UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Matemática, Estatística e Computação Científica

JOÃO VICTOR BASTOS DE FREITAS

Semi-parametric models based on the Scale Mixtures of Centred Skew Normal distribution for independent and longitudinal data

Modelos semi-paramétricos para dados independentes e longitudinais baseados em Misturas de Escala Normal Assimétrica Centralizada

Campinas

2021

João Victor Bastos de Freitas

Semi-parametric models based on the Scale Mixtures of Centred Skew Normal distribution for independent and longitudinal data

Modelos semi-paramétricos para dados independentes e longitudinais baseados em Misturas de Escala Normal Assimétrica Centralizada

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

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

Supervisor: Caio Lucidius Naberezny Azevedo

Co-supervisor: Juvêncio Santos Nobre

Este trabalho corresponde à versão final da Dissertação defendida pelo aluno João Victor Bastos de Freitas e orientada pelo Prof. Dr. Caio Lucidius Naberezny Azevedo.

Campinas

2021

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

F884s

Freitas, João Victor Bastos de, 1996-
Semi-parametric models based on the scale mixtures of centred skew normal distribution for independent and longitudinal data / João Victor Bastos de Freitas. – Campinas, SP : [s.n.], 2021.

Orientador: Caio Lucidius Naberezny Azevedo.
Coorientador: Juvêncio Santos Nobre.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Distribuição normal assimétrica. 2. Misturas de escala (Estatística). 3. Modelos aditivos generalizados. 4. Método longitudinal. 5. Equações de estimação generalizadas. I. Azevedo, Caio Lucidius Naberezny, 1979-. II. Nobre, Juvêncio Santos. III. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Modelos semi-paramétricos para dados independentes e longitudinais baseados em misturas de escala normal assimétrica centralizada
Palavras-chave em inglês:
Skew-normal distributions
Scale mixtures
Generalized additive model
Longitudinal method
Generalized estimating equations
Área de concentração: Estatística
Titulação: Mestre em Estatística
Banca examinadora:
Caio Lucidius Naberezny Azevedo [Orientador]
Alexandre Galvão Patriota
Filidor Edilfonso Vilca Labra
Data de defesa: 09-03-2021
Programa de Pós-Graduação: Estatística

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0001-6025-8087
- Currículo Lattes do autor: http://lattes.cnpq.br/9670482171354483

Dissertação de Mestrado defendida em 09 de março de 2021 e aprovada pela banca examinadora composta pelos Profs. Drs.

Prof(a). Dr(a). CAIO LUCIDIUS NABEREZNY AZEVEDO

Prof(a). Dr(a). ALEXANDRE GALVÃO PATRIOTA

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

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

A Deus pelas oportunidades e bençãos que me concedeu.
À minha esposa por todo apoio, amor, e por sempre caminhar ao meu lado.
À minha familia por todo o apoio
Aos meus mestres prof. Caio e prof. Juvêncio, por acreditarem em mim, obrigado por tudo.

Acknowledgements

Aos meus pais a quem devo tudo e que sempre foram minha principal inspiração como pessoas. Obrigado pela educação que me foi dada, pelo amor e carinho. Agradeço à minha irmã por todo amor e carinho também. Esse momento é mais uma conquista da nossa família.

À minha tia Wana e minha avó Isabel por todo carinho e por sempre cuidarem de mim.

Agradeço aos meus orientadores prof. Caio Azevedo e prof. Juvêncio Nobre por me darem a oportunidade de fazer esse trabalho, por sempre me ajudar, aconselhar e me orientar pelo caminho certo. Os levo como a maior referência acadêmica e profissional, e serei feliz se conseguir ser metade do professor que vocês são.

A todos que fazem parte do Departamento de Estatística da Unicamp. Agradeço pelos ensinamentos aos professores: Diego Bernardini, Ronaldo Dias, Aluísio Pinheiro, Caio Azevedo, Jesus Garcia, Hélio Migon, Samara Kiihl, Benilton Carvalho e Rafael Maia. Agradeço também a todos os funcionários do IMECC pela simpatia e atenção.

A todos que fazem parte do Departamento de Estatística e Matemática Aplicada da UFC.

Aos membros da minha banca: Prof. Alexandre Patriota, Prof. Filidor Labra, Prof. Gilberto Paula, Prof. Marcelo Bourguignon e Profa. Larissa Matos pela disponibilidade e pelos comentários de refinamento do presente trabalho.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo apoio financeiro através de uma bolsa de Mestrado (processo nº 2018/26780-6).

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

E a todos os familiares, colegas e amigos que não mencionei que contribuíram nessa trajetória, direta ou indiretamente.

Resumo

Nesta dissertação foram desenvolvidas duas classes flexíveis de modelos de regressão para dados contínuos, assimétricos e/ou de caudas pesadas. Uma para dados independentes e outra para dados dependentes. Consideramos uma abordagem semi-paramétrica, utilizando Modelos Lineares Parciais Aditivos Generalizados (MLPAG), para dados independentes, e MLPAG com Equações de Estimação Generalizadas (EEG), para dados dependentes. Em ambos os casos foram considerados preditores semi-paramétricos para as médias das respostas e erros (marginais) seguindo distribuições de misturas de escala normal assimétrica centralizada (MENAC). No caso de dados dependentes, as estruturas de dependência foram modeladas via EEG. Em relação às distribuições MENAC, consideramos medidas misturadoras usuais (gama, beta e binária) e outras nunca consideradas (gama generalizada, Birnbaum-Saunders e beta prime). Foram desenvolvidos métodos de estimação, medidas de qualidade de ajuste e de diagnóstico para esses modelos, sob a ótica frequentista. Foram criadas rotinas computacionais para permitir a utilização das metodologias desenvolvidas, bem como foram conduzidos estudos de simulação para verificar seus desempenhos. Também, modelagens de problemas reais, através das metodologias desenvolvidas, foram consideradas, ilustrando o potencial dos resultados obtidos.

Palavras-chave: Dados longitudinais, Misturas de escala normal assimétrica centralizada, Equações de Estimação generalizadas, Modelos Lineares Parciais Aditivos Generalizados, Inferência frequentista.

Abstract

In this dissertation, two classes of regression models for continuous, skewed and/or heavy tailed data were developed. One for independent data and another for dependent data. We considered a semi-parametric approach using Generalized Additive Partially Linear Models (GAPLM), for independent data, and GAPLM with Generalized Estimation Equations (GEE), for dependent data. In both cases, semi-parametric predictors for response means and scale mixtures of centered skew-normal (SMCSN) distributions for the (marginal) errors were considered. For dependent data, the dependence structures were modelled through GEE. Concerning the SMCSN distributions we considered either usual mixing measures (gamma, beta and binary distributions) as well as never used ones (generalized gamma, Birnbaum-Saunders and beta prime distributions). Estimation methods, goodness of model fit and diagnostic tools for these models, under the frequentist paradigm, were developed. Computational routines were created, to allow for the use of the developed methodologies, as well as simulation studies were performed to study the their performance. Also, the modelling of real problems, through such methodologies, were considered, illustrating the potential of the obtained results.

Keywords: Longitudinal data, Scale Mixtures of Centered Skew-Normal, Generalized Estimating Equations, Generalized Additive Partially Linear Models, Frequentist inference.

List of Figures

List of Tables

List of abbreviations and acronyms

pdf	probability density function
cdf	cumulative density function
SN	skew-normal
CSN	Centered skew-normal
SMN	scale mixture of Normal
SMCSN	scale mixture of Centered skew-normal
APLM	Additive partially linear model
GAPLM	Additive partially linear model
EM	Expectation-Maximization
ECME	Expectation-Conditional-Maximization-Either
SAEM	Stochastic Approximation of EM
GEE	generalized Estimating Equations
CSN	Centered skew-normal
CST	Centered Skew-t
CSS	Centered skew Slash
CSCN	Centered skew Contaminated Normal
CSGT	Centered skew generalized t
CSBPN	Centered skew Beta Prime Normal
CSBSN	Centered skew Birnbaum-Saunders Normal
CSGGN	Centered skew generalized Gamma Normal
IG	Inverse Gaussian

List of Algorithms

Contents

Chapter 1

SCALE MIXTURE OF CENTERED SKEW-NORMAL DISTRIBUTIONS

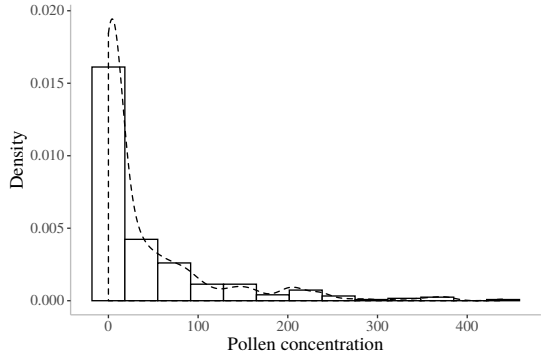
1.1 Introduction

In the following, we will introduce the first part of this work talking about a motivating example for the proposed methodologies and a brief review of the literature of methods similar to the ones we will be using.

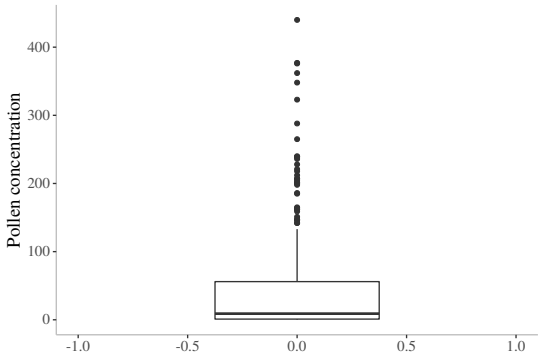
1.1.1 Motivating example

Let us consider the following study about daily ragweed pollen levels (Stark et al., 1997; Ruppert et al., 2003). This data set was collected from 1991 to 1994 during the ragweed season in Kalamazoo, Michigan, totaling 335 observations. According to Ruppert et al. (2003) the main goal is to analyze the pollen concentration (grains/m^3) as function of: the number of days in the current ragweed pollen season (days in season), the temperature in the following day (in F$^\circ$) (temperature), an indicator of significant rain the in following day (1 = at least 3 hours of steady or brief but intense rain, 0 = otherwise) and wind speed forecast for the following day (wind speed).

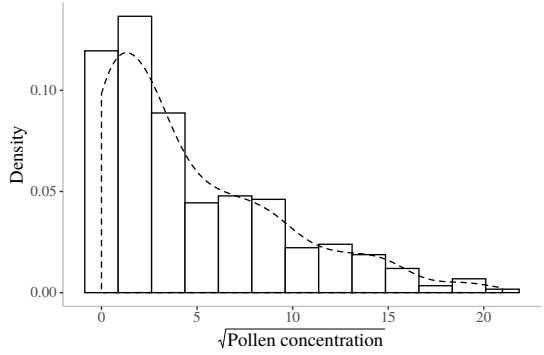
This data set was analyzed by Ferreira and Paula (2017) and we will use their results as a motivation for the first part of this dissertation. Figure 2 presents the Histogram and estimated densities in the original scale (Figure 1a) and in the square root-scale (Figure 1c). According to Ferreira and Paula (2017), the original response distribution is highly skewed, that would over complicate the data analysis. On the other hand, the square root transformation induces a lesser skewed behavior of the response, which is more easily for the modeling process. These authors also analyzed the response variable on the logarithmic scale. However, this transformation lead to non finite values. Table 1 provides descriptive statistics, which suggest an apparent skewness (1.12) and heavy tails (Kurtosis = 3.59).



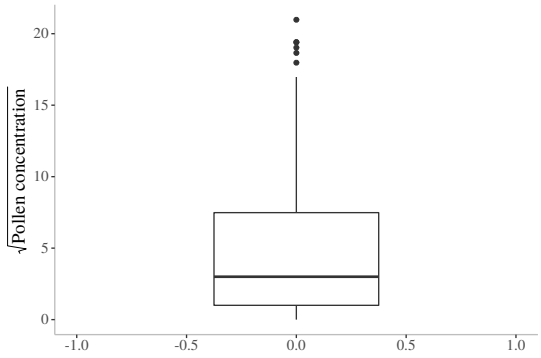
(a) Histogram and estimated density of Pollen concentration.



(b) boxplot of Pollen concentration.



(c) Histogram and estimated density of $\sqrt{\text{Pollen concentration}}$.



(d) boxplot of $\sqrt{\text{Pollen concentration}}$.

Figure 1 – Histograms, estimated densities and boxplots for Pollen concentration and $\sqrt{\text{Pollen concentration}}$.

Table 1 – Summary statistics for original scale and the square root transformation of the pollen concentration (SD is sample standard deviation and CV is coefficient of variation).

Scale	Min	Max	Mean	Median	SD	Skewness	Kurtosis	CV
Original	0	440	44.32	9	73.76	2.49	9.82	1.66
square root	0	20.98	4.71	3	4.71	1.12	3.59	0.99

Even after the square root transformation, the skewness still persists (see Chaves et al. (2020)) and, on the other hand, the original distribution is not easy to model. In this case, the use of scale mixture of skew-normal distributions can be suitable. Figures 2a-2c present the Scatter plots between the squared root of the pollen concentration and the explanatory variables along with regression curves fitted by LOESS. They indicate a nonlinear relationship between the response variable and days, in such a way that a non-parametric structure could be suitable for modeling such relation. For the other covariates, it is suitable to assume a linear relationship. Ferreira and Paula (2017) used a skew-normal partially linear model, comparing the respective results with those obtained by the fit through a normal partially linear model.

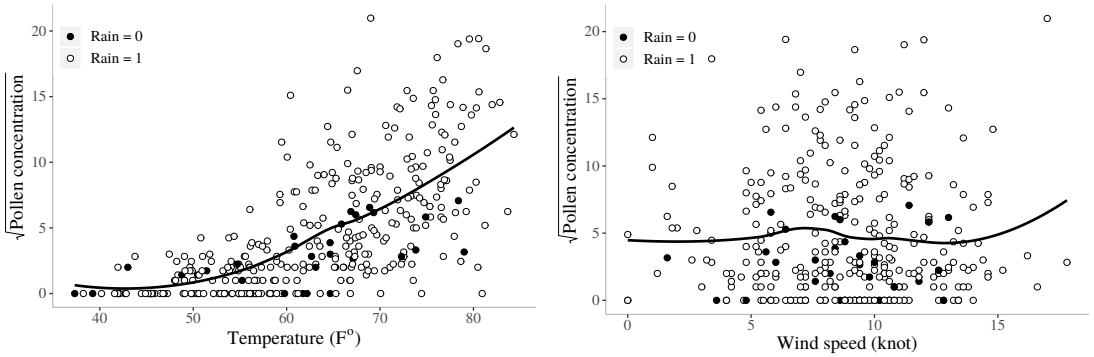
Therefore, the first part of this dissertation comprises the following contribu-

tions:

1. We review a family of distributions, namely, scale mixture of skew-normal densities, useful for modeling skewed and/or heavy-tailed data. The related centered parameterization, which presents several advantages, in terms of inferential and interpretation aspects , over the original parameterization, is presented. Considering additional mixing distributions, besides the usual ones, we developed new members for this class namely, generalized Gamma, Birnbaum-Saunders and Beta Prime, which can be useful alternatives for the usual members.

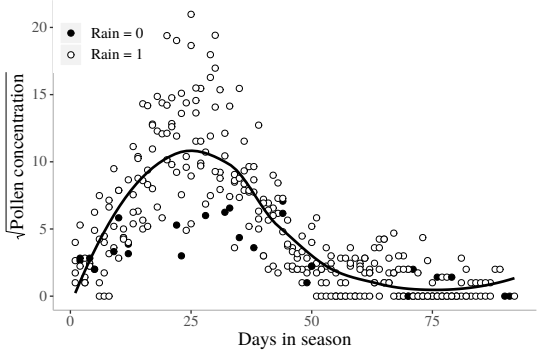
2. We develop a class of additive partially linear models based on this family.

3. We present the necessary/required inferential tooling, as well as goodness of fit measurements, diagnostic techniques of global and local influence, simulation studies and application to the ragweed pollen data showing the good performance of the developed models.



(a) Scatter plot between the squared root of the pollen concentration and temperature, fitted by LOESS.

(b) Scatter plot between the squared root of the pollen concentration and wind speed, fitted by LOESS.



(c) Scatter plot between the squared root of the pollen concentration and days in season, fitted by LOESS.

Figure 2 – Scatter plots between the squared root of the pollen concentration and the temperature, wind speed and days in season, fitted by LOESS curve.

1.1.2 Literary review

The normal distribution has been widely used in many areas of knowledge since the early days of statistical modeling, mainly due to mathematical and computational advantages over the others distributions. However, very often, the data sets present some departing from normality as: asymmetry and/or heavy/light tails. In this case the use of the normal distribution (or other symmetric distributions) may lead to misleading inference (Eling, 2012).

To handle this issue, several asymmetric and/or heavy tails distributions have been proposed in the literature. The skew-normal distribution (Azzalini, 1985), was one the first generalizations of the normal one, allowing several different shapes in terms of asymmetry. Ferreira et al. (2011) presented a class called scale mixture of skew-normal distributions (SMSN), which is an extension of the scale mixture of normal distributions (Andrews and Mallows, 1974). The SMSN distributions are very useful for analyzing skewed and/or heavy tailed data. As particular cases we have: the skew-normal distribution, skew Student-t, skew slash, skew generalized t and skew contaminated normal, besides their respective symmetrical cases. This class has been recently improved, in order to circumvent several inferential problems and limitations, for (Maioli, 2018b), using the centered version of the skew-normal distribution (Arellano-Valle and Azzalini, 2008).

Due to its flexibility the SMSN family has been widely studied and regression models have been proposed based on such class. Basso et al. (2010) showed the inference for finite mixture of SMSN. Lachos et al. (2010a) proposed a multivariate linear error-in-variables regression model based on the scale mixtures of multivariate skew-normal distributions. Cancho et al. (2011) developed a Bayesian nonlinear regression model for SMSN family. Zeller et al. (2011) developed local influence analysis for the multivariate SMSN regression models. Garay et al. (2011) proposed a non-linear regression model using SMSN distributions. da Silva Ferreira et al. (2011) developed several properties for the SMSN family of distributions, also discussing the inference process and the obtaining of the respective standard errors. Lachos et al. (2011) developed a heteroscedastic non-linear regression model for SMSN. Labra et al. (2012) built global and local influence analysis for heteroscedastic non-linear regression models based on the SMSN distributions. Zeller et al. (2012) developed the local influence analysis for multivariate error-in-variables regression models for SMSN distributions. Garay et al. (2014) presented the local influence analysis and generalized Cook's distance for nonlinear regression models based on SMSN distributions. Ferreira et al. (2015) showed the global and local influence analysis for SMSN models, as well as generalized leverage. Zeller et al. (2016) proposed a linear regression model based on finite mixture of SMSN distributions. Ferreira et al. (2016) proposed a multivariate

regression model based on SMSN distributions. Ferreira and Lachos (2016) showed a non-linear regression model for finite mixture of SMSN. Massuia et al. (2017) developed, under the Bayesian paradigm, a censored regression model for the SMSN family. Galarza Morales et al. (2017) proposed a quantile regression model for finite mixture of SMSN distributions. Mattos et al. (2018) showed a censored regression model for SMSN under the Frequentist paradigm. Castro et al. (2019) developed a semi-parametric Bayesian mixed model for the SMSN family, using wavelets penalization. da Silva Ferreira et al. (2019) developed diagnostic tools for heteroscedastic non-linear regression models based on finite mixture of SMSN. Hajrajabi and Maleki (2019) developed an autoregressive non-linear semi-parametric regression model for the SMSN class.

It is worth to note that all the literature cited above is based on the usual parameterization of the skew-normal, which can present some problems, that we will discuss later.

1.2 Centered skew-normal distribution

A random variable Y follows a skew-normal distribution with location parameter $\alpha \in \mathbb{R}$, scale parameter $\beta \in \mathbb{R}^+$ and skewness parameter $\lambda_s \in \mathbb{R}$, i.e., $Y \sim SN(\alpha, \beta^2, \lambda)$, if its probability density function (p.d.f) is given by (Azzalini, 1985):

$$f_Y(y|\alpha, \beta, \lambda_s) = \frac{2}{\beta}\phi\left(\frac{y-\alpha}{\beta}\right)\Phi\left(\lambda_s\left(\frac{y-\alpha}{\beta}\right)\right)\mathbb{1}_{(-\infty,\infty)}(y),$$

where ϕ and Φ stands for the p.d.f and cumulative function of standard Normal distribution. Note that if $\lambda_s = 0$, the normal distribution is recovered. Multivariate versions of this distribution can be found in Azzalini and Valle (1996) and Padilla et al. (2018).

On the other hand, Henze (1986) proposed an useful stochastic representation of the (original) skew-normal, which is given by:

$$Y = \alpha + \beta\left(\delta H + \sqrt{1-\delta^2}T\right), \tag{1.1}$$

where $H \sim HN(0,1) \perp T \sim N(0,1)$, $HN(0,1)$ denotes the standard half-normal distribution (i.e., a standard Normal distribution truncated below zero) and $\delta = \lambda_s/\sqrt{1+\lambda_s^2}$. From (1.1) we have that the expectation and variance of Y are given, respectively, by

$$\mathbb{E}(Y) = \alpha + \beta\delta\sqrt{\frac{2}{\pi}} \quad \text{and} \quad \mathrm{Var}(Y) = \beta^2\left(1 - \delta^2\frac{2}{\pi}\right).$$

Arellano-Vale and Azzalini (2008) presented some problems, through a practical example, induced by the above parameterization. Among them when $\lambda \to 0$, the expected Fisher information matrix is singular, even if all parameters are identifiable,

which leads to: lack of some asymptotic proprieties of the respective maximum likelihood estimators and a non quadratic shape of the log-likelihood. Even under Bayesian inference, these problems still remain, unless (very) informative priors be used Maioli (2018a); Azevedo et al. (2018). These problems can be solved if we use the centered parameterization of the skew-normal distribution proposed by Azzalini (1985).

It is said that Y_c has a skew-normal distribution under the centered parameterization (or centered skew-normal distribution) with centered parameters $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$ and $\gamma \in (-0,99527; 0,99527)$, denoted by $Y_c \sim SN_c(\mu, \sigma^2, \gamma)$, if

$$Y_c = \mu + \sigma Z_0, \tag{1.2}$$

where $Z_0 = (Z - \mu_z)/\sigma_z$, $Z \sim SN(0, 1, \lambda_s)$, $\mu_z = \sqrt{2/\pi}\delta$ and $\sigma_z = \sqrt{1-(2/\pi)\delta^2}$. We also have:

$$\mu = \mathbb{E}(Y_c) = \alpha + \beta\mu_z,\ \sigma^2 = \mathrm{Var}(Y_c) = \beta^2(1-\mu_z^2) \text{ and } \gamma = \frac{4-\pi}{2}\frac{\left(\sqrt{\frac{2}{\pi}}\delta\right)^3}{\left(1-\frac{2}{\pi}\delta^2\right)^{3/2}},$$

where γ denotes the Pearson's index of Skewness. Using the Jacobian transformation method, the p.d.f of (1.2) is given by

$$f_Y(y|\mu, \sigma^2, \gamma) = \frac{2}{\omega}\phi\left(\frac{y-\xi}{\omega}\right)\Phi\left(\lambda\left(\frac{y-\xi}{\omega}\right)\right),$$

where

$$\xi = \mu - \sigma\gamma^{1/3}s,\ s = \left(\frac{2}{4-\pi}\right)^{1/3},\ \omega = \sigma\sqrt{1+s^2\gamma^{2/3}} \text{ and } \lambda = \frac{s\gamma^{1/3}}{\sqrt{\frac{2}{\pi}+s^2\gamma^{2/3}\left(\frac{2}{\pi}-1\right)}}.$$

Notice that the $SN_c(\mu, \sigma^2, \gamma)$ is equivalent to $SN(\xi, \omega^2, \lambda)$. As in (1.1), the Henze's stochastic representation for the centered skew-normal distribution is given by

$$Y_c = \xi + \omega\left(\delta H + \sqrt{1-\delta^2}T\right), \tag{1.3}$$

with H, T, ξ, ω and δ as defined above.

1.3 Scale mixture of skew-normal distribution under the centered parameterization

In this section we present the scale mixture of centered skew-normal distributions. We considered, as mixing measure, usual choice as gamma, beta and binary distributions as well as interesting alternatives as: beta prime, Birnbaum-Saunders and generalized gamma.

A random variable Y follows a scale mixture of skew-normal distribution under the centered parameterization (or scale mixture of centered skew-normal distribution - SMCSN) with mean μ, scale parameter σ^2, skewness parameter γ, mixing measure H and shape parameters $\boldsymbol{\nu}$, i.e, $Y \sim$ SMCSN$(\mu, \sigma^2, \gamma, H, \boldsymbol{\nu})$, if it can be stochastilly represented by

$$Y = \mu + k(U)^{1/2} W, \tag{1.4}$$

where $W \sim SN_c(0, \sigma^2, \gamma) \perp U$, $k(.)$ is a positive real arbitrary function and U is a mixing distribution, with cdf $H(.; \boldsymbol{\nu})$. Also, it is possible to prove that:

- If $\mathbb{E}\left(\sqrt{k(U)}\right) < \infty$, then $\mathbb{E}(Y) = \mu$;
- If $\mathbb{E}\left(k(U)\right) < \infty$, then $\mathrm{Var}(Y) = \sigma^2 k_2$;
- If $\mathbb{E}\left(k(U)\right) < \infty$ and $\mathbb{E}\left(\sqrt{k(U)}\right) < \infty$ then the skewness coefficient is given by

$$\tilde{\mu}_3 = \frac{\mathbb{E}(Y - \mathbb{E}(Y))^3}{(\mathrm{Var}(Y))^{3/2}} = \frac{\mathbb{E}(k^{1/2}(U)W)^3}{\sigma^3 k_2^{3/2}} = \frac{k_3}{k_2^{3/2}}\gamma;$$

- If $\mathbb{E}\left(k(U)\right) < \infty$ and $\mathbb{E}\left(\sqrt{k(U)}\right) < \infty$ then the kurtosis coefficient is given by

$$\tilde{\mu}_4 = \frac{\mathbb{E}(Y - \mathbb{E}(Y))^4}{(\mathrm{Var}(Y))^2} = \frac{\mathbb{E}(k^{1/2}(U)W)^4}{\sigma^4 k_2^2} = \frac{k_4}{\sigma^4 k_2^2}\mathbb{E}(W^4)$$

  where $k_m = \mathbb{E}[k(U)^{m/2}]$.

Note that if $\gamma = 0$, the SMCSN family reduces to the scale mixture of normal distributions family. Let $Y \sim$ SMCSN$(\mu, \sigma^2, \gamma, H, \boldsymbol{\nu})$. Then, $Y|U = u \sim SN_c(\mu, \sigma^2 k(u), \gamma)$, $U \sim H(.|\boldsymbol{\nu})$. Analogous to the Henze's representation given in (1.3), we also have that $Y|U = u, H = h \sim N(\xi_u + \omega_u \delta h, \omega_u^2(1 - \delta^2))$, $H \sim HN(0, 1)$, $U \sim H(.|\boldsymbol{\nu})$, where

$$\xi_u = \mu - \sigma\sqrt{k(u)}\gamma^{1/3} s,\ s = \left(\frac{2}{4-\pi}\right)^{1/3}, \omega_u = \sigma\sqrt{k(u)}\sqrt{1 + s^2\gamma^{2/3}},\ \delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$$

$$\lambda = \frac{s\gamma^{1/3}}{\sqrt{2/\pi + s^2\gamma^{2/3}(2/\pi - 1)}}.$$

The p.d.f of (1.4) is given by

$$f(y|\mu, \sigma^2, \gamma, \boldsymbol{\nu}) = \int_0^\infty \frac{2}{\omega_u}\phi\left(\frac{y - \xi_u}{\omega_u}\right)\Phi\left(\lambda\left(\frac{y - \xi_u}{\omega_u}\right)\right) dH(u|\boldsymbol{\nu}),$$

where

$$\frac{y - \xi_u}{\omega_u} = \frac{y - (\mu - \sigma\sqrt{k(u)}\gamma^{1/3}s)}{\sigma\sqrt{k(u)}\sqrt{1 + s^2\gamma^{2/3}}} = \frac{(y - \mu) + \sigma\sqrt{k(u)}\gamma^{1/3}s}{\sigma\sqrt{k(u)}\sqrt{1 + s^2\gamma^{2/3}}}$$

$$= \frac{(y-\mu)}{\sigma\sqrt{k(u)}\sqrt{1+s^2\gamma^{2/3}}} + \frac{\sigma\sqrt{k(u)}\gamma^{1/3}s}{\sigma\sqrt{k(u)}\sqrt{1+s^2\gamma^{2/3}}} = d\frac{1}{\sqrt{k(u)}} + \frac{\gamma^{1/3}s}{\sqrt{1+s^2\gamma^{2/3}}}$$
$$= d\frac{1}{\sqrt{k(u)}} - \frac{\xi_1}{\omega_1},$$

and

$$\omega_1 = \sqrt{1+s^2\gamma^{2/3}},\ \xi_1 = -\gamma^{1/3}s \text{ and } d = \frac{y-\mu}{\sigma\sqrt{1+s^2\gamma^{2/3}}} = \frac{y-\mu}{\sigma\omega_1}.$$

1.4 Examples of Scale Mixtures of Centered Skew-Normal Distribution

The following will show the density of SMCSN distributions based on the Gamma, Beta, Binary, BP, BS and GG distributions as mixing measure. Here we consider $k(u) = 1/u$.

1.4.1 Centered skew generalized t distribution

Considering $U \sim$ gamma$(\nu_1/2, \nu_2/2)$ with pdf:

$$h(u|\nu_1,\nu_2) = \frac{(\nu_2/2)^{\nu_1/2}}{\Gamma(\nu_1/2)}u^{\nu_1/2-1}\exp\left\{-\frac{\nu_2 u}{2}\right\}\mathbb{1}_{(0,\infty)}(u), \tag{1.5}$$

we obtain the Centered skew generalized t distribution denoted by $Y \sim CSGT(\mu, \sigma^2, \gamma, \nu_1, \nu_2)$. As noted by Maioli (2018a), this distribution has some identifiability issues, since different values of $(\sigma^2, \nu_2)^\top$ can lead to the same likelihood. Maioli (2018a) proposed to use $\sigma^2 = 1$ to avoid this problem. Therefore, in this work we assume $Y \sim CSGT_{\sigma^2=1}(\mu, \gamma, \nu_1, \nu_2)$ whose p.d.f is given by:

$$f(y|\mu,\gamma,\nu_1,\nu_2) = \frac{2^{\frac{\nu_2}{2}\frac{\nu_1}{2}}}{\omega_1\sqrt{2\pi}\Gamma\left(\frac{\nu_1}{2}\right)}e^{-\frac{\xi_1^2}{2\omega_1^2}}\int_0^\infty u^{\frac{\nu_1+1}{2}-1}e^{-\frac{1}{2}\left[u\left(\frac{(u-\mu)^2}{\omega_1^2}+\nu_2\right)-2\sqrt{u}(y-\mu)\frac{\xi_1}{\omega_1^2}\right]}\times$$
$$\times\ \Phi\left(\lambda\left(\sqrt{u}\frac{(y-\mu)}{\omega_1} - \frac{\xi_1}{\omega_1}\right)\right)du,$$

where μ is the mean, γ is the skewness and ν_1 and ν_2 are the shape parameters, since $Var(Y) = \nu_2/(\nu_1 - 2)$. Note that if $\nu_1 = \nu_2 = \nu$ the CSGT distribution is reduced to Centered skew t distribution (CST) with $\sigma^2 = 1$.

1.4.2 Centered skew t distribution

Considering $U \sim$ gamma$(\nu/2, \nu/2)$ whose p.d.f is given by:

$$h(u|\nu_1,\nu_2) = \frac{(\nu_2/2)^{\nu_1/2}}{\Gamma(\nu_1/2)}u^{\nu_1/2-1}\exp\left\{-\frac{\nu_2 u}{2}\right\}\mathbb{1}_{(0,\infty)}(u),$$

we obtain the Centered skew t distribution denoted by $CST(\mu, \sigma^2, \gamma, \nu)$ whose p.d.f is given by:

$$f(y|\mu,\sigma^2,\gamma,\nu) = \frac{2(\nu/2)^{\nu/2} e^{-\frac{\xi_1^2}{2\omega_1^2}}}{\sigma\omega_1\sqrt{2\pi}\Gamma(\nu/2)} \int_0^\infty u^{\frac{\nu+1}{2}} \exp\left\{-\frac{1}{2}\left[u(d^2+\nu) - 2\sqrt{u}d\frac{\xi_1}{\omega_1}\right]\right\} \times \Phi\left(\lambda\left(\sqrt{u}d - \frac{\xi_1}{\omega_1}\right)\right) du,$$

where μ is the mean, γ is the skewness parameter, ν is the degree of freedom and $Var(Y) = \sigma^2\nu/(\nu-2)$.

1.4.3 Centered Skew-Slash distribution

Considering $U \sim$ beta$(\nu, 1)$ whose p.d.f is given by $h(u|\nu) = \nu u^{\nu-1}\mathbb{1}_{(0,1)}(u)$, we obtain the Centered Skew-Slash distribution denoted by $CSS(\mu, \sigma^2, \gamma, \nu)$ whose p.d.f is given by:

$$f(y|\mu,\sigma^2,\gamma,\nu) = \frac{2\nu}{\sigma\omega_1\sqrt{2\pi}} e^{-\frac{\xi_1^2}{2\omega_1^2}} \int_0^1 u^{\nu-\frac{1}{2}} e^{-\frac{1}{2}\left[ud^2 - 2\sqrt{u}d\frac{\xi_1}{\omega_1}\right]} \Phi\left(\lambda\left(\sqrt{u}d - \frac{\xi_1}{\omega_1}\right)\right) du,$$

where μ is the mean, γ is the skewness parameter, ν is the degree of freedom and $Var(Y) = \sigma^2\nu/(\nu-1)$.

1.4.4 Centered Skew-contaminated Normal distribution

Considering U a discrete random variable whose p.d.f is given by $h(u|\nu_1,\nu_2) = \nu_1\mathbb{1}_{u=\nu_2}(u) + (1-\nu_1)\mathbb{1}_{u=1}(u)$, we obtain the Centered Skew-contaminated Normal distribution denoted by $Y \sim$ CSCN$(\mu, \sigma^2, \gamma, \nu_1, \nu_2)$ whose p.d.f is given by:

$$f(y|\mu,\sigma^2,\gamma,\nu_1,\nu_2) = 2\left[\nu_1\frac{\sqrt{\nu_2}}{\sigma\omega_1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\sqrt{\nu_2}d - \frac{\xi_1}{\omega_1}\right)^2} \Phi\left(\lambda\left(\sqrt{\nu_2}d - \frac{\xi_1}{\omega_1}\right)\right) + (1-\nu_1)\frac{1}{\sigma\omega_1\sqrt{2\pi}} e^{-\frac{1}{2}\left(d - \frac{\xi_1}{\omega_1}\right)^2} \Phi\left(\lambda\left(d - \frac{\xi_1}{\omega_1}\right)\right)\right],$$

where μ is the mean, γ is the skewness parameter, ν_1 and ν_2 are the proportion of outliers and scale parameter, respectively, and $Var(Y) = \sigma^2(\nu_1 + \nu_2(1-\nu_1))/\nu_2$.

1.5 New distributions of scale mixture of skew-normal under the centered parameterization

In this section we develop three new members of the CSMN family based on the: Beta prime (BP) (Keeping, 1962), Birnbaum-Saunders (BS) (Birnbaum and Saunders, 1969b,a) and generalized gamma (GG) (Stacy, 1962) distributions.

1.5.1 Centred skew Beta Prime Normal

Considering $U \sim$ beta prime(ν_1, ν_2) whose p.d.f is given by

$$\frac{u^{\nu_1-1}}{B(\nu_1,\nu_2)(1+u)^{(\nu_1+\nu_2)}}\mathbb{1}_{(0,\infty)}(u),$$

we obtain the Centred skew Beta Prime Normal, denoted by $Y \sim CSBPN(\mu, \sigma^2, \nu_1, \nu_2)$, whose p.d.f is given by:

$$f(y|\mu,\sigma^2,\gamma,\nu_1,\nu_2) = \frac{2e^{-\frac{\xi_1^2}{2\omega_1^2}}}{\sigma\omega_1\sqrt{2\pi}B(\nu_1,\nu_2)}\int_0^\infty \frac{u^{\frac{\nu_1-1}{2}}}{(u+1)^{\nu_1+\nu_2}} e^{-\frac{1}{2}\left(ud^2-2\sqrt{u}d\frac{\xi_1}{\omega_1}\right)}\Phi\left[\lambda\left(\sqrt{u}d - \frac{\xi_1}{\omega_1}\right)\right]du,$$

where μ is the mean, γ is the skewness, ν_1 and ν_2 are the shape parameters and $Var(Y) = \sigma^2\nu_2/(\nu_1 - 1)$. As special case when $\gamma = 0$ we have the beta prime normal distribution, whose p.d.f is given by:

$$\begin{aligned} f(y|\mu,\sigma^2,\nu_1,\nu_2) &= \int_0^\infty \frac{\sqrt{u}}{\sqrt{2\pi}\sigma}e^{-\frac{ud}{2}}\frac{u^{\nu_1-1}}{B(\nu_1,\nu_2)(1+u)^{(\nu_1+\nu_2)}}du \\ &= \frac{1}{\sqrt{2\pi}\sigma B(\nu_1,\nu_2)}\int_0^\infty \frac{u^{\nu_1-1/2}}{(1+u)^{(\nu_1+\nu_2)}}e^{-\frac{ud}{2}}du \\ &= \frac{1}{\sqrt{2\pi}\sigma B(\nu_1,\nu_2)}\left(\frac{d}{2}\right)^{\nu_1+1/2} G_{2,1}^{1,2}\left(\frac{2}{d}\middle| \begin{matrix} 1/2-\nu_1, 1-(\nu_1+\nu_2) \\ 0 \end{matrix}\right), \end{aligned}$$

where G is the Meijer G-function (Meijer, 1936) given by:

$$G_{p,q}^{m,n}\left(x\middle| \begin{matrix} a_1,\ldots,a_p \\ b_1,\ldots,b_q \end{matrix}\right) = \frac{1}{2\pi i}\int_L \frac{\prod_{j=1}^m\Gamma(b_j+t)\prod_{j=1}^n\Gamma(1-a_j-t)}{\prod_{j=n+1}^p\Gamma(a_j+t)\prod_{j=m+1}^q\Gamma(1-b_j-t)}x^{-t}dt,$$

where $i = \sqrt{-1}$ is the complex unit and L denotes an integration path (see, Gradshteyn and Ryzhik (2014), for example).

1.5.2 skew Birnbaum-Saunders Normal

Considering $U \sim BS(\nu_1, \nu_2)$ whose p.d.f is given by

$$h(u|\nu_1,\nu_2) = \frac{1}{2\sqrt{2\pi}\nu_1\nu_2}\left[\left(\frac{\nu_2}{u}\right)^{1/2} + \left(\frac{\nu_2}{u}\right)^{3/2}\right]\exp\left\{-\frac{1}{2\nu_1^2}\left(\frac{u}{\nu_2}+\frac{\nu_2}{u}-2\right)\right\}\mathbb{1}_{(0,\infty)}(u),$$

we obtain the Centred skew Birnbaum-Saunders Normal, denoted by $Y \sim$ CSBSN$(\mu,$ $\sigma^2, \gamma, \nu_1, \nu_2)$, whose p.d.f is given by:

$$f(y|\mu,\sigma^2,\gamma,\nu_1,\nu_2) = \frac{e^{\frac{1}{\nu_1^2}-\frac{\xi_1^2}{2\omega_1^2}}}{\sigma\omega_1 2\pi\nu_1\sqrt{\nu_2}}\int_0^\infty \exp\left\{-\frac{1}{2}\left(ud^2 - 2\sqrt{u}d\frac{\xi_1}{\omega_1}+\frac{1}{\nu_1^2}\left(\frac{u}{\nu_2}+\frac{\nu_2}{u}\right)\right)\right\}\times$$

$$\times \Phi\left(\lambda\left(\sqrt{u}d - \frac{\xi_1}{\omega_1}\right)\right)\left[1 + \frac{\nu_2}{u}\right]du,$$

where μ is the mean, γ is the skewness, ν_1 and ν_2 shape parameters and $Var(Y) = \sigma^2(\nu_1^2 + 2)/(2\nu_2)$. As special case when $\gamma = 0$ we have the Birnbaum-saunders normal distribution whose p.d.f is given by:

$$f(y|\mu, \sigma^2, \nu_1, \nu_2) = \frac{e^{\frac{1}{\nu_1^2}}\sqrt{\nu_2}}{2\pi\sigma\nu_1}\left[K_0\left(\beta^*\right) + \frac{\nu_1^2}{1 + d\nu_2\nu_1^2}K_{-1}(\beta^*)\right],$$

with $\beta^* = \sqrt{1 + d\nu_2\nu_1^2}/\nu_1^2$ and $K_\eta(z)$ is the modified Bessel function given by

$$K_\eta(z) = \frac{1}{2}\left(\frac{z}{2}\right)^\eta \int_0^\infty \exp\left\{-t - \frac{z^2}{4t}\right\}\frac{1}{t^{\nu+1}}dt. \tag{1.6}$$

1.5.3 Centred skew generalized Gamma Normal

Considering $U \sim GG(\nu_1, \nu_2, \nu_3)$ whose p.d.f is given by

$$h(u|\nu_1, \nu_2, \nu_3) = \frac{\nu_2}{\nu_3\Gamma(\nu_1)}\left(\frac{u}{\nu_3}\right)^{\nu_1\nu_2 - 1}\exp\left\{-\left(\frac{u}{\nu_3}\right)^{\nu_2}\right\}\mathbb{1}_{(0,\infty)}(u),$$

we obtain the Centred skew generalized Gamma Normal, denoted by $Y \sim CSGGN(\mu, \sigma^2, \gamma, \nu_1, \nu_2, \nu_3)$, whose p.d.f is given by:

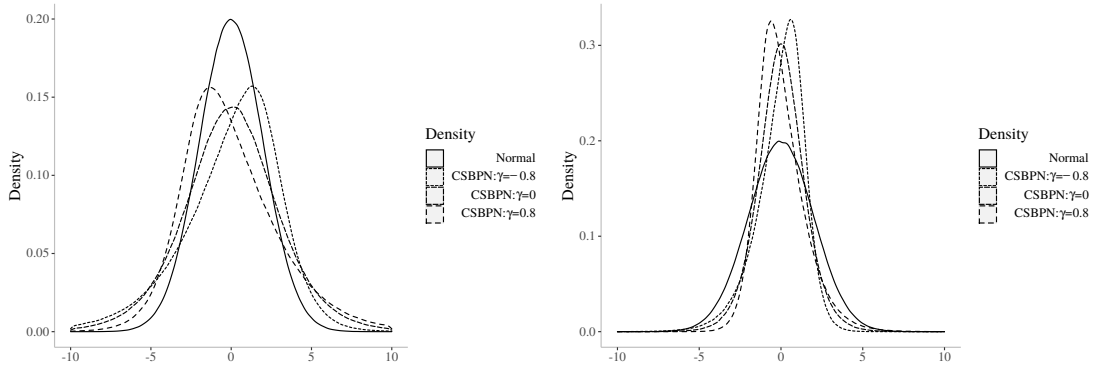
$$f(y|\mu, \sigma^2, \gamma, \nu_1, \nu_2) = \frac{2e^{-\frac{\xi_1^2}{2\omega_1^2}}\nu_2}{\sigma\omega_1\sqrt{2\pi}\nu_3^{\nu_1\nu_2}\Gamma(\nu_1)}\int_0^\infty \exp\left\{-\frac{1}{2}\left(ud^2 - 2\sqrt{u}d\frac{\xi_1}{\omega_1} + 2\left(\frac{u}{\nu_3}\right)^{\nu_2}\right)\right\}\times$$
$$\times \Phi\left(\lambda\left(\sqrt{u}d - \frac{\xi_1}{\omega_1}\right)\right)u^{\nu_1\nu_2 - 1/2}du,$$

where μ is the mean, γ is the skewness, ν_1, ν_2 and ν_3 are shape parameters and $Var(Y) = \sigma^2\Gamma(\nu_1 - 1/\nu_2)/[\nu_3\Gamma(\nu_1)]$. As special case when $\gamma = 0$ we have the generalized gamma normal distribution whose p.d.f is given by:

$$f(y|\mu, \sigma^2, \nu_1, \nu_2, \nu_3) = \frac{1}{\sqrt{2\pi}\sigma\Gamma(\nu_1)}\sum_{m=0}^{\infty}(-1)^m\frac{(\nu_3 d/2)^m}{m!}\Gamma\left(\nu_1 + \frac{m}{\nu_1} + \frac{1}{2\nu_2} - 1\right).$$

The following figures present examples of densities of aforementioned distributions . In Figures (3a) and (3b) we have the skew Beta Prime Normal with $\mu = 0$, $\sigma = 4$, varying $(\nu_1, \nu_2) = (5, 10)$ and $(\nu_1, \nu_2) = (10, 5)$ and γ by $-0.9, 0$ and 0.9. In Figures (4a) and (4b) we have the skew Birnbaum-Saunders Normal with $\mu = 0$, $\sigma = 4$ and $(\nu_1, \nu_2) = (1, 2)$ and $(\nu_1, \nu_2) = (4, 2)$, and γ by $-0.9, 0$ and 0.9. In Figures (5a) and (5b) we have the skew generalized Gamma Normal with $\mu = 0$, $\sigma = 4$ and varying $(\nu_1, \nu_2, \nu_3) = (2, 1, 4)$ and $(\nu_1, \nu_2, \nu_3) = (4, 2, 1)$, and γ by $-0.9, 0$ and 0.9.

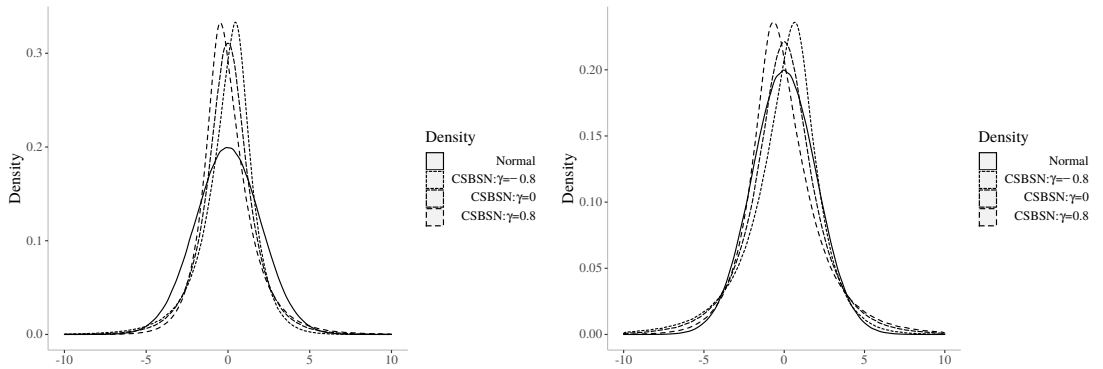
From Figures from 6 to 14, we can see that the considered members of the SMCSN family may present different behaviors in terms of variability, skewness and



(a) Probability density function of the skew Beta Prime Normal with $\nu_1 = 5$ and $\nu_2 = 10$ varying the skewness parameter in $-0.9, 0$ and 0.9.

(b) Probability density function of the skew Beta Prime Normal with $\nu_1 = 10$ and $\nu_2 = 5$ varying the skewness parameter in $-0.9, 0$ and 0.9.

Figure 3 – Probability density function of the skew Beta Prime Normal with $\mu = 0$, $\sigma = 4$ and varying the shape and skewness parameters.

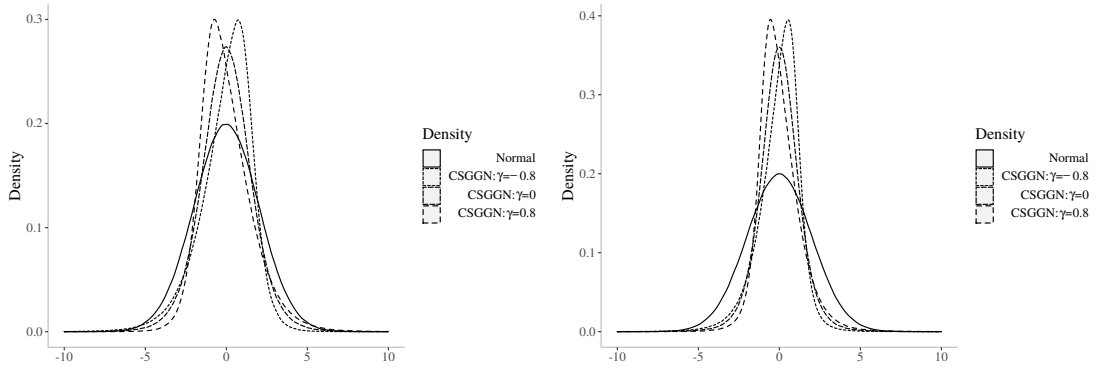


(a) Probability density function of the skew Birnbaum-Saunders Normal with $\nu_1 = 1$ and $\nu_2 = 2$ varying the skewness parameter in $-0.9, 0$ and 0.9.

(b) Probability density function of the skew Birnbaum-Saunders Normal with $\nu_1 = 4$ and $\nu_2 = 2$ varying the skewness parameter in $-0.9, 0$ and 0.9.

Figure 4 – Probability density function of the skew Birnbaum-Saunders Normal with $\mu = 0$, $\sigma = 4$ and varying the shape and skewness parameters.

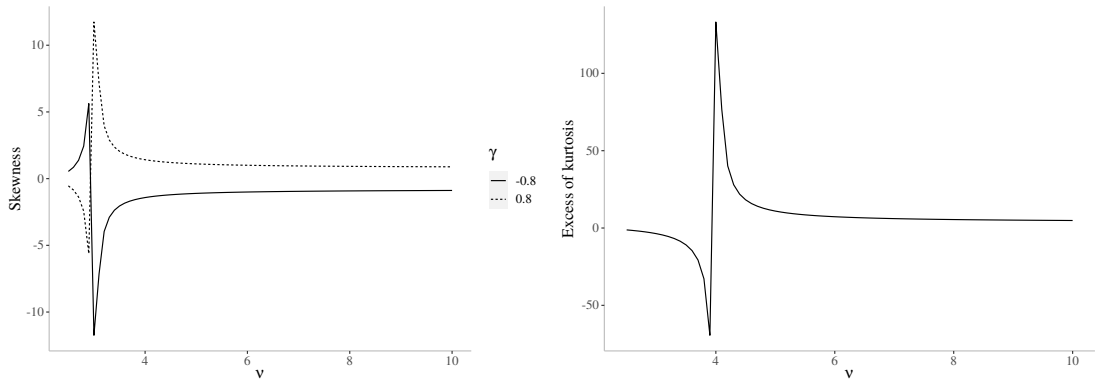
kurtosis. For example, for the CST distribution the skewness depends not only on γ but also on $\boldsymbol{\nu}$. Therefore, more than one parameter can affects the behavior of the variability, kurtosis and skewness. For the variability we can observe that the variance is directly influenced by $E(U^{-1})$. In addition, it is noted that certain distributions may have kurtosis and skewness greater (in module) than others.



(a) Probability density function of the skew generalized Gamma Normal with $\nu_1 = 2$, $\nu_2 = 1$ and $\nu_3 = 4$ varying the skewness parameter in $-0.9, 0$ and 0.9.

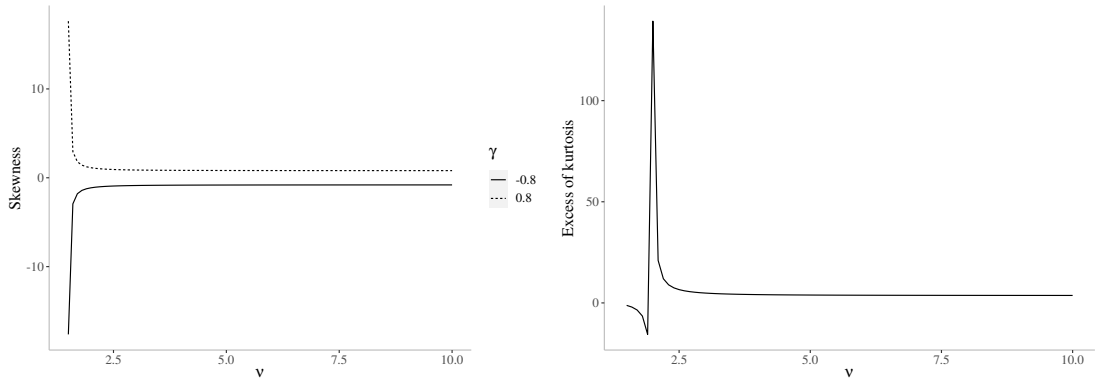
(b) Probability density function of the skew generalized Gamma Normal with $\nu_1 = 4$, $\nu_2 = 2$ and $\nu_3 = 1$ varying the skewness parameter in $-0.9, 0$ and 0.9.

Figure 5 – Probability density function of the skew generalized Gamma Normal with $\mu = 0$, $\sigma = 4$ and varying the shape and skewness parameters.



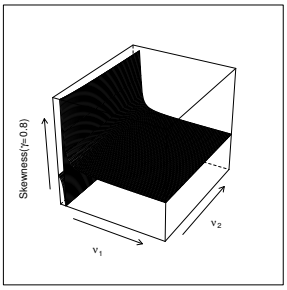
(a) (b)

Figure 6 – Skewness (a) and excess of kurtosis (b) for CST distribution with $\mu = 0$, $\sigma = 1$ and varying shape and skewness parameters.

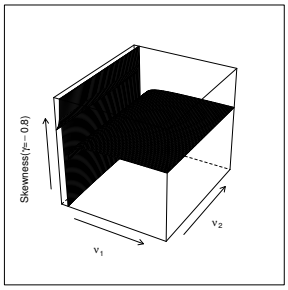


(a) (b)

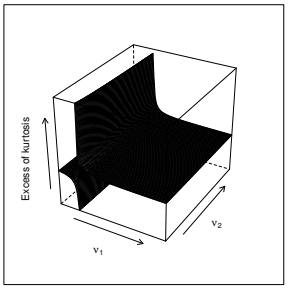
Figure 7 – Skewness (a) and excess of kurtosis (b) for the CSS distribution with $\mu = 0$, $\sigma = 1$ and varying the shape and skewness parameters.



(a)

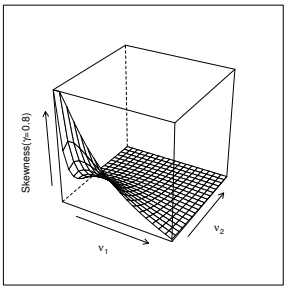


(b)

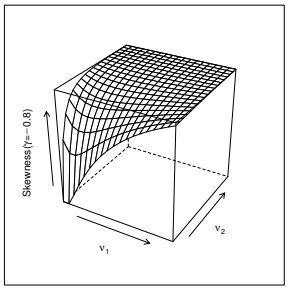


(c)

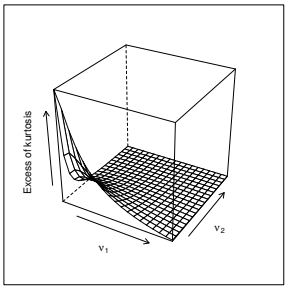
Figure 8 – Skewness (a,b) and excess of kurtosis (c) for the CSGT distribution with $\mu = 0$, $\sigma = 1$ and varying the shape and parameters.



(a)

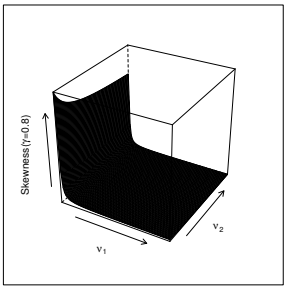


(b)

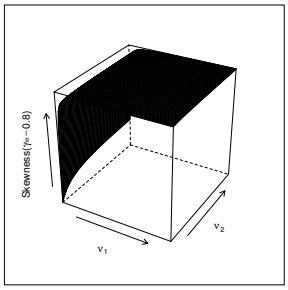


(c)

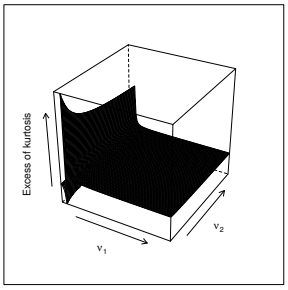
Figure 9 – Skewness (a,b) and excess of kurtosis (c) for the CSCN distribution with $\mu = 0$, $\sigma = 1$ and varying the shape and parameters.



(a)

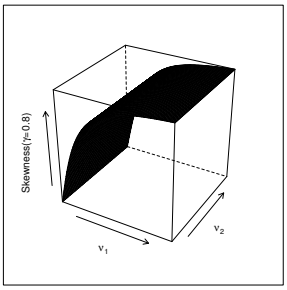


(b)

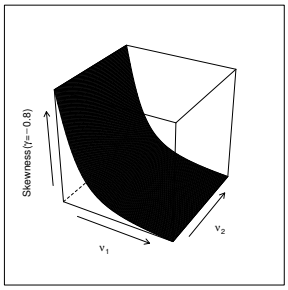


(c)

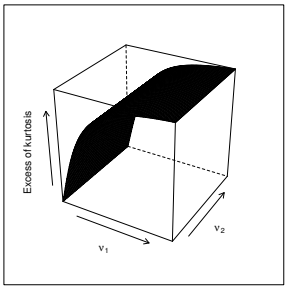
Figure 10 – Skewness (a,b) and excess of kurtosis (c) for CSBPN distribution with $\mu = 0$, $\sigma = 1$ and varying shape and parameters.



(a)

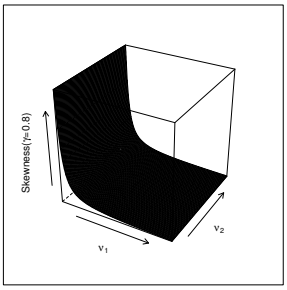


(b)

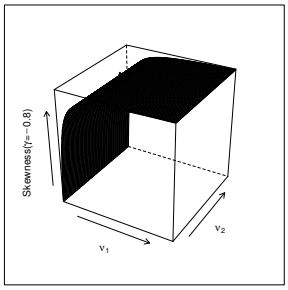


(c)

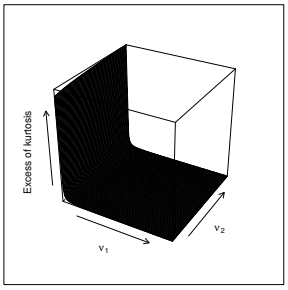
Figure 11 – Skewness (a,b) and excess of kurtosis (c) for the CSBSN distribution with $\mu = 0$, $\sigma = 1$ and varying the shape and parameters.



(a)

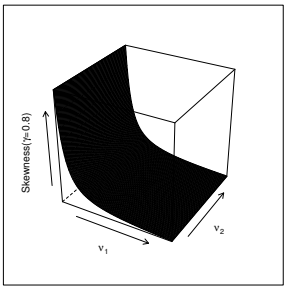


(b)

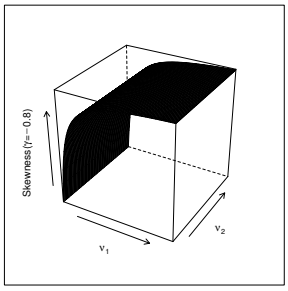


(c)

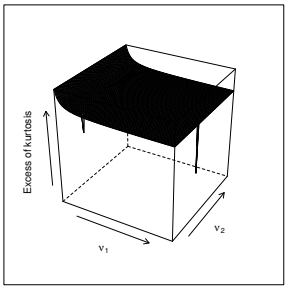
Figure 12 – Skewness (a,b) and excess of kurtosis (c) for the CSGGN distribution with $\mu = 0$, $\sigma = 1$, varying the skewness parameter, ν_1 and ν_2, and fixed $\nu_3 = 2$.



(a)

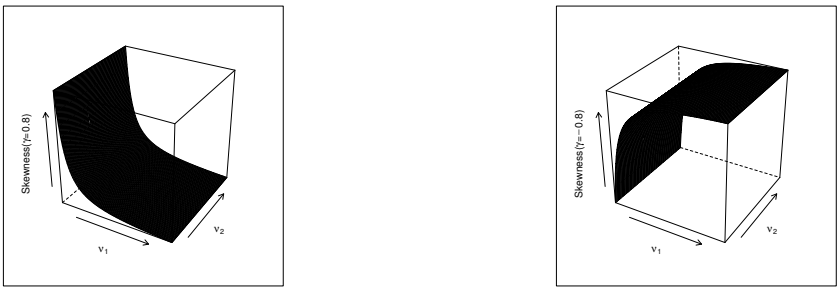


(b)

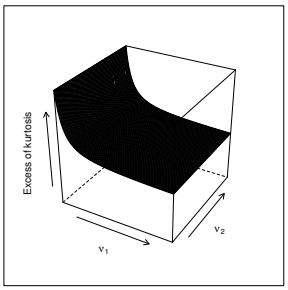


(c)

Figure 13 – Skewness (a,b) and excess of kurtosis (c) for the CSGGN distribution with $\mu = 0$, $\sigma = 1$, varying the skewness parameter, ν_1 and ν_2, and fixed $\nu_3 = 4$.



(a) (b)



(c)

Figure 14 – Skewness (a,b) and excess of kurtosis (c) for the CSGGN distribution with $\mu = 0$, $\sigma = 1$, varying the skewness parameter, ν_1 and ν_2, and fixed $\nu_3 = 6$.

1.6 Profiled log-likelihood for γ

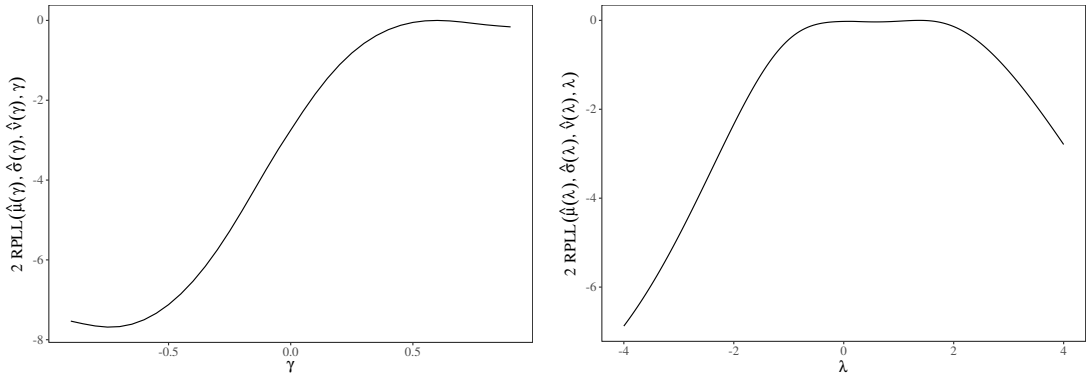
As cited before, the advantages of centered parameterization are in general an improvement in the interpretation and estimation of the parameters compared to the usual parameterization (see Arellano-Vale and Azzalini (2008) and Chaves et al. (2019), for example).

Arellano-Vale and Azzalini (2008) mentioned that under the direct parameterization of skew-normal defined in subsection 1.3, as $\lambda \to 0$ the expected Fisher information matrix becomes singular. In order to illustrate one of the advantages of the centered parameterization we follow Arellano-Vale and Azzalini (2008) and Maioli (2018a), that is, we analyze the behavior of the profile log-likelihood for the skewness parameter.

Let us consider $l_{DP}(\hat{\mu}(\lambda), \hat{\sigma}^2(\lambda), \hat{\boldsymbol{\nu}}(\lambda), \lambda)$ the profile log-likelihood for λ, where $\hat{\mu}(\lambda), \hat{\sigma}^2(\lambda), \hat{\boldsymbol{\nu}}(\lambda)$ are the respective maximum likelihood estimates for a given λ, and the Relative Profile log-likelihood (RPLL) given by $l_{DP}(\hat{\mu}(\lambda), \hat{\sigma}^2(\lambda), \hat{\boldsymbol{\nu}}(\lambda), \lambda) - l_{DP}(\hat{\mu}(\hat{\lambda}), \hat{\sigma}^2(\hat{\lambda}), \hat{\boldsymbol{\nu}}(\hat{\lambda}), \hat{\lambda})$. Similarly, for the centered parameterization the RPLL is given by $l_{DP}(\hat{\mu}(\gamma), \hat{\sigma}^2(\gamma), \hat{\boldsymbol{\nu}}(\gamma), \gamma) - l_{DP}(\hat{\mu}(\hat{\gamma}), \hat{\sigma}^2(\hat{\gamma}), \hat{\boldsymbol{\nu}}(\hat{\gamma}), \hat{\gamma})$

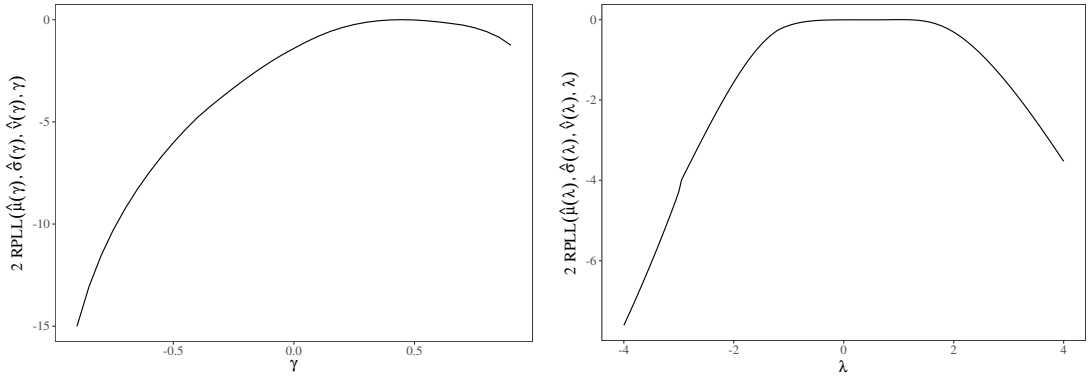
We generated a sample of 100 values from the SMCSN where: $\mu = 0$, $\sigma^2 = 1$, $\gamma = 0.7$ and $\nu = 5$ for the Centered Skew-t, $\nu = 3$ for the Centered Skew-Slash, $\boldsymbol{\nu} = (0.5, 0.5)$ for the Centered Skew-contaminated Normal, $\boldsymbol{\nu} = (6, 5)$ for the Centered skew generalized-t, $\boldsymbol{\nu} = (3, 5)$ for the Centered skew Beta Prime Normal, $\boldsymbol{\nu} = (1, 1)$ for the Centered skew Birnbaum-Saunders Normal and $\boldsymbol{\nu} = (2, 2, 2)$ for Centered skew generalized Gamma Normal.

From Figures 15-21, it is possible to see that under the direct parameterization, the relative profile log-likelihood presents a non-quadratic shape around zero, in addition it presents local maximums, which could lead to a difficult in the obtaining maximum likelihood estimates. Therefore, we can see some of the advantages of the CP under the DP (Maioli, 2018a).



