

Measurement Error Models for Replicated Data Under Asymmetric Heavy-Tailed Distributions

Chunzheng Cao^{1,2} · Yahui Wang¹ ·
Jian Qing Shi³ · Jinguan Lin⁴

Accepted: 22 May 2017 / Published online: 30 May 2017
© Springer Science+Business Media New York 2017

Abstract Replicated data with measurement errors are frequently presented in economical, environmental, chemical, medical and other fields. In this paper, we discuss a replicated measurement error model under the class of scale mixtures of skew-normal distributions, which extends symmetric heavy and light tailed distributions to asymmetric cases. We also consider equation error in the model for displaying the matching degree between the true covariate and response. Explicit iterative expressions of maximum likelihood estimates are provided via the expectation–maximization type algorithm. Empirical Bayes estimates are conducted for predicting the true covariate and response. We study the effectiveness as well as the robustness of the maximum likelihood estimations through two simulation studies. The method is applied to analyze a continuing survey data of food intakes by individuals on diet habits.

Keywords EM algorithm · Equation error · Food intakes by individuals · Replicated measurement · Robustness · Scale mixtures of skew-normal distributions

1 Introduction

With the development of social economy and the improvement of living standard, people’s dietary patterns have been greatly changed. Chronic diseases, such as hyper-

✉ Chunzheng Cao
caochunzheng@163.com

¹ School of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing 210044, China

² Department of Statistics, Seoul National University, Seoul 151-742, Korea

³ School of Mathematics and Statistics, University of Newcastle, Newcastle NE1 7RU, UK

⁴ Department of Statistics, Nanjing Audit University, Nanjing 211815, China

tension, dyslipidemia, cardiovascular disease, diabetes mellitus, dietary obesity, have often appeared among different ages in the past decades. The food-consumption behavior and the nutrition arrangement have become hot issues in economical or medical researches (Gori and Sodini 2011; le Coutre et al. 2013, among others). Continuing Survey of Food Intakes by Individuals (CSFII), conducted annually by the US Department of Agriculture (USDA), gathers regular food-consumption information through 24-h individual recalls in 48 conterminous United States. These survey data primarily aim at assessing food consumption behavior and nutritional content of diets for policy implications with respect to food production and marketing, food safety, food assistance and nutrition education (Jacobs et al. 1998). For each CSFII survey, data collection starts in April of the given year and is finished in March of the following year. Harnack et al. (1999) analyzed CSFII data using a logistic regression model, while Carroll et al. (2006) used a similar data set, the National Health and Nutrition Examination Survey data (NHANES) (Jones et al. 1987), to model measurement error directly. In addition, Sun and Empie (2007) carried out a primary statistical analysis using population dietary survey databases of USDA CSFII combined with NHANES 1999–2002.

It is noted that during the process of data collection, measurement error (ME) would inevitably occur in observations of covariates as well as response variables. It may be caused by using different measurement methods and instruments or by human or other external factors. Just like the CSFII data, the ME appears in 24-h individual recalls due to randomness and unidentifiable systematic bias. Ignoring such errors would bring a certain degree of deviations to statistical inferences. As a special tool for handling ME, measurement error models (MEMs) have been comprehensively studied and discussed in literature, see for example, Fuller (1987), Cheng and Ness (1999) and Carroll et al. (2006). According to Reiersol (1950), there exists a non-identifiability problem in normal MEMs, and we have to make some assumptions on error variances in advance, while such assumptions are usually not easy to justify. Fortunately, this problem can be solved if we have replicated data, since the error variance can be estimated either separately or together with other parameters. In CSFII data, the long-term daily intake is measured by daily diets, which is collected repeatedly for several days. An abundant literature discussed the maximum likelihood estimation (MLE) for structural replicated measurement error model (RMEM) (Chan and Mak 1979; Isogawa 1985; Lin et al. 2004; Giménez and Patat 2005; Wimmer and Witkovský 2007; Bartlett et al. 2009, among others). But these studies were all under the normal distribution assumption. However, the normality assumption can be doubtful and the related model lacks robustness against departure from the normal distribution or outlying observations, which may result in misleading conclusions in practice. Recently, Lin and Cao (2013) took advantage of scale mixtures of normal (SMN) distributions for the accommodation of extreme and outlying observations in RMEM (SMN-RMEM) and derived an iterative formulas of MLE through expectation-maximization (EM) algorithm (Dempster et al. 1977; McLachlan and Krishnan 1997). SMN distributions, as a very flexible extension of normal distribution, was developed by Andrews and Mallows (1974), and contains Student t , slash, contaminated normal and other symmetric distributions (Fang et al. 1990; Lange and Sinsheimer 1993). A comprehensive list of applications on SMN distributions in MEMs can be

