\varepsilon^2},$$

with the constraints $S_Y'^2 > \sigma_\varepsilon^2$ and $S_X'^2 > S_{XY}'^2/(S_Y'^2 - \sigma_\varepsilon^2)$.

On the other side, the penalized log-likelihood function follows after substituting the log-likelihood function into (2.4) and $\det(\boldsymbol{I_\theta}) = n^5\sigma_x^4\beta_1^4/4B^4$ into the function in (2.3), so that we get

$$\ell_p(\boldsymbol{\theta}) = c_2 - \frac{n}{2}\left[\frac{A}{B} + \frac{C}{B} + \left(1 + \frac{4}{n}\right)\log(B)\right] + \log(\sigma_x^2) + \log(\beta_1^2), \tag{2.6}$$

where c_2 is free of parameters, A, B and C are given as in (2.4). The MPL estimators for β_0, β_1, σ_x^2 and σ_u^2 are not available in closed form and estimates need to be computed numerically from the maximization of the function in (2.6), except for μ_x. In a similar way to the Case 1, the ML and MPL estimators for μ_x are the same.

Case 3: Ratio of the error variances is known. Since λ is known, it follows that $\sigma_\varepsilon^2 = \lambda\sigma_u^2$ and, consequently, the components of the vector $\boldsymbol{\theta}$ are β_0, β_1, μ_x, σ_x^2 and σ_u^2 and ML estimators are given by

$$\widehat{\mu}_x = \overline{X}, \quad \widehat{\beta}_1 = \frac{S_Y'^2 - \lambda S_X'^2 + [(S_Y'^2 - \lambda S_X'^2)^2 + 4\lambda S_{XY}'^2]^{1/2}}{2S_{XY}'}, \quad \widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1\widehat{\mu}_x,$$
$$\widehat{\sigma}_x^2 = \frac{S_{XY}'}{\widehat{\beta}_1} \quad \text{and} \quad \widehat{\sigma}_u^2 = \frac{S_Y'^2 - 2\widehat{\beta}_1 S_{XY}' + \widehat{\beta}_1^2 S_X'^2}{\lambda + \widehat{\beta}_1^2}. \tag{2.7}$$

Here, there are no constraints, for the variance estimators in (2.7) are always non-negative.

The penalized log-likelihood function follows after substituting the log-likelihood function into (2.4) and $\det(\boldsymbol{I_\theta}) = n^5\sigma_x^4(\beta_1^2+\lambda)^2/4B^4$ into the function in (2.5), so that we get

$$\ell_p(\boldsymbol{\theta}) = c_3 - \frac{n}{2}\left[\frac{A}{B} + \frac{C}{B} + \left(1+\frac{4}{n}\right)\log(B)\right] + \log(\sigma_x^2) + \log(\beta_1^2+\lambda), \tag{2.8}$$

where c_3 is free of parameters and, for this case, A, B and C are as in (2.4), but replacing σ_ε^2 by $\lambda\sigma_u^2$. The MPL estimators are found by deriving and solving the modified score equations obtained from (2.8). Next, we can establish the following result.

Proposition 1. In the structural model (2.2), assume the ratio of the error variances is known. Then, ML and MPL estimators for β_0, β_1 and μ_x are the same and are given in (2.7).

Proof. Notice that in (2.8), the term C is given by $(\overline{X}-\mu_x)^2(\beta_1^2\sigma_x^2+\lambda\sigma_u^2) - 2\beta_1\sigma_x^2(\overline{X}-\mu_x)(\overline{Y} - \beta_0 - \beta_1\mu_x) + (\overline{Y}-\beta_0-\beta_1\mu_x)^2(\sigma_x^2+\sigma_u^2)$, which is minimized and equal to 0 when $\mu_x = \overline{X}$ and $\beta_0 + \beta_1\mu_x = \overline{Y}$. These are the modified score equations to determinate the MPL estimators for μ_x and β_0, since they do not appear in the other terms in (2.8). Hence, it can be noted that the MPL and ML estimators for μ_x are the same, i.e., $\tilde{\mu}_x = \hat{\mu}_x = \overline{X}$. Next, we have to prove that the expression of $\tilde{\beta}_1$ is equal to that of $\hat{\beta}_1$ given in (2.7) and, consequently, the assertion for $\tilde{\beta}_0$ follows. The penalized log-likelihood problem is then solved by maximizing the function in (2.8) without the term C/B with respect to β_1, σ_x^2 and σ_u^2. Thus, the partial derivatives are taken and set equal to 0, giving the following system of equations:

$$\begin{aligned}
\left\{\sigma_x^2(\beta_1+\lambda)\left[n(\beta_1 S_X'^2 - S_{XY}') + (n+2)\beta_1\sigma_u^2\right] - 2\lambda\beta_1\sigma_u^4\right\}B - n\beta_1\sigma_x^2\sigma_u^2(\beta_1^2+\lambda)A &= 0,\\
\left\{n\sigma_x^2\left(S_X'^2\beta_1^2 - 2\beta_1 S_{XY}' + S_Y'^2\right) + (n+2)(\beta_1+\lambda)\sigma_x^2\sigma_u^2 - 2\lambda\sigma_u^4\right\}B - n\sigma_x^2\sigma_u^2(\beta_1^2+\lambda)A &= 0,\\
\left\{n(S_X'^2\lambda + S_Y'^2) + (n+4)\left[(\beta_1^2+\lambda)\sigma_x^2 + 2\lambda\sigma_u^2\right]\right\}B - n\left[(\beta_1^2+\lambda)\sigma_x^2 + 2\lambda\sigma_u^2\right]A &= 0.
\end{aligned} \tag{2.9}$$

Now, the strategy is to reduce the first and second equations in (2.9) to a single equation, so that the term A is eliminated. After doing that, the resulting equation is $[S_{XY}\beta_1^2 + (\lambda S_X^2 -$

$S_Y^2)\beta_1 - \lambda S_{XY}]B = 0$. Notice that B is assumed to be positive. Then, it follows that $S_{XY}\beta_1^2 + (\lambda S_X^2 - S_Y^2)\beta_1 - \lambda S_{XY} = 0$. Finally, solving this equation, the result holds. □

The MPL estimators for the variance parameters, σ_x^2 and σ_u^2, are not available in closed form. Then, estimates are computed numerically by maximizing the penalized log-likelihood function in (2.8) after substituting the MPL estimators for β_0, β_1 and μ_x given in (2.7).

Case 4: Reliability ratio is known. Since $k_x = \sigma_x^2/(\sigma_x^2 + \sigma_u^2)$ is known, it follows that $\sigma_u^2 = \tau\sigma_x^2$, where $\tau = (1-k_x)/k_x$ and, consequently, the components of the vector $\boldsymbol{\theta}$ are β_0, β_1, μ_x, σ_x^2 and σ_ε^2 with ML estimators given by

$$\widehat{\mu}_x = \overline{X}, \quad \widehat{\beta}_1 = \frac{S'_{XY}}{k_x S_X'^2}, \quad \widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1\widehat{\mu}_x, \quad \widehat{\sigma}_\varepsilon^2 = S_Y'^2 - \frac{S_{XY}'^2}{k_x S_X'^2} \text{ and } \widehat{\sigma}_x^2 = k_x S_X'^2, \tag{2.10}$$

with the constraint $S_Y'^2 > S_{XY}'^2/(k_x S_X'^2)$.

The penalized log-likelihood function for this case follows after substituting the log-likelihood function into (2.4) and $\det(\boldsymbol{I_\theta}) = n^5\sigma_x^4(1+\tau)^2/4B^4$ into the function in (2.3), so that we get

$$\ell_p(\boldsymbol{\theta}) = c_4 - \frac{n}{2}\left[\frac{A}{B} + \frac{C}{B} + \left(1 + \frac{4}{n}\right)\log(B)\right] + \log(\sigma_x^2) + \log(1+\tau), \tag{2.11}$$

where c_4 is free of parameters and, for this case, A, B and C are as in (2.4), but replacing σ_u^2 by $\tau\sigma_x^2$. Thus, the MPL estimators are found by deriving and solving the modified score equations obtained from (2.11). Next, the following result can be established.

Proposition 2. In the structural model (2.2), assume the reliability ratio is known. Then, the MPL estimators for β_0, β_1 and μ_x are the same to the ML estimators given in (2.10).

Proof. Notice that in (2.11), the term C is given by $(\overline{X} - \mu_x)^2(\beta_1^2\sigma_x^2 + \sigma_\varepsilon^2) - 2\beta_1\sigma_x^2(\overline{X} - \mu_x)(\overline{Y} - \beta_0 - \beta_1\mu_x) + \sigma_x^2(\overline{Y} - \beta_0 - \beta_1\mu_x)^2(1+\tau)$, which is minimized and equal to 0 when $\mu_x = \overline{X}$ and $\beta_0 + \beta_1\mu_x = \overline{Y}$. Here, μ_x and β_0 do not appear again in the other terms in (2.11). Also, it can be seen that the MPL and ML estimators for μ_x are the same. In a similar way to Case 3, we proceed to prove that the MPL and ML estimators for β_0 and β_1 are the same. The penalized log-likelihood problem is then solved by maximizing the function in (2.11) without the term C/B with respect to β_1, σ_x^2 and σ_ε^2. The partial derivatives are taken and set equal to 0, giving the following system of equations:

$$\begin{aligned}
\left[n\beta_1 S_X'^2 - nS'_{XY} + (n+4)\tau\beta_1\right]B - n\beta_1\sigma_x^2\tau A &= 0,\\
\left[nS_X'^2 + (n+4)(1+\tau)\sigma_x^2\right]B - n(1+\tau)\sigma_x^2 A &= 0,\\
a^*B - n\left[(1+\tau)\sigma_\varepsilon^2 + 2\tau\beta_1^2\sigma_x^2\right]A &= 0,
\end{aligned} \tag{2.12}$$

where $a^* = n\beta_1(S_X'^2\beta_1 - 2S'_{XY}) + (1+\tau)\left[nS_Y'^2 + (n+2)\sigma_\varepsilon^2\right] + 2(n+3)\tau\beta_1^2\sigma_x^2$.

Applying the same strategy as in Proposition 1, we reduce the first and second equations in (2.12) to a single equation. After some algebraic manipulations, we get the MPL estimator for β_1, which is equal to the ML estimator $\hat{\beta}_1$ given in (2.10). □

The MPL estimators for the variance parameters, σ_x^2 and σ_ε^2, are not available in closed form. Then, estimates are computed numerically by maximizing the penalized log-likelihood function in (2.11) after substituting the MPL estimators for β_0, β_1 and μ_x given into (2.10).

Case 5: Intercept is null. In this case, the components of the vector $\boldsymbol{\theta}$ are β_1, μ_x, σ_x^2, σ_u^2 and σ_ε^2 with ML estimators given by

$$\widehat{\mu}_x = \overline{X}, \quad \widehat{\beta}_1 = \frac{\overline{Y}}{\overline{X}}, \quad \widehat{\sigma}_x^2 = \frac{S'_{XY}}{\widehat{\beta}_1}, \quad \widehat{\sigma}_u^2 = S'^2_X - \frac{S'_{XY}}{\widehat{\beta}_1} \quad \text{and} \quad \widehat{\sigma}_\varepsilon^2 = S'^2_Y - \widehat{\beta}_1 S'_{XY},$$

with the constraints $\overline{X} \neq 0$, $S'^2_X > S'_{XY}/\widehat{\beta}_1$ and $S'^2_Y > \widehat{\beta}_1 S'_{XY}$.

For this case, the penalized log-likelihood function follows after substituting the log-likelihood function into (2.4) and $\det(\boldsymbol{I}_{\boldsymbol{\theta}}) = n^5\mu_x^2\beta_1^2/4B^4$ into the function in (2.5), so that we get

$$\ell_p(\boldsymbol{\theta};\boldsymbol{z}) = c_5 - \frac{n}{2}\left[\frac{A}{B} + \frac{C}{B} + \left(1 + \frac{4}{n}\right)\log B\right] + \log(\sigma_x^2) + \log(|\beta_1\mu_x|), \tag{2.13}$$

where c_5 is free of parameters and, in this case, A, B and C are as in (2.4), but replacing β_0 by 0. Thus, the MPL estimators for all the parameters involved are not available in closed form and estimates need to be computed numerically by maximizing the function in (2.13).

2.3 Interval estimation

In this section, we focus on different approaches for constructing confidence intervals for the slope parameter β_1. These confidence intervals are based on the Wald statistic and the profile penalized log-likelihood ratio test.

2.3.1 Wald-type approximate confidence interval estimation

Under standard regularity conditions, the ML estimator $\hat{\boldsymbol{\theta}}$ is approximately normal with mean vector $\boldsymbol{\theta}$ and asymptotic covariance matrix $\boldsymbol{I}_{\boldsymbol{\theta}}^{-1}$ (LEHMANN, 1998, Chapter 7). This result is often used to obtain asymptotic confidence intervals based on the Wald statistic for the components of the parameter vector $\boldsymbol{\theta}$. Moreover, a consistent estimator of the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$ is $\boldsymbol{I}_{\boldsymbol{\theta}}^{-1}$ evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, which is denoted by $\boldsymbol{I}_{\hat{\boldsymbol{\theta}}}^{-1}$. Hence, with $\gamma \in (0,1)$, an approximate $100\gamma\%$ confidence interval for the parameter of interest β_1 is obtained as

$$CI_1 = [\hat{\beta}_1 - z_{(1+\gamma)/2}\hat{\sigma}_1(\hat{\beta}_1),\ \hat{\beta}_1 + z_{(1+\gamma)/2}\hat{\sigma}_1(\hat{\beta}_1)], \tag{2.14}$$

where z_δ denotes the $100\delta\%$ percentile of the standard normal distribution and $\hat{\sigma}_1(\hat{\beta}_1)$ is the square root of the element corresponding to β_1 on the main diagonal of the estimated covariance matrix $\boldsymbol{I}_{\hat{\boldsymbol{\theta}}}^{-1}$. Thus, the inverse of the expected information matrix $\boldsymbol{I}_{\boldsymbol{\theta}}^{-1}$ associated to the model in (2.2) can also be found from Hood, Nix and Iles (1999), especially, for Cases 1–4. For Case 5, we can establish the following result.

Lemma 1. In the structural model (2.2), assume the intercept is null, $\beta_0 = 0$, with $\mu_x \neq 0$. Then,

(a) The expected information matrix is

$$\boldsymbol{I}_{\boldsymbol{\theta}} = \frac{n}{B}\begin{bmatrix} \sigma_x^4 + \mu_x^2\eta_1 + \dfrac{2\beta_1^2\sigma_x^4\sigma_u^4}{B} & \beta_1\mu_x\sigma_u^2 & \dfrac{\beta_1\sigma_x^2\sigma_u^2\eta_2}{B} & -\dfrac{\beta_1\sigma_x^4\sigma_\varepsilon^2}{B} & \dfrac{\beta_1\sigma_x^2\sigma_u^2\eta_1}{B} \\ \beta_1\mu_x\sigma_u^2 & \eta_2 & 0 & 0 & 0 \\ \dfrac{\beta_1\sigma_x^2\sigma_u^2\eta_2}{B} & 0 & \dfrac{\eta_2^2}{2B} & \dfrac{\sigma_\varepsilon^4}{2B} & \dfrac{\beta_1^2\sigma_u^4}{2B} \\ -\dfrac{\beta_1\sigma_x^4\sigma_\varepsilon^2}{B} & 0 & \dfrac{\sigma_\varepsilon^4}{2B} & \dfrac{\eta_3^2}{2B} & \dfrac{\beta_1^2\sigma_x^4}{2B} \\ \dfrac{\beta_1\sigma_x^2\sigma_u^2\eta_1}{B} & 0 & \dfrac{\beta_1^2\sigma_u^4}{2B} & \dfrac{\beta_1^2\sigma_x^4}{2B} & \dfrac{\eta_1^2}{2B} \end{bmatrix},$$

(b) $\det(\boldsymbol{I}_{\boldsymbol{\theta}}) = \dfrac{n^5\mu_x^2\beta_1^2}{4B^4}$,

(c) The inverse of the expected information matrix is

$$\boldsymbol{I}_{\boldsymbol{\theta}}^{-1} = \frac{1}{n}\begin{bmatrix} \dfrac{\eta_2}{\mu_x^2} & -\dfrac{\beta_1\sigma_u^2}{\mu_x} & -\dfrac{\sigma_x^2\eta_2}{\mu_x^2\beta_1} & \dfrac{\sigma_x^2\eta_2}{\mu_x^2\beta_1} & -\dfrac{\beta_1\sigma_x^2\eta_2}{\mu_x^2} \\ -\dfrac{\beta_1\sigma_u^2}{\mu_x} & \eta_1 & \dfrac{\sigma_u^2\sigma_x^2}{\mu_x} & -\dfrac{\sigma_u^2\sigma_x^2}{\mu_x} & \dfrac{\beta_1^2\sigma_u^2\sigma_x^2}{\mu_x} \\ -\dfrac{\sigma_x^2\eta_2}{\mu_x^2\beta_1} & \dfrac{\sigma_u^2\sigma_x^2}{\mu_x} & \dfrac{\mu_x^2\eta_4 + \sigma_x^4\eta_2}{\beta_1^2\mu_x^2} & -\dfrac{\mu_x^2\eta_5 + \sigma_x^4\eta_2}{\beta_1^2\mu_x^2} & \dfrac{\sigma_x^4\eta_2 + \mu_x^2\eta_6}{\mu_x^2} \\ \dfrac{\sigma_x^2\eta_2}{\mu_x^2\beta_1} & -\dfrac{\sigma_u^2\sigma_x^2}{\mu_x} & -\dfrac{\mu_x^2\eta_5 + \sigma_x^4\eta_2}{\beta_1^2\mu_x^2} & \dfrac{\mu_x^2\eta_7 + \sigma_x^4\eta_2}{\beta_1^2\mu_x^2} & -\dfrac{\mu_x^2\eta_8 + \sigma_x^4\eta_2}{\mu_x^2} \\ -\dfrac{\beta_1\sigma_x^2\eta_2}{\mu_x^2} & \dfrac{\beta_1^2\sigma_u^2\sigma_x^2}{\mu_x} & \dfrac{\sigma_x^4\eta_2 + \mu_x^2\eta_6}{\mu_x^2} & -\dfrac{\mu_x^2\eta_8 + \sigma_x^4\eta_2}{\mu_x^2} & \dfrac{\mu_x^2\eta_9 + \sigma_x^4\beta_1^2\eta_2}{\mu_x^2} \end{bmatrix},$$

where B is given in (2.4), but replacing β_0 by 0, $\eta_1 = \sigma_x^2 + \sigma_u^2$, $\eta_2 = \beta_1^2\sigma_u^2 + \sigma_\varepsilon^2$, $\eta_3 = \beta_1^2\sigma_x^2 + \sigma_\varepsilon^2$, $\eta_4 = B + 2\sigma_x^4\beta_1^2$, $\eta_5 = B - 2\sigma_x^2\sigma_u^2\beta_1^2$, $\eta_6 = 2\sigma_x^2\sigma_\varepsilon^2 - B$, $\eta_7 = B + 2\sigma_u^4\beta_1^2$, $\eta_8 = B - 2\sigma_u^2\sigma_\varepsilon^2$ and $\eta_9 = \beta_1^2 B + 2\sigma_\varepsilon^4$.

Next, Table 1 exhibits the expressions of the asymptotic variance of $\hat{\beta}_1$, which is denoted by $\sigma_1^2(\hat{\beta}_1)$. Thus, $\hat{\sigma}_1(\hat{\beta}_1)$ can be computed from the square root of the expressions presented in Table 1 evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

On the other side, based on the Wald statistic and the MPL estimator $\tilde{\boldsymbol{\theta}}$, other three approximate confidence intervals can be proposed. A key feature here is to obtain an consistent estimator for the asymptotic covariance matrix of $\tilde{\boldsymbol{\theta}}$. Thus, a first alternative of approximation of that matrix was suggested by Firth (1993), which can be estimated through the inverse of the

Table 1 – Asymptotic variance of the ML estimator for β_1 ($\sigma_1^2(\hat{\beta}_1)$) under the identifiability Cases 1–5.

Case 1	Case 2	Case 3	Case 4	Case 5
$\frac{B+2\beta_1^2\sigma_u^4}{n\sigma_x^4}$	$\frac{\beta_1^2 B+2\sigma_\varepsilon^4}{n\beta_1^2\sigma_x^4}$	$\frac{B}{n\sigma_x^4}$	$\frac{B}{n\sigma_x^4}$	$\frac{\beta_1^2\sigma_u^2+\sigma_\varepsilon^2}{n\mu_x^2}$

expected information matrix $\boldsymbol{I}_{\boldsymbol{\theta}}^{-1}$ evaluated at $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$ (or simply $\boldsymbol{I}_{\tilde{\boldsymbol{\theta}}}^{-1}$). However, it holds only in large samples when the penalty effect is negligible. A second alternative is to calculate the inverse of the negative of second derivatives of the penalized log-likelihood function $\ell_p(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$, which is denoted by $\boldsymbol{H}_{\tilde{\boldsymbol{\theta}}}^{-1}$, where $\boldsymbol{H}_{\tilde{\boldsymbol{\theta}}} = [\partial^2\ell_p(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top]|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}$ is the Hessian matrix and can be obtained through numerical differentiation. As a third alternative, we have the sandwich formula given by $\boldsymbol{S}_{\tilde{\boldsymbol{\theta}}} = \boldsymbol{H}_{\tilde{\boldsymbol{\theta}}}^{-1}\boldsymbol{I}_{\tilde{\boldsymbol{\theta}}}\boldsymbol{H}_{\tilde{\boldsymbol{\theta}}}^{-1}$. This matrix could provide a more reliable and robust version of the estimated asymptotic covariance matrix of $\tilde{\boldsymbol{\theta}}$ in small or moderate samples, as pointed out in Siino, Fasola and Muggeo (2016). Then, based on the covariance matrices mentioned above, we propose three approximate $100\gamma\%$ confidence intervals for β_1, which are given by

$$CI_m = [\tilde{\beta}_1 - z_{(1+\gamma)/2}\tilde{\sigma}_m(\tilde{\beta}_1),\ \tilde{\beta}_1 + z_{(1+\gamma)/2}\tilde{\sigma}_m(\tilde{\beta}_1)], \quad m \in \{2,3,4\}, \tag{2.15}$$

where $\tilde{\sigma}_2(\tilde{\beta}_1)$, $\tilde{\sigma}_3(\tilde{\beta}_1)$ and $\tilde{\sigma}_4(\tilde{\beta}_1)$ are, respectively, the square root of the elements corresponding to β_1 on the main diagonal of the estimated covariance matrices $\boldsymbol{I}_{\tilde{\boldsymbol{\theta}}}^{-1}$, $\boldsymbol{H}_{\tilde{\boldsymbol{\theta}}}^{-1}$ and $\boldsymbol{S}_{\tilde{\boldsymbol{\theta}}}$.

It is worth noticing that all the approximate confidence intervals given in (2.14) and (2.15) require the estimated asymptotic standard errors, i.e., $\hat{\sigma}_1(\hat{\beta}_1)$ and $\tilde{\sigma}_m(\tilde{\beta}_1)$, $m \in \{2,3,4\}$. On the other hand, the following section presents another alternative for computing an approximate confidence interval for β_1.

2.3.2 Profile penalized likelihood confidence interval

The penalized likelihood ratio statistic is defined as the usual likelihood ratio (LR) statistic, but instead of $\ell(\cdot)$ is used its penalized version, i.e., $\ell_p(\cdot)$. Thus, to test the hypothesis $\beta_1 = \beta_{10}$, the penalized likelihood ratio statistic is $LR = 2[\ell_p(\tilde{\beta}_1, \tilde{\boldsymbol{\xi}}) - \ell_p(\beta_{10}, \tilde{\boldsymbol{\xi}}_{\beta_{10}})]$, where $(\tilde{\beta}_1, \tilde{\boldsymbol{\xi}})^\top$ is the joint MPL estimate of $(\beta_1, \boldsymbol{\xi})^\top$ and $\tilde{\boldsymbol{\xi}}_{\beta_{10}}$ is the MPL estimate for $\boldsymbol{\xi}$ when $\beta_1 = \beta_{10}$. The components of the vector $\boldsymbol{\xi}$ are the same of $\boldsymbol{\theta}$ but omitting β_1. Hence, with $\gamma \in (0,1)$, a profile penalized likelihood (PPL) $100\gamma\%$ confidence interval for β_1 can be obtained from $LR \leq \chi_1^2(\gamma)$, where $\chi_1^2(\delta)$ denotes the $100\delta\%$ percentile of the chi-square distribution with one degree of freedom. Thus, we can represent an approximate PPL $100\gamma\%$ confidence interval for β_1 as $CI_5 = \{\beta_{10} \in \mathbb{R} : LR \leq \chi_1^2(\gamma)\}$.

To compute the PPL confidence intervals for the parameter of interest, we follow an iterative procedure given in Heinze and Ploner (2002), which is a modification of the algorithm initially proposed by Venzon and Moolgavkar (1988). Next, we describe the algorithm used in Sections 2.4 and 2.5. To do that, remind that $\boldsymbol{U}^*(\boldsymbol{\theta};\boldsymbol{z})$ is the modified score vector (or simply $\boldsymbol{U}^*(\boldsymbol{\theta})$), $\boldsymbol{H}_{\boldsymbol{\theta}}$ is the Hessian matrix of the penalized log-likelihood function $\ell_p(\boldsymbol{\theta})$ and $\ell_0 = \ell_p(\tilde{\boldsymbol{\theta}}) - \chi_1^2(\gamma)/2$, where $\tilde{\boldsymbol{\theta}}$ is the MPL estimator for $\boldsymbol{\theta}$ and let $\boldsymbol{e}_r$ be the r-th unit vector. Suppose the confidence limits for the parameter β_1 are our target, so we use the unit vector $\boldsymbol{e}_2$ in the following iterative scheme.

Step 1. With $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$, compute $\ell_p(\boldsymbol{\theta})$, $\boldsymbol{U}^*(\boldsymbol{\theta})$ and $\boldsymbol{H}_{\boldsymbol{\theta}}^{-1}$.

Step 2. Set $\kappa = 0$, compute

$$\lambda^{(\kappa)} = \pm \left\{ \frac{2[\ell_0 - \ell_p(\boldsymbol{\theta}) + 0.5\,\boldsymbol{U}^*(\boldsymbol{\theta})\boldsymbol{H}_{\theta}^{-1}\boldsymbol{U}^*(\boldsymbol{\theta})]}{\boldsymbol{e}_2^\top \boldsymbol{H}_{\boldsymbol{\theta}}^{-1}\boldsymbol{e}_2} \right\}^{1/2} \quad \text{and} \tag{2.16}$$

$$\boldsymbol{\delta}^{(\kappa)} = -\boldsymbol{H}_{\boldsymbol{\theta}}^{-1}[\boldsymbol{U}^*(\boldsymbol{\theta}) + \lambda^{(\kappa)}\boldsymbol{e}_2] \tag{2.17}$$

Step 3. Set $\kappa = 1$ and compute $\boldsymbol{\theta}^{(\kappa)} = \boldsymbol{\theta}^{(\kappa-1)} + \boldsymbol{\delta}^{(\kappa-1)}$

Step 4. With $\boldsymbol{\theta} = \boldsymbol{\theta}^{(\kappa)}$, compute $\ell_p(\boldsymbol{\theta})$, $\boldsymbol{U}^*(\boldsymbol{\theta})$ and $\boldsymbol{H}_{\boldsymbol{\theta}}^{-1}$. Thus, for sufficiently small $\varepsilon > 0$, to verify the following convergence criteria:

$$|\ell_p(\boldsymbol{\theta}) - \ell_0| < \varepsilon \text{ and } (\boldsymbol{U}^*(\boldsymbol{\theta}) + \lambda^{(\kappa)}\boldsymbol{e}_2)^\top \boldsymbol{H}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{U}^*(\boldsymbol{\theta}) + \lambda^{(\kappa)}\boldsymbol{e}_2) < \varepsilon.$$

If convergence is not yet attained, set κ to $\kappa+1$, and go to the Step 3, where $\lambda^{(\kappa)}$ and $\boldsymbol{\delta}^{(\kappa)}$ are updated alternately from the Equations (2.16) and (2.17).

Step 5. Return $\boldsymbol{e}_2^\top(\boldsymbol{\theta}^{(\kappa+1)})$ as one endpoint of the interval for β_1. For the lower confidence limit negative values of λ must be used, while for the upper confidence limit positive values of λ are used.

Computation of PPL confidence limits can be time-consuming, for the procedure has to be repeated for each endpoint of the interval of each of p parameters, resulting in $2p$ maximizations (HEINZE; PLONER, 2002). It would be the case where the practitioner has interest to compute PPL confidence intervals for all components of the vector $\boldsymbol{\theta}$.

2.4 Simulation study

In this section, we carry out two simulation studies to investigate and compare the performance of the ML and MPL estimators. We report only the results for our parameter of interest, β_1. The first simulation study consider two true values of the parameters corresponding to the model in (2.2) under the identifiability Cases 1–5 in Section 2.2.1. The second simulation study mimics the real dataset in Section 2.5.1, where σ_u^2 is known. For each true parameter setting, we simulated 10000 random samples for different sample sizes. The MPL estimates are computed numerically by using the BFGS method implemented in the R programing language (R CORE TEAM, 2017) in the `optim` function. In addition, to ensure the constraints for the different identifiability cases in Section 2.2.1, samples with estimates (in particular, of the variance parameters) violating the constraints are discarded until we obtain 10000 samples with admissible ML and MPL estimates. Moreover, samples leading to non-convergence of the iterative procedure used to compute PPL confidence intervals are also discarded. For example, for the Case 1 under the first scenario, at least 35412, 18060, 7307, 4166, 2386 samples were discarded, respectively, for sample size 20, 30, 50, 70 and 100, while for Case 2, they were 325, 29, 0, 0 and 0 rejected samples, respectively. For interval estimation, we use $\gamma = 0.95$.