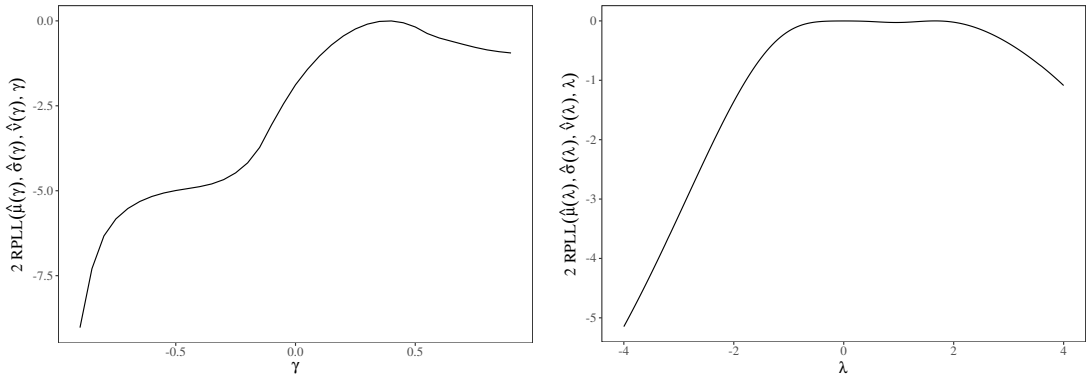
(a) (b)

Figure 15 – Profile twice the relative log-likelihood for γ in the centered parameterization (a) and for λ in the direct parameterization (b) for the skew t distribution



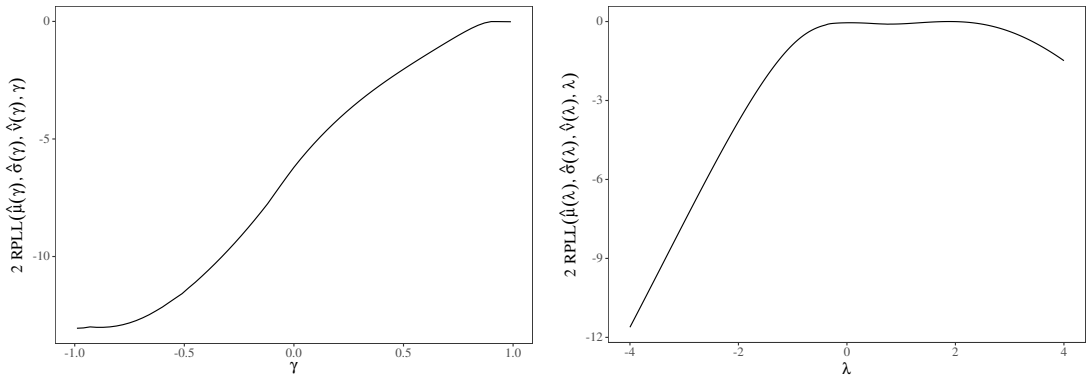
(a) (b)

Figure 16 – Profile twice the relative log-likelihood for γ in the centered parameterization (a) and for λ in the direct parameterization (b) for the skew Slash distribution



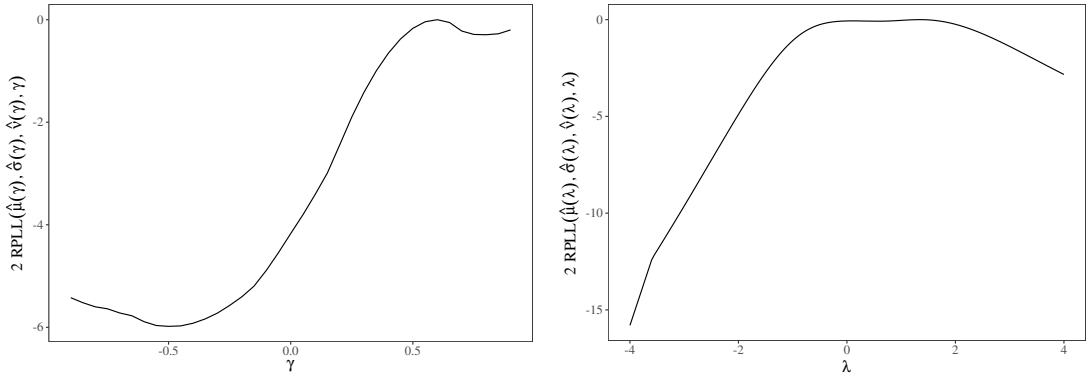
(a) (b)

Figure 17 – Profile twice the relative log-likelihood for γ in the centered parameterization (a) and for λ in the direct parameterization (b) for the skew Contaminated Normal distribution



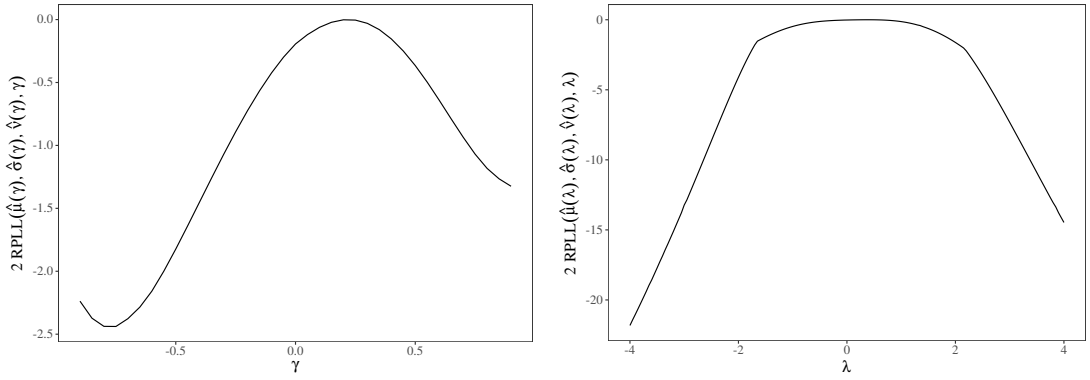
(a) (b)

Figure 18 – Profile twice the relative log-likelihood for γ in the centered parameterization (a) and for λ in the direct parameterization (b) for the skew generalized t distribution



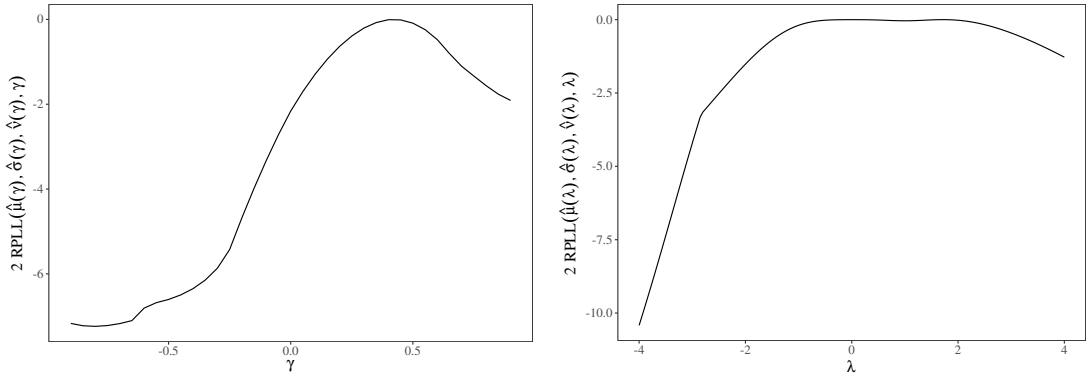
(a) (b)

Figure 19 – Profile twice the relative log-likelihood for γ in the centered parameterization (a) and for λ in the direct parameterization (b) for the skew Beta Prime Normal distribution



(a) (b)

Figure 20 – Profile twice the relative log-likelihood for γ in the centered parameterization (a) and for λ in the direct parameterization (b) for the skew Birnbaum-Saunders distribution



(a) (b)

Figure 21 – Profile twice the relative log-likelihood for γ in the centered parameterization (a) and for λ in the direct parameterization (b) for the skew generalized Gamma Normal distribution

Chapter 2

A new Semi-parametric model based on the scale mixture of Centered skew-normal

Some real problems, where some covariates (explanatory variables) are related to a response of interest, can be suitably modeled by

$$Y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \epsilon_i,\ i = 1, \ldots, n, \tag{2.1}$$

where $\boldsymbol{x}_i = (1, x_{i1}, \ldots, x_{ip})^\top$ is the vector of explanatory variables of the ith subject, Y_i is the ith observation of the response variable, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^\top$ is the vector of regression parameters and $(\epsilon_1, \ldots, \epsilon_n)^\top$ is the vector of the independent random errors with $\mathbb{E}(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$. Maioli (2018a), for example, under the structure of Equation (2.1), proposed a class of Bayesian regression models where the errors follow a SMCSN family.

However in some cases we can observe that one or more covariates does not present a clear relation with the response. An usual approach for that is to include a non-parametric component in Equation (2.1), that is:

$$Y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + g(\boldsymbol{T}_i) + \epsilon_i,\ i = 1, \ldots, n, \tag{2.2}$$

where $\boldsymbol{T}_i = (t_{i1}, \ldots, t_{iq})^\top$ is a vector of explanatory variables related to the ith experimental unit, g is an unknown function $g : \mathbb{R}^q \to \mathbb{R}$. A special case of model (2.2) is the non-parametric regression model given by

$$Y_i = g(\boldsymbol{T}_i) + \epsilon_i,\ i = 1, \ldots, n, \tag{2.3}$$

where each all components are as defined above. An important issue in estimating the related parameters is that g has an unknown shape.

In this work we develop a semi-parametric model of the form (2.2) which has as special case the non-parametric class of models of the form (2.3), where the errors belong the SMCSN family, being an extension of the parametric model proposed by Maioli (2018a). We addressed the frequentist paradigm, using a penalized likelihood approach through a combination of the SAEM (Delyon et al., 1999) and the ECME (Liu and Rubin, 1994) algorithms. Also, we proposed diagnostic measures of global and local influence and goodness-of-fit tools. In addition we performed simulation studies for parameters recovery and misspecification of the distribution and an application to a real data set.

2.1 Basis functions

Any continuous function in a function space (Kolmogorov and Fomin, 1999) can be represented as a linear combination of basis functions. Therefore, we can represent g as a linear combination of $k \leqslant n$ known basis functions $b_1, \ldots, b_K$, i.e.,

$$g(x) \approx g_k(x) = \sum_{i=1}^{k} \kappa_i b_i(x),$$

where $\kappa = \kappa_1, \ldots, \kappa_k$ are unknown coefficients (to be estimated) and k controls the flexibility (shape) of the curve (Dias, 1999). An important issue is that the base (or basis) functions must represent the target function g. Additionally, we need to seek for parsimony, in the sense of having a compromise between the value of k and the choice of b's. Therefore, we need to have a balance between the computational effort and the interpretation of the coefficients $(\kappa_1, \ldots, \kappa_k)$.

There are many proposals for choosing the basis functions, as the polynomials, Fourier series, splines and basis splines (B-splines). For more details we recommend the reading of De Boor (1978) and Kohn et al. (2001), for example. In the following, we briefly describe the splines approach.

2.1.1 Spline function

Let us suppose an interval, says, $[a, b]$ of the domain of g, and the partition into k sub-intervals, say $[\zeta_{j-1}, \zeta_j]$, $1 \leqslant j \leqslant k$, where $a < \zeta_0 < \cdots < \zeta_k < b$. Then a polynomial, say, p_j is considered to approximate the function g in each interval $[\zeta_{j-1}, \zeta_j]$, leading to a polynomial approximation function by parts, say $p_j(x)$ in the interval $[\zeta_{j-1}, \zeta_j]$, with $j = 1, \ldots, k$. The values $\zeta_0, \ldots, \zeta_k$ are known in the literature as knots, where ζ_0 and ζ_k are called external knots and $\zeta_1, \ldots, \zeta_{k-1}$ are the internal knots.

Since $p_j(x), j = 1, \ldots, k$ are constructed independently from each other, so it is not continuous over $[a, b]$ (Hastie and Tibshirani, 1990). This can be problematic,

mainly if we intend to approximate a smooth function, then, we can connect the polynomials on the internal knots $\zeta_1, \ldots, \zeta_{k-1}$, obtaining what we know as a spline function.

A spline of order m = degree + 1 with $k-1$ internal knots is such that $s(x) = \sum_{i=0}^{m-1} c_i x^i + \sum_{j=1}^{k-1} d_j (x-\zeta_j)_+^{m-1}$, where $c_0, \ldots, c_{m-1} \in \mathbb{R}$ and $d_1, \ldots, d_{k-1} \in \mathbb{R}$, and u_+^r is truncated power function of degree r given by:

$$u_+^r(u) = \begin{cases} u^r, \text{ if } u \geqslant 0 \\ 0, \text{ if } u < 0. \end{cases}$$

Notice that the spline function is a combination of $m+k$ basis functions.

2.1.2 B-splines

The set of spline functions of order m and interior knots $\zeta_1, \ldots, \zeta_{k-1}$ is called a spline space. It corresponds to a linear space of dimension $m+k$ (Schumaker, 2007). The so-called B-splines form a basis of splines spaces, i.e., they consist on pieces of polynomials, joined together in certain values, called knots. B-splines having compact support, i.e. they are non-zero within a small interval and zero outside them.

The ith B-spline of order m can be defined recursively as (De Boor, 1978)

$$B_{i,m}(x) = \frac{x-\zeta_i}{\zeta_{i+m-1}-\zeta_i} B_{i,m-1}(x) + \frac{\zeta_{i+m}-x}{\zeta_{i+m}-\zeta_{i+1}} B_{i+1,m-1}(x),$$

where

$$B_{i,1}(x) = \begin{cases} 1, \text{ if } \zeta_i \leqslant x \leqslant \zeta_{i+1} \\ 0, \text{ otherwise} \end{cases}$$

More details on splines and B-splines can be seen at De Boor (1978), Eubank (1988) and Green and Silverman (1994), for example.

2.2 Curve estimation with splines

Let us consider the model described in Equation (2.2) and that the response variable Y_i comes from a family of distributions $\mathcal{F}$, indexed by a set of q parameters $\boldsymbol{\theta} \in \mathbb{R}^q$, which includes the mean, says $\mu = f(\boldsymbol{\beta}, \boldsymbol{\kappa})$ and $\boldsymbol{\kappa}$ is a vector of non-parametric parameters.

2.2.1 Regression analysis with splines

In the regression using splines, the estimated coefficients ($\hat{\boldsymbol{\theta}}$) are obtained by maximizing the log-likelihood associated to $\mathcal{F}$. The degree of smoothing is related to

the number of basis functions (k). An usual choice for the basis functions is the cubic B-splines. The main difficulties in working with this method is the choice of both knot positions and the number of bases.

2.2.2 Smooth splines

In the smoothing splines the estimated coefficients $\hat{\boldsymbol{\theta}}$ are obtained by maximizing

$$\ell(\boldsymbol{\theta}|y) - \frac{\alpha}{2}\int_{\chi}\left[g_k^{(2)}(t)\right]^2 dt, \tag{2.4}$$

where $\ell(.)$ is the log-likelihood associated to the family $\mathcal{F}$, α is the smoothing parameter, χ is a range that covers the support of the variable of interest and the superscript symbol (k) means the kth derivative. The second component in (2.4) that penalizes the log-likelihood represents a penalty in the curvature of the estimated function. Consequently, high values of this penalty represent many oscillations in the curvature of the nonparametric function. With respect to α, large values lead to smooth curves, whereas small values imply curves with more curvature.

According to Green and Silverman (1994) the function $\hat{g}$ that maximizes (2.4) is necessarily a natural cubic spline, see also Craven and Wahba (1978) and Wahba (1981), for example. Notice that in this case the number of coefficients may be as large as the number of observations. Consequently, the computational effort becomes higher.

2.3 Semi-parametric penalized likelihood regression

Let us consider: $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ a vector of response variables, $\boldsymbol{X} = (\boldsymbol{x}_1^\top, \ldots, \boldsymbol{x}_n^\top)$, $\boldsymbol{x_i} = (1, x_{i1}, \ldots, x_{ip})^\top$ the covariates related to the ith subject, $i = 1, \ldots, n$ and $\boldsymbol{T} = (\boldsymbol{T}_1^\top, \ldots, \boldsymbol{T}_q^\top)$, $\boldsymbol{T_j} = (T_{1j}, \ldots, T_{nj})^\top$ the observations related to covariate $\boldsymbol{T}_j$ $j = 1, \ldots, q$. Then, the model (2.2) can be rewritten as:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + g(\boldsymbol{T_1}, \ldots, \boldsymbol{T_q}) + \boldsymbol{\epsilon}, \tag{2.5}$$

where $g(\boldsymbol{T_1}, \ldots, \boldsymbol{T_q}) = [g(T_{11}, \ldots, T_{1q}), \ldots, g(T_{n1}, \ldots, T_{nq})]^\top$, $\boldsymbol{\epsilon}$ is a vector of random errors with $\mathbb{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}_n$ and $\textbf{Cov}(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{I}_n$, where $\boldsymbol{I}_n$ is an identity matrix of order n and $\boldsymbol{0}_n$ is a vector of zeros of order n. We adopt the semi-parametric additive model or additive partially linear model (APLM) as it is known in the literature such that $g(T_{i1}, \ldots, T_{iq}) = \sum_{j=1}^{q} g_j(T_{ij})$, where $g_j(T_{ij}) = \sum_{l=1}^{k_j} \kappa_{lj} b_l(x_{ij})$, b_l is a known cubic B-splines basis, $l = 1, \ldots, k_j$, $j = 1, \ldots, q$. The estimation by additive models provides a more simple approximation for the function g, mainly due to the their simpler interpretability, for example when considering interaction effects between the predictors. That is, once

the model was fitted, one can examine the effect of each predictor separately (e.g. $g(T_1, T_2)$ becomes $g(T_1) + g(T_2)$). Therefore, the model (2.5) can be written as: $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \sum_{j=1}^{q} \boldsymbol{T}_j\boldsymbol{\kappa}_j + \boldsymbol{\epsilon}$, where $\boldsymbol{T}_j$ is a matrix with k_j basis functions and $\boldsymbol{\kappa}_j = (\kappa_{j1}, \ldots, \kappa_{jk_j})^\top$, $j = 1, \ldots, q$. In the usual penalized likelihood regression approach for the APLM, the estimates of $\boldsymbol{\beta}$ and $g_1, \ldots, g_q$ are obtained through the maximization of:

$$\ell(\boldsymbol{\theta}|\boldsymbol{y}) - \frac{1}{2}\sum_{j=1}^{q} \alpha_j \int_{\chi} \left[g_j^{(2)}(w)\right]^2 dw = \ell(\boldsymbol{\theta}|\boldsymbol{y}) - \frac{1}{2}\sum_{j=1}^{q} \alpha_j \boldsymbol{\kappa}_j^\top \boldsymbol{\Omega}_j \boldsymbol{\kappa}_j, \tag{2.6}$$

where $\boldsymbol{\theta}$ is related to $\mathcal{F}$, associated to the distribution function of $\boldsymbol{\epsilon}$, and $\boldsymbol{\Omega}_j$ is a matrix $(k_j \times k_j)$ based on knots. The equation (2.6) can also be simplified to $\ell(\boldsymbol{\theta}|\boldsymbol{y}) - \frac{1}{2}\boldsymbol{\kappa}^\top \boldsymbol{\Lambda}\boldsymbol{\kappa}$, where $\boldsymbol{\Lambda} = \text{diag}(\alpha_1\boldsymbol{\Omega}_1, \ldots, \alpha_q\boldsymbol{\Omega}_q)$ is a block diagonal matrix. The proof that $\int_{\chi} \left[g_j^{(2)}(w)\right]^2 dw$ can be rewritten as $\boldsymbol{\kappa}_j^\top \boldsymbol{\Omega}_j \boldsymbol{\kappa}_j$ can be seen in Appendix B.

This model can be unidentifiable, unless $\boldsymbol{g}_j^* = \boldsymbol{T}_j\boldsymbol{\kappa}_j, j = 1, ..., k$ be subject to some constraint. Vanegas and Paula (2016) and Wood (2017) proposed to use $\mathbf{1}_n^\top \boldsymbol{g}_j^* = 0$. The technical details related the using of such constrain in the estimation process can be seen in Appendix C.

2.4 SMCSN additive partial linear model

The SMCSN additive partial linear model is given by: $Y_i = \boldsymbol{x}_i\boldsymbol{\beta} + \sum_{j=1}^{q} g_j(T_{ij}) + \epsilon_i$, with all related elements defined above, assuming that $\epsilon_i \overset{ind}{\sim} SMCSN(0, u_i^{-1}\sigma^2, \gamma, \boldsymbol{\nu})$. Similarly to Maioli (2018a), we have that:

$$Y_i|U_i = u_i \overset{ind}{\sim} CSN\left(\boldsymbol{x}_i^\top\boldsymbol{\beta} + \sum_{j=1}^{q} g_j(T_{ij}), \frac{\sigma^2}{u_i}, \gamma\right),\ U_i \overset{iid}{\sim} H(.|\boldsymbol{\nu}),$$

which leads to

$$Y_i|(U_i = u_i) \overset{D}{=} \boldsymbol{x}_i^\top\boldsymbol{\beta} + \sum_{j=1}^{q} g_j(T_{ij}) + \frac{\sigma}{\sqrt{u_i}}\left(\frac{V_i - \mu_v}{\sigma_v}\right),$$

where $V_i \overset{iid}{\sim} SN(0, 1, \gamma)$ and μ_v and σ_v are the mean and the variance of V_i respectively. Then, using Henze's stochastic representation of V_i in (1.3), we have that

$$Y_i|U_i = u_i \overset{D}{=} \boldsymbol{x}_i^\top\boldsymbol{\beta} + \sum_{j=1}^{q} g_j(T_{ij}) - \frac{\sigma}{\sqrt{u_i}}\frac{\mu_v}{\sigma_v} + \frac{\sigma}{\sigma_v\sqrt{u_i}}(\delta H_i + \sqrt{1-\delta^2}Z_i),$$

where $H_i \overset{iid}{\sim} HN(0,1) \perp Z_i \overset{iid}{\sim} N(0,1), i = 1, 2, ..., n$. Setting $\mu_v = \delta b$ and $\sigma_v = \sqrt{1 - b^2\delta^2}$, it comes that

$$Y_i|U_i = u_i \overset{D}{=} \boldsymbol{x}_i^\top\boldsymbol{\beta} + \sum_{j=1}^{q} g_j(T_{ij}) - \frac{\sigma\delta b}{\sqrt{u_i}\sqrt{1-b^2\delta^2}} + \frac{\sigma\delta}{\sqrt{u_i}\sqrt{1-b^2\delta^2}}H_i + \frac{\sigma\sqrt{1-\delta^2}}{\sqrt{u_i}\sqrt{1-b^2\delta^2}}Z_i$$

$$= \boldsymbol{x}_i^\top \boldsymbol{\beta} + \sum_{j=1}^{q} g_j(T_{ij}) + \frac{\sigma\delta}{\sqrt{u_i}\sqrt{1-b^2\delta^2}}(H_i - b) + \frac{\sigma\sqrt{1-\delta^2}}{\sqrt{u_i}\sqrt{1-b^2\delta^2}} Z_i.$$

Maioli (2018a) considers $\Delta = \sigma\delta/\sqrt{1-b^2\delta^2}$ and $\tau = \sigma^2(1-\delta^2)/(1-b^2\delta^2)$, such that we can recover σ and δ through: $\delta = \Delta/\sqrt{\tau+\Delta^2}$ and $\sigma^2 = \tau + \Delta^2(1-b^2)$. Then

$$Y_i|U_i = u_i \overset{D}{=} \boldsymbol{x}_i^\top \boldsymbol{\beta} + \sum_{j=1}^{q} g_j(T_{ij}) + \frac{\Delta}{\sqrt{u_i}}(H_i - b) + \frac{\sqrt{\tau}}{\sqrt{u_i}} Z_i.$$

Finally, we have that

$$Y_i|U_i = u_i, H_i = h_i \overset{ind}{\sim} \mathcal{N}\left(\boldsymbol{x}_i^\top \boldsymbol{\beta} + \sum_{j=1}^{q} g_j(T_{ij}) + \frac{\Delta}{\sqrt{u_i}}(h_i - b), \frac{\tau}{u_i}\right), \tag{2.7}$$

$$H_i \overset{iid}{\sim} HN(0,1), \tag{2.8}$$

$$U_i \overset{iid}{\sim} h(.|\boldsymbol{\nu}). \tag{2.9}$$

In next Section we show the estimation process for $\boldsymbol{\theta}$ via penalized maximum log-likelihood using the hierarchical structure defined in Equation (2.7).

2.5 Maximum penalized log-likelihood estimation

The complete likelihood is given by:

$$\begin{aligned} L_c(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{u}.\boldsymbol{h}) &\propto \prod_{i=1}^{n} \phi\left(y_i|\mu_i, \frac{\tau}{u_i}\right) f(h_i) h(u_i|\boldsymbol{\nu}) \\ &\propto \prod_{i=1}^{n} \frac{\sqrt{u_i}}{\sqrt{\tau}} \exp\left\{-\frac{u_i}{2\tau}(y_i - \mu_i)^2\right\} \exp\left\{-\frac{h_i^2}{2}\right\} h(u_i|\boldsymbol{\nu}) \\ &\propto \frac{\prod_{i=1}^{n}\sqrt{u_i}}{\tau^{n/2}} \exp\left\{-\frac{1}{2\tau}\sum_{i=1}^{n} u_i(y_i - \mu_i)^2\right\} \exp\left\{-\frac{\sum_{i=1}^{n} h_i^2}{2}\right\} \prod_{i=1}^{n} h(u_i|\boldsymbol{\nu}), \end{aligned}$$

where $\mu_i = \boldsymbol{x}_i^\top\boldsymbol{\beta} + \sum_{j=1}^{q} g_j(T_{ij}) + \Delta(h_i - b)/\sqrt{u_i}$, remembering that $\sum_{j=1}^{q} g_j(T_{ij}) = \sum_{j=1}^{q}\sum_{l=1}^{k_j} \kappa_{lj} b_l(t_{ij})$ $= \sum_{j=1}^{n} \boldsymbol{\kappa}_j^\top \boldsymbol{b}(T_{ij})$, $\mu_i = \boldsymbol{x}_i^\top\boldsymbol{\beta} + \sum_{j=1}^{n} \boldsymbol{\kappa}_j^\top \boldsymbol{b}(T_{ij}) + \Delta(h_i - b)/\sqrt{u_i}$, $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\kappa}^\top, \Delta, \tau, \boldsymbol{\nu}^\top)^\top$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, $\boldsymbol{\kappa} = (\boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_q)^\top$, $\boldsymbol{\kappa}_j = (\kappa_{1j}, \ldots, \kappa_{k_j j})$ and $\boldsymbol{k} = (k_1, \ldots, k_q)^\top$.

For the skew generalized T model, since we set $\sigma^2 = 1$, Δ and τ depend only on $\boldsymbol{\gamma}$, in such way that is better to sample directly from δ (Maioli, 2018a). Therefore, for this model we have that $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\kappa}^\top, \delta, \boldsymbol{\nu}^\top)^\top$.

Let $\boldsymbol{y}_{obs} = (y_1, \ldots, y_n)^\top$ be an observed sample from $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top \overset{ind}{\sim} SMCSN$, and the latent vectors $\boldsymbol{u} = (u_1, \ldots, u_n)^\top$ and $\boldsymbol{h} = (h_1, \ldots, h_n)^\top$. Then the

complete penalized log-likelihood of the parameters $\boldsymbol{\theta}$ is given by:

$$l_P(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{u},\boldsymbol{h}) = \log\left(\frac{\prod_{i=1}^n \sqrt{u_i}}{\tau^{n/2}}\exp\left\{-\frac{1}{2\tau}\sum_{i=1}^n u_i(y_i-\mu_i)^2\right\}\exp\left\{-\frac{\sum_{i=1}^n h_i^2}{2}\right\}\prod_{i=1}^n h(u_i|\boldsymbol{\nu})\right) \tag{2.10}$$

$$-\frac{1}{2}\boldsymbol{\kappa}^\top\boldsymbol{\Lambda}\boldsymbol{\kappa} \tag{2.11}$$

$$= \frac{1}{2}\sum_{i=1}^n \log u_i - \frac{n}{2}\log\tau - \frac{1}{2\tau}\sum_{i=1}^n u_i(y_i-\mu_i)^2 - \frac{1}{2}\sum_{i=1}^n h_i^2 + \sum_{i=1}^n \log\left(h(u_i|\boldsymbol{\nu})\right) \tag{2.12}$$

$$-\frac{1}{2}\boldsymbol{\kappa}^\top\boldsymbol{\Lambda}\boldsymbol{\kappa}, \tag{2.13}$$

where $\mu_i = \boldsymbol{x}_i^\top\boldsymbol{\beta} + \sum_{j=1}^q g_j(T_{ij}) + \Delta(h_i - b)/\sqrt{u_i}$. The maximum penalized likelihood estimators are obtained by maximizing (2.10). However, no analytical expression are obtained from this process. The use of the Expectation-Maximization algorithm (EM algorithm), as pointed out by Ferreira et al. (2013), leads to an intractable E-step, in this case, the Expectation-Conditional-Maximization-Either algorithm (ECME algorithm) (Liu and Rubin, 1994) proposes to maximize the expected complete data function (Q-function) with CML-steps that maximize the corresponding constrained actual marginal likelihood function, which may be more treatable in our case.

However, in our case the E-step can not be calculated analytically and some approximation need to be considered (either analytical or numerical). Two usual numerical approaches to handle that are the Monte Carlo EM (MCEM), proposed by Wei et al. (1998) and the Stochastic Approximation of the EM (SAEM), proposed by Delyon et al. (1999). The first method can be computationally expensive, since a large number of simulations of the missing data is required. On the other hand, the SAEM algorithm replaces the E-step by a stochastic approximation, while the Maximization step remains the same as that of the MCEM. This algorithm has good convergence properties, see for example Kuhn and Lavielle (2005) and Allassonnière et al. (2010). In this work we consider a combination of the SAEM and the ECME, where the E- Step is done through the stochastic approximation of the SAEM algorithm, along with the CML structure of the ECME algorithm to perform the M-step. This leads to an algorithm that we name Stochastic Approximation of Expectation-Conditional-Maximization-Either (SAECME) algorithm. In the following we explain our approach.

Let us consider $\boldsymbol{u} = (u_1,\ldots,u_n)^\top$ and $\boldsymbol{t} = (t_1,\ldots,t_n)^\top$ the missing data and $\boldsymbol{y}_c = (\boldsymbol{y}^\top,\boldsymbol{u}^\top,\boldsymbol{t}^\top)^\top$. In the E-step of the ECME algorithm we must obtain the Q-function given by: $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) = \mathbb{E}\left[l_P(\boldsymbol{\theta}|\boldsymbol{y}_c)|\boldsymbol{y}_{obs},\hat{\boldsymbol{\theta}}^{(k)}\right]$, where the superscript (k) indicates the kth iteration of the algorithm. Thus, dropping out the constants, the Q-function can be written

as:

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) &= -\frac{n}{2}\log\hat{\tau}^{(k)} - \frac{1}{2\hat{\tau}^{(k)}}\sum_{i=1}^{n}\mathbb{E}\left(U_i(y_i-\hat{\mu}_i^{(k)})^2|\hat{\boldsymbol{\theta}}^{(k)}\right) - \frac{1}{2}\hat{\boldsymbol{\kappa}}^{(k)\top}\boldsymbol{\Lambda}\hat{\boldsymbol{\kappa}}^{(k)} \\
&= -\frac{n}{2}\log\hat{\tau}^{(k)} - \frac{1}{2\hat{\tau}^{(k)}}\sum_{i=1}^{n}\mathbb{E}\left(U_i\left[\left(y_i-\hat{\mu}_{ci}^{(k)}\right) - \frac{\Delta}{U_i^{1/2}}(H_i-b)\right]^2|\hat{\boldsymbol{\theta}}^{(k)}\right) \\
&- \frac{1}{2}\hat{\boldsymbol{\kappa}}^{(k)\top}\boldsymbol{\Lambda}\hat{\boldsymbol{\kappa}}^{(k)} \\
&= -\frac{n}{2}\log\hat{\tau}^{(k)} - \frac{1}{2\hat{\tau}^{(k)}}\sum_{i=1}^{n}\mathbb{E}\Big(U_i\left(y_i-\hat{\mu}_{ci}^{(k)}\right)^2 - 2\left(y_i-\hat{\mu}_{ci}^{(k)}\right)\Delta U_i^{1/2}(H_i-b) \\
&+ \Delta^2(H_i-b)^2|\hat{\boldsymbol{\theta}}^{(k)}\Big) - \frac{1}{2}\hat{\boldsymbol{\kappa}}^{(k)\top}\boldsymbol{\Lambda}\hat{\boldsymbol{\kappa}}^{(k)},
\end{aligned}
$$

where $\mu_{ci} = \boldsymbol{x}_i^\top\boldsymbol{\beta} + \sum_{j=1}^{q} g_j(T_{ij})$ and since that $H_i - b = F_i$, where H denote the Half Normal distribution, we have that

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}) &= -\frac{n}{2}\log\hat{\tau}^{(k)} - \frac{1}{2\hat{\tau}^{(k)}}\sum_{i=1}^{n}\Big[\left(y_i-\hat{\mu}_{ci}^{(k)}\right)^2\mathbb{E}(U_i|\hat{\boldsymbol{\theta}}^{(k)}) - 2\left(y_i-\hat{\mu}_{ci}^{(k)}\right)\Delta\times \\
&\times\mathbb{E}\left(U_i^{1/2}F_i|\hat{\boldsymbol{\theta}}^{(k)}\right) + \Delta^2\mathbb{E}\left(F_i^2|\hat{\boldsymbol{\theta}}^{(k)}\right)\Big] - \frac{1}{2}\hat{\boldsymbol{\kappa}}^{(k)\top}\boldsymbol{\Lambda}\hat{\boldsymbol{\kappa}}^{(k)} \\
&= -\frac{n}{2}\log\hat{\tau}^{(k)} - \frac{1}{2\hat{\tau}^{(k)}}\sum_{i=1}^{n}\Big[\left(y_i-\hat{\mu}_{ci}^{(k)}\right)^2\mathcal{E}_i^{(1;0)} - 2\left(y_i-\hat{\mu}_{ci}^{(k)}\right)\Delta\mathcal{E}_i^{(0.5;1)} + \\
&\Delta^2\mathcal{E}_i^{(0;2)}\Big] - \frac{1}{2}\hat{\boldsymbol{\kappa}}^{(k)\top}\boldsymbol{\Lambda}\hat{\boldsymbol{\kappa}}^{(k)} \\
&= -\frac{n}{2}\log\hat{\tau}^{(k)} - \frac{1}{2\hat{\tau}^{(k)}}\Big[(\boldsymbol{y}-\hat{\boldsymbol{\mu}}_c^{(k)})^\top\boldsymbol{\mathcal{E}}^{(1;0)}(\boldsymbol{y}-\hat{\boldsymbol{\mu}}_c^{(k)}) - 2\Delta(\boldsymbol{y}-\hat{\boldsymbol{\mu}}_c^{(k)})^\top\boldsymbol{\mathcal{E}}^{(0.5;1)}\mathbb{1}_n \\
&+ \Delta^2\mathbb{1}_n^\top\boldsymbol{\mathcal{E}}^{(0;2)}\mathbb{1}_n\Big] - \frac{1}{2}\hat{\boldsymbol{\kappa}}^{(k)\top}\boldsymbol{\Lambda}\hat{\boldsymbol{\kappa}}^{(k)},
\end{aligned}
$$

where $\mathcal{E}_i^{(s;l)} = \mathbb{E}\left(U_i^s F_i^l|\boldsymbol{y}_{obs},\hat{\boldsymbol{\theta}}^{(k)}\right)$ and $\boldsymbol{\mathcal{E}}^{(s;l)} = \text{diag}\left(\mathcal{E}_1^{(s;l)},\ldots,\mathcal{E}_n^{(s;l)}\right)$.

Notice that $\mathbb{E}\left(\log U_i|\boldsymbol{y}_{obs},\hat{\boldsymbol{\theta}}^{(k)}\right)$ and $\mathbb{E}\left(\log h(U_i|\boldsymbol{\nu})|\boldsymbol{y}_{obs},\hat{\boldsymbol{\theta}}^{(k)}\right)$ depend only on $\boldsymbol{\nu}$, which is assumed to be fixed at this point and we will estimate them after. Also, in this work we focus on the estimation of δ instead γ due to its ease of working with this transformation of the skewness parameter, mainly in obtaining the estimators. In order to calculate the necessary/required expectations we need to draw samples from $f(u_i,h_i|y_i)$ through the conditional method (Ripley, 2009), that consists on drawing samples from $f(u_i|y_i)$ and $f(h_i|y_i,u_i)$, sequentially. Indeed, to sample from $U_i|Y_i=y_i$ we consider the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970) as explained in Algorithm 1.

The function $g(u_i)$ in Algorithm 1 has to be a probability density function with the same support as $f(u_i|y_i)$. Notice that, since (from the Bayes theorem):$f(u_i|y_i) =$

Algorithm 1 – Algorithm for simulate $U_i|Y_i$

Given $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(k)}$, for $i = 1, \ldots, n$:

1. Starting with an initial value u_i^0 and set $m_1 = 0$;
2. Draw $u_i^* \sim g(u_i^*)$ from a proper density;
3. Generate $C \sim U(0,1)$;
4. If $C \leqslant \min\left\{1, \dfrac{f(u_i^*|\hat{\boldsymbol{\theta}}^{(k)}, y_i)h(u_i^{m_1})}{h(u_i^*)f(u_i^{m_1}|\hat{\boldsymbol{\theta}}^{(k)}, y_i)}\right\}$, $u_i^{(m_1+1)} = u_i^*$, else, $u_i^{(m_1+1)} = u_i^{(m_1)}$.
5. Return to step 2 until $m_1 = m$.

$f(y_i|u_i)h(u_i|\boldsymbol{\nu})/f(y_i) \propto f(y_i|u_i)h(u_i|\boldsymbol{\nu})$, the objective function is a product of two others (both with positive support). Therefore, a suitable choice for the proposal density can be the density of U_i itself, which leads to a simpler expression in the Metropolis-Hasting algorithm. On the other hand, for $H_i|y_i, u_i$, considering Equation (2.7) we have that

$$
\begin{aligned}
L_c(\boldsymbol{\theta}|y_i, u_i, h_i) &\propto 2\phi\left(y_i|\mu_{ci} + \frac{\Delta}{\sqrt{u_i}}h_i - \frac{\Delta}{\sqrt{u_i}}b, \frac{\tau}{u_i}\right)\phi\left(h_i|0,1\right)\mathbb{1}(h_i)_{(0,\infty)} \\
&\propto \exp\left\{-\frac{u_i}{2\tau}\left[y_i - \mu_{ci} + \frac{\Delta}{\sqrt{u_i}}b - \frac{\Delta}{\sqrt{u_i}}h_i\right]^2\right\}\exp\left\{-\frac{1}{2}h_i^2\right\}\mathbb{1}(h_i)_{(0,\infty)} \\
&\propto \exp\left\{-\frac{u_i}{2\tau}\left[\left(y_i - \mu_{ci} + \frac{\Delta}{\sqrt{u_i}}b\right)^2 - 2\left(y_i - \mu_{ci} + \frac{\Delta}{\sqrt{u_i}}b\right)\frac{\Delta}{\sqrt{u_i}}h_i + \frac{\Delta^2}{u_i}h_i^2\right]\right\} \times \\
&\exp\left\{-\frac{1}{2}h_i^2\right\}\mathbb{1}(h_i)_{(0,\infty)} \\
&\propto \exp\left\{-\frac{u_i}{2\tau}\left(y_i - \mu_{ci} + \frac{\Delta}{\sqrt{u_i}}b\right)^2\right\} \times \\
&\exp\left\{-\frac{1}{2}h_i^2 + \left(y_i - \mu_{ci} + \frac{\Delta}{\sqrt{u_i}}b\right)\frac{\sqrt{u_i}\Delta}{\tau}h_i - \frac{\Delta^2}{2\tau}h_i^2\right\}\mathbb{1}(h_i)_{(0,\infty)} \\
&\propto \exp\left\{-\frac{u_i}{2\tau}\left(y_i - \mu_{ci} + \frac{\Delta}{\sqrt{u_i}}b\right)^2\right\} \times \\
&\exp\left\{-\frac{1}{2}h_i^2\left(1 + \frac{\Delta^2}{\tau}\right) + \left(y_i - \mu_{ci} + \frac{\Delta}{\sqrt{u_i}}b\right)\frac{\sqrt{u_i}\Delta}{\tau}h_i\right\}\mathbb{1}(h_i)_{(0,\infty)} \\
&\propto \phi\left(y_i|\mu_c - \frac{\Delta}{\sqrt{u_i}}b, \frac{\tau}{u_i}\right)\phi\left(h_i|\frac{\left(y_i - \mu_{ci} + \frac{\Delta}{\sqrt{u_i}}b\right)}{\left(1 + \frac{\Delta^2}{\tau}\right)}\frac{\sqrt{u_i}\Delta}{\tau}, \frac{1}{\left(1 + \frac{\Delta^2}{\tau}\right)}\right)\mathbb{1}(h_i)_{(0,\infty)} \\
&\propto \phi\left(y_i|\mu_c - \frac{\Delta}{\sqrt{u_i}}b, \frac{\tau}{u_i}\right)\phi\left(h_i|\frac{\Delta^2 b + \Delta\sqrt{u_i}(y_i - \mu_{ci})}{\Delta^2 + \tau}, \frac{\tau}{(\tau + \Delta^2)}\right)\mathbb{1}(h_i)_{(0,\infty)}.
\end{aligned}
$$

Then

$$H_i|u_i, y_i \overset{ind}{\sim} \mathcal{N}\left(\frac{\Delta^2 b + \Delta\sqrt{u_i}(y_i - \mu_{ci})}{\Delta^2 + \tau}, \frac{\tau}{(\tau + \Delta^2)}\right) \mathbb{1}(h_i)_{(0,\infty)}, \tag{2.14}$$

which can be simulated using the `rtruncnorm` function from the `R` program (R Core Team, 2020). Therefore, we can simulate from $U_i, H_i|y_i$ according to Algorithm 2 and, then, estimate to fit the model through the Algorithm 3.

Algorithm 2 – Algorithm for simulate $U_i, H_i|Y_i$

Given $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(k)}$, for $i = 1, \ldots, n$:

1. Draw u_i^* from Algorithm 1;
2. Draw h_i^* from the distribution in 2.14, considering $u_i = u_i^*$, obtained a set of simulated values, the vector (u_i^*, h_i^*) is a simulated vector of $u_i, h_i|y_i$, $i = 1, 2, \ldots, n$.

We can obtain $\delta^{(k+1)}$ and $\hat{\sigma}^{2(k+1)}$ using: $\delta^{(k+1)} = \hat{\Delta}^{(k+1)}/\sqrt{\hat{\tau}^{(k+1)} + \hat{\Delta}^{2(k+1)}}$ and $\hat{\sigma}^{2(k+1)} = \hat{\tau}^{(k+1)} + \hat{\Delta}^{2(k+1)}(1 - b^2)$.

In Algorithm 3, δ_k^* is a smoothness parameter (Kuhn and Lavielle, 2004) which is a decreasing sequence of positive numbers such that $\sum_{k=1}^{\infty} \delta_k^* = \infty$ and $\sum_{k=1}^{\infty} \delta_k^{*^2} < \infty$. The E-step in the SAECME and MCEM algorithms are essentially the same. However, in the former a significant smaller number of simulations is necessary/required in the stochastic simulation, which is suggested to be smaller than 20. This is possible since SAEM-based algorithms use all the previous steps, weighted by the δ_k^*.

It is noteworthy that if $\delta_k^* = 1, \forall k$, the SAECME will have no memory, i.e., it will be equivalent to a combination of the MCEM and the ECME algorithms and it will converge (in distribution) to a solution in a neighborhood of the maximum likelihood estimator. On the other hand, a SAECME with memory ($0 < \delta_k^* < 1$) will converge to the ML solution. Galarza et al. (2017) suggest to use:

$$\delta_k^* = \begin{cases} 1, & \text{if } \ 1 \leqslant k \leqslant cw \\ \dfrac{1}{k - cw}, & \text{if } \ cw + 1 \leqslant k \leqslant w, \end{cases}$$

where w is the maximum number of iterations and $0 \leqslant c \leqslant 1$ is a constant that determines the percentage of initial iterations with no memory. As pointed by Galarza et al. (2017), if $c = 0$ the algorithm will have memory for all iterations, and hence it will converge to the ML estimates. If $c = 1$, the algorithm will have no memory, and then, it will converge to a solution in a neighborhood of the maximum likelihood estimator. For the first case ($c = 0$), we need a large w in order to obtain the ML estimates. For the second case ($c = 1$), the

algorithm will output a chain where after applying a burn in and thin, and apply a mean on the observations can be a reasonable estimate, for example. The choice of suitable values for w and c can be made through a graphical approach. That is, we can monitor the behavior of the consecutive estimates and/or the log-likelihood $(l(\boldsymbol{\theta}|\boldsymbol{y}_{\text{obs}}))$, using, as a convergence criterion: $||l(\boldsymbol{\theta}^{(k+1)}|\boldsymbol{y}_{\text{obs}}) - l(\boldsymbol{\theta}^{(k)}|\boldsymbol{y}_{\text{obs}})||$ or $||l(\boldsymbol{\theta}^{(k+1)}|\boldsymbol{y}_{\text{obs}})/l(\boldsymbol{\theta}^{(k)}|\boldsymbol{y}_{\text{obs}}) - 1||$, for example.

Algorithm 3 – SAECME algorithm

E-step: Given $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(k)}$, for $i = 1, \ldots, n$ do:

- **Simulation step**: To draw $(u_i^{(l,k)}, h_i^{(l,k)})$ from $f(U, H_i | Y_i = y_i)$ using the Algorithm 2, $l = 1, \ldots, m$ and given m (we discuss such choice later).
- **Stochastic approximation**: To compute the stochastic approximations for the conditional expectations necessaries to estimate $\boldsymbol{\theta}$ given by:

$$\begin{aligned}
\mathcal{E}_i^{(1;0)(k)} &= \mathcal{E}_i^{(1;0)(k-1)} + \delta_k^* \left[\frac{1}{m} \sum_{l=1}^{m} u_i^{(l,k)} - \mathcal{E}_i^{(1;0)(k-1)} \right] \\
\mathcal{E}_i^{(0.5;1)(k)} &= \mathcal{E}_i^{(0.5;1)(k-1)} + \delta_k^* \left[\frac{1}{m} \sum_{l=1}^{m} \sqrt{u_i^{(l,k)}} (h_i^{(l,k)} - b) - \mathcal{E}_i^{(0.5;1)(k-1)} \right] \\
\mathcal{E}_i^{(0;2)(k)} &= \mathcal{E}_i^{(0;2)(k-1)} + \delta_k^* \left[\frac{1}{m} \sum_{l=1}^{m} (h_i^{(l,k)} - b)^2 - \mathcal{E}_i^{(0;2)(k-1)} \right];
\end{aligned}$$

CM-step

Update $\hat{\boldsymbol{\theta}}^{(k)}$ by maximizing $Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(k)})$ over $\boldsymbol{\theta}$, which leads to obtain the following expressions:

$$\begin{aligned}
\hat{\tau}^{(k+1)} &= \frac{1}{n} \sum_{i=1}^{n} \left[\left(y_i - \hat{\mu}_{ci}^{(k)} \right)^2 \mathcal{E}_i^{(1;0)(k)} - 2 \left(y_i - \hat{\mu}_{ci}^{(k)} \right) \Delta \mathcal{E}_i^{(0.5;1)(k)} + \Delta^2 \mathcal{E}_i^{(0;2)(k)} \right] \\
\hat{\Delta}^{(k+1)} &= \left(\sum_{i=1}^{n} \mathcal{E}_i^{(0;2)(k)} \right)^{-1} \sum_{i=1}^{n} (y_i - \hat{\mu}_{ci}^{(k)}) \mathcal{E}_i^{(0.5;1)(k)} \\
\hat{\boldsymbol{\beta}}^{(k+1)} &= \left(\sum_{i=1}^{n} \boldsymbol{x}_i^\top \mathcal{E}_i^{(1;0)(k)} \boldsymbol{x}_i \right)^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i \left[y_i \mathcal{E}_i^{(1;0)(k)} - \sum_{j=1}^{q} \hat{g}_j(T_{ij}) \mathcal{E}_i^{(1;0)(k)} - \hat{\Delta}^{(k)} \mathcal{E}_i^{(0.5;1)(k)} \right] \\
&= \left(\boldsymbol{X}^\top \boldsymbol{\mathcal{E}}^{(1;0)(k)} \boldsymbol{X} \right)^{-1} \boldsymbol{X}^\top \left(\boldsymbol{\mathcal{E}}^{(1;0)(k)} \boldsymbol{y} - \boldsymbol{\mathcal{E}}^{(1;0)(k)} \boldsymbol{T} \hat{\boldsymbol{\kappa}}^{(k)} - \boldsymbol{\mathcal{E}}^{(0.5;1)(k)} \hat{\boldsymbol{\Delta}}^{(k)} \right) \\
\hat{\boldsymbol{\kappa}}_m^{(k+1)} &= \left(\sum_{i=1}^{n} \boldsymbol{T}_i^\top \mathcal{E}_i^{(1;0)(k)} \boldsymbol{T}_i + \hat{\tau}^{(k)} \boldsymbol{\Lambda}_m \right)^{-1} \sum_{i=1}^{n} \boldsymbol{T}_i \Bigg[y_i \mathcal{E}_i^{(1;0)(k)} - \boldsymbol{x}_i^\top \boldsymbol{\beta} \mathcal{E}_i^{(1;0)(k)} \\
&\quad - \sum_{j \neq m}^{q} \hat{g}_j(T_{ij}) \mathcal{E}_i^{(1;0)(k)} - \hat{\Delta}^{(k)} \mathcal{E}_i^{(0.5;1)(k)} \Bigg];, \text{ for } m = 1, \ldots, q.
\end{aligned}$$

CML-step

Update $\hat{\boldsymbol{\nu}}^{(k+1)}$ by maximizing the log-likelihood function using, for example, `optim` or `nlminb` R functions, obtaining

$$\hat{\boldsymbol{\nu}}^{(k+1)} = \arg\max_{\boldsymbol{\nu}} \left\{ \sum_{i=1}^{n} \log \left[f_{\text{SMCSN}}(y_i | \hat{\mu}_i^{(k+1)}, \hat{\tau}^{(k+1)}, \hat{\Delta}^{(k+1)}, \boldsymbol{\nu}) \right] \right\}. \tag{2.15}$$

Until some convergence criterion has been met, for example, $||\hat{\boldsymbol{\theta}}^{(k-1)} - \hat{\boldsymbol{\theta}}^{(k)}|| < c$, where $c > 0$.

2.5.1 Effective degrees of freedom

Within the context of additive partially linear models, the degrees of freedom are approximately the number of parameters related to the non-parametric components (Hastie and Tibshirani, 1990; Ibacache-Pulgar et al., 2013). The effective degrees of freedom, based on the estimator of $\boldsymbol{\kappa}$, are given by:

$$\begin{aligned} df(\alpha) &= \text{tr}\left\{\boldsymbol{T}\boldsymbol{S}_f\right\} = \text{tr}\left\{\boldsymbol{T}\left(\boldsymbol{T}^\top\boldsymbol{\mathcal{E}}^{(1;0)}\boldsymbol{T}+\hat{\tau}\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{T}^\top\right\} = \text{tr}\left\{\boldsymbol{T}^\top\left(\boldsymbol{T}^\top\boldsymbol{\mathcal{E}}^{(1;0)}\boldsymbol{T}+\hat{\tau}\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{T}\right\} \\ &= \text{tr}\left\{\left[\boldsymbol{T}^{\top^{-1}}\left(\boldsymbol{T}^\top\boldsymbol{\mathcal{E}}^{(1;0)}\boldsymbol{T}+\hat{\tau}\boldsymbol{\Lambda}\right)\boldsymbol{T}^{-1}\right]^{-1}\right\} = \text{tr}\left\{\left(\boldsymbol{\mathcal{E}}^{(1;0)}+\hat{\tau}\boldsymbol{T}^{\top^{-1}}\boldsymbol{\Lambda}\boldsymbol{T}^{-1}\right)^{-1}\right\} \\ &= \sum_{j=1}^{n}\frac{1}{L_j(\alpha)}, \end{aligned}$$

where $L_j(\alpha)$ is the jth eigenvalue of $\left(\boldsymbol{\mathcal{E}}^{(1;0)}+\hat{\tau}\boldsymbol{T}^{\top^{-1}}\boldsymbol{\Lambda}\boldsymbol{T}^{-1}\right)$.

2.5.2 Information Criteria

The Akaike Information Criterion (AIC) (Akaike, 1974), Bayesian Information Criterion (BIC) (Schwarz et al., 1978), Corrected AIC (AICc) (Hurvich and Tsai, 1989), Hannan-Quinn Information Criterion (HQIC) (Hannan and Quinn, 1979), Consistent AIC (CAIC) (Bozdogan, 1987) and Sample-size Adjusted BIC (SABIC) are given respectively by

$$\begin{aligned} \text{AIC}(\alpha) &= -2l_p(\hat{\boldsymbol{\theta}},\alpha)+2\left[p+q+df(\alpha)\right], \\ \text{BIC}(\alpha) &= -2l_p(\hat{\boldsymbol{\theta}},\alpha)+\log(n)\left[p+q+df(\alpha)\right], \\ \text{AICc}(\alpha) &= -2l_p(\hat{\boldsymbol{\theta}},\alpha)+\frac{2\left[p+q+df(\alpha)\right]\left[p+q+df(\alpha)+1\right]}{n-p-q-df(\alpha)-1}, \\ \text{HQIC}(\alpha) &= -2l_p(\hat{\boldsymbol{\theta}},\alpha)+2\log(\log(n))\left[p+q+df(\alpha)\right], \\ \text{CAIC}(\alpha) &= -2l_p(\hat{\boldsymbol{\theta}},\alpha)+(\log(n)+1)\left[p+q+df(\alpha)\right], \\ \text{SABIC}(\alpha) &= -2l_p(\hat{\boldsymbol{\theta}},\alpha)+\log\left(\frac{n+2}{24}\right)\left[p+q+df(\alpha)\right], \end{aligned}$$

where $l_p(\hat{\boldsymbol{\theta}},\alpha)$ denotes the penalized log-likelihood function available at $\hat{\boldsymbol{\theta}}$ for a fixed α. These measures can be used to select an appropriate model and/or value of α (see Ibacache-Pulgar and Reyes (2017) who used the AIC for this purpose, for example).

2.5.3 Obtaining the standard errors

There are several ways to obtain (approximate) estimates for the necessary/required standard errors, when some EM type algorithm is employed. Usually the observed information matrix is considered, see for example (Segal et al., 1994). In this work we consider the Louis principle (Louis, 1982), which relates the score

function of complete and incomplete log-likelihood through the conditional expectation $\bigtriangledown_0(\boldsymbol{\theta}) = E[\bigtriangledown_c(\boldsymbol{\theta}, \boldsymbol{Y}_c|\boldsymbol{Y}_{obs})]$, where $\bigtriangledown_0(\boldsymbol{\theta}) = \partial l_0(\boldsymbol{\theta}, \boldsymbol{Y}_{obs})/\partial\boldsymbol{\theta}$ and $\bigtriangledown_c(\boldsymbol{\theta}) = \partial l_c(\boldsymbol{\theta}, \boldsymbol{Y}_c)/\partial\boldsymbol{\theta}$ are the score function of incomplete and complete data, respectively. Also, Meilijson (1989) defined the empirical information as

$$\boldsymbol{I}_e(\boldsymbol{\theta}, \boldsymbol{y}) = \sum_{i=1}^{n} \boldsymbol{s}(y_i|\boldsymbol{\theta})\boldsymbol{s}(y_i|\boldsymbol{\theta})^\top - \frac{1}{n}\boldsymbol{S}(\boldsymbol{y}|\boldsymbol{\theta})\boldsymbol{S}(\boldsymbol{y}|\boldsymbol{\theta})^\top, \tag{2.16}$$

where $\boldsymbol{S}(\boldsymbol{y}|\boldsymbol{\theta}) = \sum_{i=1}^{n} \boldsymbol{s}(y_i|\boldsymbol{\theta})$ and $\boldsymbol{s}(y_i|\boldsymbol{\theta})$ is the empirical score function for the ith observation, which is given by

$$\boldsymbol{s}(y_i|\boldsymbol{\theta}) = E\left[\frac{\partial l_c(\boldsymbol{\theta}, \boldsymbol{y}_c)}{\partial\boldsymbol{\theta}}|\boldsymbol{y}_c\right] = \left(\boldsymbol{s}(y_i|\boldsymbol{\beta}), \boldsymbol{s}(y_i|\boldsymbol{\kappa}), \boldsymbol{s}(y_i|\sigma^2), \boldsymbol{s}(y_i|\delta)\right)^\top,$$

where

$$\begin{aligned}
\boldsymbol{s}(y_i|\boldsymbol{\beta}) &= -\frac{1}{\tau}\left[\boldsymbol{x}_i^\top\Delta\mathcal{E}_i^{0.5;1} - \boldsymbol{x}_i^\top(y_i - \mu_{ci})\mathcal{E}_i^{1;0}\right] \\
\boldsymbol{s}(y_i|\boldsymbol{\kappa}) &= -\frac{1}{\tau}\left[\boldsymbol{T}_i^\top\Delta\mathcal{E}_i^{0.5;1} - \boldsymbol{T}_i^\top(y_i - \mu_{ci})\mathcal{E}_i^{1;0}\right] - \frac{1}{2}\boldsymbol{\kappa}^\top\boldsymbol{\Lambda}.
\end{aligned}$$

Therefore

$$\begin{aligned}
\frac{\partial\Delta}{\partial\sigma^2} &= \frac{\Delta}{2\sigma^2},\ A_1 = \frac{\partial\tau}{\partial\sigma^2} = \frac{\tau}{\sigma^2}, A_2 = \frac{\partial\Delta/\tau}{\partial\sigma^2} = -\frac{\delta\sqrt{1-b^2\delta^2}}{2\sigma^3(1-\delta^2)}, \\
\boldsymbol{s}(y_i|\sigma^2) &= -\frac{n}{2}\frac{1}{\tau}A_1 - \frac{1}{2}\sum_{i=1}^{n} -\frac{1}{\tau^2}A_1(y_i - \mu_{ci})^2\mathcal{E}_i^{1;0} - 2(y_i - \mu_{ci})\mathcal{E}_i^{0.5;1}A_2 \\
&= -\frac{n}{2\sigma^2} + \frac{1}{2}\sum_{i=1}^{n}\frac{1}{\tau\sigma^2}(y_i - \mu_{ci})^2\mathcal{E}_i^{1;0} + 2(y_i - \mu_{ci})\mathcal{E}_i^{0.5;1}A_2, \\
A_3 &= \frac{\partial\Delta}{\partial\delta} = \frac{\sigma}{\sqrt{1-b^2\delta^2}}\left(1 + \frac{\delta^2b^2}{1-b^2\delta^2}\right), \\
A_4 &= \frac{\partial\tau}{\partial\delta} = \frac{2\delta\sigma^2}{1-b^2\delta^2}\left[\frac{(1-\delta^2)b^2}{1-b^2\delta^2} - 1\right], A_5 = \frac{\partial\Delta/\tau}{\partial\delta} = \frac{\tau A_3 - \Delta A_4}{\tau^2}, \\
A_6 &= \frac{\partial\Delta^2/\tau}{\partial\delta} = \frac{2\Delta\tau A_3 - \Delta^2 A_4}{\tau^2} \\
\boldsymbol{s}(y_i|\delta) &= -\frac{n}{2\tau}A_4 - \frac{1}{2}\sum_{i=1}^{n} -\frac{1}{\tau^2}A_4(y_i - \mu_{ci})^2\mathcal{E}_i^{1;0} - 2A_5(y_i - \mu_{ci})\mathcal{E}_i^{0.5;1} + A_6\mathcal{E}_i^{0;2}.
\end{aligned}$$

Replacing $\boldsymbol{\theta}$ for the respective maximum likelihood estimate, say $\tilde{\boldsymbol{\theta}}$, and assuming that $\bigtriangledown_0(\hat{\boldsymbol{\theta}}) = \boldsymbol{0}$, we have that Equation (2.16) can be rewritten as $\boldsymbol{I}_e(\hat{\boldsymbol{\theta}}, \boldsymbol{y}) = \sum_{i=1}^{n} \boldsymbol{s}(y_i|\boldsymbol{\theta})\boldsymbol{s}(y_i|\boldsymbol{\theta})^\top$.

Finally, the variance-covariance matrix of the maximum likelihood estimates can be approximated by $Cov(\hat{\boldsymbol{\theta}}) = \boldsymbol{I}_e^{-1}(\hat{\boldsymbol{\theta}}, \boldsymbol{y})$ and the respective standard errors are given by the square of root of the diagonal values of $Cov(\hat{\boldsymbol{\theta}})$.

2.5.4 Diagnostic analysis

Since the statistical models tend to be sensitive to the lack of underlying assumptions, performing diagnostic analysis is an essential step in data analysis.

Residual analysis are useful diagnostic tools for checking the departing from some specific model assumptions, as well as to provide overall subsidies concerning the goodness of model fit. The works of Cox and Snell (1968), Belsley et al. (1980) and Cook and Weisberg (1982) are seminal references. In terms of the so-called QQ-plots Atkinson (1981) proposed to build simulated based confidence bands (the so-called envelopes) to allow a better comparison between the residuals of interest and the percentiles of the reference (expected) distribution.

Another set of important techniques is the sensitive analysis, consisting on evaluating changes in the fitted model when perturbations are introduced into the data and/or (some) model assumptions. These techniques are divided into global and local influence analysis. In its turn, global influence analysis is usually divided into two approaches. The first one is the leverage analysis, which consists on studying the influence of a given observation on its respective predicted value (Hoaglin and Welsch, 1978). The second approach is the case deletion analysis, which can assess, for example, the impact of removing a particular observation on the parameter estimates of a regression model. In this case the Cook's distance (Cook, 1977) is commonly used for this purpose. On the other hand, in the local influence analyses (Cook, 1986) the effect of an infinitesimal perturbation on the data and/or on some component of the model, using a likelihood-based measure namely likelihood displacement, is considered.

For the proposed model we can use several of the aforementioned techniques and their respective generalizations. Indeed, Zhu and Lee (2001) proposed an one-step pseudo approximation for case deletion analysis which can be used to build a kind of generalized Cook's distance measure. Osorio (2006) proposed a generalized leverage measure for incomplete data. Maioli (2018b) proposed a residual that can be used for the SMCSN family. Cadigan and Farrell (2002a) proposed a generalization of likelihood displacement, which can be used to construct of perturbation schemes for local influence analysis.

2.5.5 Residual analysis

Based on the developed residuals in Maioli (2018b), we define the following residuals

$$R_i = \frac{Y_i - \boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}} - \sum_{j=1}^{n} \hat{\boldsymbol{\kappa}}_j^\top \boldsymbol{b}(T_{ij})}{\hat{\sigma}}, \tag{2.17}$$

where, according to Maioli (2018b), $R_i \overset{iid}{\sim} SMCSN(0, 1, \gamma, \boldsymbol{\nu})$.

2.5.6 Case deletion analysis

A given observation is said to be influential if, according to some criterion, it has a significant impact on the inference related to a given model. One of the most common techniques for influence analysis is the Cook's distance. It measures the impact of each observation on the respective predicted value by assessing the distance between $\hat{\boldsymbol{\theta}}$ (the estimate obtained with all observations), and without the observation y_i, namely $\hat{\boldsymbol{\theta}}_{(-i)}$. Let $l_{cp}(\boldsymbol{\theta}, \boldsymbol{y}_{c(-i)})$ be the complete-data penalized log-likelihood calculated without the observation y_i. Also, let $\hat{\boldsymbol{\theta}}_{(-i)} = \left(\hat{\boldsymbol{\beta}}_{(-i)}, \hat{\boldsymbol{\kappa}}_{(-i)}, \hat{\tau}_{(-i)}, \hat{\Delta}_{(-i)}, \hat{\boldsymbol{\nu}}_{(-i)}\right)$ be the argument that maximizes the penalized maximum likelihood using $Q_{(-i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \mathbb{E}\left[l_P(\boldsymbol{\theta}|\boldsymbol{y}_{c(-i)})|\boldsymbol{y}_{obs(-i)}, \hat{\boldsymbol{\theta}}\right]$. To obtain $\left(\hat{\boldsymbol{\theta}}_{(-i)}\right)$ we can use Algorithm 3 n times, i.e., removing each one of the observations. However, the computational effort can be quite high, mainly for large sample sizes. To circumvent that Zhu and Lee (2001) proposed the following one-step pseudo approximation:

$$\hat{\boldsymbol{\theta}}^*_{(-i)} = \hat{\boldsymbol{\theta}} + \left[-Q^{(2)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})\right]^{-1} Q^{(1)}_{(-i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}), \tag{2.18}$$

where

$$Q^{(2)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \left.\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad \text{and} \quad Q^{(1)}_{(-i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \left.\frac{\partial Q_{(-i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

are the Hessian matrix and the individual score vector evaluated at $\hat{\boldsymbol{\theta}}$, respectively. Notice that $Q^{(1)}_{(-i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = (Q^{(1)}_{(-i)\boldsymbol{\beta}}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}), Q^{(1)}_{(-i)\boldsymbol{\kappa}}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}), Q^{(1)}_{(-i)\tau}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}), Q^{(1)}_{(-i)\Delta}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}), Q^{(1)}_{(-i)\boldsymbol{\nu}}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}))^\top$ where

$$\begin{aligned}
Q^{(1)}_{(-i)\boldsymbol{\beta}}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial Q_{(-i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \frac{1}{\hat{\tau}}\sum_{l=1,l\neq i}^{n}\left[\boldsymbol{x}_l^\top(y_l - \hat{\mu}_{cl})\mathcal{E}_l^{(1;0)} - \boldsymbol{x}_l^\top\hat{\Delta}\mathcal{E}_l^{(0.5;1)}\right]\\
Q^{(1)}_{(-i)\boldsymbol{\kappa}}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial Q_{(-i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\kappa}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \frac{1}{\hat{\tau}}\sum_{l=1,l\neq i}^{n}\left[\boldsymbol{T}_l^\top(y_l - \hat{\mu}_{cl})\mathcal{E}_l^{(1;0)} - \boldsymbol{T}_l^\top\hat{\Delta}\mathcal{E}_l^{(0.5;1)} - \frac{1}{n}\boldsymbol{\Lambda}\hat{\boldsymbol{\kappa}}\right]\\
Q^{(1)}_{(-i)\tau}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial Q_{(-i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\tau}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -\frac{1}{2\hat{\tau}^2}\sum_{l=1,l\neq i}^{n}\Big[\hat{\tau} - (y_l - \hat{\mu}_{cl})^2\mathcal{E}_l^{(1;0)}\\
&\qquad + 2(y_l - \hat{\mu}_{cl})\hat{\Delta}\mathcal{E}_l^{(0.5;1)} - \hat{\Delta}^2\mathcal{E}_l^{(0;2)}\Big]\\
Q^{(1)}_{(-i)\Delta}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) &= \left.\frac{\partial Q_{(-i)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\Delta}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \frac{1}{\hat{\tau}}\sum_{l=1,l\neq i}^{n}\left[(y_l - \hat{\mu}_{cl})\mathcal{E}_l^{(0.5;1)} - \hat{\Delta}\mathcal{E}_l^{(0;2)}\right].
\end{aligned}$$

According to Zhu and Lee (2001) to measure the distance between $\hat{\boldsymbol{\theta}}_{(-i)}$ and $\hat{\boldsymbol{\theta}}$, we can compute the generalized Cook's distance:

$$GCD_i = \left(\hat{\boldsymbol{\theta}}_{(-i)} - \hat{\boldsymbol{\theta}}\right)^\top\left\{-Q^{(2)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})\right\}\left(\hat{\boldsymbol{\theta}}_{(-i)} - \hat{\boldsymbol{\theta}}\right). \tag{2.19}$$

Now, using (2.19) in (2.18) we obtain an approximation to the generalized Cook's distance, given by $GCD_i^1 = Q^{(1)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})^\top\left\{-Q^{(2)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})\right\}^{-1}Q^{(1)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$, $i = 1, \ldots, n$,

where:

$$Q^{(2)}_{\boldsymbol{\beta}}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \left.\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -\frac{1}{\hat{\tau}}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^\top\mathcal{E}_i^{(1;0)}$$

$$Q^{(2)}_{\boldsymbol{\beta\kappa}}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \left.\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\kappa}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -\frac{1}{\hat{\tau}}\sum_{i=1}^{n}\boldsymbol{T}_i\boldsymbol{x}_i^\top\mathcal{E}_i^{(1;0)}$$

$$Q^{(2)}_{\boldsymbol{\beta}\tau}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \left.\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\tau}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -\frac{1}{\hat{\tau}^2}\sum_{i=1}^{n}\left[\boldsymbol{x}_i^\top(y_l-\hat{\mu}_{ci})\mathcal{E}_i^{(1;0)} - \boldsymbol{x}_i^\top\hat{\Delta}\mathcal{E}_i^{(0.5;1)}\right]$$

$$Q^{(2)}_{\boldsymbol{\beta}\Delta}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \left.\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\Delta}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -\frac{1}{\hat{\tau}}\sum_{i=1}^{n}\boldsymbol{x}_i^\top\mathcal{E}_i^{(0.5;1)}$$

$$Q^{(2)}_{\boldsymbol{\kappa}}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \left.\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\kappa}\partial\boldsymbol{\kappa}^\top}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -\frac{1}{\hat{\tau}}\sum_{i=1}^{n}\boldsymbol{T}_i\boldsymbol{T}_i^\top\mathcal{E}_i^{(1;0)} - \boldsymbol{\Lambda}$$

$$Q^{(2)}_{\boldsymbol{\kappa}\tau}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \left.\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\kappa}\partial\tau}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -\frac{1}{\hat{\tau}^2}\sum_{i=1}^{n}\left[\boldsymbol{T}_i^\top(y_i-\hat{\mu}_{ci})\mathcal{E}_i^{(1;0)} - \boldsymbol{T}_i^\top\hat{\Delta}\mathcal{E}_i^{(0.5;1)} - \frac{1}{n}\boldsymbol{\Lambda}\hat{\boldsymbol{\kappa}}\right]$$

$$Q^{(2)}_{\boldsymbol{\kappa}\Delta}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \left.\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\kappa}\partial\Delta}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -\frac{1}{\hat{\tau}}\sum_{i=1}^{n}\boldsymbol{T}_i^\top\mathcal{E}_i^{(0.5;1)}$$

$$Q^{(2)}_{\tau}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \left.\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\tau^2}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \frac{1}{\hat{\tau}^3}\sum_{i=1}^{n}\left[\frac{\hat{\tau}}{2} - (y_i-\hat{\mu}_{ci})^2\mathcal{E}_i^{(1;0)}\right.$$
$$\left. + 2(y_i-\hat{\mu}_{ci})\hat{\Delta}\mathcal{E}_i^{(0.5;1)} - \hat{\Delta}^2\mathcal{E}_i^{(0;2)}\right]$$

$$Q^{(2)}_{\tau\Delta}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \left.\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\tau\partial\Delta}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \frac{1}{\hat{\tau}^2}\sum_{i=1}^{n}\left[\hat{\Delta}\mathcal{E}_i^{(0;2)} - (y_i-\hat{\mu}_{ci})\mathcal{E}_i^{(0.5;1)}\right]$$

$$Q^{(2)}_{\Delta}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \left.\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial\Delta^2}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = -\frac{1}{\hat{\tau}}\sum_{i=1}^{n}\mathcal{E}_i^{(0;2)}.$$

2.5.7 Generalized leverage

In Normal Linear models the main idea of the leverage analysis is to study the influence of the ith observation y_i on the fitted value itself $\hat{y}_i$ through h_{ii}, the elements of main diagonal of the Hat matrix $\boldsymbol{H}$, once $h_{ii} = \partial\hat{y}_i/\partial y_i$, i.e., h_{ii} is the variation of $\hat{y}_i$ when y_i is increased by an infinite one. For more details about the leverage analysis, see for example, Hoaglin and Welsch (1978), Emerson et al. (1984) and Wei et al. (1998).