found, for instance, in [Osorio et al. \(2009\)](#), [Lachos et al. \(2011\)](#) and [Zeller et al. \(2012\)](#). [Cao et al. \(2015\)](#) showed appealing robustness compared to normal ones for the multivariate RMEM. However, the SMN distributions may be violated when a data set contains asymmetric outcomes. Currently, a new class of robust distributions, called scale mixtures of skew-normal (SMSN) distributions, has been developed by [Branco and Dey \(2001\)](#) and becomes very attractive since it can model skewness and heavy tails simultaneously. It includes the whole symmetric family of SMN distributions and its skew version. More details about SMSN distributions can be found in [Genton \(2004\)](#). SMSN distributions have been applied into many models in recent years, for example, linear mixed models ([Arellano-Valle et al. 2005](#); [Jara et al. 2008](#); [Lachos et al. 2010a](#)), nonlinear regression models ([Xie et al. 2008](#); [Cancho et al. 2008](#); [Cao et al. 2014](#)), MEMs ([Lachos et al. 2010b](#); [Montenegro et al. 2010](#); [Zeller et al. 2014](#)).

In this paper, we will develop the RMEM under SMSN distributions (SMSN-RMEM). In particular, the equation error is considered to cope with the random relationship between true covariates and response. In practice, the non-equation error model is often used in natural science, for instance, physics and chemistry, while the equation error model is frequently used in econometrics and the medical sciences ([Cheng and Riu 2006](#)). The presence of equation error could significantly affect the estimation results, as what we will discuss later in this paper. Using EM algorithm, we will calculate MLEs of SMSN-RMEMs with or without equation error. The proposed model shows its effectiveness and robustness of MLEs, which is confirmed by simulation studies and the application.

The rest of this paper is organized as follows. Section 2 gives a brief description of the SMSN distributions. Section 3 proposes the SMSN-RMEM and presents iterative formulas using EM algorithm. The observed information matrix as well as empirical Bayes estimations and predictions are also reported. In Sect. 4, the performance of the model and the importance of equation error are examined via simulation studies. Section 5 applies the model to analyze the inner relationship between saturated fat and caloric intake in CSFII data. Some conclusions are given in Sect. 6.

2 The Scale Mixtures of Skew-Normal Distributions

In this section, we give a brief description of the definition and properties of SMSN distributions. The details can be found in [Branco and Dey \(2001\)](#), [Genton \(2004\)](#) and [Basso et al. \(2010\)](#). An m -dimensional random vector \mathbf{Y} has a skew-normal (SN) distribution, denoted by $\mathbf{Y} \sim \text{SN}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, if its probability density function (pdf) is given by

$$f(\mathbf{y}) = 2\phi_m(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi(A), \quad (1)$$

where $\boldsymbol{\mu}$ is an m -dimensional location vector, $\boldsymbol{\Sigma}$ is a covariance matrix, $\boldsymbol{\lambda}$ is a skewness parameter vector and $A = \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$; ϕ_m is the probability density function (pdf) of m -dimensional normal distribution while Φ is the cumulative distribution function (cdf) of the standard normal distribution. The SMSN random vector \mathbf{Y} can be expressed as

$$Y = \mu + \kappa^{1/2}(U)Z,$$

where $Z \sim \text{SN}_m(\mathbf{0}, \Sigma, \lambda)$, $\kappa(\cdot)$ is a strictly positive weight function, and U is a positive random variable with cdf $H(\cdot; \mathbf{v})$ and pdf $h(\cdot; \mathbf{v})$, independent of Z . Here \mathbf{v} is a scalar or vector indexing the distribution of U .

We now use the notation $Y \sim \text{SMSN}_m(\mu, \Sigma, \lambda; H)$ to stand for scale mixtures of skew-normal distributions. Given $U = u$, the conditional distribution of Y is a multi-SN distribution, i.e., $Y|U = u \sim \text{SN}_m(\mu, \kappa(u)\Sigma, \lambda)$. When $\lambda = \mathbf{0}$, SMSN distributions are deteriorated to SMN distributions $\text{SMN}_m(\mu, \Sigma; H)$.

Note that when $\kappa(u) \equiv 1$, SMSN is a SN distribution. Another special case is the one with $\kappa(u) = 1/u$, i.e., the skew-normal/independent distribution (Lachos et al. 2010a). We define the conditional moments by

$$u_r = E[\kappa^{-r}(U)|y], \quad \eta_r = E\left[\kappa^{-r/2}(U)W_\Phi\left(\kappa^{-1/2}(U)A\right)|y\right],$$

where $W_\Phi(x) = \phi(x)/\Phi(x)$, $x \in \mathbb{R}$. The pdf of Y and the conditional moments under some heavy-tailed SMSN distributions, such as the multivariate skew-Student t (ST), skew-slash (SS), skew-contaminated normal (SCN) distributions, are given in ‘‘Appendix A’’.