2.4.1 First simulation study

In this study, we generate data from the model in (2.1) with sample size $n = 20$, 30, 50, 70 and 100 under the following scenarios. In the first scenario, we chose the true parameter values $(\beta_0, \beta_1, \mu_x, \sigma_x^2, \sigma_u^2, \sigma_\varepsilon^2)^\top = (0, -2, 4, 1, 1, 1)^\top$, while in the second one $(\beta_0, \beta_1, \mu_x, \sigma_x^2, \sigma_u^2, \sigma_\varepsilon^2)^\top = (0, -8, 16, 9, 1, 9)^\top$. Notice that the reliability ratio k_x for the first and second scenarios is 0.5 and 0.9, respectively, so that $k_x = 0.5$ can be considered a moderate reliability and $k_x = 0.9$ as high reliability. The true values of the parameters in our scenarios were chosen following Wang and Sivaganesan (2013).

In order to assess the point estimates obtained by the ML and MPL methods, Figure 1 displays the simulated bias of the estimates for β_1 and Tables 2 and 3 show the sampling standard deviation (SD), the root mean squared error (RMSE) and the average of the estimated asymptotic standard errors ($\hat{\sigma}_1(\hat{\beta}_1)$, $\tilde{\sigma}_2(\tilde{\beta}_1)$, $\tilde{\sigma}_3(\tilde{\beta}_1)$ and $\tilde{\sigma}_4(\tilde{\beta}_1)$) under the first and second scenarios, respectively.

In Figure 1, it can be observed that the bias decreases as the sample size increases for all cases, as expected. It can be also seen that in the first scenario ($k_x = 0.5$) and for Case 1, the simulated bias of the ML and MPL estimates for β_1 are quite similar. For Case 2, the biases of the ML estimates are smaller than those of the MPL estimates. For Case 5, there was a slight bias reduction by the MPL method when $n = 20$. Also, in the second scenario ($k_x = 0.9$) and for Cases 1 and 2, there was a slight reduction of bias by the MPL method in small and moderated samples and, for Case 5, it occurs when $n \in \{20, 30\}$. As expected, in most cases both of the estimation methods are quite similar when the sample size increases. The Cases 3 and 4 are not presented since the ML and MPL estimators for β_1 are the same (as proved in Propositions 1 and 2).

In Tables 2 and 3, it can be seen that the values of SD and RMSE decrease when the sample size increases, as expected. From the Table 2 and for Case 1 (known σ_u^2), since the bias of the estimates is not negligible for the majority of the entries (see, e.g., Figure 1–(a)), the values of SD and RMSE differ, especially when the sample size is small or moderate, whichever the estimation method. The presence of bias is not surprising for moderated reliability ratios ($k_x = 0.5$) and small samples. With respect to Case 2 (known σ_ε^2), the SD and RMSE values are reasonably close to the average of the estimated asymptotic standard errors for both estimation

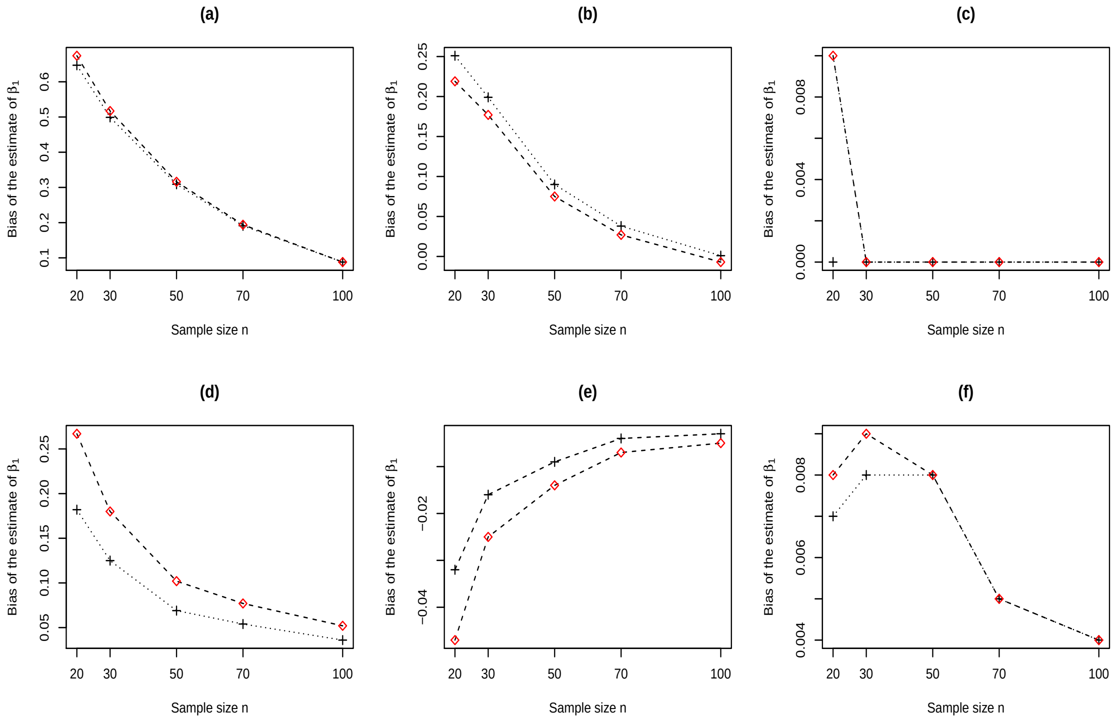

Figure 1 – Simulated bias of the estimates for β_1. First scenario: (a) Case 1, (b) Case 2 and (c) Case 5. Second scenario: (d) Case 1, (e) Case 2 and (f) Case 5. Estimation methods: ML ($\Diamond$) and MPL ($+$).

methods (in particular when the sample size is larger than 30). A similar performance to the previous case was observed under the identifiability Cases 3–5. Moreover, in the second scenario ($k_x = 0.9$) for Cases 1–5, Table 3 shows that the values of RMSE, SD and the average of the asymptotic standard errors ($\hat{\sigma}_1(\hat{\beta}_1)$, $\tilde{\sigma}_2(\tilde{\beta}_1)$, $\tilde{\sigma}_3(\tilde{\beta}_1)$ and $\tilde{\sigma}_4(\tilde{\beta}_1)$) do not differ so much, especially when the sample size is larger than 30.

Next, Tables 4 and 5 display the empirical coverage probability (CP) and the average interval length (IL) of the five approximate 95% confidence intervals for β_1 (described in Section 2.3) under the first and second scenarios, respectively. For the PPL confidence intervals, we compute the confidence limits using the algorithm described in Section 2.3.2. Also, the interpretation becomes easier if the focus is placed first on coverage probability and after that on the average interval length. Thus, the best results of the coverage probability are those whose values are quite close to the nominal value of 0.95 and, among those of the best approximate confidence intervals, we focus on those intervals that have the shortest average lengths.

In the sequel, some valuable conclusions are drawn from the Tables 4 and 5 in what follows:

Case 1: σ_u^2 is known. Overall, for the first scenario ($k_x = 0.5$), the best results seem to be achieved with the CI_5 interval, which presents much closer coverage probabilities to the nominal level than the other intervals, as shown in Table 4. In the second scenario ($k_x = 0.9$), we observe in

Table 2 – First scenario: $(\beta_0, \beta_1, \mu_x, \sigma_x^2, \sigma_u^2, \sigma_\varepsilon^2)^\top = (0, -2, 4, 1, 1, 1)^\top$ ($k_x = 0.5$). Sampling standard deviation (SD), root mean square error (RMSE) and average of the estimated asymptotic standard errors ($\hat{\sigma}_1(\hat{\beta}_1)$, $\tilde{\sigma}_2(\tilde{\beta}_1)$, $\tilde{\sigma}_3(\tilde{\beta}_1)$ and $\tilde{\sigma}_4(\tilde{\beta}_1)$).

Case	n	Method	SD	RMSE	$\hat{\sigma}_1(\hat{\beta}_1)$	$\tilde{\sigma}_2(\tilde{\beta}_1)$	$\tilde{\sigma}_3(\tilde{\beta}_1)$	$\tilde{\sigma}_4(\tilde{\beta}_1)$
	20	ML	0.281	0.730	0.453	-	-	-
		MPL	0.289	0.708	-	0.448	0.433	0.418
	30	ML	0.277	0.586	0.435	-	-	-
		MPL	0.282	0.573	-	0.431	0.420	0.410
1	50	ML	0.259	0.409	0.408	-	-	-
		MPL	0.259	0.403	-	0.404	0.397	0.390
	70	ML	0.261	0.325	0.388	-	-	-
		MPL	0.259	0.322	-	0.384	0.378	0.373
	100	ML	0.261	0.276	0.357	-	-	-
		MPL	0.257	0.272	-	0.353	0.349	0.345
	20	ML	0.377	0.437	0.422	-	-	-
		MPL	0.374	0.450	-	0.400	0.385	0.371
	30	ML	0.349	0.391	0.381	-	-	-
		MPL	0.346	0.399	-	0.366	0.357	0.349
2	50	ML	0.305	0.314	0.336	-	-	-
		MPL	0.304	0.317	-	0.327	0.322	0.317
	70	ML	0.284	0.285	0.301	-	-	-
		MPL	0.283	0.285	-	0.295	0.292	0.289
	100	ML	0.258	0.258	0.261	-	-	-
		MPL	0.258	0.258	-	0.258	0.256	0.254
	20	ML	0.379	0.442	0.382	-	-	-
		MPL	0.379	0.442	-	0.355	0.344	0.333
	30	ML	0.333	0.364	0.361	-	-	-
		MPL	0.333	0.364	-	0.342	0.335	0.328
3	50	ML	0.304	0.307	0.328	-	-	-
		MPL	0.304	0.307	-	0.316	0.313	0.309
	70	ML	0.288	0.288	0.296	-	-	-
		MPL	0.288	0.288	-	0.288	0.286	0.284
	100	ML	0.254	0.255	0.254	-	-	-
		MPL	0.254	0.255	-	0.249	0.248	0.246
	20	ML	0.524	0.566	0.591	-	-	-
		MPL	0.524	0.566	-	0.567	0.540	0.514
	30	ML	0.419	0.444	0.471	-	-	-
		MPL	0.419	0.444	-	0.458	0.443	0.428
4	50	ML	0.321	0.332	0.356	-	-	-
		MPL	0.321	0.332	-	0.350	0.343	0.336
	70	ML	0.273	0.278	0.299	-	-	-
		MPL	0.273	0.278	-	0.295	0.290	0.286
	100	ML	0.234	0.236	0.248	-	-	-
		MPL	0.234	0.236	-	0.246	0.243	0.241
	20	ML	0.124	0.124	0.122	-	-	-
		MPL	0.124	0.124	-	0.111	0.111	0.111
	30	ML	0.101	0.101	0.100	-	-	-
		MPL	0.101	0.101	-	0.094	0.094	0.094
5	50	ML	0.078	0.078	0.078	-	-	-
		MPL	0.078	0.078	-	0.075	0.075	0.075
	70	ML	0.068	0.068	0.066	-	-	-
		MPL	0.068	0.068	-	0.064	0.064	0.064
	100	ML	0.055	0.055	0.056	-	-	-
		MPL	0.055	0.055	-	0.055	0.055	0.055

Table 3 – Second scenario: $(\beta_0,\beta_1,\mu_x,\sigma_x^2,\sigma_u^2,\sigma_\varepsilon^2)^\top = (0,-8,16,9,1,9)^\top$ ($k_x = 0.9$). Sampling standard deviation (SD), root mean square error (RMSE) and average of the estimated asymptotic standard errors ($\hat{\sigma}_1(\hat{\beta}_1)$, $\tilde{\sigma}_2(\tilde{\beta}_1)$, $\tilde{\sigma}_3(\tilde{\beta}_1)$ and $\tilde{\sigma}_4(\tilde{\beta}_1)$).

Case	n	Method	SD	RMSE	$\hat{\sigma}_1(\hat{\beta}_1)$	$\tilde{\sigma}_2(\tilde{\beta}_1)$	$\tilde{\sigma}_3(\tilde{\beta}_1)$	$\tilde{\sigma}_4(\tilde{\beta}_1)$
	20	ML	0.668	0.719	0.823	-	-	-
		MPL	0.684	0.708	-	0.811	0.777	0.745
	30	ML	0.534	0.564	0.636	-	-	-
		MPL	0.543	0.557	-	0.629	0.611	0.594
1	50	ML	0.422	0.434	0.473	-	-	-
		MPL	0.426	0.432	-	0.470	0.462	0.454
	70	ML	0.354	0.362	0.392	-	-	-
		MPL	0.357	0.361	-	0.390	0.385	0.380
	100	ML	0.297	0.302	0.323	-	-	-
		MPL	0.299	0.301	-	0.322	0.319	0.316
	20	ML	0.720	0.721	0.656	-	-	-
		MPL	0.718	0.719	-	0.627	0.599	0.571
	30	ML	0.563	0.563	0.528	-	-	-
		MPL	0.562	0.562	-	0.512	0.496	0.481
2	50	ML	0.420	0.420	0.410	-	-	-
		MPL	0.420	0.420	-	0.402	0.394	0.386
	70	ML	0.354	0.355	0.344	-	-	-
		MPL	0.354	0.354	-	0.339	0.335	0.330
	100	ML	0.292	0.292	0.288	-	-	-
		MPL	0.292	0.292	-	0.285	0.283	0.280
	20	ML	0.711	0.714	0.658	-	-	-
		MPL	0.711	0.714	-	0.628	0.600	0.573
	30	ML	0.553	0.554	0.530	-	-	-
		MPL	0.553	0.554	-	0.513	0.497	0.482
3	50	ML	0.419	0.419	0.409	-	-	-
		MPL	0.419	0.419	-	0.400	0.393	0.385
	70	ML	0.355	0.355	0.344	-	-	-
		MPL	0.355	0.355	-	0.339	0.334	0.330
	100	ML	0.294	0.294	0.288	-	-	-
		MPL	0.294	0.294	-	0.285	0.282	0.279
	20	ML	0.727	0.741	0.739	-	-	-
		MPL	0.727	0.741	-	0.709	0.675	0.643
	30	ML	0.553	0.564	0.581	-	-	-
		MPL	0.553	0.564	-	0.565	0.546	0.528
4	50	ML	0.419	0.426	0.435	-	-	-
		MPL	0.419	0.426	-	0.428	0.419	0.410
	70	ML	0.350	0.355	0.362	-	-	-
		MPL	0.350	0.355	-	0.357	0.352	0.347
	100	ML	0.291	0.294	0.299	-	-	-
		MPL	0.291	0.294	-	0.296	0.293	0.290
	20	ML	0.119	0.119	0.117	-	-	-
		MPL	0.119	0.119	-	0.106	0.106	0.106
	30	ML	0.097	0.097	0.096	-	-	-
		MPL	0.097	0.097	-	0.090	0.090	0.090
5	50	ML	0.075	0.076	0.075	-	-	-
		MPL	0.075	0.076	-	0.072	0.072	0.072
	70	ML	0.063	0.063	0.063	-	-	-
		MPL	0.063	0.063	-	0.061	0.061	0.061
	100	ML	0.053	0.053	0.053	-	-	-
		MPL	0.053	0.053	-	0.052	0.052	0.052

Table 4 – First scenario: $(\beta_0, \beta_1, \mu_x, \sigma_x^2, \sigma_u^2, \sigma_\varepsilon^2)^\top = (0, -2, 4, 1, 1, 1)^\top$ $(k_x = 0.5)$. Empirical coverage probability (CP) and average interval length (IL) of the approximate 95% confidence intervals for β_1 under the identifiability Cases 1–5.

		CP					IL				
Case	n	CI_1	CI_2	CI_3	CI_4	CI_5	CI_1	CI_2	CI_3	CI_4	CI_5
1	20	0.700	0.712	0.690	0.668	0.885	1.777	1.755	1.696	1.640	2.113
	30	0.807	0.813	0.802	0.790	**0.922**	1.707	1.688	1.646	1.606	**1.989**
	50	0.894	0.896	0.892	0.887	**0.954**	1.600	1.583	1.554	1.528	**1.811**
	70	0.921	0.923	0.921	0.918	**0.963**	1.521	1.504	1.482	1.461	**1.688**
	100	0.943	0.944	0.943	0.942	**0.971**	1.400	1.385	1.369	1.354	**1.522**
2	20	0.856	0.824	0.812	0.800	**0.924**	1.655	1.567	1.510	1.455	**1.914**
	30	0.864	0.846	0.838	0.831	**0.925**	1.495	1.437	1.401	1.367	**1.687**
	50	0.910	0.899	0.895	0.891	**0.943**	1.315	1.282	1.263	1.244	**1.436**
	70	0.926	0.918	0.916	0.914	**0.947**	1.180	1.157	1.145	1.133	**1.263**
	100	0.937	0.932	0.931	0.929	**0.946**	1.025	1.011	1.003	0.996	**1.078**
3	20	0.841	0.816	0.807	0.797	0.898	1.499	1.392	1.348	1.306	1.622
	30	0.889	0.875	0.869	0.864	**0.932**	1.415	1.339	1.313	1.287	**1.529**
	50	0.921	0.912	0.911	0.907	**0.940**	1.285	1.239	1.226	1.213	**1.372**
	70	0.935	0.930	0.929	0.927	**0.944**	1.161	1.130	1.121	1.113	**1.224**
	100	**0.942**	0.938	0.937	0.936	**0.941**	**0.994**	0.975	0.970	0.965	1.033
4	20	**0.943**	**0.933**	0.921	0.908	**0.932**	**2.317**	2.222	2.115	2.014	**2.203**
	30	**0.951**	**0.947**	**0.940**	0.931	**0.946**	1.847	1.794	1.735	1.679	**1.785**
	50	**0.960**	**0.957**	**0.954**	**0.950**	**0.957**	1.396	1.371	1.343	**1.317**	1.368
	70	**0.964**	**0.962**	**0.960**	**0.956**	**0.962**	1.171	1.155	1.139	**1.123**	1.154
	100	**0.963**	**0.961**	**0.959**	**0.957**	**0.961**	0.972	0.963	0.953	**0.944**	0.962
5	20	**0.927**	0.902	0.902	0.902	**0.919**	0.477	0.434	0.434	0.434	**0.458**
	30	**0.939**	0.927	0.927	0.927	**0.932**	0.393	0.369	0.369	0.369	**0.383**
	50	**0.948**	0.938	0.938	0.938	**0.943**	0.307	0.295	0.295	0.295	**0.303**
	70	**0.940**	0.934	0.934	0.934	**0.937**	0.260	0.252	0.252	0.252	**0.257**
	100	**0.951**	0.947	0.947	0.947	**0.949**	0.218	0.214	0.214	0.214	**0.216**

Table 5 that the coverage probability presented by the five approximate confidence intervals (CI_1–CI_5) tend to give similar results and close to the nominal level. Furthermore, the CI_2–CI_5 intervals' lengths are slightly shorter than those of the CI_1 interval and the CI_3 interval stands out when the sample size is 20. The CI_4 interval can be recommended for having the shortest average lengths when the sample size is larger than 20.

Case 2: σ_ε^2 is known. In the first scenario ($k_x = 0.5$), it can be observed in Table 4 that the CI_5 interval has the best results in terms of coverage probability being these values much closer to the nominal level than in the other intervals. Also, the CI_5 interval's average lengths are slightly larger than those of the other intervals. For second scenario ($k_x = 0.9$) and when the sample size is less than 30, it is observed in Table 5 that the CI_1 interval presents much closer coverage probabilities the nominal level, but their average lengths were larger than those of the others intervals. Furthermore, when the sample size is greater than 30, in terms of coverage probability, the performance of the C_1, C_2 and CI_5 intervals is similar and close to the nominal level. Also, the average lengths of the CI_2 and CI_5 intervals are slightly shorter than the ones of the CI_1

Table 5 – Second scenario: $(\beta_0, \beta_1, \mu_x, \sigma_x^2, \sigma_u^2, \sigma_\varepsilon^2)^\top = (0, -8, 16, 9, 1, 9)^\top$ ($k_x = 0.9$). Empirical coverage probability (CP) and average interval length (IL) of the approximate 95% confidence intervals for β_1 under the identifiability Cases 1–5.

		CP					IL				
Case	n	CI_1	CI_2	CI_3	CI_4	CI_5	CI_1	CI_2	CI_3	CI_4	CI_5
	20	**0.950**	**0.954**	**0.946**	0.938	**0.959**	3.224	3.179	**3.047**	2.921	3.193
	30	**0.952**	**0.957**	**0.950**	**0.946**	**0.959**	2.493	2.466	2.396	**2.327**	2.469
1	50	**0.959**	**0.961**	**0.958**	**0.955**	**0.964**	1.856	1.843	1.810	**1.779**	1.843
	70	**0.958**	**0.960**	**0.958**	**0.955**	**0.960**	1.536	1.528	1.508	**1.489**	1.528
	100	**0.961**	**0.963**	**0.961**	**0.959**	**0.963**	1.267	1.263	1.251	**1.240**	1.263
	20	**0.920**	0.906	0.890	0.876	0.907	**2.573**	2.459	2.347	2.240	2.514
	30	**0.928**	0.919	0.909	0.900	0.917	**2.072**	2.007	1.945	1.884	2.036
2	50	**0.942**	**0.937**	0.932	0.926	**0.935**	1.605	**1.574**	1.544	1.515	1.588
	70	**0.939**	**0.935**	0.930	0.927	**0.936**	1.350	**1.331**	1.312	1.294	1.339
	100	**0.944**	**0.942**	0.940	0.938	**0.941**	1.130	**1.119**	1.108	1.097	1.124
	20	**0.923**	0.912	0.900	0.885	0.908	**2.579**	2.462	2.351	2.245	2.511
	30	**0.934**	**0.926**	0.917	0.908	**0.922**	2.077	**2.011**	1.949	1.889	2.037
3	50	**0.942**	**0.938**	0.933	0.929	**0.939**	1.601	**1.570**	1.540	1.511	1.582
	70	**0.938**	**0.934**	0.931	0.928	**0.937**	1.347	**1.328**	1.310	1.292	1.336
	100	**0.943**	**0.941**	0.939	0.936	**0.942**	1.128	**1.116**	1.106	1.095	1.121
	20	**0.941**	0.932	0.918	0.903	0.930	**2.896**	2.780	2.647	2.520	2.754
	30	**0.952**	**0.945**	0.937	0.929	**0.944**	2.278	**2.214**	2.141	2.071	2.202
4	50	**0.953**	**0.949**	**0.944**	**0.940**	**0.948**	1.706	1.676	1.642	**1.609**	1.671
	70	**0.956**	**0.953**	**0.950**	**0.946**	**0.952**	1.418	1.400	1.380	**1.360**	1.397
	100	**0.952**	**0.951**	**0.950**	**0.946**	**0.951**	1.171	1.160	1.148	**1.137**	1.159
	20	**0.932**	0.905	0.905	0.905	0.918	**0.458**	0.417	0.417	0.417	0.436
	30	**0.938**	0.924	0.924	0.924	**0.931**	0.376	0.353	0.353	0.353	**0.364**
5	50	**0.942**	0.932	0.932	0.932	**0.937**	0.293	0.281	0.281	0.281	**0.287**
	70	**0.948**	0.940	0.940	0.940	**0.944**	0.248	0.241	0.241	0.241	**0.244**
	100	**0.949**	0.945	0.945	0.945	**0.947**	0.208	0.204	0.204	0.204	**0.206**

interval, here the CI_2 interval stands out.

Case 3: Ratio of the error variances is known. In the first scenario ($k_x = 0.5$), the known value of $\lambda = \sigma_u^2/\sigma_\varepsilon^2$ is 1, for this situation, similar findings to Cases 1 and 2 are found in Table 4, i.e., the best results were achieved by the CI_5 interval with much closer coverage probabilities to the nominal level, in particular, when the sample size is greater than 20. Also, the average lengths of the CI_5 interval are slightly larger than those of the other intervals. In the second scenario ($k_x = 0.9$), the known value of λ is 3, for this situation, a similar performance is observed in Table 5 for the coverage for the coverage probability of the CI_1, CI_2 and CI_5 intervals when the sample size is greater than 20. Moreover, among these intervals, the CI_2 interval (followed by the CI_5 interval) outperformed the CI_1 interval, having the shortest average length.

Case 4: Reliability ratio is known. For the first scenario ($k_x = 0.5$), Table 4 shows that the CI_1, CI_2 and CI_5 intervals have a good performance in terms of coverage probability when the sample size is less than 30, where the CI_5 interval stands out with shorter average lengths than the other intervals. For sample sizes greater than 30, the CI_4, CI_3, CI_2 and CI_5 intervals, in this order, presented shorter average lengths than the CI_1 interval. Hence, the CI_4 interval can be recommended, mainly because its average lengths are shorter than those of the other intervals. In Table 5, under second scenario ($k_x = 0.9$), the CI_1 interval has a good coverage probability performance when the sample size is 20, followed by the CI_2 interval. And, if the sample size is greater than 20, the five approximate confidence intervals (CI_1–CI_5) present similar results of the coverage probability and are close to the nominal value. Furthermore, in terms of the average interval lengths, the CI_4 interval outperformed the other intervals by having the shortest average lengths. The CI_3 interval can be recommended as the second best.

Case 5: Intercept is null. In this case, it is observed a good performance of the coverage probability presented by the CI_1 and CI_5 intervals, which tend to give similar and close results to the nominal level regardless of the scenario (see the Tables 4 and 5). Also, the average lengths presented by the CI_1 and CI_5 intervals are not so different.

Lastly, for large sample size ($n = 100$) in all scenarios presented in this section and for most of entries in Tables 4 and 5, it can be observed a similar performance of the coverage probability for the five approximate confidence intervals (CI_1–CI_5), because the penalty effect introduced into the log-likelihood function became negligible. As expected, the increase of sample size reduces the average length of the five approximate confidence intervals.

2.4.2 Second simulation study

In this simulation study, the true values of the parameters are based on the ML estimates of the dataset analyzed in Section 2.5.1, i.e., $(\beta_0, \beta_1, \mu_x, \sigma_x^2, \sigma_u^2, \sigma_\varepsilon^2)^\top = (67, 0.43, 70.6, 220.1, 57, 38.4)^\top$ where σ_u^2 is considered known (Case 1). Observe that the reliability ratio is $k_x = 0.8$. Here, $n = 10$, 20, 30, 50 and 100.

In Figure 2, we can observe that the simulated bias of the MPL estimates presented a bias reduction (in module) when $n \in \{10, 20\}$. In addition, both the MPL and ML methods deliver quite similar estimates when the sample size is greater than 50.

Table 6 shows the results related to the performance measures for point estimates in this study. For the bias reduction provided by the MPL method when $n \in \{20, 30\}$, we noted that there was a slight inflation of the values of RMSE. For the remaining samples sizes, the two estimation methods do not differ too much with respect to SD and RMSE and in turn are quite similar to the average of $\hat{\sigma}_1(\hat{\beta}_1)$, $\tilde{\sigma}_2(\tilde{\beta}_1)$, $\tilde{\sigma}_3(\tilde{\beta}_1)$ and $\tilde{\sigma}_4(\tilde{\beta}_1)$), as desired.

Finally, Table 7 shows that the CI_1 and CI_5 intervals have the highest coverage probability values, especially, when the sample size is 10 and the CI_1 interval delivers slightly shorter average length than the CI_5 interval. As expected, the performance for all interval estimates is improved when the sample size increases. Also, in terms of average interval lengths, the CI_2–CI_4 intervals can be recommended when the sample size is large because their average lengths are slightly shorter than the other intervals.

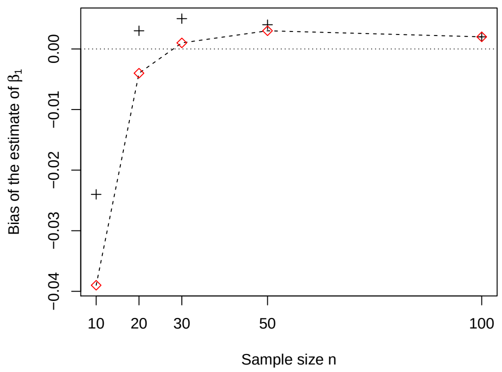

Figure 2 – Simulated bias of the estimates for β_1. Case 1 (known σ_u^2). Estimation methods: ML ($\Diamond$) and MPL ($+$).

Table 6 – True parameter values $(\beta_0, \beta_1, \mu_x, \sigma_x^2, \sigma_u^2, \sigma_\varepsilon^2)^\top = (67, 0.4, 7.6, 220.1, 57, 38.4)^\top$ ($k_x = 0.8$). Sampling standard deviation (SD), root mean square error (RMSE) and average of the estimated asymptotic standard errors ($\hat{\sigma}_1(\hat{\beta}_1)$, $\tilde{\sigma}_2(\tilde{\beta}_1)$, $\tilde{\sigma}_3(\tilde{\beta}_1)$ and $\tilde{\sigma}_4(\tilde{\beta}_1)$).

Case	n	Method	SD	RMSE	$\hat{\sigma}_1(\hat{\beta}_1)$	$\tilde{\sigma}_2(\tilde{\beta}_1)$	$\tilde{\sigma}_3(\tilde{\beta}_1)$	$\tilde{\sigma}_4(\tilde{\beta}_1)$
	10	ML	0.145	0.150	0.150	-	-	-
		MPL	0.151	0.153	-	0.145	0.135	0.126
	20	ML	0.120	0.120	0.121	-	-	-
		MPL	0.123	0.123	-	0.117	0.113	0.110
	30	ML	0.101	0.101	0.099	-	-	-
		MPL	0.102	0.102	-	0.097	0.095	0.094
	50	ML	0.078	0.078	0.077	-	-	-
		MPL	0.078	0.078	-	0.075	0.075	0.074
	100	ML	0.054	0.054	0.054	-	-	-
		MPL	0.053	0.053	-	0.053	0.053	0.053

Table 7 – True parameter values $(\beta_0, \beta_1, \mu_x, \sigma_x^2, \sigma_u^2, \sigma_\varepsilon^2)^\top = (67, 0.4, 7.6, 220.1, 57, 38.4)^\top$. Empirical coverage probability (CP) and average interval length (IL) of the approximate 95% confidence intervals for β_1 under the identifiability Case 1 (known σ_u^2).