Wei et al. (1998) proposed a leverage measure for a general class of regression models, namely, generalized leverage. Following this idea, Osorio (2006) introduced the generalized leverage for incomplete data and Ferreira et al. (2015) extended this concept to the SMSN regression models. The generalized leverage is given by $\boldsymbol{GH}(\hat{\boldsymbol{\theta}}) = \boldsymbol{D}_{\boldsymbol{\theta}}\left[-\ddot{\boldsymbol{Q}}_{\boldsymbol{\theta}}\right]^{-1}\ddot{\boldsymbol{Q}}_{\boldsymbol{\theta},\boldsymbol{y}}$, where $\boldsymbol{D}_{\boldsymbol{\theta}} = \partial\boldsymbol{\mu}/\partial\boldsymbol{\theta}^\top$ and $\ddot{\boldsymbol{Q}}_{\boldsymbol{\theta},\boldsymbol{y}} = \partial^2\boldsymbol{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{y}^\top$ are the Hessian matrix given in section 2.5.6. Thus, the elements of $\boldsymbol{D}_{\boldsymbol{\theta}}$ are given by $\boldsymbol{D}_{\beta} = \boldsymbol{X}$, $\boldsymbol{D}_{\kappa} = \boldsymbol{T}$ and $\boldsymbol{D}_{\Delta} = \boldsymbol{D}_{\tau} = \boldsymbol{0}$, and the elements of $\ddot{\boldsymbol{Q}}_{\boldsymbol{\theta},\boldsymbol{y}}$ are given by $\ddot{\boldsymbol{Q}}_{\boldsymbol{\beta},\boldsymbol{y}} = \frac{1}{\tau}\boldsymbol{X}^\top\boldsymbol{\mathcal{E}}^{(1;0)}$, $\ddot{\boldsymbol{Q}}_{\boldsymbol{\kappa},\boldsymbol{y}} = \frac{1}{\tau}\boldsymbol{T}^\top\boldsymbol{\mathcal{E}}^{(1;0)}$, $\ddot{\boldsymbol{Q}}_{\boldsymbol{\Delta},\boldsymbol{y}} = \frac{1}{\tau}\boldsymbol{\mathcal{E}}^{(0.5;1)}$, $\ddot{\boldsymbol{Q}}_{\boldsymbol{\tau},\boldsymbol{y}} = \frac{1}{\tau^2}\left[(\boldsymbol{y}-\boldsymbol{\mu}_c)\boldsymbol{\mathcal{E}}^{(1;0)} - \Delta\boldsymbol{\mathcal{E}}^{(0.5;1)}\right]$.

2.5.8 Local influence

The local influence method proposed by Cook (1986) consists on the assessment of the model under small perturbations on the model assumptions and/or on the data. If such perturbations lead to significance inferential changes, the model could not be robust against them. Therefore, the model could be not suitable to analyze the data. Through this methodology we can assess the model robustness against outliers and/or lacking of model assumptions, as heteroscedasticidity. Also, the behavior of covariates in the regression structures, can be analyzed.

The most usual measure of influence is the likelihood displacement proposed by Cook (1986). However, for our case we can build a displacement measure using the function Q, following the idea of Zhu and Lee (2001) and Cadigan and Farrell (2002b) we can define the Q-displacement and we can get various perturbation schemes using it.

Let us consider a perturbation vector, say $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)^\top$ restricted to some open subset $\boldsymbol{\Omega} \in \mathbb{R}^m$. Let $l_{cp}(\boldsymbol{\theta}, \boldsymbol{\omega}|\boldsymbol{y}_c)$ be the penalized complete log-likelihood of the perturbed model. Also, let $\hat{\boldsymbol{\theta}}(\boldsymbol{\omega}) = \left(\hat{\boldsymbol{\beta}}(\boldsymbol{\omega})^\top, \hat{\boldsymbol{\kappa}}(\boldsymbol{\omega})^\top, \hat{\tau}(\boldsymbol{\omega})^\top, \hat{\Delta}(\boldsymbol{\omega})^\top\right)^\top$ be the maximum of the function $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \mathbb{E}\left[l_P(\boldsymbol{\theta}, \boldsymbol{\omega}|\boldsymbol{y}_c)|\boldsymbol{y}_{obs}, \hat{\boldsymbol{\theta}}\right]$. Then the Q-displacement is given by $QD(\boldsymbol{\omega}) = 2\left\{Q(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) - Q(\hat{\boldsymbol{\theta}}(\boldsymbol{\omega})|\hat{\boldsymbol{\theta}})\right\}$.

Cook (1986) proposed to study the local behavior of $QD(\boldsymbol{\omega})$ for any value of $\boldsymbol{\omega}$ in a neighborhood of $\boldsymbol{\omega}_0$, which represents the null perturbation vector, such that $Q(\hat{\boldsymbol{\theta}}(\boldsymbol{\omega}_0)|\hat{\boldsymbol{\theta}}) = Q(\hat{\boldsymbol{\theta}}|\hat{\boldsymbol{\theta}}) \Longrightarrow QD(\boldsymbol{\omega}_0) = 0$. The length of the vector $\boldsymbol{\omega}$, m, depends on the proposed perturbation scheme. It is considered a $(m+1)$-dimensional surface called Influence Graph, which is function of $\boldsymbol{\alpha}_\omega = \left[\boldsymbol{\omega}^\top, QD(\boldsymbol{\omega})\right]^\top$ when $\boldsymbol{\omega} \in \boldsymbol{\Omega}$. Then, the local influence method consists on evaluating how the surface $\boldsymbol{\alpha}_\omega$ deviates from the tangent plane in $\boldsymbol{\omega}_0$. Such analysis can be done by studying the curvatures of the normal surface sections $\boldsymbol{\alpha}_\omega$ in $\boldsymbol{\omega}_0$. Verbeke and Molenberghs (2000) illustrated through Figure 22 the normal curvature for a surface $\boldsymbol{\alpha}_\omega$.

The intersection between the normal section and the tangent plane T_0 is named projected line. Cook (1986) suggests to study the normal curvature (Bates and Watts, 1980) of the projected line on the plot $QD(\boldsymbol{\omega}_0 + a\boldsymbol{d}) \times a$, where $a \in \mathbb{R}$ and $\boldsymbol{d}$ is an arbitrary direction of norm equals to one ($||d|| = 1$). It can be shown that the normal curvature in the $\boldsymbol{d}$ direction is given by (Cook, 1986):

$$C_{QD,\boldsymbol{d}} = 2|\boldsymbol{d}^\top \boldsymbol{\Upsilon}_{\boldsymbol{\theta},\boldsymbol{\omega}_0}^\top \left\{Q^{(2)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})\right\}^{-1} \boldsymbol{\Upsilon}_{\boldsymbol{\theta},\boldsymbol{\omega}_0}\boldsymbol{d}|, \tag{2.20}$$

leading to $-Q^{(2)}_{\boldsymbol{\omega}_0} = \boldsymbol{\Upsilon}_{\boldsymbol{\theta},\boldsymbol{\omega}_0}^\top \left\{Q^{(2)}(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})\right\}^{-1} \boldsymbol{\Upsilon}_{\boldsymbol{\theta},\boldsymbol{\omega}_0}$, where $\boldsymbol{\Upsilon}_{\boldsymbol{\theta},\boldsymbol{\omega}_0} = \partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\omega}^\top = \left(\boldsymbol{\Upsilon}_{\boldsymbol{\beta},\boldsymbol{\omega}_0}^\top, \boldsymbol{\Upsilon}_{\boldsymbol{\kappa},\boldsymbol{\omega}_0}^\top, \boldsymbol{\Upsilon}_{\tau,\boldsymbol{\omega}_0}^\top, \boldsymbol{\Upsilon}_{\Delta,\boldsymbol{\omega}_0}^\top\right)^\top$ at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. According to Cook (1986) the information provided by $-Q^{(2)}_{\boldsymbol{\omega}_0}$ is fundamental for detecting influential observations in $QD(\boldsymbol{\omega})$. We can

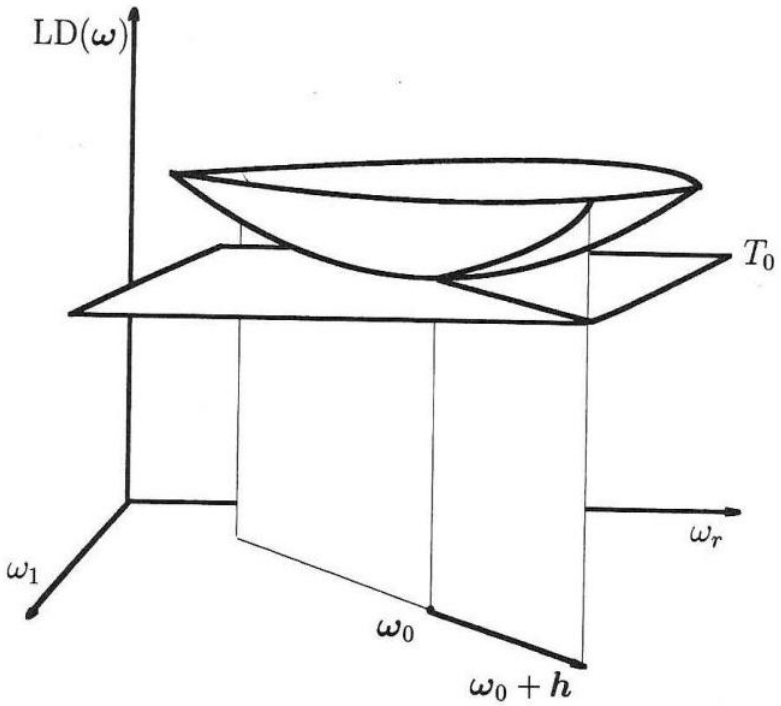


Figure 22 – Normal curvature for a surface α_ω and unitary direction h.

Source: Verbeke and Molenberghs (2000).

analyze the direction of the d_{max}, the eigenvector corresponding to the projected line of largest curvature C_{max} associated with the highest eigenvalue of the matrix $-Q^{(2)}_{\omega_0}$. The most usual plot for local influence analysis is the index versus d_{max}.

We developed local influence schemes for case-weight perturbation, the response variable perturbation, the scale parameter perturbation, the skewness parameter perturbation and continuous covariate perturbation. In a general way, the case-weight perturbation can be interpreted as a perturbation in the variance of each experimental unit. The response variable disturbance can be seen as a tool for identifying outliers (Schwarzmann, 1991a). The individual perturbation scheme of covariates helps to evaluate the influence of each one of them on the estimation process. However, this scheme makes sense only if the covariate is continuous. Disturbances in the scale parameter and skewness are useful for checking the model sensitivity to the lacking of the homogeneity of these parameters along the observations.

2.5.8.1 Case-weight perturbation

Let us consider the following the perturbation scheme:

$$Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \omega_i \mathbb{E}\left[l_{cpi}(\boldsymbol{\theta}|\boldsymbol{y}_c)|\boldsymbol{y}_{obs}, \hat{\boldsymbol{\theta}}\right] = \sum_{i=1}^{n} \omega_i Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) - \frac{1}{2}\boldsymbol{\kappa}^\top \boldsymbol{\Lambda}\boldsymbol{\kappa},$$

where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$ and $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$. The elements of $\boldsymbol{\Upsilon}_{\boldsymbol{\theta},\boldsymbol{\omega}_0}$ of Equation (2.20) are given by:

$$\begin{aligned}
\boldsymbol{\Upsilon}_{\boldsymbol{\beta},\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\omega_i} = \frac{1}{\hat{\tau}}\boldsymbol{x}_i^\top\left[(y_i-\hat{\mu}_{ci})\mathcal{E}_i^{(1;0)} - \hat{\Delta}\mathcal{E}_i^{(0.5;1)}\right]\\
\boldsymbol{\Upsilon}_{\boldsymbol{\kappa},\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\kappa}\partial\omega_i} = \frac{1}{\hat{\tau}}\boldsymbol{T}_i^\top\left[(y_i-\hat{\mu}_{ci})\mathcal{E}_i^{(1;0)} - \hat{\Delta}\mathcal{E}_i^{(0.5;1)}\right]\\
\boldsymbol{\Upsilon}_{\sigma^2,\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\omega_i} = -\frac{1}{2\sigma^2} + \frac{1}{2\tau\sigma^2}(y_i-\mu_{ci})^2\mathcal{E}_i^{(1;0)} + A_2(y_i-\mu_{ci})\mathcal{E}_i^{(0.5;1)} - \frac{1}{2}A_7\mathcal{E}_i^{(0;2)}\\
\boldsymbol{\Upsilon}_{\delta,\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\delta\partial\omega_i} = -\frac{A_4}{2\tau} + \frac{A_4}{2\tau^2}(y_i-\mu_{ci})^2\mathcal{E}_i^{(1;0)} + A_5(y_i-\mu_{ci})\mathcal{E}_i^{(0.5;1)} - \frac{1}{2}A_6\mathcal{E}_i^{(0;2)},
\end{aligned}$$

where

$$A_7 = \frac{\partial\Delta^2/\tau}{\partial\sigma^2} = \frac{\Delta^2\tau}{\sigma^6} - \frac{A_1\Delta^2}{\sigma^4}.$$

2.5.8.2 Scale perturbation

We assume that $Y_i \overset{ind}{\sim} SMCSN(\mu_i, \omega_i^{-1}\sigma^2, \gamma, \boldsymbol{\nu})$, $\omega_i > 0$ for $i = 1, \ldots, n$. The perturbed Q-function, under this scheme, is given by:

$$\begin{aligned}
Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) &= \sum_{i=1}^{n} -\frac{1}{2}\log\left(\frac{\hat{\tau}}{\omega_i}\right) - \frac{\omega_i}{2\hat{\tau}}\Bigg[\left(y_i-\hat{\mu}_{ci}^{(k)}\right)^2\mathcal{E}_i^{(1;0)} - 2\left(y_i-\hat{\mu}_{ci}^{(k)}\right)\frac{\Delta}{\sqrt{\omega_i}}\mathcal{E}_i^{(0.5;1)}\\
&\quad + \frac{\Delta^2}{\omega_i}\mathcal{E}_i^{(0;2)}\Bigg] - \frac{1}{2}\hat{\boldsymbol{\kappa}}^{(k)\top}\boldsymbol{\Lambda}\hat{\boldsymbol{\kappa}}^{(k)}\\
&= \sum_{i=1}^{n} -\frac{1}{2}\log\left(\frac{\hat{\tau}}{\omega_i}\right) - \frac{1}{2\hat{\tau}}\Bigg[\omega_i\left(y_i-\hat{\mu}_{ci}^{(k)}\right)^2\mathcal{E}_i^{(1;0)} - 2\left(y_i-\hat{\mu}_{ci}^{(k)}\right)\sqrt{\omega_i}\Delta\mathcal{E}_i^{(0.5;1)}\\
&\quad + \Delta^2\mathcal{E}_i^{(0;2)}\Bigg] - \frac{1}{2}\hat{\boldsymbol{\kappa}}^{(k)\top}\boldsymbol{\Lambda}\hat{\boldsymbol{\kappa}}^{(k)},
\end{aligned}$$

where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^\top$, $\Delta_{\omega_i} = \Delta/\sqrt{\omega_i}$, $\tau_{\omega_i} = \tau/\omega_i$ and $\boldsymbol{\omega}_0 = (1, \ldots, 1)^\top$. The elements of $\boldsymbol{\Upsilon}_{\boldsymbol{\theta},\boldsymbol{\omega}_0}$ in Equation (2.20) are given by:

$$\begin{aligned}
\boldsymbol{\Upsilon}_{\boldsymbol{\beta},\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\omega_i} = \frac{1}{\hat{\tau}}\boldsymbol{x}_i^\top\left[(y_i-\hat{\mu}_{ci})\mathcal{E}_i^{(1;0)} - \frac{1}{2}\hat{\Delta}\mathcal{E}_i^{(0.5;1)}\right]\\
\boldsymbol{\Upsilon}_{\boldsymbol{\kappa},\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\kappa}\partial\omega_i} = \frac{1}{\hat{\tau}}\boldsymbol{T}_i^\top\left[(y_i-\hat{\mu}_{ci})\mathcal{E}_i^{(1;0)} - \frac{1}{2}\hat{\Delta}\mathcal{E}_i^{(0.5;1)}\right]\\
\boldsymbol{\Upsilon}_{\sigma^2,\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\omega_i} = \frac{A_1}{2\tau^2}(y_i-\mu_{ci})^2\mathcal{E}_i^{(1;0)} + \frac{A_2}{2}(y_i-\mu_{ci})\mathcal{E}_i^{(0.5;1)}\\
\boldsymbol{\Upsilon}_{\gamma,\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\gamma\partial\omega_i} = \frac{A_4}{2\tau^2}(y_i-\mu_{ci})^2\mathcal{E}_i^{(1;0)} + \frac{A_5}{2}(y_i-\mu_{ci})\mathcal{E}_i^{(0.5;1)}.
\end{aligned}$$

2.5.8.3 Skewness perturbation

We assume that $\delta_{\omega_i} = \delta/\omega_i$, $\omega_i > 0$ for $i = 1, \ldots, n$. The perturbed Q-function, under this scheme, is given by:

$$
\begin{aligned}
Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) &= \sum_{i=1}^{n} -\frac{1}{2}\log(\hat{\tau}_{\omega_i}) - \frac{1}{2\hat{\tau}_{\omega_i}}\Bigg[\left(y_i-\hat{\mu}_{ci}^{(k)}\right)^2\mathcal{E}_i^{(1;0)} - 2\left(y_i-\hat{\mu}_{ci}^{(k)}\right)\Delta_{\omega_i}\mathcal{E}_i^{(0.5;1)} \\
&\quad + \Delta_{\omega_i}^2\mathcal{E}_i^{(0;2)}\Bigg] - \frac{1}{2}\hat{\boldsymbol{\kappa}}^{(k)\top}\boldsymbol{\Lambda}\hat{\boldsymbol{\kappa}}^{(k)} \\
&= \sum_{i=1}^{n} -\frac{1}{2}\log(\hat{\tau}_{\omega_i}) + \Bigg[-\frac{1}{2\hat{\tau}_{\omega_i}}\left(y_i-\hat{\mu}_{ci}^{(k)}\right)^2\mathcal{E}_i^{(1;0)} + \frac{\Delta_{\omega_i}}{\hat{\tau}_{\omega_i}}\left(y_i-\hat{\mu}_{ci}^{(k)}\right)\mathcal{E}_i^{(0.5;1)} \\
&\quad - \frac{\Delta_{\omega_i}^2}{2\hat{\tau}_{\omega_i}}\mathcal{E}_i^{(0;2)}\Bigg] - \frac{1}{2}\hat{\boldsymbol{\kappa}}^{(k)\top}\boldsymbol{\Lambda}\hat{\boldsymbol{\kappa}}^{(k)},
\end{aligned}
$$

where $\boldsymbol{\omega} = (\omega_1,\ldots,\omega_n)^\top$ and $\boldsymbol{\omega}_0 = (1,\ldots,1)^\top$. The elements of $\boldsymbol{\Upsilon}_{\boldsymbol{\theta},\boldsymbol{\omega}_0}$ in Equation (2.20) are given by:

$$
\begin{aligned}
\boldsymbol{\Upsilon}_{\boldsymbol{\beta},\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\omega_i} = \delta\boldsymbol{x}_i^\top\left[A_5\mathcal{E}_i^{(0.5;1)} - \frac{A_4}{\tau^2}(y_i-\mu_{ci})\mathcal{E}_i^{(1;0)}\right] \\
\boldsymbol{\Upsilon}_{\boldsymbol{\kappa},\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\kappa}\partial\omega_i} = \delta\boldsymbol{T}_i^\top\left[A_5\mathcal{E}_i^{(0.5;1)} - \frac{A_4}{\tau^2}(y_i-\mu_{ci})\mathcal{E}_i^{(1;0)}\right] \\
\boldsymbol{\Upsilon}_{\sigma^2,\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\omega_i} = \frac{A_4\delta}{2\sigma^2\tau^2}(y_i-\mu_{ci})^2\mathcal{E}_i^{(1;0)} - \delta A_2^\delta(y_i-\mu_{ci})\mathcal{E}_i^{(0.5;1)} \\
\boldsymbol{\Upsilon}_{\delta,\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\delta\partial\omega_i} = \frac{\delta}{2\tau^2}\left(A_4^\delta\tau - A_4^2\right) + \frac{\delta}{2\tau^4}\left(A_4^\delta\tau^2 - 2A_4^2\tau\right)(y_i-\mu_{ci})^2\mathcal{E}_i^{(1;0)} - \\
&\quad \delta A_5^\delta(y_i-\mu_{ci})\mathcal{E}_i^{(0.5;1)} + \frac{\delta A_6^\delta}{2}\mathcal{E}_i^{0;2},
\end{aligned}
$$

where

$$
\begin{aligned}
A_2^\delta &= \frac{\partial A_2}{\partial\delta} = -\frac{1}{2\sigma^3(1-\delta^2)^2}\left(\sqrt{1-b^2\delta^2} + \frac{\delta^2 b^2}{\sqrt{1-b^2\delta^2}}\right) \\
A_4^\delta &= \frac{\partial A_4}{\partial\delta} = \frac{2\sigma^2}{(1-b^2\delta^2)^2}\left\{\frac{b^2}{1-b^2\delta^2}\left[(1-\delta^2)(1-b^2\delta^2) + 4\delta^2b^2(1-\delta^2)\right] - (1+b^2\delta^2)\right\} \\
A_3^\delta &= \frac{\partial A_3}{\partial\delta} = \frac{\sigma\delta b^2}{(1-b^2\delta^2)^{3/2}}\left\{1 + \frac{1}{1-b^2\delta^2}\left[2(1-b^2\delta^2) + 3\delta^2b^2\right]\right\} \\
\frac{\partial^2\Delta}{\partial\delta^2} &= \sigma\left[-\frac{1}{2(1-b^2\delta^2)(3/2)} + \frac{b^2}{(1-b^2\delta^2)^3}\left(2\delta(1-b^2\delta^2)^{3/2} + 3\delta^3b^2(1-b^2\delta^2)\right)\right] \\
A_5^\delta &= \frac{\partial A_5}{\partial\delta} = \frac{1}{\tau^3}\left[\tau\left(\tau A_3^\delta - \Delta A_4^\delta\right) - 2A_4^\delta(\tau A_3 - \Delta A_4)\right] \\
A_6^\delta &= \frac{\partial A_6}{\partial\delta} = \frac{1}{\tau^3}\left[\tau(2A_3^2\tau + 2\Delta\tau A_3^\delta - \Delta^2A_4^\delta) - 2A_4(2\Delta\tau A_3 - \Delta^2A_4)\right].
\end{aligned}
$$

2.5.8.4 Continuous explanatory variable perturbation

Let $\boldsymbol{\omega} = (\omega_1,\ldots,\omega_n)^\top$ be the perturbation vector. We consider an additive perturbation to the r-th covariate given by $\boldsymbol{x}_{i\boldsymbol{\omega}}^\top = \boldsymbol{x}_i^\top + \omega_i S_r\boldsymbol{e}_r^\top$, $i = 1,\ldots,n$ and $r = 1,\ldots,p$, where S_r is the standard deviation of the rth covariate vector and $\boldsymbol{e}_r^\top$ is a p-vector with 1 in r-th position and 0's elsewhere. In this scheme we have:

$$Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = -\frac{n}{2}\log\hat{\tau}^{(k)} - \frac{1}{2\hat{\tau}^{(k)}}\sum_{i=1}^{n}\Bigg[\left(y_i-\hat{\mu}_{ci}^{*(k)}\right)^2\mathcal{E}_i^{(1;0)} - 2\left(y_i-\hat{\mu}_{ci}^{*(k)}\right)\Delta\mathcal{E}_i^{(0.5;1)} + \Delta^2\mathcal{E}_i^{(0;2)}\Bigg] - \frac{1}{2}\hat{\boldsymbol{\kappa}}^{(k)\top}\boldsymbol{\Lambda}\hat{\boldsymbol{\kappa}}^{(k)},$$

where $\hat{\mu}_{ci}^{*} = \boldsymbol{x}_{i\boldsymbol{\omega}}^{\top}\boldsymbol{\beta} + \sum_{j=1}^{k}\boldsymbol{T}_{ij}\boldsymbol{\kappa}_j$. In this case the non perturbation vector is given by $\boldsymbol{\omega}_0 = (0,\ldots,0)^{\top}$. The elements of $\boldsymbol{\Upsilon}_{\boldsymbol{\theta},\boldsymbol{\omega}_0}$ in Equation (2.20) are given by:

$$\begin{aligned}
\boldsymbol{\Upsilon}_{\boldsymbol{\beta},\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\omega_i} = \frac{1}{\tau}\left\{\left[S_r\boldsymbol{e}_r^{\top}(y_i-\mu_{ci}) - \boldsymbol{x}_i^{\top}S_r\boldsymbol{e}_r^{\top}\boldsymbol{\beta}\right]\mathcal{E}_i^{(1;0)} - S_r\boldsymbol{e}_r^{\top}\Delta\mathcal{E}_i^{(0.5;1)}\right\}\\
\boldsymbol{\Upsilon}_{\boldsymbol{\kappa},\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\kappa}\partial\omega_i} = -\frac{S_r\boldsymbol{e}_r^{\top}\boldsymbol{\beta}}{\hat{\tau}}\boldsymbol{T}_i^{\top}\mathcal{E}_i^{(1;0)}\\
\boldsymbol{\Upsilon}_{\sigma^2,\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\omega_i} = \frac{1}{\tau\sigma^2}S_r\boldsymbol{e}_r^{\top}(y_i-\mu_{ci})\mathcal{E}_i^{(1;0)} - S_r\boldsymbol{e}_r^{\top}A_2\mathcal{E}_i^{(0.5;1)}\\
\boldsymbol{\Upsilon}_{\gamma,\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\gamma\partial\omega_i} = \frac{1}{\tau^2}A_4S_r\boldsymbol{e}_r^{\top}(y_i-\mu_{ci})\mathcal{E}_i^{(1;0)} - S_r\boldsymbol{e}_r^{\top}A_5\mathcal{E}_i^{(0.5;1)}.
\end{aligned}$$

2.5.8.5 Response variable perturbation

In this case we consider again an additive perturbation scheme given by $y_{i\omega_i} = y_i + S_y\omega_i$, for $i = 1,\ldots,n$. The perturbation and non-perturbation vectors are given by $\boldsymbol{\omega} = (\omega_1,\ldots,\omega_n)^{\top}$ and $\boldsymbol{\omega}_0 = (0,\ldots,0)^{\top}$, respectively. The perturbed Q-function is given by

$$Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = -\frac{n}{2}\log\hat{\tau}^{(k)} - \frac{1}{2\hat{\tau}^{(k)}}\sum_{i=1}^{n}\Bigg[\left(y_{i\omega_i}-\hat{\mu}_{ci}^{(k)}\right)^2\mathcal{E}_i^{(1;0)} - 2\left(y_{i\omega_i}-\hat{\mu}_{ci}^{(k)}\right)\Delta\mathcal{E}_i^{(0.5;1)} + \Delta^2\mathcal{E}_i^{(0;2)}\Bigg] - \frac{1}{2}\hat{\boldsymbol{\kappa}}^{(k)\top}\boldsymbol{\Lambda}\hat{\boldsymbol{\kappa}}^{(k)},$$

the elements of $\boldsymbol{\Upsilon}_{\boldsymbol{\theta},\boldsymbol{\omega}_0}$, in Equation (2.20), are given by:

$$\begin{aligned}
\boldsymbol{\Upsilon}_{\boldsymbol{\beta},\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\omega_i} = \frac{S_y}{\hat{\tau}}\boldsymbol{x}_i^{\top}\mathcal{E}_i^{(1;0)}\\
\boldsymbol{\Upsilon}_{\boldsymbol{\kappa},\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\kappa}\partial\omega_i} = \frac{S_y}{\hat{\tau}}\boldsymbol{T}_i^{\top}\mathcal{E}_i^{(1;0)}\\
\boldsymbol{\Upsilon}_{\sigma^2,\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\sigma^2\partial\omega_i} = -\frac{S_y}{\tau\sigma^2}(y_i-\mu_{ci})\mathcal{E}_i^{(1;0)} + A_2S_y\mathcal{E}_i^{(0.5;1)}\\
\boldsymbol{\Upsilon}_{\delta,\boldsymbol{\omega}_0} &= \frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\omega}|\hat{\boldsymbol{\theta}})}{\partial\delta\partial\omega_i} = -\frac{A_4}{\tau^2}S_y(y_i-\mu_{ci})\mathcal{E}_i^{(1;0)} + A_5S_y\mathcal{E}_i^{(0.5;1)}.
\end{aligned}$$

2.5.9 Simulation studies

Following, we will analyze the performance of the proposed models and the estimators through two simulation studies: one for the parameters recovery and another for the misspecification of the distribution of the response variable.

2.5.9.1 Simulation study 1

This simulation study aims to investigate asymptotic properties of the proposed model. We generate 100 Monte Carlo replicas from: $Y_i = x_{1i}+2x_{2i}+cos(t_i)+\epsilon_i$, $i = 1,\ldots,n$, where $x_{11},\ldots,x_{1n} \overset{iid}{\sim} U(0,1)$ and $x_{21},\ldots,x_{2n} \overset{iid}{\sim} U(0,2)$, considering $n = 50, 100$ and 300, $t_i \in (0, 3\pi)$, 300 iterations of the SAEM algorithm, $k = 1/3$ and the following scenarios, chosen in such a way that we have error distributions with high asymmetry and heavy tails: $\epsilon_i \overset{iid}{\sim} CSN(0,1,0.8)$, $\epsilon_i \overset{iid}{\sim} CST(0,1,-0.8,5)$, $\epsilon_i \overset{iid}{\sim} CSS(0,1,0.8,3)$, $\epsilon_i \overset{iid}{\sim} CSCN(0,1,-0.8,0.5,0.5)$, $\epsilon_i \overset{iid}{\sim} CSGT(0,1,0.8,15,5)$, $\epsilon_i \overset{iid}{\sim} CSBP(0,1,-0.8,3,3)$, $\epsilon_i \overset{iid}{\sim} CSBS(0,1,0.8,1,1)$ and $\epsilon_i \overset{iid}{\sim} CSGG(0,1,-0.8,2,1,0.66)$, independent for $i = 1,\ldots,n$.

We decided to set the smoothness parameter α at 0.01, since for certain replicas under low sample size, the poor estimation of the scale parameter lead to non-smooth non-parametric curves. Indeed, small values of α can lead to an overfitting (Hastie and Tibshirani, 1990). However, since in the simulations we only want to check how close the estimated and curves are, we can consider that value.

For the CSGGN distribution we noticed a bias in its estimates that increase as the sample size increases. Maybe this is due to some identifiability problem, probably, between the scale parameter and the shape parameters. As example, when $\nu_2 = 1$ the mixing distribution generalized gamma reduces to the gamma distribution and in this case the CSGGN distribution reduces to CSGT distribution that has this identifiability problem. Therefore, we decided, also, for the CSGGN distribution assume that $\sigma^2 = 1$, for simulations and in the rest of this work.

The results for all models can be found in Appendix E. It can be seen from the boxplots that as the sample size increases, the bias and variability of $\beta_0, \beta_1, \sigma^2$ and δ estimates, decrease, except for β_0 in the CSGT model, where we can observe an underestimation. For the CST and CSGT distributions, both bias and variability for ν estimates decrease as the sample size increase. For the other distributions $\boldsymbol{\nu}$ is either overestimated or underestimated. However, the respective bias decreases as the sample size increases, as shown by the $\sqrt{MSE}/\nu$ plots. For the non parametric curves, the variability of the estimates over the replicas decreases as the sample size increases, for all models.

2.5.9.2 Simulation study 2

To evaluate the proposed model on the robustness of the estimates against model misspecification, we generated only one Monte Carlo replica from: $Y_i = 2x_{1i} + sin(t_i^2 - t_i + 2) + \epsilon_i, i = 1, \ldots, 300$, where $x_{1i} \overset{iid}{\sim} U(0,1)$, $t_i \in (0,3)$, 500 iterations of the SAEM algorithm, $k = 2/3$ and the following scenarios, chosen in such way that we have situations of high asymmetry and high heavy tails: $\epsilon_i \overset{iid}{\sim} CST(0,1,0.8,3)$ and $\epsilon_i \overset{iid}{\sim} CSS(0,1,-0.8,2)$. Then, we fitted all models of the SMCSN and CSN distributions, for this simulated data.

From Figures 23 and 26, we can see that the nonparametric fitted curves are close to the actual ones. For the data set generated from the CST model, the regression parameters estimates are close to the actual ones, except for the CSN and CSS distributions. Also, the variance is clearly overestimated for the CSS model. For the model generated from the CSS we notice that the regression parameters and the estimated variances are close to the real values. These results give us a strong indication that a distribution within the SMCSN class may be better than the others in certain situations. However, we strongly recommended to fit and compare all developed models in this work, for a given data set.

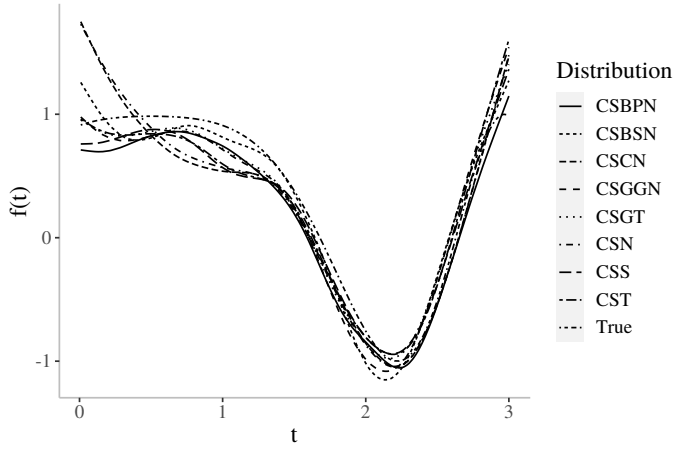


Figure 23 – Fitted and actual curves by distribution for the simulation study 2 generated by CST distribution.

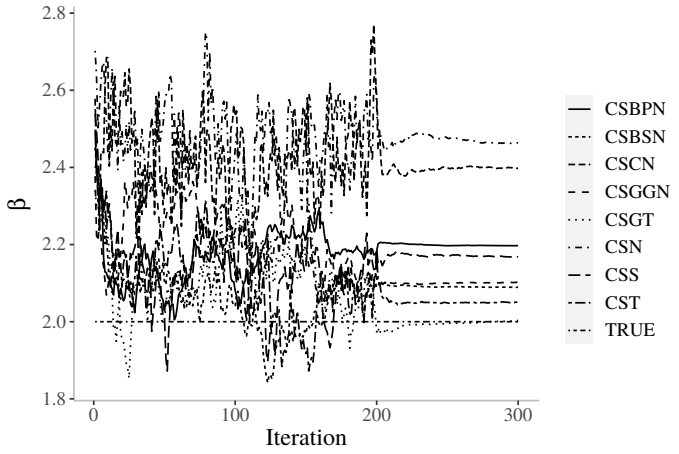


Figure 24 – $\hat{\beta}$ along the iterations of the SAEM algorithm for each fitted model - simulation study 2 with the data set generated by CST distribution.

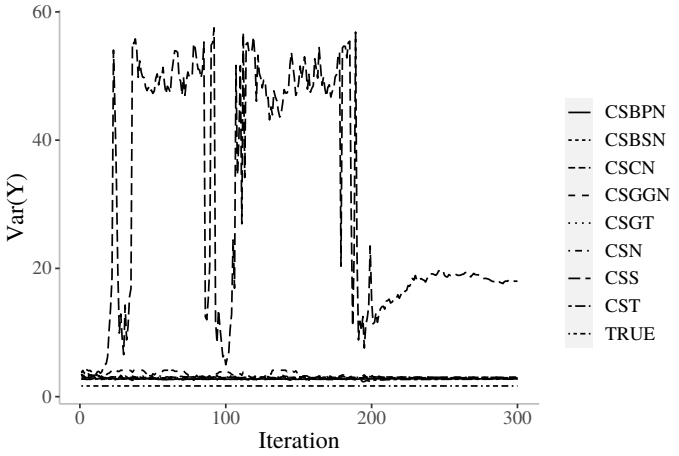


Figure 25 – $\hat{\mathrm{Var}}(Y)$ along the iterations of the SAEM algorithm for each fitted model - simulation study 2 with the data set generated by CST distribution.

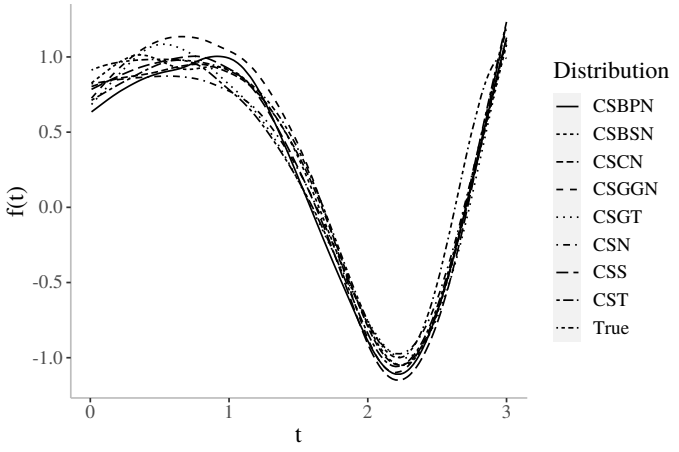


Figure 26 – Fitted and actual curves by distribution and the actual curve for the simulation study 2 generated by CSS distribution.

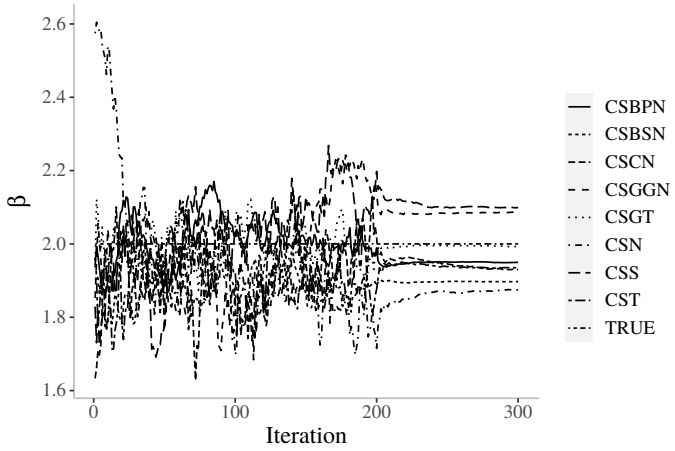


Figure 27 – $\hat{\beta}$ along the iterations of the SAEM algorithm for each fitted model - simulation study 2 with the data set generated by CSS distribution.

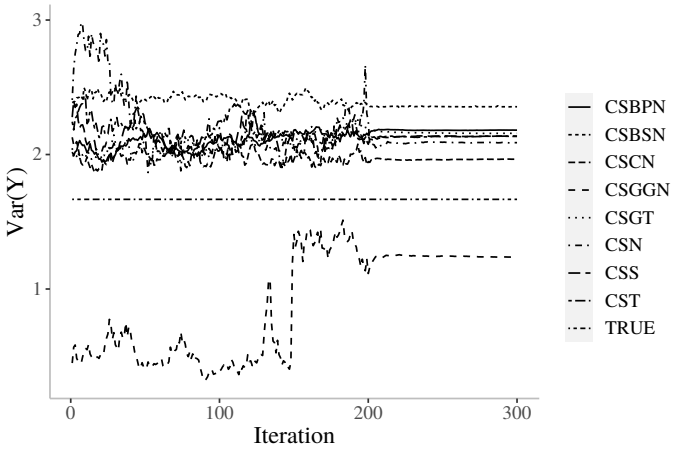


Figure 28 – $\hat{\mathrm{Var}}(Y)$ along the iterations of the SAEM algorithm for each fitted model - simulation study 2 with the data set generated by CSS distribution.

2.5.10 Application to Ragweed data

The data description can be found in Section 1. We fitted several models within the developed class, to explain the variability of the square root of the ragweed pollen concentration as a function of those mentioned environmental variables: $\sqrt{Y_i} = \beta_1 \texttt{rain}_i + \beta_2 \texttt{temperature}_i + \beta_3 \texttt{windspeed}_i + f(\texttt{days}_i) + \epsilon_i$, $i = 1, \ldots, n$, where y_i denotes observed ragweed pollen concentration in the ith day and ϵ_i are independent errors following some scale mixture of centered skew-normal distribution, including the Centered skew-normal distribution, for comparative purposes. We use 67 knots equally spaced which to be an adequate number to represent the non-parametric curves, in the sense that the inclusion of one more knot would not positively impact the goodness-of-fit measures. We used 500 iterations of the SAEM algorithm and $k = 1/2$, leading to the convergence of all parameters. Some inferential results and information criteria are given in Table 2. All models lead to similar conclusions, with all the effects being significant (at a significance level of 0.10) except for the CSBSN model, which indicates no significance for the covariate rain. The information criteria indicate that the Centered skew Birnbaum-Saunders Normal presents the best fit.

Figure 29 presents quantile-quantile plots with confidence bands of 95% for the residuals presented in Equation (2.17). We can notice that the Centered skew-normal and Centered skew generalized Gamma Normal presents many observations outside of the confidence bands, whereas the others show a good fit, mainly the CST and CSBPN models. From Figure 30, we have the confidence bands for each point of the nonparametric fitted curve between days in season and pollen ragweed. We may notice that the CSBSN model estimate a different shape in its curve in the end of season. Given all the above results, we selected the CSBSN model.

Since under the CSBSN model the covariate rain was not significant, a respective reduced model was fitted. The results are presented in Table 3. The information criteria still indicate that the CSBSN model is the best when compared to the models in the Table 2. Also, no significant changes are observed in the nonparametric fit as well as the envelope plot (both still indicating a well model fit), as can bee seen in Figure 31. In addition, in Table 2 we added confidence intervals for the parameters and a nullity test for the skewness parameter, and the most interesting thing to note is that the final model estimated a value for the skewness parameter very high and significantly different from 0, this information and the ν values reinforce the use of an asymmetric model of heavy tails.

In Figure 32 we have the generalized Cook's Distance and the Diagonal of the generalized Leverage matrix for the selected model. They suggest the absence of aberrant points. From Figure 33 we see no evidence of the presence of influential observations. These two results stress the well model fit to the data. In Appendix E We

Table 2 – Estimates, Standard errors (SE), p-values and results for the parameters of ragweed levels model.

Model	Parameter	Est.	SE	p-value (Wald)
	β_0	1.3660	0.5075	0.0071
	β_1	0.0773	0.0268	0.0039
	β_2	0.2226	0.0484	< 0.0001
	σ^2	4.5758	0.7139	-
CSN	δ	0.8768	0.0749	-
	α	210.6303	-	-
	$df(\alpha)$	10.7114	-	-
	AIC	1494.403	-	-
	BIC	1554.328	-	-
	AICc	1464.630	-	-
	HQIC	1518.293	-	-
	CAIC	1570.040	-	-
	SABIC	1504.490	-	-

Model	Parameter	Est.	SE	p-value (Wald)
	β_0	1.4707	0.4454	0.0009
	β_1	0.0755	0.0199	0.0001
	β_2	0.1820	0.0406	< 0.0001
	σ^2	2.9001	0.0165	-
CST	δ	0.9901	0.0120	-
	ν	4.4098	-	-
	α	354.4412	-	-
	$df(\alpha)$	17.3291	-	-
	AIC	1502.844	-	-
	BIC	1591.824	-	-
	AICc	1459.838	-	-
	HQIC	1538.317	-	-
	CAIC	1615.153	-	-
	SABIC	1517.822	-	-

Model	Parameter	Est.	SE	p-value (Wald)
	β_0	0.8186	0.4244	0.054
	β_1	0.0821	0.0210	< 0.0001
	β_2	0.2235	0.0404	< 0.0001
	σ^2	1.9789	0.0105	-
CSS	δ	0.9174	0.0117	-
	ν	1.4696	-	-
	α	806.2733	-	-
	$df(\alpha)$	23.3605	-	-
	AIC	1536.495	-	-
	BIC	1648.480	-	-
	AICc	1483.626	-	-
	HQIC	1581.140	-	-
	CAIC	1677.841	-	-
	SABIC	1555.346	-	-

Model	Parameter	Est.	SE	p-value (Wald)
	β_0	1.1465	0.4403	0.0092
	β_1	0.0842	0.0197	< 0.0001
	β_2	0.1840	0.0395	< 0.0001
CSGT	δ	0.9915	0.0104	-
	ν_1	4.6615	-	-
	ν_2	15.0000	-	-
	α	1682.355	-	-
	$df(\alpha)$	39.4994	-	-
	AIC	1597.428	-	-
	BIC	1768.968	-	-
	AICc	1519.096	-	-
	HQIC	1664.613	-	-
	CAIC	1814.468	-	-
	SABIC	1624.640	-	-

Model	Parameter	Est.	SE	p-value (Wald)
	β_0	1.3305	0.5570	0.0170
	β_1	0.0836	0.0288	0.0037
	β_2	0.2362	0.0515	< 0.0001
	σ^2	2.3866	0.0868	-
CSCN	δ	0.9380	0.0274	-
	ν_1	0.3877	-	-
	ν_2	0.2830	-	-
	α	556.6058	-	-
	$df(\alpha)$	19.0927	-	-
	AIC	1519.58	-	-
	BIC	1619.122	-	-
	AICc	1471.977	-	-
	HQIC	1559.264	-	-
	CAIC	1645.220	-	-
	SABIC	1536.335	-	-

Model	Parameter	Est.	SE	p-value (Wald)
	β_0	1.0777	0.3701	0.0036
	β_1	0.0777	0.0194	< 0.0001
	β_2	0.1949	0.0401	< 0.0001
	σ^2	2.7412	0.0166	-
CSBPN	δ	0.9889	0.0136	-
	ν_1	2.9095	-	-
	ν_2	3.5315	-	-
	α	600.0000	-	-
	$df(\alpha)$	13.7178	-	-
	AIC	1505.157	-	-
	BIC	1584.177	-	-
	AICc	1466.594	-	-
	HQIC	1536.660	-	-
	CAIC	1604.895	-	-
	SABIC	1518.458	-	-

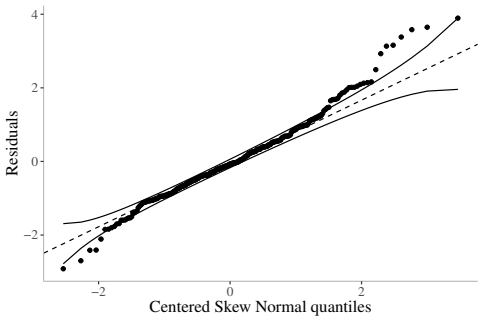
Model	Parameter	Est.	SE	p-value (Wald)
	β_0	0.3430	0.2379	0.15
	β_1	0.0586	0.0115	< 0.0001
	β_2	0.1543	0.0054	< 0.0001
	σ^2	21.5736	0.1202	-
CSBSN	δ	0.9930	0.0101	-
	ν_1	2.0000	-	-
	ν_2	9.1914	-	-
	α	70	-	-
	$df(\alpha)$	1.5771	-	-
	AIC	1467.285	-	-
	BIC	1499.999	-	-
	AICc	1450.635	-	-
	HQIC	1480.327	-	-
	CAIC	1508.576	-	-
	SABIC	1472.791	-	-

Model	Parameter	Est.	SE	p-value (Wald)
	β_0	1.0907	0.1760	< 0.0001
	β_1	0.0571	0.0079	< 0.0001
	β_2	0.1897	0.0169	< 0.0001
CSGGN	δ	0.9976	0.0030	-
	ν_1	1.5	-	-
	ν_2	1.5	-	-
	ν_3	1.5	-	-
	α	857.01	-	-
	$df(\alpha)$	25.6044	-	-
	AIC	1796.578	-	-
	BIC	1924.749	-	-
	AICc	1737.111	-	-
	HQIC	1847.676	-	-
	CAIC	1958.354	-	-
	SABIC	1818.153	-	-

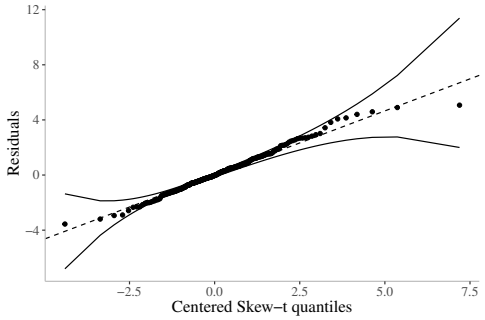
have the line plots showing the convergence of all parameters.

Table 3 – Estimates, Standard errors (SE), p-values and results for the parameters of ragweed levels CSBSN model without rain.

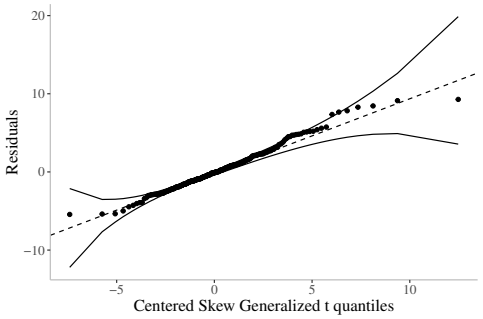
Parameter	Est.	SE	p-value (Wald)	CI (95%)
β_2	0.065	0.011	<0.001	(0.0434,0.0866)
β_3	0.164	0.025	<0.001	(0.115,0.213)
σ^2	23.075	0.039	-	(22.9986, 23.1514)
δ	0.998	0.003	<0.001	(0.9921,1.0000)
ν	(2.000, 9.644)	-	-	
α	70	-	-	
$df(\alpha)$	1.924	-	-	
AIC	1468.176	-	-	
BIC	1498.400	-	-	
AICc	1452.761	-	-	
HQIC	1480.225	-	-	
CAIC	1506.324	-	-	
SABIC	1473.263	-	-	



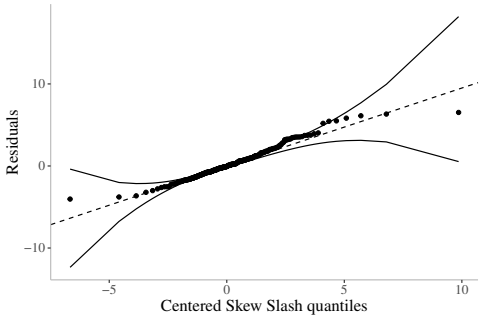
(a) Quantile-Quantile envelope for residuals of Additive partially linear Centered skew Normal model.



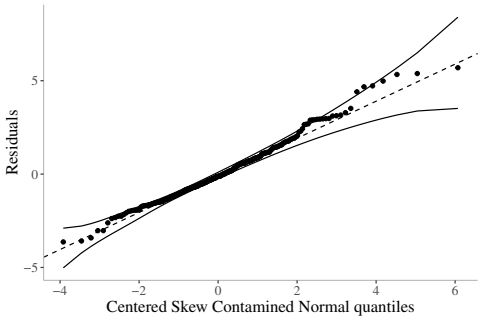
(b) Quantile-Quantile envelope for residuals of Additive partially linear Centered skew t model.



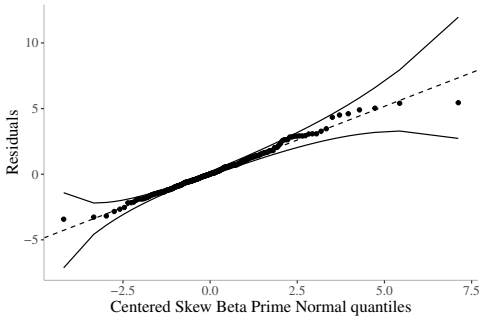
(c) Quantile-Quantile envelope for residuals of Additive partially linear Centered skew generalized t model.



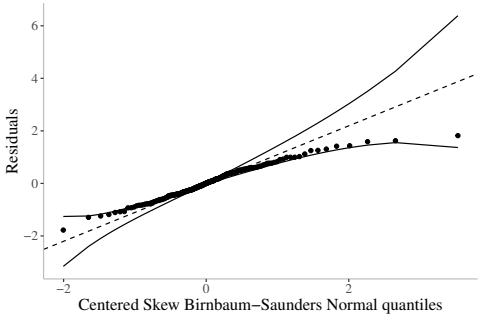
(d) Quantile-Quantile envelope for residuals of Additive partially linear Centered skew Slash model.



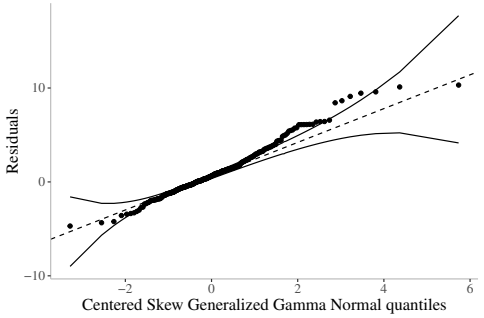
(e) Quantile-Quantile envelope for residuals of Additive partially linear Centered skew Contaminated Normal model.



(f) Quantile-Quantile envelope for residuals of Additive partially linear Centered skew Beta Prime Normal model.

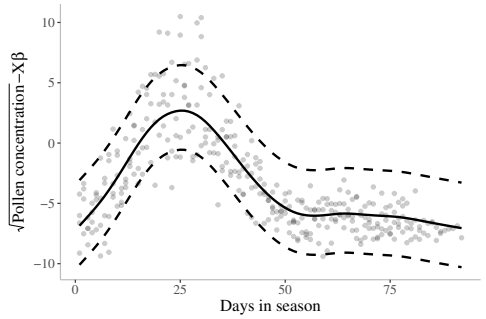


(g) Quantile-Quantile envelope for residuals of Additive partially linear Centered skew Birnbaum-Saunders Normal model.

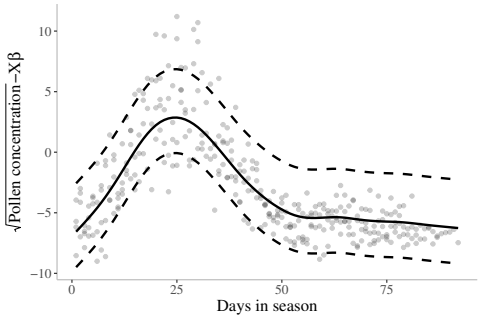


(h) Quantile-Quantile envelope for residuals of Additive partially linear Centered skew generalized Gamma Normal model.

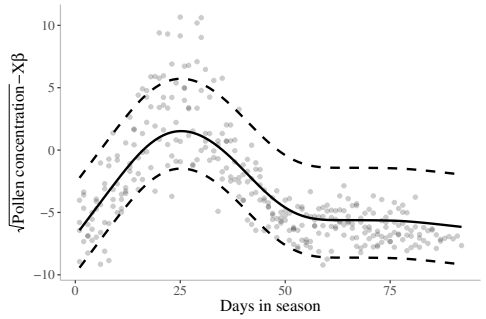
Figure 29 – Quantile-Quantile envelopes for fitted models to Ragweed data.



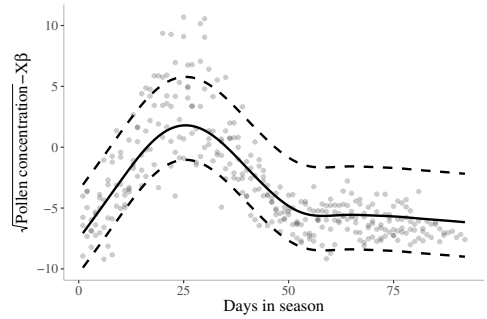
(a) 95% pointwise confidence bands for f(Days in season) of Additive partially linear Centered skew Normal model.



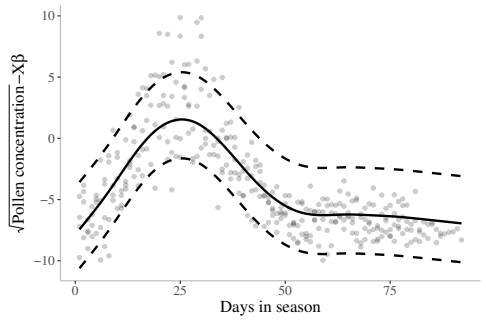
(b) 95% pointwise confidence bands for f(Days in season) of Additive partially linear Centered skew t model.



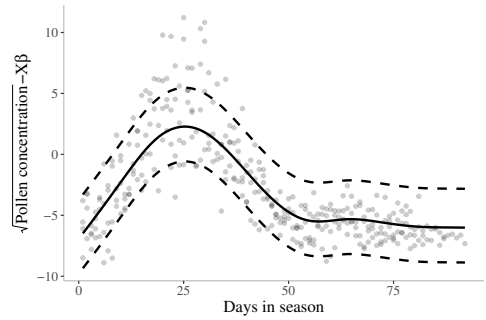
(c) 95% pointwise confidence bands for f(Days in season) of Additive partially linear Centered skew generalized t model.



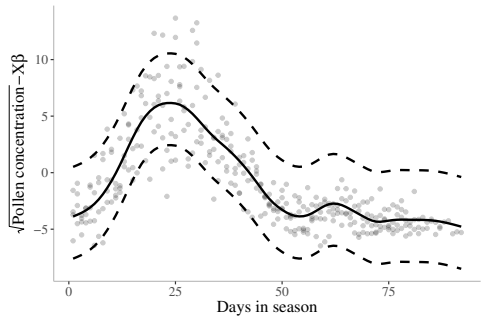
(d) 95% pointwise confidence bands for f(Days in season) of Additive partially linear Centered skew Slash model.



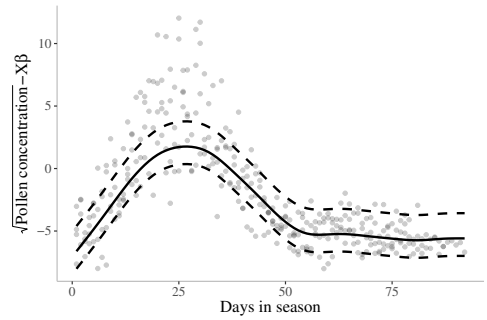
(e) 95% pointwise confidence bands for f(Days in season) of Additive partially linear Centered skew Contaminated Normal model.



(f) 95% pointwise confidence bands for f(Days in season) of Additive partially linear Centered skew Beta Prime Normal model.

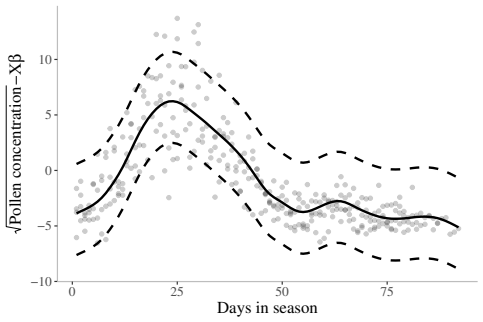


(g) 95% pointwise confidence bands for f(Days in season) of Additive partially linear Centered skew Birnbaum-Saunders Normal model.

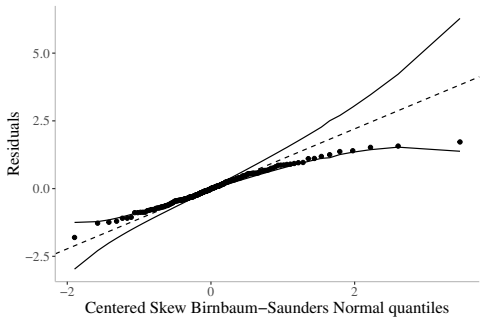


(h) 95% pointwise confidence bands for f(Days in season) of Additive partially linear Centered skew generalized Gamma Normal model.

Figure 30 – 95% pointwise confidence bands for f(Days in season) of fitted models.

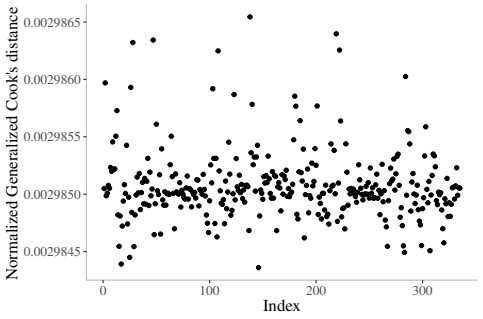


(a) 95% pointwise confidence bands for f(Days in season) of Additive partially linear Centered skew Birnbaum-Saunders Normal model.

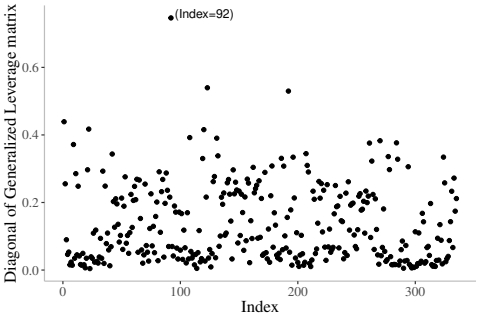


(b) Pointwise Quantile-Quantile envelope for residuals of Additive partially linear Centered skew Birnbaum-Saunders Normal model.

Figure 31 – 95% pointwise confidence bands for f(Days in season) and pointwise Quantile-Quantile envelope of CSBSN model.

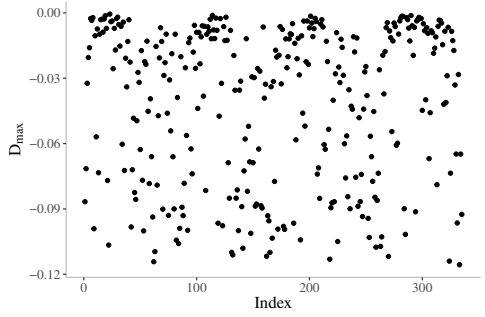


(a) Generalized Cook's Distance of Additive partially linear Centered skew Birnbaum-Saunders Normal model.

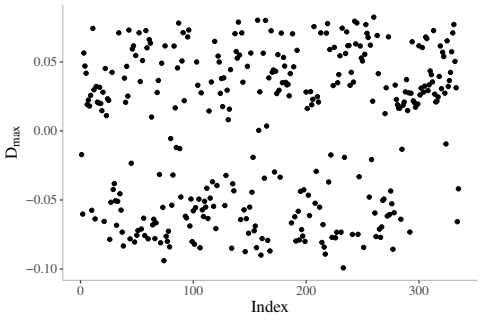


(b) Diagonal of generalized Leverage matrix for Additive partially linear Centered skew Birnbaum-Saunders Normal model.

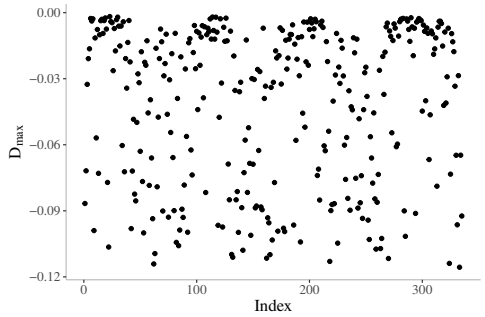
Figure 32 – Generalized Cook's Distance and Diagonal of generalized Leverage matrix for the selected model.



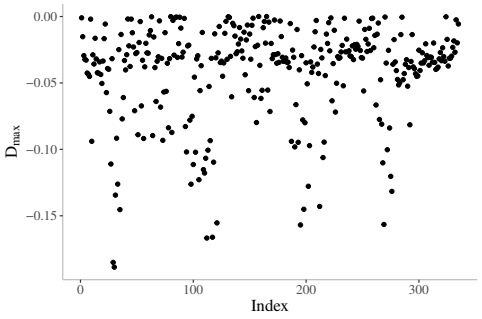
(a) $D_{\max}$ versus Index for case-weight perturbation.



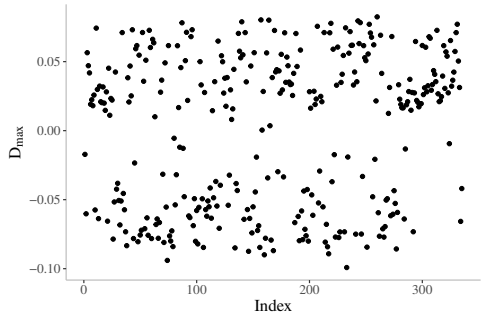
(b) $D_{\max}$ versus Index for response perturbation.



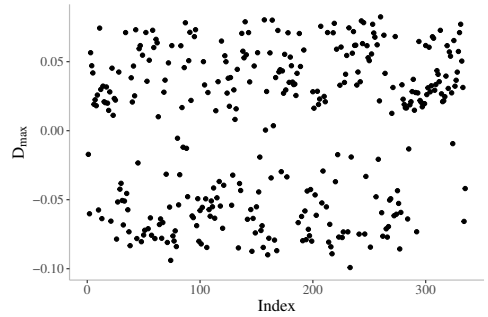
(c) $D_{\max}$ versus Index for scale perturbation.



(d) $D_{\max}$ versus Index for skewness perturbation.



(e) $D_{\max}$ versus Index for temperature perturbation.



(f) $D_{\max}$ versus Index for wind speed perturbation.

Figure 33 – Local influence for the selected model.

Chapter 3

Longitudinal data modeling using semi-parametric SMCSN model

3.1 Introduction

In many situations of fields of research it is common to carry out experiment where several observations of the outcome(s) of interest are made on the same experimental units, over the so-called conditions of evaluation. These situations are named repeated measurements. The respective data sets/methods of analysis are named repeated measurement data/analysis. When these evaluation conditions can not be mutually randomized, for example, when they correspond to time-points, we have the so-called longitudinal data. Longitudinal data studies are a powerful research strategy, since it is possible to characterize and evaluate global and individual changes over time, relating them to a set of covariates of interest (besides the time-points). Due to the longitudinal structure of the data it is expected to observe within-subject dependence. When we do not take into account such feature in a proper way, misleading inference can be obtained, as to underestimate or overestimate, for example. For more details on longitudinal studies, among others, see: Ware (1985), Diggle et al. (1994), Vonesh and Chinchilli (1996), Singer and Andrade (2000), Demidenko (2013) and Singer et al. (2017).

There are many challenges in the analysis of longitudinal data, as to handle with multivariate response and complex correlation structures. Therefore, it is important to consider appropriate techniques to handle all these features, properly. Indeed, there are a wide variety of techniques, many of them based on the so-called mixed models, as: Normal Linear Mixed Models (Henderson, 1953; Henderson et al., 1959), nonlinear Mixed Models (Lindstrom and Bates, 1990), generalized Linear Mixed Models (Breslow and Clayton, 1993), Semi-parametric Linear Mixed Models (Diggle and Zeger, 1994), skew Linear Mixed Models (Arellano-Vale et al., 2005) and Elliptical Linear Mixed

Models (Savalli et al., 2006). They are very important when there is an interesting in measuring the between-subjects heterogeneity and/or to account for extra (unknown) sources of variability.

Another approach is based on the so-called estimation functions (Godambe, 1991). They are particularly important when the interest is more focused on the population variations (than the individual ones) and/or in modeling the marginal response distributions. An estimation function depend on the data and the parameters (of interest). Also, under some regularity conditions they have good properties such as consistency and known asymptotic distribution. Liang and Zeger (1986), based on estimating functions, proposed the analysis of repeated measurement data using generalized Linear Models (GLM), introducing the Generalized Estimation Equations (GEE) estimation method. Based on some properties of the estimation functions and under some regularity conditions, these authors obtain consistent estimators for the regression parameters, provided that the marginal distributions are correctly specified.

The GEE methodology is easy to implement and very flexible, since it is only necessary to specify the regression structure, the correlation matrix and some characteristics of the marginal distributions (it is not required to specify the whole multivariate distribution). Under some regularity conditions (Sen and Singer, 1993), consistent estimators can be derived to the regression coefficients and its covariance matrix, even if that the response correlation matrix does not match the true underlying one. In addition, GEE may be preferred to the mixed models, when the interest lies on the marginal (population) characteristics. Besides the work of Liang and Zeger (1986), that focus on the exponential family of distributions, we can cite the GEE-based models for rates and proportions (Song et al., 2004; Freitas et al., 2021b), for positive data (Tsuyuguchi et al., 2020; Freitas et al., 2021a) and for count data (Kong et al., 2015; Sarvi et al., 2019), among others.