3 Model Description and Parameter Estimation

Let x, y be the unobserved (true) covariate and response, respectively. They satisfy an incomplete linear relationship $y = \alpha + \beta x + e$, in which the equation error e means that the true variables x and y are not perfectly related if other factors besides x are also accountable for the variation in y . We usually observe $X = x + \delta$ and $Y = y + \varepsilon$ with measurement errors δ and ε , as the surrogates of x and y . Suppose x and y are respectively observed p and q times to bring out replicated observations $X_t^{(i)}, i = 1, \dots, p$ and $Y_t^{(j)}, j = 1, \dots, q$. Then the SMSN-RMEM with equation error is given by

$$\begin{aligned} X_t^{(i)} &= x_t + \delta_t^{(i)}, \quad i = 1, \dots, p, \\ Y_t^{(j)} &= y_t + \varepsilon_t^{(j)}, \quad j = 1, \dots, q, \\ y_t &= \alpha + \beta x_t + e_t, \quad t = 1, \dots, n, \end{aligned} \tag{2}$$

where, $\delta_t^{(i)}, \varepsilon_t^{(j)}$ and e_t are uncorrelated each other, and

$$\begin{aligned} U_t &\stackrel{iid}{\sim} H(u; \mathbf{v}), \quad t = 1, \dots, n, \\ x_t|U_t &= u_t \stackrel{ind}{\sim} \text{SN}(\mu_x, \kappa(u_t)\phi_x, \lambda_x), \quad \text{i.e., } x_t \stackrel{iid}{\sim} \text{SMSN}(\mu_x, \phi_x, \lambda_x; H), \\ \delta_t^{(i)}|U_t &= u_t \stackrel{ind}{\sim} \text{N}(0, \kappa(u_t)\phi_\delta), \quad \text{i.e., } \delta_t^{(i)} \stackrel{iid}{\sim} \text{SMN}(0, \phi_\delta; H), \end{aligned}$$

$$\begin{aligned} \varepsilon_t^{(j)}|U_t = u_t &\stackrel{ind}{\sim} N(0, \kappa(u_t)\phi_\varepsilon), \text{ i.e., } \varepsilon_t^{(j)} \stackrel{iid}{\sim} \text{SMN}(0, \phi_\varepsilon; H), \\ e_t|U_t = u_t &\stackrel{ind}{\sim} N(0, \kappa(u_t)\phi_e), \text{ i.e., } e_t \stackrel{iid}{\sim} \text{SMN}(0, \phi_e; H). \end{aligned}$$

Note that the observation vector is $\mathbf{Z}_t = (\mathbf{X}_t^\top, \mathbf{Y}_t^\top)^\top$, where $\mathbf{X}_t = (X_t^{(1)}, \dots, X_t^{(p)})^\top$ and $\mathbf{Y}_t = (Y_t^{(1)}, \dots, Y_t^{(q)})^\top$. Motivated by [Montenegro et al. \(2010\)](#) and [Zeller et al. \(2011\)](#), we express model (2) by the following hierarchical representation

$$\begin{aligned} \mathbf{Z}_t|x_t, U_t = u_t &\stackrel{ind}{\sim} N_m(\mathbf{a} + \mathbf{b}x_t, \kappa(u_t)\boldsymbol{\Sigma}_1), \\ x_t|U_t = u_t, V_t = v_t &\stackrel{ind}{\sim} N(\mu_x + \tau_x v_t, \kappa(u_t)\gamma_x), \\ V_t|U_t = u_t &\stackrel{ind}{\sim} \text{HN}(0, \kappa(u_t)), \\ U_t &\stackrel{iid}{\sim} H(u; \mathbf{v}), \end{aligned} \tag{3}$$

where $m = p + q$, $\mathbf{a} = (\mathbf{0}_p^\top, \alpha \mathbf{1}_q^\top)^\top$, $\mathbf{b} = (\mathbf{1}_p^\top, \beta \mathbf{1}_q^\top)^\top$, $\mathbf{c} = (\mathbf{0}_p^\top, \mathbf{1}_q^\top)^\top$, $\boldsymbol{\Sigma}_1 = \mathbf{D}(\boldsymbol{\phi}) + \phi_e \mathbf{c} \mathbf{c}^\top$ with $\boldsymbol{\phi} = (\phi_\delta \mathbf{1}_p^\top, \phi_e \mathbf{1}_q^\top)^\top$, $\tau_x = \phi_x^{1/2} \delta_x$ with $\delta_x = \lambda_x (1 + \lambda_x^2)^{-1/2}$, $\gamma_x = (1 - \delta_x^2) \phi_x = \phi_x / (1 + \lambda_x^2)$, and $\mathbf{D}(\cdot)$ denotes a diagonal matrix whose diagonal elements are formed by a vector.