	CP					IL				
n	CI_1	CI_2	CI_3	CI_4	CI_5	CI_1	CI_2	CI_3	CI_4	CI_5
10	**0.919**	0.911	0.889	0.864	**0.921**	0.588	0.569	0.528	0.492	0.607
20	**0.944**	**0.935**	0.926	0.913	**0.936**	0.473	0.459	0.445	0.432	0.489
30	**0.941**	**0.934**	0.930	0.925	**0.937**	0.389	0.379	0.372	0.367	0.401
50	**0.950**	**0.945**	**0.942**	0.938	**0.948**	0.302	0.295	0.293	0.292	0.312
100	**0.948**	**0.949**	**0.948**	**0.947**	**0.956**	0.213	0.209	0.209	0.209	0.224

2.5 Data analysis

In this section, in order to illustrate the proposed methodology, we consider two datasets previously analyzed in the literature. Some conclusions are presented as well.

2.5.1 Yields of corn and determinations of soil nitrogen

The dataset was taken from Fuller (1987, p. 18). The data refer to yields of corn (Y) and determinations of available soil nitrogen (X) collected at 11 sites on Marshal soil in Iowa (unknown units). Note that the estimates of soil nitrogen incorporate measurement error due to sampling error associated with the sample of soil as well as it often occurs in chemical analysis due to that many factors may affect a measurement (for example, instrument, observer, etc). Here, the error variance is $\sigma_u^2 = 57$ and is regarded as known. Also, the estimated reliability ratio k_x is 0.8.

In this case, ML and MPL estimates for β_1 are 0.433 and 0.435, respectively, which do not differ so much. The estimated asymptotic standard error of the ML estimate for β_1 is $\hat{\sigma}_1(\hat{\beta}_1) = 0.163$, while for the MPL estimate we have $\tilde{\sigma}_2(\tilde{\beta}_1) = 0.142$ and $\tilde{\sigma}_3(\tilde{\beta}_1) = \tilde{\sigma}_4(\tilde{\beta}_1) = 0.143$. Next, Table 8 displays the approximate 95% confidence intervals for β_1 and their interval lengths by using the five interval estimators described in Section 2.3. We can see that the CI_2, CI_3 and CI_4 estimates have, in this order, shorter interval lengths than the other intervals. However, the simulation study in Section 2.4.2 indicated that these interval estimates have a poor coverage probability performance under a similar scenario (Section 2.4.2 with $n = 10$). Thus, the CI_1 estimate gives a good interval estimate for β_1 followed by the CI_5 estimate.

2.5.2 Methods comparison study

The dataset was extracted from Strike (1991, pp. 318–323). The data are from an assay comparison study involving two methods for 56 measurements of serum gentamicin concentrations (in μmol/l), where X denotes the observed test results using an enzyme-mediated immunoassay (EMIT) and Y as the observed test results using a fluoro-immunoassay (FIA). Based on repeatability standard deviations, the ratio of the variances has been estimated to be $\lambda = 0.236$ and is considered as known. Moreover, the estimated reliability ratio is high ($k_x = 0.95$).

Here, ML and MPL estimates for β_1 are the same, i.e., $\hat{\beta}_1 = \tilde{\beta}_1 = 0.950$. The estimated asymptotic standard error of the ML estimate is $\hat{\sigma}_1(\hat{\beta}_1) = 0.034$, while for the MPL estimate we have $\tilde{\sigma}_2(\tilde{\beta}_1) = \tilde{\sigma}_3(\tilde{\beta}_1) = 0.033$, and $\tilde{\sigma}_4(\tilde{\beta}_1) = 0.032$. In interval estimation, although the CI_3 and CI_4 estimates have the shortest lengths, as shown in Table 8, they could not be considered, due to that in a scenario under similar conditions showed a bad coverage probability performance (see, e.g., Table 5 - Case 3 for $n \in \{50, 70\}$). Consequently, the CI_2 and CI_5 estimates can be recommended.

2.6 Conclusions

In this chapter, Firth's method was applied to measurement error models under the assumption of normality for the identifiability Cases 1–5. The penalty function used was the

Table 8 – Approximate 95% confidence intervals for β_1 and their lengths.

	Section 2.5.1 (known σ_u^2)		Section 2.5.2 (known λ)	
Estimator	Interval	Length	Interval	Length
CI_1	[0.113, 0.753]	0.640	[0.883, 1.017]	0.134
CI_2	[0.156, 0.714]	0.558	[0.885, 1.016]	0.131
CI_3	[0.155, 0.715]	0.560	[0.886, 1.014]	0.128
CI_4	[0.154, 0.716]	0.562	[0.887, 1.013]	0.126
CI_5	[0.153, 0.807]	0.654	[0.888, 1.019]	0.131

Jeffrey's prior in a Bayesian context. It was proved that for the model in (2.2) for the assumption of the ratio of error variances known (Case 3), the point estimators based on the ML and MPL methods for the parameters β_0, β_1 and μ_x are the same. This result also holds true when the reliability ratio k_x is known (Case 4). Now, within the scope of the scenarios in Section 2.4, we summarize some conclusions. For point estimation, the MPL method provided less biased and more precise estimates for the slope parameter β_1 than the ML method under the following conditions: High reliability ratio ($k_x = 0.9$) and, for identifiability Cases 1 and 2, small and moderated sample sizes. On the other hand, outside these conditions, both ML and MPL estimators for β_1 yield similar results.

With respect to interval estimation for β_1 and under our different scenarios, the results are summarized as follows:

(i) When σ_u^2 is known, the CI_5 estimator stands out in terms of coverage probability when the reliability ratio is moderate ($k_x = 0.5$) and the sample size is larger than 20. For $k_x = 0.9$ and sample size larger than 20, the CI_4 estimator outperforms the other intervals, having the shortest average length.

(ii) When σ_ε^2 is known, the CI_5 estimator outperforms the other interval estimators, especially in terms of coverage probability, when $k_x = 0.5$ and the sample size larger than 20. For $k_x = 0.9$ and the sample size larger than 30, the CI_1 estimator can be indicated followed by the CI_2 and CI_5 estimators.

(iii) When the ratio of error variances is known, if $k_x = 0.5$, the CI_5 estimator can be used in moderated sample sizes, its coverage probability is the closest to the nominal value. For $k_x = 0.9$ and sample size larger than 20, the CI_1 estimator can be recommended followed by the CI_2 and CI_5 estimators, in this order.

(iv) When the reliability ratio is known, when $k_x = 0.5$, the CI_5 estimator can be used followed by the CI_2 estimator in small or moderated sample sizes, while for sample sizes greater than 50, the CI_4 estimator can be indicated for having the shortest average length. For $k_x = 0.9$, the CI_1 estimator outperforms the other interval estimators, especially in terms of coverage probability when the sample size is 20 and 30. The CI_4 estimator outperforms the other intervals, having the shortest average length for sample sizes larger than 30.

(v) When the intercept is null, the CI_1 and CI_5 estimators can be used irrespective of the reliability ratio, because the performance in terms of the coverage probability and average interval length presented quite similar and close results.

Our findings led us to conclude that Firth's method can offer an improvement over some asymptotic properties of the ML estimators. As mentioned earlier, this improvement, in terms of bias reduction, mean square error and coverage probability, does not occur simultaneously as was observed in our simulation studies applied to MEM. Moreover, it is important to note that bias reduction in a particular model is not always achieved. As commented by Firth (1993), the merits of this method will depend on some factors, for instance, the choice of the parameter of interest and the skewness of the ML estimates, among others.

On the other hand, estimation methods based on generalized fiducial inference have been recently investigated and reviewed in Hannig *et al.* (2016) and, recent published papers of several authors have showed an improvement in interval estimation, showing good frequentist properties. Thus, some works related to these estimation methods involving the model in (2.2) under the identifiability Cases 1, 3 and 4 have been studied, respectively, in Yan and Xu (2017), Yan, Wang and Xu (2017b) and Yan, Wang and Xu (2017a). The new fiducial intervals for the slope parameter proposed by these authors have showed a good coverage probability performance and shorter average lengths with respect to that of the frequentist approach, especially, in small samples. For this reason, in the next chapter, we will construct two estimation methods based on the generalized fiducial inference to a special case of measurement error models, which is known as the Grubbs model.

CHAPTER

3

GENERALIZED FIDUCIAL INFERENCE FOR THE GRUBBS MODEL

In this chapter, we develop generalized fiducial inference for the precision of each measuring instrument and the variability product without available replications on the observations under the Grubbs model considering the two-instrument case. Thus, we briefly review two existing estimation approaches for the Grubbs model and then we construct the two new estimation methods built on generalized fiducial inference. We compare these two new methods with existing approaches through a simulation study and illustrate with a dataset from a methods comparison study. Finally, some remarks are also presented. A modification of this chapter allowed us to obtain as result the following paper, see Tomaya and Castro (2018b).

3.1 Introduction

The Grubbs model is often used in methods comparison studies to assess the relative agreement between two or more analytical instruments (or methods) designed to measure the same quantity of interest (GRUBBS, 1983). The main aim is to see whether or not two or more methods produce measurement errors of similar magnitude.

The Grubbs model is a particular case of the MEM given in (2.1) with the specification of $\beta_1 = 1$. Next, the model formulation is described as follows. Consider at our disposal two instruments for measuring a quantity of interest x and a set of readings of experimental units. Thus, let x_i be the unobservable measurement, while X_i and Y_i are the measured values obtained by the instruments A' and B', respectively, in the i-th unit. Relating these variables, we have

$$Y_i = \beta_0 + x_i + \varepsilon_i \quad \text{and} \quad X_i = x_i + u_i, \quad i = 1, \dots, n, \tag{3.1}$$

where β_0 is the bias or systematic error of the instrument B' with respect to the instrument A'. We assume in (3.1) that x_i are independent and distributed as the $\mathcal{N}(\mu_x, \sigma_x^2)$ distribution and

ε_i and u_i are independent random errors following the $\mathcal{N}(0,\sigma_\varepsilon^2)$ and $\mathcal{N}(0,\sigma_u^2)$ distributions, respectively. Moreover, x_i, ε_j and u_k are independent for all $i,j,k=1,\ldots,n$. Hence, under these assumptions, the observations $(X_i,Y_i)^\top$ are independent and distributed as

$$(X_i,Y_i)^\top \sim \mathcal{N}_2(\boldsymbol{\mu},\boldsymbol{\Sigma}), \tag{3.2}$$

with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ given by

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_x \\ \beta_0 + \mu_x \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix} = \begin{bmatrix} \sigma_x^2 + \sigma_u^2 & \sigma_x^2 \\ \sigma_x^2 & \sigma_x^2 + \sigma_\varepsilon^2 \end{bmatrix}, \tag{3.3}$$

where $\beta_0, \mu_x, \sigma_x^2, \sigma_u^2$ and σ_ε^2 are unknown parameters.

For the two-instrument case, Bayesian inference for some parameters of interest was studied in Draper and Guttman (1975) as well as estimation for the systematic error β_0 was investigated in Castro and Vidal (2017). Inference on the precision parameters (σ_u^2 and σ_ε^2) and product variability (σ_x^2) under such model was studied in Grubbs (1948), Grubbs (1983), among others. Later on, Maloney and Rastogi (1970), Shukla (1973) and Jaech (1985) proposed different approaches for testing hypothesis on these parameters. Recently, a new two-instrument estimator for the precision of a measuring instrument was provided by Lombard and Potgieter (2012). This estimator improves upon the Grubbs estimator only if the ratio σ_x^2/σ_u^2 is previously specified.

On the other hand, a methodology for constructing hypotheses tests based on generalized statistical inference was introduced by Tsui and Weerahandi (1989), while Weerahandi (1993) proposed a generalized pivotal quantity (GPQ) with its corresponding generalized confidence interval. A direct connection between generalized confidence intervals and fiducial intervals was established in Hannig, Iyer and Patterson (2006). They proved the asymptotic frequency correctness of such intervals. They also recognized a subclass of the GPQ, termed as fiducial generalized pivotal quantity (FGPQ). In a similar way, Xu and Li (2006) and Li, Xu and Li (2007) also found a connection between the Fisher (1930)'s fiducial inference and the generalized statistical inference and showed a general method to derive the fiducial distribution and the generalized test variable, respectively. This last result was called by them as the generalized fiducial distribution (GFD). In the past 10 years, a unification of the main results in parametric problems based on the generalized fiducial inference was investigated and reviewed in Hannig (2009b) and Hannig *et al.* (2016). Recent works have showed an improvement in the interval estimation when the generalized fiducial inference is adopted. For example, in measurement error models, Wang and Iyer (2008) provided procedures for constructing uncertainty regions for the intercept and slope parameters using a FGPQ. Moreover, Yan, Wang and Xu (2017a), Yan, Wang and Xu (2017b) and Yan and Xu (2017) proposed two confidence intervals based on FGPQ and GFD for the slope parameter with good empirical frequentist properties.

Motived by the recent interest in fiducial methods, we construct two new estimation methods for the precision parameters and product variability under the Grubbs model. The methods are based on generalized fiducial inference. Furthermore, we compare them with two other existing approaches, which we call Grubbs estimation method (G) (GRUBBS, 1948; JAECH, 1985) and Thompson's estimation method (T) (THOMPSON, 1963).

3.2 Estimation methods

In this section, we describe two existing estimation approaches for the parameters of interest under the Grubbs model and after that other two new estimation methods based on generalized fiducial inference are built as well.

3.2.1 Existing approaches

Next, point and interval estimation, for the precision parameters and the product variability, computed by the Grubbs method and the Thompson's method are described as follows.

3.2.1.1 Grubbs estimators and interval estimation

Based on the method of moments, (GRUBBS, 1948) provided the expressions for the point estimators of the product variability and the precision parameters, given by

$$\widetilde{\sigma}_x^2 = S_{XY}, \quad \widetilde{\sigma}_u^2 = S_X^2 - S_{XY} \quad \text{and} \quad \widetilde{\sigma}_\varepsilon^2 = S_Y^2 - S_{XY}, \tag{3.4}$$

where $S_X^2 = \sum_{i=1}^n (X_i - \overline{X})^2/(n-1)$, $S_Y^2 = \sum_{i=1}^n (Y_i - \overline{Y})^2/(n-1)$, $S_{XY} = \sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})/(n-1)$, $\overline{X} = \sum_{i=1}^n X_i/n$ and $\overline{Y} = \sum_{i=1}^n Y_i/n$. Note that negative estimates can be obtained from (3.4). In addition, the expressions for the variances of $\widetilde{\sigma}_x^2$, $\widetilde{\sigma}_u^2$ and $\widetilde{\sigma}_\varepsilon^2$ in (3.4) are

$$\text{Var}(\widetilde{\sigma}_x^2) = \frac{2\sigma_x^4 + D}{n-1}, \quad \text{Var}(\widetilde{\sigma}_u^2) = \frac{2\sigma_u^4 + D}{n-1} \quad \text{and} \quad \text{Var}(\widetilde{\sigma}_\varepsilon^2) = \frac{2\sigma_\varepsilon^4 + D}{n-1}, \tag{3.5}$$

where $D = \sigma_x^2\sigma_u^2 + \sigma_x^2\sigma_\varepsilon^2 + \sigma_u^2\sigma_\varepsilon^2$.

To estimate the variances in (3.5), we can replace the unknown parameters involved by their estimates. Although (3.5) gives the sampling variances of the estimators in (3.4), it does not indicate their distributions. Then, under certain conditions (JAECH, 1985), $\widetilde{\sigma}_x^2$, $\widetilde{\sigma}_u^2$ and $\widetilde{\sigma}_\varepsilon^2$ are approximately distributed as chi-square random variables with ν_x, ν_u and ν_ε degrees of freedom, respectively. For interval estimation of σ_u^2, first we compute $\nu_u = 2\sigma_u^4/\text{Var}(\widetilde{\sigma}_u^2)$ using the estimates of σ_u^4 and $\text{Var}(\widetilde{\sigma}_u^2)$ from (3.4) and (3.5). In a similar way, ν_x and ν_ε are computed. Then, with $\gamma \in (0,1)$, an approximate two-sided $100\gamma\%$ confidence interval for σ_u^2 has lower and upper limits given by $\nu_u\widetilde{\sigma}_u^2/\chi^2_{\nu_u}((1+\gamma)/2)$ and $\nu_u\widetilde{\sigma}_u^2/\chi^2_{\nu_u}((1-\gamma)/2))$, respectively, where $\chi^2_\nu(\delta)$ represents the $100\delta\%$ percentile of the χ^2_ν distribution. Similar steps are taken to construct the approximate two-sided $100\gamma\%$ confidence intervals for σ_x^2 and σ_ε^2.

3.2.1.2 Thompson's estimators and simultaneous confidence region

Using the restricted maximum likelihood method, (THOMPSON, 1963) proposed the point estimators for the precision parameters and product variability displayed in Table 9. These estimators take only nonnegative values.

Table 9 – Thompson's point estimators.

	Parameter		
Conditions	σ_x^2	σ_u^2	σ_ε^2
$S_X^2, S_Y^2 \geq S_{XY} \geq 0$	S_{XY}	$S_X^2 - S_{XY}$	$S_Y^2 - S_{XY}$
$S_Y^2 \geq S_{XY} \geq S_X^2$	S_X^2	0	$S_X^2 + S_Y^2 - 2S_{XY}$
$S_X^2 \geq S_{XY} \geq S_Y^2$	S_Y^2	$S_X^2 + S_Y^2 - 2S_{XY}$	0
$S_{XY} < 0$	0	S_X^2	S_Y^2

Furthermore, Thompson (1963) also derived an exact method to construct a simultaneous confidence region for the three parameters (σ_x^2, σ_u^2 and σ_ε^2). The inequalities

$$\begin{aligned} |\sigma_x^2 - C_{XY}K| &\leq MC_XC_Y, \\ |\sigma_u^2 - (C_X^2 - C_{XY})K| &\leq MC_X(C_X^2 + C_Y^2 - 2C_{XY})^{1/2} \\ \text{and} \quad |\sigma_\varepsilon^2 - (C_Y^2 - C_{XY})K| &\leq MC_Y(C_X^2 + C_Y^2 - 2C_{XY})^{1/2} \end{aligned}$$

hold true simultaneously with probability at the least γ, where $C_X^2 = (n-1)S_X^2$, $C_Y^2 = (n-1)S_Y^2$ and $C_{XY} = (n-1)S_{XY}$. The values of K and M are given in (THOMPSON, 1963) for γ equal to 0.99 and 0.95. For these simultaneous intervals, all negative lower limits are replaced by 0.

3.2.2 Generalized fiducial inference

The idea of generalized fiducial inference (GFI) is motived as a unification of the theory of the generalized confidence intervals proposed by Weerahandi (1993) and on the surrogate variable method for obtaining confidence intervals for the variance components in balanced mixed linear models studied by Chiang (2001). From these new procedures, Hannig (2009b) found that there is a direct connection with the fiducial inference introduced by Fisher (1930). Thus, through a series of works, this unification has been found to have excellent theoretical and empirical properties for several practical applications, under fairly general conditions, see, e.g., the papers of Hannig, Iyer and Patterson (2006), Wang and Iyer (2008), Hannig (2009a), Hannig (2009b), Hannig (2013), Hannig *et al.* (2016), among others.

For a better understanding on GFI, Hannig *et al.* (2016) defines it as a data dependent measure on the parameter space by carefully using an inverse of the deterministic data generating equation without the use of Bayes theorem. The resulting generalized fiducial distribution (GFD) is a data dependent distribution on the parameter space and can be seen, rather than to a point and interval estimator, as a distribution estimator of the unknown parameter of interest. Moreover, the GFD can be treated mathematically in a similar way to the Bayesian posteriors to define approximate confidence intervals (sets). All the aforementioned papers in the literature presented that the GFD has showed desirable empirical properties, e.g., conservative coverages or (and) shorter or comparable expected lengths than others competing approaches (for example, the frequentist and Bayesian procedures). Some discussions of philosophical controversies on the development of GFI are discussed in Hannig *et al.* (2016).

On the basis of generalized fiducial inference, we describe two new estimation methods. The first method is based on the generalized fiducial pivotal quantity studied in Hannig, Iyer and Patterson (2006) with an extension presented in Hannig (2009b). The second method is built using the generalized fiducial distribution developed in Hannig *et al.* (2016).

3.2.2.1 Estimation by using a fiducial generalized pivotal quantity

We briefly review the concept of FGPQ and refer readers to Hannig, Iyer and Patterson (2006) for more details. Let $\boldsymbol{S} \in \mathbb{R}^k$ be an observable random vector with distribution indexed by a parameter $\boldsymbol{\eta} \in \mathbb{R}^r$, and $\boldsymbol{S}^*$ represents an independent copy of $\boldsymbol{S}$. Let $\boldsymbol{s}$ and $\boldsymbol{s}^*$ represent the observations of $\boldsymbol{S}$ and $\boldsymbol{S}^*$, respectively. Suppose our parameter of interest is $\boldsymbol{\xi} = \boldsymbol{\pi}(\boldsymbol{\eta}) \in \mathbb{R}^q$. $\boldsymbol{R_\xi}(\boldsymbol{S}, \boldsymbol{S}^*, \boldsymbol{\eta})$ (or simply $\boldsymbol{R_\xi}$) is called a FGPQ for a parameter $\boldsymbol{\xi}$, if $\boldsymbol{R_\xi}(\boldsymbol{S}, \boldsymbol{S}^*, \boldsymbol{\eta})$ satisfies the following two conditions:

(i) The distribution of $\boldsymbol{R_\xi}(\boldsymbol{S}, \boldsymbol{S}^*, \boldsymbol{\eta})|\boldsymbol{S} = \boldsymbol{s}$ is free of $\boldsymbol{\eta}$

(ii) For every allowable $\boldsymbol{s} \in \mathbb{R}^k$, $\boldsymbol{R_\xi}(\boldsymbol{s}, \boldsymbol{s}, \boldsymbol{\eta}) = \boldsymbol{\xi}$.

In the sequel, we describe the so-called structural method (HANNIG; IYER; PATTERSON, 2006) to construct a FGPQ for $\boldsymbol{\xi}$. Suppose there exits mappings $f_1, \ldots, f_r$, with $f_i : \mathbb{R}^k \times \mathbb{R}^r \to \mathbb{R}$, such that if $E_i = f_i(\boldsymbol{S}, \boldsymbol{\eta})$, for $i = 1, \ldots, r$, then $\boldsymbol{E} = (E_1, \ldots, E_r)^\top$ has a joint distribution that is free of $\boldsymbol{\eta}$, we say that $\boldsymbol{f}(\boldsymbol{S}, \boldsymbol{\eta})$ is a pivotal quantity for $\boldsymbol{\eta}$, where $\boldsymbol{f} = (f_1, \ldots, f_r)^\top$. Suppose that the mapping $\boldsymbol{f}(\boldsymbol{s}, \cdot)$ is invertible for every $\boldsymbol{s}$. We then say that $\boldsymbol{f}(\boldsymbol{S}, \boldsymbol{\eta})$ is an invertible pivotal quantity for $\boldsymbol{\eta}$. Let $\boldsymbol{g}(\boldsymbol{s}, \cdot) = (g_1(\boldsymbol{s}, \cdot), \ldots, g_r(\boldsymbol{s}, \cdot))^\top$ be the inverse mapping, so that whenever $\boldsymbol{e} = \boldsymbol{f}(\boldsymbol{s}, \boldsymbol{\eta})$, we have $\boldsymbol{g}(\boldsymbol{s}, \boldsymbol{e}) = \boldsymbol{\eta}$. Thus, define

$$\boldsymbol{R_\xi} = \boldsymbol{R_\xi}(\boldsymbol{S}, \boldsymbol{S}^*, \boldsymbol{\eta}) = \boldsymbol{\pi}(g_1(\boldsymbol{S}, \boldsymbol{f}(\boldsymbol{S}^*, \boldsymbol{\eta})), \ldots, g_r(\boldsymbol{S}, \boldsymbol{f}(\boldsymbol{S}^*, \boldsymbol{\eta}))) = \boldsymbol{\pi}(g_1(\boldsymbol{S}, \boldsymbol{E}^*), \ldots, g_r(\boldsymbol{S}, \boldsymbol{E}^*)),$$

where $\boldsymbol{E}^* = \boldsymbol{f}(\boldsymbol{S}^*, \boldsymbol{\eta})$ is an independent copy of $\boldsymbol{E}$, and then $\boldsymbol{R_\xi}$ is a FGPQ for $\boldsymbol{\xi}$. When $\boldsymbol{\xi}$ is a scalar parameter, a point estimate may be obtained by the mean or the median of the distribution of $\boldsymbol{R_\xi}$ conditional on $\boldsymbol{S} = \boldsymbol{s}$. Hence, with $\gamma \in (0, 1)$, an equal-tailed two-sided $100\gamma\%$ fiducial generalized confidence interval for ξ is given by $[\hat{\xi}_{(1-\gamma)/2}, \hat{\xi}_{(1+\gamma)/2}]$, where $\hat{\xi}_\delta$ denotes the $100\delta\%$ percentile of the distribution of $\boldsymbol{R_\xi}$ conditional on $\boldsymbol{S} = \boldsymbol{s}$.
Following the method described above, we construct a FGPQ for the product variability and the

precision parameters, i.e., $\boldsymbol{\xi} = (\sigma_x^2, \sigma_u^2, \sigma_\varepsilon^2)^\top$. Let $\boldsymbol{T}$ be the sample covariance matrix, that is,

$$\boldsymbol{T} = \begin{bmatrix} S_X^2 & S_{XY} \\ S_{XY} & S_Y^2 \end{bmatrix}.$$

Hence, from (3.2), it follows that $(n-1)\boldsymbol{T} \sim \mathscr{W}(n-1, \boldsymbol{\Sigma})$, where $\mathscr{W}(n-1, \boldsymbol{\Sigma})$ denotes the Wishart distribution with $n-1$ degrees of freedom and mean matrix $(n-1)\boldsymbol{\Sigma}$. With $\boldsymbol{V}$ denoting the upper triangular positive definite matrix such that $\boldsymbol{\Sigma} = \boldsymbol{V}\boldsymbol{V}^\top$, let the structural equation be $(n-1)\boldsymbol{T} = \boldsymbol{V}\boldsymbol{W}\boldsymbol{V}^\top$, where

$$\boldsymbol{V} = \begin{bmatrix} (\sigma_X^2\sigma_Y^2 - \sigma_{XY}^2)^{1/2}/\sigma_Y & \sigma_{XY}/\sigma_Y \\ 0 & \sigma_Y \end{bmatrix} \quad \text{and} \quad \boldsymbol{W} = \begin{bmatrix} W_{11} & W_{12} \\ W_{12} & W_{22} \end{bmatrix} \sim \mathscr{W}(n-1, \boldsymbol{I}_2), \tag{3.6}$$

with $\boldsymbol{I}_r$ denoting the $r \times r$ unit matrix.

Let $\boldsymbol{\eta} = (\sigma_X^2, \sigma_{XY}, \sigma_Y^2)^\top$, $\boldsymbol{S} = (S_X^2, S_{XY}, S_Y^2)^\top$ and $\boldsymbol{E} = (W_{11}, W_{12}, W_{22})^\top$. By the above structural equation, it follows that $\boldsymbol{W} = (n-1)\boldsymbol{V}^{-1}\boldsymbol{T}(\boldsymbol{V}^{-1})^\top$, so this yields the components of $\boldsymbol{E} = \boldsymbol{f}(\boldsymbol{S}, \boldsymbol{\eta})$, which are given by

$$\begin{aligned}
W_{11} &= \frac{\sigma_Y^2(n-1)S_X^2}{\sigma_X^2\sigma_Y^2 - \sigma_{XY}^2} - 2\frac{\sigma_{XY}(n-1)S_{XY}}{\sigma_X^2\sigma_Y^2 - \sigma_{XY}^2} + \frac{\sigma_{XY}^2(n-1)S_Y^2}{\sigma_Y^2(\sigma_X^2\sigma_Y^2 - \sigma_{XY}^2)}, \\
W_{12} &= \frac{(n-1)S_X^2}{\sqrt{\sigma_X^2\sigma_Y^2 - \sigma_{XY}^2}} - \frac{\sigma_{XY}(n-1)S_Y^2}{\sigma_Y^2\sqrt{\sigma_X^2\sigma_Y^2 - \sigma_{XY}^2)}} \quad \text{and} \\
W_{22} &= \frac{(n-1)S_Y^2}{\sigma_Y^2}.
\end{aligned}$$

Consequently, using the inverse mapping of $\boldsymbol{E}$, the components of $\boldsymbol{\eta} = \boldsymbol{g}(\boldsymbol{S}, \boldsymbol{E})$ are as follows

$$\begin{aligned}
\sigma_X^2 &= \frac{n-1}{W_{22}}\left[\frac{W_{22}^2(S_X^2S_Y^2 - S_{XY}^2)}{S_{YY}(W_{11}W_{22} - W_{12}^2)} + \frac{1}{S_Y^2}\left(S_{XY} - \frac{\sqrt{S_X^2S_Y^2 - S_{XY}^2}}{\sqrt{W_{11}W_{22} - W_{12}^2}}W_{12}\right)^2\right], \\
\sigma_{XY} &= \frac{n-1}{W_{22}}\left(S_{XY} - \frac{\sqrt{S_X^2S_Y^2 - S_{XY}^2}}{\sqrt{W_{11}W_{22} - W_{12}^2}}W_{12}\right) \quad \text{and} \\
\sigma_Y^2 &= \frac{(n-1)S_Y^2}{W_{22}}.
\end{aligned} \tag{3.7}$$

To compute the components of $\boldsymbol{E}$ and $\boldsymbol{\eta}$ we follow the steps provided in Yan, Wang and Xu (2017b). Now, due to (3.3), we have $\boldsymbol{\xi} = \boldsymbol{\pi}(\boldsymbol{\eta}) = (\sigma_{XY}, \sigma_X^2 - \sigma_{XY}, \sigma_Y^2 - \sigma_{XY})^\top$ and applying the structural method, a FGPQ for $\boldsymbol{\xi}$ is given by

$$\boldsymbol{R}_\xi = (\sigma_{XY}^*, \sigma_X^{2*} - \sigma_{XY}^*, \sigma_Y^{2*} - \sigma_{XY}^*)^\top, \tag{3.8}$$

where σ_X^{2*}, σ_Y^{2*} and σ_{XY}^* are, respectively, σ_X^2, σ_Y^2 and σ_{XY} given in (3.7) replacing $\boldsymbol{E}$ for $\boldsymbol{E}^*$, where $\boldsymbol{E}^* = (W_{11}^*, W_{12}^*, W_{22}^*)^\top$ is an independent copy of $\boldsymbol{E} = (W_{11}, W_{12}, W_{22})^\top$. The quantities σ_X^{2*}, σ_Y^{2*} and σ_{XY}^* are free of parameters.