Recently, Manghi et al. (2019) proposed the use of generalized Additive Partially Linear Models with GEE, for modeling correlated data under linear and/or nonlinear relations of covariates with the response variable. There are few works related to this approach in the literature. Most of them under rely on the assumption of exponential family for the marginal distributions. To the best of our knowledge, there are no works considering SMCSN distributions (including the skew normal distribution).

The second part of this dissertation deals with the following contributions:

1. We developed a family of generalized additive partially linear models based on the scale mixtures of centered skew-normal distributions using GEE.

2. We developed appropriate estimation methods, information criteria for model comparison, residual analysis, global and local influence tools.

3. We performed simulation studies and real data set analysis. We show that our proposal is more flexible in relation to its competitors in exponential family.

3.2 Estimating functions

We say that a function $\boldsymbol{\psi}$ is an estimation function associated with the random vector $\boldsymbol{Y}$ and parameters of interest $\boldsymbol{\theta}$, if, for each $\boldsymbol{\theta} \in \Theta$, $\boldsymbol{\psi(\theta; Y)} = (\psi_1, ..., \psi_p)^\top$ is a random variable, where $\Theta \subseteq \mathbb{R}^p$ is the parameter space. In this work we will consider the regular cases, that is, Θ is compact with finite dimension (p), and the true parameter $\boldsymbol{\theta_0}$ is an interior point of Θ.

Assuming a random sample of n independent random vectors $\boldsymbol{Y}_i = (Y_{i1}, ..., Y_{it_i})^\top$, $i = 1, ..., n$ and that each sample vector is related to an estimation function, say $\boldsymbol{\psi}_i$, then a sample estimating function $\boldsymbol{\Psi}_n(\boldsymbol{\theta})$ is given by $\boldsymbol{\Psi}_n(\boldsymbol{Y}; \boldsymbol{\theta}) = \sum_{i=1}^{n} \boldsymbol{\psi}_i(\boldsymbol{Y}_i; \boldsymbol{\theta})$, where $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, ..., \boldsymbol{Y}_n^\top)^\top$. We also restrict our attention to estimation functions whose roots are estimators of the parameters of interest, i.e., $\boldsymbol{\Psi}_n(\boldsymbol{Y}; \hat{\boldsymbol{\theta}}) = \mathbf{0}$.

Let $Y_1, ..., Y_n$ be a random sample with $\mathbb{E}(Y_i|\boldsymbol{\theta}) = \mu_i(\boldsymbol{\theta})$, where μ_i doubly differentiable regarding $\boldsymbol{\theta}$ and Var$(y_i) = \sigma^2$. Then

$$\boldsymbol{\Psi}_n(\boldsymbol{y}; \boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} [y_i - \mu_i(\boldsymbol{\theta})] = \mathbf{0} \tag{3.1}$$

is an estimating equation.

On the other hand, $\Psi_n(\boldsymbol{\theta})$ is an unbiased function if $\mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{\Psi}_n(\boldsymbol{\theta})] = \mathbf{0}, \ \ \forall \boldsymbol{\theta} \in \Theta$. If all estimating functions $\boldsymbol{\psi}_i$ are unbiased, then $\boldsymbol{\Psi}_n$ will be also unbiased. Furthermore, let $\boldsymbol{\Psi}_n$ be an unbiased estimating function then, the related variability and sensibility matrices (both $p \times p$ and square), are given, respectively, by:

$$V_\Psi(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{\Psi}_n(\boldsymbol{\theta}) \boldsymbol{\Psi}_n^\top(\boldsymbol{\theta})], \ S_\Psi(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}} \left[\frac{\partial}{\partial \boldsymbol{\theta}^\top} \boldsymbol{\Psi}_n(\boldsymbol{\theta}) \right]. \tag{3.2}$$

Let (Ω,$\mathcal{A}$,$\mathcal{P}$) be a probability space, $\Omega \subset \mathbb{R}$ a sample space $\mathcal{P} = \{P_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p\}$, for some $p \in \mathbb{N}$. An estimating function $\boldsymbol{\Psi}_n(\boldsymbol{\theta}) : \Omega \times \Theta \to \mathbb{R}^p$ is said to be regular if $\forall \boldsymbol{\theta} \in \Theta$ and $i, j = 1, ..., p$,

1. $\boldsymbol{\Psi}_n(\boldsymbol{\theta})$ is an unbiased estimating function;

2. The derivative $\partial \boldsymbol{\Psi}_n(\boldsymbol{\theta}) / \partial \theta_i \ \exists$ and is almost sure continuous $\forall \boldsymbol{y} \in \Omega$;

3. It is possible to exchange the integration and derivation operators as follows:

$$\frac{\partial}{\partial \theta_i} \int_\Omega \boldsymbol{\Psi}_n(\boldsymbol{\theta}, \boldsymbol{y}) d\mathbb{P}_{\boldsymbol{\theta}} = \int_\Omega \frac{\partial}{\partial \theta_i} \left[\boldsymbol{\Psi}_n(\boldsymbol{\theta}, \boldsymbol{y}) \right] d\mathbb{P}_{\boldsymbol{\theta}}.$$

The fact of $\boldsymbol{\psi}(\boldsymbol{\theta})$ be integrable as a function of $\boldsymbol{y}$ for each θ_i, the propriety 2, and supposing that $\partial\boldsymbol{\psi}(\boldsymbol{\theta},\boldsymbol{y})/\partial\theta_i$ is **dominated by an integrable function**, guarantees this propriety;

4. $\mathbb{E}_{\boldsymbol{\theta}}[\boldsymbol{\Psi}_i(\boldsymbol{\theta})\boldsymbol{\Psi}_j(\boldsymbol{\theta})] \in \mathbb{R}$ and $V_{\Psi}(\boldsymbol{\theta})$ is positive definite;
5. $\mathbb{E}_{\boldsymbol{\theta}}\left[\dfrac{\partial}{\theta_l}\boldsymbol{\Psi}_i(\boldsymbol{\theta})\dfrac{\partial}{\theta_k}\boldsymbol{\Psi}_j(\boldsymbol{\theta})\right] \in \mathbb{R}$, where $l,k = 1,...,p$ and $S_{\Psi}(\boldsymbol{\theta})$ is non singular.

On the other hand, The Godambe Information Matrix of $\boldsymbol{\theta}$, associated to a regular estimating function $\boldsymbol{\Psi}_n$ is given by: $\boldsymbol{J}_{\boldsymbol{\Psi}}(\boldsymbol{\theta}) = \boldsymbol{S}_{\boldsymbol{\Psi}}^{\top}(\boldsymbol{\theta})\boldsymbol{V}_{\boldsymbol{\Psi}}^{-\mathbf{1}}(\boldsymbol{\theta})\boldsymbol{S}_{\boldsymbol{\Psi}}(\boldsymbol{\theta})$.

The Godambe information matrix plays a similar role to the Fisher information matrix, i.e., the former is related to the information about the variability of the estimators. Notice that if $\boldsymbol{S}_{\Psi}(\boldsymbol{\theta}) = -\boldsymbol{V}_{\Psi}(\boldsymbol{\theta})$, then the Godambe Information Matrix coincides with the Fisher Information Matrix.

Let $Q_i(\boldsymbol{\theta})$, $i = 1,...,n$ be non stochastic matrices and $u_i = u_i(y_i;\boldsymbol{\theta})$ zero mean vectors mutually independents. Then an estimating function class is said to be additive or linear if (Crowder, 1987): $\boldsymbol{\Psi_n(\theta)} = \sum_{i=1}^{n} Q_i(\boldsymbol{\theta})u_i(y_i;\boldsymbol{\theta})$.

A regular estimating function is said to be optimal if its associated estimators have minimal asymptotic variance. The element within class of linear estimating functions, according to Crowder (1987), is given by: $\boldsymbol{\Psi}_n^*(\boldsymbol{\theta}) = \sum_{i=1}^{n} Q_i^*(\boldsymbol{\theta})u_i(y_i;\boldsymbol{\theta})$, where

$$Q_i^*(\boldsymbol{\theta}) = \mathbb{E}\left(\frac{\partial u_i}{\partial \boldsymbol{\theta}^{\top}}\right)^{\top} \mathbf{Cov}(u_i)^{-1},$$

and $\mathbf{Cov}(u_i) = \text{diag}\{\mathbf{Var}(u_i)^{1/2}\}\boldsymbol{R}^v(u_i)\text{diag}\{\mathbf{Var}(u_i)^{1/2}\}$, being $\boldsymbol{R}^v(u_i)$ the correlation matrix of u_i, for $i = 1,\ldots,n$.

In the following we show the conditions that guarantee the asymptotic normality of the estimators obtained from the regular estimating functions.

Following Jørgensen and Labouriau (1994), let $\Psi : \Omega \times \Theta \to \mathbb{R}^p$ be a regular estimating function and $\{\hat{\theta}_n\}_{n\geqslant 1}$ a sequence of estimators of an estimating equations, and suppose that $\exists\ \theta \in \Theta$ such that $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$, where $\hat{\theta}_n$ is asymptotically Normal, i.e., $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} \mathcal{N}(0, \bar{J}_{\Psi}^{-1}(\boldsymbol{\theta}))$, where

$$\bar{J}_{\Psi}(\boldsymbol{\theta}) = \lim_{n\to\infty} \frac{1}{n}\{S_{\Psi}^{\top}(\boldsymbol{\theta})V_{\Psi}^{-1}(\boldsymbol{\theta})S_{\Psi}(\boldsymbol{\theta})\},$$

here the symbol "$\xrightarrow{D}$" stands for the convergence in distribution (related to $\mathbb{P}_{\theta}$) and "$\xrightarrow{\mathbb{P}}$" the convergence. The conditions that guarantee the above propriety are:

1. $y_i, i = 1,...,n$ are independent t_i-random vectors;

2. $\boldsymbol{\psi}_i(\boldsymbol{\theta}) = (\psi_{i1}, ..., \psi_{ip})^\top, i = 1, ..., n$, are regular estimating functions;

3. $\boldsymbol{\Psi}_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \boldsymbol{\psi}_i(\boldsymbol{\theta})$;

4. For $\delta > 0$:
$$\mathbb{E}_{\boldsymbol{\theta}} \left\{ \sup_{\boldsymbol{h}:||\boldsymbol{h}||\leqslant\delta} \left|\left| \frac{\partial}{\partial \boldsymbol{\theta}^\mathsf{T}} \boldsymbol{\psi}_i(\boldsymbol{\theta} + \boldsymbol{h}) - \frac{\partial}{\partial \boldsymbol{\theta}^\mathsf{T}} \boldsymbol{\psi}_i(\boldsymbol{\theta}) \right|\right| \right\} \xrightarrow{\mathbb{P}} \phi_\delta,$$
according to $n \to \infty$, $\phi_\delta \to \infty$ when $\delta \to \infty$ and $\phi_\delta \to 0$ when $\delta \to 0$;

5. As $n \to \infty$:
$$\frac{1}{n} \frac{\partial \boldsymbol{\Psi}_n}{\partial \boldsymbol{\theta}^\top}(\boldsymbol{\theta}) \xrightarrow{\mathbb{P}} \boldsymbol{S}_\Psi(\boldsymbol{\theta});$$

6. We have that
$$\frac{1}{n} \sum_{i=1}^{n} \text{Cov}(\boldsymbol{\psi}_i) \to V(\boldsymbol{\theta}) \text{ positive definite};$$

7. As $n \to \infty$:
$$\frac{\boldsymbol{\Psi}_n(\boldsymbol{\theta})}{\sqrt{n}} \xrightarrow{D} \mathcal{N}_p(0, \boldsymbol{V}_\Psi(\boldsymbol{\theta}));$$

8. $\hat{\boldsymbol{\theta}}_n$ is the solution of $\boldsymbol{\Psi}_n(\boldsymbol{w}) = \boldsymbol{0}$, $\boldsymbol{w} \in \Theta$;

In addition under conditions that guarantee the existence of a sequence of roots of $\boldsymbol{\Psi}_n(\boldsymbol{w})$ that are limited in probability, or restricted to a compact set almost certainly when $n \to \infty$, it comes that $\hat{\boldsymbol{\theta}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\theta}$ and $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{D} \mathcal{N}(0, \bar{\boldsymbol{J}}_\Psi^{-1}(\boldsymbol{\theta}))$. A proof for this theorem can be seen in Jørgensen and Labouriau (1994), for example.

In practice, the correlation matrix $\boldsymbol{R}^v$ is unknown. Liang and Zeger (1986) proposed to use a so-called working correlation matrix $\boldsymbol{R}_i(\boldsymbol{\rho})$, which depends on $\boldsymbol{\rho}$. We will use this approach with work correlation matrices.

3.3 GEE for Additive Partially Linear scale mixture of centered skew-normal regression models

Let $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{it_i})^\top$ be the individual profile of the ith experimental unit, $i = 1, \ldots, n$. Let us assume that the marginal density of Y_{ij} follows a SMCSN distribution, i.e., $Y_{ij} \overset{ind}{\sim}$ SMCSN$(\mu_{ij}, \sigma^2, \gamma, \boldsymbol{\nu})$, where σ^2, γ, $\boldsymbol{\nu}$ are assumed to be constant over the N observations, where $N = \sum_i t_i$. For modeling the marginal means (μ_{ij}) we assume that $f(\mu_{ij}) = \eta_{ij} = \boldsymbol{x}_{ij}^\top \boldsymbol{\beta} + \sum_{k=1}^{q} g_k(T_{ijk}) = \boldsymbol{x}_{ij}^\top \boldsymbol{\beta} + \sum_{k=1}^{q} \boldsymbol{\kappa}_k \boldsymbol{b}(T_{ijk})$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a vector of regression parameters $(p < n)$, $g_j(T_{ijk}) = \sum_{l=1}^{k_k^*} \kappa_{lk} b_l(x_{ijkl})$, b_l is a cubic

B-spline, $l = 1, \ldots, k_k^*$, $k = 1, \ldots, q$, $f(\cdot)$ is called link function, which is assumed to be monotonous and at least doubly differentiable, $\boldsymbol{x}_{ij} = (x_{ij1}, \ldots, x_{ijp})^\top$ and $\boldsymbol{T}_k = (T_{11k}, \ldots, T_{Nk})$ are fixed and known covariate vectors related to the jth observation of the ith experimental unit.

For developing the GEE approach of interest, let us $\boldsymbol{u}_i = \boldsymbol{u}_i(\boldsymbol{y}_i, \boldsymbol{\beta}) = (u_{i1}, \ldots, u_{it_i})^\top, i = 1, \ldots, n$ a zero mean vector mutually independent, within the context of Additive Partially Linear SMCSN regression models for repeated measurement we propose to use:

$$\boldsymbol{u}_i = \frac{1}{\tau}(\boldsymbol{y}_i - \boldsymbol{\mu}_{ci}) = \frac{(1 - b^2\delta^2)}{\sigma^2(1 - \delta^2)}(\boldsymbol{y}_i - \boldsymbol{\mu}_{ci}), \tag{3.3}$$

with $i = 1, \ldots, n$. Due to the good results obtained from the developed approach for the independent data, we considered the Score Vector related to the developed EM algorithm, properly transformed, as propose for $\boldsymbol{u}_i$. Also, we have that $E(u) = (1/\tau)(E(\boldsymbol{Y}_i) - \boldsymbol{\mu}_{ci}) = 0$ and $\mathrm{Var}(u_{\Psi ij}) = (1/\tau^2)\mathrm{Var}(\boldsymbol{Y}_i)$.

The idea, based on Liang and Zeger (1986), is to consider a suitable working correlation matrix, say $\boldsymbol{R}(\boldsymbol{\rho})$, in $\mathrm{Var}(\boldsymbol{U}_i)$, for modeling the within-subject dependency, i.e., $\mathrm{Cov}(\boldsymbol{u}_i) = \boldsymbol{\Sigma}_i = \mathrm{Var}(\boldsymbol{u}_i)^{1/2} R(\boldsymbol{\rho}) \mathrm{Var}(\boldsymbol{u}_i)^{1/2}$ where $\boldsymbol{\rho}$ is a vector of correlation parameters of $\boldsymbol{u}_i$. Therefore, we can define a penalized GEE (Manghi et al., 2019) for $\boldsymbol{\xi} = (\boldsymbol{\beta}^\top, \boldsymbol{\kappa}^\top)^\top$, as we did for log-likelihood in independent case, given by:

$$\begin{aligned}\boldsymbol{\Psi}(\boldsymbol{\xi}) &= \sum_{i=1}^{n} \left[-E\left(\frac{\partial \boldsymbol{u}_i}{\partial \boldsymbol{\xi}^\top} \right)^\top \right] \boldsymbol{\Sigma}_i^{-1} \boldsymbol{u}_i - \boldsymbol{P}(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \boldsymbol{M}_i^\top \boldsymbol{\Lambda}_i \boldsymbol{\Sigma}_i^{-1} \boldsymbol{u}_{\Psi i} - \boldsymbol{P}(\boldsymbol{\alpha}) \\ &= \sum_{i=1}^{n} \boldsymbol{M}_i^\top \boldsymbol{W}_i \boldsymbol{\Lambda}_i^{-1} \boldsymbol{u}_{\Psi i} - \boldsymbol{P}(\boldsymbol{\alpha}),\end{aligned}$$

where $\boldsymbol{M}_i = (\boldsymbol{x}_i, \boldsymbol{N}_{i1}, \ldots, \boldsymbol{N}_{iq})$, $\boldsymbol{W}_i = \boldsymbol{\Lambda}_i^\top \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\Lambda}_i$, $\boldsymbol{\Lambda}_i = E(\partial \boldsymbol{u}_i / \partial \boldsymbol{\eta}_i) = \mathrm{diag}(a_{i1}, \ldots, a_{it_i})$, $a_{ij} = (1/\tau)\partial \mu_{ij} / \partial \eta_{ij}$, $i = 1, \ldots, n$, and $\boldsymbol{P}(\boldsymbol{\alpha}) = (\boldsymbol{0}_p^\top, \alpha_1 \boldsymbol{\kappa}_1^\top \boldsymbol{\Omega}_1, \ldots, \alpha_q, \boldsymbol{\kappa}_q^\top \boldsymbol{\Omega}_q)$.

The estimators $\hat{\boldsymbol{\xi}}$ of $\boldsymbol{\xi}$ are obtained by solving $\boldsymbol{\Psi}(\boldsymbol{\xi}) = \boldsymbol{0}$, for this we need to calculate the related sensibility and variability matrices, which are also useful for obtaining the related standard errors. Indeed, the sensibility matrix for $\Psi(\Xi)$ is given by

$$\boldsymbol{S}_{\boldsymbol{\xi}} = E\left(\frac{\partial}{\partial \boldsymbol{\xi}^\top} \Psi(\boldsymbol{\xi}) \right) = -\sum_{i=1}^{n} \boldsymbol{M}_i^\top \boldsymbol{W}_i \boldsymbol{M} - \boldsymbol{M}(\boldsymbol{\alpha}),$$

where $\boldsymbol{M}(\boldsymbol{\alpha}) = \mathrm{diag}(\boldsymbol{0}_{pp}, \alpha_1 \boldsymbol{\Omega}_1, \ldots, \alpha_q \boldsymbol{\Omega}_q)$. On the other hand, the variability matrix for $\Psi(\Xi)$ is given by $\boldsymbol{V}_{\boldsymbol{\xi}} = \mathrm{Cov}\{\Psi(\boldsymbol{\xi})\Psi(\boldsymbol{\xi}^\mathsf{T})\} = \sum_{i=1}^{n} \boldsymbol{M}_i^\top \boldsymbol{W}_i \boldsymbol{\Lambda}_i^{-1} \mathrm{Cov}(\boldsymbol{u}_i) \boldsymbol{\Lambda}_i^{-1} \boldsymbol{W}_i^\top \boldsymbol{M}_i$.

We have that the asymptotic distribution of $\hat{\boldsymbol{\xi}}$ is given by (Jørgensen and Knudsen, 2004; Godambe, 1991): $\hat{\boldsymbol{\xi}} \sim \mathcal{N}\left(\boldsymbol{\xi}, \boldsymbol{S}_{\boldsymbol{\xi}}^{-1} \boldsymbol{V}_{\boldsymbol{\xi}} \boldsymbol{S}_{\boldsymbol{\xi}}^{-\top}\right)$, where $(.)^{-\top}$ stands for the inverse of the transpose of a given matrix. Furthermore, this result is valid when

$t_i \to \infty, \forall i$, $n \to \infty$ or both. The sandwich variance covariance matrix estimator of $\widehat{\boldsymbol{\xi}}$ (this is referred to as "robust estimator" in the literature) is given (Liang and Zeger, 1986) by:

$$\mathrm{Var}(\widehat{\boldsymbol{\xi}}) = \boldsymbol{S}_{\xi}^{-1}\left(\sum_{i=1}^{n}\boldsymbol{M}_i^{\top}\boldsymbol{W}_i\boldsymbol{\Lambda}_i^{-1}\boldsymbol{u}_i^{\top}\boldsymbol{u}_i\boldsymbol{\Lambda}_i^{-1}\boldsymbol{W}_i^{\top}\boldsymbol{M}_i\right)\boldsymbol{S}_{\xi}^{-1}.$$

Using the Gauss-Seidel method (see Hastie and Tibshirani (1990) and Manghi et al. (2019)), a current estimate of $\boldsymbol{\xi}$ at $(k+1)$-th iteration is updated, given $(\hat{\sigma}^{2(k)}, \hat{\gamma}^{(k)}, \hat{\boldsymbol{\nu}}^{(k)})$, by:

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \left(\sum_{i=1}^{n}\boldsymbol{x}_i^{\top}\boldsymbol{W}_i^{(k)}\boldsymbol{x}_i\right)^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i^{\top}\boldsymbol{W}_i^{(k)}\left\{\boldsymbol{z}_i^{(k)} - \sum_{s>0}\boldsymbol{N}_{is}\boldsymbol{\kappa}_s^{(k+1)}\right\}$$

$$\hat{\boldsymbol{\kappa}}_m^{(k+1)} = \left(\sum_{i=1}^{n}\boldsymbol{T}_i^{\top}\boldsymbol{W}_i^{(k)}\boldsymbol{T}_i + \alpha_m\Omega_m\right)^{-1}\sum_{i=1}^{n}\boldsymbol{T}_i^{\top}\boldsymbol{W}_i^{(k)}\left\{\boldsymbol{z}_i^{(k)} - \sum_{s\neq m}\boldsymbol{N}_{is}\boldsymbol{\kappa}_s^{(k+1)} - \sum_{i=1}^{n}\boldsymbol{x}_i^{\top}\boldsymbol{\beta}\right\},$$

$m = 1, \ldots, q$, where $\boldsymbol{z}_i = \boldsymbol{\eta}_i + \boldsymbol{\Lambda}_i^{-1}\boldsymbol{u}_i$. On the other hand, we can update $(\hat{\boldsymbol{\sigma}}^{2^{(k+1)}}, \hat{\boldsymbol{\gamma}}^{(k+1)}, \hat{\boldsymbol{\nu}}^{(k+1)})$ by maximizing the log-likelihood, given $\hat{\boldsymbol{\mu}}^{(k+1)}$, associated to a SMCSN distribution. Also, in order to have a faster algorithm, we estimate $\boldsymbol{\varphi} = 1/\boldsymbol{\nu}$ instead $\boldsymbol{\nu}$, since that $\boldsymbol{\varphi} \in (0,1)$ the respective algorithm is speeded up due to its parametric space being reduced. Thus, we have:

$$(\hat{\boldsymbol{\sigma}}^{2^{(k+1)}}, \hat{\boldsymbol{\gamma}}^{(k+1)}, \hat{\boldsymbol{\varphi}}^{(k+1)}) = \arg\max_{(\sigma^2,\gamma,\boldsymbol{\varphi})}\left\{\sum_{i=1}^{n}\log\left[f_{\mathsf{SMCSN}}(y_i|\hat{\mu}_i^{(k+1)}, \sigma^2, \gamma, \boldsymbol{\varphi})\right]\right\}, \quad (3.4)$$

$$\hat{\boldsymbol{\nu}}^{(k+1)} = 1/\boldsymbol{\varphi}, s \quad (3.5)$$

For the correlation parameters ($\boldsymbol{\rho}$) we can use consistent estimators based on the method of the moments using the $Cov(\boldsymbol{u}_i)$ which can be $\sqrt{n}$-consistently estimated by $\sum_i \boldsymbol{u}_i\boldsymbol{u}_i^{\top}/n$. In the following, we present the specific estimators for some structured working correlation matrices commonly considered.

Unstructured

In this case we have $t(t-1)/2$ parameters to be estimated. Let $\rho_{jj'}$ be the (j, j') element of $\boldsymbol{\Omega}$, for $j \neq j'$, which may be estimated by

$$\hat{\rho}_{jj'} = \frac{\sum_{i=1}^{n} u_{ij}u_{ij'}}{\sqrt{\sum_{i=1}^{n} u_{ij}^2}\sqrt{\sum_{i=1}^{n} u_{ij'}^2}}.$$

Exchangeable

Here the diagonal elements of $\boldsymbol{\Omega}$ are 1 and the others are ρ, i.e., it is assumed that the correlation between any two observations of the same individual is always the same. Thus ρ can be estimated by

$$\hat{\rho} = \frac{\sum_{i=1}^{n}\sum_{j>j'} u_{ij}u_{ij'}}{\sum_{j=1}^{t_i} u_{ij}^2}\frac{N}{\sum_{i=1}^{n}\frac{t_i(t_i-1)}{2}}.$$

First-order autoregressive (AR-1)

In this case the diagonal elements of $\boldsymbol{\Omega}$ are 1 and for the jth line and j'th column are $\rho^{|j-j'|}$, for $j \neq j'$. That is, we assume that the within-subject correlation decreases as the distance between the time-points increases. In this case ρ can be estimated by

$$\hat{\rho} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{t_i-1} u_{ij}u_{i(j+1)}}{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{t_i-1} u_{ij}^2 \sum_{i=1}^{n}\sum_{j=2}^{t_i} u_{ij}^2}}.$$

3.3.1 Effective degrees of freedom related to the nonparametric components

In context of GEE and generalized Additive Partially Linear models, following the idea of Manghi et al. (2019), the degrees of freedom can be derived from the solution of linear predictors given by

$$\hat{\boldsymbol{\eta}} = \boldsymbol{M}\hat{\boldsymbol{\xi}} = \hat{\boldsymbol{S}}_\alpha \hat{\boldsymbol{z}} = \boldsymbol{M}(\boldsymbol{M}^\top \hat{\boldsymbol{W}}\boldsymbol{M} + \boldsymbol{M}(\boldsymbol{\alpha}))\boldsymbol{M}^\top \hat{\boldsymbol{W}}\hat{\boldsymbol{z}}, \tag{3.6}$$

where $\hat{\boldsymbol{S}}_\alpha$ may be interpreted as a projection matrix or smoother matrix (Manghi et al., 2019) and the effective degrees of freedom are given by $df(\boldsymbol{\alpha}) = \text{tr}(\hat{\boldsymbol{S}}_\alpha)$ and $\boldsymbol{M} = \text{diag}(\boldsymbol{M}_1, \ldots, \boldsymbol{M}_n)$, $\boldsymbol{W} = \text{diag}(\boldsymbol{W}_1, \ldots, \boldsymbol{W}_n)$.

Therefore, we can consider information criteria for model selection based on some of the quasi-likelihood measure for goodness of fit. For the SMCSN family, the quasi-likelihoods is given by

$$Q(\boldsymbol{\mu}, \boldsymbol{y}) = \sum_{i=1}^{n}\sum_{j=1}^{t_i}\int_{y_{ij}}^{\mu_{ij}} \frac{y_{ij}-a}{\text{Var(Y)}}da = -\sum_{i=1}^{n}\frac{(y_{ij}-\mu_{ij})^2}{2\text{Var(Y)}}.$$

Then, we can define the quasi Akaike Information Criterium (QAIC) and quasi Bayesian Information Criterium (QBIC) given by:

$$QAIC = 2(p + df(\boldsymbol{\alpha}) + q) - 2Q(\boldsymbol{\mu}, \boldsymbol{y})$$
$$QBIC = \log(N)(p + df(\boldsymbol{\alpha}) + q) - 2Q(\boldsymbol{\mu}, \boldsymbol{y}),$$

These measures can be used to select the best model and the smoothness parameter α, since the lower the value the better is the model fit.

3.4 Simulation study

In this section we present a simulation study to analyze some asymptotic proprieties of the proposed estimators. We generate 100 Monte Carlo replicas based

on a given distribution of scale mixtures of centered skew-normal distribution (error distribution, see below) considering $\mu_{ij} = 2x_{ij} + cos(tt_{ij}),\ i = 1,\ldots,n, j = 1,\ldots,t$, where $x_{ij} \overset{iid}{\sim} U(0,1)$ and $tt_{ij} \in (0, 3\pi)$, $i = 1,\ldots,n$, $j = 1,\ldots,t$ and a suitable working correlation matrix. Due to the within-experimental unity dependence structure, the simulated responses were obtained via Student-t copulas (Demarta and McNeil, 2005). The covariate values were kept constant during along the replicas. We also considered $n = 10, 50$, $t = 3, 10$ and $\rho = 0.3, 0.8$. Also, we set $\sigma^2 = 1$, $\alpha = 10$ and considered: CSN($\gamma = 0.8$), CSGT($\gamma = -0.8, \nu_1 = 15, \nu_2 = 5$), CST($\gamma = 0.8, \nu = 5$), CSCN($\gamma = -0.8, \nu_1 = 0.5, \nu_2 = 0.5$), CSS($\gamma = 0.8, \nu = 3$), CSBPN($\gamma = -0.8, \nu_1 = 3, \nu_2 = 3$), CSBSN($\gamma = 0.8, \nu_1 = 1, \nu_2 = 1$) and CSGGN($\gamma = -0.8, \nu_1 = 2, \nu_2 = 1, \nu_2 = 0.66$), as the error distribution. For each case the SMCSN GEE models were fitted using different correlation matrices (AR(1), exchangeable and unstructured) including the true working correlation matrix one.

The results for all scenarios can be found in Appendix E. In general we can notice that the results were similar to those presented for the independent data (see subsection 2.5.9). The "sample sizes" are presented as: (n, t). For the sample size (10,10), the β and σ^2 estimates under the unstructured correlation matrix present a high variability that impaired the visualization of the boxplots, so we decided to remove the (10,10) case from some plots.

It can be seen from boxplots and MSE plots that as the sample size increases, the bias and variability of $\beta_0, \beta_1, \sigma^2$, δ and $\boldsymbol{\nu}$ estimates decrease, except for the CSBSN distribution, where we can observe a bias in the estimates of σ^2, γ and ν_2. For the non parametric curves plots, the variability over the replicas decreases as the sample size increases along the subjects over the replicas, for all models. The only caveat is for the case under (10,10) where we can notice that for some replicas the shape of the estimated and actual nonparametric curves presented do not match, indicating that, for a low number of experimental units, regardless of the number of repeated measurements, we may have unreliable inferences and estimates for non-parametric curves.

3.5 Diagnostic analysis

Additionally to all issues mentioned for the diagnostic analysis for models for independent data, here we have the needing of checking the suitableness of the working correlation matrix too. Venezuela et al. (2007), based on developed tools for generalized Linear Models, developed some concepts of leverage analysis for the GEE setup, as well as case deletion (Cook's Distance) and residual analysis based on the proposals of Preisser and Qaqish (1996) and Tan et al. (1997). Venezuela et al. (2011) extended the idea of local influence (Cook, 1986) using the generalized local influence (Cadigan

and Farrell, 2002b). Later Manghi et al. (2019) presented diagnostic measures for GEE Additive Partially Linear Models.

Here, we adapt the measures developed by Manghi et al. (2019), for our approach.

3.5.1 Leverage analysis

From Equation (3.6) we have that the $\boldsymbol{S}_\alpha$ can be used to construct a leverage measure considering that:

$$\begin{aligned}\operatorname{tr}(\boldsymbol{S}_\alpha) &= \operatorname{tr}\left(\boldsymbol{M}(\boldsymbol{M}^\top\boldsymbol{W}\boldsymbol{M}+\boldsymbol{M}(\boldsymbol{\alpha}))^{-1}\boldsymbol{M}^\top\boldsymbol{W}\right)\\ &= \operatorname{tr}\left(\boldsymbol{W}^{1/2}\boldsymbol{M}(\boldsymbol{M}^\top\boldsymbol{W}\boldsymbol{M}+\boldsymbol{M}(\boldsymbol{\alpha}))^{-1}\boldsymbol{M}^\top\boldsymbol{W}^{1/2}\right)\\ &= \operatorname{tr}(\boldsymbol{H}(\boldsymbol{\alpha})).\end{aligned}$$

Then $\boldsymbol{H}(\boldsymbol{\alpha})$ plays a role of an orthogonal projection matrix of vectors in $\mathbb{R}^n$ in the subspace generated by the columns of the matrix $\boldsymbol{W}^{1/2}\boldsymbol{M}$. Then, a given observation is said to be leverage if it is an influential point with respect to the values of $\boldsymbol{W}^{1/2}\boldsymbol{M}$. The index plot of the values of the diagonal of $\boldsymbol{H}(\boldsymbol{\alpha})$ provide an easy identification of candidates to be influential observations.

3.5.2 Residual analysis

The ordinary residuals are given by (Venezuela et al., 2007; Manghi et al., 2019): $\hat{\boldsymbol{e}} = \hat{\boldsymbol{W}}^{1/2}(\hat{\boldsymbol{z}}-\hat{\boldsymbol{\eta}}) = [\boldsymbol{I}-\hat{\boldsymbol{H}}(\boldsymbol{\alpha})]\hat{\boldsymbol{W}}^{1/2}\hat{\boldsymbol{z}}$, which have zero mean with $\boldsymbol{e} = (\boldsymbol{e}_1^\top,\ldots,\boldsymbol{e}_n^\top)^\top$ and $\boldsymbol{e}_i = (e_{i1},\ldots,e_{it_i})^\top$. Next, we will talk about their variance and standardization.

3.5.2.1 Naive residual

Given that $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}+O(1)$ then $\hat{\boldsymbol{e}} = \boldsymbol{e}+O(1) = [\boldsymbol{I}-\boldsymbol{H}(\boldsymbol{\alpha})]\boldsymbol{W}^{1/2}\boldsymbol{z}+O(1)$, and it follows that $\operatorname{Cov}(\hat{\boldsymbol{e}}) = \operatorname{Cov}(\boldsymbol{e})+O(1) = [\boldsymbol{I}-\boldsymbol{H}(\boldsymbol{\alpha})]\boldsymbol{W}^{1/2}\boldsymbol{\Lambda}^{-1}\operatorname{Cov}(\boldsymbol{u})\boldsymbol{\Lambda}^{-1}\boldsymbol{W}^{1/2}[\boldsymbol{I}-\boldsymbol{H}(\boldsymbol{\alpha})]+O(1)$.

If we assume that $\boldsymbol{R}_i(\boldsymbol{\rho})$ is the true correlation matrix of $\boldsymbol{Y}_i, \forall i = 1,\ldots,n$, we have that:

$$\begin{aligned}\widehat{\operatorname{Cov}}(\hat{\boldsymbol{e}}) &= [\boldsymbol{I}-\hat{\boldsymbol{H}}(\boldsymbol{\alpha})]\hat{\boldsymbol{W}}^{1/2}\hat{\boldsymbol{\Lambda}}^{-1}[\widehat{\operatorname{Var}}(\boldsymbol{u})^{1/2}\boldsymbol{R}(\hat{\boldsymbol{\rho}})\widehat{\operatorname{Var}}(\boldsymbol{u})^{1/2}]\hat{\boldsymbol{\Lambda}}^{-1}\hat{\boldsymbol{W}}^{1/2}[\boldsymbol{I}-\hat{\boldsymbol{H}}(\boldsymbol{\alpha})]\\ &= [\boldsymbol{I}-\hat{\boldsymbol{H}}(\boldsymbol{\alpha})][\boldsymbol{I}-\hat{\boldsymbol{H}}(\boldsymbol{\alpha})].\end{aligned}$$

Therefore, the naive residuals are given by $\hat{r}_{ij}^N = \hat{e}_{ij}/(1-\hat{h}_{ijj})$, $i = 1,\ldots,n$ and $j = 1,\ldots,t_i$.

3.5.2.2 Robust residuals

In this case if the true covariance matrix of $\boldsymbol{u}_i$ is unknown we can use a robust estimator for this and we have that $\widehat{\mathrm{Cov}}(\hat{\boldsymbol{e}})^R = [\boldsymbol{I} - \hat{\boldsymbol{H}}(\boldsymbol{\alpha})]\hat{\boldsymbol{W}}^{1/2}\hat{\boldsymbol{\Lambda}}^{-1}\hat{\boldsymbol{u}}\hat{\boldsymbol{u}}^\top\hat{\boldsymbol{\Lambda}}^{-1}\hat{\boldsymbol{W}}^{1/2}[\boldsymbol{I} - \hat{\boldsymbol{H}}(\boldsymbol{\alpha})]$. Then, the robust residual is given by $\hat{r}_{ij}^R = \hat{e}_{ij}/\widehat{\mathrm{Cov}}(\hat{\boldsymbol{e}})_{ij}^R$, for $i = 1, \ldots, n$ and $j = 1, \ldots, t_i$.

Another usual approach is the quantile residual given, under the SMCSN distributions, by $r_{ij}^q = \Phi^{-1}(F_{SMCSN}(y_{ij}, \mu_{ij}, \sigma^2, \gamma, \boldsymbol{\nu}))$, where F_{SMCSN} denotes the related cdf of a distribution of SMCSN family of distributions. We performed simulations to evaluate the mean, standard deviation, skewness and excess of kurtosis of naive, robust and quantile residuals. We shall only outline the main conclusions . As expected, the quantile residuals distribution is very well approximated by the standard normal one: the mean, skewness and excess of kurtosis are closer to 0 and the standard deviation to 1. The mean and the standard deviation of the naive and robust residuals are close to zero and one, respectively, but they displays considerable skewness and excess of kurtosis. Therefore, we can analyze the quantile residuals in the usual way, that is, through histograms, box-plots, index plots, and fitted values plot. Also, QQplot with envelopes can be made quickly, since the respective confidence bands can be simulated from the standard normal distributions

3.5.3 Local influence

Venezuela et al. (2011) proposed a generalized displacement measure to any, which is given, in our case, by $\mathrm{FD}(\omega) = 2\left\{\mathcal{F}(\widehat{\boldsymbol{\xi}}) - \mathcal{F}(\widehat{\boldsymbol{\xi}}_\omega)\right\}$, where $\mathcal{F}$ is a function, doubly differentiable, such that the estimator of $\boldsymbol{\xi}$, denoted by $\widehat{\boldsymbol{\xi}}$, is the solution of

$$\frac{\partial \mathcal{F}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = \mathbf{0},$$

where $\widehat{\boldsymbol{\xi}}$ and $\widehat{\boldsymbol{\xi}}_\omega$ are the estimated value for the original and perturbed model, respectively, with $\mathrm{FD}(\omega) \geqslant 0$. Also $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)^\top$ is a perturbation vector, where m depends on the proposed perturbation scheme. Based on Cook (1986), the idea is to study the local behavior of $\mathrm{FD}(\omega)$ for any value of $\boldsymbol{\omega}$ in a neighborhood of $\boldsymbol{\omega}_0$, which represents the null perturbation vector, such that $\mathcal{F}(\widehat{\boldsymbol{\xi}}_{\boldsymbol{\omega}_0}) = \mathcal{F}(\widehat{\boldsymbol{\xi}}) \Rightarrow \mathrm{FD}(\omega_0) = 0$.

Essentially, Venezuela et al. (2011) generalized the proposal of Cadigan and Farrell (2002b) to the GEE framework, replacing the likelihood equations by the estimation functions, such that, for a given perturbed estimating equation $\boldsymbol{\Psi}(\hat{\boldsymbol{\xi}}_\omega|\boldsymbol{\omega}) = \mathbf{0}$, there is a null perturbation vector such that $\boldsymbol{\Psi}(\boldsymbol{\xi}_\omega|\boldsymbol{\omega}_0) = \boldsymbol{\Psi}(\boldsymbol{\xi})$. Venezuela et al. (2011) also proposed a local influence measure for the GEE approach given by the eigenvector $d_{\max}$, corresponding to the largest eigenvalue of the matrix $\widehat{\boldsymbol{B}}_G = -\boldsymbol{\Delta}\boldsymbol{S}^{-1}\boldsymbol{\Delta}$, where $\boldsymbol{\Delta} = \partial\boldsymbol{\Psi}_1(\boldsymbol{\xi}|\boldsymbol{\omega})/\partial\boldsymbol{\omega}^\top$, $\boldsymbol{S} = \partial\boldsymbol{\Psi}_1(\boldsymbol{\xi})/\partial\boldsymbol{\xi}^\top = \boldsymbol{M}^\top(\dot{\boldsymbol{\Lambda}}\boldsymbol{\Sigma}^{-1}\boldsymbol{u} - \boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1})\boldsymbol{M} - \boldsymbol{M}(\boldsymbol{\alpha})$,

$\dot{\boldsymbol{\Lambda}} = \text{diag}(\dot{a}_{ij}) = \text{diag}(\partial a_{ij}/\partial\eta_{ij})$, $\dot{a}_{ij} = (1/\tau)\partial^2\mu_{ij}/\partial\eta_{ij}^2$, which all are evaluated at $\boldsymbol{\xi} = \widehat{\boldsymbol{\xi}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. In this work we calculate the matrix $\boldsymbol{\Delta}$ for six different perturbation schemes based on the proposals of Venezuela et al. (2011) and Manghi et al. (2019), namely: case-weight perturbation, response perturbation, single-covariate perturbation, scale parameter perturbation, skewness parameter perturbation, shape parameters perturbation and working correlation matrix perturbation.

In general, flagged observations under the case-weight perturbation scheme can be interpreted as a perturbation in variance of each experimental unit (Venezuela et al., 2011). Perturbations in the response variable can be seen as an alternative way of identifying outliers (Schwarzmann, 1991b). The single-covariate perturbation scheme helps to evaluate the influence of each continous covariate in the estimating process. Perturbations in the scale, shape and skewness parameters are useful for checking the model sensitivity to the lacking of the homogeneity of these parameters, along the observations. Working correlation matrix perturbations could indicate, for example, the misspecification of such structure.

3.5.3.1 Case-weight perturbation

Let us consider the following perturbation scheme (Manghi et al., 2019): $\boldsymbol{\Psi}(\boldsymbol{\xi}|\boldsymbol{\omega}) = \boldsymbol{M}^\top\boldsymbol{W}\boldsymbol{\Lambda}^{-1}\text{diag}(\boldsymbol{\omega})\boldsymbol{u} - \boldsymbol{P}(\boldsymbol{\alpha})$, where $\boldsymbol{\omega} = (\boldsymbol{\omega}_1^\top, \ldots, \boldsymbol{\omega}_n^\top)^\top$, $\boldsymbol{\omega}_i = (\omega_{i1}, \ldots, \omega_{it_i})^\top$, $i = 1, \ldots, n$ and ω_0 is a vector of 1's. Therefore $\boldsymbol{\Delta} = \boldsymbol{M}^\top\boldsymbol{W}\boldsymbol{\Lambda}^{-1}\boldsymbol{u}$.

3.5.3.2 Response perturbation

Let us consider an additive perturbation scheme for the response variable Y_{ij} (Venezuela et al., 2011), $i = 1, \ldots, n$ and $j = 1, \ldots, t_i$ given by $y_{\omega ij} = y_{ij} + \omega_{ij}\sqrt{\text{Var}(y_{ij})}$, where the non-perturbation vector is $\omega_{ij} = 0$, i.e., $\boldsymbol{\omega}_0 = \boldsymbol{0}$. Thus, the perturbed estimating function is given by $\boldsymbol{\Psi}(\boldsymbol{\xi}|\boldsymbol{\omega}) = \boldsymbol{M}^\top\boldsymbol{W}\boldsymbol{\Lambda}^{-1}\dfrac{1}{\tau}(\boldsymbol{y}_\omega - \boldsymbol{\mu}) - \boldsymbol{P}(\boldsymbol{\alpha})$. In this case $\boldsymbol{\Delta} = \boldsymbol{M}^\top\boldsymbol{W}\boldsymbol{\Lambda}^{-1}\sqrt{\text{Var}(y_{ij})/\tau}$.

3.5.3.3 Explanatory variable perturbation

Based on Thomas and Cook (1989), we propose an additive perturbation scheme on the k-th column of the design matrix $\boldsymbol{X}$, that is, $\boldsymbol{x}_k = (x_{11k}, x_{12k}, \ldots, x_{Nk})^\top$, where each component of the perturbed vector $\boldsymbol{x}_{\boldsymbol{\omega}\boldsymbol{k}}$ is given by: $x_{\omega ijk} = x_{ijk} + \omega_{ij}s_{x_k}$, where s_{x_k} is the standard deviation of $\boldsymbol{x}_k$, where $i = 1, \ldots, n$ and $j = 1, \ldots, t_i$. Here, the non-perturbation vector is $\boldsymbol{\omega}_0 = \boldsymbol{0}$.

The perturbed estimating functions is given by: $\boldsymbol{\Psi}(\boldsymbol{\xi}|\boldsymbol{\omega}) = \boldsymbol{M}_\omega^\top\boldsymbol{\Lambda}_\omega\boldsymbol{\Sigma}^{-1}\boldsymbol{u}_\omega - \boldsymbol{P}(\boldsymbol{\alpha})$, where $\boldsymbol{M}_\omega = \boldsymbol{M} + B_{k\omega}$, $B_{k\omega}$ is a matrix with $\boldsymbol{\omega}s_{x_k}$ in the k-th column and zeros elsewhere, $\boldsymbol{\Lambda}_\omega = \text{diag}(\boldsymbol{a}_\omega)$, $a_{ij\omega} = (1/\tau)\partial\mu_{ij\omega}/\partial\eta_{ij\omega}$ and $\boldsymbol{u}_\omega = (1/\tau)(\boldsymbol{y} - \boldsymbol{\mu}_\omega)$. Noticing

that $\mu_{\omega_{ij}} = g^{-1}(\eta_{\omega_{ij}})$ and $\eta_{\omega_{ij}} = \beta_1 x_{1ij} + \cdots + \beta_k(x_{kij} + \omega_{ij}s_k) + \beta_p x_{pij} + \sum_{l=1}^{q} g_k(T_{ijl})$, we have that $\boldsymbol{\Delta}_{ij} = \dot{\boldsymbol{M}}_{\omega_{ij}}^{\top}\boldsymbol{\Lambda}_{\omega_{ij}}\boldsymbol{\Sigma}^{-1}\boldsymbol{u}_{\omega_{ij}} + \boldsymbol{M}_{\omega_{ij}}^{\top}\left[\dot{\boldsymbol{\Lambda}}_{\omega_{ij}}\boldsymbol{\Sigma}^{-1}\boldsymbol{u}_{\omega_{ij}} + \boldsymbol{\Lambda}_{\omega_{ij}}\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{u}}_{\omega_{ij}}\right]$, where $\dot{\boldsymbol{M}}_{\omega_{ij}} = s_{x_k}(\dot{\boldsymbol{M}}_{\omega_{1ij}}^{\top}, \ldots, \dot{\boldsymbol{M}}_{\omega_{nij}}^{\top})^{\top}$, $\dot{\boldsymbol{\Lambda}}_{\omega_{ij}} = \dfrac{\partial \boldsymbol{\Lambda}_{\omega_{ij}}}{\partial \omega_{ij}} = \dfrac{\beta_k s_{x_k}}{\tau}\mathsf{diag}\{0, \ldots, \partial^2\mu_{\omega_{ij}}/\partial\eta_{\omega_{ij}}^2, \ldots, 0\}$, $\dfrac{\partial \boldsymbol{u}_{\omega_{ij}}}{\partial \omega_{ij}} = -\dfrac{\beta_k s_{x_k}}{\tau}\mathsf{diag}\{0, \ldots, \partial\mu_{\omega_{ij}}/\partial\eta_{\omega_{ij}}, \ldots, 0\}$

3.5.3.4 Scale parameter perturbation

Let us consider a multiplicative perturbation scheme for the scale parameter σ^2 given by: $\sigma^2_{\omega ij} = \sigma^2/\omega_{ij}$, where $i = 1, \ldots, n$ and $j = 1, \ldots, t_i$. In this case, $\boldsymbol{\omega}_0 = 1$, the perturbed estimating function is given by: $\boldsymbol{\Psi}(\boldsymbol{\xi}|\boldsymbol{\omega}) = \boldsymbol{M}^{\top}\boldsymbol{\Lambda}_{\boldsymbol{\omega}}\boldsymbol{\Sigma}_{\boldsymbol{\omega}}^{-1}\boldsymbol{u}_{\boldsymbol{\omega}} - \boldsymbol{P}(\boldsymbol{\alpha})$, the $\boldsymbol{\Delta}$ matrix is given by:

$$\boldsymbol{\Delta} = \frac{\partial \boldsymbol{\Lambda}_{\boldsymbol{\omega}}}{\partial \boldsymbol{\omega}}\boldsymbol{\Sigma}_{\boldsymbol{\omega}}^{-1}\boldsymbol{u}_{\boldsymbol{\omega}} + \boldsymbol{\Lambda}_{\boldsymbol{\omega}}\left[\frac{\partial \boldsymbol{\Sigma}_{\omega_{ij}}^{-1}}{\partial \omega_{ij}}\boldsymbol{u}_{\boldsymbol{\omega}} + \boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{u}_{\omega_{ij}}}{\partial \omega_{ij}}\right],$$

where:

$$\begin{aligned}
\frac{\partial \boldsymbol{\Lambda}_{\omega_{ij}}}{\partial_{\omega_{ij}}} &= -\frac{\sigma^2}{\omega_{ij}^2}\tau^*(\partial\mu_{ij}/\partial\eta_{ij}),\ \tau^* = \frac{\partial\tau^{-1}}{\partial\sigma^2} = \frac{1-\delta^2}{1-b^2\delta^2},\ \frac{\partial \boldsymbol{\Sigma}_{\omega_{ij}}^{-1}}{\partial \omega_{ij}} = -\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}_{\omega_{ij}}}{\partial \omega_{ij}}\boldsymbol{\Sigma}^{-1},\\
\frac{\partial \boldsymbol{\Sigma}_{\omega_{ij}}}{\partial \omega_{ij}} &= \frac{\partial \mathsf{Var}(u_{ij})^{1/2}}{\partial \omega_{ij}}\boldsymbol{R}(\boldsymbol{\rho})\mathsf{Var}(u_{ij})^{1/2} + \frac{\partial \mathsf{Var}(u_{ij})^{1/2}}{\partial \omega_{ij}}\boldsymbol{R}(\boldsymbol{\rho})\mathsf{Var}(u_{ij})^{1/2},\\
\frac{\partial \mathsf{Var}(u_{ij})^{1/2}}{\partial \omega_{ij}} &= -\frac{\sigma^3_{\omega_{ij}}\sqrt{E(U^{-1})}}{\omega_{ij}^2}\frac{1-\delta^2}{1-b^2\delta^2},\ \frac{\partial \boldsymbol{u}_{\omega_{ij}}}{\omega_{ij}} = -\frac{\sigma^2}{\omega_{ij}^2}\tau^*(y_{ij} - \mu_{ij}).
\end{aligned}$$

3.5.3.5 Skewness parameter perturbation

To make the related calculations easier, we consider the following parameterization of the skewness parameter:

$$\delta = \frac{\lambda}{\sqrt{1+\lambda^2}},\ \lambda = \frac{s\gamma^{1/3}}{\sqrt{\frac{2}{\pi} + s^2\gamma^{2/3}\left(\frac{2}{\pi} - 1\right)}}.$$

Also, a multiplicative perturbation scheme is given by $\delta_{\omega ij} = \delta/\omega_{ij}$, where $i = 1, \ldots, n$ and $j = 1, \ldots, t_i$ and $\boldsymbol{\omega}_0 = 1$. The perturbed estimating function is given by: $\boldsymbol{\Psi}(\boldsymbol{\xi}|\boldsymbol{\omega}) = \boldsymbol{M}^{\top}\boldsymbol{\Lambda}_{\boldsymbol{\omega}}\boldsymbol{\Sigma}_{\boldsymbol{\omega}}^{-1}\boldsymbol{u}_{\boldsymbol{\omega}} - \boldsymbol{P}(\boldsymbol{\alpha})$, where

$$\boldsymbol{\Delta} = \frac{\partial \boldsymbol{\Lambda}_{\boldsymbol{\omega}}}{\partial \boldsymbol{\omega}}\boldsymbol{\Sigma}_{\boldsymbol{\omega}}^{-1}\boldsymbol{u}_{\boldsymbol{\omega}} + \boldsymbol{\Lambda}_{\boldsymbol{\omega}}\left[\frac{\partial \boldsymbol{\Sigma}_{\omega_{ij}}^{-1}}{\partial \omega_{ij}}\boldsymbol{u}_{\boldsymbol{\omega}} + \boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{u}_{\omega_{ij}}}{\partial \omega_{ij}}\right],$$

and

$$\frac{\partial \boldsymbol{\Lambda}_{\omega_{ij}}}{\partial_{\omega_{ij}}} = -\frac{\delta}{\omega_{ij}^2}\tau^{\dagger}(\partial\mu_{ij}/\partial\eta_{ij}),\ \tau^* = \frac{\partial\tau^{-1}}{\partial\delta} = -2\frac{\sigma^2\delta(1-b^2)}{(1-b^2\delta^2)^2},$$

$$\frac{\partial \boldsymbol{\Sigma}_{\omega_{ij}}^{-1}}{\partial \omega_{ij}} = -\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}_{\omega_{ij}}}{\partial \omega_{ij}}\boldsymbol{\Sigma}^{-1},$$
$$\frac{\partial \boldsymbol{\Sigma}_{\omega_{ij}}}{\partial \omega_{ij}} = \frac{\partial \text{Var}(u_{ij})^{1/2}}{\partial \omega_{ij}}\boldsymbol{R}(\boldsymbol{\rho})\text{Var}(u_{ij})^{1/2} + \frac{\partial \text{Var}(u_{ij})^{1/2}}{\partial \omega_{ij}}\boldsymbol{R}(\boldsymbol{\rho})\text{Var}(u_{ij})^{1/2},$$
$$\frac{\partial \text{Var}(u_{ij})^{1/2}}{\partial \omega_{ij}} = \frac{2\sigma^9\delta^2(1-\delta^2)^3(1-b^2)\sqrt{E(U^{-1})}}{\omega_{ij}^2(1-b^2\delta^2)^5}, \quad \frac{\partial \boldsymbol{u}_{\omega_{ij}}}{\omega_{ij}} = -\frac{\delta}{\omega_{ij}^2}\tau^*(y_{ij}-\mu_{ij}),$$

which all quantities are evaluated at $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ and $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

3.5.3.6 Shape parameter perturbation

Let us consider a multiplicative perturbation scheme for the kth shape parameter ν_k given by: $\nu_{k\omega ij} = \frac{\nu_k}{\omega_{ij}}$, where $i = 1, \ldots, n$, $j = 1, \ldots, t_i$, $k = 1$ for the CST and CSS distributions, $k = 1, 2$ for the CSGT, CSCN, CSBPN and CSBSN distributions, and $k = 1, 2, 3$ for the CSGGN distribution. In this case, $\boldsymbol{\omega}_0 = 1$ and the perturbed estimating function is given by: $\boldsymbol{\Psi}(\boldsymbol{\xi}|\boldsymbol{\omega}) = \boldsymbol{M}^\top\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\boldsymbol{\omega}}^{-1}\boldsymbol{u} - \boldsymbol{P}(\boldsymbol{\alpha})$, and

$$\boldsymbol{\Delta} = -\boldsymbol{M}^\top\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}_{\omega_{ij}}}{\partial \omega_{ij}}\boldsymbol{\Sigma}^{-1}\boldsymbol{u},$$

where

$$\frac{\partial \boldsymbol{\Sigma}_{\omega_{ij}}}{\partial \omega_{ij}} = \frac{\partial \text{Var}(u_{ij})^{1/2}}{\partial \omega_{ij}}\boldsymbol{R}(\boldsymbol{\rho})\text{Var}(u_{ij})^{1/2} + \frac{\partial \text{Var}(u_{ij})^{1/2}}{\partial \omega_{ij}}\boldsymbol{R}(\boldsymbol{\rho})\text{Var}(u_{ij})^{1/2},$$
$$\frac{\partial \text{Var}(u_{ij})^{1/2}}{\partial \omega_{ij}} = -\frac{\nu_{k\omega_{ij}}}{\omega_{ij}^2}\frac{1}{2\sqrt{Var(Y_{ij})}}\frac{\sigma^2}{\tau}\frac{\partial E(U^{-1})_{\omega_{ij}}}{\partial \nu_{k\omega_{ij}}},$$

which all quantities are evaluated at $\omega = \omega_0$ and $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

The derivatives $\partial E(U^{-1})/\nu_k$ are presented in Table 4, remembering that:

$$\dot{\Gamma}(x) = \frac{\partial \Gamma(x)}{\partial x} = \int_0^\infty t^{x-1}e^{-t}\log(t)dt, \quad \mathcal{R}(x) > 0,$$

where $\mathcal{R}(x)$ is the real part of x.

3.5.3.7 Working correlation matrix perturbation

Let $\boldsymbol{R}(\boldsymbol{\rho})$ be a working correlation matrix indexed by a $\binom{t_i}{2}$ dimensional vector $\boldsymbol{\rho} = \left(\rho_{12}, \ldots, \rho_{\binom{t_i}{2}}\right)^\top$. Since each experimental unit may have a specific working correlation matrix, Venezuela et al. (2011) proposed a related perturbation scheme given by $\rho_{\omega i(jl)} = \rho_{jl}/\omega_{i(jl)}$, where $i = 1, \ldots, n$, $j < l$ and $j, l = 1, \ldots, t_i$. For this perturbation scheme $\boldsymbol{\omega} = (\omega_{1(12)}, \ldots, \omega_{1((t_i-1)t_i)}, \ldots, \omega_{n(12)}, \ldots, \omega_{n((t_i-1)t_i)})^\top$ is a perturbation vector and $\boldsymbol{\omega}_0 = \mathbf{1}$. The perturbation estimating equation is given by: $\boldsymbol{\Psi}(\boldsymbol{\xi}|\boldsymbol{\omega}) = \boldsymbol{M}^\top\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\boldsymbol{\omega}}^{-1}\boldsymbol{u} - \boldsymbol{P}(\boldsymbol{\alpha})$.

Table 4 – Derivatives of $\partial E(U_i^{-1})/\partial\nu_k$ for each distribution of the SMCSN family.

Distribution	$\partial E(U_i^{-1})/\partial\nu_1$	$\partial E(U_i^{-1})/\partial\nu_2$	$\partial E(U_i^{-1})/\partial\nu_3$
CST	$-\frac{2}{(\nu-2)^2}$ $(\nu_1=\nu)$	-	-
CSS	$-\frac{1}{(\nu-1)^2}$ $(\nu_1=\nu)$	-	-
CSGT	$-\frac{\nu_2}{(\nu_1-2)^2}$	$\frac{1}{\nu_1-2}$	-
CSCN	$\frac{1-\nu_2}{\nu_2}$	$-\frac{\nu_1}{\nu_2^2}$	-
CSBPN	$-\frac{\nu_2}{(\nu_1-1)^2}$	$\frac{1}{\nu_1-1}$	-
CSBSN	$\frac{\nu_1}{\nu_2}$	$-\frac{\nu_1^2+2}{2\nu_2^2}$	-
CSGGN	$\frac{\dot\Gamma(\nu_1-1/\nu_2)\Gamma(\nu_1)-\Gamma(\nu_1-1/\nu_2)\dot\Gamma(\nu_1)}{\nu_3\Gamma(\nu_1)^2}$	$\frac{\dot\Gamma(\nu_1-1/\nu_2)}{\nu_3\nu_2^2\Gamma(\nu_1)}$	$-\frac{\Gamma(\nu_1-1/\nu_2)}{\nu_3^2\Gamma(\nu_1)}$

Each column of the matrix $\boldsymbol{\Delta}$ can be expressed by:

$$\frac{\partial\boldsymbol{\Psi}(\boldsymbol{\xi}|\boldsymbol{\omega})}{\partial\boldsymbol{\omega}_{(jl)}^\top}=-\boldsymbol{M}^\top\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}_\omega}{\partial\omega_{ij}^\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{u},$$

where the ith diagonal block of $\boldsymbol{\Sigma}_\omega$ is $\boldsymbol{\Sigma}_{\omega i}=\sqrt{\text{Var}(\boldsymbol{u}_i)}\backslash R(\boldsymbol{\rho}_{\omega i})\sqrt{\text{Var}(\boldsymbol{u}_i)}$, with $\boldsymbol{\rho}_{\omega i}=\left(\rho_{\omega i(12)},\ldots,\rho_{\omega i((t_i-1)t_i)}\right)$ and $i=1,\ldots,n$. Furthermore, we have:

$$\frac{\partial\boldsymbol{\Sigma}_{\omega i}}{\partial\omega_{i(jl)}}=\sqrt{\text{Var}(\boldsymbol{u}_i)}\frac{\partial\boldsymbol{R}(\boldsymbol{\rho}_{\omega i})}{\partial\omega_{i(jl)}}\sqrt{\text{Var}(\boldsymbol{u}_i)},$$

where $\partial\boldsymbol{R}(\boldsymbol{\rho}_{\omega_i})/\partial\omega_{i(jl)}$ is a symmetric matrix with null diagonal and jl and lj elements equal to $-\alpha_{jl}$, $i=1,\ldots,n$ and $j,l=1,\ldots,t_i$.

3.6 Framingham cholesterol data analysis

This data set is related to an unbalanced longitudinal experiment with respect to the number of repeated measurements involving 200 randomly selected subjects. The main goal is to examine the role of serum cholesterol as a risk factor for the evolution of cardiovascular diseases (Zhang and Davidian, 2001). The response variable is defined as the cholesterol level for each patient, and the covariates are: age in years (age), sex (0=female, 1=male) and years elapsed since the start of the study ($year$). Due to repeated measurement over the subjects, it is expected to observe within-subject dependence. This data set has been analyzed under different models, as: Zhang and Davidian (2001) that used linear mixed models with a flexible density for the random

effects, Lachos et al. (2010b) which used a linear mixed model with skew-normal random effects and Galarza et al. (2017) that proposed a quantile mixed model. This data set can be obtained through the package `qrLMM` (Galarza and Lachos, 2020) of the `R` software (R Core Team, 2020).

Figure 34 shows the relationship between years and cholesterol level by sex. We can see a difference on the response distribution between the males and females. Also, we can observe a higher variability among the individuals profiles, mainly for the females.

From Figure 35, we can observer a serial pattern with high values for the within-subject correlations. Then, among the correlation matrices presented in this work, the AR-1 seems the be the best option, even though we will compare the results under all structures. Based on Figure 36 we can not see a clear relationship between age and cholesterol level, suggesting that a non-parametric regression structure could be suitable to relate them.

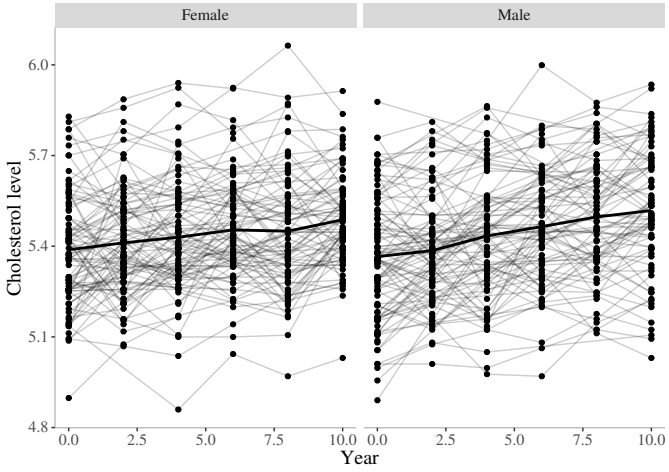


Figure 34 – Individual and average profiles for cholesterol level by sex.

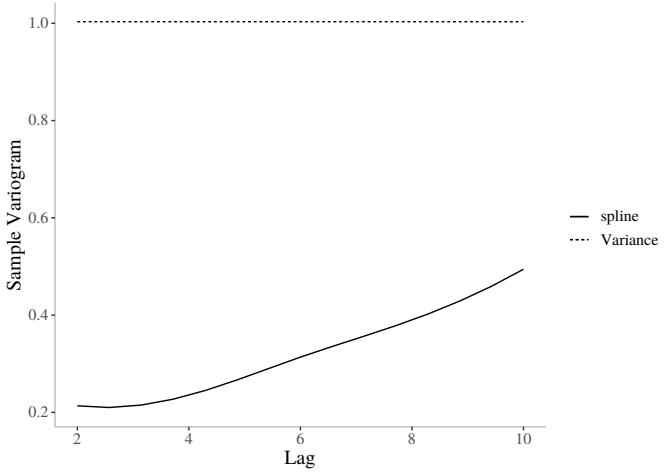


Figure 35 – Sample variogram for cholesterol level by year.

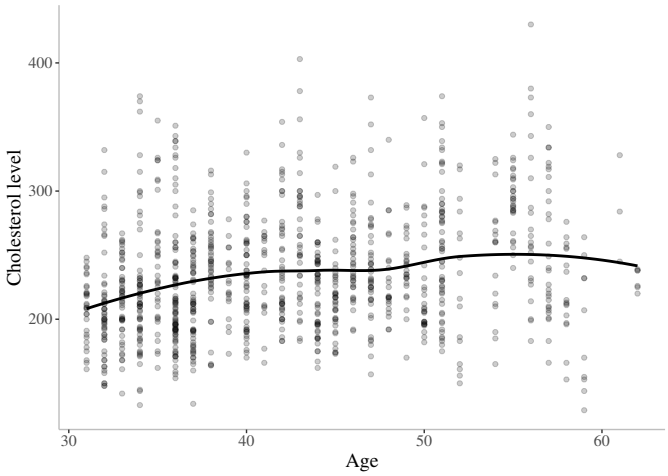


Figure 36 – Scatter plot between the cholesterol level and age, fitted by LOESS

We fitted several models, comparing our approach with some usual suggestions found in the literature, more specifically, we fitted a model for cholesterol level based on GEE and gamma, inverse gaussian (IG), CSN and SMCSN distributions using, exchangeable, AR-1 and unstructured working correlation matrices and considering:

$$
\begin{aligned}
Y_{ij} &\sim gamma(\mu_{ij}, \phi),\\
Y_{ij} &\sim IG(\mu_{ij}, \phi),\\
Y_{ij} &\sim CSN(\mu_{ij}, \sigma^2, \gamma),\\
Y_{ij} &\sim SMCSN(\mu_{ij}, \sigma^2, \gamma, \boldsymbol{\nu}),\\
\mu_{ij} &= \beta_1 \mathtt{sex} + \beta_2 \mathtt{year}^*_{ij} + f(age_{ij}),
\end{aligned}
$$

where Y_{ij} is the cholesterol level divided by 100 at the jth time point for the ith subject (Zhang and Davidian, 2001), $\mathtt{year}^*_{ij} = (\mathtt{year}_{ij} - 5)/10$, and SMCSN is a member of that family presented in this work.

From Table 5 the best working correlation matrices for each model is: exchangeable for Gamma, IG and CSBPN, unstructured for CSN, CST, CSCN, CSGT and CSBSN, and AR-1 for CSS and CSGGN.

Under the selected working correlation matrices, the results of Normal probability plots with 95% confidence simulated envelope for the quantile residuals are given in Figure 37. We have that the GEE CST, CSS, CSCN, CSGT and CSBPN models present a good fit, since there are no points outside the envelopes. On the other hand, for the Gamma, IG, CSN, CSBSN and CSGGN, we can see some points outside the envelope, indicating a poor fit. In Figure 38 we have the estimated nonparametric curves for the selected models, according to the QQ-plots. The dotted and dashed lines refer to 95% confidence bands built under the naive and robust variances, respectively. We can noticed similar behavior among all models, even though for the CST, CSCN and CSBPN models present bands with larger width, compared to the CSS and CSGT models. The

Table 5 – QAIC and QBIC of the fitted models for cholesterol level data.

Distribution \ Correlation	QAIC			QBIC		
	Exchangeable	Unstructured	AR-1	Exchangeable	Unstructured	AR-1
Gamma	**1348.668**	1353.073	1353.947	**1394.079**	1398.638	1401.068
Inverse Gaussian	**1524.998**	1568.22	1563.14	**1570.422**	1593.086	1615.362
CSN	1078.371	**1077.681**	1079.68	1127.987	**1126.916**	1131.419
CST	1072.220	**1072.081**	1072.176	1126.739	**1126.588**	1128.806
CSS	1034.767	1034.311	**1032.835**	1090.345	1089.941	**1089.724**
CSCN	**1024.716**	1029.911	1034.544	**1085.168**	1090.451	1096.382
CSGT	**1015.315**	1018.762	1024.203	**1070.769**	1074.293	1081.035
CSBPN	**1029.329**	1031.264	1030.945	**1089.810**	1091.813	1092.759
CSBSN	**854.3898**	856.124	857.945	**913.6731**	915.473	918.525
CSGGN	**1390.302**	15731.770	12240.310	**1452.812**	15815.210	12323.150

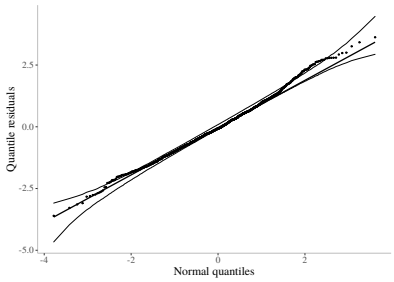
results of the respective fitted models are given in Table 6. From the naive and robust standard errors and the p-values for the individual test for nullity (within parentheses), we have that all coefficients were significant under for all models.