It is straightforward to know that $\mathbf{Z}_t \sim \text{SMSN}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$, and its conditional distribution $\mathbf{Z}_t|u_t \sim \text{SN}_m(\boldsymbol{\mu}, \kappa(u_t)\boldsymbol{\Sigma}, \boldsymbol{\lambda})$, where the location, scale and skewness parameters can be expressed as $\boldsymbol{\mu} = \mathbf{a} + \mu_x \mathbf{b}$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 + \phi_x \mathbf{b} \mathbf{b}^\top$ and $\boldsymbol{\lambda} = \frac{\lambda_x \phi_x \boldsymbol{\Sigma}^{-1/2} \mathbf{b}}{\sqrt{\phi_x + \lambda_x^2 \Lambda_x}}$ respectively, with $\Lambda_x = \phi_x / c$ and $c = 1 + \phi_x \mathbf{b}^\top \boldsymbol{\Sigma}_1^{-1} \mathbf{b}$.

3.1 EM Algorithm

It is intractable to calculate MLEs effectively through common likelihood method. In the assumption of SMN distributions, [Lin and Cao \(2013\)](#) proposed an iterative approach using EM algorithm in RMEM without assuming equation errors. Here we devote to get the MLEs of model (2) by EM algorithm.

Denoting the parameter vector of model (2) by $\boldsymbol{\theta} = (\mu_x, \alpha, \beta, \lambda_x, \phi_x, \phi_\delta, \phi_\varepsilon, \phi_e)^\top$, the estimates of $\boldsymbol{\theta}$ at the k -th iteration by $\widehat{\boldsymbol{\theta}}^{(k)}$ and the complete data set of model (2) by $\mathbf{Z}_c = \{\mathbf{Z}, \mathbf{x}, \mathbf{v}, \mathbf{u}\}$, where $\mathbf{Z} = \{\mathbf{Z}_1, \dots, \mathbf{Z}_n\}$, $\mathbf{x} = (x_1, \dots, x_n)^\top$, $\mathbf{v} = (v_1, \dots, v_n)^\top$, $\mathbf{u} = (u_1, \dots, u_n)^\top$, then the complete log-likelihood function based on \mathbf{Z}_c is given by

$$l(\boldsymbol{\theta}|\mathbf{Z}_c) = \sum_{t=1}^n (l_{x_t|v_t, u_t} + l_{\mathbf{Z}_t|x_t, u_t}),$$

where,

$$l_{x_t|v_t,u_t} = -\frac{1}{2} \log(\gamma_x) - \frac{1}{2} \gamma_x^{-1} \kappa^{-1}(u_t) (x_t - \mu_x - \tau_x v_t)^2,$$

$$l_{\mathbf{Z}_t|x_t,u_t} = -\frac{1}{2} \log(|\Sigma_1|) - \frac{1}{2} \kappa^{-1}(u_t) (\mathbf{Z}_t - \mathbf{a} - \mathbf{b}x_t)^\top \Sigma_1^{-1} (\mathbf{Z}_t - \mathbf{a} - \mathbf{b}x_t).$$

with $\Sigma_1^{-1} = \text{diag}\{\frac{1}{\phi_\delta} \mathbf{I}_p, \frac{1}{\phi_\varepsilon} \mathbf{I}_q - \frac{\phi_\varepsilon}{\phi_\varepsilon(\phi_\varepsilon + q\phi_e)} \mathbf{1}_q \mathbf{1}_q^\top\}$ and $|\Sigma_1| = \phi_\delta^p \phi_\varepsilon^{q-1} (\phi_\varepsilon + q\phi_e)$. We ignore the constant which is independent of θ in above expressions. The EM algorithm is constructed as follows.