Furthermore, for each component of $\boldsymbol{R_\xi}$, two other FGPQ's may be obtained from (3.8). These are $\max(0, R_{\xi_j})$ and $|R_{\xi_j}|$, where R_{ξ_j} denotes the j-th component of $\boldsymbol{R_\xi}$, $j \in \{1,2,3\}$, so that these quantities take only nonnegative values, as pointed out in Hannig, Iyer and Patterson (2006). These modified FGPQ's are asymptotically equivalent to R_{ξ_j}.

Therefore, a point estimate for each component of $\boldsymbol{\xi}$ may be obtained by the mean or the median of the distribution of $\boldsymbol{R_\xi}$ conditional on $\boldsymbol{S} = \boldsymbol{s}$. An equal-tailed two-sided $100\gamma\%$ fiducial generalized confidence interval for ξ_j is given by $[\hat{\xi}_{j,(1-\gamma)/2}, \hat{\xi}_{j,(1+\gamma)/2}]$, where $\hat{\xi}_{j,\delta}$ denotes the $100\delta\%$ percentile of the distribution of R_{ξ_j} (or the two other modified FGPQ's) conditional on $\boldsymbol{S} = \boldsymbol{s}$, $j \in \{1,2,3\}$.

Notice the distribution of the FGPQ in (3.8) may not be obtained in closed form. Samples can be generated by using a Monte Carlo technique and from them we can obtain point and interval estimates. One can compute them through the following scheme:

Step 1. Given a realization $\boldsymbol{s}$ of $\boldsymbol{S} = (S_X^2, S_{XY}, S_Y^2)^\top$, which is computed using the observations in (3.2). Select a large number, say $N = 10000$. For $m = 1, \ldots, N$, carry out the following steps:

Step 2. For $m = 1$, generate a realization $\boldsymbol{e}_m^*$ of $\boldsymbol{E}_m^*$ from (3.6).

Step 3. Calculate $\boldsymbol{R}_{\boldsymbol{\xi},m}$ using the expression in (3.8) by replacing $\boldsymbol{E}_m^*$ and $\boldsymbol{S}$ by $\boldsymbol{e}_m^*$ and $\boldsymbol{s}$, respectively. Repeat steps 2–3 until $m = N$.

Step 4. For $j \in \{1,2,3\}$, put $R_{\xi_{j,1}}, \ldots, R_{\xi_{j,N}}$ in ascending order. Choose the median as a point estimate as well as the $N(1-\gamma)/2$-th and $N(1+\gamma)/2$-th elements of the ordered values as the limits of the equal-tailed two-sided confidence interval.

3.2.2.2 Estimation by using the generalized fiducial distribution

Next, we briefly review the definition and construction about the GFD and refer to readers to Hannig (2009a), Hannig (2009b) and Hannig *et al.* (2016), for more details, and then we apply it for our model in (3.2). Let $\boldsymbol{Z} \in \mathbb{R}^{2n}$ denote the random vector of the observations. Let the data generating equation be

$$\boldsymbol{Z} = \boldsymbol{G}(\boldsymbol{\theta}, \boldsymbol{U}), \tag{3.9}$$

where $\boldsymbol{G}(\cdot,\cdot)$ is a deterministic function, $\boldsymbol{U}$ is a completely known random vector and $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$ is a parameter vector. The distribution of $\boldsymbol{Z}$ is determined by $\boldsymbol{U}$ and for any fixed value of the parameter $\boldsymbol{\theta}$. After observing $\boldsymbol{Z}$ we can use to infer a distribution on $\boldsymbol{\theta}$. This allows us to define a probability measure on the parameter space $\boldsymbol{\Theta}$. If the generating equation in (3.9) can be inverted then the inverse will be denoted as $\boldsymbol{G}^{-1}(\cdot,\cdot)$. Thus, for an observed $\boldsymbol{z}$ and $\boldsymbol{u}$ it is possible to calculate $\boldsymbol{\theta}$ from $\boldsymbol{\theta} = \boldsymbol{G}^{-1}(\boldsymbol{z}, \boldsymbol{u})$. Through this inverse relationship and a random sample of $\boldsymbol{u}_1', \ldots, \boldsymbol{u}_Q'$ we can compute a random sample for $\boldsymbol{\theta}$ so that $\boldsymbol{\theta}_1' = \boldsymbol{G}^{-1}(\boldsymbol{z}, \boldsymbol{u}_1'), \ldots, \boldsymbol{\theta}_Q' = \boldsymbol{G}^{-1}(\boldsymbol{z}, \boldsymbol{u}_Q')$.

This sample is termed a fiducial sample and can be used to calculate both point and confidence interval estimators for the true parameter $\boldsymbol{\theta}_0$.

Furthermore, as pointed out in Hannig (2009b), the inverse generating equation $\boldsymbol{G}^{-1}(\cdot,\cdot)$ may not exist due to two reasons: (i) There is no $\boldsymbol{\theta}$ that satisfies $\boldsymbol{z}=\boldsymbol{G}(\boldsymbol{\theta},\boldsymbol{u})$ or (ii) There is more than one $\boldsymbol{\theta}$ for some value of $\boldsymbol{u}$ and $\boldsymbol{z}$ that satisfies (3.9). To overcome these issues, Hannig (2009b) recommended the following alternatives: To deal with (i), he suggests to eliminate the values of $\boldsymbol{u}$ for which there is no solution from the sample space and then re-normalizing the probabilities. That is reasonable because the data was generated using $\boldsymbol{z}_0$ and $\boldsymbol{u}_0$ so at least one solution for $\boldsymbol{\theta}=\boldsymbol{G}^{-1}(\boldsymbol{z},\boldsymbol{u})$ exists. Therefore, $\boldsymbol{u}$'s considered are those that allow to the function $\boldsymbol{G}^{-1}(\cdot,\cdot)$ to exist. For (ii), the suggestion of Hannig (2009b) is to select one of the several solutions for $\boldsymbol{\theta}$ using a possible random mechanics. Besides, this choice of the parameter has only an effect of second order in statistic inference.

Recently, Hannig *et al.* (2016) provided an attractive and refined definition of the GFD, which can be applied in several situations where the data follow a continuous distribution. Thus, under some differentiability conditions (HANNIG *et al.*, 2016), the generalized fiducial distribution for $\boldsymbol{\theta}$ has a density function given by

$$f_F(\boldsymbol{\theta})=\kappa L(\boldsymbol{\theta};\boldsymbol{z})\boldsymbol{J}(\boldsymbol{z},\boldsymbol{\theta}), \tag{3.10}$$

where κ is a normalizing constant given by $\int_{\boldsymbol{\Theta}} L(\boldsymbol{\theta}';\boldsymbol{z})\boldsymbol{J}(\boldsymbol{z},\boldsymbol{\theta}')d\boldsymbol{\theta}'$, $L(\boldsymbol{z};\boldsymbol{\theta})$ is the likelihood and the function

$$\boldsymbol{J}(\boldsymbol{z},\boldsymbol{\theta})=\mathscr{D}\left(\frac{\boldsymbol{d}}{\boldsymbol{d\theta}}\boldsymbol{G}(\boldsymbol{\theta},\boldsymbol{u})\bigg|_{\boldsymbol{u}=\boldsymbol{G}^{-1}(\boldsymbol{\theta},\boldsymbol{z})}\right). \tag{3.11}$$

If (i) $n=p$ then $\mathscr{D}(\boldsymbol{A}^*)=|\det(\boldsymbol{A}^*)|$. Otherwise the function $\mathscr{D}(\boldsymbol{A}^*)$ depends on the norm used; (ii) The l_∞ norm gives $\mathscr{D}(\boldsymbol{A}^*)=\sum_{\boldsymbol{j}\in\mathscr{J}}|\det(\boldsymbol{A}^*)_{\boldsymbol{j}}|$, where $\mathscr{J}$ contains the $\binom{n}{p}$ p-tuples of indices $\boldsymbol{j}=(1\leq j_1<\ldots<j_p\leq n)^\top$. For any $n\times p$ matrix $\boldsymbol{A}^*$, the submatrix $(\boldsymbol{A}^*)_{\boldsymbol{j}}$ is the $p\times p$ matrix containing the rows $\boldsymbol{j}=(j_1,\ldots,j_p)^\top$ of $\boldsymbol{A}^*$; (iii) Under an additional assumption, the l_2 norm gives $\mathscr{D}(\boldsymbol{A}^*)=[\det(\boldsymbol{A}^{*\top}\boldsymbol{A}^*)]^{1/2}$. Among these functions defined for $\mathscr{D}(\boldsymbol{A}^*)$, Hannig *et al.* (2016) recommended using (ii).

A noteworthy characteristic of the GFD is that its density could change with transformations of the data generating equation, as remarked in Hannig (2009b) and Hannig *et al.* (2016). Thus, assume that an observed data set $\boldsymbol{Z}'$ is obtained through an one-to-one smooth transformation of the observed data $\boldsymbol{Z}$, which is denoted by $\boldsymbol{Z}'=\boldsymbol{H}(\boldsymbol{Z})$. Hence, using the chain rule, the GFD baned on the new data generating equation and considering the observed data $\boldsymbol{z}'=\boldsymbol{H}(\boldsymbol{z})$ has its density function given by (3.10) with the Jacobian function (3.11) reduced to

$$\boldsymbol{J_H}(\boldsymbol{z}',\boldsymbol{\theta})=\mathscr{D}\left(\frac{\boldsymbol{d}}{\boldsymbol{dz}}\boldsymbol{H}(\boldsymbol{z})\,\frac{\boldsymbol{d}}{\boldsymbol{d\theta}}\boldsymbol{G}(\boldsymbol{u},\boldsymbol{\theta})\bigg|_{\boldsymbol{u}=\boldsymbol{G}^{-1}(\boldsymbol{z},\boldsymbol{\theta})}\right), \tag{3.12}$$

where $\boldsymbol{z}$ is written instead of $\boldsymbol{H}^{-1}(\boldsymbol{z}')$.

A special transformation is described as follows. Let $\boldsymbol{Z}' = (\boldsymbol{S}, \boldsymbol{U})$ be an one-to-one smooth transformation, where $\boldsymbol{S}$ is a p-dimensional statistic and $\boldsymbol{U}$ is an ancillary statistic. Let $\boldsymbol{s} = \boldsymbol{S}(z)$ and $\boldsymbol{u} = \boldsymbol{U}(z)$ represent the observations of $\boldsymbol{S}$ and $\boldsymbol{U}$, respectively. Since $\boldsymbol{dU}/\boldsymbol{d\theta} = \boldsymbol{0}$, the function $\mathscr{D}$ in (3.12) reduces to the absolute value of the determinant of the $p \times p$ non-zero sub-matrix

$$J(z', \boldsymbol{\theta}) = \left| \det \left(\frac{\boldsymbol{d}}{\boldsymbol{d\theta}} \boldsymbol{S}(\boldsymbol{G}(\boldsymbol{U}, \boldsymbol{\theta})) \Big|_{\boldsymbol{u}=\boldsymbol{G}^{-1}(z,\boldsymbol{\theta})} \right) \right|. \tag{3.13}$$

Let $Q_{\boldsymbol{s}}(\boldsymbol{u}) = \boldsymbol{\theta}$ be the solution of the equation $\boldsymbol{s} = \boldsymbol{S}(\boldsymbol{G}(\boldsymbol{u}, \boldsymbol{\theta}))$. A direct calculation shows that the density function (3.10) with the Jacobian function (3.13) is the conditional distribution of $Q_{\boldsymbol{s}}(\boldsymbol{U}^*)|\boldsymbol{A}(\boldsymbol{U}^*) = \boldsymbol{a}$, that is, the GFD based on $\boldsymbol{S}$ conditional on the observed ancillary $\boldsymbol{U} = \boldsymbol{u}$ (IYER; PATTERSON, 2002; HANNIG *et al.*, 2016).

Next, it follows in details the construction of the GFD for the parameters of interest in the Grubbs model. Assume $\boldsymbol{Z}_i = (X_i, Y_i)^\top$ are independent random vectors following the $\mathscr{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, where $\boldsymbol{\mu} = (\mu_X, \mu_Y)^\top$ and

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix},$$

with $\rho \in (-1, 1)$ is the Pearson correlation coefficient. Following the steps of the method described above, Wandler and Hannig (2011) obtained the GFD for $\boldsymbol{\theta}' = (\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)^\top$. The specific steps are described as follows. First, the data generating equation is given by

$$\boldsymbol{Z}_i = \boldsymbol{G}(\boldsymbol{U}_i, \boldsymbol{\theta}') = \boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{U}_i, \quad i = 1, \ldots, n, \tag{3.14}$$

where $\boldsymbol{U}_i = (U_{1i}, U_{2i})^\top$ are independent random vectors following the $\mathscr{N}_2(\boldsymbol{0}_2, \boldsymbol{I}_2)$ distribution and $\boldsymbol{L}$ is the lower triangle Cholesky decomposition of $\boldsymbol{\Sigma}$ given by

$$\boldsymbol{L} = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} = \begin{bmatrix} \sigma_X & 0 \\ \rho\sigma_Y & \sigma_Y(1-\rho^2)^{1/2} \end{bmatrix}.$$

The inverse of the data generating equation $\boldsymbol{U}_i = \boldsymbol{G}^{-1}(\boldsymbol{Z}_i, \boldsymbol{\theta})$ is obtained from (3.14) with components given by $U_{1i} = (X_i - \mu_X)/L_{11}$ and $U_{2i} = (Y_i - \mu_Y - L_{21}U_{1i})/L_{22}$, $i = 1, \ldots, n$. Since there are five parameters, it is necessary to have the same number of generating equations to determinate each term in (ii). To find that, Wandler and Hannig (2011) used the following five data generating equations using the first three samples, i.e., $U_{11}, U_{21}, U_{12}, U_{22}$ and U_{23}. Moreover, in this case, $\mathscr{J}$ contains the $\binom{n}{2,1,n-3}$ p-tuples of indices $\boldsymbol{j} = (1 \leq j_1 < \ldots < j_p \leq n)^\top$. Hence, Wandler and Hannig (2011) obtained the GFD for $\boldsymbol{\theta}'$ given by

$$f_F(\boldsymbol{\theta}') \propto \frac{1}{[\det(\boldsymbol{\Sigma})]^{n/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} (z_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (z_i - \boldsymbol{\mu}) \right) \times \binom{n}{2, 1, n-3}^{-1} \sum_{\boldsymbol{j} \in \mathscr{J}} \frac{|v(\boldsymbol{z_j})|}{2^2(\sigma_X^2)^{3/2}(\sigma_Y^2)^{1/2}(1-\rho^2)}, \tag{3.15}$$

where $|v(\boldsymbol{z_j})|$ is the absolute value of a data function $v(\boldsymbol{z_j})$ and will not be specified here due to that this function is scorned later.
Now, the vector of parameters in our model in (3.2) is $\boldsymbol{\theta} = (\beta_0, \mu_x, \sigma_x^2, \sigma_u^2, \sigma_\varepsilon^2)^\top$ and to obtain its GFD, we need a mapping that transforms $\boldsymbol{\theta}'$ to $\boldsymbol{\theta}$. Hannig (2013) demonstrated that the GFD is invariant under smooth re-parameterization. Thus, let $\boldsymbol{h} : \mathbb{R}^5 \to \mathbb{R}^5$ be an one-to-one transformation defined by $\boldsymbol{h}(\boldsymbol{\theta}') = (\mu_Y - \mu_X, \mu_X, \rho\sigma_X\sigma_Y, \sigma_X^2 - \rho\sigma_X\sigma_Y, \sigma_Y^2 - \rho\sigma_X\sigma_Y)^\top = \boldsymbol{\theta}$. Consequently, the Jacobian matrix of the transformation from $\boldsymbol{\theta}'$ to $\boldsymbol{\theta}$ is given by

$$\boldsymbol{J_h}(\boldsymbol{\theta}) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & \dfrac{\sigma_u^2\sigma_\varepsilon^2 + D}{2(\sigma_x^4 + D)^{3/2}} & -\dfrac{\sigma_x^2(\sigma_x^2+\sigma_\varepsilon^2)^{-1/2}}{2(\sigma_x^2+\sigma_u^2)^{3/2}} & -\dfrac{\sigma_x^2(\sigma_x^2+\sigma_u^2)^{-1/2}}{2(\sigma_x^2+\sigma_\varepsilon^2)^{3/2}} \end{bmatrix},$$

where D is given in (3.5). And, since the determinant of $\boldsymbol{J_h}(\boldsymbol{\theta})$ is $(\sigma_x^4 + D)^{-1/2}$ and by using the GFD for $\boldsymbol{\theta}'$ given in (3.15), it follows that the GFD for $\boldsymbol{\theta}$ is given by

$$\begin{aligned} f_F(\boldsymbol{\theta}) \propto & \frac{1}{[\det(\boldsymbol{\Sigma})]^{n/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{z}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu})\right) \\ & \times \binom{n}{2,1,n-3}^{-1} \sum_{\boldsymbol{j}\in\mathscr{C}} \frac{|v(\boldsymbol{z_j})|(\sigma_x^2+\sigma_\varepsilon^2)^{1/2}}{2^2(\sigma_x^2+\sigma_u^2)^{1/2}D} \times \frac{1}{(\sigma_x^4+D)^{1/2}}, \end{aligned} \tag{3.16}$$

Here, $\boldsymbol{\mu} = (\mu_x, \beta_0 + \mu_x)^\top$, while $\boldsymbol{\Sigma}$ and D are given in (3.3) and (3.5), respectively. Next, we calculate the GFD for the parameter of interest, i.e, $\boldsymbol{\xi} = (\sigma_x^2, \sigma_u^2, \sigma_\varepsilon^2)^\top$. Therefore, the marginal GFD for $\boldsymbol{\xi}$ is

$$\begin{aligned} f_F(\boldsymbol{\xi}) &\propto \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\boldsymbol{R_\theta}}(\boldsymbol{\theta})d\beta_0 d\mu_x \\ &\propto \exp\left(-\frac{(n-1)\left[(S_X^2 - 2S_{XY} + S_Y^2)\sigma_x^2 + \sigma_\varepsilon^2 S_X^2 + \sigma_u^2 S_Y^2\right]}{2D}\right) \\ &\quad \times \frac{(\sigma_x^2+\sigma_\varepsilon^2)^{1/2}}{D^{(n+1)/2}\left[(\sigma_x^2+\sigma_u^2)(\sigma_x^4+D)\right]^{1/2}}. \end{aligned} \tag{3.17}$$

Finally, for $f_F(\boldsymbol{\xi})$ in (3.17), the Metropolis algorithm can be used to generate Markov chain Monte Carlo (MCMC) samples and to obtain both point and interval estimates. It is well-known that the Metropolis algorithm is often used when the domain of the GFD is the whole Euclidean space. Thus, we apply a similar technique given in Wang (2004). Since $\sigma_x^2 > 0$, $\sigma_u^2 > 0$ and $\sigma_\varepsilon^2 > 0$, we use for convenience the following transformations: $\lambda_1 = \log(\sigma_x^2)$, $\lambda_2 = \log(\sigma_u^2)$ and $\lambda_3 = \log(\sigma_\varepsilon^2)$. After this transformation, the domain of the GFD for $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)^\top$ is the

Euclidean space $\mathbb{R}^3$, and so the GFD for $\boldsymbol{\lambda}$ is

$$f_F^*(\boldsymbol{\lambda}) \propto \exp\left(-\frac{(n-1)\left[(S_X^2 - 2S_{XY} + S_Y^2)\sigma_x^2 + \sigma_\varepsilon^2 S_X^2 + \sigma_u^2 S_Y^2\right]}{2D}\right) \times \frac{(\sigma_x^2+\sigma_\varepsilon^2)^{1/2}}{D^{(n+1)/2}\left[(\sigma_x^2+\sigma_u^2)(\sigma_x^4+D)\right]^{1/2}} \exp(\lambda_1+\lambda_2+\lambda_3), \tag{3.18}$$

where $\sigma_x^2 = \exp(\lambda_1)$, $\sigma_u^2 = \exp(\lambda_2)$ and $\sigma_\varepsilon^2 = \exp(\lambda_3)$ are functions of λ_1, λ_2 and λ_3.

From (3.18), we generate a random sequence $(\lambda_1^{(t)}, \lambda_2^{(t)}, \lambda_3^{(t)})^\top$, $(t = 1,\ldots,N',\ldots)$ whose distributions converge to $f_F^*(\boldsymbol{\lambda})$ and then to calculate both point and interval estimates. To calculate the point and interval estimates for our parameters of interest we merely apply the inverse transformation. The proposal distribution for $\boldsymbol{\lambda}$ is a $\mathcal{N}_3(\mathbf{0}_3, a^2\boldsymbol{I}_3)$ random walk, where a is tuned to adjust the acceptance rate. Therefore, samples from the GFD for $\boldsymbol{\lambda}$ in (3.18) are drawn through the following algorithm:

Step 1. Choose a starting point $(\lambda_1^{(0)}, \lambda_2^{(0)}, \lambda_3^{(0)})^\top$

Step 2. For $t = 1,2,\ldots,N'+N_0$, iteratively do:

(a) Generate independent random values w_1, w_2, w_3 from the $\mathcal{N}(0,a^2)$ distribution.

(b) Let $(w_1', w_2', w_3')^\top = (\lambda_1^{(t-1)}, \lambda_2^{(t-1)}, \lambda_3^{(t-1)})^\top + (w_1, w_2, w_3)^\top$

(b) Compute the ratio of the densities $r^* = f_F^*(w_1', w_2', w_3')/f_F^*(w_1, w_2, w_3)$.

(c) Generate a uniform random value u on $(0,1)$.
If $u \le r^*$, then $(\lambda_1^{(t)}, \lambda_2^{(t)}, \lambda_3^{(t)})^\top = (w_1', w_2', w_3')^\top$,
else $(\lambda_1^{(t)}, \lambda_2^{(t)}, \lambda_3^{(t)})^\top = (\lambda_1^{(t-1)}, \lambda_2^{(t-1)}, \lambda_3^{(t-1)})^\top$.

Step 3. For $j \in \{1,2,3\}$, to do the transformation $\xi_j^{(t)} = \exp(\lambda_j^{(t)})$, $t = 1,\ldots,N'+N_0$ and put $\xi_j^{(1)},\ldots,\xi_j^{(N'+N_0)}$ in ascending order. Discard the first N_0 samples and calculate for the sample $\xi_j^{(N_0+1)},\ldots,\xi_j^{(N'+N_0)}$ the median as a point estimate as well as the $(1-\gamma)/2$-th and $(1+\gamma)/2$-th elements of the ordered values as the limits of the equal-tailed two-sided confidence interval.

In our simulation and application in Sections 3.3 and 3.4, we generate $N'+N_0 = 55000$ samples for each Markov chain, where $N_0 = 5000$ and we take a spacing of size 5. Consequently, inferences are based on 10000 samples.

Lastly, the GFD for $\boldsymbol{\theta}$ delivers an additional consequence. The second row of the product in (3.16) provides the expression for $\boldsymbol{J}(\boldsymbol{z},\boldsymbol{\theta})$ and, in turn, it can be factorized as the product of two functions, i.e., $\boldsymbol{J}(\boldsymbol{z},\boldsymbol{\theta}) = v^*(\boldsymbol{z})l(\boldsymbol{\theta})$, where

$$v^*(\boldsymbol{z}) = \binom{n}{2,1,n-3}^{-1} \sum_{\boldsymbol{j}\in\mathscr{J}} \frac{|v(\boldsymbol{z_j})|}{2^2} \quad \text{and} \quad l(\boldsymbol{\theta}) = \frac{(\sigma_x^2+\sigma_\varepsilon^2)^{1/2}}{\left[(\sigma_x^2+\sigma_u^2)(\sigma_x^4+D)\right]^{1/2} D}.$$

Due to this factorization, the GFD for $\boldsymbol{\theta}$ in (3.16) is also a Bayesian posterior with respect to the priori $l(\boldsymbol{\theta})$ along with constant priors for β_0 and μ_x, as pointed out by Hannig (2009b).

3.3 Simulation study

In this section, a simulation study is conducted to assess the performance of the point estimators for each procedure, as well as interval estimation in terms of the empirical coverage probability and average interval length. The confidence coefficient is set at $\gamma = 0.95$. Recall G denotes the Grubbs estimation method, T is the Thompson's estimation method, FGPQ denotes the estimation method by using a fiducial generalized pivotal quantity and GFD is the estimation method using the generalized fiducial distribution. In the case of the FGPQ method, we use the modified FGPQ estimator given by $\max(0, R_{\xi_j})$, $j \in \{1,2,3\}$. We generate 1000 random samples of size n = 10, 15, 20, 30 and 50 from the model described in Section 3.1. The data $\boldsymbol{Z}_i$ are generated from (3.2) with the following parameter setting: $\beta_0 = 0.220$, $\mu_x = 4.400$, $\sigma_x^2 = 0.035$, $\sigma_u^2 = 0.010$ and $\sigma_\varepsilon^2 = 0.034$. These values mimic the estimates obtained from the real data set in Section 3.4. The computational implementation was developed in the R language (R CORE TEAM, 2017).

Based on 1000 random samples, Table 10 shows the average of the point estimates, the sampling standard deviation (SD) and the root mean squared error (RMSE) for the estimation procedures described in Section 3.2. The medians obtained by the FGPQ and GFD procedures are adopted as point estimates.

In Table 10, the values of SD and RMSE decrease when the sample size increases, as expected. Since the bias of the estimates is small for the majority of the scenarios, the values of SD and RMSE are similar. Moreover, for most of the entries, the values of SD and RMSE for the GFD method are slightly smaller in comparison with the other procedures. With respect to σ_x^2 and σ_u^2, when the sample size is less than 30, most of the results indicate the point estimates have some bias, but when $n \geq 30$ the bias decreases. We also see that for σ_ε^2, the estimates from the GFD method achieve better performance than the other methods in terms of bias.

The results of the empirical coverage probability and the average interval length of the 95% confidence intervals are displayed in Table 11. For G method, in some replications the 2.5%-percentile of the chi-square distribution is very small because the estimate of the degrees of freedom from Section 3.2.1.1 is smaller than one. Consequently, the upper limit of the confidence interval is very big and the average interval length is greater than 10^4. Overall, the G and T methods behave poorly delivering the widest confidence intervals with coverage probability far away from the nominal confidence coefficient. With respect to σ_x^2, it can be seen that the GFD method has coverage probability that is the closest to the nominal value among the four methods. The FGPQ method is the second best. It should be noted that in the case of σ_u^2 and σ_ε^2, the FGPQ and GFD methods yield coverage probabilities close to the nominal value even for small and moderate sample sizes. It is clear that, on average, the shortest confidence intervals correspond to the GFD method and as the second best was the FGPQ method.

Due to the negative estimates from (3.4), in order to obtain 1000 replications with admissible variance estimates for all procedures, 474, 277, 254, 141 and 70 replications were discarded when the sample sizes is 10, 15, 20, 30 and 50, respectively.

Table 10 – Average, sampling standard deviation (SD) and root mean squared error (RMSE) of the estimates from the G, T, FGPQ and GFD methods. True values are $\sigma_x^2 = 0.035$, $\sigma_u^2 = 0.010$ and $\sigma_\varepsilon^2 = 0.034$.

		Point estimates				SD				RMSE			
Parameter	n	G	T	FGPQ	GFD	G	T	FGPQ	GFD	G	T	FGPQ	GFD
σ_x^2	10	0.030	0.030	0.032	0.028	0.019	0.019	0.020	0.016	0.019	0.019	0.020	0.017
	15	0.032	0.032	0.033	0.030	0.016	0.016	0.017	0.015	0.016	0.016	0.017	0.015
	20	0.033	0.033	0.034	0.031	0.014	0.014	0.014	0.013	0.014	0.014	0.014	0.013
	30	0.033	0.033	0.034	0.032	0.011	0.011	0.012	0.011	0.011	0.011	0.012	0.011
	50	0.034	0.034	0.035	0.034	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009
σ_u^2	10	0.016	0.016	0.025	0.018	0.012	0.012	0.016	0.010	0.014	0.014	0.022	0.013
	15	0.014	0.014	0.019	0.016	0.010	0.010	0.011	0.008	0.010	0.010	0.014	0.010
	20	0.013	0.013	0.016	0.015	0.009	0.009	0.010	0.007	0.009	0.009	0.012	0.009
	30	0.012	0.012	0.014	0.013	0.007	0.007	0.007	0.006	0.007	0.007	0.009	0.007
	50	0.011	0.011	0.012	0.012	0.006	0.006	0.006	0.005	0.006	0.006	0.006	0.005
σ_ε^2	10	0.029	0.029	0.030	0.035	0.017	0.017	0.019	0.018	0.018	0.018	0.019	0.019
	15	0.031	0.031	0.031	0.034	0.015	0.015	0.016	0.016	0.015	0.015	0.016	0.016
	20	0.032	0.032	0.032	0.034	0.014	0.014	0.014	0.014	0.014	0.014	0.014	0.014
	30	0.032	0.032	0.033	0.034	0.012	0.012	0.012	0.012	0.012	0.012	0.012	0.012
	50	0.033	0.033	0.034	0.034	0.008	0.008	0.009	0.008	0.008	0.008	0.009	0.008

Table 11 – Empirical coverage probability (CP) and average interval length (IL) of the 95% confidence intervals from the G, T, FGPQ and GFD methods.