All models well fitted to the data, according to the QQ-plots, also present a well fit for the perturbation measures, then we select the model that showed the lowest QAIC and QBIC values. Even though these criterias based on the quasi-likelihood are not suitable for model selection concerning the error distribution (see Wang (2014), for example), we consider such results as an indicative. In this case, the CSGT model is the best one. Through the Figure 39 we can notice that for all perturbation schemes there are no influential points, showing the good fit of the selected model to the data.

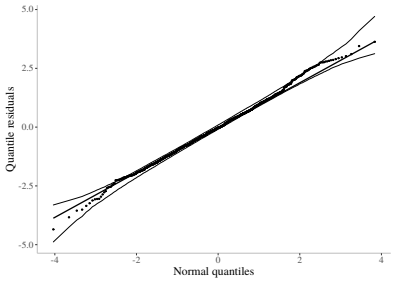
Analyzing the respective estimates we can notice that men have, on average, higher cholesterol levels than women. Also, there is an increasing in cholesterol levels over time. From the inspection of the non-parametric curves, we can see that patients around 55 years old tend to have a higher cholesterol level (in average), whereas patients over 60 and around 30 years old, tend to have low cholesterol levels.

Table 6 – Estimates, Standard errors (SE), p-values of Wald test and results for the parameters of gee models.

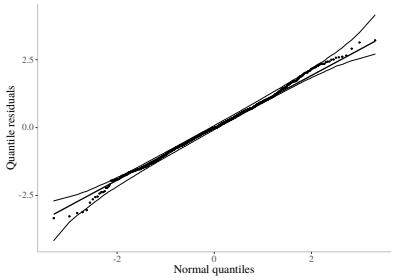
Model	Parameter	Est.	SE naive	SE robust
Gamma	β_1	-0.0053	0.0079(0.51)	0.0003(< 0.01)
	β_2	0.1205	0.0021(< 0.01)	< 0.0001(< 0.01)
	σ^2	0.0252	-	-
	$df(\alpha)$	6.1726	-	-
IG	β_1	-0.0222	0.0051(< 0.01)	0.0003(< 0.01)
	β_2	0.1212	0.0013(< 0.01)	< 0.0001(< 0.01)
	σ^2	0.0096	-	-
	$df(\alpha)$	6.1750	-	-
CSN	β_1	0.0126	0.0013(< 0.01)	0.0004(< 0.01)
	β_2	0.1197	0.0006(< 0.01)	< 0.0001(< 0.01)
	σ^2	0.1803	-	-
	γ	0.5486	-	-
	$df(\alpha)$	6.0201	-	-
CST	β_1	0.0126	0.0229(0.58)	0.0004(< 0.01)
	β_2	0.1197	0.0101(< 0.01)	< 0.0001(< 0.01)
	σ^2	0.1673	-	-
	γ	0.6018	-	-
	ν	25.1381	-	-
	$df(\alpha)$	6.0098	-	-
CSS	β_1	0.0054	0.0190(0.78)	< 0.0001(< 0.01)
	β_2	0.1230	0.0134(< 0.01)	< 0.0001(< 0.01)
	σ^2	0.1478	-	-
	γ	0.6164	-	-
	ν	4.5119	-	-
	$df(\alpha)$	6.4908	-	-
CSGT	β_1	0.0330	0.0213(0.12)	0.0002(< 0.01)
	β_2	0.1209	0.0055(< 0.01)	< 0.0001(< 0.01)
	γ	0.6626	-	-
	ν	(20.0000, 3.5396)	-	-
	$df(\alpha)$	6.2009	-	-
CSCN	β_1	0.0330	0.0212(0.12)	0.0002(< 0.01)
	β_2	0.1209	0.0055(< 0.01)	< 0.0001(< 0.01)
	σ^2	0.1	-	-
	γ	0.6540	-	-
	ν	(0.7762, 0.4492)	-	-
	$df(\alpha)$	6.2105	-	-
CSBPN	β_1	0.0330	0.0211(0.12)	0.0002(< 0.01)
	β_2	0.1209	0.0055(< 0.01)	< 0.0001(< 0.01)
	σ^2	0.1884	-	-
	γ	0.6358	-	-
	ν	(33.3333,33.3333)	-	-
	$df(\alpha)$	6.2164	-	-
CSBSN	β_1	0.0204	0.0005(< 0.01)	0.0003(< 0.01)
	β_2	0.1214	0.0002(< 0.01)	< 0.0001(< 0.01)
	σ^2	0.3220	-	-
	γ	0.9	-	-
	ν	(1.0000,2.0753)	-	-
	$df(\alpha)$	5.9877	-	-
CSGGN	β_1	0.0340	0.0014(< 0.01)	0.0012(< 0.01)
	β_2	0.1208	0.0004(< 0.01)	0.0001(< 0.01)
	σ^2	0.2111	-	-
	γ	0.6113	-	-
	ν	(5.577702, 1.000000, 19.304003)	-	-
	$df(\alpha)$	10.9715	-	-



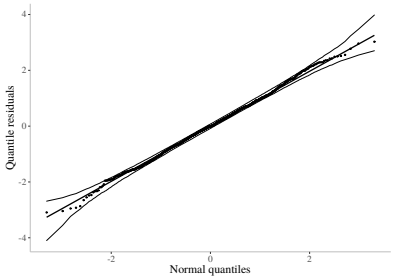
(a) Quantile-Quantile envelope for quantile residuals of Additive partially linear gamma model



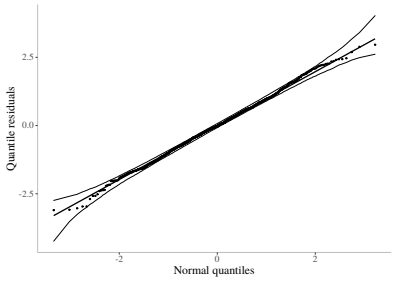
(b) Quantile-Quantile envelope for quantile residuals of Additive partially linear IG model



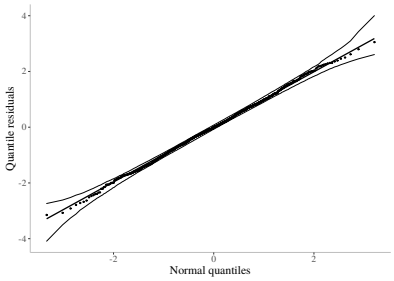
(c) Quantile-Quantile envelope for quantile residuals of Additive partially linear CSN model



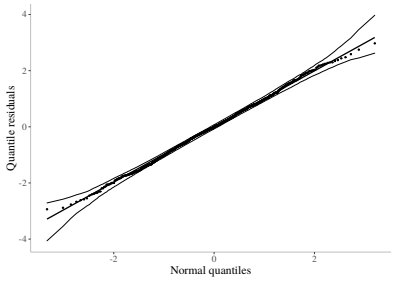
(d) Quantile-Quantile envelope for quantile residuals of Additive partially linear CST model



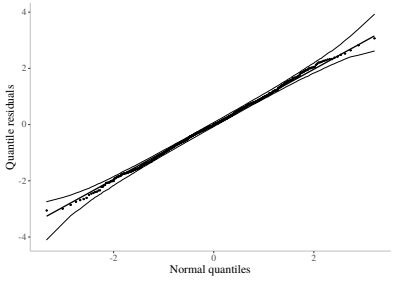
(e) Quantile-Quantile envelope for quantile residuals of Additive partially linear CSS model



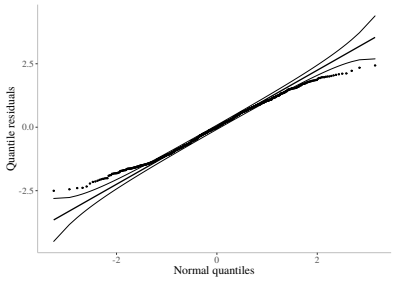
(f) Quantile-Quantile envelope for quantile residuals of Additive partially linear CSCN model



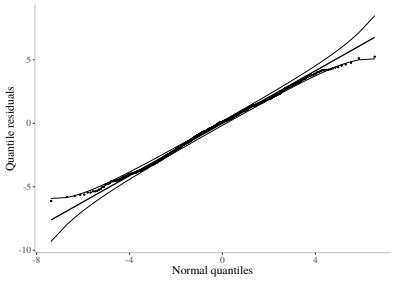
(g) Quantile-Quantile envelope for quantile residuals of Additive partially linear CSGT model



(h) Quantile-Quantile envelope for quantile residuals of Additive partially linear CSBPN model

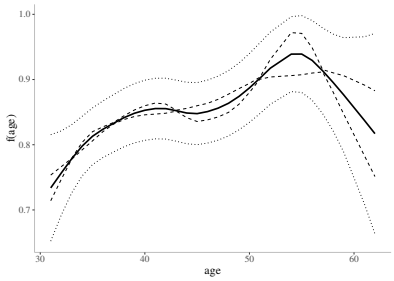


(i) Quantile-Quantile envelope for quantile residuals of Additive partially linear CSBSN model

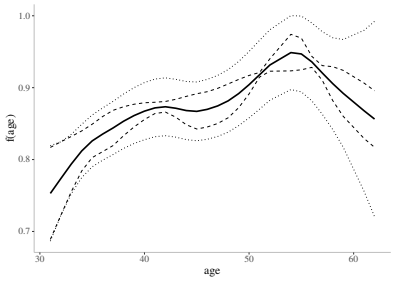


(j) Quantile-Quantile envelope for quantile residuals of Additive partially linear CSGGN model

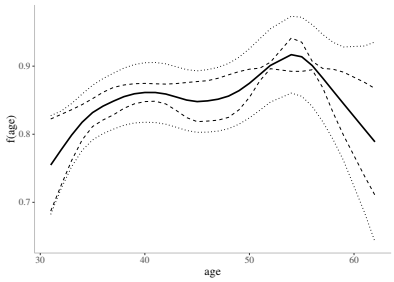
Figure 37 – Quantile-Quantile envelopes for fitted models to cholesterol data



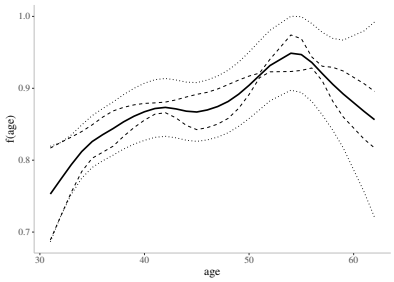
(a) 95% confidence bands for f(age) of Additive partially linear CST model.



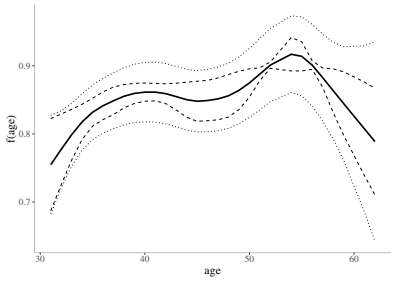
(b) 95% confidence bands for f(age) of Additive partially linear CSS model.



(c) 95% confidence bands for f(age) of Additive partially linear CSCN model.

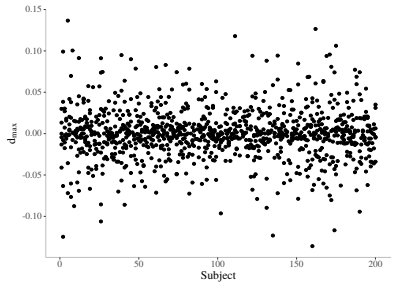


(d) 95% confidence bands for f(age) of Additive partially linear CSGT model.

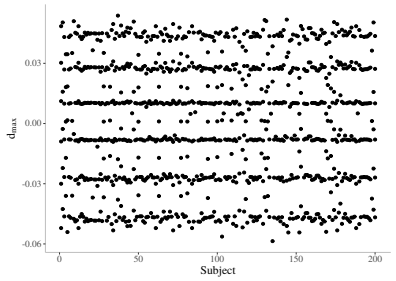


(e) 95% confidence bands for f(age) of Additive partially linear CSBPN model.

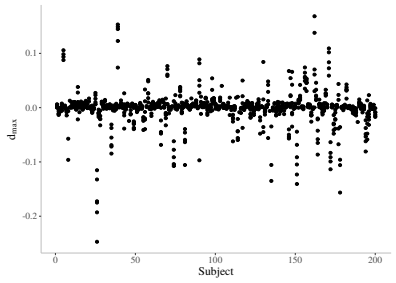
Figure 38 – 95% pointwise confidence bands for f(age) of fitted models.



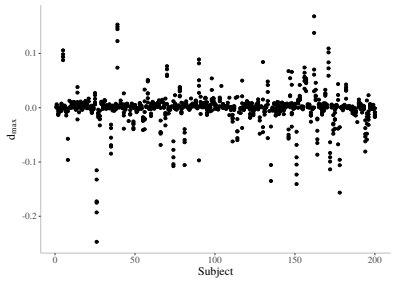
(a) d_{max} versus subject for case-weight perturbation.



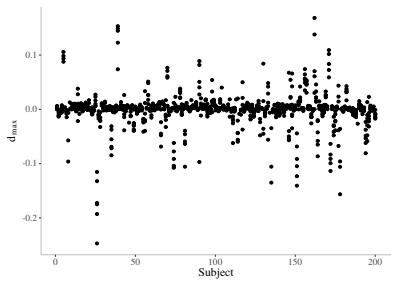
(b) d_{max} versus subject for response perturbation.



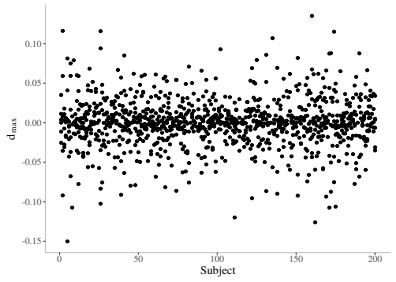
(c) d_{max} versus subject for skewness parameter perturbation.



(d) d_{max} versus subject for ν_1 perturbation.



(e) d_{max} versus subject for ν_2 perturbation.



(f) d_{max} versus subject for working correlation matrix perturbation.

Figure 39 – Local influence for GEE CSGT Model.

Chapter 4

Conclusions

Throughout this work we developed a class of regression models based on the scale mixtures of centred skew-normal distributions, as a natural extension of the normal regression model and as a more properly approach than the non-centred parameterization, skewed and/or heavy tailed data. Additionally we consider cases where we have independent and correlated response. In this sense, we propose regression models where the response distribution is a mixture between a skew-normal distribution and a mixing measure. We consider usual mixture distributions such as the beta, gamma and binary as well as never used models, namely beta prime, Birnbaum-Saunders and generalized gamma. We also propose a regression structure for the response mean through semi-parametric linear predictors.

We developed estimation methods under the frequentist approach for independent data using the SAEM algorithm, and for correlated data using generalized Estimation Equations. Through simulation studies, we find that the estimates approach the true values as the sample size increase, even though, the estimates for the shape parameters still need to be improved.

For model fit assessment, for independent data, we develop tools for global and local influence diagnostic analysis. We propose a residual based on the SMCSN family of distributions and we also present AIC and BIC criteria. Generalized leverage and Cook distance measures were obtained. In addition, local influence perturbation schemes were developed for: case-weight perturbation, scale parameter perturbation, skewness parameter perturbation, continuous covariate perturbation and response variable perturbation.

On the other hand, for correlated data, we also develop tools for global and local influence diagnostic analysis. We propose the use of quantile residuals and present AIC and BIC criteria based on the the respective quasi-likelihood. Local influence perturbation schemes were developed for: case-weight perturbation, scale parameter perturbation, skewness parameter perturbation, continuous covariate perturbation,

response variable perturbation and working correlation matrix perturbation.

Besides the advantages in terms of parameter interpretation, likelihood behavior and maximum likelihood estimator performance, in the real data analysis it was shown that our proposal overcomes the usual ones.

4.1 Future works

As future work we suggest the following research topics:

1. To improve the estimators for the shape parameters, proposing new estimation method such as the method of moments.
2. To propose models for independent and correlated data considering regression structures for the scale and skewness parameters.
3. To developed Bayesian analysis, including model fit assessment, model comparison and influence diagnostics, for the developed class of regression models.

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723. Citado na página 55.

Allassonnière, S., Kuhn, E., Trouvé, A., et al. (2010). Construction of bayesian deformable models via a stochastic approximation algorithm: a convergence study. *Bernoulli*, 16(3):641–678. Citado na página 49.

Andrews, D. F. and Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(1):99–102. Citado na página 25.

Arellano-Vale and Azzalini, A. (2008). The centred parametrization for the multivariate skew-normal distribution. *Journal of Multivariate Analysis*, 99(7):1362–1382. Citado 2 vezes nas páginas 26 and 39.

Arellano-Vale, B., R., Bolfarine, H., and Lachos, V. H. (2005). Skew-normal linear mixed models. *Journal of Data Science*, 3:415–438. Citado na página 76.

Arellano-Valle, R. B. and Azzalini, A. (2008). The centred parametrization for the multivariate skew-normal distribution. *Journal of multivariate analysis*, 99(7):1362–1382. Citado na página 25.

Atkinson, A. C. (1981). Two graphical displays for outlying and influential observations in regression. *Biometrika*, 68(1):13–20. Citado na página 57.

Azevedo, C. L. N., Bazan, J. L., and Farias, R. B. A. (2018). *Binary regression models: Inference and applications*. Associação Brasilieria de Estatística, mini-curso apresentado no 23^o sinape edition. Citado na página 27.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12:171–178. Citado 3 vezes nas páginas 25, 26, and 27.

Azzalini, A. and Valle, A. D. (1996). The multivariate skew-normal distribution. *Biometrika*, 83(4):715–726. Citado na página 26.

Basso, R. M., Lachos, V. H., Cabral, C. R. B., and Ghosh, P. (2010). Robust mixture modeling based on scale mixtures of skew-normal distributions. *Computational Statistics & Data Analysis*, 54(12):2926–2941. Citado na página 25.

Bates, D. M. and Watts, D. G. (1980). Relative curvature measures of nonlinearity. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42(1):1–25. Citado na página 60.

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Wiley Series in Probability and Statistics. John Wiley & Sons, New York. Citado na página 57.

Birnbaum, Z. W. and Saunders, S. C. (1969a). Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability*, 6(2):328–347. Citado na página 30.

Birnbaum, Z. W. and Saunders, S. C. (1969b). A new family of life distributions. *Journal of Applied Probability*, 6(2):319–327. Citado na página 30.

Bozdogan, H. (1987). Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika*, 52(3):345–370. Citado na página 55.

Breslow, N. and Clayton, D. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88:9–25. Citado na página 76.

Cadigan, N. G. and Farrell, P. J. (2002a). Generalized local influence with applications to fish stock cohort analysis. *Journal of the Royal Statistical Society Series C*, 51:469–483. Citado na página 57.

Cadigan, N. G. and Farrell, P. J. (2002b). Generalized local influence with applications to fish stok cohort analysis. *Journal of the Royal Statistical Society. Series C*, 51(4):469–483. Citado 3 vezes nas páginas 60, 84, and 86.

Cancho, V. G., Dey, D. K., Lachos, V. H., and Andrade, M. G. (2011). Bayesian nonlinear regression models with scale mixtures of skew-normal distributions: Estimation and case influence diagnostics. *Computational Statistics & Data Analysis*, 55(1):588–602. Citado na página 25.

Castro, L. M., Wang, W.-L., Lachos, V. H., Inácio de Carvalho, V., and Bayes, C. L. (2019). Bayesian semiparametric modeling for hiv longitudinal data with censoring and skewness. *Statistical methods in medical research*, 28(5):1457–1476. Citado na página 26.

Chaves, N. L., Azevedo, C. L., Vilca-Labra, F., and Nobre, J. S. (2019). A new birnbaum-saunders type distribution based on the skew-normal model under a centered parameterization. *Chilean Journal of Statistics (ChJS)*, 10(1). Citado na página 39.

Chaves, N. L., Azevedo, C. L., Vilca-Labra, F., and Nobre, J. S. (2020). A log birnbaum–saunders regression model based on the skew-normal distribution under the centred parameterization. *Statistics and Its Interface*, 13(3):335–346. Citado na página 23.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1):15–18. Citado na página 57.

Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(2):133–169. Citado 4 vezes nas páginas 57, 60, 84, and 86.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*. Retrieved from the University of Minnesota Digital Conservancy, http://hdl.handle.net/11299/37076. Chapman and Hall, New York. Citado na página 57.

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Wiley for the Royal Statistical Society*, 30(2):248–275. Citado na página 57.

Craven, P. and Wahba, G. (1978). Smoothing noisy data with spline functions. *Numerische mathematik*, 31(4):377–403. Citado na página 46.

Crowder, M. (1987). On linear and quadratic estimating functions. *Biometrika*, 74(3):591–597. Citado na página 79.

da Silva Ferreira, C., Bolfarine, H., and Lachos, V. H. (2011). Skew scale mixtures of normal distributions: properties and estimation. *Statistical Methodology*, 8(2):154–171. Citado na página 25.

da Silva Ferreira, C., Lachos, V. H., and Garay, A. M. (2019). Inference and diagnostics for heteroscedastic nonlinear regression models under skew scale mixtures of normal distributions. *Journal of Applied Statistics*, pages 1–30. Citado na página 26.

De Boor, C. (1978). *A practical guide to splines*, volume 27. springer-verlag New York. Citado 2 vezes nas páginas 44 and 45.

Delyon, B., Lavielle, M., and Moulines, E. (1999). Convergence of a stochastic approximation version of the em algorithm. *Annals of statistics*, pages 94–128. Citado 2 vezes nas páginas 44 and 49.

Demarta, S. and McNeil, A. J. (2005). The t copula and related copulas. *International statistical review*, 73(1):111–129. Citado na página 84.

Demidenko, E. (2013). *Mixed Models: Theory and Applications with R, 2nd Edition*. Wiley Series in Probability and Statistics. John Wiley & Sons, New York. Citado na página 76.

Desmond, A. (1986). On the relationship between two fatigue-life models. *IEEE Transactions on Reliability*, 35(2):167–169. Citado na página 111.

Dias, R. (1999). Sequential adaptive nonparametric regression via h-splines. *Communications in Statistics-Simulation and Computation*, 28(2):501–515. Citado na página 44.

Diggle, P., Heagerty, P., Liang, K.-Y., and Zeger, L. S. (1994). *Analysis of Longitudinal Data*, volume 25. Oxford. Citado na página 76.

Diggle, P. and Zeger, S. (1994). Semiparametric models for longitudinal data with application to cd4 cell numbers in hiv seroconverters. *Biometrics*, 50:689–99. Citado na página 76.

Eling, M. (2012). Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models? *Insurance: Mathematics and Economics*, 51(2):239–248. Citado na página 25.

Emerson, J. D., Hoaglin, D. C., and Kempthorne, P. J. (1984). Leverage in least squares additive-plus-multiplicative fits for two-way tables. *Journal of the American Statistical Association*, 79(386):329–335. Citado na página 59.

Eubank, R. L. (1988). *Spline smoothing and nonparametric regression*. Number 04; QA278. 2, E8. Citado na página 45.

Ferreira, C. d. S., Bolfarine, H., and Lachos, V. H. (2011). Skew scale mixtures of normal distributions: Properties and estimation. *Statistical Methodology*, 8(2):154 – 171. Citado na página 25.

Ferreira, C. S. and Lachos, V. H. (2016). Nonlinear regression models under skew scale mixtures of normal distributions. *Statistical Methodology*, 33:131–146. Citado na página 26.

Ferreira, C. S., Lachos, V. H., and Bolfarine, H. (2015). Inference and diagnostics in skew scale mixtures of normal regression models. *Journal of Statistical Computation and Simulation*, 85(3):517–537. Citado 2 vezes nas páginas 25 and 59.

Ferreira, C. S., Lachos, V. H., and Bolfarine, H. (2016). Likelihood-based inference for multivariate skew scale mixtures of normal distributions. *AStA Advances in Statistical Analysis*, 100(4):421–441. Citado na página 25.

Ferreira, C. S. and Paula, G. A. (2017). Estimation and diagnostic for skew-normal partially linear models. *Journal of Applied Statistics*, 44(16):3033–3053. Citado 2 vezes nas páginas 22 and 23.

Ferreira, G., Castro, L. M., Lachos, V. H., and Dias, R. (2013). Bayesian modeling of autoregressive partial linear models with scale mixture of normal errors. *Journal of Applied Statistics*, 40(8):1796–1816. Citado na página 49.

Freitas, J. V. B. d., Nobre, J. S., Bouguignon, M., and Santos-Neto, M. (2021+a). A new approach to modelling positive random variables with repeated measures. *Submitted to Journal of Applied Statistics*. Citado na página 77.

Freitas, J. V. B. d., Nobre, J. S., Espinheira, P. L., and Rêgo, L. C. (2021+b). Marginal regression models to analyzing bounded correlated data with varying precision/dispersion. *Submitted to TEST*. Citado na página 77.

Galarza, C. E. and Lachos, V. H. (2020). *qrLMM: Quantile Regression for Linear Mixed-Effects Models*. R package version 2.1. Citado na página 91.

Galarza, C. E., Lachos, V. H., and Bandyopadhyay, D. (2017). Quantile regression in linear mixed models: a stochastic approximation em approach. *Statistics and its Interface*, 10(3):471. Citado 2 vezes nas páginas 52 and 91.

Galarza Morales, C., Lachos Davila, V., Barbosa Cabral, C., and Castro Cepero, L. (2017). Robust quantile regression using a generalized class of skewed distributions. *Stat*, 6(1):113–130. Citado na página 26.

Garay, A. M., Lachos, V. H., and Abanto-Valle, C. A. (2011). Nonlinear regression models based on scale mixtures of skew-normal distributions. *Journal of the Korean Statistical Society*, 40(1):115–124. Citado na página 25.

Garay, A. M., Lachos, V. H., Labra, F. V., and Ortega, E. M. (2014). Statistical diagnostics for nonlinear regression models based on scale mixtures of skew-normal distributions. *Journal of Statistical Computation and Simulation*, 84(8):1761–1778. Citado na página 25.

Godambe, V. P., editor (1991). *Estimating Functions*. Oxford University Press, Oxford. Citado 2 vezes nas páginas 77 and 81.

Gradshteyn, I. S. and Ryzhik, I. M. (2014). *Table of integrals, series, and products*. Academic press. Citado na página 31.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A roughness penalty approach*. Chapman & Hall, New York. Citado 2 vezes nas páginas 45 and 46.

Hajrajabi, A. and Maleki, M. (2019). Nonlinear semiparametric autoregressive model with finite mixtures of scale mixtures of skew normal innovations. *Journal of Applied Statistics*, pages 1–20. Citado na página 26.

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(2):190–195. Citado na página 55.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman & Hall, London. Citado 4 vezes nas páginas 44, 55, 65, and 82.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. Citado na página 50.

Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics*, 9(2):226–252. Citado na página 76.

Henderson, C. R., Kempthorne, O., Searle, S. R., and von Krosigk, C. M. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics*, 15(2):192–218. Citado na página 76.

Henze, N. (1986). A probabilistic representation of the 'skew-normal' distribution. *Scandinavian Journal of Statistics*, 13(4):271–275. Citado na página 26.

Hoaglin, D. C. and Welsch, R. E. (1978). The hat matrix in regression and anova. *The American Statistician*, 32(1):17–22. Citado 2 vezes nas páginas 57 and 59.

Hurvich, C. M. and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307. Citado na página 55.

Ibacache-Pulgar, G., Paula, G. A., and Cysneiros, F. J. A. (2013). Semiparametric additive models under symmetric distributions. *Test*, 22(1):103–121. Citado na página 55.

Ibacache-Pulgar, G. and Reyes, S. (2017). Local influence for elliptical partially varying-coefficient model. *Statistical Modelling*, 18(2):149–174. Citado na página 55.

Jørgensen, B. and Knudsen, S. J. (2004). Parameter orthogonality and bias adjustment for estimating functions. *Scandinavian Journal of Statistics*, 31(1):93–114. Citado na página 81.

Jørgensen, B. and Labouriau, R. (1994). *Exponential Families and Theoretical Inference*. Leture notes, University of British Columbia, Vancouver. Citado 2 vezes nas páginas 79 and 80.

Keeping, E. (1962). Introduction to statistical inference, vol. 26. *DoverPublications. com*. Citado na página 30.

Kohn, R., Smith, M., and Chan, D. (2001). Nonparametric regression using linear combinations of basis functions. *Statistics and Computing*, 11(4):313–322. Citado na página 44.

Kolmogorov, A. and Fomin, S. (1999). *Elements of the Theory of Functions and Functional Analysis*. Number v. 1 in Dover books on mathematics. Dover. Citado na página 44.

Kong, M., Xu, S., Levy, S. M., and Datta, S. (2015). Gee type inference for clustered zero-inflated negative binomial regression with application to dental caries. *Computational statistics & data analysis*, 85:54–66. Citado na página 77.

Kuhn, E. and Lavielle, M. (2004). Coupling a stochastic approximation version of em with an mcmc procedure. *ESAIM: Probability and Statistics*, 8:115–131. Citado na página 52.

Kuhn, E. and Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Computational statistics & data analysis*, 49(4):1020–1038. Citado na página 49.

Labra, F. V., Garay, A. M., Lachos, V. H., and Ortega, E. M. (2012). Estimation and diagnostics for heteroscedastic nonlinear regression models based on scale mixtures of skew-normal distributions. *Journal of Statistical Planning and Inference*, 142(7):2149–2165. Citado na página 25.

Lachos, V., Labra, F., Bolfarine, H., and Ghosh, P. (2010a). Multivariate measurement error models based on scale mixtures of the skew–normal distribution. *Statistics*, 44(6):541–556. Citado na página 25.

Lachos, V. H., Bandyopadhyay, D., and Garay, A. M. (2011). Heteroscedastic nonlinear regression models based on scale mixtures of skew-normal distributions. *Statistics & probability letters*, 81(8):1208–1217. Citado na página 25.

Lachos, V. H., Ghosh, P., and Arellano-Valle, R. B. (2010b). Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica*, pages 303–322. Citado na página 91.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22. Citado 4 vezes nas páginas 77, 80, 81, and 82.

Lindstrom, M. J. and Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, 46(3):673–687. Citado na página 76.

Liu, C. and Rubin, D. B. (1994). The ecme algorithm: a simple extension of em and ecm with faster monotone convergence. *Biometrika*, 81(4):633–648. Citado 2 vezes nas páginas 44 and 49.

Louis, T. A. (1982). Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):226–233. Citado na página 55.

Maioli, M. C. (2018a). Univariate and bivariate regression models based on centered skew scale mixture of normal distributions. *Dissertação (Dissertação em Estatística) - Universidade Estadual de Campinas*. Citado 7 vezes nas páginas 27, 29, 39, 43, 44, 47, and 48.

Maioli, M. C. (2018b). Univariate and bivariate regression models based on centered skew scale mixture of normal distributions: Modelos de regressão univariados e bivariados baseados nas distribuições de mistura de escala normal assimétrica sob a parametrização centrada. Citado 2 vezes nas páginas 25 and 57.

Manghi, R. F., Cysneiros, F. J. A., and Paula, G. A. (2019). Generalized additive partial linear models for analyzing correlated data. *Computational Statistics and Data Analysis*, 128:47–60. Citado 6 vezes nas páginas 77, 81, 82, 83, 85, and 87.

Massuia, M. B., Garay, A. M., Lachos, V. H., and Cabral, C. R. (2017). Bayesian analysis of censored linear regression models with scale mixtures of skew-normal distributions. *Statistics and Its Interface*, 10(3):425–439. Citado na página 26.

Mattos, T. d. B., Garay, A. M., and Lachos, V. H. (2018). Likelihood-based inference for censored linear regression models with scale mixtures of skew-normal distributions. *Journal of Applied Statistics*, 45(11):2039–2066. Citado na página 26.

Meijer, C. (1936). Uber whittakersche bezw. besselsche funktionen und deren produkte. *Nieuw Archief voor Wiskunde*, 18(2):10–29. Citado na página 31.

Meilijson, I. (1989). A fast improvement to the em algorithm on its own terms. *Journal of the Royal Statistical Society: Series B (Methodological)*, 51(1):127–138. Citado na página 56.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092. Citado na página 50.

Osorio, F. (2006). Diagnóstico de influência em modelos elıpticos com efeitos mistos. *Unpublished Ph. D. dissertation.(Dept. of Statistics, University of Sao Paulo, Brazil, 2006)*. Citado 2 vezes nas páginas 57 and 59.

Padilla, J. L., Azevedo, C. L., and Lachos, V. H. (2018). Multidimensional multiple group irt models with skew normal latent trait distributions. *Journal of Multivariate Analysis*, 167:250–268. Citado na página 26.

Preisser, J. S. and Qaqish, B. F. (1996). Deletion diagnostics for generalised estimating equations. *Biometrika*, 83(3):551–562. Citado na página 84.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Citado 2 vezes nas páginas 52 and 91.

Ripley, B. D. (2009). *Stochastic simulation*, volume 316. John Wiley & Sons. Citado na página 50.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric regression*. Number 12. Cambridge university press. Citado na página 22.

Sarvi, F., Moghimbeigi, A., and Mahjub, H. (2019). Gee-based zero-inflated generalized poisson model for clustered over or under-dispersed count data. *Journal of Statistical Computation and Simulation*, 89(14):2711–2732. Citado na página 77.

Savalli, C., Paula, G. A., and Cysneiros, F. J. A. (2006). Assessment of variance components in elliptical linear mixed models. *Statistical Modelling*, 6:59–76. Citado na página 77.

Schumaker, L. (2007). *Spline functions: basic theory*. Cambridge University Press. Citado na página 45.

Schwarz, G. et al. (1978). Estimating the dimension of a model. *Annals of statistics*, 6(2):461–464. Citado na página 55.

Schwarzmann, B. (1991a). A connection between local-influence analysis and residual diagnostics. *Technometrics*, 33(1):103–104. Citado na página 61.

Schwarzmann, B. (1991b). A connection between local-influence analysis and residual diagnostics. *Technometrics*, 33(1):103–104. Citado na página 87.

Segal, M. R., Bacchetti, P., and Jewell, N. P. (1994). Variances for maximum penalized likelihood estimates obtained via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(2):345–352. Citado na página 55.

Sen, P. K. and Singer, J. M. (1993). *Large Sample Methods in Statistics: an introdution with applications*. Boca Raton. Citado na página 77.

Singer, J., Nobre, J., and Rocha, F. (2017). *Análise de Dados Longitudinais (versão parcial preliminar)*. Citado na página 76.

Singer, J. M. and Andrade, D. F. (2000). 5 analysis of longitudinal data. In *Bioenvironmental and Public Health Statistics*, volume 18 of *Handbook of Statistics*, pages 115 – 160. Elsevier. Citado na página 76.

Song, P. X.-K., Qiu, Z., and Tan, M. (2004). Modelling heterogeneous dispersion in marginal models for longitudinal proportional data. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 46(5):540–553. Citado na página 77.

Stacy, E. (1962). A generalization of the gamma distribution. *Annals of Mathematical Statistics*, 33(3):1187–1192. Citado na página 30.

Stark, P. C., Ryan, L. M., McDonald, J. L., and Burge, H. A. (1997). Using meteorologic data to predict daily ragweed pollen levels. *Aerobiologia*, 13(3):177. Citado na página 22.

Tan, M., Qu, Y., and H.Kutner, M. (1997). Model diagnostics for marginal regression analysis of correlated binary data. *Communications in Statistics - Simulation and Computation*, 26(2):539–558. Citado na página 84.

Thomas, W. and Cook, R. D. (1989). Assessing influence on regression coefficients in generalized linear models. *Biometrika*, 76(4):741–749. Citado na página 87.

Tsuyuguchi, A. B., Paula, G. A., and Barros, M. (2020). Analysis of correlated birnbaum–saunders data based on estimating equations. *TEST*, 29(3):661–681. Citado na página 77.

Vanegas, L. H. and Paula, G. A. (2016). An extension of log-symmetric regression models: R codes and applications. *Journal of Statistical Computation and Simulation*, 86:1709–1735. Citado 2 vezes nas páginas 47 and 115.

Venezuela, M. K., Botter, D. A., and Sandoval, M. C. (2007). Diagnostic techniques in generalized estimating equations. *Journal of Statistical Computation and Simulation*, 77(10):879–888. Citado 2 vezes nas páginas 84 and 85.

Venezuela, M. K., Sandoval, M. C., and Botter, D. A. (2011). Local influence in estimating equations. *Computational Statistics & Data Analysis*, 55(4):1867 – 1883. Citado 4 vezes nas páginas 84, 86, 87, and 89.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York. Citado 2 vezes nas páginas 60 and 61.

Vonesh, E. and Chinchilli, V. M. (1996). *Linear and nonlinear models for the analysis of repeated measurements*. CRC press, Boca Raton. Citado na página 76.

Wahba, G. (1981). Data-based optimal smoothing of orthogonal series density estimates. *The annals of statistics*, pages 146–156. Citado na página 46.

Wang, M. (2014). Generalized estimating equations in longitudinal data analysis: a review and recent developments. *Advances in Statistics*, 2014. Citado na página 93.

Ware, J. H. (1985). Linear models for the analysis of longitudinal studies. *The American Statistician*, 39(2):95–101. Citado na página 76.

Wei, B.-C., Hu, Y.-Q., and Fung, W.-K. (1998). Generalized leverage and its applications. *Scandinavian Journal of Statistics*, 25(1):25–37. Citado 2 vezes nas páginas 49 and 59.

Wood, S. N. (2017). *Generalized Additive Models: An introduction with R*. CRC Press, Boca Raton, second edition. Citado 2 vezes nas páginas 47 and 115.

Zeller, C. B., Cabral, C. R., and Lachos, V. H. (2016). Robust mixture regression modeling based on scale mixtures of skew-normal distributions. *Test*, 25(2):375–396. Citado na página 25.

Zeller, C. B., Carvalho, R. R., and Lachos, V. H. (2012). On diagnostics in multivariate measurement error models under asymmetric heavy-tailed distributions. *Statistical Papers*, 53(3):665–683. Citado na página 25.

Zeller, C. B., Lachos, V. H., and Vilca-Labra, F. E. (2011). Local influence analysis for regression models with scale mixtures of skew-normal distributions. *Journal of Applied Statistics*, 38(2):343–368. Citado na página 25.

Zhang, D. and Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*, 57(3):795–802. Citado 2 vezes nas páginas 90 and 92.

Zhu, H.-T. and Lee, S.-Y. (2001). Local influence for incomplete data models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(1):111–126. Citado 3 vezes nas páginas 57, 58, and 60.

APPENDIX A

Details to obtain the Scale Mixture of Normal distributions

A.1 Distribution 2

Has noted by Desmond (1986) the BS can be represented by Inverse Gaussian Distribution (IG). Consider two random variables $X_1 \sim IG(\mu_I, \lambda_I)$ and $X_2^{-1} \sim IG(\mu_I^{-1}, \lambda\mu_I^2)$, then the p.d.f. of U is given by

$$h(u|\nu_1, \nu_2) = \frac{1}{2} f_{X_1}(u|\mu_I, \lambda_I) + \frac{1}{2} f_{X_2}(u|\mu_I, \lambda_I),$$

with

$$f_{X_1}(u|\mu_I, \lambda_I) = \left(\frac{\lambda_I}{2\pi u^3}\right)^{1/2} \exp\left\{-\frac{\lambda_I}{2\mu_I^2 u}(u-\mu_I)^2\right\},$$

and $f_{X_2}(u|\mu_I, \lambda_I) = u f_{X_1}(u|\mu_I, \lambda_I)/\mu_I$, where $\nu_1 = \sqrt{\mu_I/\lambda_I}$ and $\nu_2 = \mu_I$, that implies $\lambda_I = \nu_2/\nu_1^2$. That said, we have that the distribution 2, denoted by $D2(\mu, \sigma^2, \nu_1, \nu_2)$, and p.d.f:

$$\begin{aligned} f(y|\mu, \sigma^2, \nu_1, \nu_2) &= \int_0^\infty \frac{\sqrt{u}}{\sqrt{2\pi}\sigma} e^{-\frac{ud}{2}} \frac{1}{2} \left[f_{X_1}(u|\nu_1, \nu_2) + f_{X_2}(u|\nu_1, \nu_2)\right] du \\ &= \int_0^\infty \frac{\sqrt{u}}{\sqrt{2\pi}\sigma} e^{-\frac{ud}{2}} \frac{1}{2} f_{X_1}(u|\nu_1, \nu_2) du + \int_0^\infty \frac{\sqrt{u}}{\sqrt{2\pi}\sigma} e^{-\frac{ud}{2}} \frac{1}{2} f_{X_2}(u|\nu_1, \nu_2) du \\ &= f_1(y|\mu, \sigma^2, \nu_1, \nu_2) + f_2(y|\mu, \sigma^2, \nu_1, \nu_2), \end{aligned}$$

$$\begin{aligned} f_1(y|\mu, \sigma^2, \nu_1, \nu_2) &= \int_0^\infty \frac{\sqrt{u}}{\sqrt{2\pi}\sigma} e^{-\frac{ud}{2}} \frac{1}{2} \left(\frac{\lambda_I}{2\pi u^3}\right)^{1/2} \exp\left\{-\frac{\lambda_I}{2\mu_I^2 u}(u-\mu_I)^2\right\} du \\ &= \frac{1}{2\sqrt{2\pi}\sigma} \int_0^\infty e^{-\frac{ud}{2}} \left(\frac{\lambda_I}{2\pi u^2}\right)^{1/2} \exp\left\{-\frac{\lambda_I u}{2\mu_I^2} + \frac{\lambda_I}{\mu_I} - \frac{\lambda_I}{2u}\right\} du \end{aligned}$$

$$
\begin{aligned}
&= \frac{e^{\frac{\lambda_I}{\mu_I}}}{2\sqrt{2\pi}\sigma}\int_0^\infty \left(\frac{\lambda_I}{2\pi u^2}\right)^{1/2}\exp\left\{-\frac{\lambda_I u}{2\mu_I^2}-\frac{ud}{2}-\frac{\lambda_I}{2u}\right\}du\\
&= \frac{e^{\frac{\lambda_I}{\mu_I}}}{2\sqrt{2\pi}\sigma}\int_0^\infty \left(\frac{\lambda_I}{2\pi u^2}\right)^{1/2}\exp\left\{-\left(\frac{1}{2\mu_I^2}+\frac{d}{2\lambda_I}\right)\lambda_I u-\frac{\lambda_I}{2u}\right\}du\\
\beta^2 &= 2\lambda_I^2\left(\frac{1}{2\mu_I^2}+\frac{d}{2\lambda_I}\right), x=\left(\frac{1}{2\mu_I^2}+\frac{d}{2\lambda_I}\right)\lambda_I u=\frac{\beta^2}{2\lambda_I}u\\
&= \frac{e^{\frac{\lambda_I}{\mu_I}}}{4\pi\sigma}\int_0^\infty \left(\frac{\beta^4}{x^2 4\lambda_I}\right)^{1/2}\exp\left\{-x-\frac{\beta^2}{4x}\right\}\frac{2\lambda_I}{\beta^2}dx\\
&= \frac{e^{\frac{\lambda_I}{\mu_I}}\sqrt{\lambda_I}}{4\pi\sigma}\int_0^\infty \exp\left\{-x-\frac{\beta^2}{4x}\right\}\frac{1}{x}dx\\
&= \frac{e^{\frac{\lambda_I}{\mu_I}}\sqrt{\lambda_I}}{2\pi\sigma}K_0(\beta) = \frac{e^{\frac{\lambda_I}{\mu_I}}\sqrt{\lambda_I}}{2\pi\sigma}K_0\left(\lambda_I\sqrt{2\left(\frac{1}{2\mu_I^2}+\frac{d}{2\lambda_I}\right)}\right)\\
&= \frac{e^{\frac{\lambda_I}{\mu_I}}\sqrt{\lambda_I}}{2\pi\sigma}K_0\left(\lambda_I\sqrt{2\left(\frac{1}{2\mu_I^2}+\frac{d}{2\lambda_I}\right)}\right),
\end{aligned}
$$

where $K_\eta(z)$ is the modified Bessel function given in (1.6).

$$
\begin{aligned}
f_2(y|\mu,\sigma^2,\nu_1,\nu_2) &= \int_0^\infty \frac{\sqrt{u}}{\sqrt{2\pi}\sigma}e^{-\frac{ud}{2}}\frac{1}{2}\frac{u}{\mu_I}\left(\frac{\lambda_I}{2\pi u^3}\right)^{1/2}\exp\left\{-\frac{\lambda_I}{2\mu_I^2 u}(u-\mu_I)^2\right\}du\\
&= \frac{1}{2\sqrt{2\pi}\sigma\mu_I}\int_0^\infty e^{-\frac{ud}{2}}\left(\frac{\lambda_I}{2\pi}\right)^{1/2}\exp\left\{-\frac{\lambda_I u}{2\mu_I^2}+\frac{\lambda_I}{\mu_I}-\frac{\lambda_I}{2u}\right\}du\\
&= \frac{e^{\frac{\lambda_I}{\mu_I}}\sqrt{\lambda_I}}{4\pi\sigma\mu_I}\int_0^\infty \exp\left\{-\frac{\lambda_I u}{2\mu_I^2}-\frac{ud}{2}-\frac{\lambda_I}{2u}\right\}du\\
&= \frac{e^{\frac{\lambda_I}{\mu_I}}\sqrt{\lambda_I}}{4\pi\sigma\mu_I}\int_0^\infty \exp\left\{-\left(\frac{1}{2\mu_I^2}+\frac{d}{2\lambda_I}\right)\lambda_I u-\frac{\lambda_I}{2u}\right\}du\\
\beta^2 &= 2\lambda_I^2\left(\frac{1}{2\mu_I^2}+\frac{d}{2\lambda_I}\right), x=\left(\frac{1}{2\mu_I^2}+\frac{d}{2\lambda_I}\right)\lambda_I u=\frac{\beta^2}{2\lambda_I}u\\
&= \frac{e^{\frac{\lambda_I}{\mu_I}}\sqrt{\lambda_I}}{4\pi\sigma\mu_I}\int_0^\infty \exp\left\{-x-\frac{\beta^2}{4x}\right\}\frac{2\lambda_I}{\beta^2}dx = \frac{e^{\frac{\lambda_I}{\mu_I}}\lambda_I^{3/2}}{2\pi\sigma\mu_I\beta^2}K_{-1}(\beta)
\end{aligned}
$$

$$
\begin{aligned}
f(y|\mu,\sigma^2,\nu_1,\nu_2) &= f_1(y|\mu,\sigma^2,\nu_1,\nu_2)+f_2(y|\mu,\sigma^2,\nu_1,\nu_2)\\
&= \frac{e^{\frac{\lambda_I}{\mu_I}}\sqrt{\lambda_I}}{2\pi\sigma}K_0(\beta)+\frac{e^{\frac{\lambda_I}{\mu_I}}\lambda_I^{3/2}}{2\pi\sigma\mu_I\beta^2}K_{-1}(\beta) = \frac{e^{\frac{\lambda_I}{\mu_I}}\lambda_I^{1/2}}{2\pi\sigma}\left[K_0(\beta)+\frac{\lambda_I}{\mu_I\beta^2}K_{-1}(\beta)\right]\\
&= \frac{e^{\frac{1}{\nu_1^2}}\sqrt{\nu_2}}{2\pi\sigma\nu_1}\left[K_0(\beta^*)+\frac{\nu_1^2}{1+d\nu_2\nu_1^2}K_{-1}(\beta^*)\right],
\end{aligned}
$$

with $\beta^* = \sqrt{1+d\nu_2\nu_1^2}/\nu_1^2$.

A.2 Distribution 3

$$
\begin{aligned}
f(y|\mu,\sigma^2,\nu_1,\nu_2) &= \int_0^\infty \frac{\sqrt{u}}{\sqrt{2\pi}\sigma} e^{-\frac{ud}{2}} \frac{\nu_2}{\nu_3\Gamma(\nu_1)} \left(\frac{u}{\nu_3}\right)^{\nu_1\nu_2-1} \exp\left\{-\left(\frac{u}{\nu_3}\right)^{\nu_2}\right\} du \\
&= \frac{\nu_2}{\sqrt{2\pi}\sigma\nu_3\Gamma(\nu_1)} \int_0^\infty e^{-\frac{ud}{2}} \left(\frac{u}{\nu_3}\right)^{\nu_1\nu_2-1/2} \exp\left\{-\left(\frac{u}{\nu_3}\right)^{\nu_2}\right\} du \\
& x = u/\nu_3, \\
&= \frac{\nu_2}{\sqrt{2\pi}\sigma\Gamma(\nu_1)} \int_0^\infty e^{-x\nu_3\frac{d}{2}} x^{\nu_1\nu_2-1/2} \exp\left\{-x^{\nu_2}\right\} dx \\
&= \frac{\nu_2}{\sqrt{2\pi}\sigma\Gamma(\nu_1)} \int_0^\infty e^{-x\nu_3\frac{d}{2}} x^{\nu_1\nu_2-1/2} \exp\left\{-x^{\nu_2}\right\} dx \\
&= \frac{\nu_2}{\sqrt{2\pi}\sigma\Gamma(\nu_1)} \sum_{m=0}^\infty (-1)^m \frac{(\nu_3 d/2)^m}{m!} \int_0^\infty x^{\nu_1\nu_2+m-1/2} \exp\left\{-x^{\nu_2}\right\} dx, \\
& w = x^{\nu_2} \\
&= \frac{\nu_2}{\sqrt{2\pi}\sigma\Gamma(\nu_1)} \sum_{m=0}^\infty (-1)^m \frac{(\nu_3 d/2)^m}{m!} \int_0^\infty \frac{1}{\nu_2} w^{\nu_1+m/\nu_2+1/(2\nu_2)-1} \exp\left\{-w\right\} dw, \\
&= \frac{1}{\sqrt{2\pi}\sigma\Gamma(\nu_1)} \sum_{m=0}^\infty (-1)^m \frac{(\nu_3 d/2)^m}{m!} \Gamma\left(\nu_1 + \frac{m}{\nu_1} + \frac{1}{2\nu_2} - 1\right),
\end{aligned}
$$

APPENDIX B

Matrix form of $\int_{\chi} \left[g_j^{(2)}(w) \right]^2 dw$

Let $\boldsymbol{b} = (b_1, \ldots, b_k)^\top$ be a vector of know $\boldsymbol{k}$ basis functions. Consider a function g that can be approximated as a linear combination of these bases, that is, $g(t) = \boldsymbol{\kappa}^\top \boldsymbol{b}(t)$, where $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_k)^\top$. Since $g(t)^{(2)}$ is a constant, obtaining the matrix form of the penalty term to the non-smoothness of the curve is given by:

$$\begin{aligned}
\int_{\chi} \left[g_j^{(2)}(w) \right]^2 dw &= \int_{\chi} \left[\boldsymbol{\kappa}^\top \boldsymbol{b}^{(2)}(w) \right]^2 dw \\
&= \int_{\chi} \boldsymbol{\kappa}^\top \boldsymbol{b}^{(2)}(w) \boldsymbol{b}^{(2)}(w) \boldsymbol{\kappa} dw \\
&= \boldsymbol{\kappa}^\top \left[\int_{\chi} \boldsymbol{b}^{(2)}(w) \boldsymbol{b}^{(2)}(w) dw \right] \boldsymbol{\kappa} \\
&= \boldsymbol{\kappa}^\top \boldsymbol{\Omega} \boldsymbol{\kappa},
\end{aligned}$$

where $\boldsymbol{\Omega}$ is a square matrix $(k \times k)$ with elements: $\Omega_{ij} = \int_{\chi} b_i^{(2)}(w) b_j^{(2)}(w) dw$. Considering the observed values $(x_1, \ldots, x_n)$, we have to: $\Omega_{ij} = \sum_{l=1}^{n} b_i^{(2)}(x_l) b_j^{(2)}(x_l)$.

APPENDIX C

Additive Partial Linear Model Identifiability

Consider the additive partial linear model given by

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \sum_{j=1}^{q} \boldsymbol{f}_j + \boldsymbol{\epsilon} = \boldsymbol{X}\boldsymbol{\beta} + \sum_{j=1}^{q} \boldsymbol{T}_j\boldsymbol{\theta}_j + \boldsymbol{\epsilon} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{T}\boldsymbol{\theta} + \boldsymbol{\epsilon}.$$

This model may be unidentifiable, for this we need to impose a restriction on the parameters $\boldsymbol{\theta}$. As seen in Wood (2017) and Vanegas and Paula (2016) an appropriate constraint would be $\mathbf{1}_n^\top \boldsymbol{f}_j = \mathbf{1}_n^\top \boldsymbol{T}_j\boldsymbol{\theta}_j = 0$, where $\mathbf{1}_n^\top$ is a n-vector of ones.

Note that this is a restriction of type $\boldsymbol{C}\boldsymbol{\theta} = \mathbf{0}$. To use this constraint we can apply the QR decomposition:

$$\boldsymbol{C}^\top = \boldsymbol{Q} \times (\boldsymbol{R}, \mathbf{0})^\top$$

where $\boldsymbol{Q}$ is an orthogonal matrix $(k \times k)$ and $\boldsymbol{R}$ is an upper triangular matrix $(q \times q)$, with $k = 1 + k_1 + \cdots + k_q$. The matrix $\boldsymbol{Q}$ can be partitioned as $\boldsymbol{Q} \equiv [\boldsymbol{D} : \boldsymbol{Z}]$, where $\boldsymbol{Z}$ is a matrix $k \times (k - q)$.

In this way, $\boldsymbol{\theta} = \boldsymbol{Z}\boldsymbol{\theta}_z$ will satisfy the constraints for any vector $\boldsymbol{\theta}_z$ of dimension $k - p$, because:

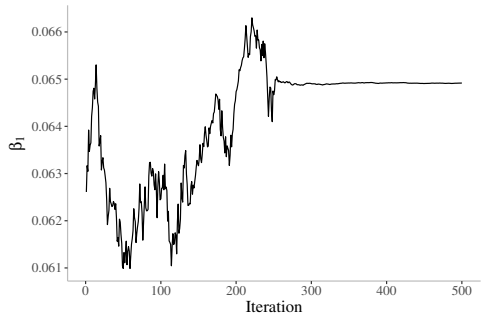
$$\boldsymbol{C}\boldsymbol{\theta} = (\boldsymbol{R}^\top, \mathbf{0}) \times (\boldsymbol{D}^\top, \boldsymbol{Z}^\top)^\top \times \boldsymbol{Z}\boldsymbol{\theta}_z = (\boldsymbol{R}^\top, \mathbf{0}) \times (\mathbf{0}, \boldsymbol{I}_{k-q})\boldsymbol{\theta}_z = \mathbf{0}.$$

That said, we obtain the QR decomposition to $\boldsymbol{C}^\top$ and defines $\boldsymbol{Z}$ as being the $k - q$ last columns of the orthogonal array $\boldsymbol{Q}$. We estimate the parameters of interest using $\tilde{\boldsymbol{T}} = \boldsymbol{T}\boldsymbol{Z}$ and $\tilde{\boldsymbol{\Omega}} = \boldsymbol{Z}^\top\boldsymbol{\Omega}\boldsymbol{Z}$ and obtain $\hat{\boldsymbol{\theta}} = \boldsymbol{Z}\hat{\boldsymbol{\theta}}_z$.

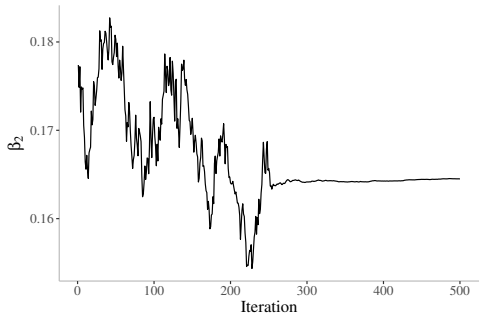
APPENDIX D

Convergence plots

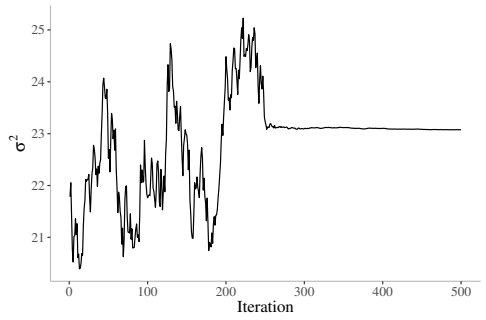
D.1 Convergence plots for ragweed pollen Centered Skew Distribution 2 model



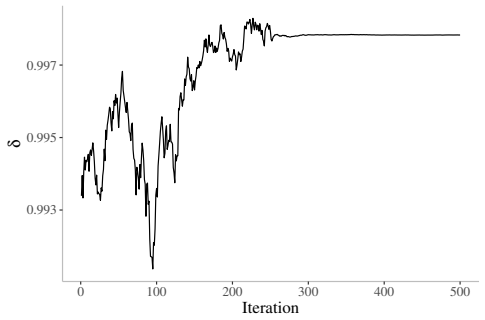
(a) Convergence plot for β_1.



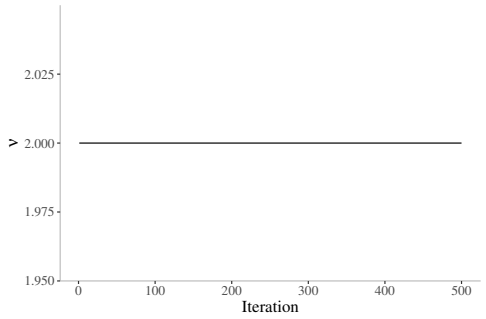
(b) Convergence plot for β_2.



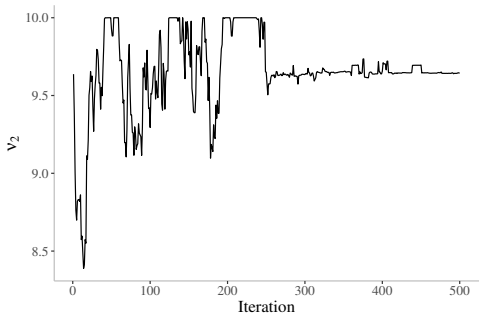
(c) Convergence plot for σ^2.



(d) Convergence plot for γ.



(e) Convergence plot for ν_1.



(f) Convergence plot for ν_2.

Figure 40 – Convergence plots for ragweed pollen Centered Skew Birnbaum-Saunders Normal model parameters.

APPENDIX E

Simulation study results

E.1 Chapter 2: simulation study 1

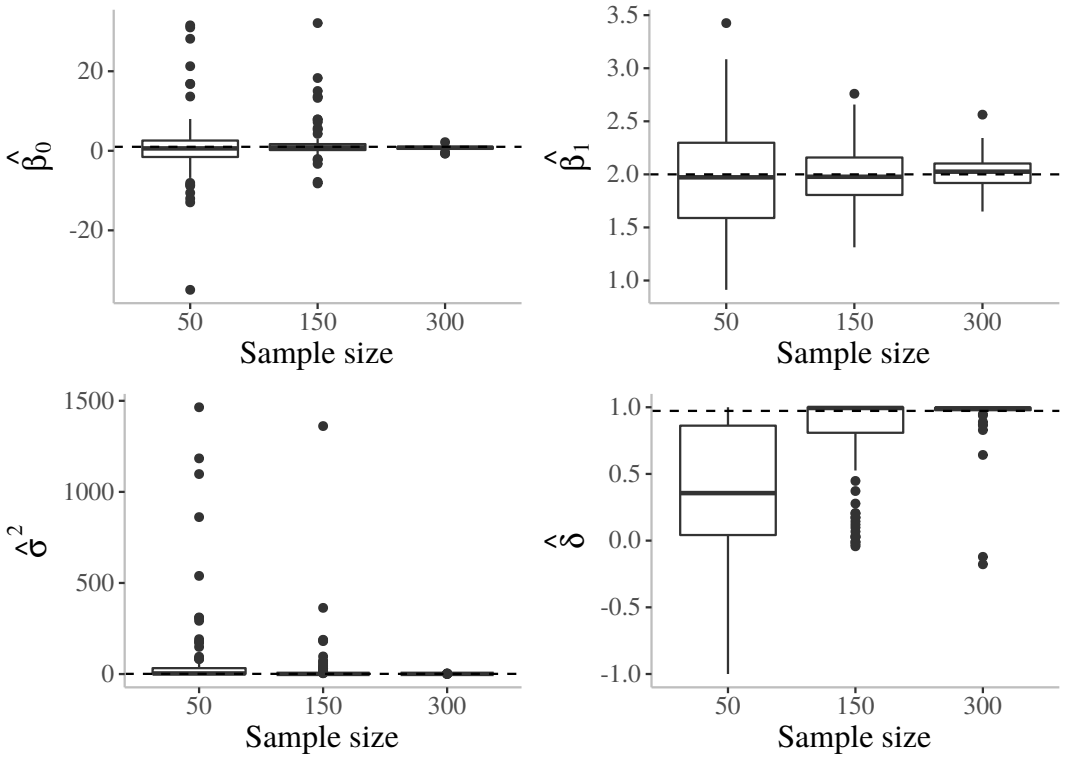


Figure 41 – Box-plots of the bias for Centered Skew Normal model.

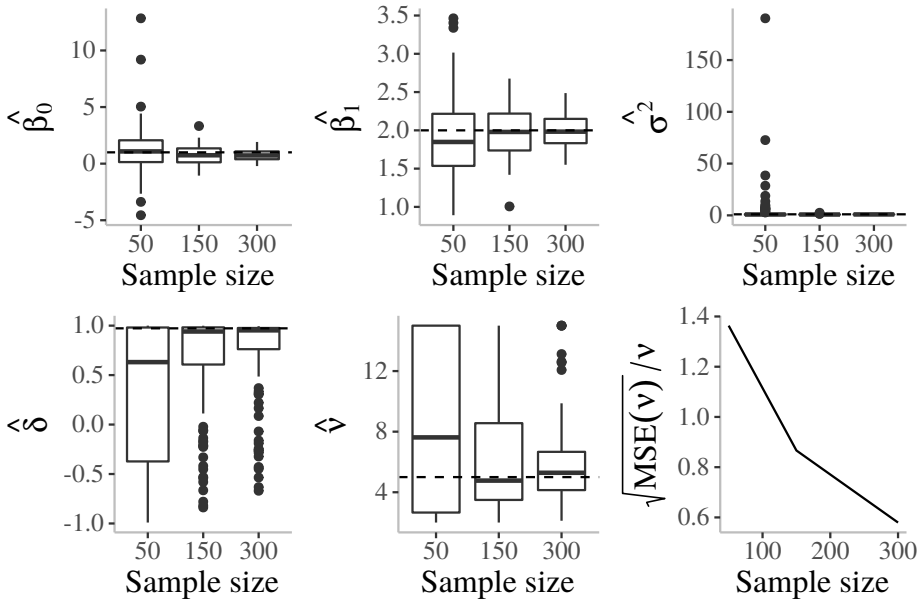


Figure 42 – Box-plots of the bias for Centered Skew-t model.

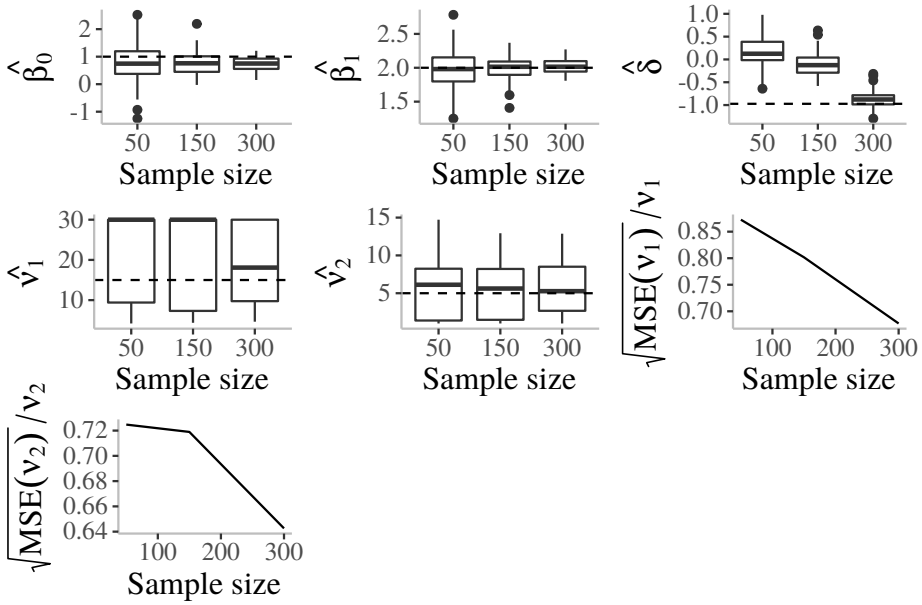


Figure 43 – Box-plots of the bias for Centered Skew Generalized t model.

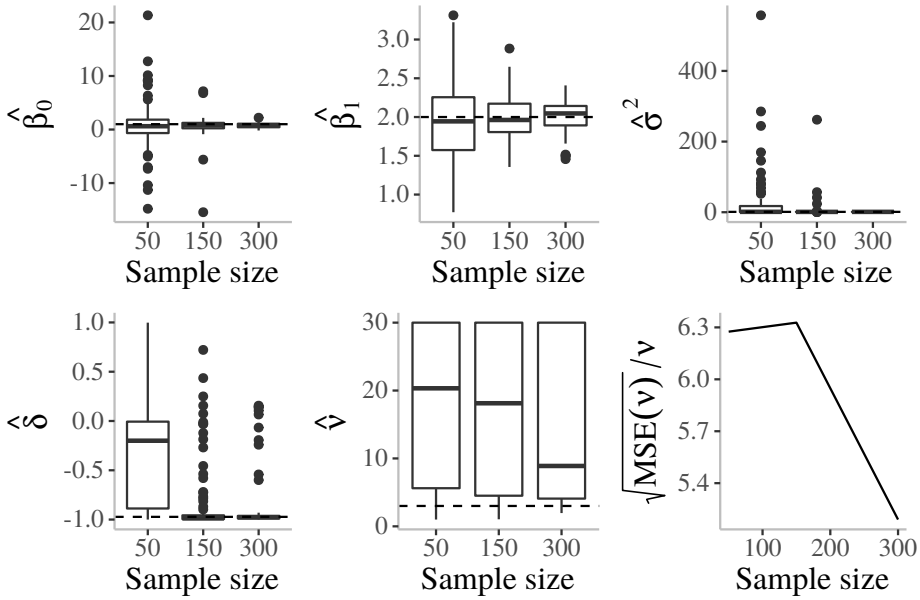


Figure 44 – Box-plots of the bias for Centered Skew Slash model.

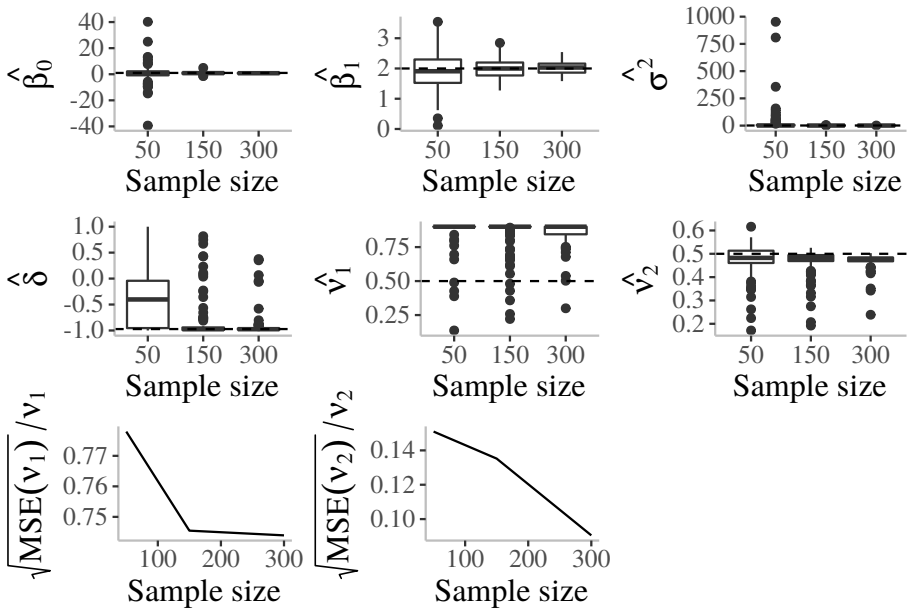


Figure 45 – Box-plots of the bias for Centered Skew Contamined Normal model.