E-step: Given the value of $\widehat{\theta}^{(k)}$, we calculate the Q -function by using the property of conditional expectation $E[l(\theta|\mathbf{Z}_c)|\widehat{\theta}^{(k)}, \mathbf{Z}]$. The Q -function takes the following form

$$Q(\theta|\widehat{\theta}^{(k)}) = \sum_{t=1}^n [Q_{1t}(\theta|\widehat{\theta}^{(k)}) + Q_{2t}(\theta|\widehat{\theta}^{(k)})], \tag{4}$$

where

$$Q_{1t}(\theta|\widehat{\theta}^{(k)}) = -\frac{1}{2} \log(\gamma_x) - \frac{1}{2} \gamma_x^{-1} (\widehat{u}x_t^{(k)} + \mu_x^2 \widehat{u}_t^{(k)} + \tau_x^2 \widehat{u}v_t^{(k)} - 2\mu_x \widehat{u}x_t^{(k)} - 2\tau_x \widehat{u}x_t^{(k)}v_t^{(k)} + 2\mu_x \tau_x \widehat{u}v_t^{(k)}),$$

$$Q_{2t}(\theta|\widehat{\theta}^{(k)}) = -\frac{1}{2} \log(|\Sigma_1|) - \frac{1}{2} [\widehat{u}_t^{(k)} (\mathbf{Z}_t - \mathbf{a})^\top \Sigma_1^{-1} (\mathbf{Z}_t - \mathbf{a}) - 2\widehat{u}x_t^{(k)} (\mathbf{Z}_t - \mathbf{a})^\top \Sigma_1^{-1} \mathbf{b} + \widehat{u}x_t^{(k)} \mathbf{b}^\top \Sigma_1^{-1} \mathbf{b}],$$

with $\widehat{u}_t^{(k)} = E[\kappa^{-1}(U_t)|\widehat{\theta}^{(k)}, \mathbf{Z}_t]$, $\widehat{u}v_t^{(k)} = E[\kappa^{-1}(U_t)v_t|\widehat{\theta}^{(k)}, \mathbf{Z}_t]$, $\widehat{u}v_t^{(k)2} = E[\kappa^{-1}(U_t)v_t^2|\widehat{\theta}^{(k)}, \mathbf{Z}_t]$, $\widehat{u}x_t^{(k)} = E[\kappa^{-1}(U_t)x_t|\widehat{\theta}^{(k)}, \mathbf{Z}_t]$, $\widehat{u}x_t^{(k)2} = E[\kappa^{-1}(U_t)x_t^2|\widehat{\theta}^{(k)}, \mathbf{Z}_t]$, $\widehat{u}x_t^{(k)}v_t^{(k)} = E[\kappa^{-1}(U_t)x_tv_t|\widehat{\theta}^{(k)}, \mathbf{Z}_t]$, and they can be readily evaluated by

$$\begin{aligned} \widehat{u}v_t^{(k)} &= \widehat{u}_t^{(k)} \widehat{\mu}_{v_t}^{(k)} + \widehat{M}_v^{(k)} \widehat{\eta}_{1,t}^{(k)}, \\ \widehat{u}v_t^{(k)2} &= \widehat{u}_t^{(k)} \widehat{\mu}_{v_t}^{(k)2} + \widehat{M}_v^{(k)2} + \widehat{M}_v^{(k)} \widehat{\mu}_{v_t}^{(k)} \widehat{\eta}_{1,t}^{(k)}, \\ \widehat{u}x_t^{(k)} &= \widehat{u}_t^{(k)} \widehat{r}_t^{(k)} + \widehat{u}v_t^{(k)} \widehat{s}^{(k)}, \\ \widehat{u}x_t^{(k)2} &= \widehat{u}_t^{(k)} \widehat{r}_t^{(k)2} + 2\widehat{u}v_t^{(k)} \widehat{r}_t^{(k)} \widehat{s}^{(k)} + \widehat{u}v_t^{(k)2} \widehat{s}^{(k)2} + \widehat{\gamma}_x^{(k)} / \widehat{c}_1^{(k)}, \\ \widehat{u}x_t^{(k)}v_t^{(k)} &= \widehat{u}v_t^{(k)} \widehat{r}_t^{(k)} + \widehat{u}v_t^{(k)2} \widehat{s}^{(k)}, \end{aligned}$$

with $\widehat{\mu}_{v_t} = \widehat{\tau}_x \widehat{a}_t / \widehat{c}$, $\widehat{M}_v = \sqrt{\widehat{c}_1 / \widehat{c}}$, $\widehat{\eta}_{1,t}^{(k)} = E[\kappa^{-1/2}(U_t) W_\Phi(\kappa^{-1/2}(U_t) \widehat{\mu}_{v_t} / \widehat{M}_v) | \widehat{\theta}^{(k)}, \mathbf{Z}_t]$, $\widehat{r}_t = \widehat{\mu}_x + \widehat{\gamma}_x \widehat{a}_t / \widehat{c}_1$, $\widehat{s} = \widehat{\tau}_x / \widehat{c}_1$, $a_t = (\mathbf{Z}_t - \boldsymbol{\mu})^\top \Sigma_1^{-1} \mathbf{b}$, $c_1 = 1 + \gamma_x \mathbf{b}^\top \Sigma_1^{-1} \mathbf{b}$.