		CP				IL			
Parameter	n	G	T	FGPQ	GFD	G	T	FGPQ	GFD
σ_x^2	10	0.983	1.000	0.979	0.940	$>10^4$	0.234	0.122	0.088
	15	0.974	1.000	0.960	0.944	$>10^4$	0.150	0.088	0.071
	20	0.980	0.999	0.968	0.951	$>10^4$	0.119	0.071	0.061
	30	0.979	0.999	0.962	0.959	0.116	0.090	0.054	0.048
	50	0.967	0.995	0.951	0.947	0.042	0.065	0.040	0.037
σ_u^2	10	0.852	1.000	0.937	0.980	$>10^4$	0.170	0.125	0.069
	15	0.861	1.000	0.944	0.977	$>10^4$	0.100	0.071	0.052
	20	0.874	1.000	0.948	0.971	$>10^4$	0.075	0.052	0.043
	30	0.890	1.000	0.957	0.975	$>10^4$	0.055	0.038	0.033
	50	0.921	1.000	0.964	0.975	$>10^4$	0.040	0.026	0.025
σ_ε^2	10	0.985	1.000	0.986	0.979	$>10^4$	0.229	0.120	0.121
	15	0.985	0.998	0.962	0.970	$>10^4$	0.145	0.085	0.085
	20	0.983	0.999	0.961	0.968	$>10^4$	0.116	0.069	0.068
	30	0.975	0.998	0.950	0.951	0.089	0.088	0.053	0.052
	50	0.979	0.999	0.970	0.974	0.041	0.065	0.039	0.038

3.4 Data analysis

In this section, a real data set is analyzed to illustrate the proposed methodology and to compare it with two existing approaches, as described in Section 3.2.

This real dataset was extracted from Jaech (1985, p. 22). The results were obtained from an assay comparison study involving two methods for measuring the density of 43 cylindrical nuclear reactor fuel pellets of sintered uranium. The first method is called geometric method, which consists of weighing the pellet and finding its volume and thus obtaining the density, while the immersion method uses the change of the weight of the pellet when it is weighed in the air and in a certain liquid.

Here, X_i and Y_i represent the measurements obtained from the geometric and immersion methods, respectively, for $i = 1,\ldots,43$. Each measurement was expressed as the percentage theoretical density minus 90% for convenience. The measurements for the Geometric and Immersion methods range from 3.950 to 4.900 and from 3.940 to 5.220, respectively. Also, the average density of cylindrical nuclear reactor fuel pellets obtained from the Immersion method (4.397) was slightly smaller than that of the Geometric method (4.620).

Next, Table 12 displays the point estimates, the confidence intervals and their lengths with a confidence coefficient of 95% for the parameters of interest, i.e., the product variability and the precision parameters. The point estimates of σ_x^2, σ_u^2 and σ_ε^2 do not differ so much irrespective of the estimation method. For the interval estimation of these three parameters, the GFD procedure has the shortest interval length, followed by the FGPQ method. These results are consistent with the simulation study carried out in Section 3.3.

Table 12 – Point estimates, 95% confidence intervals and their lengths from the G, T, FGPQ and GFD methods.

Parameter	Estimate	G	T	FGPQ	GFD
σ_x^2	Point	0.035	0.035	0.035	0.034
	Interval	[0.021, 0.068]	[0.007, 0.082]	[0.019, 0.063]	[0.019, 0.059]
	Length	0.047	0.075	0.044	0.040
σ_u^2	Point	0.010	0.010	0.011	0.011
	Interval	[0.004, 0.082]	[0.000, 0.042]	[0.000, 0.030]	[0.002, 0.028]
	Length	0.078	0.042	0.030	0.026
σ_ε^2	Point	0.034	0.034	0.035	0.036
	Interval	[0.021, 0.067]	[0.007,0.081]	[0.018, 0.062]	[0.019, 0.062]
	Length	0.046	0.074	0.044	0.043

3.5 Conclusion

Two new estimation procedures based on the generalized fiducial inference were proposed for the Grubbs model. In our work, a FGPQ and a GFD were established for the product variability and precision parameters. Good empirical frequentist properties were obtained from the numerical results in Section 3.3. We found that the GFD and FGPQ methods are more suitable than the two other existing approaches and can achieve a good performance. Within the scope of our simulation study in Section 3.3, we believe the following conclusions are reliable: (i) For point estimation and when the sample size is 30 or larger, the four estimation procedures tend to give similar results and (ii) The FGPQ or the GFD method is recommended for the interval estimation because they give a good empirical coverage probability with shorter interval lengths than the other two existing approaches, when the sample size considered is small or moderate.

CHAPTER

4

A SKEW-t CENTERED STUDENT'S t MEASUREMENT ERROR MODEL

In this chapter, we propose a new MEM that extends the assumption of normality, which in this work is designated as skew-t centered Student's t MEM. We briefly review some recent works, the motivation for our model and possible applications. Next, we formulate the new heteroscedastic MEM and develop an EM-type algorithm for estimation. We conduct a simulation study to gauge the performance of the maximum likelihood (ML) estimators and illustrate using a dataset from a methods comparison study. Lastly, concluding remarks are also considered. On the basis of this chapter, we get to present and publish the main results in Tomaya and Castro (2018a).

4.1 Introduction

Among the main references about measurement error models cited in Chapter 1, many published works deal with homoscedastic measurement errors models (see, e.g., the model in Section 2.2.1). In practice, it frequently happens that the variability of a measurement may change across observations (BUONACCORSI, 2010, Section 6.4.5) and it is common to find datasets that allow the reliable estimation of the error variances (CHENG; RIU, 2006). To deal with such data, heteroscedastic measurement error models with known error variances have been studied in several areas such as Medicine (KULATHINAL; KUULASMAA; GASBARRA, 2002; CASTRO; GALEA; BOLFARINE, 2008), Analytical Chemistry (GALEA-ROJAS *et al.*, 2003) and Astrophysics (KELLY, 2007), among others.

A heteroscedastic linear MEM under the normality assumption can be formulated as

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \text{and} \quad X_i = x_i + u_i,\ i = 1, \ldots, n, \tag{4.1}$$

where β_0 and β_1 are the intercept and slope parameters, respectively, and X_i, Y_i, x_i and y_i

are defined as in (2.1). Assume in (4.1) that x_i are distributed as random variables from the $\mathcal{N}(\mu_x, \sigma_x^2)$ distribution, u_i and ε_i are independent and distributed as the $\mathcal{N}(0, \sigma_{u_i}^2)$ and $\mathcal{N}(0, \sigma_{\varepsilon_i}^2)$ distributions, respectively, $i = 1, \ldots, n$, i.e., ε_i and u_i are the heteroscedastic additive error terms. When normality is assumed, this model poses identifiability problems (as mentioned in Section 2.2.1) and to cope with them, a common alternative is to impose a side condition on one of the error variances (CHENG; NESS, 1999, Section 1.2).

The normality assumption is often debatable. For instance, it is often doubtful and suffers from a lack of robustness against influential observations on the parameter estimates. For this reason, several works have considered relaxations of the normality assumption, such as considering asymmetric distributions, see Arellano-Valle *et al.* (2005), Kheradmandi and Rasekh (2015) and models based on distributions with tails heavier than the ones of a normal distribution, see Cao, Lin and Zhu (2012), Melo, Ferrari and Patriota (2014).

On the other side, the skew-*t* distribution (AZZALINI; CAPITANIO, 2003) has been adopted as an attractive and flexible distribution. As considered in the literature, this distribution permits simultaneously to capture the skewness and the heavy-tailedness in the data. Some papers based on this distribution have been applied to MEM, for example, Lachos, Cancho and Aoki (2010), Lachos *et al.* (2010) and Cabral, Lachos and Zeller (2014). But, the same degrees of freedom is specified for both the distributions of the unobserved covariate and the error terms. Besides, Lachos, Cancho and Aoki (2010) presented under a Bayesian approach considering the degrees of freedom like known. It is worth noticing that the applicability of these models mentioned is limited in the sense that the degrees of freedom for the random variables are considered as known or (and) are the same. Hence, these models could not accommodate different levels of heaviness in the tails of the unobserved covariate and random errors. This issue occurs, for example, when the unobserved covariate and random errors have heavy tails and can be different. It is observed in the dataset related to a methods comparison study that we illustrate in Section 4.4.

In contrast to the works mentioned previously and under a frequentist approach, Choudhary, Sengupta and Casey (2014) proposed a linear mixed model based on the skew-*t* distribution allowing different degrees of freedom for the random variables. Thus, this brings us to the main purpose of this chapter, which is to present a MEM framework based on distributions with possible different degrees of freedom. We extend the classical normal MEM by modeling the unobserved covariate by a skew-*t* (ST) distribution and for the random errors, we assume the centered Student's *t* (cT) distribution (SUTRADHAR, 1993). As mentioned in Choudhary, Sengupta and Casey (2014), it is necessary to take into account the independence of both distributions, which allows their degrees of freedom to be different. Therefore, this approach enables us to model the data with great flexibility, accommodating skewness, heavy tails and a parameter representing the mean of the true covariate. The above formulation could also be proposed under a Bayesian approach, but this is not the scope in this chapter and it is part of possible future work.

4.2 Model and parameter estimation

In this section, we formulate the proposed model and give more details on estimation procedure through of the development of an EM-type algorithm to compute of the maximum likelihood estimates for the model parameters. Moreover, new theoretical results for the proposed model are development throughout this section.

4.2.1 Representation of the proposed model

The MEM in (4.1) can be written, in a similar form to Fuller (1987, p. 352), as

$$\boldsymbol{Z}_i = \boldsymbol{a} + \boldsymbol{b}x_i + \boldsymbol{e}_i, \tag{4.2}$$

where $\boldsymbol{Z}_i = (X_i, Y_i)^\top$, $\boldsymbol{a} = (0, \beta_0)^\top$, $\boldsymbol{b} = (1, \beta_1)^\top$ and $\boldsymbol{e}_i = (u_i, \varepsilon_i)^\top$, $i = 1, \ldots, n$. We assume in (4.2) that $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ are independent together with the assumptions

$$x_i \sim \mathscr{S}t(\xi, \omega^2, \lambda, \eta_x) \quad \text{and} \quad \boldsymbol{e}_i \sim ct_2\left(\boldsymbol{0}_2, \boldsymbol{\Sigma}_i, \eta_e\right), \tag{4.3}$$

with x_i independent of $\boldsymbol{e}_j$, for $i, j = 1, \ldots, n$. Since we are using the centered parameterization of the Student's t distribution in Sutradhar (1993), $\boldsymbol{\Sigma}_i = \mathrm{diag}(\sigma^2_{u_i}, \sigma^2_{\varepsilon_i})$ is the known covariance matrix of $\boldsymbol{e}_i$. We emphasize that heteroscedastic MEM's with known error variances are a common setup in many applications, see, e.g., Cheng and Riu (2006). The centered version of the Student's t distribution brings more flexibility with respect to the normal distribution and the interpretation for the mean vector and covariance matrix is direct (see Section 1.1). Further, the parameter λ introduces skewness in the unobserved covariate x_i, and consequently, in the vector of the observations $\boldsymbol{Z}_i$.

Rather than using the degrees of freedom ν_x and ν_e, we use $\eta_x = 1/\nu_x$ and $\eta_e = 1/\nu_e$, as commented in Lange, Little and Taylor (1989), these transformations improve the inference procedure. We call the model (4.2)-(4.3) as the skew-t centered Student's t measurement error model (STcT MEM). Notice that the degrees of freedom ν_x and ν_e can be different.

Now, according to Azzalini and Capitanio (2003), x_i and $\boldsymbol{e}_i$ can be represented stochastically. To do that, consider x_i and $\boldsymbol{e}_i$ as in (4.3), $W_{x_i} \sim \mathscr{G}(1/2\eta_x, 1/2\eta_x)$ and $W_{e_i} \sim \mathscr{G}(1/2\eta_e, 1/2c(\eta_e))$ with $c(\eta_e) = \eta_e/(1-2\eta_e)$. Let $x_i^* \sim \mathscr{SN}(0, \omega^2, \lambda)$, $G_{1i} \sim \mathscr{N}(0,1)$, $\boldsymbol{G}_{2i} \sim \mathscr{N}_2(\boldsymbol{0}_2, \boldsymbol{I}_2)$ and $D_i^* \sim \mathscr{TN}(0, 1; (0, +\infty))$. Assume that $G_{1i}, \boldsymbol{G}_{2i}, D_i^*, W_{x_i}$ and W_{e_i} are mutually independent. Thus, the stochastic representations of x_i and $\boldsymbol{e}_i$ are as follows:

$$x_i \overset{d}{=} \xi + x_i^*/W_{x_i}^{1/2} \quad \text{and} \quad \boldsymbol{e}_i \overset{d}{=} \boldsymbol{\Sigma}_i^{1/2}\boldsymbol{G}_{2i}/W_{e_i}^{1/2}, \tag{4.4}$$

with $x_i^* \overset{d}{=} \omega[\delta D_i^* + (1-\delta^2)^{1/2} G_{1i}]$ and "$\overset{d}{=}$" means equality in distribution. Next, we transform (ω^2, λ) to (ψ, γ), where

$$\gamma = \omega\delta \quad \text{and} \quad \psi = \omega^2(1-\delta^2), \quad \text{with} \quad \delta = \lambda/(1+\lambda^2)^{1/2}. \tag{4.5}$$

Using (4.5), an alternative stochastic representation of x_i^* is obtained, i.e., $x_i^* \stackrel{d}{=} \gamma D_i^* + \psi^{1/2} G_{1i}$. The importance of the transformation (4.5) is that it helps us to derive some theoretical results and the implementation of the EM-type algorithm (as described in Section 4.2.2).

In the sequel we present propositions needed to derive some properties and the marginal pdf of $\boldsymbol{Z}_i$, and consequently, the log-likelihood function for the proposed model. The proofs are given as follows.

Proposition 3. Consider $\boldsymbol{Z}_i$ as defined in (4.2) along with (4.5) and the stochastic representations given in (4.4).

(a) A hierarchical representation for $\boldsymbol{Z}_i$ is given as follows:

$$\begin{aligned}
&\boldsymbol{Z}_i|x_i, W_{e_i} \sim \mathcal{N}_2(\boldsymbol{a}+\boldsymbol{b}x_i, \boldsymbol{\Sigma}_i/W_{e_i}), \quad x_i|D_i, W_{x_i} \sim \mathcal{N}(\xi+\gamma D_i, \psi/W_{x_i}),\\
&D_i|W_{x_i} \sim \mathcal{TN}(0, 1/W_{x_i}; (0,\infty)), \quad W_{x_i} \sim \mathcal{G}\left(\frac{1}{2\eta_x}, \frac{1}{2\eta_x}\right) \quad \text{and}\\
&W_{e_i} \sim \mathcal{G}\left(\frac{1}{2\eta_e}, \frac{1}{2\,c(\eta_e)}\right) \quad \text{with} \quad c(\eta_e) = \eta_e/(1-2\eta_e).
\end{aligned} \tag{4.6}$$

This representation can also be written by transforming $(W_{x_i}, W_{e_i})^\top$ to $(U_i = W_{x_i}, V_i = W_{x_i}/W_{e_i})^\top$.

(b) The mean vector and the covariance matrix of $\boldsymbol{Z}_i$ are

$$\begin{aligned}
&E[\boldsymbol{Z}_i] = \boldsymbol{a}+\boldsymbol{b}\mu_x, \quad 0<\eta_x,\ 0<\eta_e<1/2 \quad \text{and}\\
&Var[\boldsymbol{Z}_i] = \left(\frac{\psi+\gamma^2}{1-2\eta_x} - \tau_{\eta_x}^2\gamma^2\right)\boldsymbol{b}\boldsymbol{b}^\top + \boldsymbol{\Sigma}_i, \quad 0<\eta_x, \eta_e<1/2,
\end{aligned} \tag{4.7}$$

where $\mu_x = \xi + \tau_{\eta_x}\gamma$, with $\tau_{\eta_x} = \left[\Gamma((1-2\eta_x)/2\eta_x)/\Gamma(1/2\eta_x)\right]/\sqrt{\eta_x\pi}$.

Proof. The result in (a) follows from (4.4) and (4.5). For (b), the result follows directly from (a) upon applying the law of iterated expectations and using well-known results about moments involving normal and gamma random variables. □

The following lemma is crucial in our theory, which was adapted for the measurement error model framework in (4.2) from the result for linear mixed model given in Arellano-Valle, Bolfarine and Lachos (2005) (Corollary 2 of Theorem 1). The lemma will be used to prove the following proposition. Moreover, other results found in the literature were also applied for the proofs and will be cited whenever needed.

Lemma 2. Let $\boldsymbol{Z}_i$ as defined in (4.2). Suppose $\boldsymbol{Z}_i|x_i \sim \mathcal{N}_2(\boldsymbol{a}+\boldsymbol{b}x_i, \boldsymbol{\Sigma}_i)$ and $x_i \sim \mathcal{SN}(\xi, \omega^2, \lambda)$. Then, $\boldsymbol{Z}_i \sim \mathcal{SN}_2(\boldsymbol{a}+\boldsymbol{b}\xi, \boldsymbol{\Pi}_i, \boldsymbol{\lambda}_i^*)$, where $\boldsymbol{\Pi}_i = \omega^2\boldsymbol{b}\boldsymbol{b}^\top + \boldsymbol{\Sigma}_i$ and $\boldsymbol{\lambda}_i^* = \boldsymbol{\Pi}_i^{-1/2}\boldsymbol{b}\omega\lambda/[1+\lambda^2\omega^{-2}(\omega^{-2}+\boldsymbol{b}^\top\boldsymbol{\Sigma}_i^{-1}\boldsymbol{b})^{-1}]^{1/2}$.

Proof. To obtain the marginal distribution de $\boldsymbol{Z}_i$, we compute

$$\begin{aligned} f(\boldsymbol{z}_i) &= \int_{-\infty}^{\infty} f(\boldsymbol{z}_i|x_i)f(x_i)dx_i \\ &= \int_{-\infty}^{\infty} 2\phi_2(\boldsymbol{z}_i;\boldsymbol{a}+\boldsymbol{b}x_i,\boldsymbol{\Sigma}_i)\phi(x_i;\xi,\omega^2)\Phi(\lambda(x_i-\xi)/\omega)dx_i \\ &= 2\phi_2(\boldsymbol{z}_i;\boldsymbol{a}+\boldsymbol{b}\xi,\boldsymbol{\Pi}_i)\int_{-\infty}^{\infty}\phi(x_i;\xi+\Lambda_i\boldsymbol{b}^\top\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{z}_i-\boldsymbol{a}-\boldsymbol{b}\xi),\Lambda_i)\Phi(\lambda(x_i-\xi)/\omega)dx_i, \end{aligned} \tag{4.8}$$

where $\boldsymbol{\Pi}_i = \omega^2\boldsymbol{b}\boldsymbol{b}^\top + \boldsymbol{\Sigma}_i$ and $\Lambda_i = (\omega^{-2}+\boldsymbol{b}^\top\boldsymbol{\Sigma}_i^{-1}\boldsymbol{b})^{-1}$. The third equality is obtained after applying the Lemma 2 provided in Arellano-Valle, Bolfarine and Lachos (2005) in the second equality in (4.8). Besides that, the last integral is the expectation of $\Phi(\lambda(x_i-\xi)/\omega)$, wherein $x_i \sim \mathcal{N}(\xi+\Lambda_i\boldsymbol{b}^\top\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{z}_i-\boldsymbol{a}-\boldsymbol{b}\xi),\Lambda_i)$ and using the Lemma 1 in Arellano-Valle, Bolfarine and Lachos (2005), it follows that

$$\begin{aligned} f(\boldsymbol{z}_i) &= 2\phi_2(\boldsymbol{z}_i;\boldsymbol{a}+\boldsymbol{b}\xi,\boldsymbol{\Pi}_i)\Phi(-\lambda\xi/\omega;-\lambda(\xi+\Lambda_i\boldsymbol{b}^\top\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{z}_i-\boldsymbol{a}-\boldsymbol{b}\xi))/\omega,1+\lambda^2\omega^{-2}\Lambda_i) \\ &= 2\phi_2(\boldsymbol{z}_i;\boldsymbol{a}+\boldsymbol{b}\xi,\boldsymbol{\Pi}_i)\Phi\big([\lambda\Lambda_i\boldsymbol{b}^\top\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{z}_i-\boldsymbol{a}-\boldsymbol{b}\xi)]/[\omega(1+\lambda^2\omega^{-2}\Lambda_i)^{1/2}]\big). \end{aligned} \tag{4.9}$$

Now, we must verify if the argument of the cdf of the $\mathcal{N}(0,1)$ distribution in (4.9) is equal to $\boldsymbol{\lambda}_i^{*\top}\boldsymbol{\Pi}_i^{-1/2}(\boldsymbol{z}_i-\boldsymbol{a}-\boldsymbol{b}\xi)$. To do that, one can see that the matrix $\boldsymbol{\Pi}_i$ does not appear in the second term of the second equality in (4.9), so using a well-known matrix property, which is given by $\boldsymbol{I}_2 = \boldsymbol{\Pi}_i\boldsymbol{\Pi}_i^{-1} = \boldsymbol{\Pi}_i{\boldsymbol{\Pi}_i^{-1/2}}^\top\boldsymbol{\Pi}_i^{-1/2}$, we get

$$\frac{\lambda\Lambda_i\boldsymbol{b}^\top\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Pi}_i{\boldsymbol{\Pi}_i^{-1/2}}^\top}{\omega(1+\lambda^2\omega^{-2}\Lambda_i)^{1/2}}\boldsymbol{\Pi}_i^{-1/2}(\boldsymbol{z}_i-\boldsymbol{a}-\boldsymbol{b}\xi) = \left[\frac{\boldsymbol{\Pi}_i^{-1/2}\boldsymbol{\Pi}_i\boldsymbol{\Sigma}_i^{-1}\boldsymbol{b}\Lambda_i\omega^{-1}\lambda}{(1+\lambda^2\omega^{-2}\Lambda_i)^{1/2}}\right]^\top\boldsymbol{\Pi}_i^{-1/2}(\boldsymbol{z}_i-\boldsymbol{a}-\boldsymbol{b}\xi). \tag{4.10}$$

After some algebraic manipulations, we obtain that $\boldsymbol{\Pi}_i\boldsymbol{\Sigma}_i^{-1}\boldsymbol{b}\Lambda_i\omega^{-1} = \boldsymbol{b}\omega$ and replacing within the brackets in (4.10) follows the expression of $\boldsymbol{\lambda}_i^*$. Consequently, the result holds. □

Turning out to our model, to derive the marginal pdf of $\boldsymbol{Z}_i$, we define, for $v_i > 0$,

$$\begin{aligned} \boldsymbol{\Pi}_{v_i} &= \omega^2\boldsymbol{b}\boldsymbol{b}^\top + v_i\boldsymbol{\Sigma}_i, \quad \boldsymbol{\Delta}_i = \boldsymbol{Z}_i-\boldsymbol{a}-\boldsymbol{b}\xi, \\ \boldsymbol{\lambda}_{v_i} &= \frac{\boldsymbol{\Pi}_{v_i}^{-1/2}\boldsymbol{b}\omega\lambda}{[1+\lambda^2\omega^{-2}(\omega^{-2}+\boldsymbol{b}^\top\boldsymbol{\Sigma}_i^{-1}\boldsymbol{b}/v_i)^{-1}]^{1/2}}, \quad \xi_{v_i} = \boldsymbol{\lambda}_{v_i}^\top\boldsymbol{\Pi}_{v_i}^{-1/2}\boldsymbol{\Delta}_i, \\ \varsigma_{v_i} &= \boldsymbol{\Delta}_i^\top\boldsymbol{\Pi}_{v_i}^{-1}\boldsymbol{\Delta}_i + \frac{1}{\eta_x} + \frac{1}{c(\eta_e)v_i} \quad \text{and} \quad \eta^* = 2+\frac{1}{\eta_x}+\frac{1}{\eta_e}. \end{aligned} \tag{4.11}$$

It is worth noticing that expressions such as $\boldsymbol{\Pi}_{v_i}$ and $\boldsymbol{\lambda}_{v_i}$ defined in (4.11) were adapted from Lemma 2.

Proposition 4. Consider $\boldsymbol{Z}_i$ as defined in (4.2), (4.6) and the transformation $(U_i = W_{x_i}, V_i = W_{x_i}/W_{e_i})^\top$.

(a) A hierarchical representation for $\boldsymbol{Z}_i$ is given as follows:

$$\boldsymbol{Z}_i|U_i,V_i \sim \mathscr{SN}_2(\boldsymbol{a}+\boldsymbol{b}\xi,\, \boldsymbol{\Pi}_{V_i}/U_i, \boldsymbol{\lambda}_{V_i}),$$
$$U_i \sim \mathscr{G}\left(\frac{1}{2\eta_x}, \frac{1}{2\eta_x}\right) \quad \text{and} \quad U_i/V_i \sim \mathscr{G}\left(\frac{1}{2\eta_e}, \frac{1}{2\,c(\eta_e)}\right).$$

(b) For $\boldsymbol{z}_i \in \mathbb{R}^2$, $u_i, v_i > 0$, the joint pdf of $(\boldsymbol{Z}_i^\top, U_i, V_i)^\top$ is

$$f(\boldsymbol{z}_i, u_i, v_i; \boldsymbol{\theta}) = \frac{2\pi^{-1}}{\eta_x^{1/2\eta_x} c(\eta_e)^{1/2\eta_e}} \frac{[\det(\boldsymbol{\Pi}_{v_i})]^{-1/2}}{\varsigma_{v_i}^{\eta^*/2} v_i^{1/2\eta_e+1}} \frac{\Gamma(\eta^*/2)}{\Gamma(1/2\eta_x)\Gamma(1/2\eta_e)} \times \Phi(\xi_{v_i} u_i^{1/2}) f(u_i; \eta^*/2, \varsigma_{v_i}/2), \tag{4.12}$$

where $f(\cdot;\eta^*/2,\varsigma_{v_i}/2)$ is the pdf of the $\mathscr{G}(\eta^*/2,\varsigma_{v_i}/2)$ distribution.

(c) For $\boldsymbol{z}_i \in \mathbb{R}^2$, $v_i > 0$, the joint pdf of $(\boldsymbol{Z}_i^\top, V_i)^\top$ is

$$f(\boldsymbol{z}_i, v_i; \boldsymbol{\theta}) = \frac{2\pi^{-1}}{\eta_x^{1/2\eta_x} c(\eta_e)^{1/2\eta_e}} \frac{[\det(\boldsymbol{\Pi}_{v_i})]^{-1/2}}{\varsigma_{v_i}^{\eta^*/2} v_i^{1/2\eta_e+1}} \frac{\Gamma(\eta^*/2)}{\Gamma(1/2\eta_x)\Gamma(1/2\eta_e)} \times F_t(\xi_{v_i}(\eta^*/\varsigma_{v_i})^{1/2}; \eta^*). \tag{4.13}$$

(d) For $\boldsymbol{z}_i \in \mathbb{R}^2$, the marginal pdf of $\boldsymbol{Z}_i$ is given by

$$f(\boldsymbol{z}_i; \boldsymbol{\theta}) = \int_0^\infty f(\boldsymbol{z}_i, v_i; \boldsymbol{\theta}) dv_i. \tag{4.14}$$

Proof. First, consider the expressions in (4.11). For the result in (a), we use (4.4) to write $\boldsymbol{Z}_i$ as

$$\boldsymbol{Z}_i|x_i, W_{e_i} \sim \mathscr{N}_2(\boldsymbol{a}+\boldsymbol{b}x_i, \boldsymbol{\Sigma}/W_{e_i}) \quad \text{and} \quad x_i|W_{x_i} \sim \mathscr{SN}(\xi, \omega^2/W_{x_i}, \lambda),$$

where $W_{x_i} \sim \mathscr{G}(1/2\eta_x, 1/2\eta_x)$ and $W_{e_i} \sim \mathscr{G}(1/2\eta_e, 1/2c(\eta_e))$ with $c(\eta_e) = \eta_e/(1-2\eta_e)$. The result now holds from an application of the Lemma 2 and after transforming $(W_{x_i}, W_{e_i})^\top$ to $(U_i = W_{x_i}, V_i = W_{x_i}/W_{e_i})^\top$. For (b), we use (a) to write the joint pdf of $(\boldsymbol{Z}_i^\top, U_i, V_i)^\top$ as $f(\boldsymbol{z}_i, u_i, v_i; \boldsymbol{\theta}) = f(\boldsymbol{z}_i|u_i, v_i; \boldsymbol{\theta}) f(u_i, v_i; \boldsymbol{\theta})$ and after simplification, the result follows. For (c), the result is obtained by integrating $f(\boldsymbol{z}_i, u_i, v_i; \boldsymbol{\theta})$ with respect to u_i by using Lemma 1 in Azzalini and Capitanio (2003). Finally, for (d), we integrate out v_i from $f(\boldsymbol{z}_i, v_i; \boldsymbol{\theta})$ given in (c). □

Notice that the marginal pdf of $\boldsymbol{Z}_i$ is not available in a closed form and requires one-dimensional numerical integration. The results in Proposition 4 facilitate the computational implementation of the estimation method. Further, it follows that the log-likelihood function of $\boldsymbol{\theta}$ given the observed data $\boldsymbol{z} = (\boldsymbol{z}_1^\top, \ldots, \boldsymbol{z}_n^\top)^\top$ is given by

$$\begin{aligned}\ell(\boldsymbol{\theta}) &= \log(L(\boldsymbol{\theta}; \boldsymbol{z})) = \log(f(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n; \boldsymbol{\theta})) = \sum_{i=1}^n \log(f(\boldsymbol{z}_i; \boldsymbol{\theta})) \\ &= \sum_{i=1}^n \log\left(\int_0^\infty f(\boldsymbol{z}_i, v_i; \boldsymbol{\theta}) dv_i\right),\end{aligned} \tag{4.15}$$

with $f(\boldsymbol{z}_i, v_i; \boldsymbol{\theta})$ as in Proposition 4 and $\boldsymbol{\theta} = (\beta_0, \beta_1, \xi, \omega^2, \lambda, \eta_x, \eta_e)^\top$ encapsulates the parameters in the model.