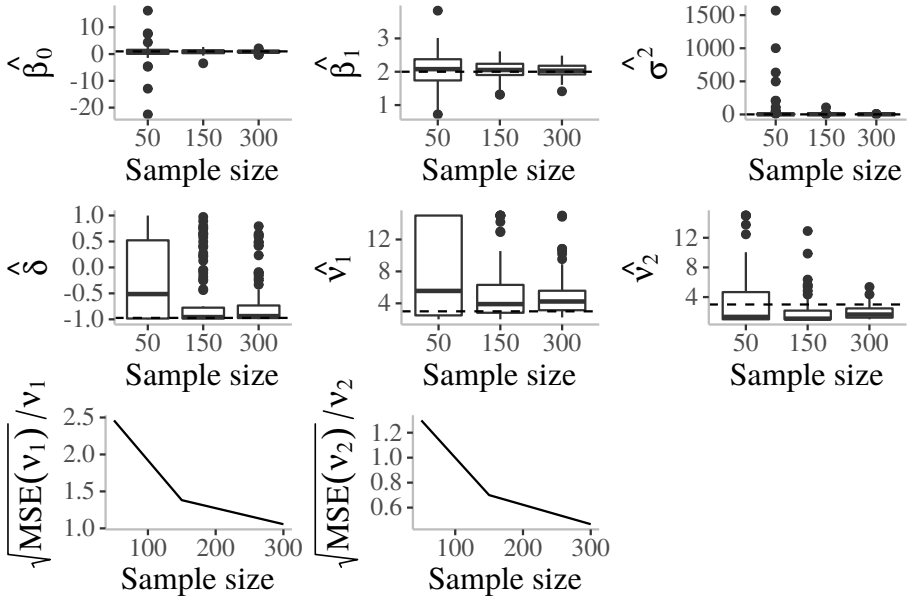


Figure 46 – Box-plots of the bias for Centered Skew Beta Prime Normal model.

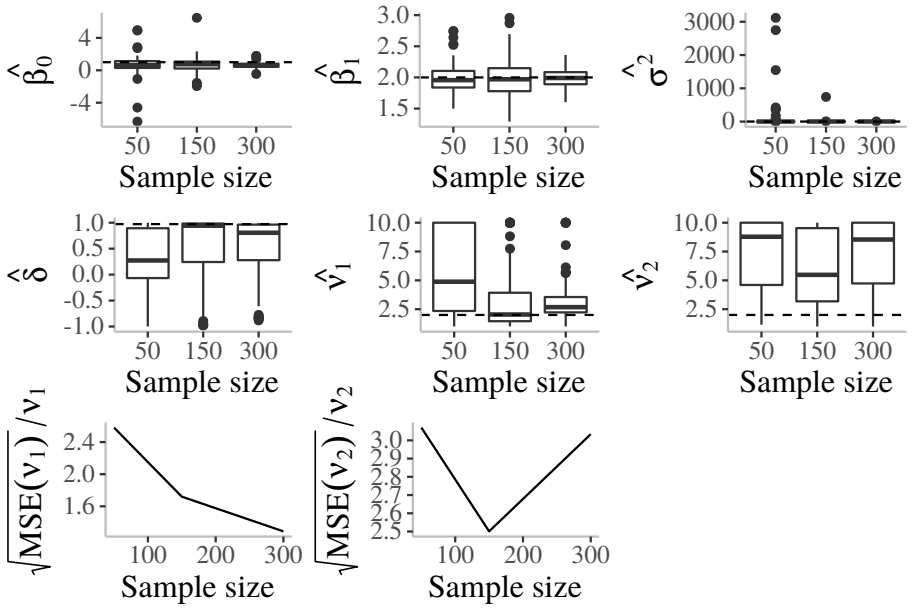


Figure 47 – Box-plots of the bias for Centered Skew Birnbaum-Saunders Normal model.

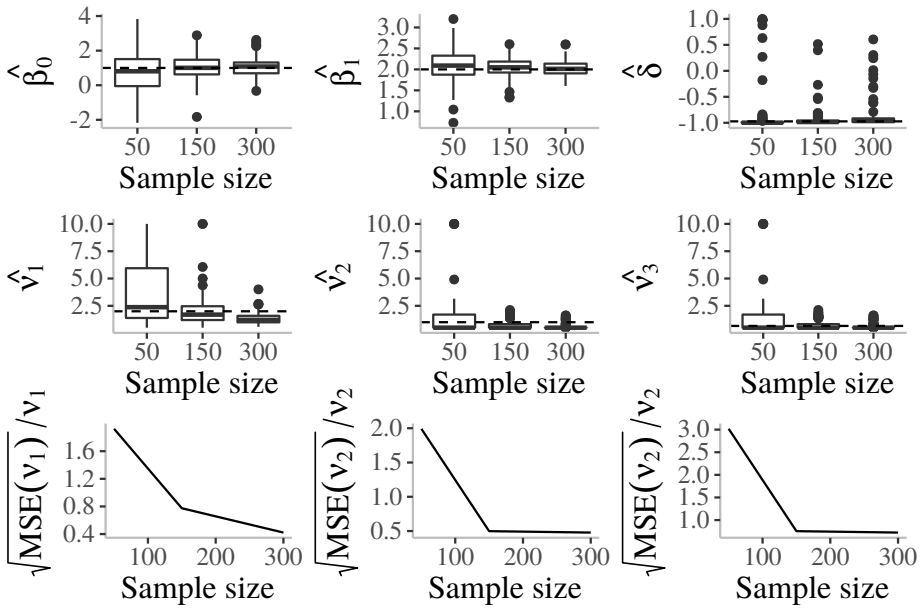


Figure 48 – Box-plots of the bias for Centered Skew Generalized Gamma Normal model.

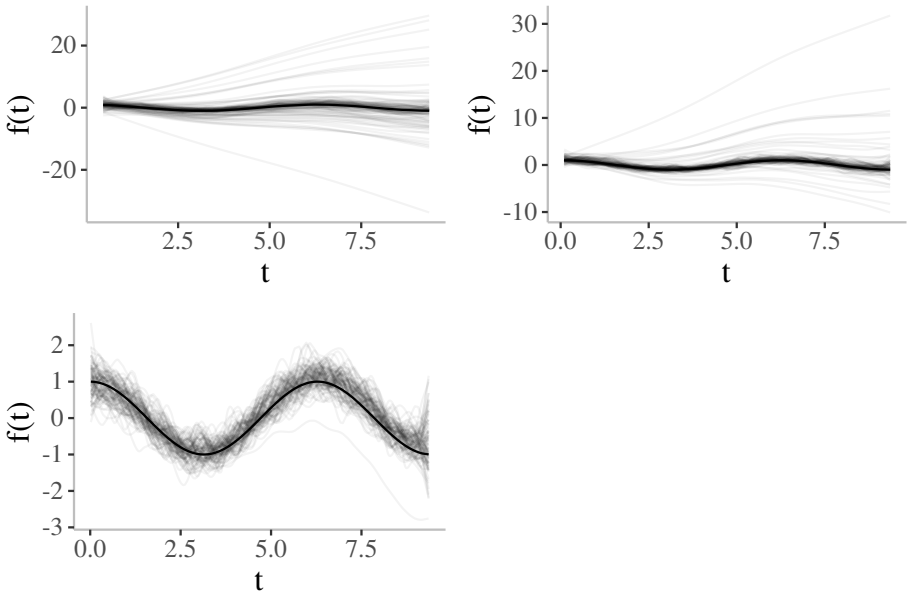


Figure 49 – fitted curves (gray lines) and actual curves (black lines) for Centered Skew Normal model.

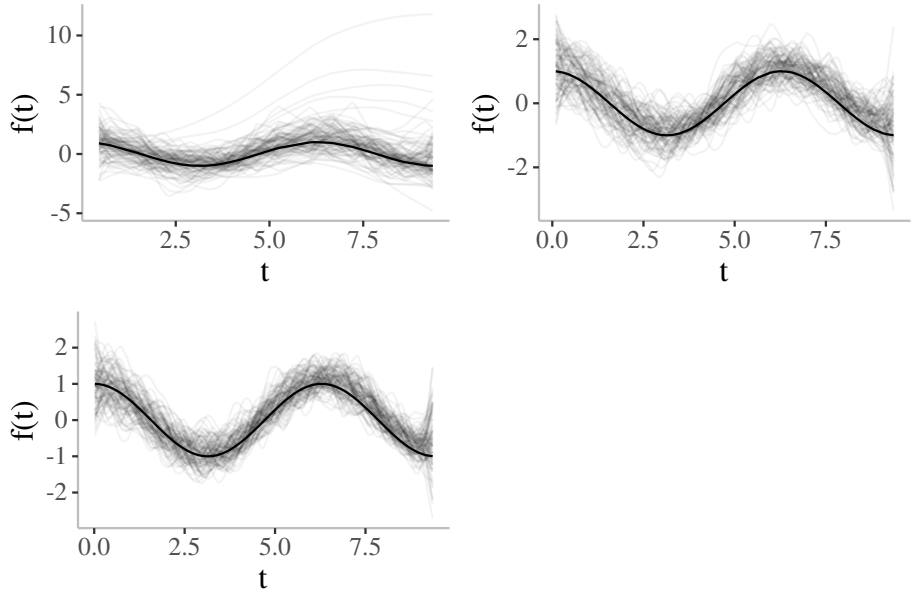


Figure 50 – fitted curves (gray lines) and actual curves (black lines) for Centered Skew-t model.

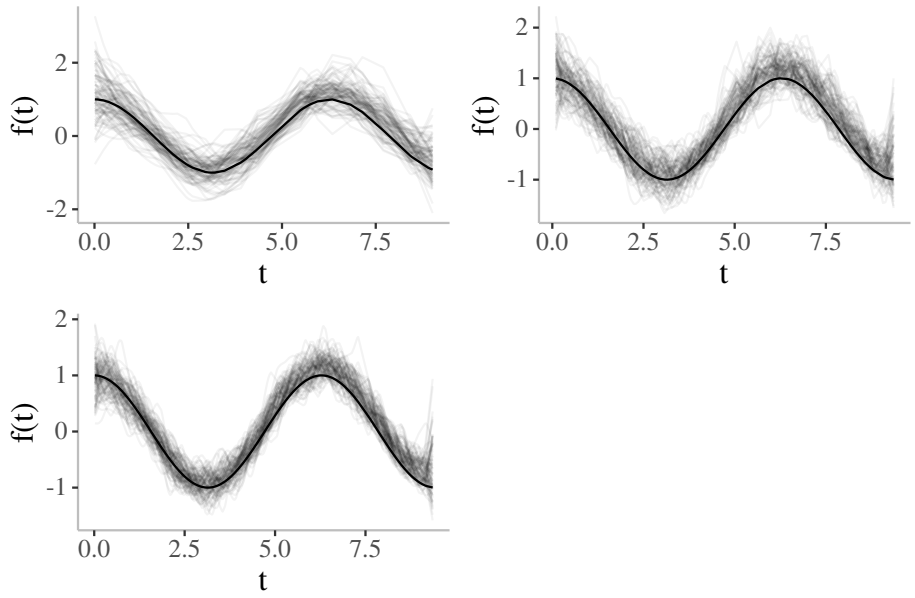


Figure 51 – fitted curves (gray lines) and actual curves (black lines) for Centered Skew Generalized t model.

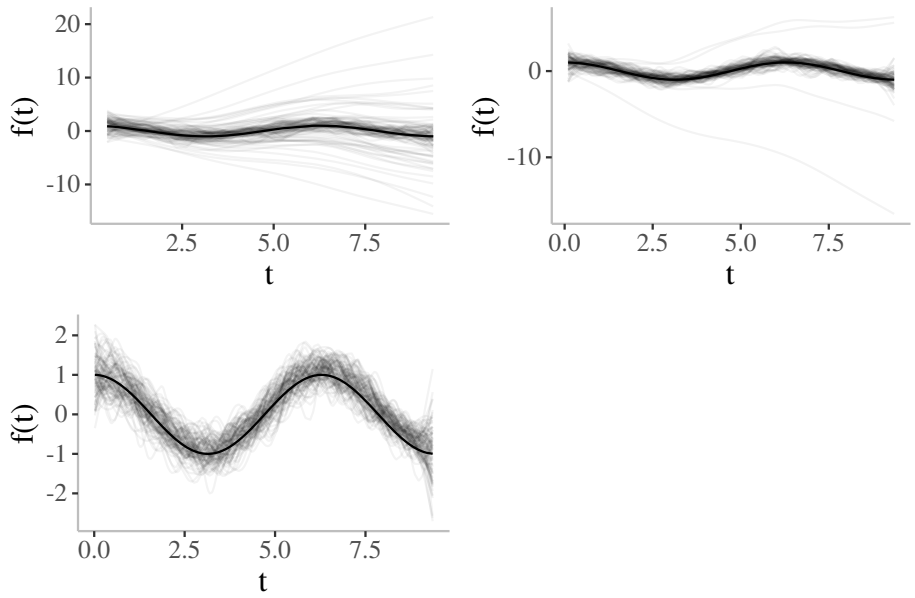


Figure 52 – fitted curves (gray lines) and actual curves (black lines) for Centered Skew Slash model.

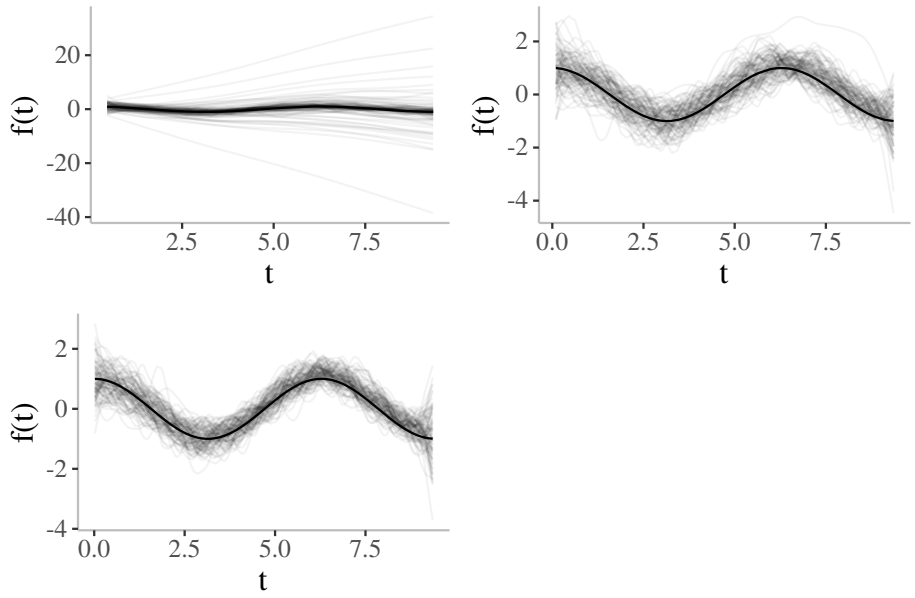


Figure 53 – fitted curves (gray lines) and actual curves (black lines) for Centered Skew Contamined Normal model.

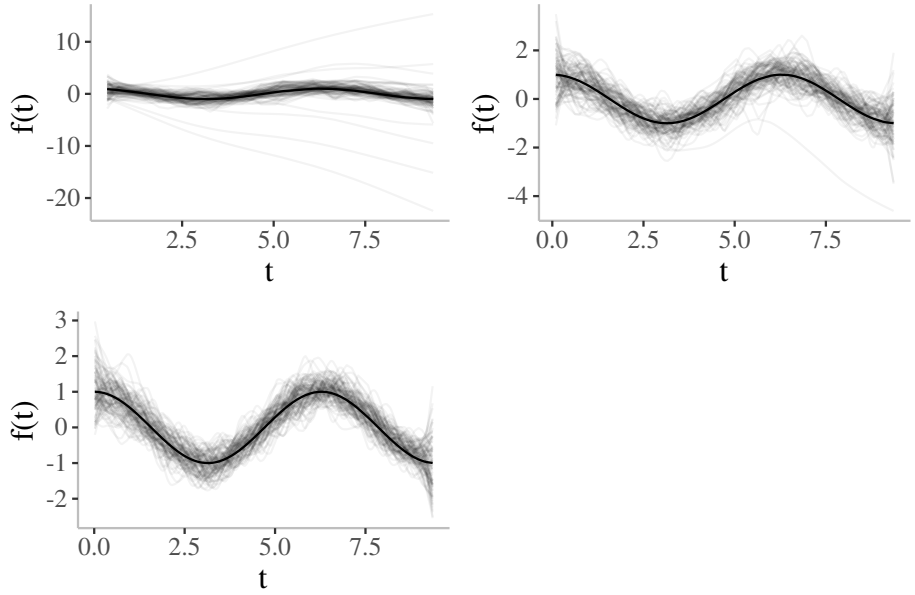


Figure 54 – fitted curves (gray lines) and actual curves (black lines) for Centered Skew Beta Prime Normal model.

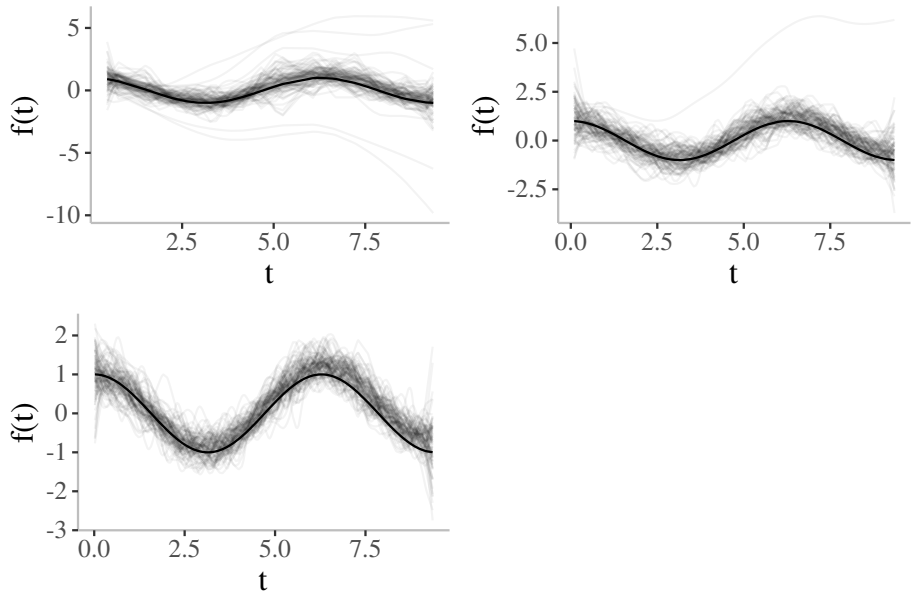


Figure 55 – fitted curves (gray lines) and actual curves (black lines) for Centered Skew Birnbaum-Saunders Normal model.

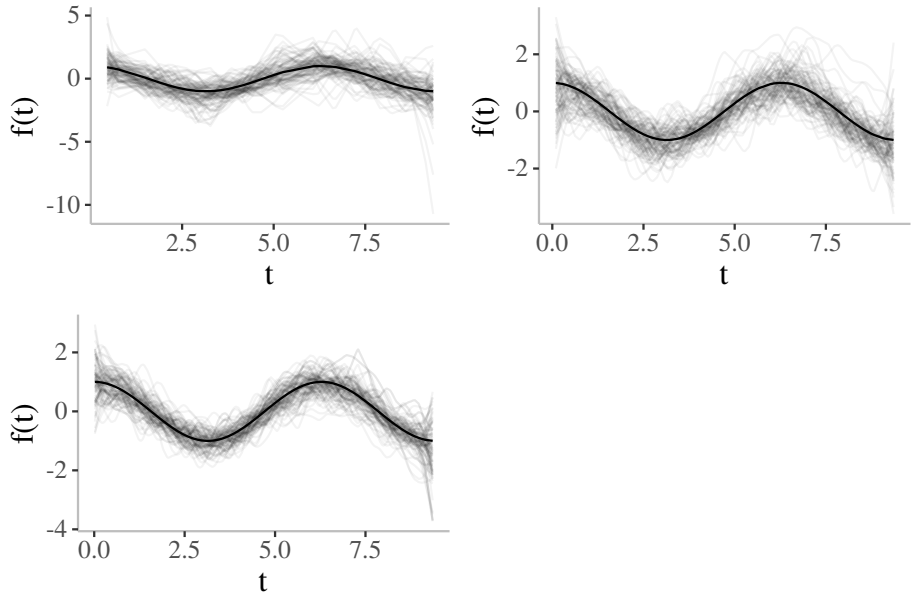


Figure 56 – fitted curves (gray lines) and actual curves (black lines) for Centered Skew Generalized Gamma Normal model.

E.2 Chapter 3

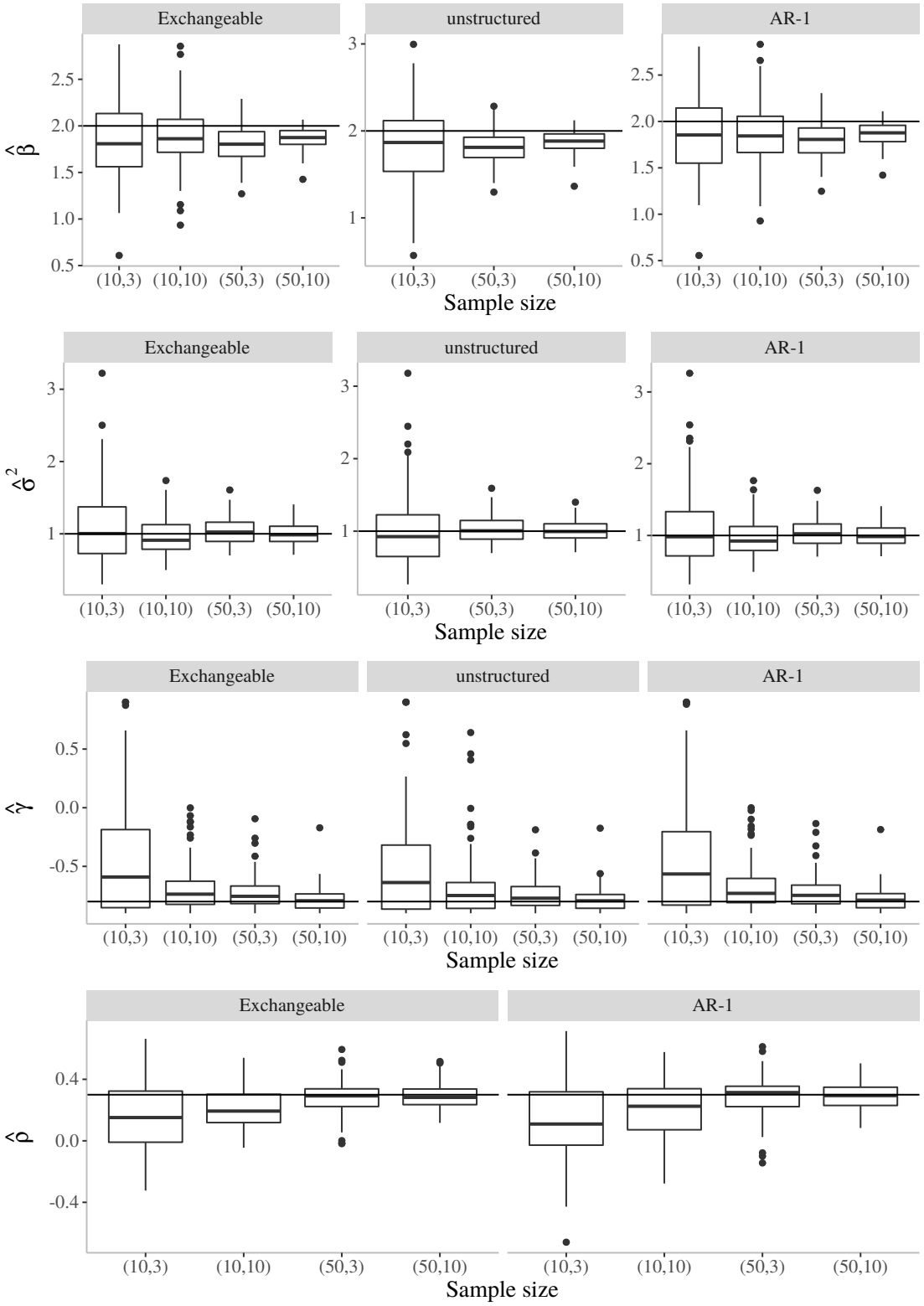


Figure 57 – Simulation study: estimated parameters for GEE-based CSN model with $\rho = 0.3$ (exchangeable) by working correlation matrices.

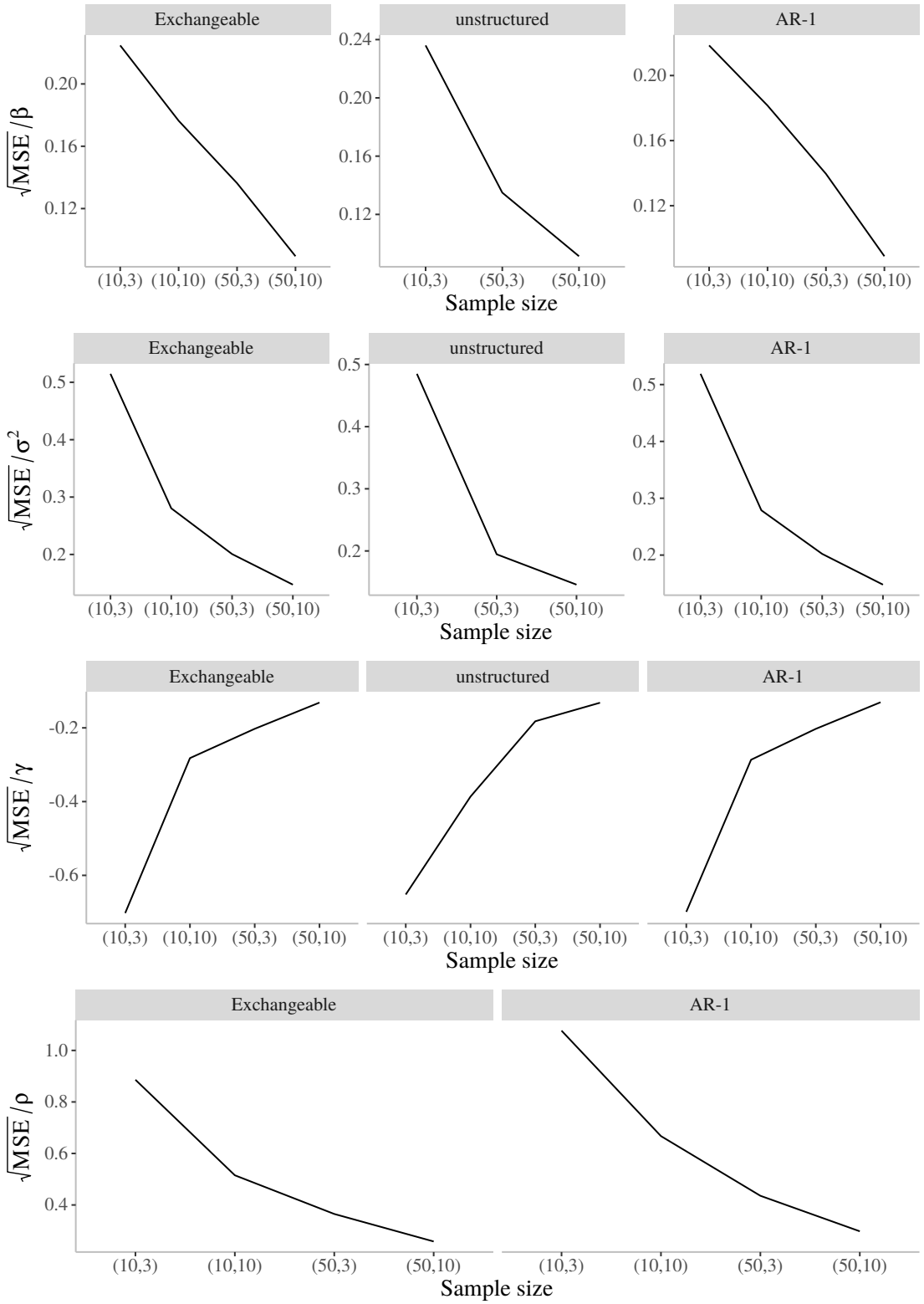


Figure 58 – Simulation study: relative mean square error of the parameters for GEE-based CSN model with $\rho = 0.3$ (exchangeable) by working correlation matrices.

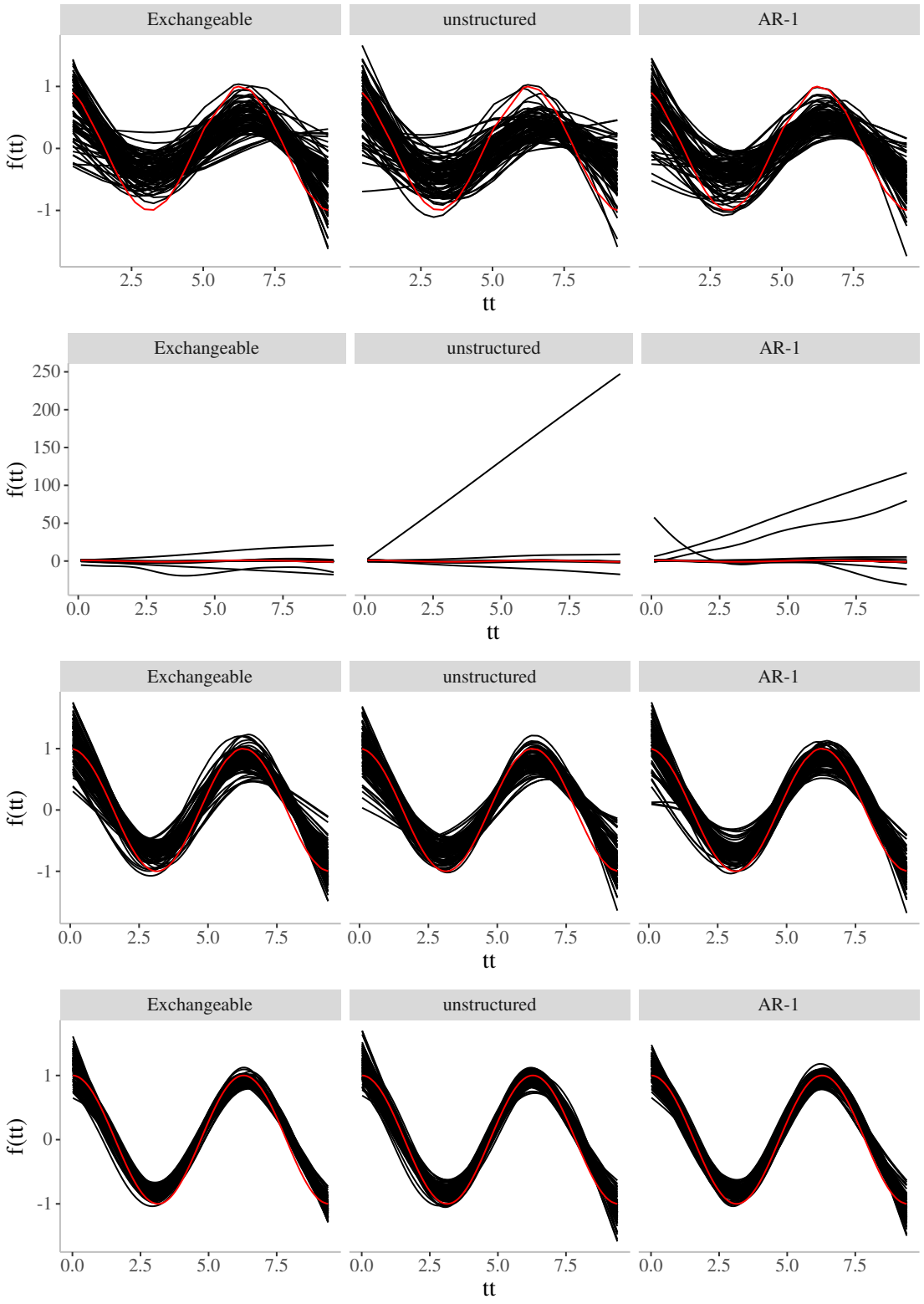


Figure 59 – Simulation study: nonparametric curves for GEE-based CSN model for (10,3), (10,10), (50,3) and (50,10), respectively, with $\rho = 0.3$ (exchangeable) by working correlation matrices.

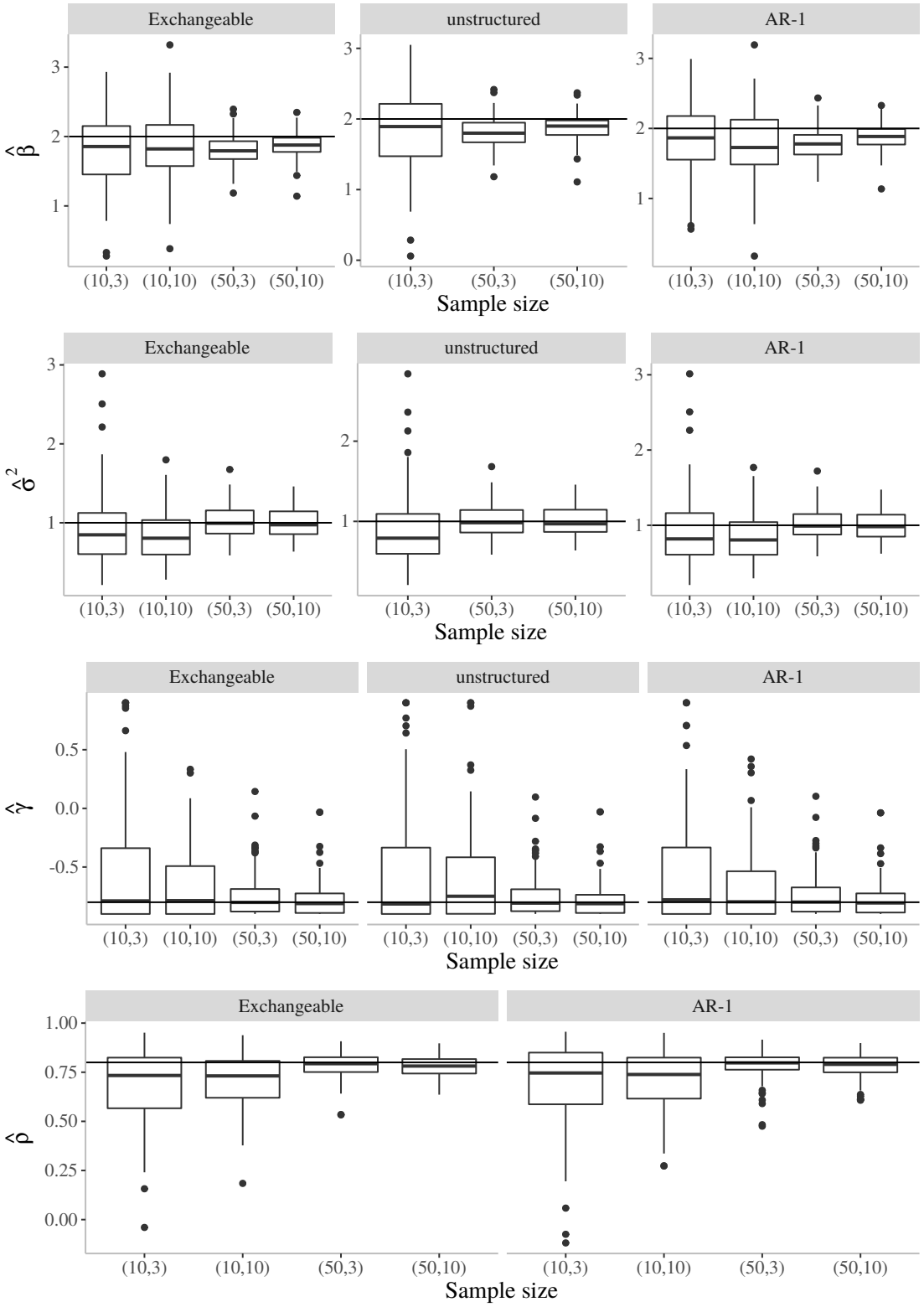


Figure 60 – Simulation study: estimated parameters for GEE-based CSN model with $\rho = 0.8$ (exchangeable) by working correlation matrices.

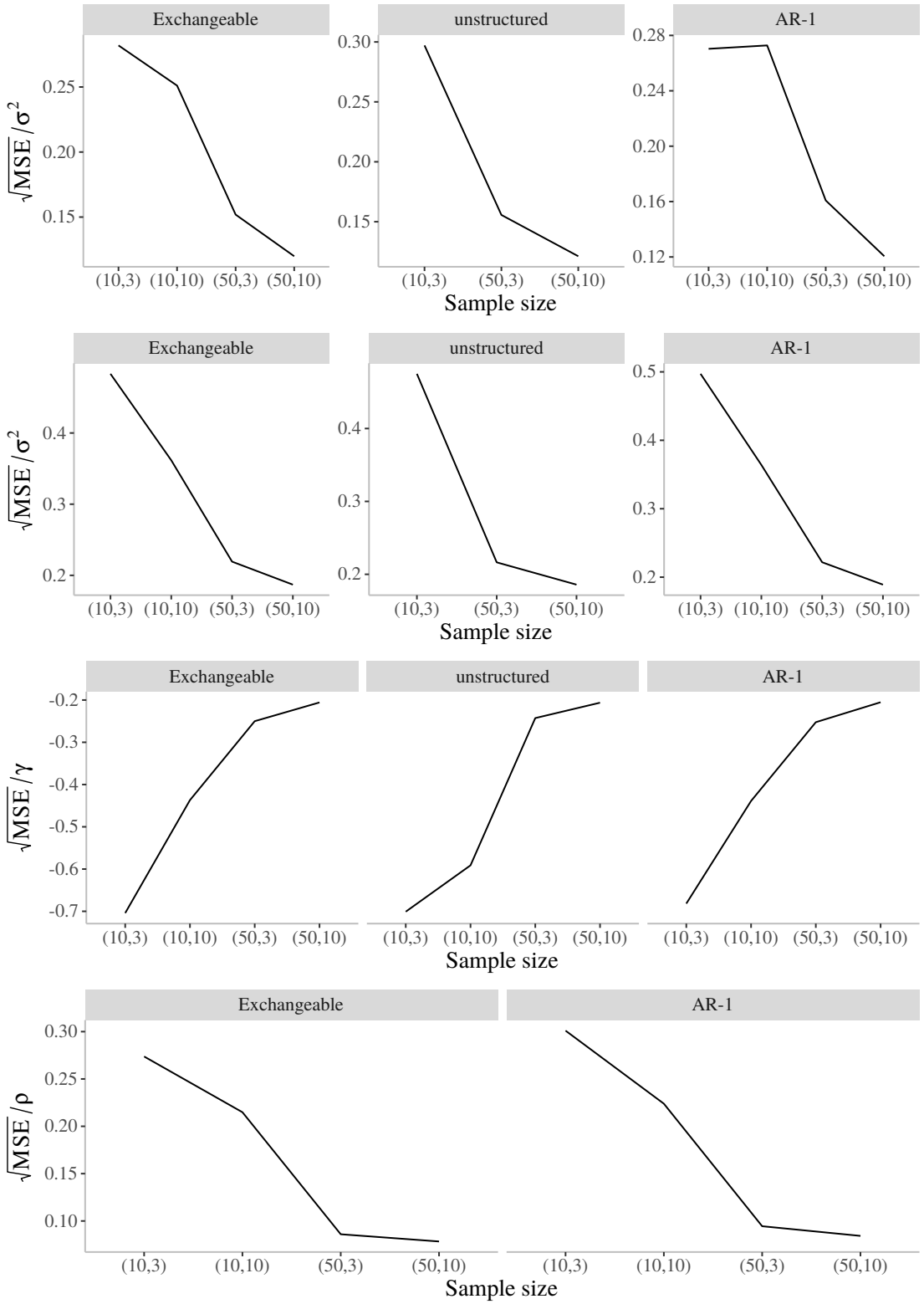


Figure 61 – Simulation study: relative mean square error of the parameters for GEE-based CSN model with $\rho = 0.8$ (exchangeable) by working correlation matrices.

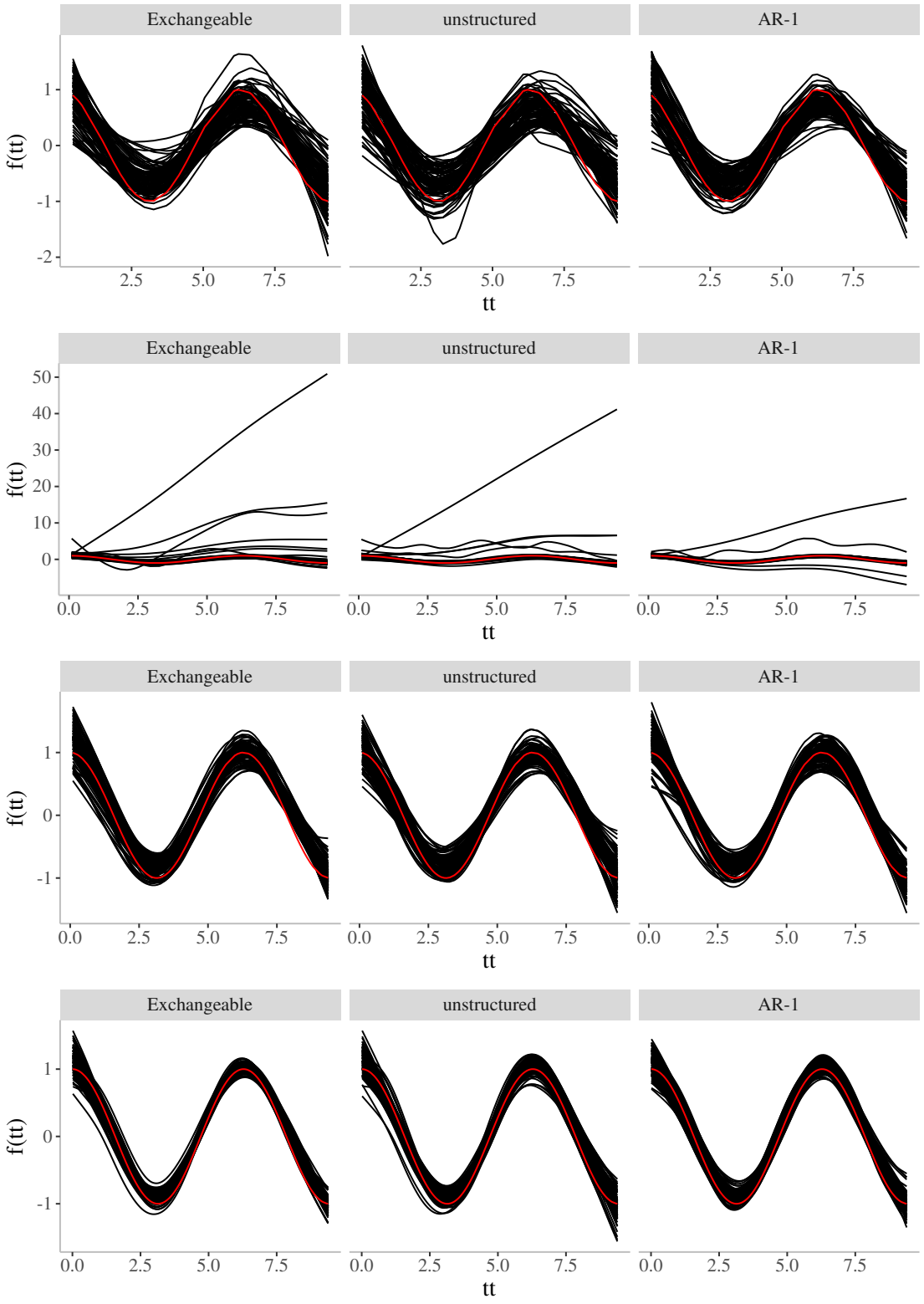


Figure 62 – Simulation study: nonparametric curves for GEE-based CSN model for (10,3), (10,10), (50,3) and (50,10), respectively, with $\rho = 0.8$ (exchangeable) by working correlation matrices.

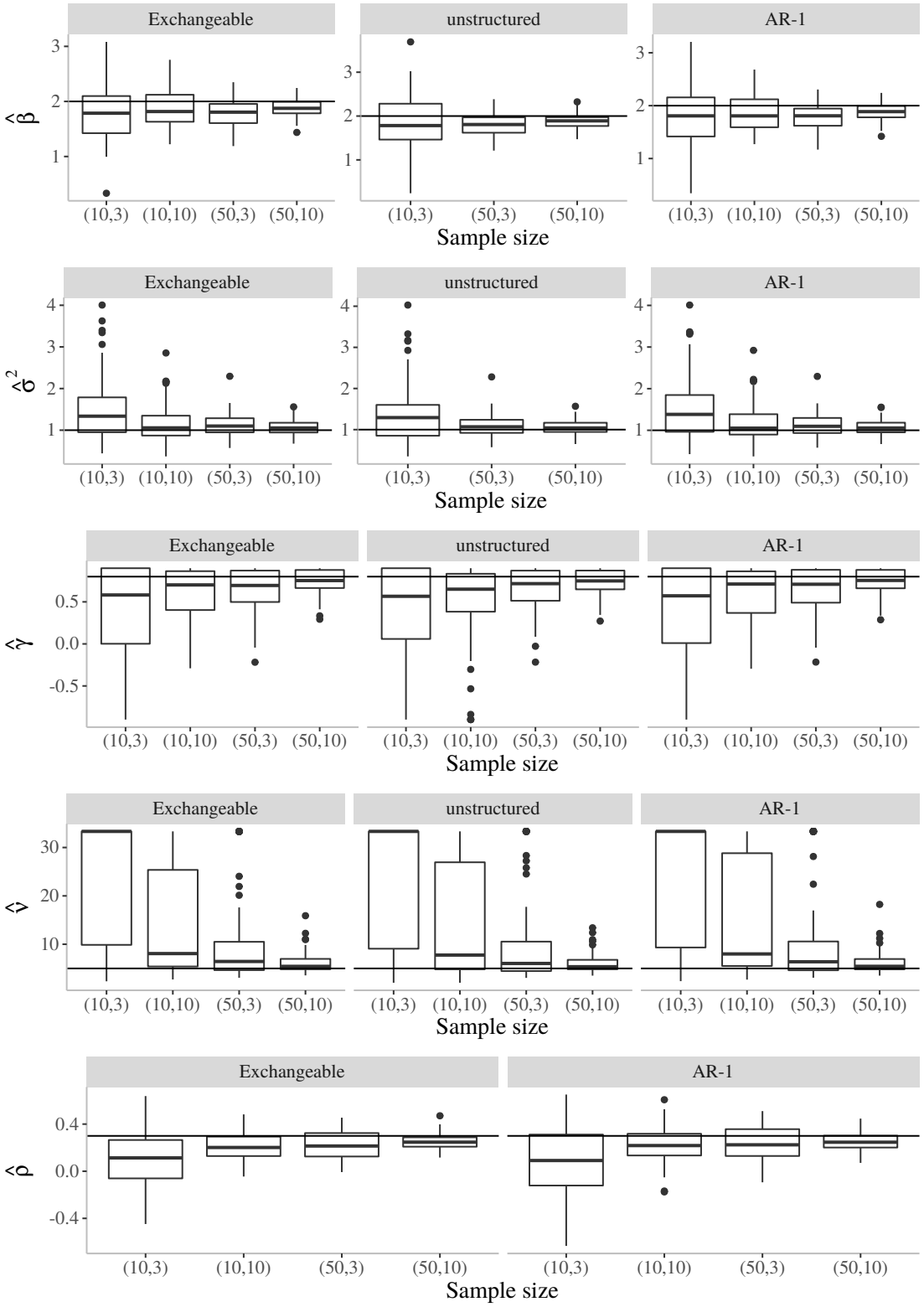


Figure 63 – Simulation study: estimated parameters for GEE-based CST model with $\rho = 0.3$ (exchangeable) by working correlation matrices.

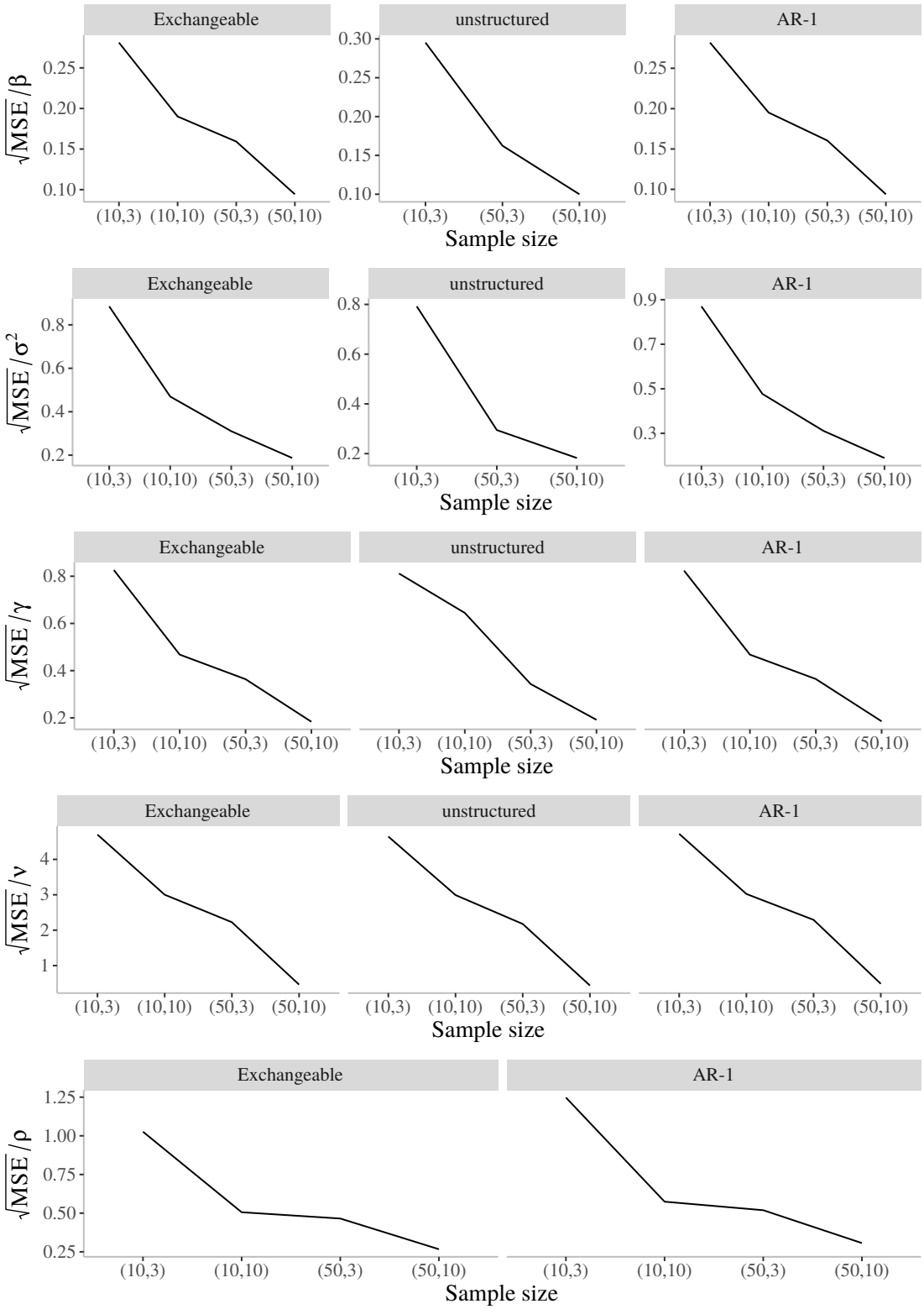


Figure 64 – Simulation study: relative mean square error of the parameters for GEE-based CST model with $\rho = 0.3$ (exchangeable) by working correlation matrices.

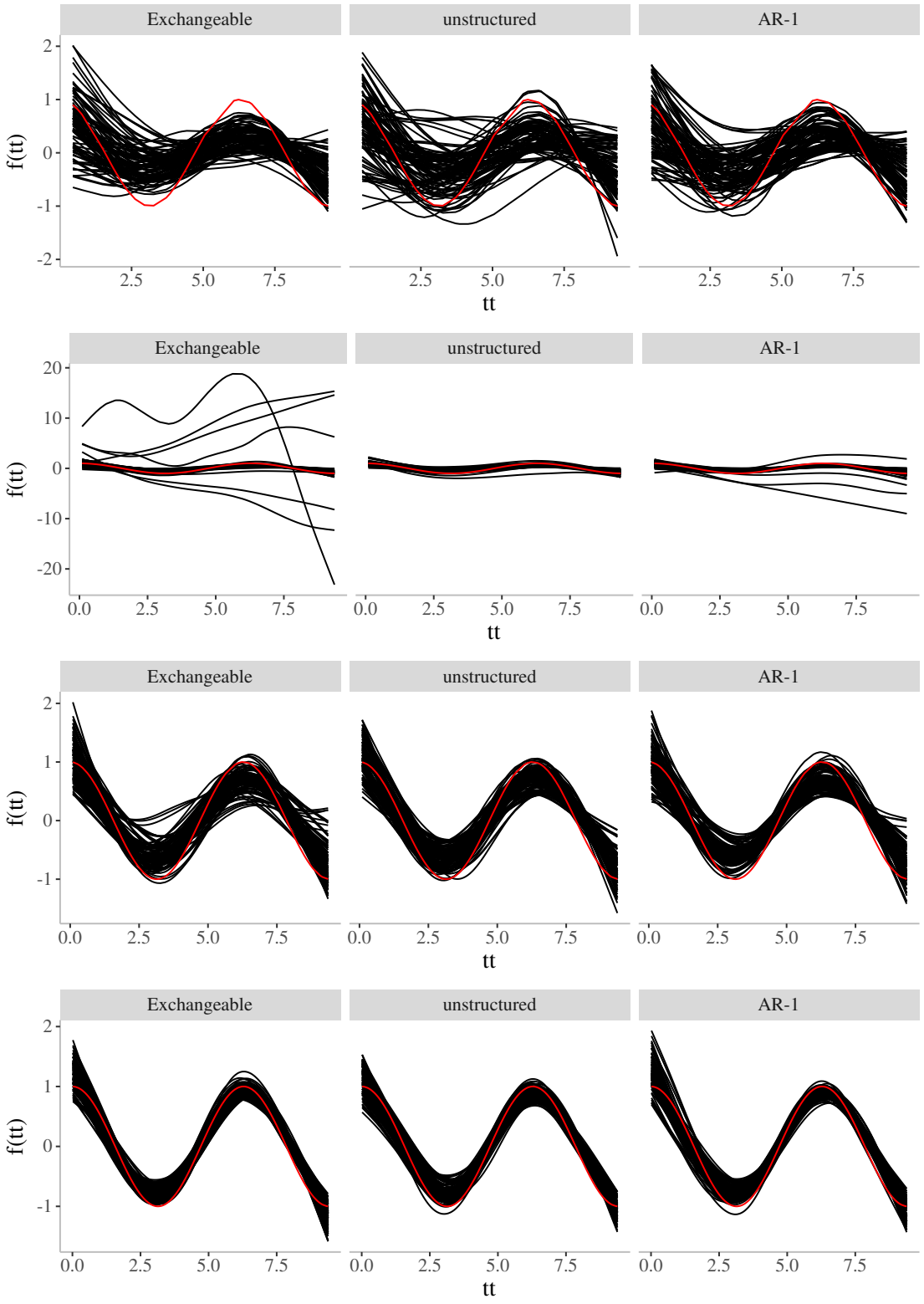


Figure 65 – Simulation study: nonparametric curves for GEE-based CST model for (10,3), (10,10), (50,3) and (50,10), respectively, with $\rho = 0.8$ (exchangeable) by working correlation matrices.

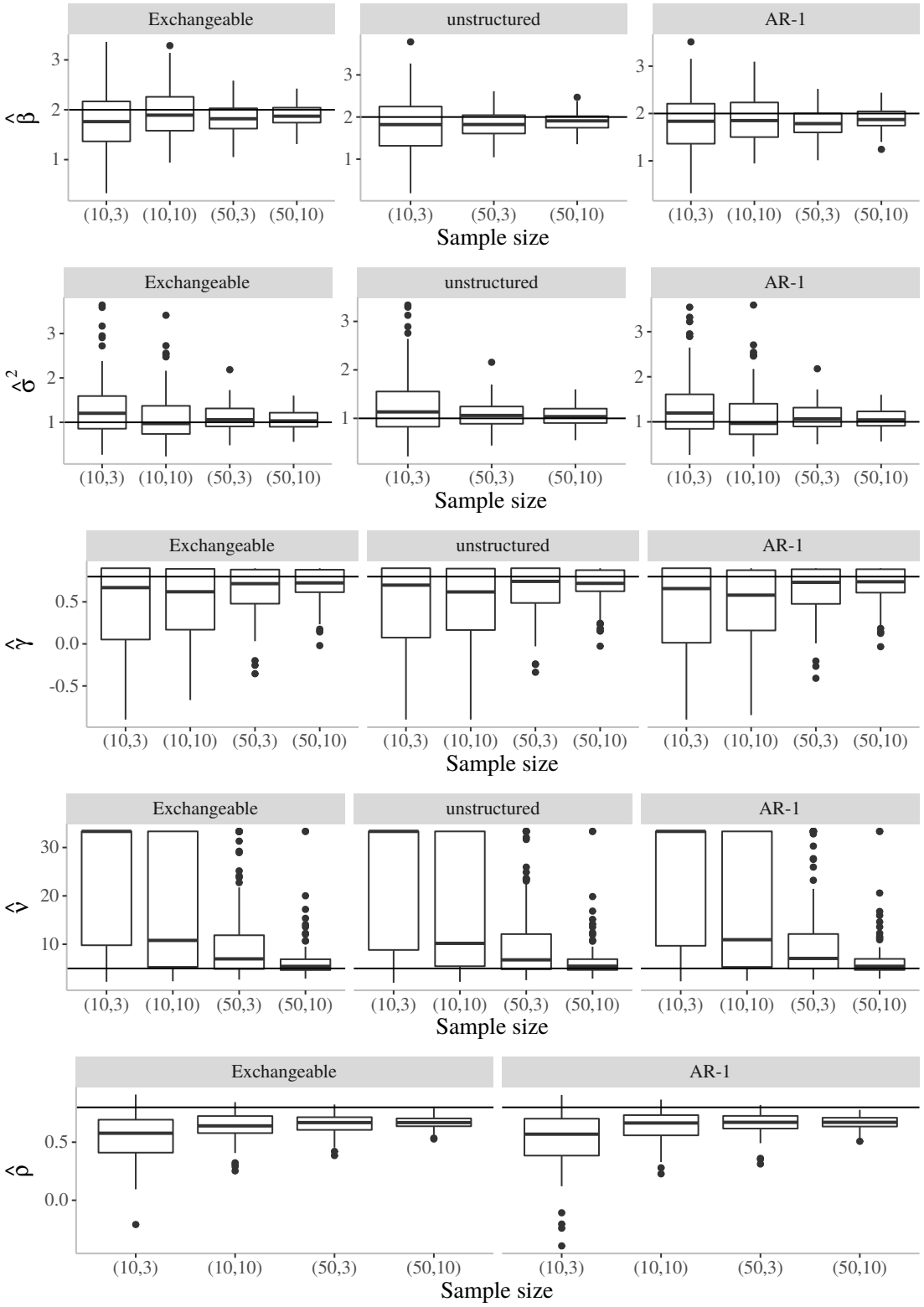


Figure 66 – Simulation study: estimated parameters for GEE-based CST model with $\rho = 0.8$ (exchangeable) by working correlation matrices.

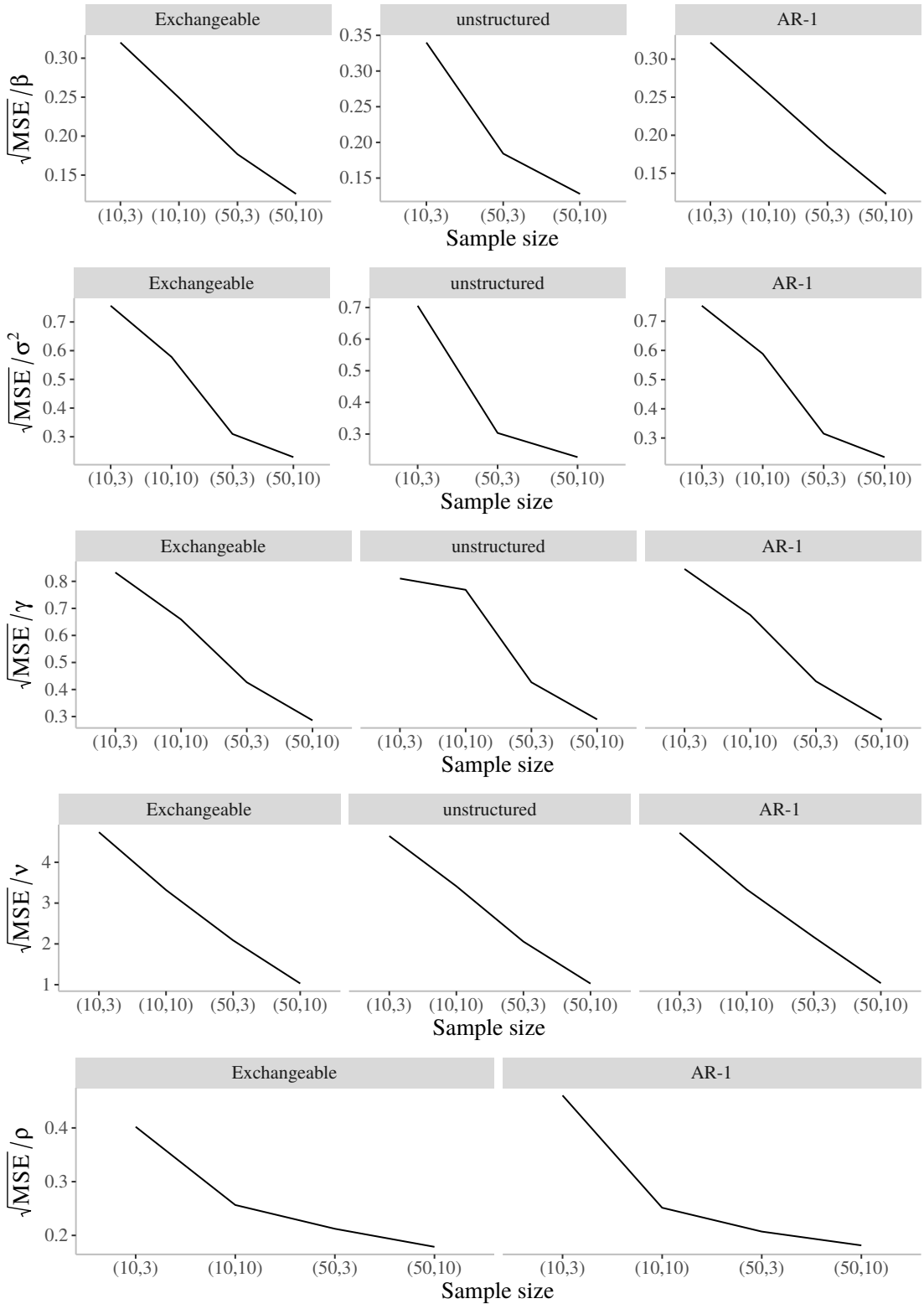


Figure 67 – Simulation study: relative mean square error of the parameters for GEE-based CST model with $\rho = 0.8$ (exchangeable) by working correlation matrices.

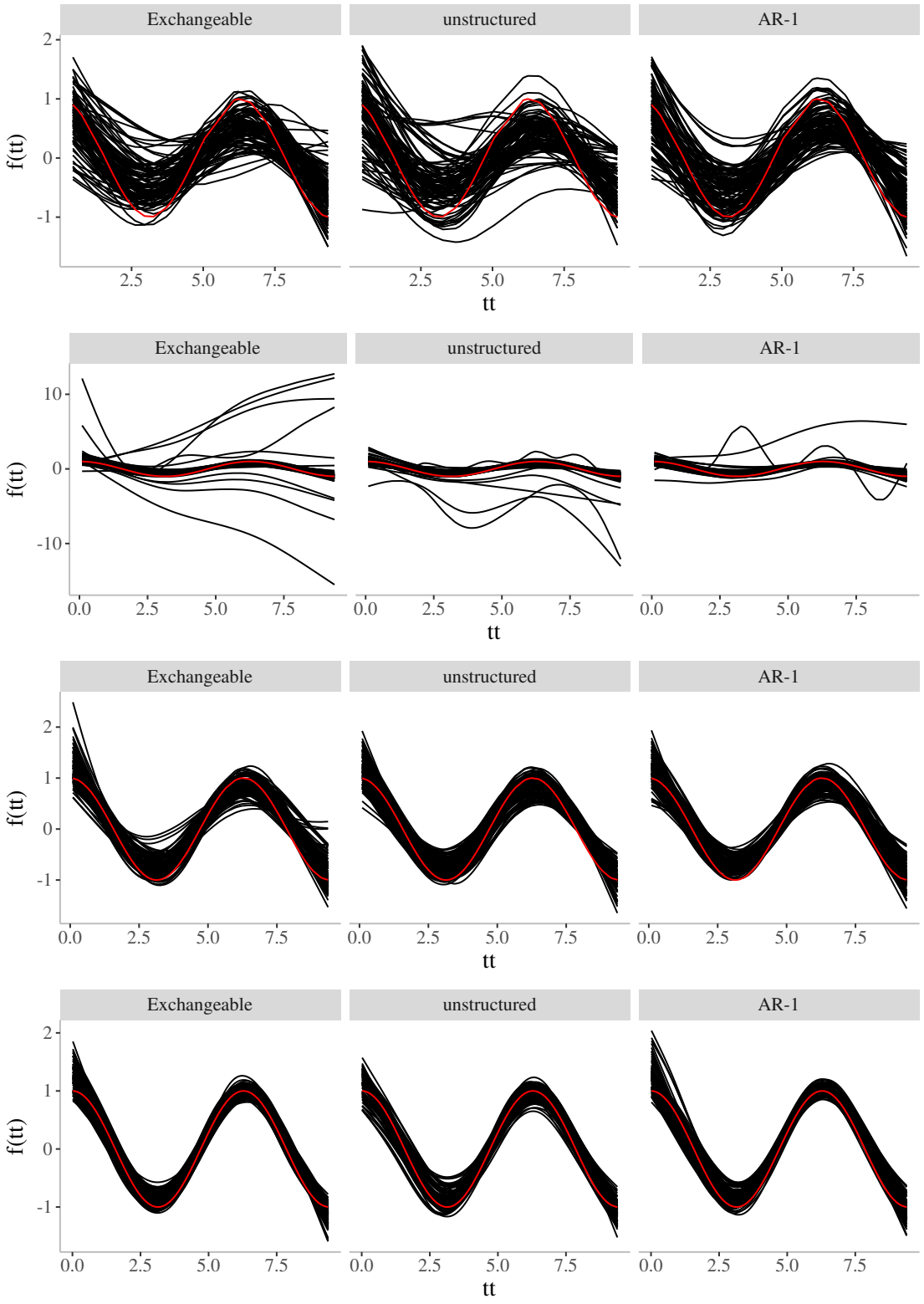


Figure 68 – Simulation study: nonparametric curves for GEE-based CST model for (10,3), (10,10), (50,3) and (50,10), respectively, with $\rho = 0.8$ (exchangeable) by working correlation matrices.

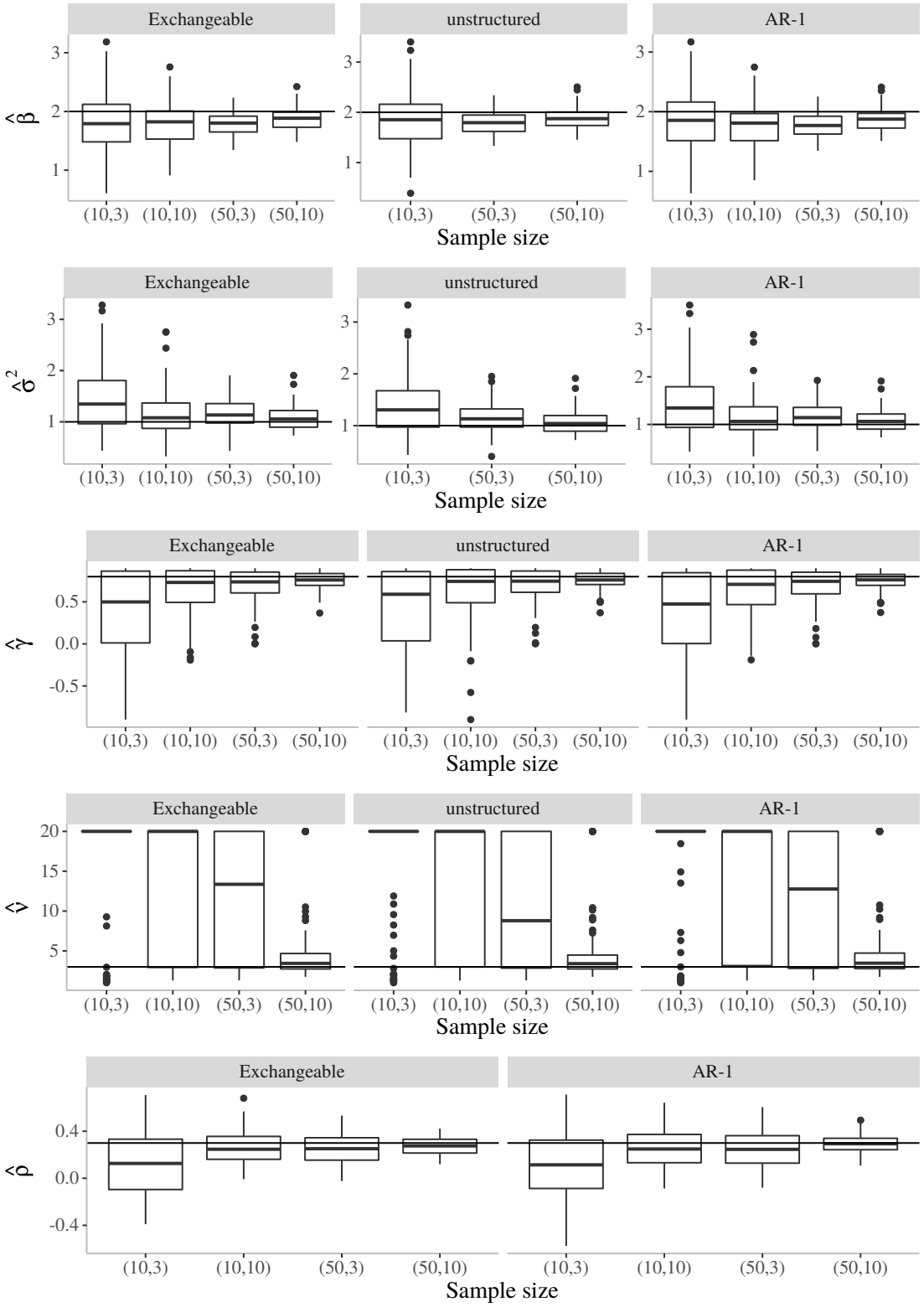


Figure 69 – Simulation study: estimated parameters for GEE-based CSS model with $\rho = 0.3$ (exchangeable) by working correlation matrices.