M-step: Maximizing $Q(\theta|\hat{\theta}^{(k)})$ with respect to θ , we achieve the updated estimates $\hat{\theta}^{(k+1)}$ by the following iterative equations:

$$\begin{aligned} \hat{\mu}_x^{(k+1)} &= \frac{\sum_{t=1}^n \widehat{u}x_t^{(k)} \sum_{t=1}^n \widehat{u}v_t^{2(k)} - \sum_{t=1}^n \widehat{u}v_t^{(k)} \sum_{t=1}^n \widehat{u}xv_t^{(k)}}{\sum_{t=1}^n \widehat{u}_t^{(k)} \sum_{t=1}^n \widehat{u}v_t^{2(k)} - \left(\sum_{t=1}^n \widehat{u}v_t^{(k)}\right)^2}, \\ \hat{\tau}_x^{(k+1)} &= \frac{\sum_{t=1}^n \widehat{u}_t^{(k)} \sum_{t=1}^n \widehat{u}xv_t^{(k)} - \sum_{t=1}^n \widehat{u}v_t^{(k)} \sum_{t=1}^n \widehat{u}x_t^{(k)}}{\sum_{t=1}^n \widehat{u}_t^{(k)} \sum_{t=1}^n \widehat{u}v_t^{2(k)} - \left(\sum_{t=1}^n \widehat{u}v_t^{(k)}\right)^2}, \\ \hat{\gamma}_x^{(k+1)} &= \frac{1}{n} \sum_{t=1}^n \left(\widehat{u}x_t^{2(k)} + \left(\hat{\mu}_x^{(k+1)}\right)^2 \widehat{u}_t^{(k)} + \left(\hat{\tau}_x^{(k+1)}\right)^2 \widehat{u}v_t^{2(k)} \right. \\ &\quad \left. - 2\hat{\mu}_x^{(k+1)}\widehat{u}x_t^{(k)} - 2\hat{\tau}_x^{(k+1)}\widehat{u}xv_t^{(k)} + 2\hat{\mu}_x^{(k+1)}\hat{\tau}_x^{(k+1)}\widehat{u}v_t^{(k)} \right), \\ \hat{\alpha}^{(k+1)} &= \frac{\sum_{t=1}^n \left(\widehat{u}_t^{(k)} \bar{Y}_t\right) \sum_{t=1}^n \widehat{u}x_t^{2(k)} - \sum_{t=1}^n \widehat{u}x_t^{(k)} \sum_{t=1}^n \left(\widehat{u}x_t^{(k)} \bar{Y}_t\right)}{\sum_{t=1}^n \widehat{u}_t^{(k)} \sum_{t=1}^n \widehat{u}x_t^{2(k)} - \left(\sum_{t=1}^n \widehat{u}x_t^{(k)}\right)^2}, \\ \hat{\beta}^{(k+1)} &= \frac{\sum_{t=1}^n \widehat{u}_t^{(k)} \sum_{t=1}^n \left(\widehat{u}x_t^{(k)} \bar{Y}_t\right) - \sum_{t=1}^n \widehat{u}x_t^{(k)} \sum_{t=1}^n \left(\widehat{u}_t^{(k)} \bar{Y}_t\right)}{\sum_{t=1}^n \widehat{u}_t^{(k)} \sum_{t=1}^n \widehat{u}x_t^{2(k)} - \left(\sum_{t=1}^n \widehat{u}x_t^{(k)}\right)^2}, \\ \hat{\phi}_\delta^{(k+1)} &= \frac{1}{n} \sum_{t=1}^n \left[\frac{1}{p} \widehat{u}_t^{(k)} \sum_{i=1}^p \left(X_t^{(i)}\right)^2 - 2\widehat{u}x_t^{(k)} \bar{X}_t + \widehat{u}x_t^{2(k)} \right], \\ \hat{\phi}_\varepsilon^{(k+1)} &= \frac{1}{n(q-1)} \left(\sum_{t=1}^n \hat{\pi}_{1t}^{(k+1)} - q \sum_{t=1}^n \hat{\pi}_{2t}^{(k+1)} \right), \\ \hat{\phi}_e^{(k+1)} &= \frac{1}{n(q-1)} \left(q \sum_{t=1}^n \hat{\pi}_{2t}^{(k+1)} - \frac{1}{q} \sum_{t=1}^n \hat{\pi}_{1t}^{(k+1)} \right), \end{aligned}$$