4.2.2 Maximum likelihood estimation via the ECM algorithm

We call attention to the fact that the log-likelihood function in (4.15) can be directly maximized to get the ML estimates $\widehat{\boldsymbol{\theta}}$. However, this is generally not practical due to the dimension of $\boldsymbol{\theta}$ or the complexity of the log-likelihood function, as in (4.15). For this reason, we consider the ECM algorithm (MENG; RUBIN, 1993), which is a modification of the EM algorithm (DEMPSTER; LAIRD; RUBIN, 1977). The M step of the EM algorithm is replaced by a sequence of simpler constrained maximization (CM) steps. Each iteration of this algorithm increases the likelihood function to achieve convergence at a local or a global maximum.
Next, we denote $\boldsymbol{Z}_{\text{mis},i} = (x_i, D_i, U_i, V_i)^\top$ as the missing data and $\boldsymbol{Z}_{\text{comp},i} = (\boldsymbol{Z}_i^\top, \boldsymbol{Z}_{\text{mis},i}^\top)^\top$ as the complete data for the i-th subject. The random quantities x_i, D_i, U_i and V_i are from (4.6) with $(W_{x_i}, W_{e_i})^\top$ transformed to $(U_i = W_{x_i}, V_i = W_{x_i}/W_{e_i})^\top$. Besides that, there is an one-to-one correspondence between $(\omega^2, \lambda)^\top$ and $(\gamma, \psi)^\top$, so that for this algorithm, we consider $\boldsymbol{\theta}^* = (\beta_0, \beta_1, \xi, \gamma, \psi, \eta_x, \eta_e)^\top$ as the parameter vector to be estimated, and using the inverse transformation of (4.5), $\boldsymbol{\theta}$ can be estimated. Consequently, the complete data log-likelihood function for $\boldsymbol{z}_{\text{comp}} = (\boldsymbol{z}_{\text{comp},1}^\top, \ldots, \boldsymbol{z}_{\text{comp},n}^\top)^\top$ takes the form

$$\ell_c(\boldsymbol{\theta}^*) = \sum_{i=1}^n \ell_{c,i}(\boldsymbol{\theta}^*) = \log\big(L(\boldsymbol{\theta}^*; \boldsymbol{z}_{\text{comp}})\big) = \sum_{i=1}^n \log\big(f(\boldsymbol{z}_{\text{comp},i}; \boldsymbol{\theta}^*)\big), \tag{4.16}$$

where

$$\begin{aligned}
\ell_{c,i}(\boldsymbol{\theta}^*) =& c - \frac{1}{2}(\boldsymbol{z}_i - \boldsymbol{a})^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{z}_i - \boldsymbol{a})(u_i/v_i) + (\boldsymbol{z}_i - \boldsymbol{a})^\top \boldsymbol{\Sigma}_i^{-1}\boldsymbol{b}x_i(u_i/v_i) - \frac{1}{2}\boldsymbol{b}^\top \boldsymbol{\Sigma}_i^{-1}\boldsymbol{b}x_i^2(u_i/v_i) \\
& - \frac{1}{2}\log(\psi) - \frac{1}{2\psi}\left[x_i^2 + \xi^2 + \gamma^2 d_i^2 - 2(\xi x_i + \gamma x_i d_i - \xi\gamma d_i)\right]u_i \\
& + \frac{1}{2\eta_x}\log(1/2\eta_x) - \log(\Gamma(1/2\eta_x)) + \frac{1}{2\eta_x}[\log(u_i) - u_i] + \frac{1}{2\eta_e}\log(1/[2\,c(\eta_e)]) \\
& - \log(\Gamma(1/2\eta_e)) + \frac{1}{2\eta_e}[\log(u_i/v_i) - (1-2\eta_e)(u_i/v_i)].
\end{aligned}$$

where c is free of parameters. From (4.16) and ignoring the terms that do not involve $\boldsymbol{\theta}^*$, the expected log-likelihood function in the r-th iteration of the ECM algorithm, $Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{*(r)}) = E\left[\ell_c(\boldsymbol{\theta}^*)|\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n; \boldsymbol{\theta}^{*(r)}\right]$, can be written as

$$Q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{*(r)}) = \sum_{i=1}^n \left\{Q_{i1}\big(\beta_0, \beta_1|\boldsymbol{\theta}^{*(r)}\big) + Q_{i2}\big(\gamma, \psi, \xi|\boldsymbol{\theta}^{*(r)}\big) + Q_{i3}\big(\eta_x, \eta_e|\boldsymbol{\theta}^{*(r)}\big)\right\}, \tag{4.17}$$

where

$$\begin{aligned}
Q_{i1}\big(\beta_0, \beta_1|\boldsymbol{\theta}^{*(r)}\big) =& -\frac{1}{2}(\boldsymbol{z}_i - \boldsymbol{a})^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{z}_i - \boldsymbol{a})E_r[U_i/V_i] \\
& + (\boldsymbol{z}_i - \boldsymbol{a})^\top \boldsymbol{\Sigma}_i^{-1}\boldsymbol{b}E_r[x_iU_i/V_i] - \frac{1}{2}\boldsymbol{b}^\top \boldsymbol{\Sigma}_i^{-1}\boldsymbol{b}E_r[x_i^2U_i/V_i], \\
Q_{i2}\big(\gamma, \psi, \xi|\boldsymbol{\theta}^{*(r)}\big) =& -\frac{1}{2}\log(\psi) - \frac{1}{2\psi}\{E_r[x_i^2U_i] + \xi^2E_r[U_i] + \gamma^2E_r[D_i^2U_i]
\end{aligned}$$

$$
\begin{aligned}
&\quad -2(\xi E_r[x_iU_i]+\gamma E_r[x_iD_iU_i]-\xi\gamma E_r[D_iU_i])\},\\
Q_{i3}\big(\eta_x,\eta_e|\boldsymbol{\theta}^{*(r)}\big) &= \frac{1}{2\eta_x}\log(1/2\eta_x)-\log(\Gamma(1/2\eta_x))+\frac{1}{2\eta_x}\{E_r[\log(U_i)]-E_r[U_i]\}\\
&\quad +\frac{1}{2\eta_e}\log(1/[2c(\eta_e)])-\log(\Gamma(1/2\eta_e))+\frac{1}{2\eta_e}E_r[\log(U_i/V_i)]\\
&\quad -\frac{1}{2c(\eta_e)}E_r[U_i/V_i]
\end{aligned}
$$

and $E_r[\cdot]$ denotes the expectation under the conditional distribution of $\boldsymbol{Z}_{\text{mis},i}|\ \boldsymbol{Z}_i$ evaluated at $\boldsymbol{\theta}^*=\boldsymbol{\theta}^{*(r)}$.

The E and the CM steps of the ECM algorithm are given in detail as follows:

E step Compute all the conditional expectations involved in (4.17) (see Section 4.2.2.1 for more details).

CM step 1 Fix $\beta_1=\beta_1^{(r)}$ and update β_0 by maximizing $\sum_{i=1}^n Q_{i1}(\beta_0,\beta_1^{(r)}|\ \boldsymbol{\theta}^{*(r)})$ with respect to β_0, yielding

$$
\beta_0^{(r+1)}=\left(\sum_{i=1}^n\sigma_{\varepsilon_i}^{-2}E_r[U_i/V_i]\right)^{-1}\sum_{i=1}^n\sigma_{\varepsilon_i}^{-2}\big(Y_iE_r[U_i/V_i]-\beta_1^{(r)}E_r[x_iU_i/V_i]\big).
$$

CM step 2 Fix $\beta_0=\beta_0^{(r+1)}$ and update β_1 by maximizing $\sum_{i=1}^n Q_{i1}\big(\beta_0^{(r+1)},\beta_1|\boldsymbol{\theta}^{*(r)}\big)$ with respect to β_1, yielding

$$
\beta_1^{(r+1)}=\left(\sum_{i=1}^n\sigma_{\varepsilon_i}^{-2}E_r[x_i^2U_i/V_i]\right)^{-1}\sum_{i=1}^n\sigma_{\varepsilon_i}^{-2}\big(Y_i-\beta_0^{(r+1)}\big)E_r[x_iU_i/V_i].
$$

CM step 3 Fix $(\psi,\xi)^\top=(\psi^{(r)},\xi^{(r)})^\top$ and update γ by maximizing $\sum_{i=1}^n Q_{i2}\big(\gamma,\psi^{(r)},\xi^{(r)}|\boldsymbol{\theta}^{*(r)}\big)$ with respect to γ, yielding

$$
\gamma^{(r+1)}=\left(\sum_{i=1}^n E_r[D_i^2U_i]\right)^{-1}\sum_{i=1}^n\big(E_r[x_iD_iU_i]-\xi^{(r)}E_r[D_iU_i]\big).
$$

CM step 4 Fix $(\gamma,\xi)^\top=(\gamma^{(r+1)},\xi^{(r)})^\top$ and update ψ by maximizing $\sum_{i=1}^n Q_{i2}\big(\gamma^{(r+1)},\psi,\xi^{(r)}|\ \boldsymbol{\theta}^{*(r)}\big)$ with respect to ψ, yielding

$$
\begin{aligned}
\psi^{(r+1)}=&\frac{1}{n}\sum_{i=1}^n\big[E_r[x_i^2U_i]+(\xi^{(r)})^2E_r[U_i]+\big(\gamma^{(r+1)}\big)^2E_r[D_i^2U_i]-2(\xi^{(r)}E_r[x_iU_i]\\
&+\gamma^{(r+1)}E_r[x_iD_iU_i]-\xi^{(r)}\gamma^{(r+1)}E_r[D_iU_i])\big].
\end{aligned}
$$

CM step 5 Fix $(\gamma,\psi)^\top=(\gamma^{(r+1)},\psi^{(r+1)})^\top$ and update ξ by maximizing $\sum_{i=1}^n Q_{i2}(\gamma^{(r+1)},\psi^{(r+1)},\xi|\boldsymbol{\theta}^{*(r)})$ with respect to ξ, yielding

$$
\xi^{(r+1)}=\left(\sum_{i=1}^n E_r[U_i]\right)^{-1}\sum_{i=1}^n\big(E_r[x_iU_i]-\gamma^{(r+1)}E_r[D_iU_i]\big).
$$

CM step 6 Update (η_x, η_e) by numerically maximizing $\sum_{i=1}^{n} Q_{i3}\left(\eta_x, \eta_e | \boldsymbol{\theta}^{*(r)}\right)$ over (η_x, η_e) to get $(\eta_x^{(r+1)}, \eta_e^{(r+1)})$.

It is worth emphasizing that this ECM algorithm can be adapted to fit particular cases of the STcT MEM, as described in Section 4.2.3. Additionally, the algebraic expressions for $\beta_0^{(r+1)}$, $\beta_1^{(r+1)}$, $\gamma^{(r+1)}$, $\psi^{(r+1)}$ and $\xi^{(r+1)}$ are derived in Section 4.2.2.2. When a numerical maximization is required, a good idea is to transform the parameters from a constrained space to an unconstrained one. Moreover, it is well-known that for this type of algorithm, one needs to run the ECM algorithm with several starting points to have some assurance that the algorithm converges to a global maximum $\hat{\boldsymbol{\theta}}$.

Next, let $\boldsymbol{I} = -\partial^2 \ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ denote the observed information matrix for $\boldsymbol{\theta}$, which is evaluated by numerically differentiating $\ell(\boldsymbol{\theta})$ in (4.15) using, for instance, the `numDeriv` package (GILBERT; VARADHAN, 2016). From the asymptotic properties of the ML estimators, we have that $\hat{\boldsymbol{\theta}}$ is approximately normal with mean vector $\boldsymbol{\theta}$ and asymptotic covariance matrix $\boldsymbol{I}^{-1}$ (LEHMANN, 1998, Chapter 7). Consequently, this result can be used to compute asymptotic standard errors for the ML estimators, hypothesis testing and construct asymptotic confidence intervals.

4.2.2.1 Computing expectations in the E step

In this section, we describe how to compute all the expectations involved in (4.17). We omit the subscript of each random variable to simplify the notation. Next, our strategy is to obtain the joint distribution of the missing vector $(x, D, U, V)^\top$ conditional on the observed $\boldsymbol{Z}$ and then, use it to derive the desired expectations in the E step of the ECM algorithm. Define, for $v > 0$,

$$\boldsymbol{\Omega}_v = \psi\boldsymbol{b}\boldsymbol{b}^\top + v\boldsymbol{\Sigma},\ \rho_v = (\psi^{-1} + \boldsymbol{b}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{b}/v)^{-1} \quad \text{and} \quad \zeta_v^2 = (1+\gamma^2\boldsymbol{b}^\top\boldsymbol{\Omega}_v^{-1}\boldsymbol{b})^{-1}. \tag{4.18}$$

Using well-known matrix results (SEBER; LEE, 2003, p. 467), one can verify that $\boldsymbol{\Omega}_v$ and ρ_v are related by $\rho_v = \psi - \psi^2\boldsymbol{b}^\top\boldsymbol{\Omega}_v^{-1}\boldsymbol{b}$ and $\rho_v\boldsymbol{b}^\top\boldsymbol{\Sigma}^{-1}/v = \psi\boldsymbol{b}^\top\boldsymbol{\Omega}_v^{-1}$. In addition, ξ_v given in (4.11) can be written in terms of (4.18) as $\xi_v = \zeta_v\gamma\boldsymbol{b}^\top\boldsymbol{\Omega}_v^{-1}\boldsymbol{\Delta}$. Therefore, the following proposition can be established.

Proposition 5. Consider $\boldsymbol{Z}$ as defined in (4.2), (4.18) and the quantities given in Proposition 4.

(a) The conditional pdf of $V|\boldsymbol{Z}$ is $f(v|\boldsymbol{z};\boldsymbol{\theta}) = f(\boldsymbol{z}, v;\boldsymbol{\theta})/f(\boldsymbol{z};\boldsymbol{\theta})$, $v > 0$.

(b) The conditional pdf of $U|\boldsymbol{Z}, V$ is $f(u|\boldsymbol{z}, v;\boldsymbol{\theta}) = f(\boldsymbol{z}, u, v;\boldsymbol{\theta})/f(\boldsymbol{z}, v;\ \boldsymbol{\theta})$, $u, v > 0$.

(c) $D|\boldsymbol{Z}, U, V \sim \mathscr{T}\mathscr{N}(\xi_V\zeta_V, \zeta_V^2/U; (0,\infty))$.

(d) $x|\boldsymbol{Z}, D, U, V \sim \mathscr{N}(\xi + \psi\boldsymbol{b}^\top\boldsymbol{\Omega}_V^{-1}(\boldsymbol{Z} - \boldsymbol{a} - \boldsymbol{b}\xi) + \rho_V\psi^{-1}\gamma D,\ \rho_V/U)$.

Proof. For (a) and (b), we apply the definition of conditional pdf and using the density functions (b)–(d) given in Proposition 4, so the results hold. For (c), from (4.6), we obtain $\boldsymbol{Z}|D, U, V \sim$

$\mathcal{N}_2(\boldsymbol{a}+\boldsymbol{b}\xi+\boldsymbol{b}\gamma D, \boldsymbol{\Omega}_V/U)$ and $D|U,V \sim \mathcal{TN}(0,1/U;(0,+\infty))$. Now, by applying the definition of conditional pdf to $D|\boldsymbol{Z},U,V$, it follows that $f(d|\boldsymbol{z},u,v;\boldsymbol{\theta}) \propto f(\boldsymbol{z}|d,u,v;\boldsymbol{\theta})f(d|u,v;\boldsymbol{\theta}) = 2\phi_2(\boldsymbol{z};\boldsymbol{a}+\boldsymbol{b}\xi+\boldsymbol{b}\gamma d, \boldsymbol{\Omega}_v/u)\phi(d;0,1/u)I(d>0)$ and by applying Lemma 2 in Arellano-Valle, Bolfarine and Lachos (2005) to the product on the right side of equality, we obtain $f(d|\boldsymbol{z},u,v;\boldsymbol{\theta}) \propto \phi_2(\boldsymbol{z};\boldsymbol{a}+\boldsymbol{b}\xi,(\Omega_v+\gamma^2\boldsymbol{b}\boldsymbol{b}^\top)/u)2\phi(d;\xi_v\zeta_v,\zeta_v^2/u)I(d>0)$. Then, we can see that the pdf of $D|\boldsymbol{Z},U,V$ is proportional to the pdf of the $\mathcal{TN}(\xi_V\zeta_V\zeta_V^2/U;(0,\infty))$ distribution evaluated at d. Therefore, this establishes the desired conditional distribution. For (d), from a similar manner to (c), we use (4.6) to obtain $\boldsymbol{Z}|x,D,U,V \sim \mathcal{N}_2(\boldsymbol{a}+\boldsymbol{b}x, V\boldsymbol{\Sigma}/U)$ and $x|D,U \sim \mathcal{N}(\xi+\gamma D,\psi/U)$. By applying the definition of conditional pdf to $x|\boldsymbol{Z},D,U,V$, it follows that $f(x|\boldsymbol{z},d,u,v;\boldsymbol{\theta}) \propto f(\boldsymbol{z}|x,d,u,v;\boldsymbol{\theta})f(x|d,u;\boldsymbol{\theta})$. Then, applying Lemma 2 in Arellano-Valle, Bolfarine and Lachos (2005), we obtain that $f(x|\boldsymbol{z},d,u,v;\boldsymbol{\theta}) \propto \phi_2(\boldsymbol{z};\boldsymbol{a}+\boldsymbol{b}\xi+\boldsymbol{b}\gamma d,\Omega_v/u)\ \phi(x;\xi+\gamma d+\rho_v\boldsymbol{b}^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{a}-\boldsymbol{b}\xi-\boldsymbol{b}\gamma d)/v,\rho_v/u)$. Thus, using (4.18), we simplify the mean in the second pdf. We notice that the pdf of $x|\boldsymbol{Z},D,U,V$ is proportional to the pdf of the $\mathcal{N}(\xi+\psi\boldsymbol{b}^\top\boldsymbol{\Omega}_v^{-1}(\boldsymbol{z}-\boldsymbol{a}-\boldsymbol{b}\xi)+\rho_V\psi^{-1}\gamma d,\rho_V/U)$ distribution evaluated at x. Therefore, it follows that this must also be the conditional distribution of $x|\boldsymbol{Z},D,U,V$ and the result holds. □

Proposition 6. Consider $\boldsymbol{Z}$ as defined in (4.2), (4.6) and the quantities in Proposition 5. Then, we have the following conditional expectations on $\boldsymbol{Z}$ and V.

(a) For an integer k such that $\eta^*+2k>0$,

$$E[U^k|\boldsymbol{Z},V] = 2^k\frac{\Gamma((\eta^*+2k)/2)}{\Gamma(\eta^*/2)}\frac{F_t(\xi_V[(\eta^*+2k)/\varsigma_V]^{1/2};\eta^*+2k)}{\varsigma_V^kF_t(\xi_V[\eta^*/\varsigma_V]^{1/2};\eta^*)}.$$

(b) For an integer k such that $\eta^*+k>0$,

$$E\left[U^{k/2}\frac{\phi(\xi_VU^{1/2})}{\Phi(\xi_VU^{1/2})}|\boldsymbol{Z},V\right] = \frac{2^{(k-1)/2}}{\pi^{1/2}}\frac{\Gamma((\eta^*+k)/2)}{\Gamma(\eta^*/2)} \times\frac{\varsigma_V^{\eta^*/2}}{(\varsigma_V+\xi_V^2)^{(\eta^*+k)/2}F_t(\xi_V[\eta^*/\varsigma_V]^{1/2};\eta^*)}.$$

(c) For an integer k such that $\eta^*+2k>1$,

$$E[DU^k|\boldsymbol{Z},V] = \zeta_V\left\{\xi_VE[U^k|\boldsymbol{Z},V]+E\left[U^{(2k-1)/2}\frac{\phi(\xi_VU^{1/2})}{\Phi(\xi_VU^{1/2})}|\boldsymbol{Z},V\right]\right\}.$$

(d) For an integer k such that $\eta^*+2k>2$,

$$E[D^2U^k|\boldsymbol{Z},V] = \zeta_V^2E[U^{k-1}|\boldsymbol{Z},V]+\xi_V\zeta_VE[DU^k|\boldsymbol{Z},V].$$

(e) $E[xDU|\boldsymbol{Z},V] = \left(\xi+\psi\boldsymbol{b}^\top\boldsymbol{\Omega}_V^{-1}\boldsymbol{\Delta}\right)E[DU|\boldsymbol{Z},V]+\rho_V\psi^{-1}\gamma E[D^2U|\boldsymbol{Z},V]$.

(f) $E[xU|\boldsymbol{Z},V] = \left(\xi+\psi\boldsymbol{b}^\top\boldsymbol{\Omega}_V^{-1}\boldsymbol{\Delta}\right)E[U|\boldsymbol{Z},V]+\rho_V\psi^{-1}\gamma E[DU|\boldsymbol{Z},V]$.

(g) $E[x^2U|\mathbf{Z},V] = (\xi + \mathbf{\Delta}^\top\mathbf{\Sigma}^{-1}\boldsymbol{b}\rho_V/V)E[xU|\mathbf{Z},V] + \{1+\psi^{-1}\gamma E[xDU|\mathbf{Z},V]\}\rho_V.$

Proof. First, using the part (b) of Proposition 5 together with parts (b) and (c) of Proposition 4, it follows that

$$f(u|\boldsymbol{z},v,\boldsymbol{\theta}) = \frac{\Phi(\xi_v u^{1/2})}{F_t(\xi_v[\eta^*/\varsigma_v]^{1/2},\eta^*)} f(u;\eta^*/2,\varsigma_v/2), \tag{4.19}$$

where $f(u;\eta^*/2,\varsigma_v/2)$ is the pdf of the $\mathscr{G}(\eta^*/2,\varsigma_v/2)$ distribution evaluated at u. Thus, for (a), from the definition of expectation and the pdf in (4.19), we can write

$$E[U^k|\boldsymbol{z},v] = \int_0^\infty u^k f(u|\boldsymbol{z},v,\boldsymbol{\theta})du = \frac{1}{F_t(\xi_v[\eta^*/\varsigma_v]^{1/2},\eta^*)}\int_0^\infty u^k\Phi(\xi_v u^{1/2})f(u;\eta^*/2,\varsigma_v/2)du.$$

Note that the integral on the right side is proportional to the expectation of $\Phi(\xi_v U^{1/2})$, wherein the random variable U follows the $\mathscr{G}((\eta^*+2k)/2),\varsigma_v/2)$ distribution. Then, by applying Lemma 1 in Azzalini and Capitanio (2003), we have after lengthy algebra that the result follows.

For (b), from the definition of conditional expectation and the pdf in (4.19), it follows that

$$\begin{aligned} E\left[U^{k/2}\frac{\phi(\xi_v U^{1/2})}{\Phi(\xi_v U^{1/2})}|\boldsymbol{z},v\right] &= \int_0^\infty u^{k/2}\frac{\phi(\xi_v u^{1/2})}{\Phi(\xi_v u^{1/2})} f(u|\boldsymbol{z},v,\boldsymbol{\theta})du \\ &= \frac{1}{F_t(\xi_v[\eta^*/\varsigma_v]^{1/2},\eta^*)}\int_0^\infty u^{k/2}\phi(\xi_v u^{1/2})f(u;\eta^*/2,\varsigma_v/2)du, \end{aligned}$$

so the integral in the second equality is proportional to the pdf of the $\mathscr{G}((\eta^*+k)/2),(\varsigma_v+\xi_v)^2/2)$ distribution evaluated at u. After some algebraic manipulations, the result follows by making use of the gamma integral.

For (c), using the law of iterated expectations, we can write

$$E[DU^k|\boldsymbol{z},v] = E[U^k E[D|\boldsymbol{z},v,U]|\boldsymbol{z},v]. \tag{4.20}$$

Using part (c) of Proposition 5 and properties about moments of a truncated normal random variable (JOHNSON; KOTZ; BALAKRISHNAN, 1994, Section 10.1), we can write the inner expectation as

$$E[D|\boldsymbol{z},v,u] = \zeta_v\left(\xi_v + u^{-1/2}\frac{\phi(\xi_v u^{1/2})}{\Phi(\xi_v u^{1/2})}\right).$$

The result now follows after substituting this expression into (4.20).

For (d), as in (c), we can write

$$E[D^2U^k|\boldsymbol{z},v] = E[U^k E[D^2|\boldsymbol{z},v,U]|\boldsymbol{z},v]. \tag{4.21}$$

Then, applying part (c) of Proposition 6 and properties about moments of a truncated normal random variable (JOHNSON; KOTZ; BALAKRISHNAN, 1994, Section 10.1), the inner expectation can be written as the recursive relationship, $E[D^2|\boldsymbol{z},v,u] = \zeta_v^2/u + \xi_v\zeta_v E[D|\boldsymbol{z},v,u]$. The result now follows after substituting this expression into (4.21).

For (e), using the law of iterated expectations, we can write

$$E[xDU|\boldsymbol{z},v] = E[U\, E[D\, E[\, x\,|\boldsymbol{z},v,D,U]|\boldsymbol{z},v,U]|\boldsymbol{z},v]. \tag{4.22}$$

Moreover, part (d) of Proposition 5 yields

$$E[x|\boldsymbol{z},v,u,d] = \xi + \psi\boldsymbol{b}^\top\boldsymbol{\Omega}_v^{-1}(\boldsymbol{z}-\boldsymbol{a}-\boldsymbol{b}\xi)+\rho_v\psi^{-1}\gamma d.$$

In this equality, we proceed by multiplying both sides by d and taking expectation with respect to the conditional distribution of $D|\boldsymbol{z},v,u$, it follows that

$$E[xD|\boldsymbol{z},v,u] = \xi E[D|\boldsymbol{z},v,u]+\rho_v\psi^{-1}\gamma E[D^2|\boldsymbol{z},v,u]+\psi\boldsymbol{b}^\top\Omega_v^{-1}(\boldsymbol{z}-\boldsymbol{a}-\boldsymbol{b}\xi)E[D|\boldsymbol{z},v,u].$$

The result follows from substituting the expression above into (4.22). In a similar manner, we can proceed to establish (f).

For (g), as in (4.22), we can write

$$E[x^2U|\boldsymbol{z},v] = E[U\, E[\, E[\, x^2\,|\boldsymbol{z},v,D,U]|\boldsymbol{z},v,U]|\boldsymbol{z},v]. \tag{4.23}$$

Next, in the inner expectation of (4.23), we apply the following property,

$$E[\, x^2\,|\boldsymbol{z},v,D,U] = Var[x|\boldsymbol{z},v,D,U]+(E[x|\boldsymbol{z},v,D,U])^2,$$

and applying part (d) of Proposition 5, it follows that

$$E[\, x^2\,|\boldsymbol{z},v,D,U] = \rho_v/u+\{\xi+\rho_v\psi^{-1}\gamma d+\psi\boldsymbol{b}^\top\boldsymbol{\Omega}_v^{-1}(\boldsymbol{z}-\boldsymbol{a}-\boldsymbol{b}\xi)\}E[x|\boldsymbol{z},d,u,v].$$

Lastly, after substituting this expression into (4.23), we obtain

$$E[x^2U|\boldsymbol{z},v] = \rho_v+\xi E[xU|\boldsymbol{z},v]+\rho_v\psi^{-1}\gamma E[xU|\boldsymbol{z},v]+\psi\boldsymbol{b}^\top\boldsymbol{\Omega}_v^{-1}(\boldsymbol{z}-\boldsymbol{a}-\boldsymbol{b}\xi)E[xDU|\boldsymbol{z},v],$$

and using $\boldsymbol{\Omega}_v^{-1}\boldsymbol{b}\psi = \boldsymbol{\Sigma}^{-1}\boldsymbol{b}\rho_v/v$, the result follows after simplification. □

The Propositions 5 and 6 suggest that the expectations involved in the E step can be computed as follows. The two expectations, $\mathrm{E}[\log(V)|\boldsymbol{Z}] = \int_0^\infty \log(v)f(v|z,\boldsymbol{\theta})dv$ and $\mathrm{E}[\log(U)|\boldsymbol{Z}] = \int_0^\infty[\int_0^\infty \log(u)f(u|\boldsymbol{z},v,\boldsymbol{\theta})du]f(v|\boldsymbol{z},\boldsymbol{\theta})dv$, need to be computed numerically with the pdfs given in Proposition 5. Finally, to compute expectations required in the E step of the ECM algorithm in Section 4.2.2, first we compute the expectation conditional on $(\boldsymbol{Z}^\top,V)^\top$ using Proposition 6 and then average it over the conditional distribution of $(V|\boldsymbol{Z})$. The computation requires a unidimensional numerical integration and here it was computed using the `statmod` package (GINER; SMYTH, 2016) implemented in the R programing language (R CORE TEAM, 2017). All the expectations are evaluated at $\boldsymbol{\theta}^* = \boldsymbol{\theta}^{*(r)}$.