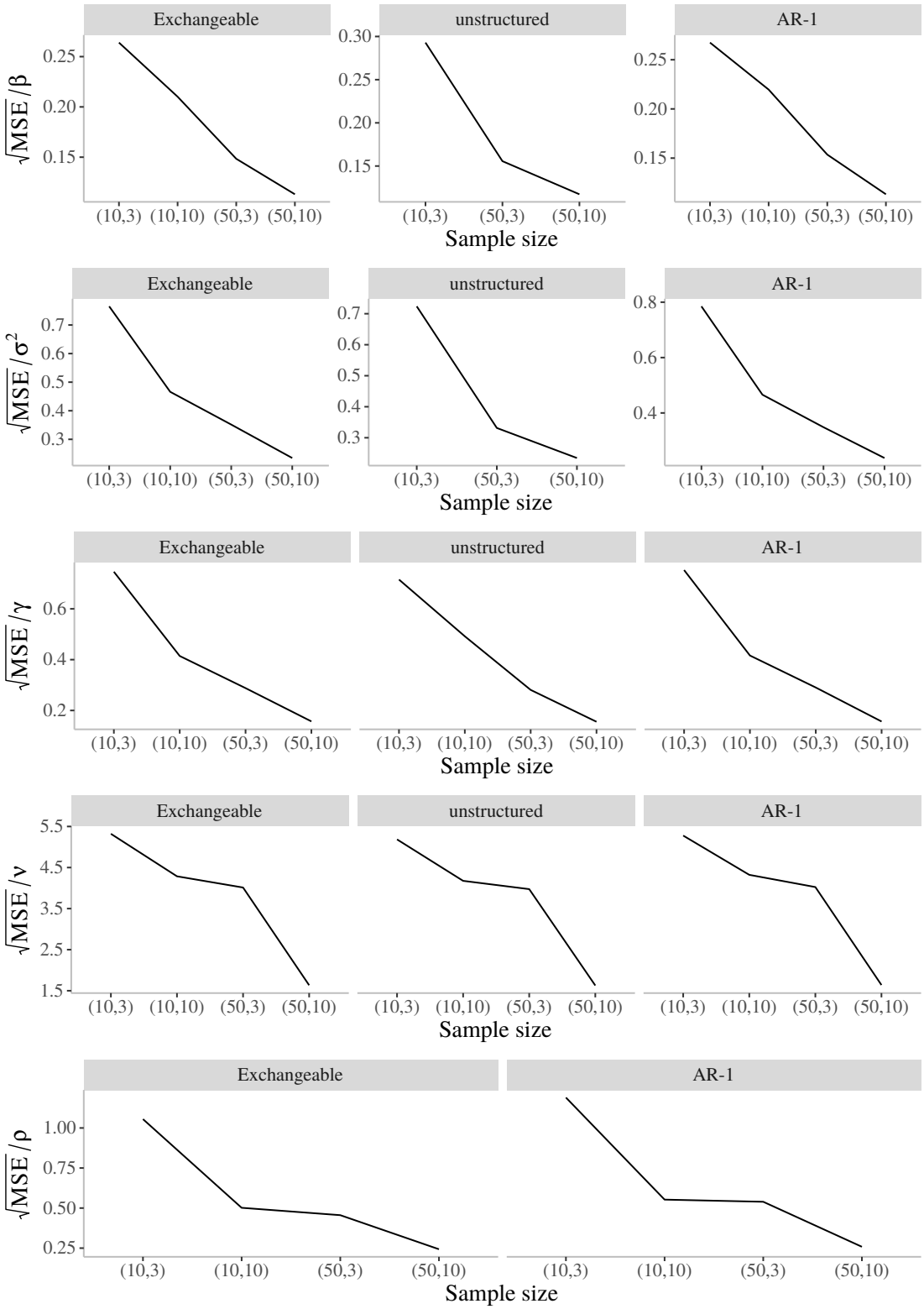


Figure 70 – Simulation study: relative mean square error of the parameters for GEE-based CSS model with $\rho = 0.3$ (exchangeable) by working correlation matrices.

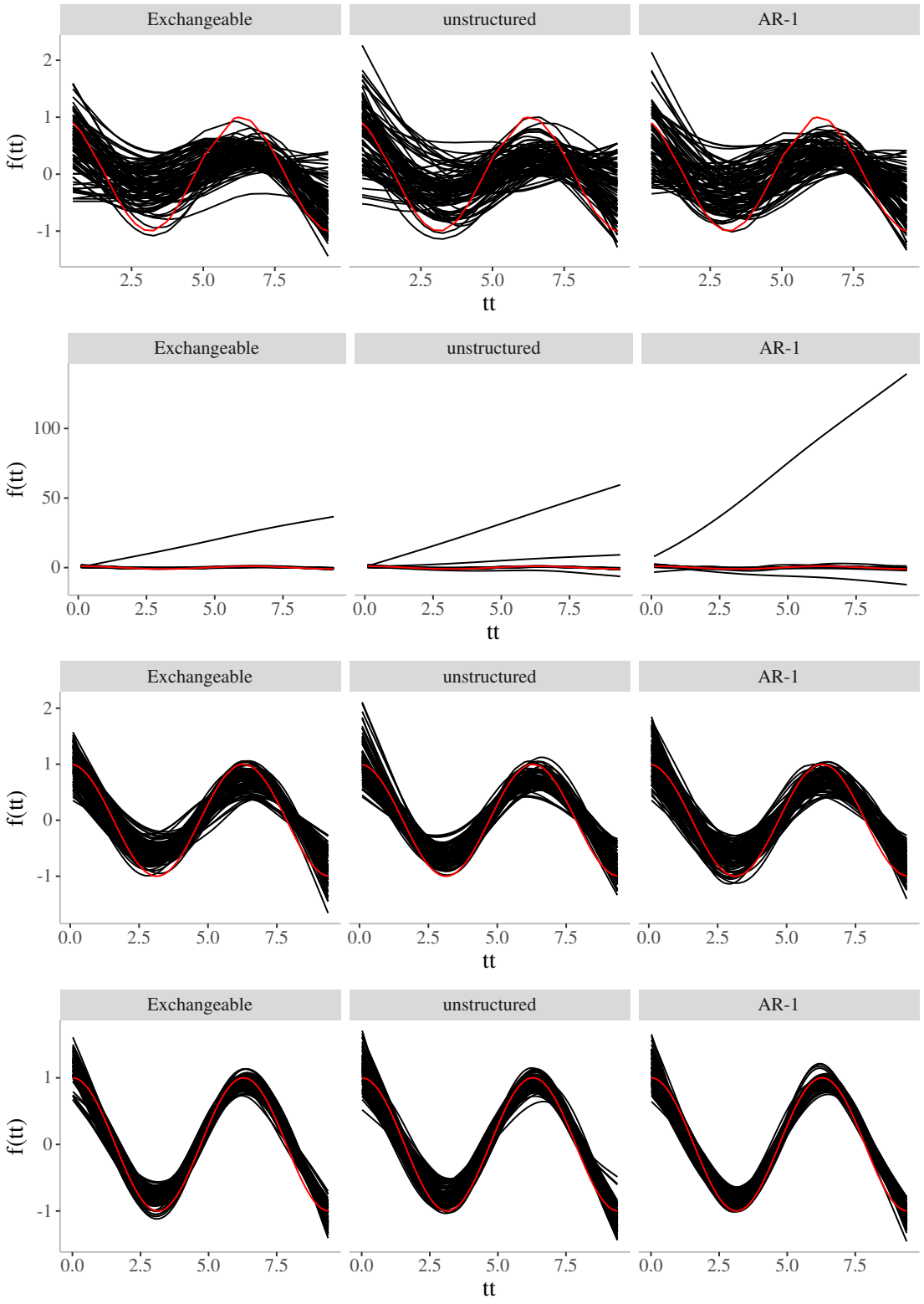


Figure 71 – Simulation study: nonparametric curves for GEE-based CSS model for (10,3), (10,10), (50,3) and (50,10), respectively, with $\rho = 0.3$ (exchangeable) by working correlation matrices.

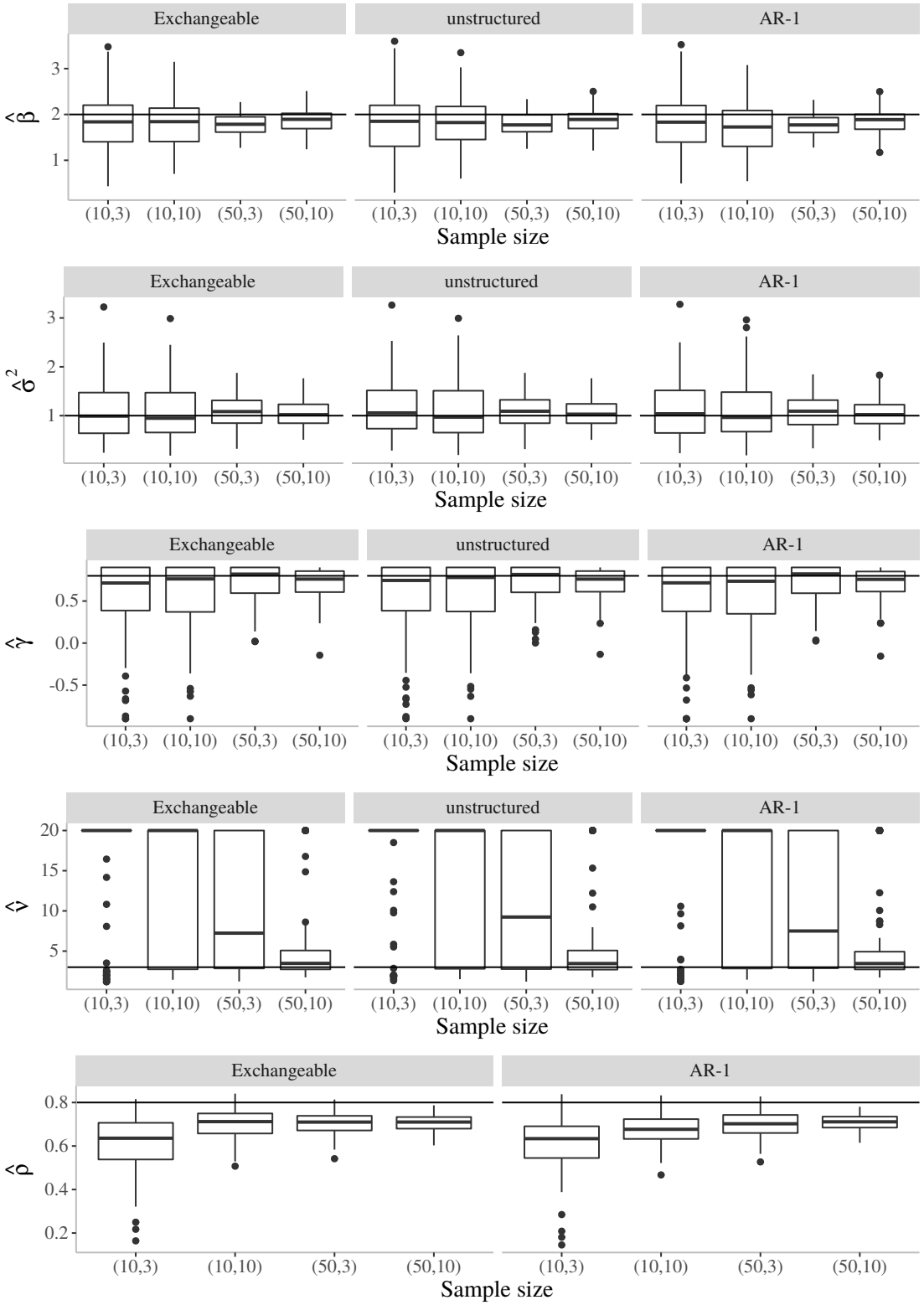


Figure 72 – Simulation study: estimated parameters for GEE-based CSS model with $\rho = 0.8$ (exchangeable) by working correlation matrices.

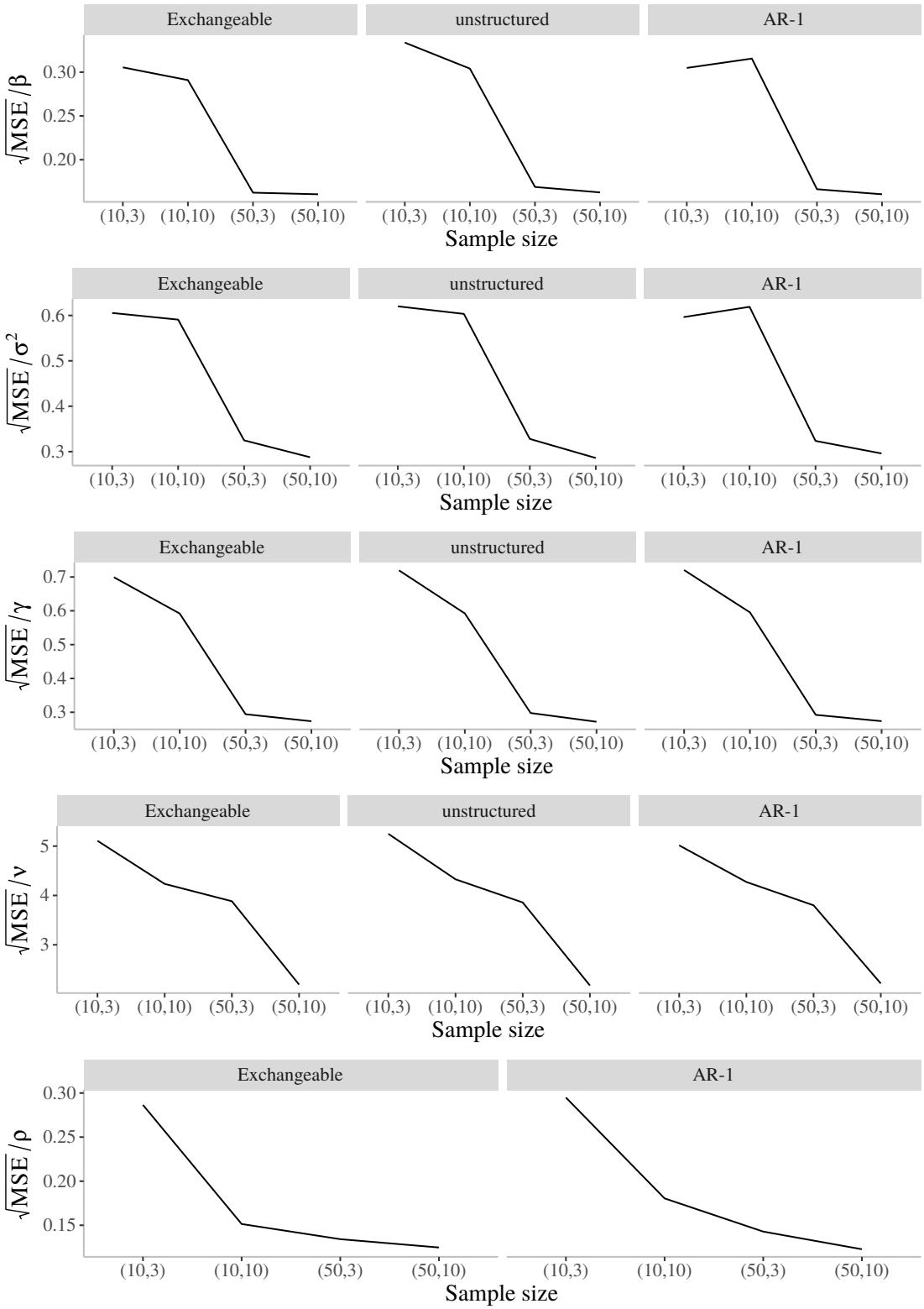


Figure 73 – Simulation study: relative mean square error of the parameters for GEE-based CSS model with $\rho = 0.8$ (exchangeable) by working correlation matrices.

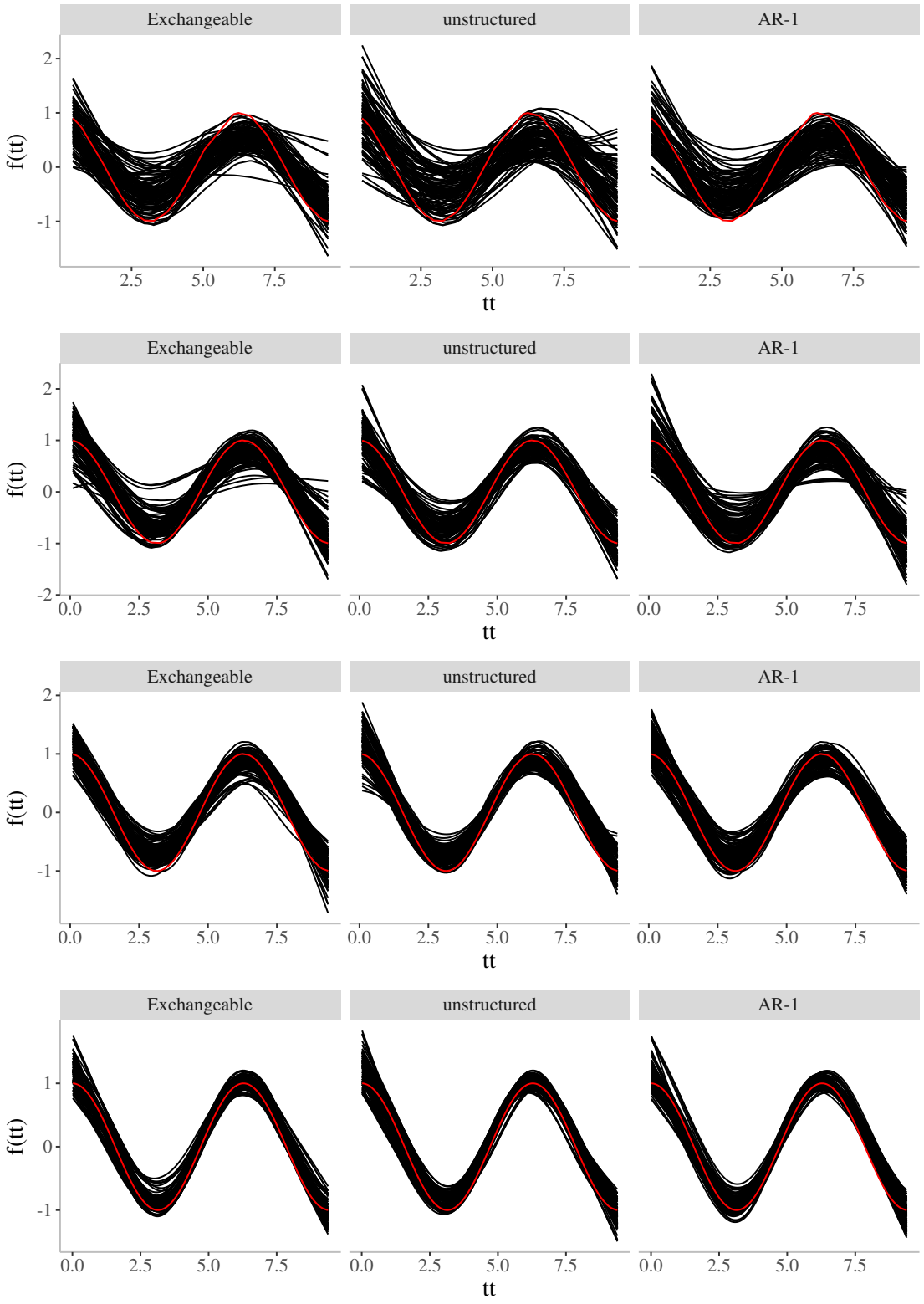


Figure 74 – Simulation study: nonparametric curves for GEE-based CSS model for (10,3), (10,10), (50,3) and (50,10), respectively, with $\rho = 0.8$ (exchangeable) by working correlation matrices.

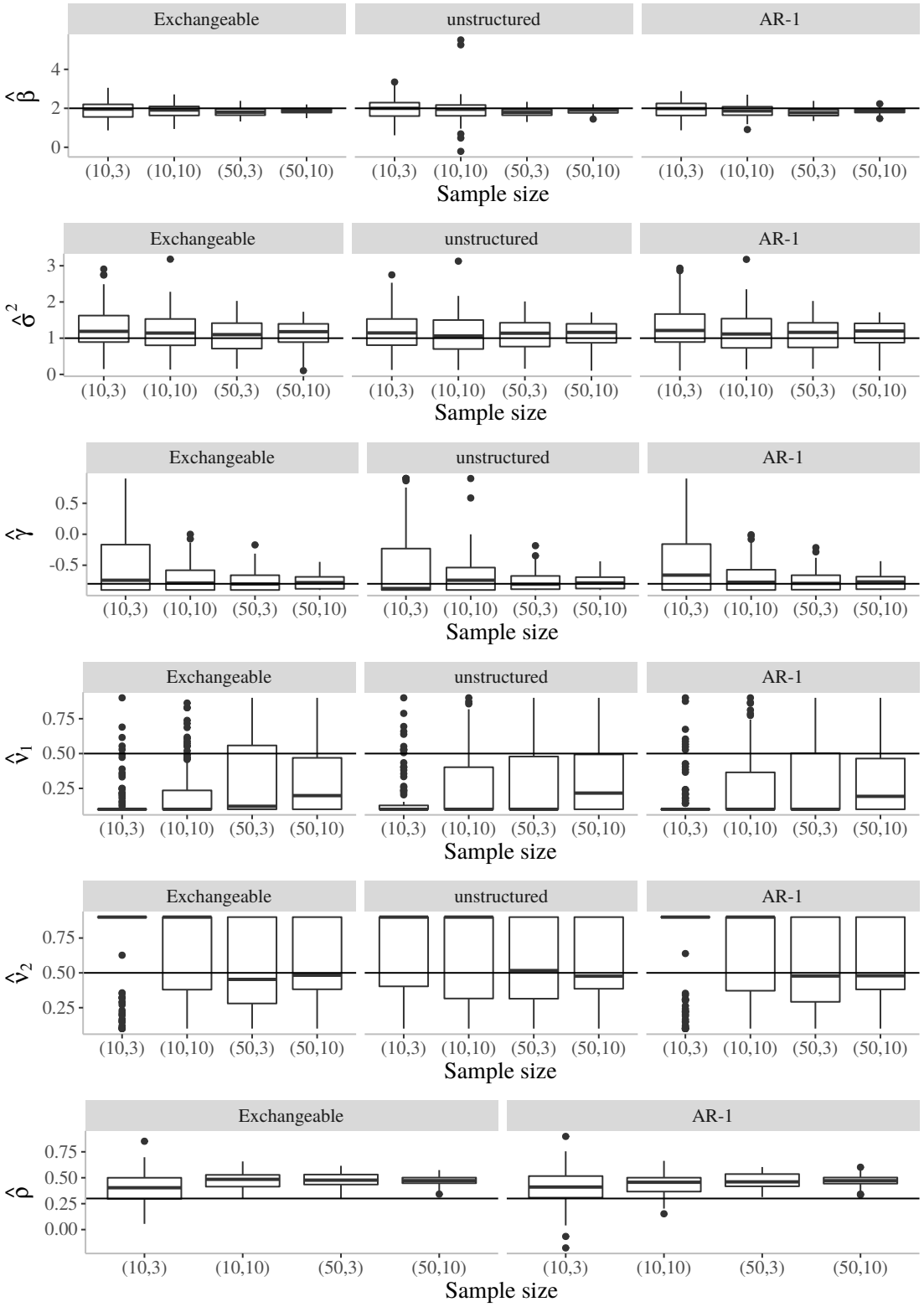


Figure 75 – Simulation study: estimated parameters for GEE-based CSCN model with $\rho = 0.3$ (exchangeable) by working correlation matrices.

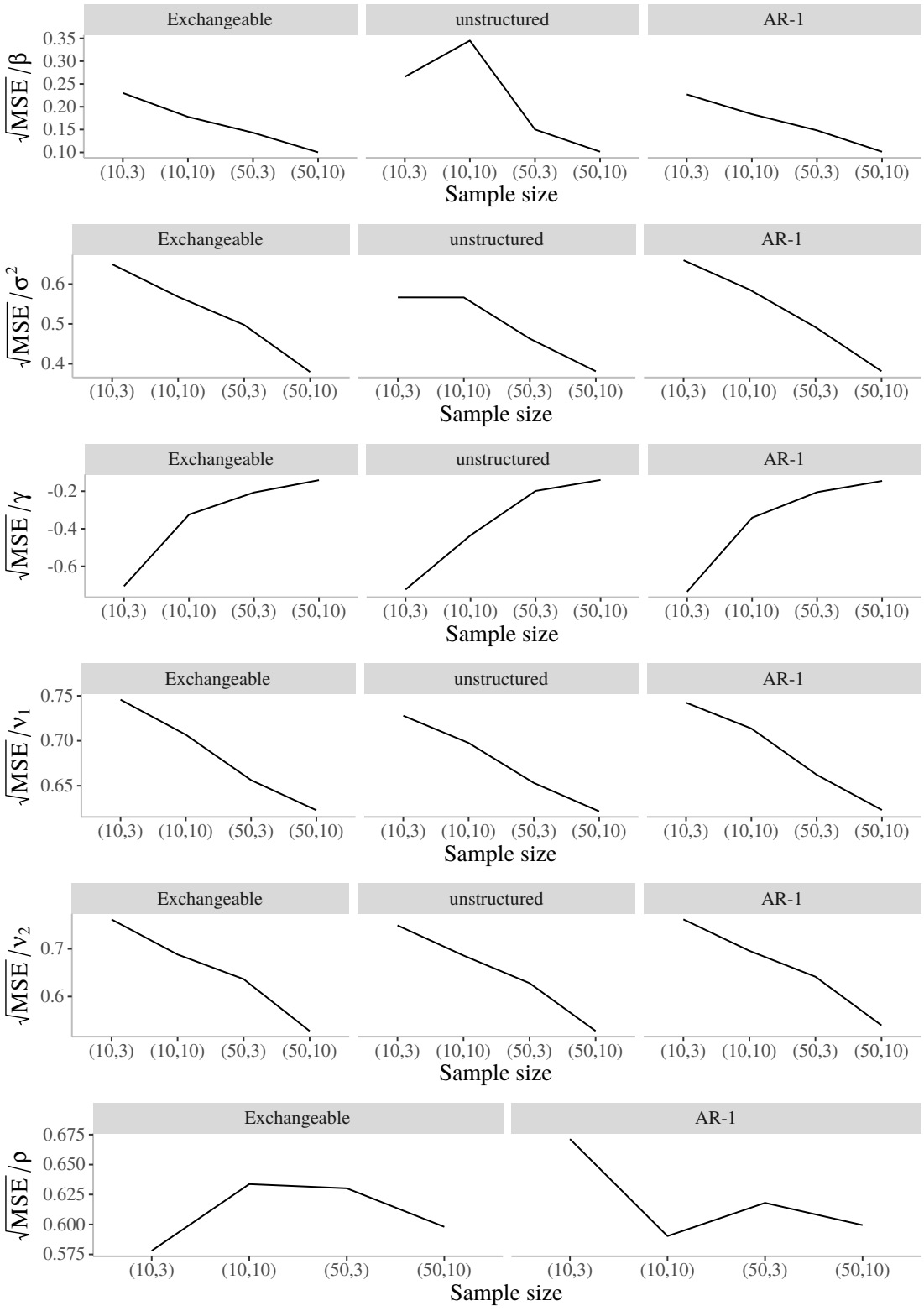


Figure 76 – Simulation study: relative mean square error of the parameters for GEE-based CSCN model with $\rho = 0.3$ (exchangeable) by working correlation matrices.

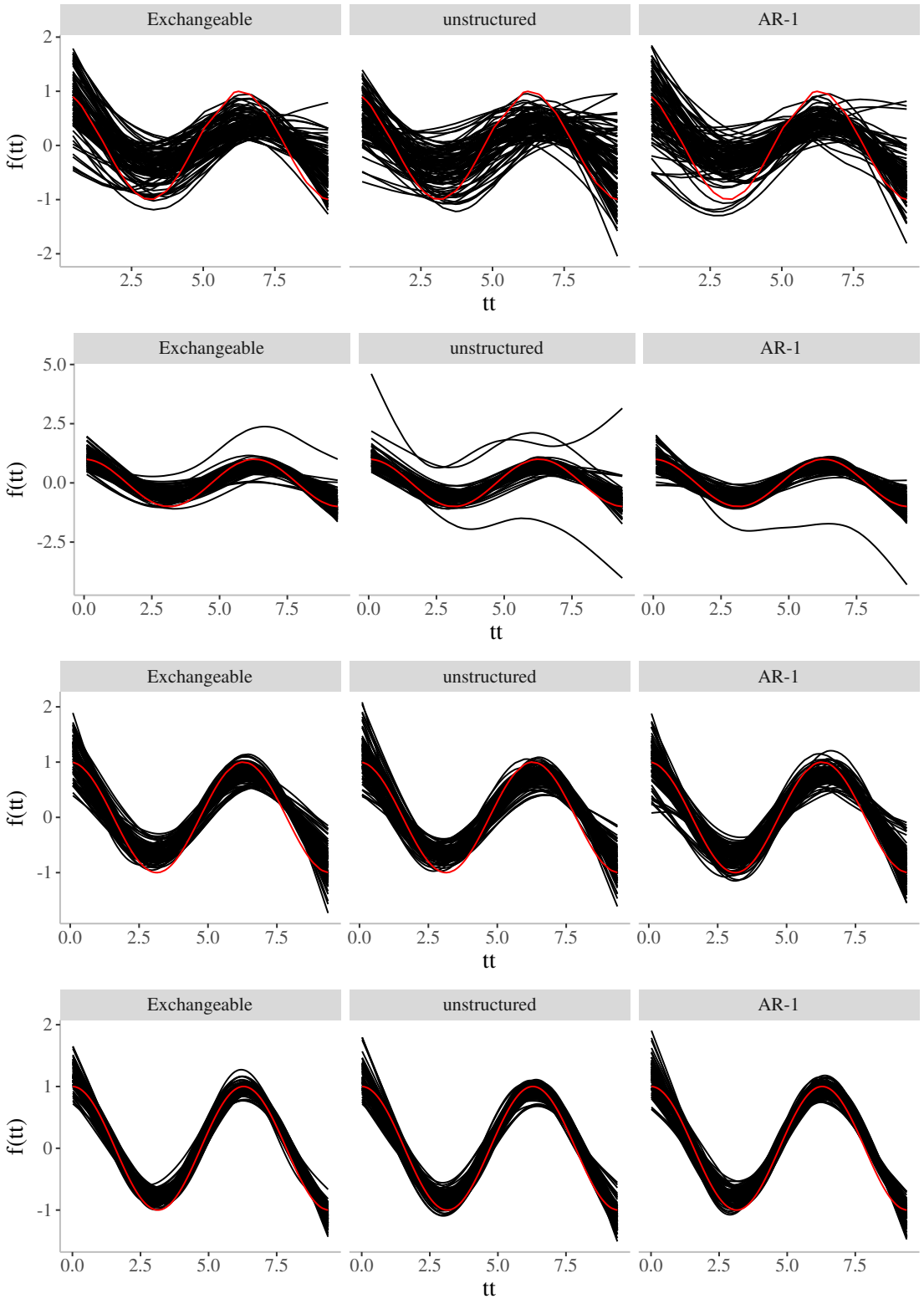


Figure 77 – Simulation study: nonparametric curves for GEE-based CSCN model for (10,3), (10,10), (50,3) and (50,10), respectively, with $\rho = 0.3$ (exchangeable) by working correlation matrices.

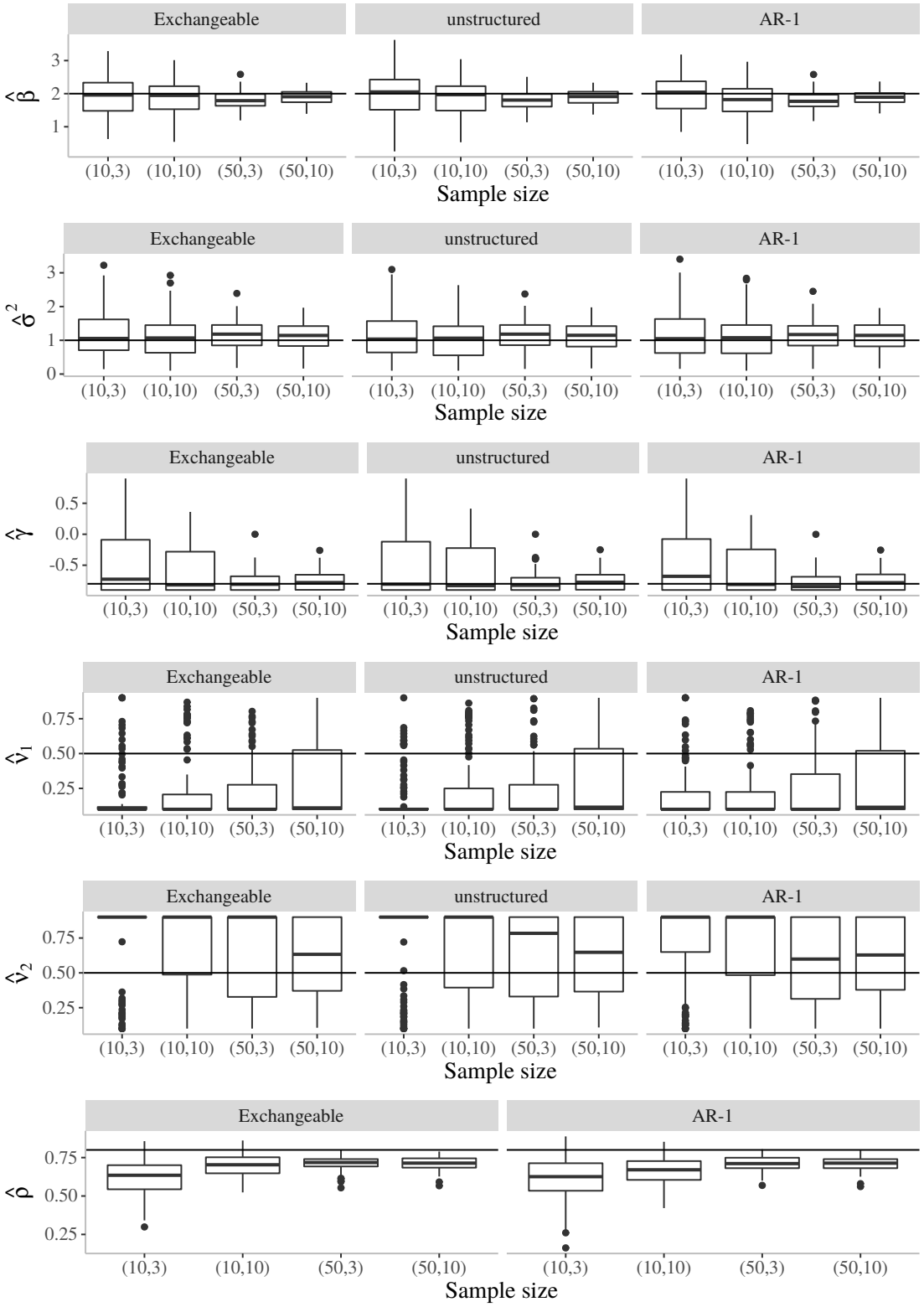


Figure 78 – Simulation study: estimated parameters for GEE-based CSCN model with $\rho = 0.8$ (exchangeable) by working correlation matrices.

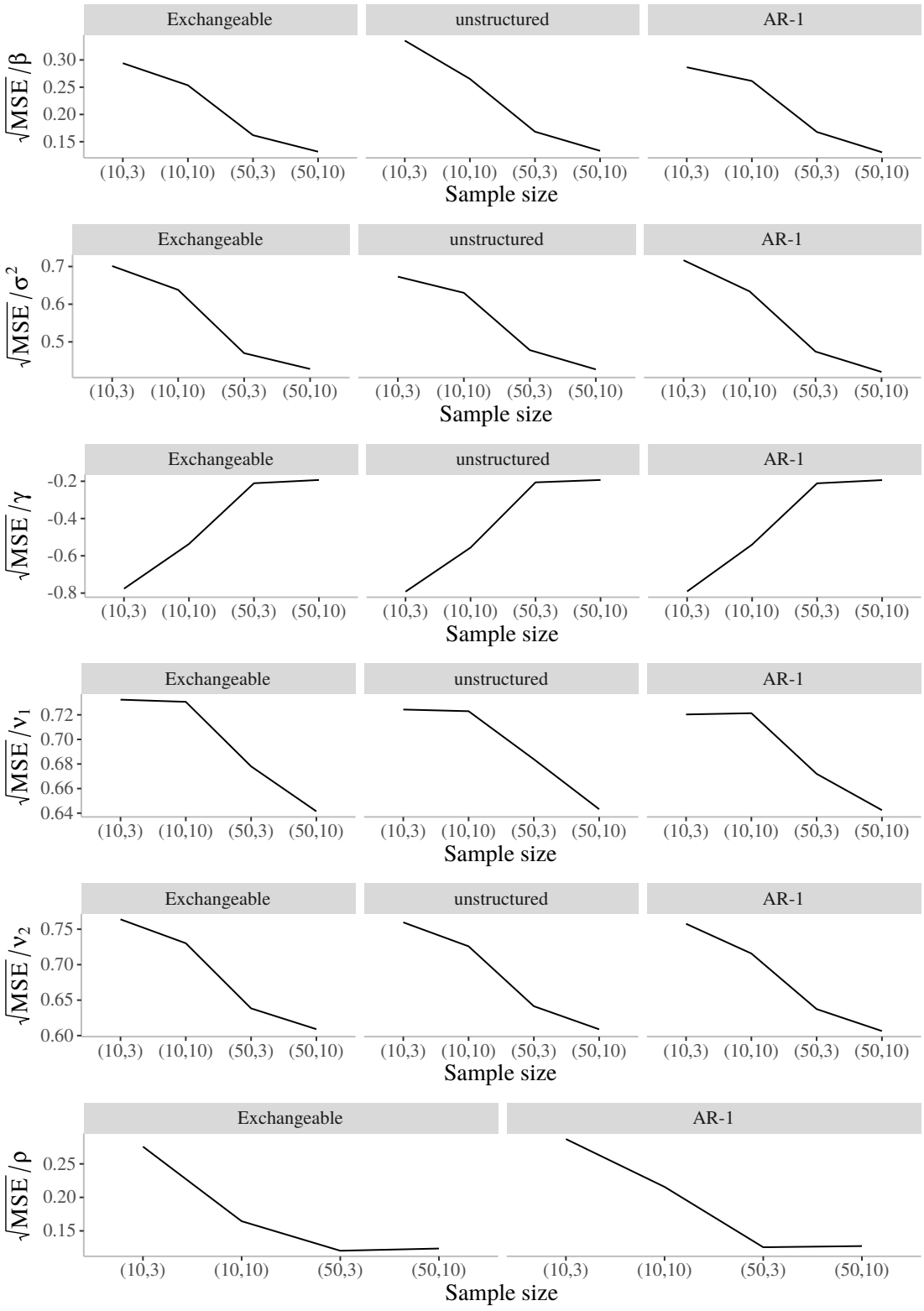


Figure 79 – Simulation study: relative mean square error of the parameters for GEE-based CSCN model with $\rho = 0.8$ (exchangeable) by working correlation matrices.

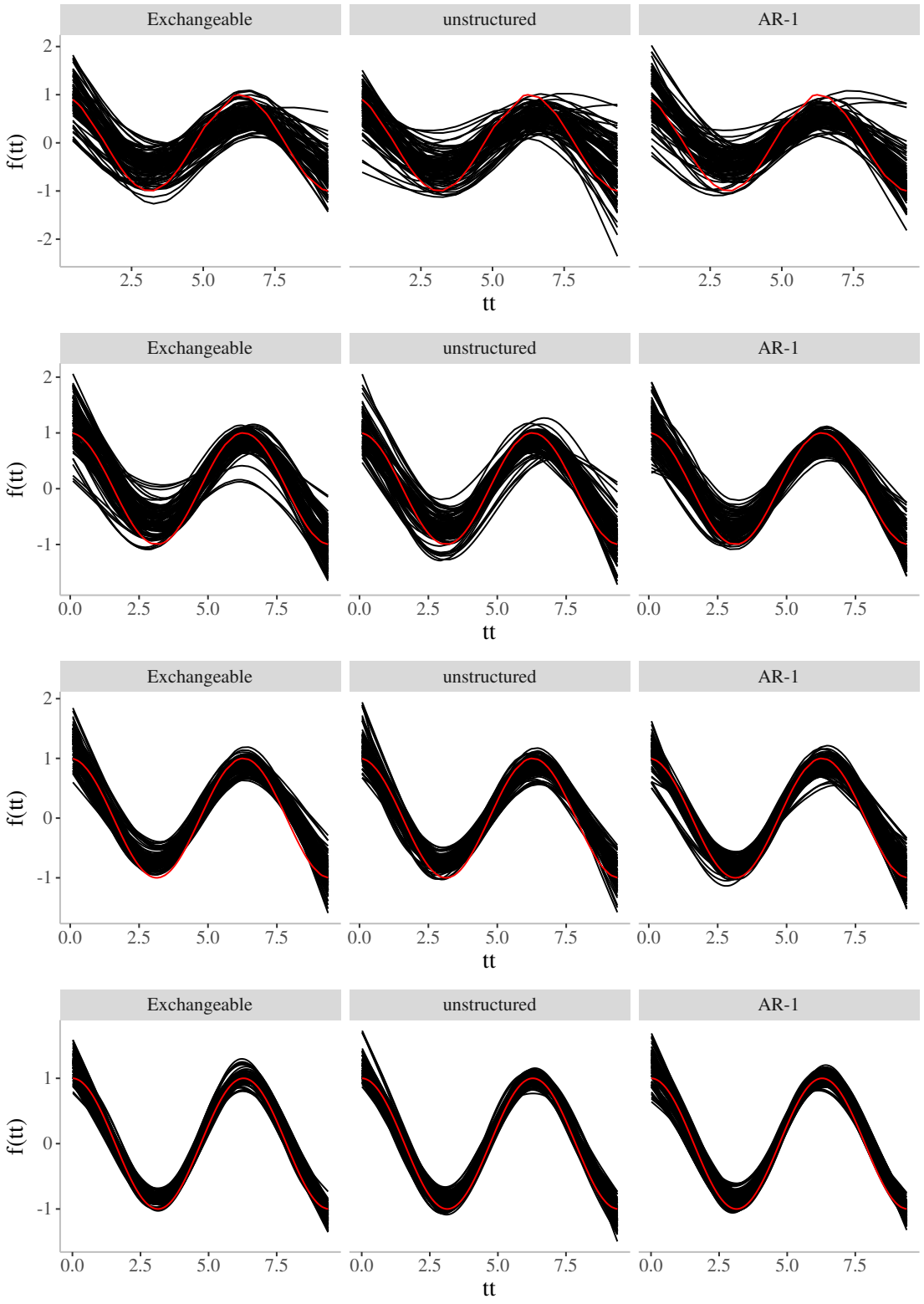


Figure 80 – Simulation study: nonparametric curves for GEE-based CSCN model for (10,3), (10,10), (50,3) and (50,10), respectively, with $\rho = 0.8$ (exchangeable) by working correlation matrices.

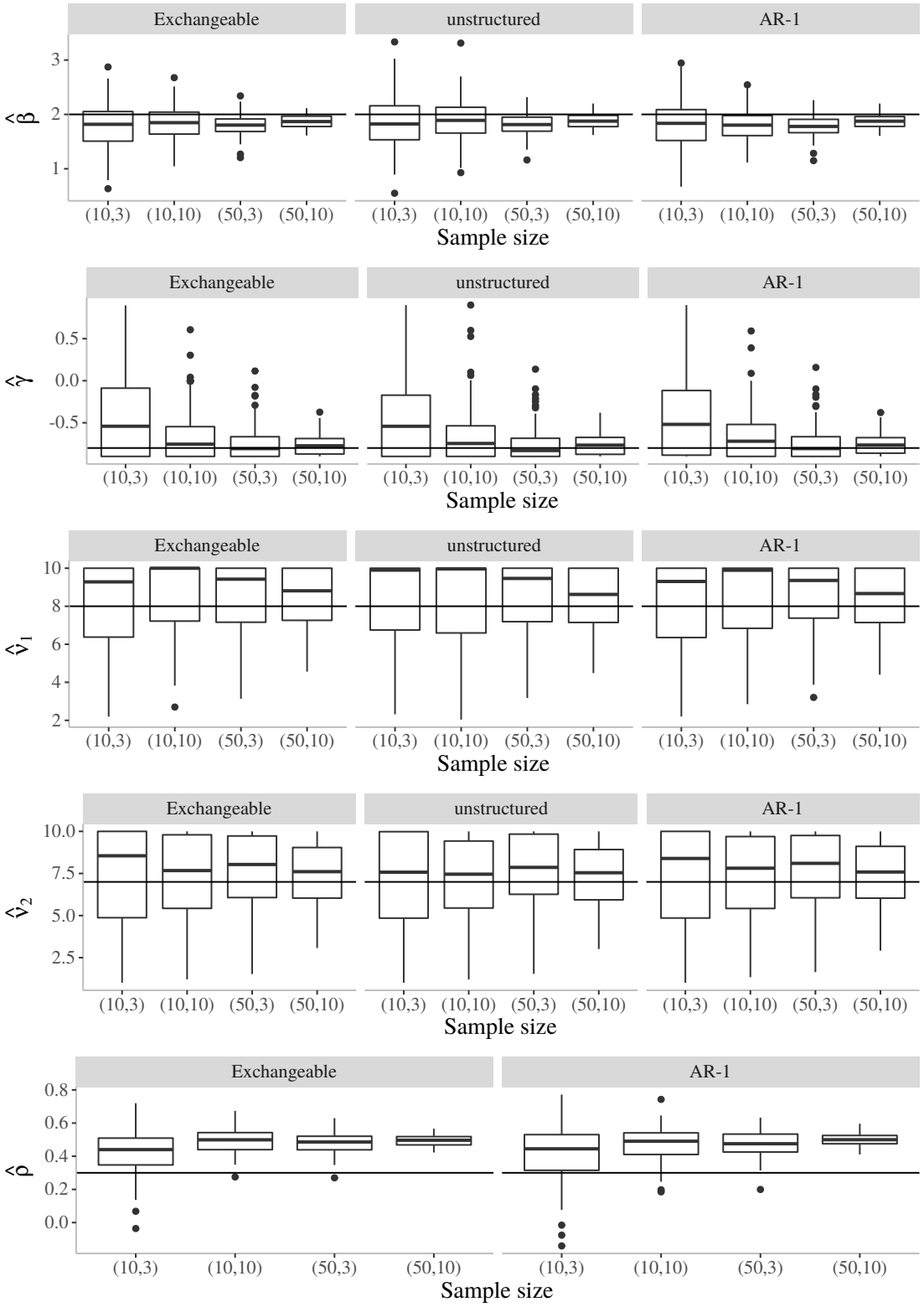


Figure 81 – Simulation study: estimated parameters for GEE-based CSGT model with $\rho = 0.3$ (exchangeable) by working correlation matrices.

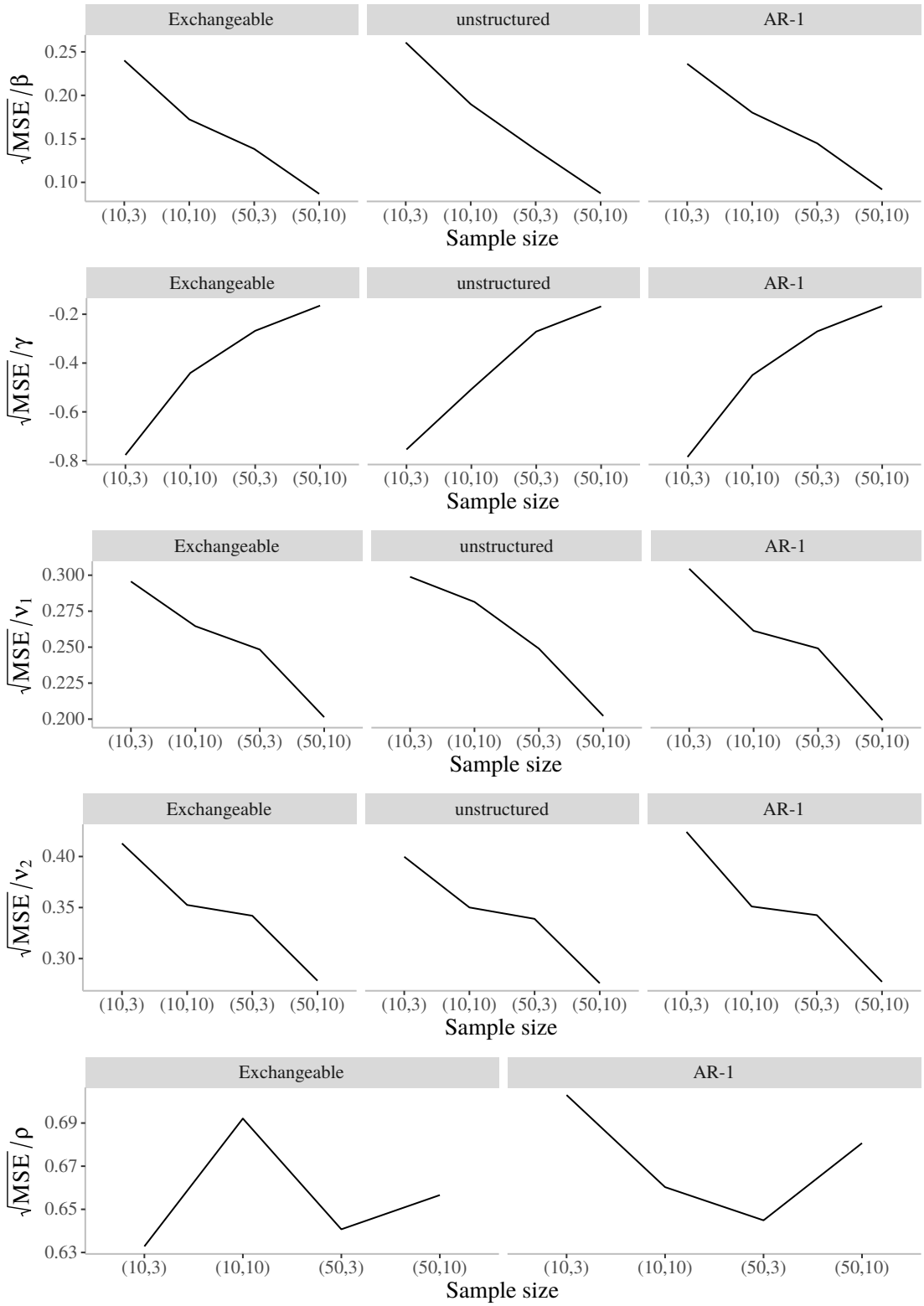


Figure 82 – Simulation study: relative mean square error of the parameters for GEE-based CSGT model with $\rho = 0.3$ (exchangeable) by working correlation matrices.

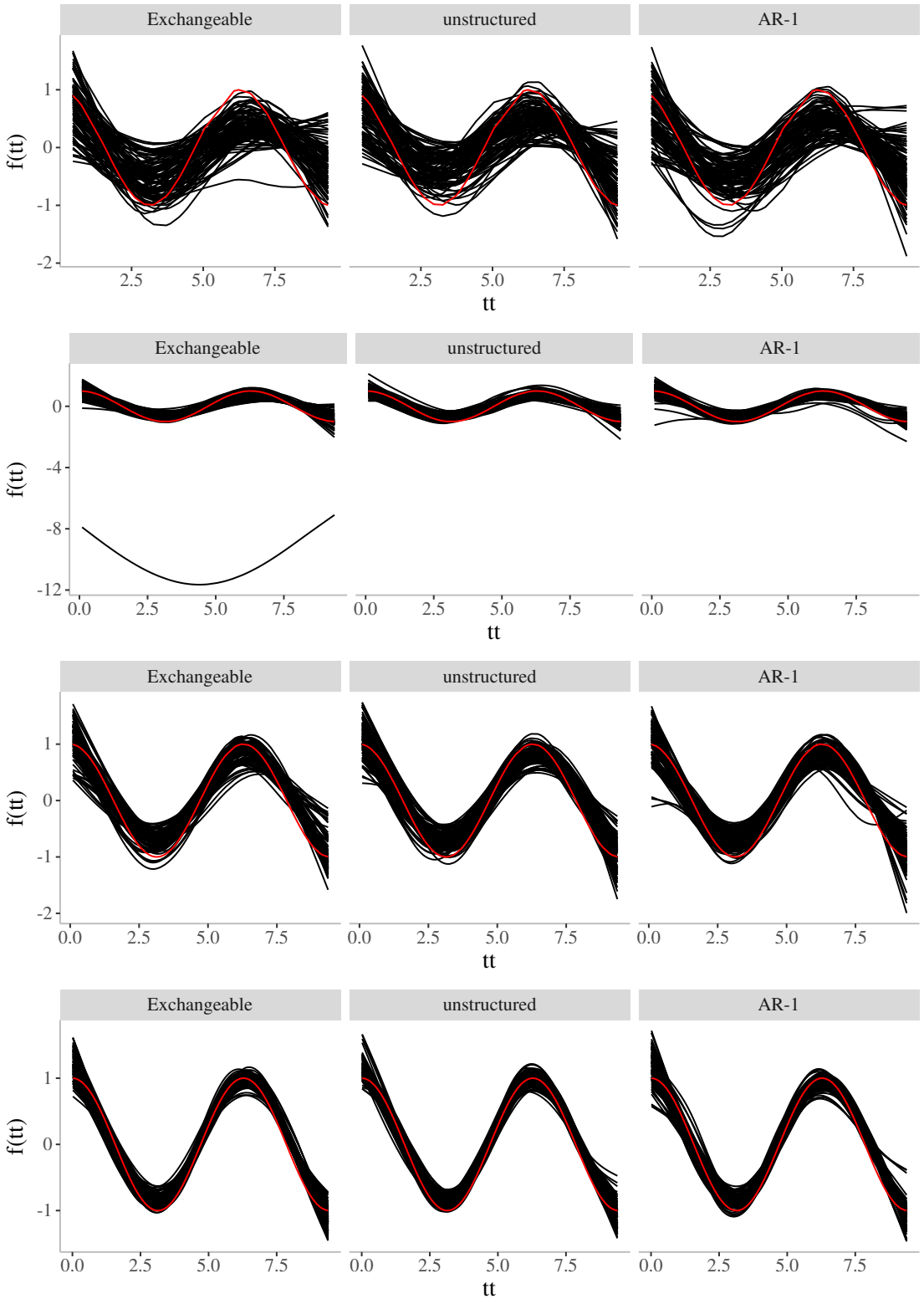


Figure 83 – Simulation study: nonparametric curves for GEE-based CSGT model for (10,3), (10,10), (50,3) and (50,10), respectively, with $\rho = 0.3$ (exchangeable) by working correlation matrices.

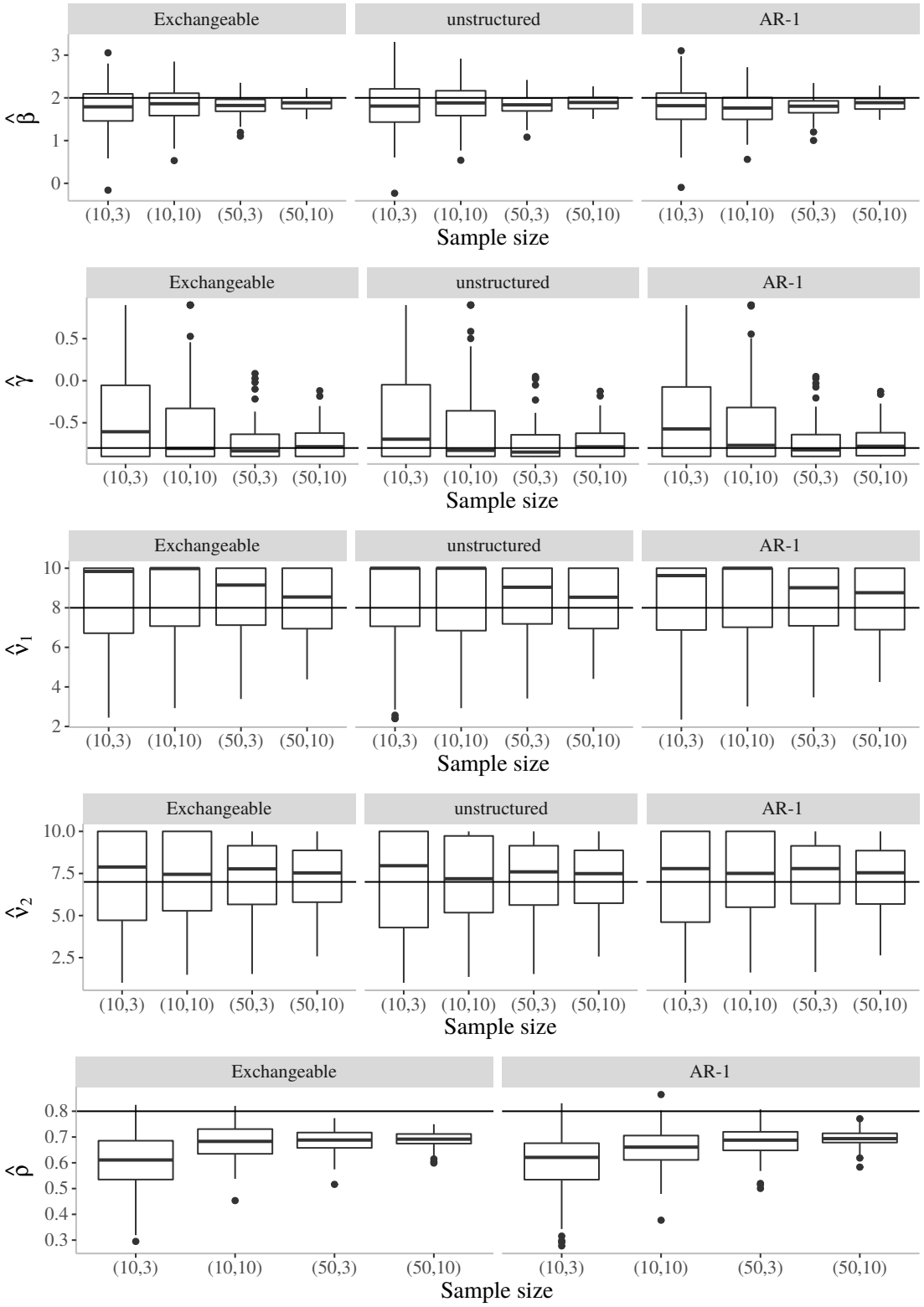


Figure 84 – Simulation study: estimated parameters for GEE-based CSGT model with $\rho = 0.8$ (exchangeable) by working correlation matrices.

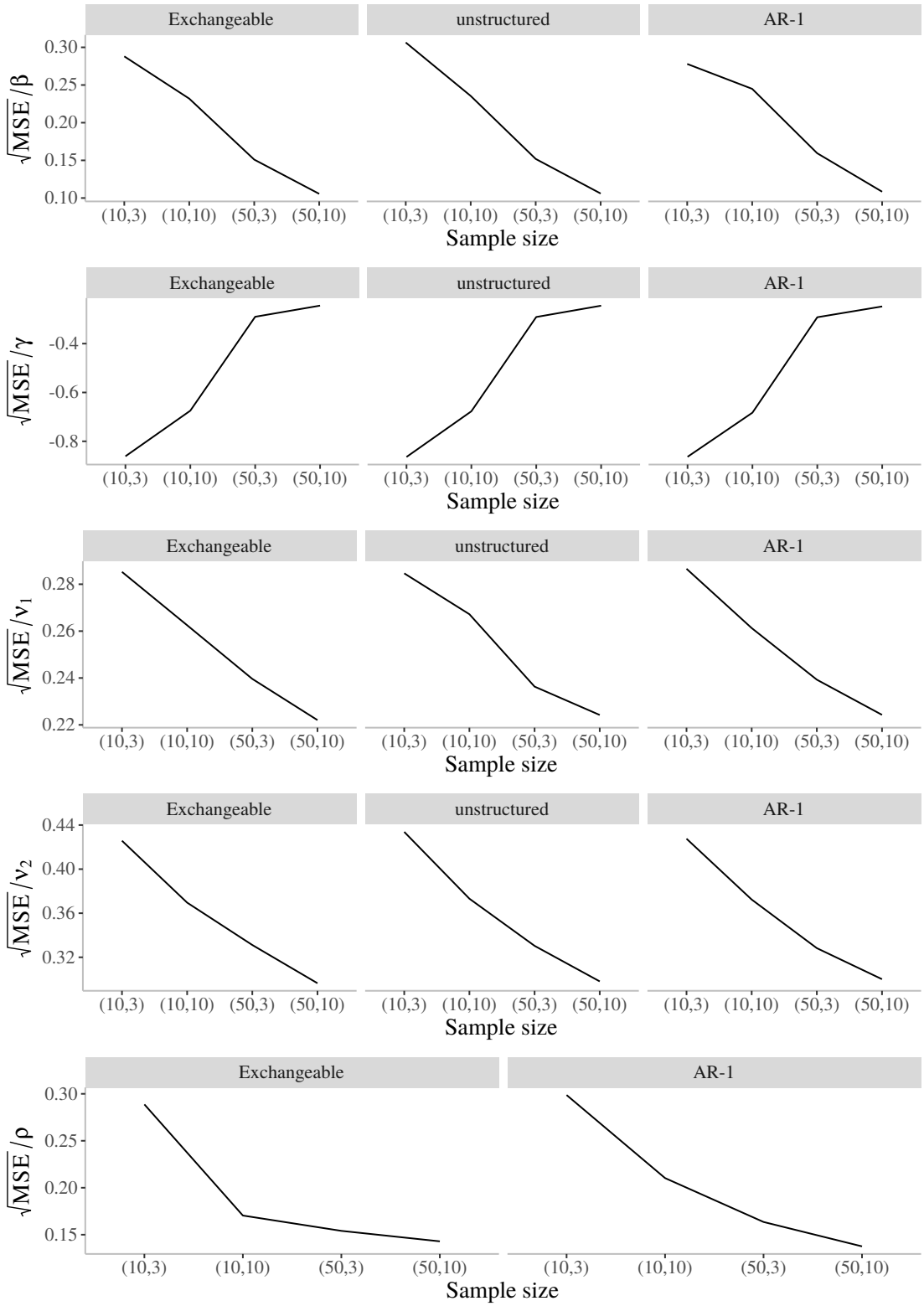


Figure 85 – Simulation study: relative mean square error of the parameters for GEE-based CSGT model with $\rho = 0.8$ (exchangeable) by working correlation matrices.

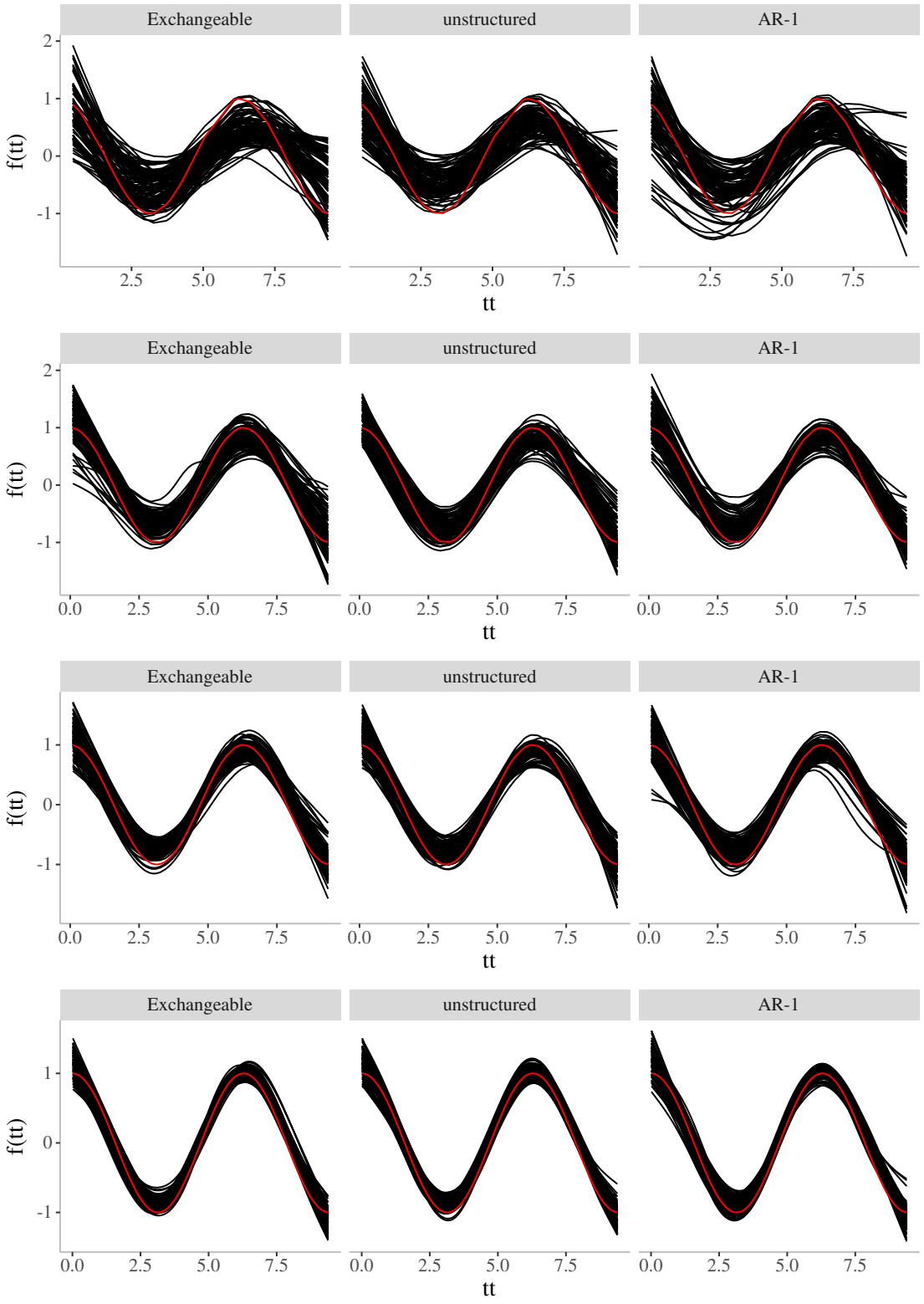


Figure 86 – Simulation study: nonparametric curves for GEE-based CSGT model for (10,3), (10,10), (50,3) and (50,10), respectively, with $\rho = 0.8$ (exchangeable) by working correlation matrices.