where,

$$\begin{aligned} \bar{X}_t &= \frac{1}{p} \sum_{i=1}^p X_t^{(i)}, \bar{Y}_t = \frac{1}{q} \sum_{j=1}^q Y_t^{(j)}, \\ \hat{\pi}_{1t}^{(k+1)} &= \widehat{u}_t^{(k)} \sum_{j=1}^q \left(Y_t^{(j)} - \hat{\alpha}^{(k+1)} \right)^2 - 2q\widehat{u}x_t^{(k)} \left(\bar{Y}_t - \hat{\alpha}^{(k+1)} \right) \hat{\beta}^{(k+1)} \\ &\quad + q\widehat{u}x_t^{2(k)} \left(\hat{\beta}^{(k+1)} \right)^2, \\ \hat{\pi}_{2t}^{(k+1)} &= \widehat{u}_t^{(k)} \left(\bar{Y}_t - \hat{\alpha}^{(k+1)} \right)^2 - 2\widehat{u}x_t^{(k)} \left(\bar{Y}_t - \hat{\alpha}^{(k+1)} \right) \hat{\beta}^{(k+1)} + \widehat{u}x_t^{2(k)} \left(\hat{\beta}^{(k+1)} \right)^2. \end{aligned}$$

Note that the estimators $\widehat{\lambda}_x$ and $\widehat{\phi}_x$ can be inferred from the one-to-one transformation $\lambda_x = \tau_x / \sqrt{\gamma_x}$ and $\phi_x = \gamma_x + \tau_x^2$. Thus the iterative expressions for these two parameters can be written as

$$\begin{aligned} \widehat{\lambda}_x^{(k+1)} &= \widehat{\tau}_x^{(k+1)} / \sqrt{\widehat{\gamma}_x^{(k+1)}}, \\ \widehat{\phi}_x^{(k+1)} &= \widehat{\gamma}_x^{(k+1)} + \left(\widehat{\tau}_x^{(k+1)}\right)^2. \end{aligned}$$

Starting from suitable initial values (for example, the MLEs from the SMN-RMEM or the moment estimators), the iterations of the above EM algorithm between E-step and M-step are repeated until a suitable convergence rule is satisfied, e.g., $\|\widehat{\boldsymbol{\theta}}^{(k+1)} - \widehat{\boldsymbol{\theta}}^{(k)}\|$ is sufficiently small. We will apply a closed form (Harville 1997) to compute the inverse of matrix $\boldsymbol{\Sigma}_1$ as

$$\boldsymbol{\Sigma}_1^{-1} = \mathbf{D}^{-1}(\boldsymbol{\phi}) - \frac{\phi_e \mathbf{D}^{-1}(\boldsymbol{\phi}) \mathbf{c} \mathbf{c}^\top \mathbf{D}^{-1}(\boldsymbol{\phi})}{1 + \phi_e \mathbf{c}^\top \mathbf{D}^{-1}(\boldsymbol{\phi}) \mathbf{c}}.$$

This can guarantee a positive covariance matrix.

3.2 The Observed Information Matrix

Asymptotic confidence intervals of the MLEs can be constructed using the observed information matrix. We have inferred $\mathbf{Z}_t \sim \text{SMSN}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; \text{H})$, so the log-likelihood function of \mathbf{Z} respect to $\boldsymbol{\theta}$ is given by $l(\boldsymbol{\theta}) = \sum_{t=1}^n l_t(\boldsymbol{\theta})$, where

$$l_t(\boldsymbol{\theta}) = \log 2 - \frac{m}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}| + \log(K_t), \tag{5}$$

with $K_t = \int_0^\infty \kappa^{-m/2}(u_t) \exp\{-\kappa^{-1}(u_t)d_t/2\} \Phi(\kappa^{-1/2}(u_t)A_t) d\text{H}(u_t)$, $d_t = (\mathbf{Z}_t - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Z}_t - \boldsymbol{\mu})$, and $A_t = \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{Z}_t - \boldsymbol{\mu})$.