4.2.2.2 Computation of maximizers in CM steps

In this section, we describe how to derive the maximizers in CM steps presented in Section 4.2.2. The expressions for $\beta_0^{(r+1)}$ and $\beta_1^{(r+1)}$ in the CM steps 1 and 2, respectively, are obtained by solving the equations

$$\begin{aligned}0 =& \frac{\partial}{\partial\beta_0}\sum_{i=1}^{n} Q_{i1}(\beta_0,\beta_1^{(r)}|\boldsymbol{\theta}^{*(r)})\\ =& \sum_{i=1}^{n}\left(\sigma_{\varepsilon_i}^{-2}Y_i-\sigma_{\varepsilon_i}^{-2}\beta_0\right)E_r[U_i/V_i]-\beta_1^{(r)}\sum_{i=1}^{n}\sigma_{\varepsilon_i}^{-2}E_r[x_iU_i/V_i]\\ =& \sum_{i=1}^{n}\sigma_{\varepsilon_i}^{-2}\left(Y_iE_r[U_i/V_i]-\beta_1^{(r)}E_r[x_iU_i/V_i]\right)-\beta_0\sum_{i=1}^{n}\sigma_{\varepsilon_i}^{-2}E_r[U_i/V_i],\end{aligned}$$

for β_0, and

$$0=\frac{\partial}{\partial\beta_1}\sum_{i=1}^{n} Q_{i1}(\beta_0^{(r+1)},\beta_1|\boldsymbol{\theta}^{*(r)})=\sum_{i=1}^{n}\sigma_{\varepsilon_i}^{-2}(Y_i-\beta_0^{(r+1)})E_r[x_iU_i/V_i]-\beta_1\sum_{i=1}^{n}\sigma_{\varepsilon_i}^{-2}E_r[x_i^2U_i/V_i],$$

for β_1. Similarly, to get the expression for $\gamma^{(r+1)}$ in CM step 3, we solve for γ the equation,

$$0=\frac{\partial}{\partial\gamma}\sum_{i=1}^{n} Q_{i2}(\gamma,\psi^{(r)},\xi^{(r)}|\boldsymbol{\theta}^{*(r)})=(\psi^{(r)})^{-1}\sum_{i=1}^{n}\left(E_r[x_iD_iU_i]-\xi^{(r)}E_r[D_iU_i]\right)-\gamma\sum_{i=1}^{n}E_r[D_i^2U_i],$$

after some algebraic manipulations the result follows by multiplying both sides by $\psi^{(r)}$. To derive the expression for $\psi^{(r+1)}$ in CM step 4, we solve the following equation

$$0=\frac{\partial}{\partial\psi}\sum_{i=1}^{n} Q_{i2}(\gamma^{(r+1)},\psi,\xi^{(r)}|\boldsymbol{\theta}^{*(r)})=-\frac{n}{\psi}+\frac{1}{\psi^2}\sum_{i=1}^{n}A_i,$$

where $A_i=\{E_r[x_i^2U_i]+(\gamma^{(r+1)})^2E_r[D_i^2U_i]+(\xi^{(r)})^2E_r[U_i]-2(\xi^{(r)}E_r[x_iU_i]+\gamma^{(r+1)}E_r[x_iD_iU_i]-\xi^{(r)}\gamma^{(r+1)}E_r[D_iU_i])\}$, upon simplification. The result follows by multiplying both sides with $(\psi^{(r)})^2$.

Lastly, the expression for $\xi^{(r+1)}$ in CM step 5 is obtained by solving the equation,

$$0=\frac{\partial}{\partial\xi}\sum_{i=1}^{n} Q_{i2}(\gamma^{(r+1)},\psi^{(r+1)},\xi|\boldsymbol{\theta}^{*(r)})=(\psi^{(r+1)})^{-1}\sum_{i=1}^{n}(-E_r[U_i]\xi+E_r[x_iU_i]-\gamma^{(r+1)}E_r[D_iU_i]).$$

4.2.3 Particular models

From the stochastic representations in (4.4), it becomes easy to obtain some particular models of the STcT MEM. According to the distributions of the true covariate and the error terms, we label each model by the initial letter of the assumed distribution. In this way, we describe them in detail in what follows.

(a) Skew-t normal (STN) MEM. We take $\eta_e \to 0$. Thus, $\boldsymbol{e}_i \sim \mathscr{N}_2(\mathbf{0}_2,\boldsymbol{\Sigma}_i)$ and $x_i \sim \mathscr{S}t(\xi,\omega^2,\lambda,\eta_x)$. A hierarchical representation is given by

$$\boldsymbol{Z}_i|U_i \sim \mathscr{SN}_2(\boldsymbol{a}+\boldsymbol{b}\xi,\ \boldsymbol{\Pi}_{U_i}/U_i,\boldsymbol{\lambda}_{U_i}) \quad \text{and} \quad U_i \sim \mathscr{G}\left(\frac{1}{2\eta_x},\frac{1}{2\eta_x}\right),$$

where, we define, for $u_i > 0$, $\boldsymbol{\Pi}_{u_i} = u_i\boldsymbol{\Sigma}_i + \omega^2\boldsymbol{b}\boldsymbol{b}^\top$ and $\boldsymbol{\lambda}_{u_i} = \boldsymbol{\Pi}_{u_i}^{-1/2}\boldsymbol{b}\omega\lambda/[1+\lambda^2\omega^{-2}(\omega^{-2}+\boldsymbol{b}^\top\boldsymbol{\Sigma}_i^{-1}\boldsymbol{b}/u_i)^{-1}]^{1/2}$. In this case, the joint pdf of $(\boldsymbol{Z}_i^\top, U_i)^\top$ is

$$f(\boldsymbol{z}_i, u_i; \boldsymbol{\theta}) = \frac{\pi^{-1}}{(2\eta_x)^{1/2\eta_x}} \frac{[\det(\boldsymbol{\Pi}_{u_i})]^{-1/2}}{\Gamma(1/2\eta_x)} u_i^{1/2\eta_x} \exp\left(-\frac{1}{2}\varsigma_{u_i}u_i\right)\Phi(\xi_{u_i}u_i^{1/2}),$$

where $\xi_{u_i} = \boldsymbol{\lambda}_{u_i}^\top\boldsymbol{\Pi}_{u_i}^{-1/2}\boldsymbol{\Delta}_i$, $\varsigma_{u_i} = \boldsymbol{\Delta}_i^\top\boldsymbol{\Pi}_{u_i}^{-1}\boldsymbol{\Delta}_i + 1/\eta_x$ and $\boldsymbol{\Delta}_i$ as in (4.11). Therefore, the marginal pdf of $\boldsymbol{Z}_i$ is obtained by integrating $f(\boldsymbol{z}_i, u_i; \boldsymbol{\theta})$ with respect to u_i.

(b) Student's t centered Student's t (TcT) MEM. We take $\lambda = 0$, meaning that there is no skewness, so that $x_i \sim t(\xi, \omega^2, \eta_x)$ and $\boldsymbol{e}_i \sim ct_2(\boldsymbol{0}_2, \boldsymbol{\Sigma}_i, \eta_e)$ implying that $(\gamma, \psi)^\top = (0, \omega^2)^\top$. Thus, substituting $\lambda = 0$ into the last term of the right-hand side in (4.13), we obtain $1/2$. After simplification, the joint pdf of $(\boldsymbol{Z}_i, V_i)$ follows for this case. Consequently, the marginal pdf of $\boldsymbol{Z}_i$ is obtained by integrating the resulting joint pdf with respect to v_i.

(c) Student's t normal (TN) MEM. We take $\lambda = 0$ and $\eta_e \to 0$. Thus, $x_i \sim t(\xi, \omega^2, \eta_x)$ and $\boldsymbol{e}_i \sim \mathcal{N}_2(\boldsymbol{0}_2, \boldsymbol{\Sigma}_i)$. The joint pdf of $(\boldsymbol{Z}_i, U_i)$ is given by

$$f(\boldsymbol{z}_i, u_i; \boldsymbol{\theta}) = \frac{\pi^{-1}[\det(\boldsymbol{\Pi}_{u_i})]^{-1/2}}{2^{1/2\eta_x+1}\eta_x^{1/2\eta_x}} \frac{u_i^{1/2\eta_x}}{\Gamma(1/2\eta_x)} \exp\left(-\frac{1}{2}\varsigma_{u_i}u_i\right),$$

and consequently, by integrating $f(\boldsymbol{z}_i, u_i; \boldsymbol{\theta})$ with respect to u_i we get the marginal pdf of $\boldsymbol{Z}_i$. Here, $\boldsymbol{\Pi}_{u_i}$ and ς_{u_i} are defined as in (a).

(d) Skew-normal centered Student's t (SNcT) MEM: We take $\eta_x \to 0$. Thus, $x_i \sim \mathcal{SN}(\xi, \omega^2, \lambda)$ and $\boldsymbol{e}_i \sim ct_2(\boldsymbol{0}_2, \boldsymbol{\Sigma}_i, \eta_e)$. A hierarchical representation is given by

$$\boldsymbol{Z}_i|V_i \sim \mathcal{SN}_2(\boldsymbol{a}+\boldsymbol{b}\xi,\, \boldsymbol{\Pi}_{v_i}, \boldsymbol{\lambda}_{v_i}) \quad \text{and} \quad \frac{1}{V_i} \sim \mathcal{G}\left(\frac{1}{2\eta_e}, \frac{1}{2c(\eta_e)}\right),$$

and the joint pdf of $(\boldsymbol{Z}_i, V_i)$ is as follows

$$f(\boldsymbol{z}_i, v_i; \boldsymbol{\theta}) = \frac{\pi^{-1}[\det(\boldsymbol{\Pi}_{v_i})]^{-1/2}}{[2c(\eta_e)]^{1/2\eta_e}\Gamma(1/2\eta_e)} \frac{1}{v_i^{1/2\eta_e+1}} \exp\left(-\frac{1}{2}\left[\varsigma_{v_i}^* + \frac{1}{c(\eta_e)}\frac{1}{v_i}\right]\right)\Phi(\xi_{v_i}),$$

where $\varsigma_{v_i}^* = \boldsymbol{\Delta}_i^\top\boldsymbol{\Pi}_{v_i}^{-1}\boldsymbol{\Delta}_i$ with $\boldsymbol{\Pi}_{v_i}^{-1}$ and $\boldsymbol{\Delta}_i$ as in (4.11). The marginal pdf of $\boldsymbol{Z}_i$ is obtained integrating $f(\boldsymbol{z}_i, v_i; \boldsymbol{\theta})$ with respect to v_i.

(e) Skew-normal normal (SNN) MEM: We take $\eta_x \to 0$ and $\eta_e \to 0$. Thus, $x_i \sim \mathcal{SN}(\xi, \omega^2, \lambda)$ and $\boldsymbol{e}_i \sim \mathcal{N}_2(\boldsymbol{0}_2, \boldsymbol{\Sigma}_i)$. In this case, the marginal pdf of $\boldsymbol{Z}_i$ is reduced to

$$f(\boldsymbol{z}_i; \boldsymbol{\theta}) = 2\phi(\boldsymbol{z}_i; \boldsymbol{a}+\boldsymbol{b}\xi, \boldsymbol{\Pi}_i, \boldsymbol{\lambda}_i)\Phi(\boldsymbol{\lambda}_i^\top\boldsymbol{\Pi}_i^{-1/2}(\boldsymbol{z}_i - \boldsymbol{a} - \boldsymbol{b}\xi)),$$

where $\boldsymbol{\Pi}_i = \omega^2\boldsymbol{b}^\top\boldsymbol{b} + \boldsymbol{\Sigma}_i$ and $\boldsymbol{\lambda}_i$ is obtained by using $\boldsymbol{\lambda}_{v_i}$ in (4.11) replacing $\boldsymbol{\Pi}_{v_i}$ by $\boldsymbol{\Pi}_i$. This model was also described as a particular case of the MEM in Arellano-Valle *et al.* (2005).

(f) Normal centered Student's t MEM (NcT MEM): We take $\lambda = 0$ and $\eta_x \to 0$, so that $x_i \sim \mathcal{N}(\xi, \omega^2)$ and $\boldsymbol{e}_i \sim ct_2(\mathbf{0}_2, \boldsymbol{\Sigma}_i, \eta_e)$. In this case, $\gamma = 0$ and the gamma distribution for W_{x_i} is not needed ($W_{x_i} = U_i \equiv 1$). A hierarchical representation is given by

$$\boldsymbol{Z}_i|x_i, V_i \sim \mathcal{N}_2(\boldsymbol{a} + \boldsymbol{b}\xi, V_i\boldsymbol{\Sigma}_i), \quad x_i \sim \mathcal{N}(\xi, \omega^2) \quad \text{and} \quad \frac{1}{V_i} \sim \mathcal{G}\left(\frac{1}{2\eta_e}, \frac{1}{2c(\eta_e)}\right).$$

The joint pdf of $(\boldsymbol{Z}_i, V_i)$ is as follows

$$f(\boldsymbol{z}_i, v_i; \boldsymbol{\theta}) = \frac{\pi^{-1}}{2^{1/2\eta_e+1}} \frac{[\det(\boldsymbol{\Pi}_{v_i})]^{-1/2}}{c(\eta_e)^{1/2\eta_e}\Gamma(1/2\eta_e)} \frac{1}{v_i^{1/2\eta_e+1}} \exp\left(-\frac{1}{2}\left[\varsigma^*_{v_i} + \frac{1}{c(\eta_e)}\frac{1}{v_i}\right]\right),$$

where $\boldsymbol{\Pi}_{v_i}$ as in (4.11) and $\varsigma^*_{v_i}$ as in (d). Hence, the marginal pdf of $\boldsymbol{Z}_i$ is obtained by integrating $f(\boldsymbol{z}_i, v_i; \boldsymbol{\theta})$ with respect to v_i.

(g) Normal normal (NN) MEM. We take $\lambda = 0$ and $\eta_x, \eta_e \to 0$. Thus, $x_i \sim \mathcal{N}(\xi, \omega^2)$ and $\boldsymbol{e}_i \sim \mathcal{N}_2(\mathbf{0}, \boldsymbol{\Sigma}_i)$. This case is the heteroscedastic MEM model under the normality assumption. The pdf of $\boldsymbol{Z}_i$ is given by $\phi_2(\boldsymbol{z}_i; \boldsymbol{a} + \boldsymbol{b}\xi, \boldsymbol{\Pi}_i)$, where $\boldsymbol{\Pi}_i = \omega^2\boldsymbol{b}\boldsymbol{b}^\top + \boldsymbol{\Sigma}_i$. As mentioned before, this model has attracted many attention (KULATHINAL; KUULASMAA; GASBARRA, 2002; BUONACCORSI, 2010).

It is worth emphasizing that up to the best of our knowledge, the particular models given in (a), (b), (c), (d) and (f) have not yet been proposed in the literature, so that they can be considered as flexible alternatives to the normality assumption. Moreover, it is important to note that the particular models given in (a), (c), (d)–(g) are limiting cases of the proposed model, while the model in (b) is a nested model.

Next, we briefly point out how the ECM algorithm was modified to fit each one of the particular models used in the Sections 4.3 and 4.4. Thus, we describe for the following models: (a) STN MEM, (b) TcT MEM, (c) TN MEM and (g) NN MEM. In case (a), the gamma distribution for $W_{e_i} = U_i/V_i$ is not needed, so that we can apply the algorithm after setting $U_i/V_i \equiv 1$ in the expected log-likelihood given in (4.16). In case (b), the algorithm works by setting $\gamma = 0$ in the expected log-likelihood given in (4.16) and omitting the CM step 3. In case (c), the modifications we need are as in cases (a) and (b) and can be applied after modifying the expected log-likelihood in (4.16). Lastly, in case (g), the gamma distribution for $W_{x_i} = U_i$ is not required, so that we can apply the ECM algorithm after setting $U_i \equiv 1$ and by using the modifications as in cases (a) and (b). For cases (a), (c) and (g), we assign to η_x and η_e, very small values to indicate that they tend to be close to 0. For instance, for an ECM algorithm to fit the NN MEM, it was considered $\lambda = 0$ and $\eta_x, \eta_e = 0.01$.

Finally, there exists a variety of methodologies used to compare competing models for a given dataset. The aim is to try to select the one that best fits the data. In this chapter, we consider the Akaike information criterion (AIC) and the Bayesian criterion selection (BIC) to compare the different models. These measures are computed by $AIC = -2\log L(\widehat{\theta};z) + 2k$ and $BIC = -2\log L(\widehat{\theta};z) + k\log(n)$, where $\widehat{\boldsymbol{\theta}}$ is the ML estimate, k is the number of parameters to be estimated and n denotes the sample size.

4.3 Simulation study

In this section, we carry out a simulation study to gauge the performance of the proposed methodology. We generated 1000 random samples of size $n = 50$, 100 and 500 from a STcT MEM. The observed data $\boldsymbol{Z}_i$ are generated from the representation (4.6) with the following parameter values: $\beta_0 = 0.1$, $\beta_1 = 1.1$, $\xi = 1$, $\omega^2 = 0.5$, $\lambda = 5$. The variances of the measurement errors, that is, $\sigma^2_{\varepsilon_i}$ and $\sigma^2_{u_i}$, are generated from uniform distributions on $(0.15, 0.25)$ and $(0.20, 0.35)$, respectively, and then $\boldsymbol{\Sigma}_i = \text{diag}(\sigma^2_{\varepsilon_i}, \sigma^2_{u_i})$ is assumed known, $i = 1, \ldots, n$. For the inverse of the degrees of freedom, we take $\eta_x = 1/10$ and $\eta_e = 1/5$, keeping them fixed throughout the simulations in order to save computing time. The value for β_1 was motivated by ML estimate from the real data set analyzed in Section 4.4.

For the STcT MEM, numerical summaries – including the simulated bias of the ML estimates, the average of the asymptotic standard errors (SE), the root mean squared error of the estimates (RMSE), the standard deviation (SD) of estimates and the coverage probabilities (CP) of the nominal 95% asymptotic confidence intervals – are reported in Table 13. Recall that the asymptotic standard errors are the square root of the main diagonal elements of the inverse of the observed information matrix ($\boldsymbol{I}^{-1}$). For each generated sample, ML estimates of the parameters were computed by using the ECM algorithm described in Section 4.2.2. The numerical derivatives and the expectations needed in the E step of ECM algorithm are computed, respectively, using the `numDeriv` package (GILBERT; VARADHAN, 2016) and the `statmod` package (GINER; SMYTH, 2016). All computations were developed in the R language (R CORE TEAM, 2017).

Table 13 shows that the simulated bias of the ML estimates decreases when the sample size increases, as desirable. There is a substantial reduction in the values of RMSE and SD and these are fairly close to the average of the asymptotic standard errors (SE), as expected. Note also that all CP's are fairly close to the nominal value 95%. However, for the estimation of the skewness parameter λ, a larger sample size is needed, as can be seen for $n = 500$. Regarding the estimation of this parameter, Arellano-Valle *et al.* (2005) and Kheradmandi and Rasekh (2015) mentioned that in some samples, the probability of having $\widehat{\lambda} = \infty$ can be positive. Alternative methods of estimation of λ are part of ongoing work.

For the sake of model comparison, the 1000 simulated data sets were also fitted under the following competing MEM's: STN, TcT , TN and NN. The efficiencies of STcT MEM-based estimators relative to the competing models are computed by dividing the mean squared error under the competing model by those under the STcT MEM. These measures are displayed in

Table 13 – Simulated bias, sample standard deviation (SD), root mean square error (RMSE), average asymptotic standard error (SE) of ML estimates and coverage probability (CP) of the nominal 95% asymptotic confidence intervals for the STcT MEM. True values: $\beta_0 = 0.1$, $\beta_1 = 1.1$, $\xi = 1$, $\omega^2 = 0.5$ and $\lambda = 5$.

n	Parameter	Bias	SD	RMSE	SE	CP
50	β_0	-0.027	0.380	0.381	0.381	0.941
	β_1	0.013	0.233	0.233	0.232	0.939
	ξ	0.088	0.237	0.253	0.227	0.964
	ω^2	-0.022	0.214	0.215	0.238	0.918
	λ	3.637	7.980	8.769	21.526	0.871
100	β_0	-0.008	0.240	0.240	0.255	0.955
	β_1	0.008	0.148	0.148	0.156	0.954
	ξ	0.034	0.134	0.138	0.141	0.967
	ω^2	-0.022	0.152	0.154	0.169	0.943
	λ	2.201	5.796	6.200	11.993	0.874
500	β_0	0.002	0.105	0.105	0.108	0.952
	β_1	0.000	0.063	0.063	0.066	0.958
	ξ	0.006	0.047	0.047	0.055	0.970
	ω^2	-0.006	0.064	0.064	0.076	0.963
	λ	0.388	2.200	2.233	3.368	0.937

Table 14. We also compute the Akaike information criterion (AIC) and the Bayesian information criterion (BIC).

Table 14 – Efficiencies of STcT MEM-based estimators relative to the competing models when the true model is STcT MEM.

n	Quantity	STN MEM	TcT MEM	TN MEM	NN MEM
50	β_0	1.661	0.990	1.503	1.448
	β_1	1.721	1.008	1.558	1.500
	$\mathrm{E}[X]$	1.138	4.174	4.340	4.810
	$\mathrm{E}[Y]$	1.179	4.628	4.851	5.405
100	β_0	1.945	1.048	1.928	1.719
	β_1	1.979	1.051	1.951	1.726
	$\mathrm{E}[X]$	1.371	10.43	10.99	12.38
	$\mathrm{E}[Y]$	1.471	12.54	13.06	14.71
500	β_0	1.521	1.093	1.695	1.578
	β_1	1.526	1.083	1.678	1.556
	$\mathrm{E}[X]$	1.385	67.60	70.90	80.49
	$\mathrm{E}[Y]$	1.349	82.00	85.89	97.44

Table 14 shows that relative efficiency improves as n increases. It is worth to note that the quantities $\beta_0, \beta_1, E[X]$ and $E[Y]$ have the same interpretation in any fitted models. Notice that for β_0 and β_1, the efficiency of STcT MEM-based estimators relative to the TcT MEM is close to 1. This tells us that models based on heavy-tails distributions also can deliver good estimates (CAO; LIN; ZHU, 2012). For the remaining quantities, the gain in efficiency is quite substantial, that is, the relative efficiencies values are greater than 1 in all cases, giving support to the STcT

MEM. Moreover, we report in Table 15 the percentage of times that AIC and BIC selected each model when the data generating model is the STcT MEM. As mentioned before, the TcT MEM can be considered as a competing model with respect to the STcT MEM. Observe that BIC was not favorable on the selection of the STcT MEM when $n = 50$. For this sample size, the penalty term into the expression of the BIC was not able to correctly identify the true model, for the maximum values of the likelihood function computed by the STcT MEM and TcT MEM were very close. When n increases, for most of the samples AIC and BIC selected the proposed model as the best one over the other competing models. Thus, it is important to note that, based on our results, these criteria will be reliable if the sample size is larger than 50.

Table 15 – Percentage of samples for which AIC and BIC selected each model when the STcT MEM is the true model.

		Model				
n	Criterion	STcT	STN	TcT	TN	NN
50	AIC	50.20	12.10	28.40	3.00	6.30
	BIC	32.60	6.20	45.60	6.10	9.50
100	AIC	75.60	9.70	12.70	0.40	1.60
	BIC	56.10	7.30	32.20	1.60	2.80
500	AIC	99.80	0.20	0.00	0.00	0.00
	BIC	99.20	0.20	0.60	0.00	0.00

Finally, to compare the robustness of ML estimators of the parameters of the proposed model against the competing models, we mimic the scenario in Choudhary, Sengupta and Casey (2014), wherein the data are generated from mixtures of normal distributions that represent the contaminated normal distributions. We conducted a simulation study involving different patterns of $x-$ and $\boldsymbol{e}-$outliers. Thus, the MEM used to simulate the data is given by (4.2) and with the following distributions:

$$
\begin{aligned}
x_i &\sim (1-p_x)\mathcal{N}(\xi,\omega^2) + p_x\mathcal{N}(\xi,c^2\omega^2),\\
\boldsymbol{e}_i &\sim (1-p_e)\mathcal{N}_2(\mathbf{0},\boldsymbol{\Sigma}_i) + p_e\mathcal{N}_2(\mathbf{0},c^2\boldsymbol{\Sigma}_i),\ i=1,\ldots,100,
\end{aligned} \tag{4.24}
$$

where p_x and p_e denote the respective expected proportions of the $x-$ and $\boldsymbol{e}-$outliers in the data and c denote the contamination factor. We take $p_x, p_e = 0, 0.05, 0.10, 0.25$ and $c = 4$ (distant contamination pattern), so that 16 combinations were used in this simulation study. In a similar way as Table 14, we consider only the STN MEM and the NN MEM to perform such comparisons with the proposed model. For these models, the practitioner may be interested in inference on (β_0, β_1) or $\mathrm{E}[\boldsymbol{Z}_1]$. The interpretation of these parameters is the same in the three models. Therefore, a total of 500 Monte Carlo replications were obtained for each (p_x, p_e, c) combination. The values of the parameters involved in (4.2) and (4.24) are in what follows: $\beta_0 = 0.1$, $\beta_1 = 1.1$, $\xi = 0$ and $\omega^2 = 0.5$. $\sigma^2_{\varepsilon_i}$ and $\sigma^2_{u_i}$, are generated from uniform distributions on $(0.15, 0.25)$ and $(0.20, 0.35)$, respectively, and then $\boldsymbol{\Sigma}_i = \mathrm{diag}(\sigma^2_{\varepsilon_i}, \sigma^2_{u_i})$ is assumed known. In this simulation study, the inverse of degrees of freedom are estimated using the CM step 6 in Section 4.2.2.

The STcT MEM, the STN MEM and the NN MEM are fitted for each data set and the ML estimates for $(\beta_0, \beta_1)^\top$ and $E[\boldsymbol{Z}_1]$ are computed. The target value of $E[\boldsymbol{Z}_1]$ is $(0.1, 0)^\top$. Next, the relative efficiencies are presented in Table 16. To ease the interpretation of these results, following Choudhary, Sengupta and Casey (2014), a gain or loss of up to 20% is too modest to be important for a practitioner, especially if the change is not in the same direction for all parameters. Observe that the models are equally efficient when there are no presence of outliers. When outliers are present, most efficiencies are one or more, some of them are 0.9 and 0.8. This suggests that although STcT MEM may lose some efficiency over STN MEM and NN MEM, the loss is never substantial from a practical viewpoint. In contrast, the STcT MEM may offer gains in efficiency over the STN MEM and the NN MEM, especially for β_0 and β_1, depending on the proportion of the outliers.

Table 16 – Efficiencies of STcT MEM-based estimators relative to the two competing models when the true model is based on contaminated normal distribution with distant contamination pattern.

	$p_x = 0$				$p_x = 0.05$				$p_x = 0.10$				$p_x = 0.25$			
	p_e				p_e				p_e				p_e			
Quantity	0	0.05	0.10	0.25	0	0.05	0.10	0.25	0	0.05	0.10	0.25	0	0.05	0.10	0.25
	STN MEM															
β_0	1.0	1.1	1.2	1.7	1.0	1.1	1.2	1.4	1.0	1.1	1.2	1.5	1.0	1.1	1.2	1.5
β_1	1.0	1.5	2.0	5.6	1.0	1.4	1.7	2.4	1.0	1.2	1.4	2.0	1.0	1.1	1.2	1.5
E$[Y_1]$	1.0	1.0	1.0	1.2	1.0	0.8	1.0	1.0	1.0	0.8	0.9	1.1	0.9	0.8	0.9	1.0
E$[X_1]$	1.0	1.0	1.1	1.2	1.0	0.8	1.0	1.1	0.9	0.8	0.8	1.1	0.9	0.8	0.9	1.0
	NN MEM															
β_0	1.0	1.1	1.2	1.6	1.0	1.0	1.2	1.3	1.0	1.1	1.2	1.4	1.0	1.1	1.2	1.5
β_1	1.0	1.3	1.8	4.5	0.9	1.2	1.4	2.1	0.9	1.1	1.2	1.7	0.9	1.0	1.1	1.4
E$[Y_1]$	1.0	1.0	1.0	1.2	1.0	0.8	1.0	1.1	1.0	0.8	0.9	1.2	1.0	0.8	1.0	1.0
E$[X_1]$	1.0	0.9	1.0	1.1	0.9	0.8	1.0	1.1	0.9	0.8	0.8	1.1	1.0	0.8	1.0	1.0

The STcT MEM's gain for $(\beta_0, \beta_1)^\top$ is remarkable for $p_e = 0.25$ and $p_x < p_e$, reaching its maximum at $(70\%, 460\%)^\top$ and $(60\%, 350\%)^\top$ with respect to the STN MEM and the NN MEM, respectively. Moreover, the results suggest that from a practical viewpoint, there may not be much difference among the three models for estimation of E$[\boldsymbol{Z}_1]$. On the other hand, the STcT MEM may outperform the STN MEM and the NN MEM for estimating $(\beta_0, \beta_1)^\top$ when we have a large proportion of x- and $\boldsymbol{e}$-outliers.

4.4 Data analysis

Our approach is illustrated with the analysis of a real dataset from a methods comparison study in Galea-Rojas *et al.* (2003, Example 4). Small amounts of coarse gold are present in a cooper deposit. In order to check the content of the gold particles, two measurement methods were used to analyze the samples. These are known as Classical and Screen Fire Assay (FA) methods. It is essential to know that the main emphasis in methods comparison studies is to check whether the two measurement techniques are comparable. It can be noted that the measurements were collected from a chemical laboratory and many factors may affect a measurement, such as observer, position of subject and laboratory. Consequently, these measurements are error-prone and we can apply the proposed model.

We denote X_i as the measurement taken by the Classical method and Y_i as the measurement obtained by the Screen FA method on the ith subject, $i = 1, \ldots, 501$. Moreover, measurement errors have known standard deviations, as our approach requires, see Galea-Rojas *et al.* (2003) for a detailed description.