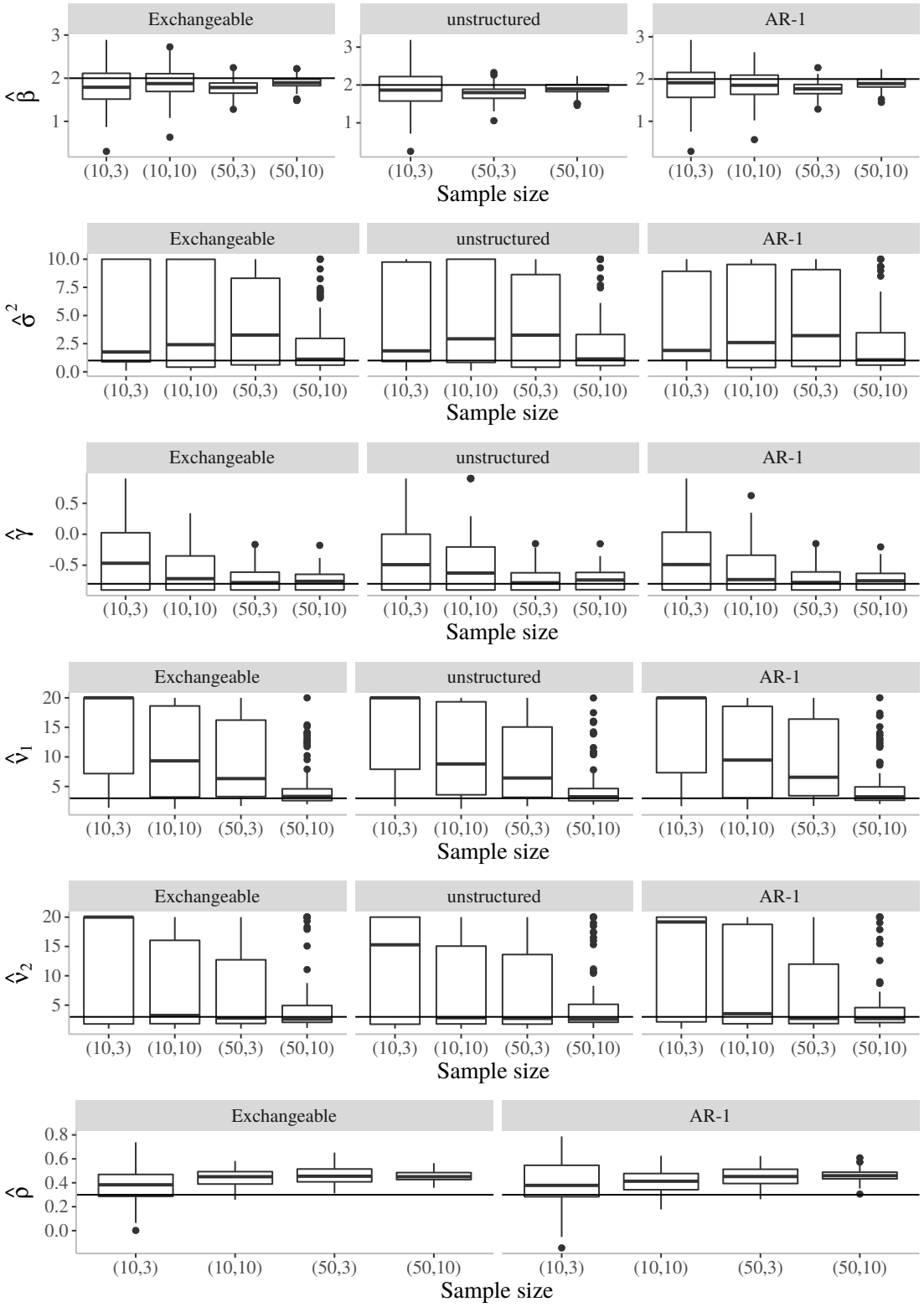


Figure 87 – Simulation study: estimated parameters for GEE-based CSBPN model with $\rho = 0.3$ (exchangeable) by working correlation matrices.

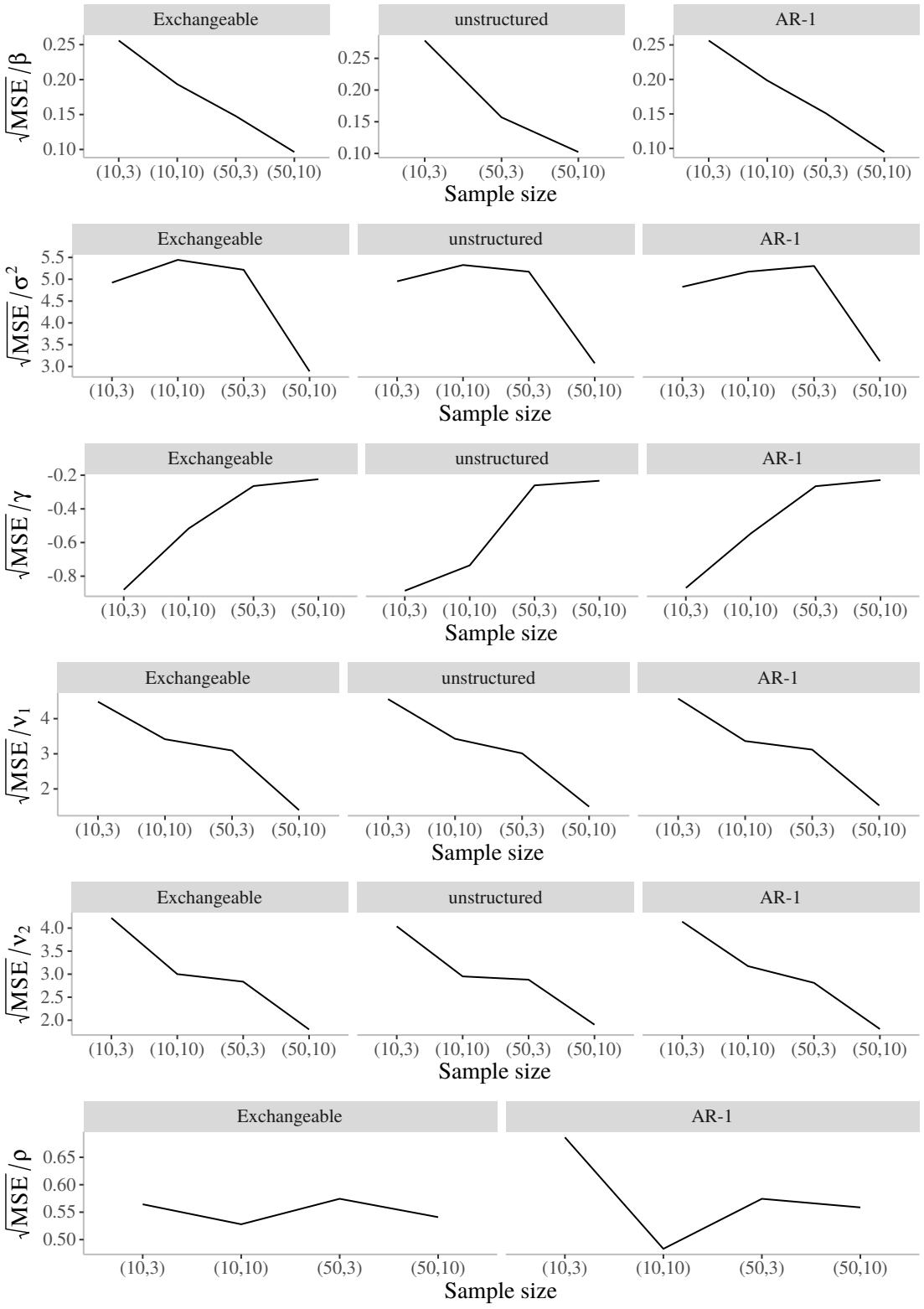


Figure 88 – Simulation study: relative mean square error of the parameters for GEE-based CSBPN model with $\rho = 0.3$ (exchangeable) by working correlation matrices.

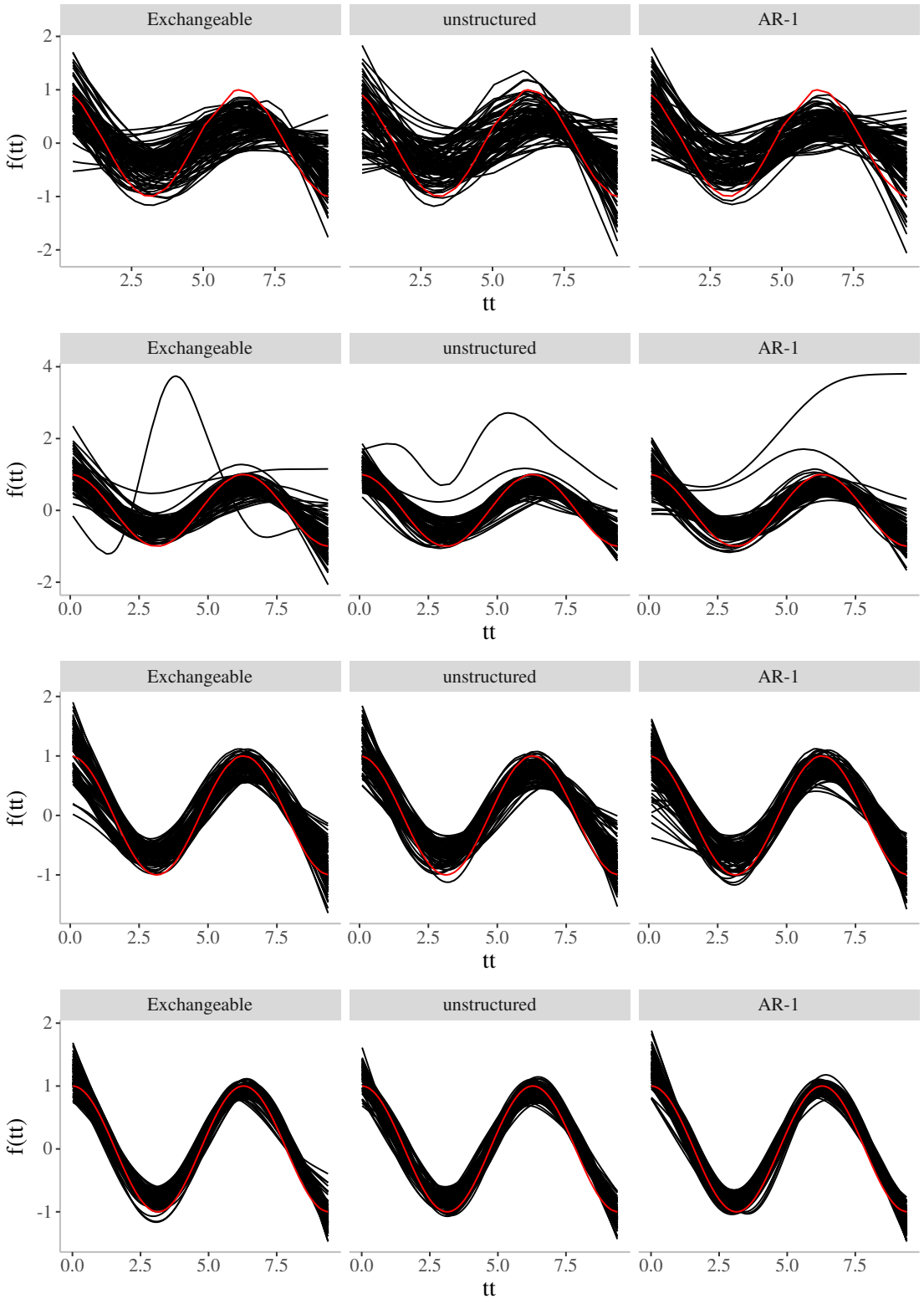


Figure 89 – Simulation study: nonparametric curves for GEE-based CSBPN model for (10,3), (10,10), (50,3) and (50,10), respectively, with $\rho = 0.3$ (exchangeable) by working correlation matrices.

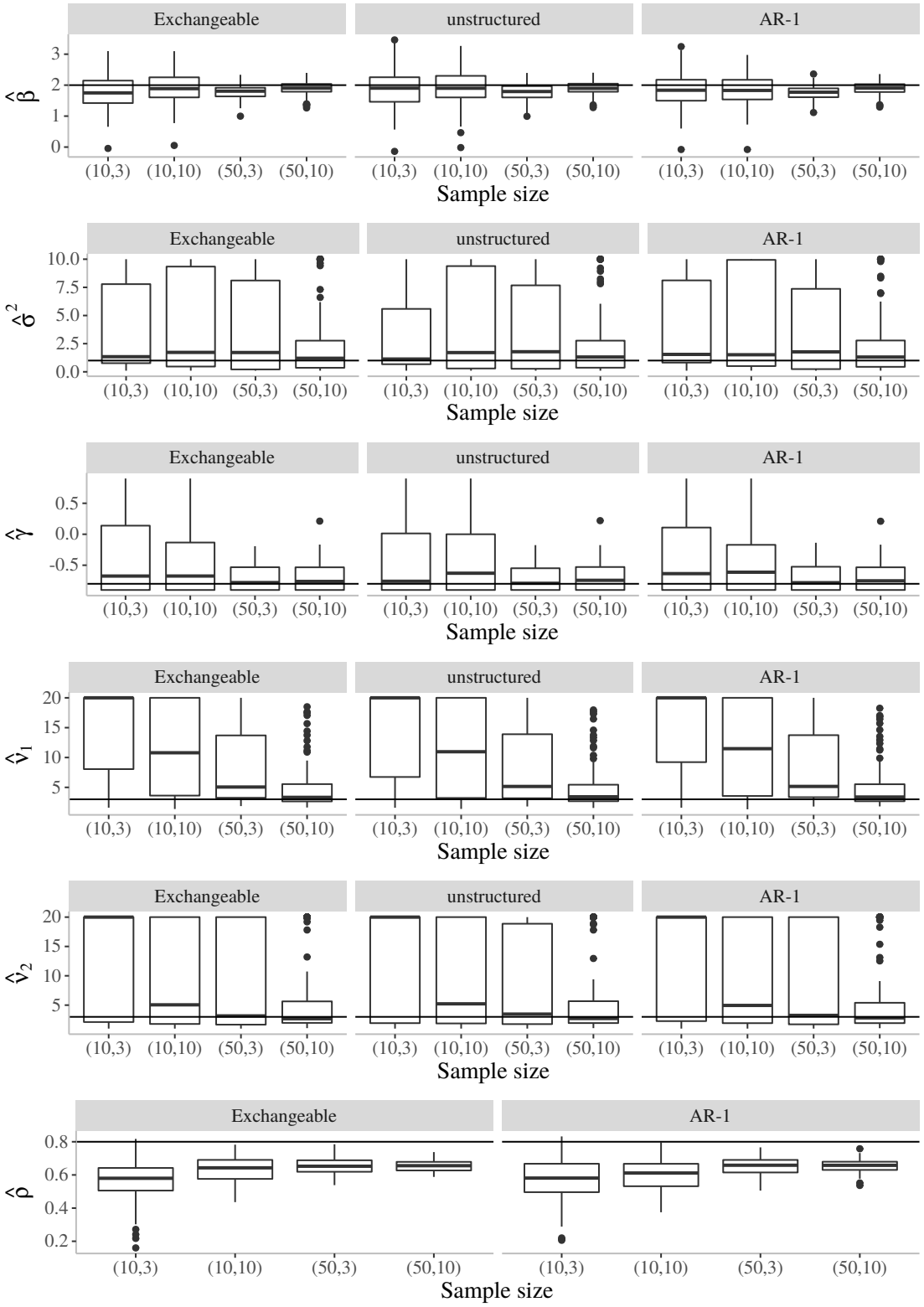


Figure 90 – Simulation study: estimated parameters for GEE-based CSBPN model with $\rho = 0.8$ (exchangeable) by working correlation matrices.

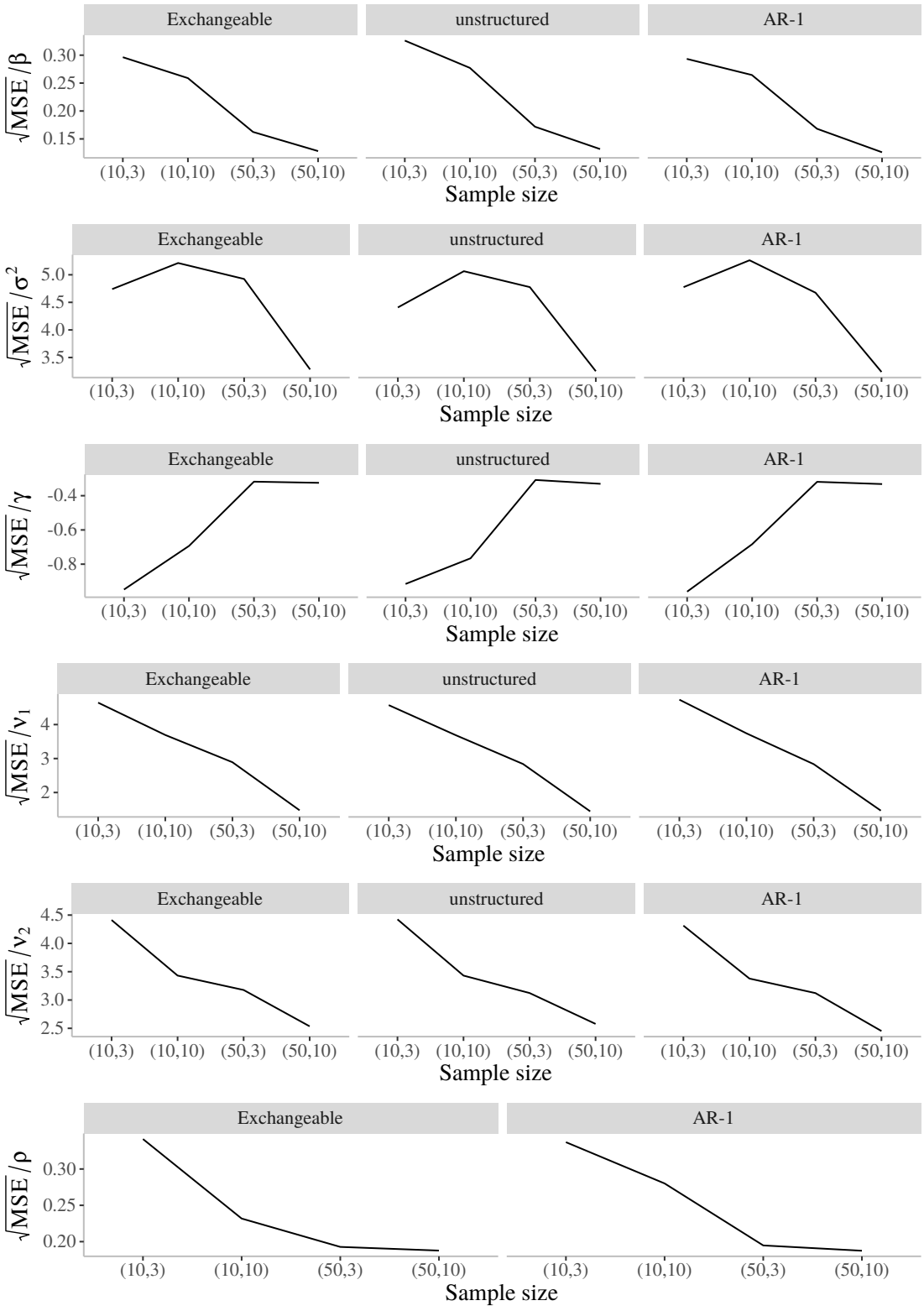


Figure 91 – Simulation study: relative mean square error of the parameters for GEE-based CSBPN model with $\rho = 0.8$ (exchangeable) by working correlation matrices.

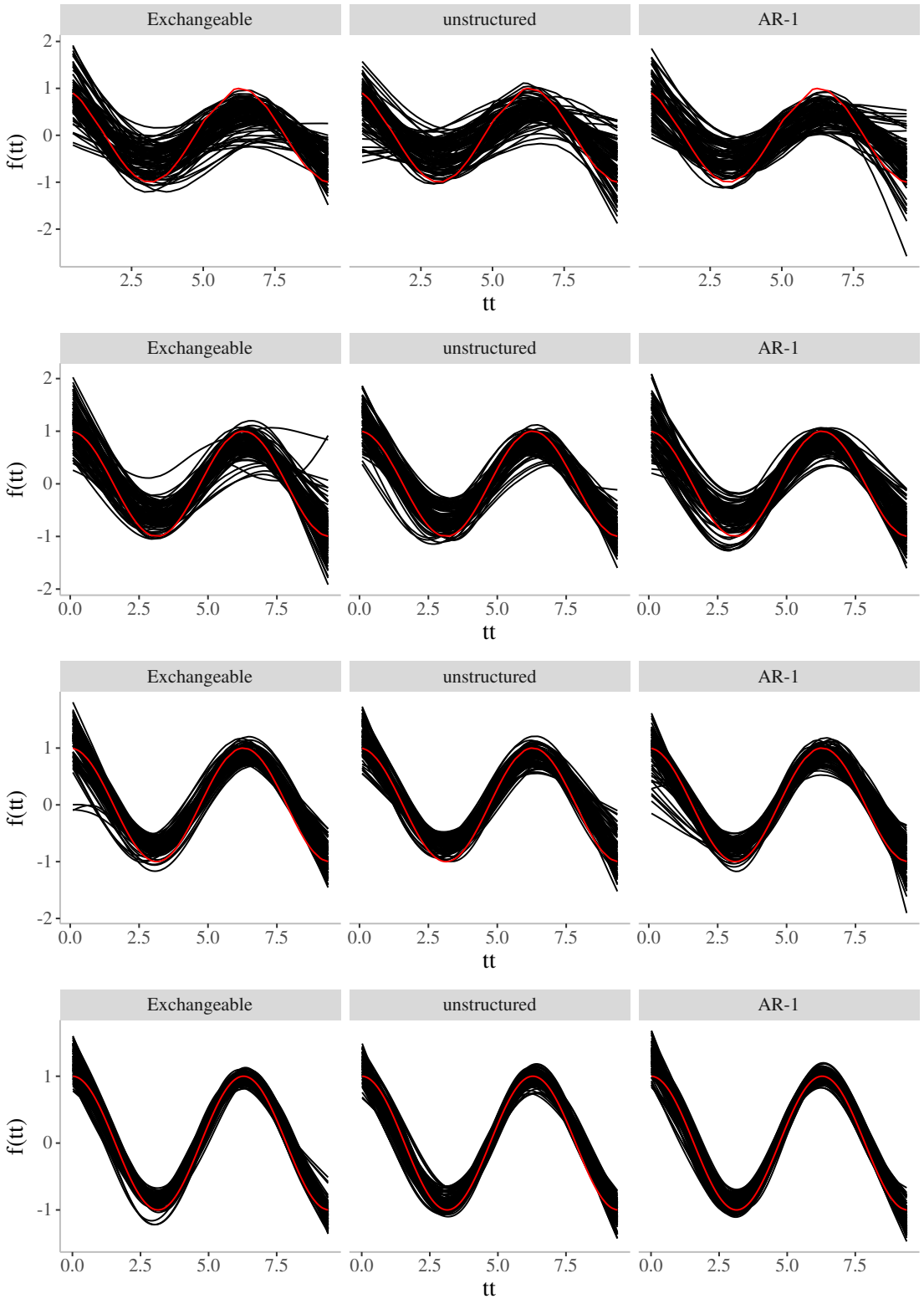


Figure 92 – Simulation study: nonparametric curves for GEE-based CSBPN model for (10,3), (10,10), (50,3) and (50,10), respectively, with $\rho = 0.8$ (exchangeable) by working correlation matrices.

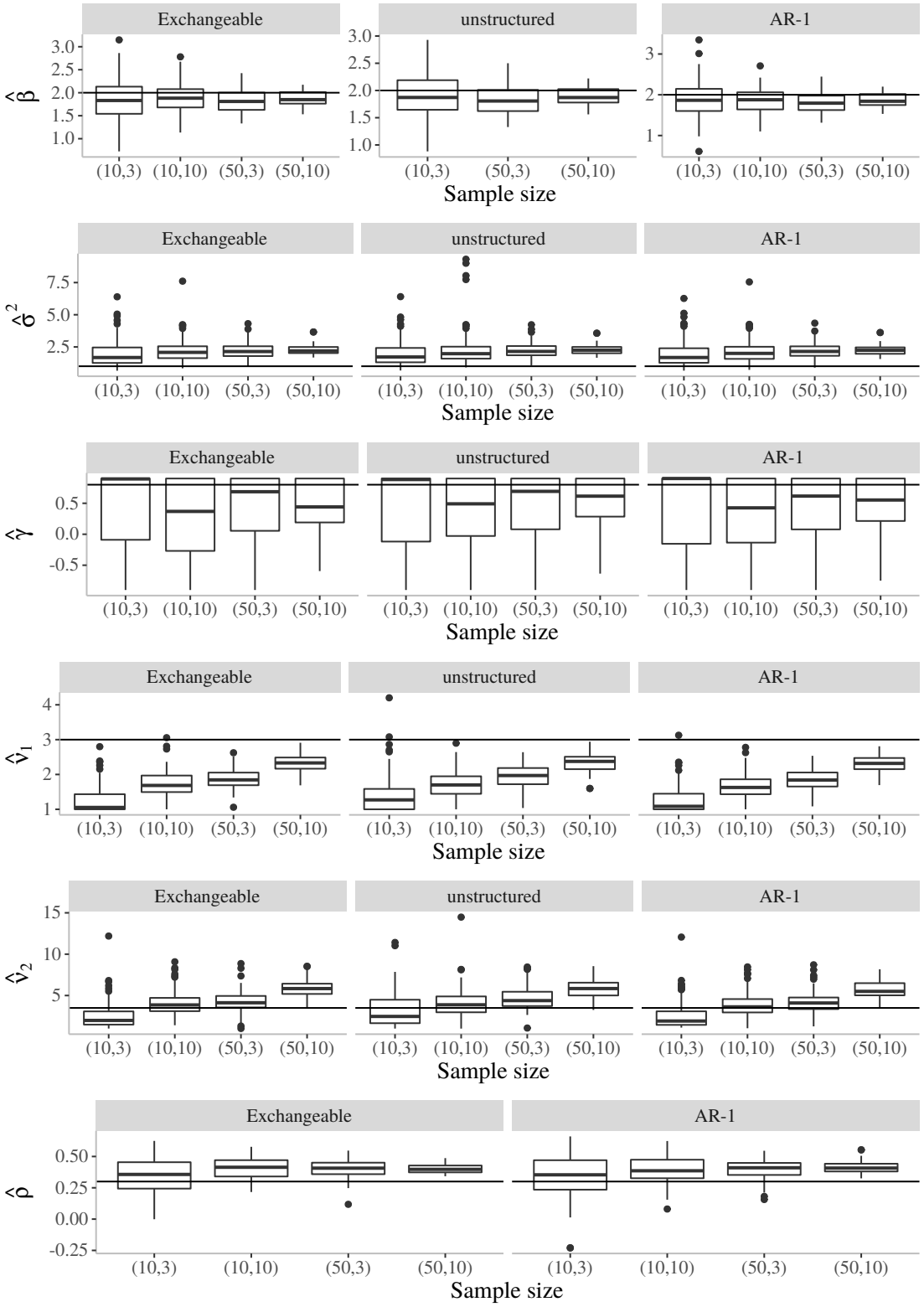


Figure 93 – Simulation study: estimated parameters for GEE-based CSBSN model with $\rho = 0.3$ (exchangeable) by working correlation matrices.

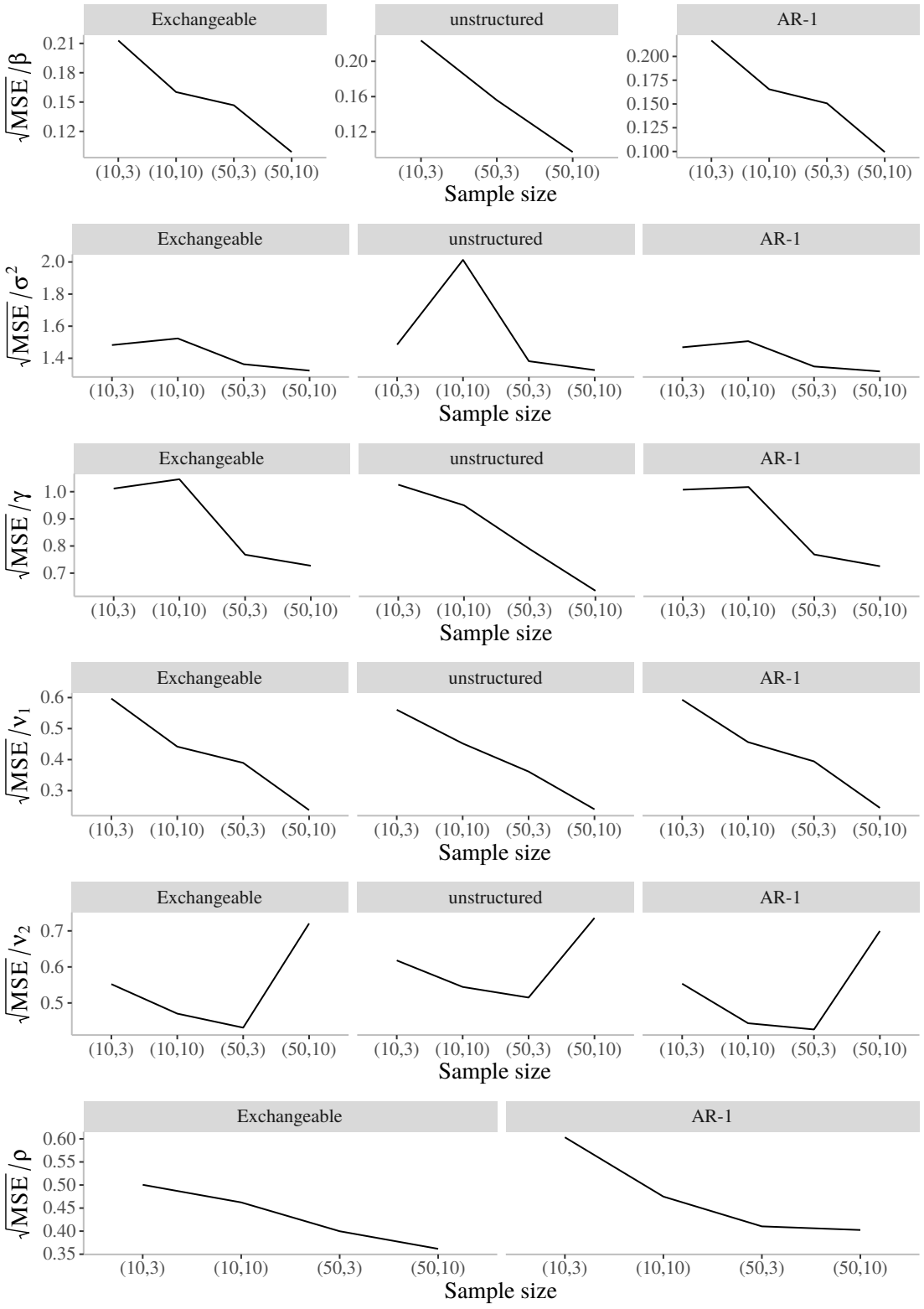


Figure 94 – Simulation study: relative mean square error of the parameters for GEE-based CSBSN model with $\rho = 0.3$ (exchangeable) by working correlation matrices.

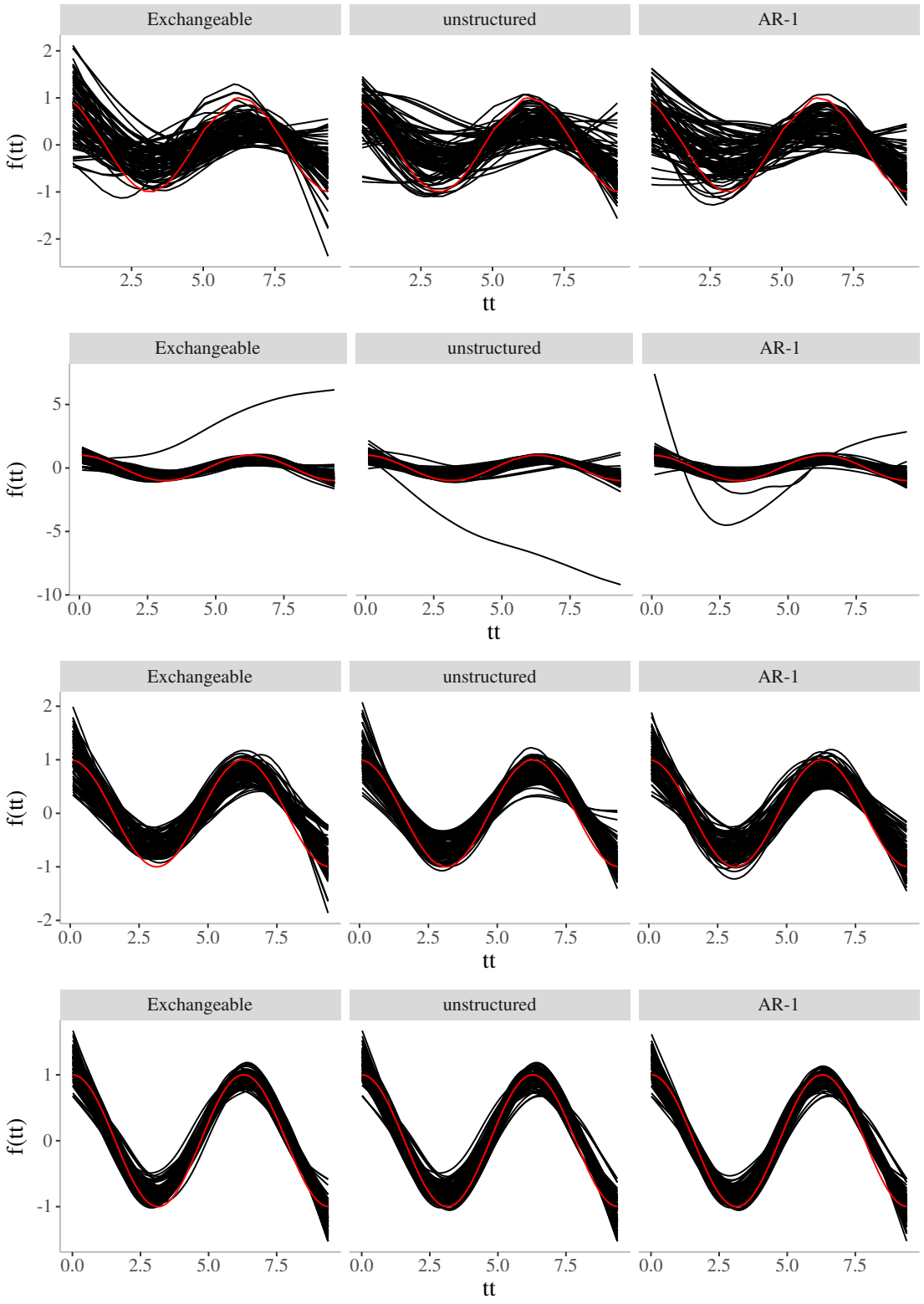


Figure 95 – Simulation study: nonparametric curves for GEE-based CSBSN model for (10,3), (10,10), (50,3) and (50,10), respectively, with $\rho = 0.3$ (exchangeable) by working correlation matrices.

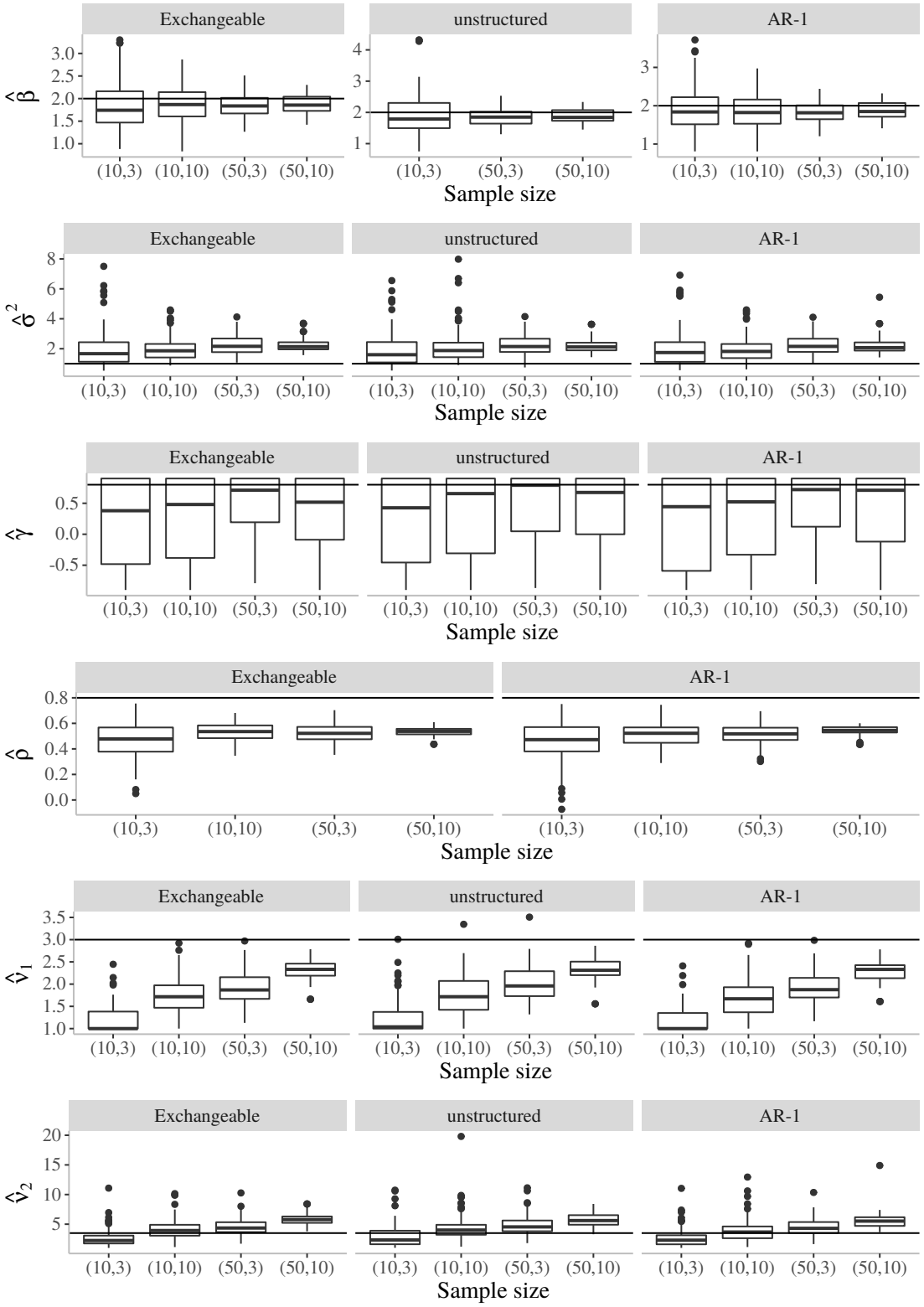


Figure 96 – Simulation study: estimated parameters for GEE-based CSBSN model with $\rho = 0.8$ (exchangeable) by working correlation matrices.

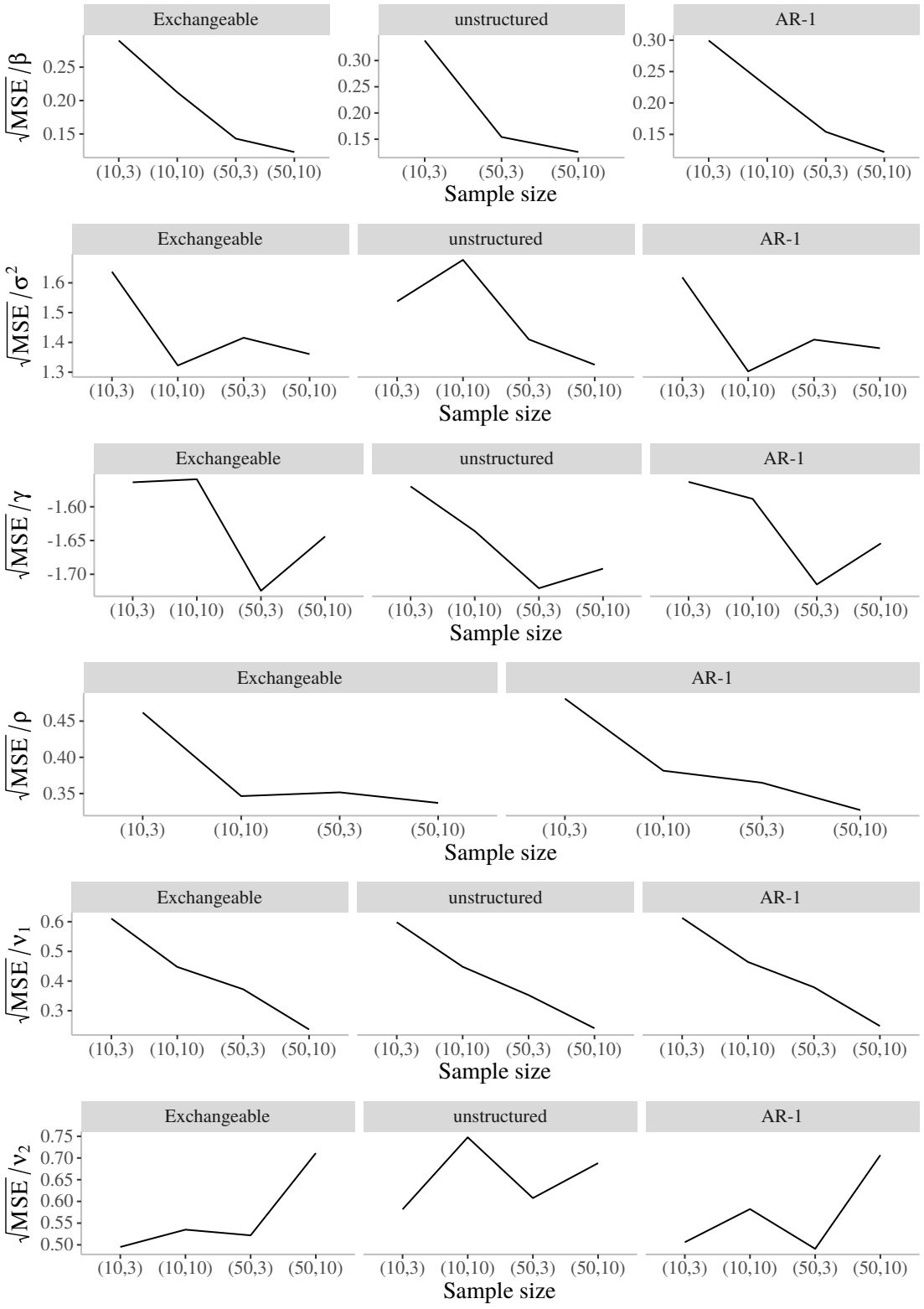


Figure 97 – Simulation study: relative mean square error of the parameters for GEE-based CSBSN model with $\rho = 0.8$ (exchangeable) by working correlation matrices.

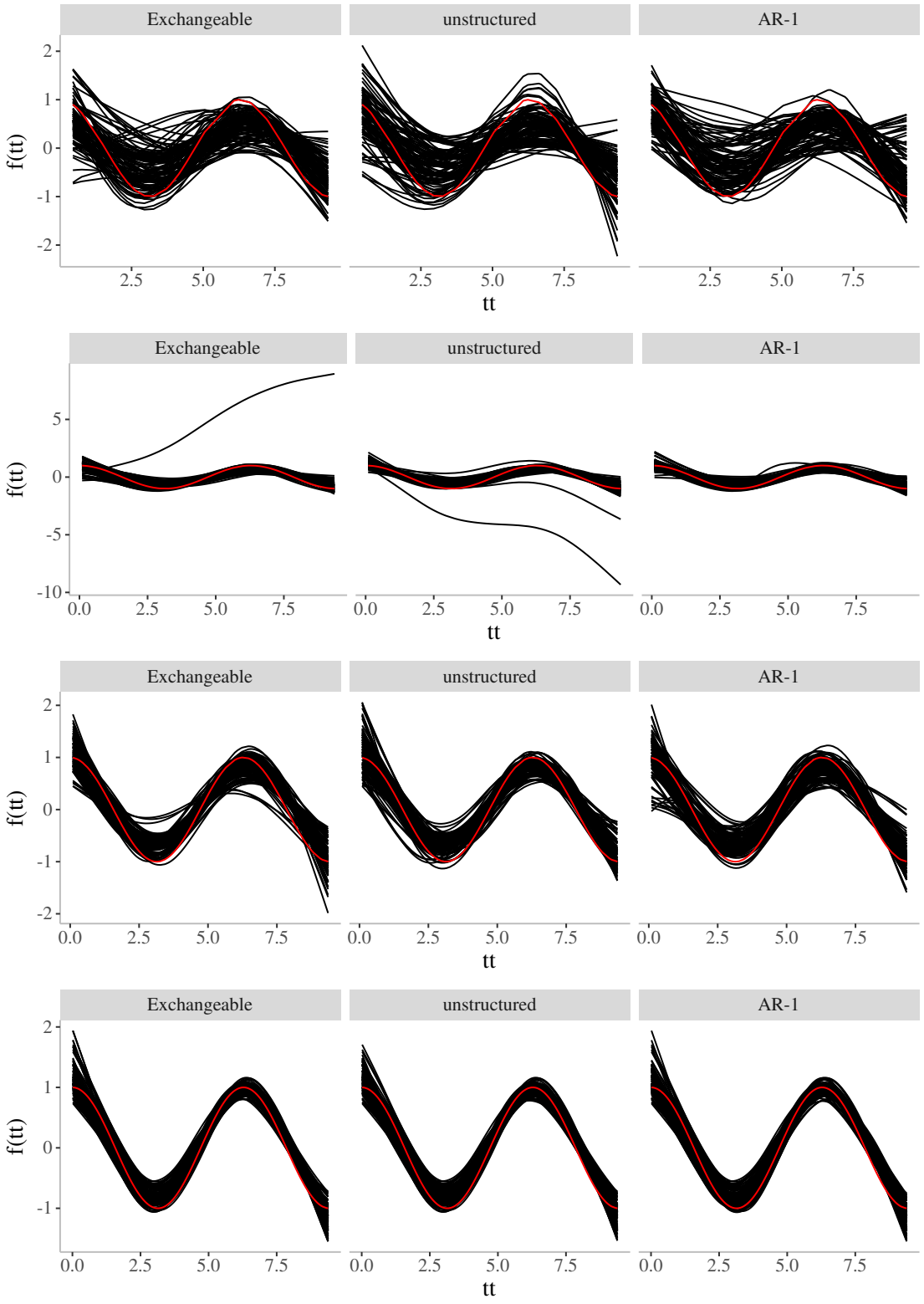


Figure 98 – Simulation study: nonparametric curves for GEE-based CSBSN model for (10,3), (10,10), (50,3) and (50,10), respectively, with $\rho = 0.8$ (exchangeable) by working correlation matrices.

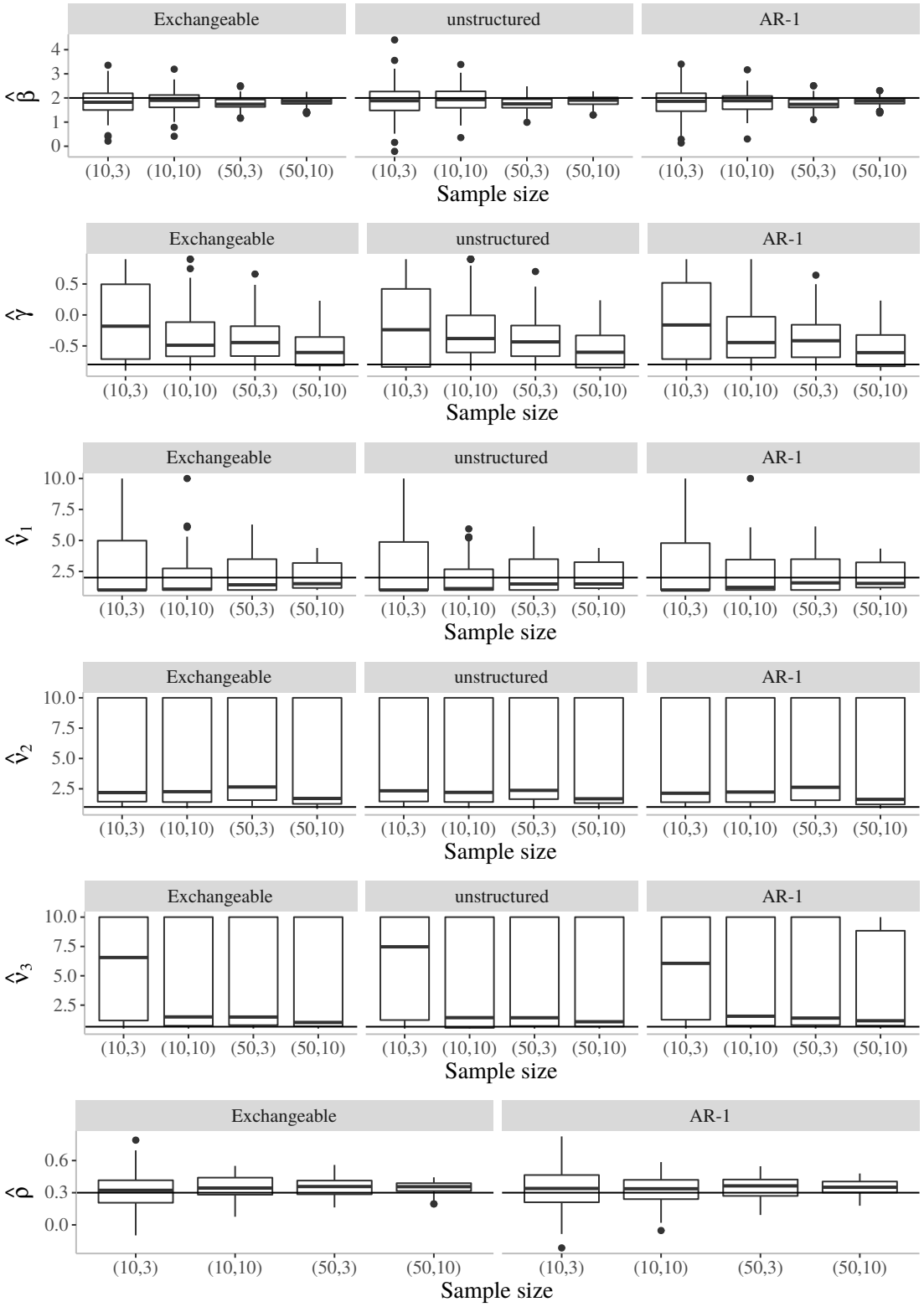


Figure 99 – Simulation study: estimated parameters for GEE-based CSGGN model with $\rho = 0.3$ (exchangeable) by working correlation matrices.

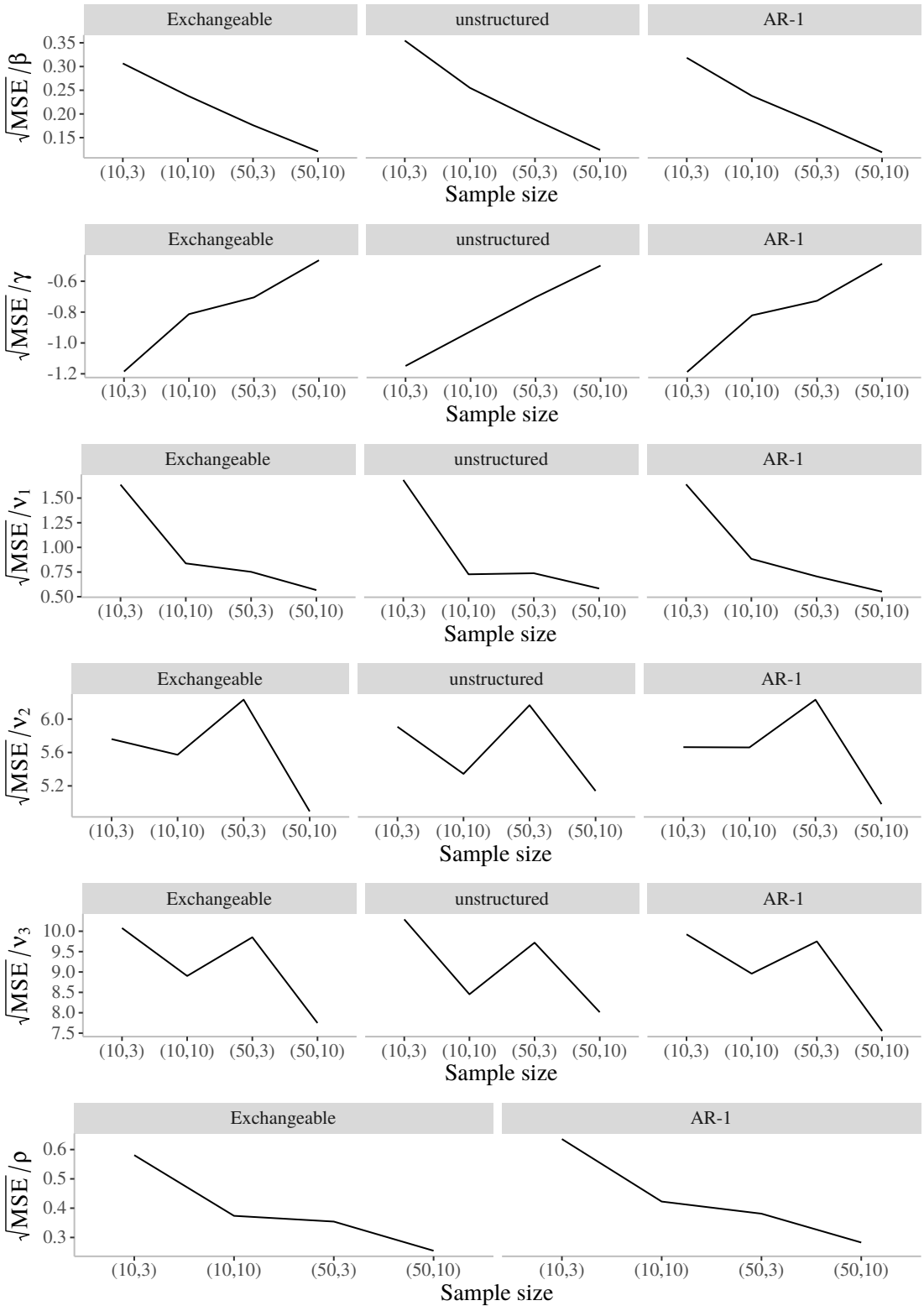


Figure 100 – Simulation study: relative mean square error of the parameters for GEE-based CSGGN model with $\rho = 0.3$ (exchangeable) by working correlation matrices.

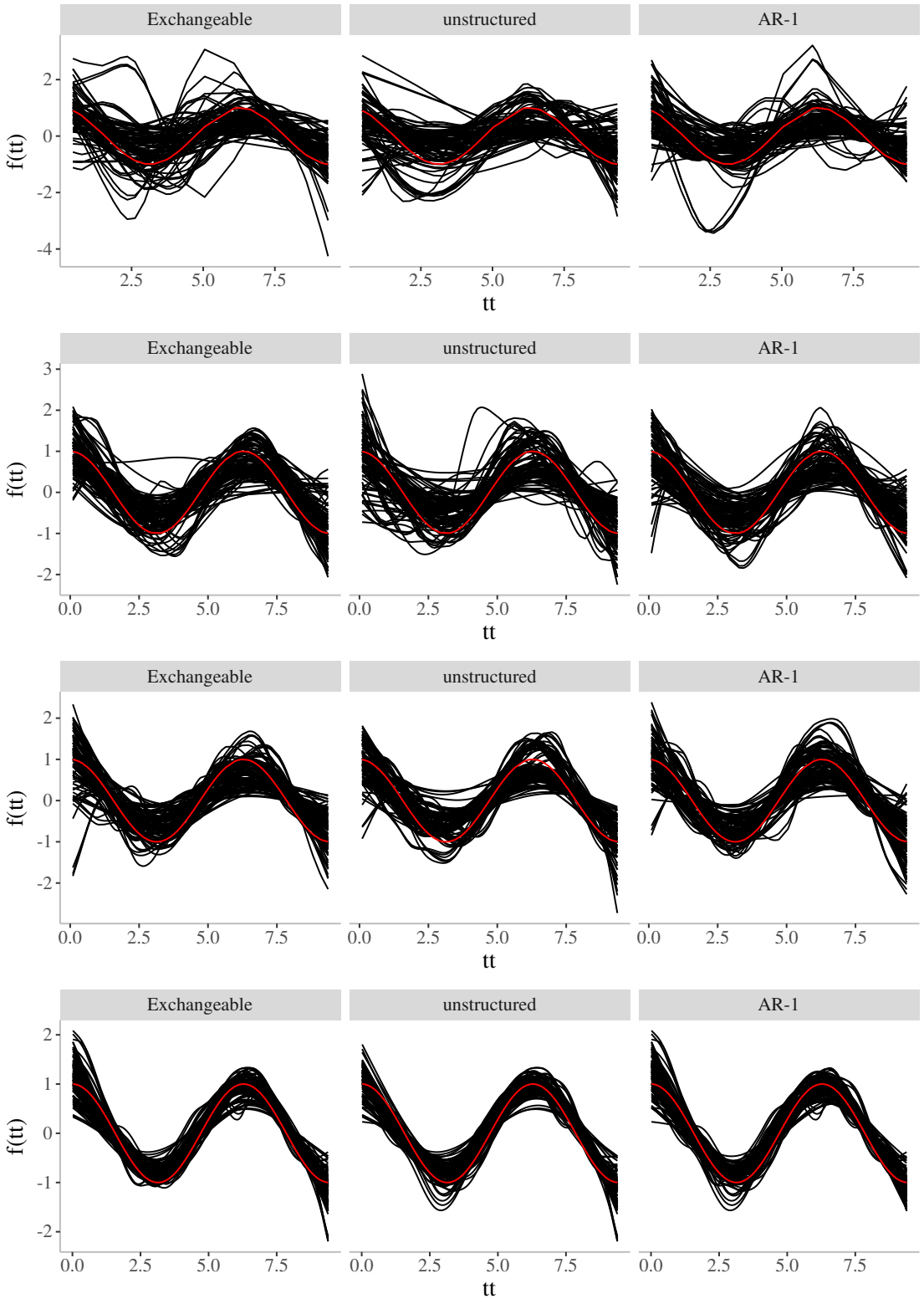


Figure 101 – Simulation study: nonparametric curves for GEE-based CSGGN model for (10,3), (10,10), (50,3) and (50,10), respectively, with $\rho = 0.3$ (exchangeable) by working correlation matrices.

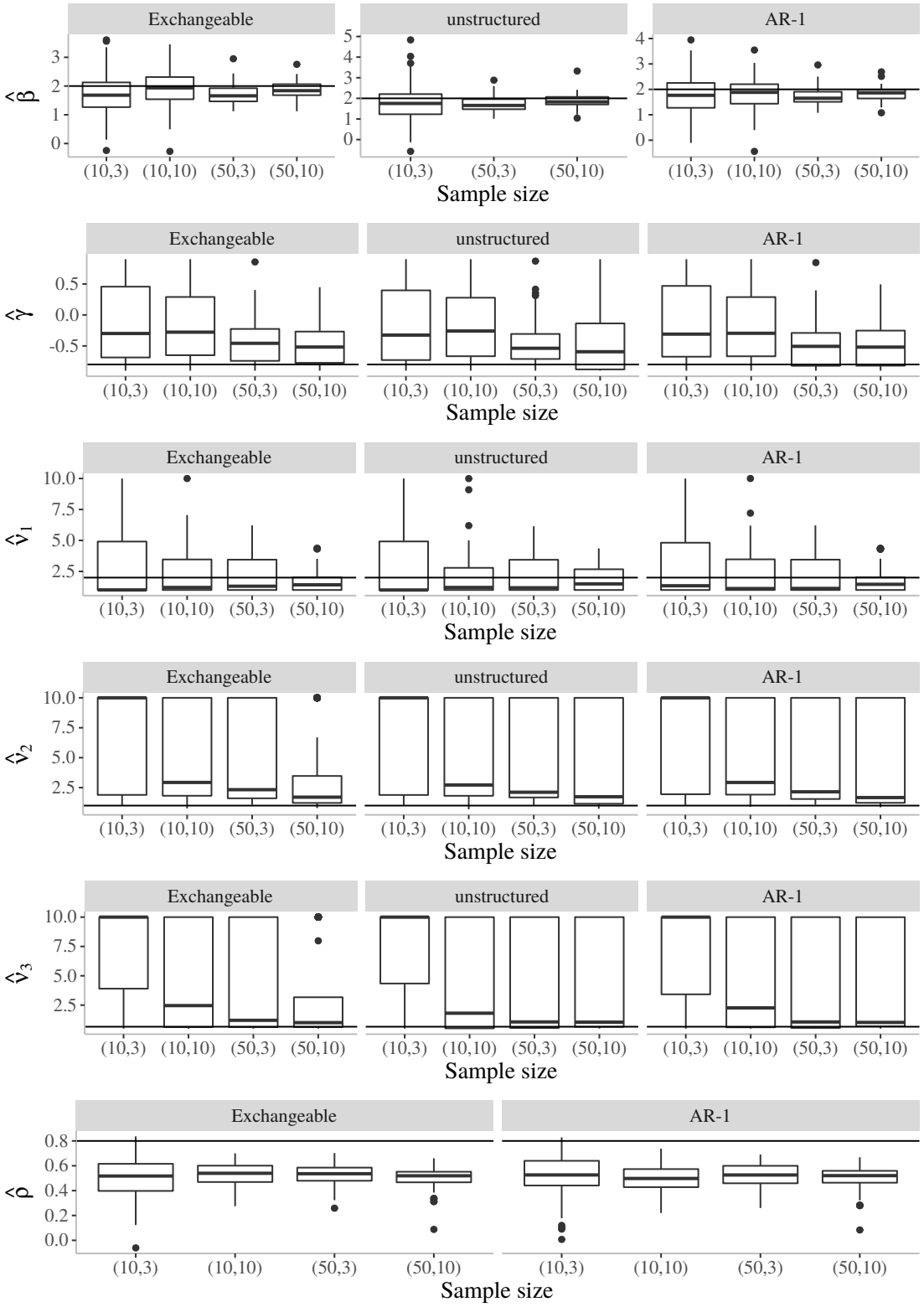


Figure 102 – Simulation study: estimated parameters for GEE-based CSGGN model with $\rho = 0.8$ (exchangeable) by working correlation matrices.

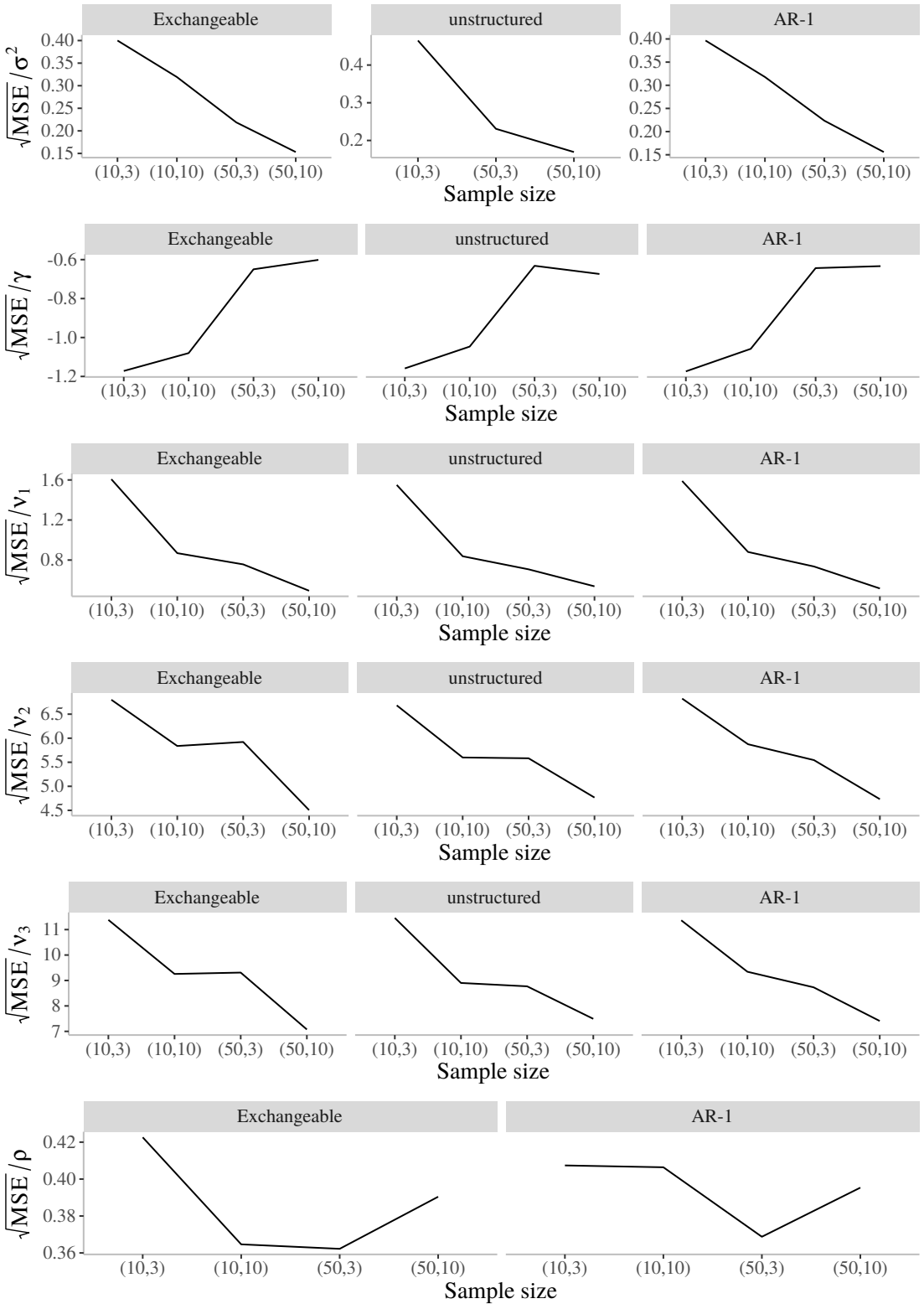


Figure 103 – Simulation study: relative mean square error of the parameters for GEE-based CSGGN model with $\rho = 0.8$ (exchangeable) by working correlation matrices.

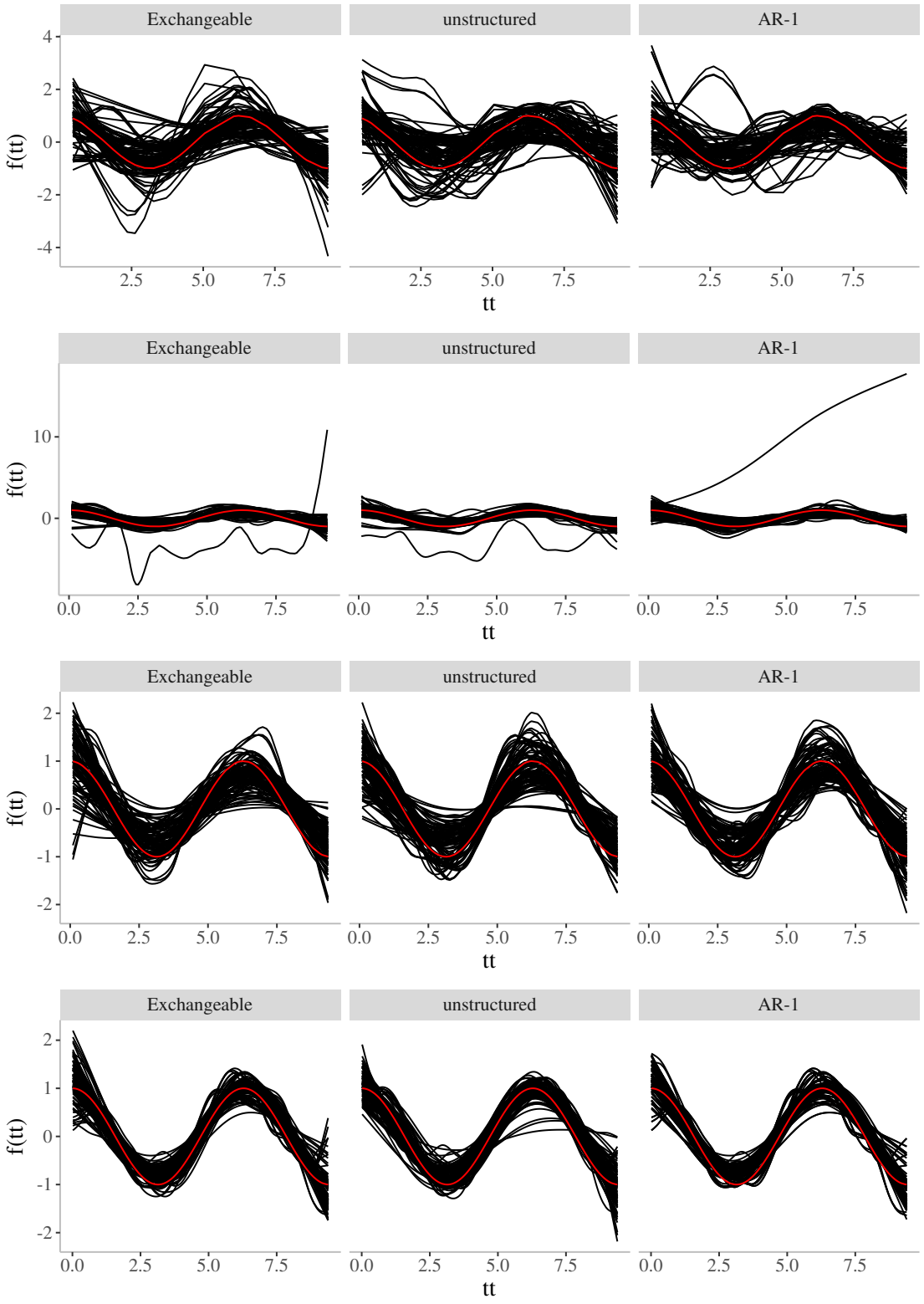


Figure 104 – Simulation study: nonparametric curves for GEE-based CSGGN model for (10,3), (10,10), (50,3) and (50,10), respectively, with $\rho = 0.8$ (exchangeable) by working correlation matrices.