Table 17 exhibits descriptive statistics for the measurements (in g/t) taken by the Classical and Screen FA methods. We include the mean (Mean), median (MD), standard deviation (SD), coefficient of skewness (CS), coefficient of kurtosis (CK), minimum (Min) and maximum (Max) values. In Table 17, concentrations for the Classical and Screen FA methods range from 0.040 to 4.523 g/t and from 0.038 to 2.650 g/t, respectively. Also, we can observe that CS and CK indicate a positively skewed nature and high kurtosis level of the distributions for Screen FA and Classical methods. The skewed nature and high kurtosis level can be confirmed by the histograms in Figure 3.

Table 17 – Descriptive statistics of the concentrations of the gold particles (in g/t) taken by two measurement methods.

Variable (Method)	Mean	SD	MD	Min	Max	CS	CK
X (Classical)	0.289	0.424	0.151	0.040	4.523	5.306	41.158
Y (Screen FA)	0.294	0.337	0.176	0.038	2.650	3.045	12.639

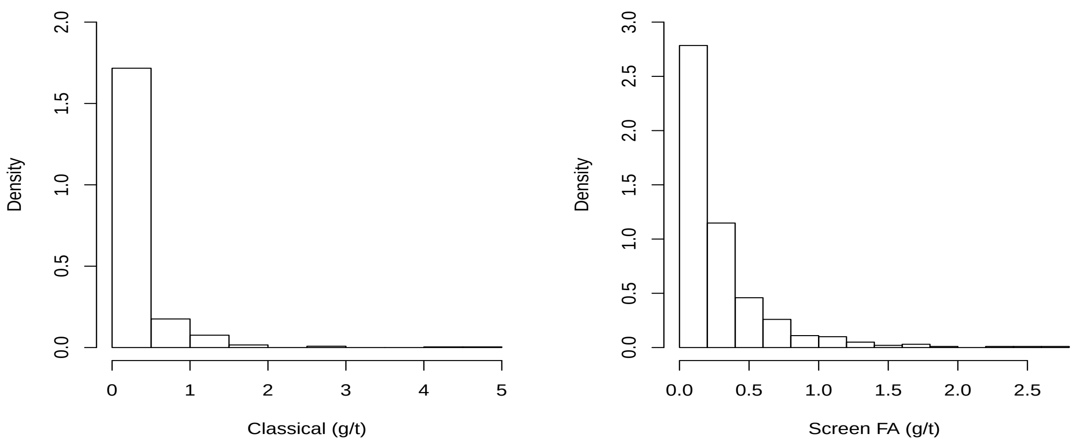

Figure 3 – Histogram of the concentrations of the gold particles (in g/t) taken by two measurement methods. The Classical method (left) and Screen FA method (right).

Next, we consider the model (4.2) in an equivalent representation $\boldsymbol{Z}_i^* = \boldsymbol{a}^* + \boldsymbol{b}x_i' + \boldsymbol{e}_i$, where $\boldsymbol{Z}_i^* = (X_i^*, Y_i)^\top$, $\boldsymbol{a}^* = (0, \beta_0^*)^\top$, $\boldsymbol{b} = (1, \beta_1)^\top$ and $\boldsymbol{e}_i = (u_i, \varepsilon_i)^\top$, $i = 1, \ldots, 501$. Here, $\beta_0^* = \beta_0 + \beta_1 R$, $X_i^* = X_i - R$ and $R = \bar{X} = \sum_{i=1}^n X_i/n$ (known constant). In addition, $x_i' \sim$

$\mathscr{S}t(\xi^*, \omega^2, \lambda, \eta_x)$ where $\xi^* = \xi - R$. To compute the asymptotic standard errors for the estimator of $\boldsymbol{\theta}$, the delta method was used. The centered covariate X_i^* makes the computation of the observed information matrix more stable.

In the sequel, we fit the STcT MEM as well as the competing models in Section 4.3. Here, the degrees of freedom are considered as unknown parameters. These (low) estimates give strong support against the normal distribution as an alternative to (4.3). Furthermore, Table 18 shows that the STcT MEM is selected by AIC and BIC over other competing models, followed by the STN MEM and the TcT MEM in that order. The remaining models yield a much worse fit. Table 19 displays the ML estimates and their asymptotic standard errors for all the MEM considered in this section.

Table 18 – Model selection criteria.

	Model				
Criterion	STcT	STN	TcT	TN	NN
AIC	-1319.8	-1081.5	-1067.0	-753.7	-480.6
BIC	-1290.3	-1056.2	-1041.7	-732.6	-463.7

Table 19 – ML estimates and asymptotic standard errors (SE).

	Model									
	STcT		STN		TcT		TN		NN	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_0 [a]	-0.001	0.006	0.002	0.006	-0.004	0.006	0.002	0.006	0.001	0.006
β_1 [a]	1.123	0.026	1.142	0.026	1.163	0.028	1.138	0.025	1.157	0.026
ξ	0.028	0.003	0.026	0.004	0.171	0.009	0.172	0.001	0.239	0.012
ω^2	0.091	0.028	0.048	0.008	0.017	0.002	0.019	0.003	0.067	0.006
λ	192.3	160.9	118.3	98.40	-	-	-	-	-	-
η_x	0.209	0.053	0.250	0.037	0.250	0.034	0.250	0.028	-	-
η_e	0.250	0.047	-	-	0.250	0.023	-	-	-	-

[a] Statistical significance of the parameters (p-value): β_0 (0.868) and β_1 (< 0.0001).

Furthermore, as the TcT MEM and STcT MEM are two nested models (see Section 4.2.3), we test the null hypothesis H_0 : the TcT MEM is preferable (or simply $\lambda = 0$) against the alternative hypothesis H_a : the STcT MEM is preferable (or simply $\lambda \neq 0$). Testing the hypothesis H_0 is of interest to verify whether or not the inclusion of the skewness parameter is significant. We consider the log-likelihood ratio (LR) test statistic given by $LR = 2[\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}_0]$ for testing H_0, which is asymptotically distributed according to the χ_1^2 distribution, where $\ell(\hat{\boldsymbol{\theta}})$ and $\ell(\hat{\boldsymbol{\theta}}_0)$ are the log-likelihood functions evaluated at the ML estimates based on the full and constrained models, respectively. Applying the test, we obtain $LR = 254.810$ (p-value < 0.0001). Therefore, the parameter λ must be considered and there is evidence that the distributions of the true unobserved x_i's are asymmetric.

The scatter plot of the individual observations together with their standard deviations and regression lines are shown in Figure 4. Notice that the ML estimates for β_0 and β_1 given in Table 19 and fitted by the different models, were fairly similar including their estimated asymptotic standard errors. Consequently, due to a slight difference among the ML estimates for β_1, the regression lines were almost close. It can also be noted that the estimates for η_x and η_e indicated different levels of heaviness in the tails of the unobserved covariate and the random errors distributions.

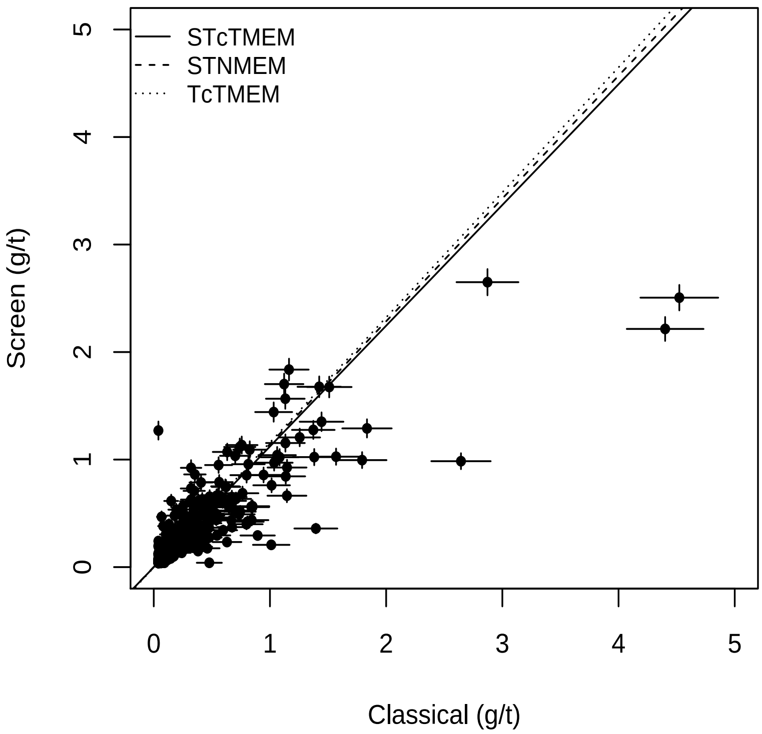

Figure 4 – Individual observations and their standard deviations together with regression lines under three different models.

4.5 Conclusion

In this chapter, we proposed an extension of the structural MEM under the assumption of the normality, named STcT MEM, which offers a great flexibility for accommodating skewness and heavy tails in the data. A special attractive characteristic of the proposed model is that it can incorporate simultaneously different levels of heaviness in the tails of the true covariate and the error terms distributions. We developed an EM-type algorithm to perform maximum likelihood estimation of the parameters involved in the proposed model, which with some adaptations, was also useful to fit other models we dealt with in this chapter. The simulation results indicate that the STcT MEM-based ML estimates presented good properties in small to moderate sample sizes. When the data are generated from a MEM based on contaminated normal distributions with distant contamination factor, the STcT MEM's gain for estimating of $(\beta_0, \beta_1)^\top$ is more efficient than the other models, when we have a considerable proportion of the x- and $\boldsymbol{e}$-outliers.

CHAPTER

5

FUTURE WORK

In this chapter, we briefly describe some problems that we plan to address as future work. We have studied maximum penalized likelihood method in a normal MEM using the Jeffreys' prior as the penalty function. We would like to use an orthogonalized version of the Firth's method or other penalty functions, see, e.g., Bolfarine and Cordani (1993), Azzalini and Arellano-Valle (2013) and Lima and Cribari-Neto (2017).

On the other side, FGPQ's for the Grubbs model were obtained by means of the structural method, but they are not unique, which means that different FGPQ's can be constructed by using the other techniques outlined in Hannig, Iyer and Patterson (2006). Moreover, another generalized fiducial distributions can be derived (HANNIG *et al.*, 2016).

Instead of the variance parameters in the Grubbs model, FGPQ's and GFD's could be built for other quantities of interest, for instance, the reliability of an instrument (σ_x^2/σ_u^2 or $\sigma_x^2/\sigma_\varepsilon^2$). Generalized fiducial estimation methods for the three-instrument case might be investigated.

In our simulation study related to the STcT MEM, we observed that it is necessary a larger sample size for the estimation of the skewness parameter. Alternative methods of estimation (for instance, penalized likelihood-based estimation) is subject of future work as well as assessment of local influence and hypotheses testing procedures. Moreover, the distributions considered in Lachos *et al.* (2010) could also be adapted in an analogous way to this work. For the case of the assumption on the known error variances, some extension for the case of replicated observations could be thought. Thus, this would permit us to consider the unknown heteroscedastic variances estimable from replications (LIN; CAO, 2013; CAO *et al.*, 2017). A Bayesian approach also could be proposed as part of possible future work.

The observed information matrix was computed using numerical differentiation. Alternatives to this approach are the method of Louis (1982) and other approaches reviewed in McLachlan and Krishnan (2007, Section 4.7). Further work is needed to develop and compare these methods. Lastly, we have developed an EM-type algorithm to fit the STcT MEM. We would like to implement this algorithm in a more efficient way so that it takes less time to converge.

BIBLIOGRAPHY

ARELLANO-VALLE, R.; BOLFARINE, H.; LACHOS, V. H. Skew-normal linear mixed models. **Journal of Data Science**, v. 3, p. 415–438, 2005. Citations on pages 60, 61, and 66.

ARELLANO-VALLE, R. B.; OZAN, S.; BOLFARINE, H.; LACHOS, V. H. Skew normal measurement error models. **Journal of Multivariate Analysis**, v. 96, p. 265–281, 2005. Citations on pages 17, 58, 70, and 72.

ARRUÉ, J.; ARELLANO-VALLE, R. B.; GÓMEZ, H. W. Bias reduction of maximum likelihood estimates for a modified skew-normal distribution. **Journal of Statistical Computation and Simulation**, v. 86, p. 2967–2984, 2016. Citation on page 20.

AZZALINI, A.; ARELLANO-VALLE, R. B. Maximum penalized likelihood estimation for skew-normal and skew-t distributions. **Journal of Statistical Planning and Inference**, v. 143, p. 419–433, 2013. Citations on pages 20 and 81.

AZZALINI, A.; CAPITANIO, A. Statistical application of the multivariate skew-normal distribution. **Journal of the Royal Statistical Society, Series B**, v. 61, p. 579–609, 1999. Citation on page 17.

____. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. **Journal of the Royal Statistical Society, Series B**, v. 65, p. 367–389, 2003. Citations on pages 17, 58, 59, 62, and 67.

BIRCH, M. W. A note on the maximum likelihood estimation of a linear structural relationship. **Journal of the American Statistical Association**, v. 59, p. 1175–1178, 1964. Citation on page 22.

BOLFARINE, H.; CORDANI, L. K. Estimation of a structural linear regression model with a known reliability ratio. **Annals of the Institute of Statistical Mathematics**, v. 45, p. 531–540, 1993. Citation on page 81.

BUONACCORSI, J. P. **Measurement Error: Models, Methods and Applications**. Boca Raton: Chapman & Hall/CRC, 2010. Citations on pages 15, 57, and 71.

CABRAL, C. R. B.; LACHOS, V. H.; ZELLER, C. B. Multivariate measurement error models using finite mixtures of skew-Student t distributions. **Journal of Multivariate Analysis**, v. 124, p. 179–198, 2014. Citation on page 58.

CAO, C.; CHEN, M.; WANG, Y.; SHI, J. Heteroscedastic replicated measurement error models under asymmetric heavy-tailed distributions. **Computational Statistics**, v. 33, p. 319–338, 2017. Citation on page 81.

CAO, C.-Z.; LIN, J.-G.; ZHU, X.-X. On estimation of a heteroscedastic measurement error model under heavy-tailed distributions. **Computational Statistics & Data Analysis**, v. 56, p. 438–448, 2012. Citations on pages 58 and 73.

CARROLL, R. J.; RUPPERT, D.; STEFANSKI, L. A.; CRAINICEANU, C. M. **Measurement Error in Nonlinear Models: A Modern Perspective**. 2nd. ed. Boca Raton: Chapman & Hall/CRC, 2006. Citation on page 15.

CASTRO, M. de; GALEA, M.; BOLFARINE, H. Hypothesis testing in an errors-in-variables model with heteroscedastic measurement errors. **Statistics in Medicine**, v. 27, p. 5217–5234, 2008. Citation on page 57.

CASTRO, M. de; VIDAL, I. Bayesian inference in measurement error models from objective priors for the bivariate normal distribution. To appear in *Statistical Papers*, DOI:10.1007/s00362-016-0863-7. 2017. Citation on page 42.

CHAN, L. K.; MAK, T. K. On the maximum likelihood estimation of a linear structural relationship when the intercept is know. **Journal of Multivariate Analysis**, v. 9, p. 304–312, 1979. Citation on page 22.

CHENG, C.-L.; NESS, J. W. V. **Statistical Regression with Measurement Error**. London: Arnold, 1999. Citations on pages 15, 21, 22, 23, and 58.

CHENG, C.-L.; RIU, J. On estimating linear relationships when both variables are subject to heteroscedastic measurement errors. **Technometrics**, v. 48, p. 511–519, 2006. Citations on pages 57 and 59.

CHIANG, A. K. L. A simple general method for constructing confidence intervals for functions of variance components. **Technometrics**, v. 43, p. 356–367, 2001. Citation on page 44.

CHOUDHARY, P. K.; SENGUPTA, D.; CASEY, P. A general skew-*t* mixed model that allows different degrees of freedom for random effects and error distributions. **Journal of Statistical Planning and Inference**, v. 147, p. 235–247, 2014. Citations on pages 58, 74, and 75.

CORDEIRO, G. M.; MCCULLAGH, P. Bias correction in generalized linear models. **Journal of the Royal Statistical Society. Series B**, v. 53, p. 629–643, 1991. Citation on page 19.

COX, D. R.; SNELL, E. J. A general definition of residuals. **Journal of the Royal Statistical Society, Series B**, v. 30, p. 248–275, 1968. Citation on page 19.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm (with discussion). **Journal of the Royal Statistical Society, Series B**, v. 39, p. 1–38, 1977. Citation on page 63.

DRAPER, N.; GUTTMAN, I. Two simultaneous measurement procedures: A Bayesian approach. **Journal of the American Statistical Association**, v. 70, p. 43–46, 1975. Citation on page 42.

EFRON, B. Defining the curvature of a statistical problem (with applications to second order efficiency). **The Annals of Statistics**, v. 3, p. 1189–1217, 1975. Citation on page 19.

FIRTH, D. Bias reduction of maximum likelihood estimates. **Biometrika**, v. 80, p. 27–38, 1993. Citations on pages 9, 11, 15, 19, 20, 22, 27, and 40.

FISHER, R. Inverse probability. **Proceedings of the Cambridge Philosophical Society**, xxvi, p. 528–535, 1930. Citations on pages 42 and 44.

FULLER, W. A. **Measurement Error Models**. New York: Wiley, 1987. Citations on pages 15, 22, 38, and 59.

GALEA-ROJAS, M.; CASTILHO, M. V. D.; BOLFARINE, H.; CASTRO, M. de. Detection of analytical bias. **Analyst**, v. 128, p. 1073–1081, 2003. Citations on pages 57, 75, and 76.

GILBERT, P.; VARADHAN, R. **numDeriv: Accurate Numerical Derivatives**. [S.l.], 2016. R package version 2016.8-1. Available from: http://CRAN.R-project.org/package=numDeriv (Accessed 10 February 2016). Available: <http://CRAN.R-project.org/package=numDeriv>. Citations on pages 65 and 72.

GILLARD, J. An overview of linear structural models in errors in variables regression. **Revstat Statistical Journal**, v. 8, p. 57–80, 2010. Citation on page 22.

GINER, G.; SMYTH, G. K. statmod: Probability calculations for the inverse Gaussian distribution. **The R Journal**, v. 8, p. 339–351, 2016. Citations on pages 68 and 72.

GRUBBS, F. E. On estimating precision of measuring instruments and product variability. **Journal of the American Statistical Association**, v. 43, p. 243–264, 1948. Citations on pages 42 and 43.

____. Grubbs's estimator. In: KOTZ, S. (Ed.). **Encyclopedia of Statistical Sciences**. New York: Wiley, 1983. v. 3, p. 542–549. Citations on pages 41 and 42.

HANNIG, J. **On Asymptotic Properties of Generalized Fiducial Inference for Discretized Data**. Chapel Hill, USA, 2009. Citations on pages 44 and 47.

____. On generalized fiducial inference. **Statistica Sinica**, v. 19, p. 491–544, 2009. Citations on pages 42, 44, 45, 47, 48, and 52.

____. Generalized fiducial inference via discretization. **Statistica Sinica**, v. 23, p. 489–514, 2013. Citations on pages 44 and 50.

HANNIG, J.; IYER, H.; LAI, R. C. S.; LEE, T. C. M. Generalized fiducial inference: A review and new results. **Journal of the American Statistical Association**, v. 111, p. 1346–1361, 2016. Citations on pages 40, 42, 44, 45, 47, 48, 49, and 81.

HANNIG, J.; IYER, H.; PATTERSON, P. Fiducial generalized confidence intervals. **Journal of the American Statistical Association**, v. 101, p. 254–269, 2006. Citations on pages 42, 44, 45, 47, and 81.

HEFLEY, T. J.; HOOTEN, M. B. On the existence of the maximum likelihood estimates for presence-only data. **Methods in Ecology and Evolution**, v. 6, p. 648–655, 2015. Citation on page 20.

HEINZE, G.; PLONER, M. Sas and splus programs to perform cox regression without convergence problems. **Computer Methods and Programs in Biomedicine**, v. 67, p. 217–223, 2002. Citations on pages 28 and 29.

HEINZE, G.; SCHEMPER, M. A solution to the problem of monotone likelihood in Cox regression. **Biometrics**, v. 57, p. 114–119, 2001. Citation on page 20.

____. A solution to the problem of separation in logistic regression. **Statistics in Medicine**, v. 21, p. 2409–2419, 2002. Citation on page 20.

HOOD, K.; NIX, B. A. J.; ILES, T. C. Asymptotic information and variance-covariance matrices for the linear strutural model. **The Statistician**, v. 48, p. 477–493, 1999. Citations on pages 22, 23, and 27.

IYER, H. K.; PATTERSON, P. **A recipe for constructing generalized pivotal quantities and generalized confidence intervals**. Tech. rep., Department of Statistics, Colorado State University, 2002. Citation on page 49.

JAECH, J. **Statistical Analysis of Measurement Errors**. New York: Wiley, 1985. Citations on pages 42, 43, and 55.

JOHNSON, N. L.; KOTZ, S.; BALAKRISHNAN, N. **Continuous Univariate Distributions**. 2nd. ed. New York: Wiley, 1994. Citation on page 67.

KELLY, B. C. Some aspects of measurement error in linear regression of astronomical data. **The Astrophysical Journal**, v. 665, p. 1489–1506, 2007. Citation on page 57.

KHERADMANDI, A.; RASEKH, A. Estimation in skew-normal linear mixed measurement error models. **Journal of Multivariate Analysis**, v. 136, p. 1–11, 2015. Citations on pages 58 and 72.

KOSMIDIS, I. Improved estimation in cumulative link models. **Journal of the Royal Statistical Society, Series B**, v. 76, p. 169–196, 2014. Citation on page 20.

KOSMIDIS, I.; FIRTH, D. Bias reduction in exponential family nonlinear models. **Biometrika**, v. 96, p. 793–804, 2009. Citation on page 20.

____. Multinomial logit bias reduction via the Poisson log-linear models. **Biometrika**, v. 98, p. 755–759, 2011. Citation on page 20.

KULATHINAL, S. B.; KUULASMAA, K.; GASBARRA, D. Estimation of an errors-in-variables regression model when the variances of the measurement errors vary between the observations. **Statistics in Medicine**, v. 21, p. 1089–1101, 2002. Citations on pages 57 and 71.

LACHOS, V. H.; CANCHO, V. G.; AOKI, R. Bayesian analysis of skew-t multivariate null intercept measurement error model. **Statistical Papers**, v. 51, p. 531–545, 2010. Citation on page 58.

LACHOS, V. H.; LABRA, F. V.; BOLFARINE, H.; GHOSH, P. Multivariate measurement error models based on scale mixtures of the skew-normal distribution. **Statistics**, v. 44, p. 541–556, 2010. Citations on pages 17, 58, and 81.

Lagos Álvarez, B.; Jiménez Gamero, M. A note on bias reduction of maximum likelihood estimates for the scalar skew t distribution. **Journal of Statistical Planning and Inference**, v. 142, p. 608–612, 2012. Citation on page 20.

LANGE, K. L.; LITTLE, R. J. A.; TAYLOR, J. M. G. Robust statistical modeling using the t distribution. **Journal of the American Statistical Association**, v. 84, p. 881–896, 1989. Citation on page 59.

LEHMANN, E. L. **Elements of Large-Sample Theory**. New York: Springer, 1998. Citations on pages 26 and 65.

LI, X.-M.; XU, X.-Z.; LI, G.-Y. A fiducial argument for generalized p-value. **Science in China Series A**, v. 50, p. 957–966, 2007. Citation on page 42.

LIMA, V. M.; CRIBARI-NETO, F. **Penalized Maximum Likelihood Estimation in the Modified Extended Weibull Distribution**. 2017. Citation on page 81.

LIN, J.-G.; CAO, C.-Z. On estimation of measurement error models with replication under heavy-tailed distributions. **Computational Statistics**, v. 28, p. 809–829, 2013. Citation on page 81.

LOMBARD, F.; POTGIETER, C. Another look at the Grubbs estimators. **Chemometrics and Intelligent Laboratory Systems**, v. 110, p. 74–80, 2012. Citation on page 42.

LOUIS, T. A. Finding the observed information matrix when using the em algorithm. **Journal of the Royal Statistical Society, Series B**, v. 44, p. 226–233, 1982. Citation on page 81.

MADANSKY, W. The fitting of straight lines when both variables are subject to error. **Journal of the American Statistical Association**, v. 54, p. 173–205, 1959. Citation on page 22.

MALONEY, C. J.; RASTOGI, S. C. Significance test for Grubbs's estimators. **Biometrics**, v. 26, p. 671–676, 1970. Citation on page 42.

MCLACHLAN, G. J.; KRISHNAN, T. **The EM Algorithm and Extensions**. 2nd. ed. New York: Wiley, 2007. Citation on page 81.

MELO, T. F. N.; FERRARI, S. L. P.; PATRIOTA, A. G. Modified likelihood ratio tests in heteroskedastic multivariate regression models with measurement error. **Journal of Statistical Computation and Simulation**, v. 84, p. 2233–2247, 2014. Citation on page 58.

MENG, X.-L.; RUBIN, D. B. Maximum likelihood estimation via the ECM algorithm: a general framework. **Biometrika**, v. 80, p. 267–278, 1993. Citation on page 63.

MEYVISCH, P. Evaluation of the addition of the Firth's penalty term to the Bradley-Terry likelihood. **Journal of Modern Applied Statistical Methods**, v. 15, p. 224–237, 2016. Citation on page 20.

OSPINA, R.; CRIBARI-NETO, F.; VASCONCELLOS, K. L. P. Improved point and interval estimation for a beta regression model. **Computational Statistic & Data Analysis**, v. 51, p. 960–981, 2006. Citation on page 20.

PETTITT, A. N.; KELLY, J. M.; GAO, J. T. Bias correction for censored data with exponential lifetimes. **Statistica Sinica**, v. 8, p. 941–963, 1998. Citation on page 20.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2017. Available: <https://www.R-project.org/>. Citations on pages 29, 52, 68, and 72.

SARTORI, N. Bias prevention of maximum likelihood estimates for scalar skew normal and skew *t* distributions. **Journal of Statistical Planning and Inference**, v. 136, p. 4259–4275, 2006. Citation on page 20.

SCHAEFER, R. L. Bias correction in maximum likelihood logistic regression. **Statistics in Medicine**, Wiley Online Library, v. 2, p. 71–78, 1983. Citation on page 19.

SEBER, G. A. F.; LEE, A. J. **Linear Regression Analysis**. New York: Wiley, 2003. Citation on page 65.

SHUKLA, G. K. Some exact tests of hypotheses about Grubbs's estimators. **Biometrics**, v. 29, p. 373–377, 1973. Citation on page 42.

SIINO, M.; FASOLA, S.; MUGGEO, V. M. R. Inferential tools in penalized logistic regression for small and sparse data: A comparative study. **Statistical Methods in Medical Research**, v. 27, p. 1365–1375, 2016. Citations on pages 20 and 28.

STRIKE, P. Assay method comparison studies. In: STRIKE, P. (Ed.). **Statistical Methods in Laboratory Medicine**. UK: Butterworth-Heinemann, 1991. p. 307 – 330. Citation on page 38.

SUTRADHAR, B. C. Score test for the covariance matrix of the elliptical t-distribution. **Journal of Multivariate Analysis**, v. 46, p. 1–12, 1993. Citations on pages 17, 58, and 59.

THOMPSON, J. R.; CARTER, R. L. An overview of normal theory structural measurement error models. **International Statistical Review**, v. 75, p. 183–198, 2007. Citation on page 22.

THOMPSON, W. A. J. Precision of simultaneous measurement procedures. **Journal of the American Statistical Association**, v. 58, p. 474–479, 1963. Citations on pages 43 and 44.

TOMAYA, L. C.; CASTRO, M. de. A heteroscedastic measurement error model based on skew and heavy-tailed distributions with known error variances. **Journal of Statistical Computation and Simulation**, v. 88, p. 2185–2200, 2018. Citation on page 57.

____. New estimation methods for the grubbs model. **Chemometrics and Intelligent Laboratory Systems**, v. 176, p. 119–125, 2018. Citation on page 41.

TSUI, K.-W.; WEERAHANDI, S. Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. **Journal of the American Statistical Association**, v. 84, p. 602–607, 1989. Citation on page 42.

VENZON, D. J.; MOOLGAVKAR, S. H. A method for computing profile-likelihood-based confidence intervals. **Journal of the Royal Statistical Society, Series C**, v. 37, p. 87–94, 1988. Citation on page 28.

WANDLER, D. V.; HANNIG, J. Fiducial inference on the largest mean of a multivariate normal distribution. **Journal of Multivariate Analysis**, v. 102, p. 87–104, 2011. Citation on page 49.

WANG, C. M.; IYER, H. K. Fiducial approach for assessing agreement between two instruments. **Metrologia**, v. 45, p. 415–421, 2008. Citations on pages 42 and 44.

WANG, G. **Some Bayesian methods in the estimation of parameters in the measurement error models and crossover trial**. Phd Thesis — Departament of Mathematical Sciences, University of Cincinnati, Cincinnati, USA, 2004. Citations on pages 23 and 50.

WANG, G.; SIVAGANESAN, S. Objective priors for parameters in a normal linear regression with measurement error. **Communications in Statistics - Theory and Methods**, v. 42, p. 2694–2713, 2013. Citation on page 30.

WEERAHANDI, S. Generalized confidence intervals. **Journal of the American Statistical Association**, v. 88, p. 899–905, 1993. Citations on pages 42 and 44.

XU, X.; LI, G. Fiducial inference in the pivotal family of distributions. **Science in China Series A**, v. 49, p. 410–432, 2006. Citation on page 42.

YAN, L.; WANG, R.; XU, X. Fiducial inference in the classical errors-in-variables model. **Metrika**, v. 80, p. 93–114, 2017. Citations on pages 40 and 42.

____. A new confidence interval in errors-in-variables model with known error variance. **Journal of Applied Statistics**, v. 44, p. 2204–2221, 2017. Citations on pages 40, 42, and 46.

YAN, L.; XU, X. A new confidence interval in measurement error model with the reliability ratio known. **Communications in Statistics - Theory and Methods**, v. 46, p. 9636–9650, 2017. Citations on pages 40 and 42.