We use the observed information matrix with the help of score function for $\boldsymbol{\theta}$. It is expressed by

$$\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{1}{2} \frac{\partial \log |\boldsymbol{\Sigma}|}{\partial \boldsymbol{\theta}} + \frac{1}{K_t} \frac{\partial K_t}{\partial \boldsymbol{\theta}}, \tag{6}$$

where, $\frac{\partial K_t}{\partial \boldsymbol{\theta}} = I_t^\phi \left(\frac{m+1}{2}\right) \frac{\partial A_t}{\partial \boldsymbol{\theta}} - \frac{1}{2} I_t^\phi \left(\frac{m+2}{2}\right) \frac{\partial d_t}{\partial \boldsymbol{\theta}}$, and in which,

$$\begin{aligned} I_t^\phi(w) &= \int_0^\infty \kappa^{-w}(u_t) \exp\left\{-\kappa^{-1}(u_t)d_t/2\right\} \Phi\left(\kappa^{-1/2}(u_t)A_t\right) d\text{H}(u_t), \\ I_t^\phi(w) &= \int_0^\infty \kappa^{-w}(u_t) \exp\left\{-\kappa^{-1}(u_t)d_t/2\right\} \phi\left(\kappa^{-1/2}(u_t)A_t\right) d\text{H}(u_t). \end{aligned}$$

Then the asymptotic covariance of the MLEs can be estimated by the inverse of the observed information matrix $I(\hat{\theta}) = \sum_{t=1}^n \left(\frac{\partial l_t(\theta)}{\partial \theta} \right) \left(\frac{\partial l_t(\theta)}{\partial \theta} \right)^\top |_{\theta=\hat{\theta}}$. Explicit expressions of $I_t^\Phi(w)$ and $I_t^\phi(w)$ for some SMSN distributions can be found in Lachos et al. (2010a) and Zeller et al. (2011). The derivatives of the score function are given in ‘‘Appendix B’’.

3.3 Empirical Bayes Estimation and Prediction

In this section, we consider an empirical Bayes approach to estimate the true variable x_t and to predict Y_t given X_t . From the definition of the models, it is not difficult to derive that the conditional distribution of x_t given Z_t and u_t belongs to the extended SN (ESN) distributions (Azzalini and Capitanio 1999), with the conditional pdf as

$$f(x_t | Z_t, u_t) = \phi(x_t; \mu_x + \Lambda_x a_t, \kappa(u_t) \Lambda_x) \frac{\Phi\left(\kappa^{-1/2}(u_t) \lambda_x \phi_x^{-1/2}(x_t - \mu_x)\right)}{\Phi(\kappa^{-1/2}(u_t) A_t)}.$$

Thus, we get the conditional expectation of x_t given Z_t and u_t as

$$E[x_t | Z_t, u_t] = \mu_x + \Lambda_x a_t + \kappa^{1/2}(u_t) \frac{\lambda_x \Lambda_x}{\sqrt{1 + \lambda_x^2 \Lambda_x}} W_\Phi\left(\kappa^{-1/2}(u_t) A_t\right).$$

Then

$$\begin{aligned} \hat{x}_t &= E[x_t | Z_t] = E\{E[x_t | Z_t, u_t] | Z_t\} \\ &= \mu_x + \Lambda_x a_t + \frac{\lambda_x \Lambda_x}{\sqrt{1 + \lambda_x^2 \Lambda_x}} \eta_{-1,t}, \end{aligned}$$

where $\eta_{-1,t} = E[\kappa^{1/2}(U_t) W_\Phi(\kappa^{-1/2}(U_t) A_t) | Z_t]$.

Next, we will predict Y_t given X_t . From the previous conclusions, we know that $Z_t \sim \text{SMSN}_m(\mu, \Sigma, \lambda; H)$, and $Z_t | u_t \sim \text{SN}_m(\mu, \kappa(u_t) \Sigma, \lambda)$. Now, partition Σ as

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where $\Sigma_{11} = \phi_\delta I_p + \phi_x \mathbf{1}_p \mathbf{1}_p^\top$, $\Sigma_{12} = \beta \phi_x \mathbf{1}_p \mathbf{1}_q^\top$, $\Sigma_{21} = \Sigma_{12}^\top$, $\Sigma_{22} = \phi_\varepsilon I_q + (\phi_e + \beta^2 \phi_x) \mathbf{1}_q \mathbf{1}_q^\top$. Then we have marginal distribution

$$X_t \sim \text{SMSN}_p(\mu_x \mathbf{1}_p, \Sigma_{11}, \lambda_{1t}; H),$$

where $\lambda_{1t} = \frac{\lambda_x \phi_x}{\sqrt{1 + \lambda_x^\top \lambda}} \Sigma_{11,t}^{-1/2} \mathbf{1}_{p_t} / \sqrt{1 - \frac{\rho^2 p_t}{\phi_{\delta_t} + p_t \phi_x}}$, i.e., $X_t | u_t \sim \text{SN}_{p_t}(\mu_x \mathbf{1}_{p_t}, \kappa(u_t) \Sigma_{11,t}, \lambda_{1t})$.

Denote $A_{1,t} = \lambda_1^\top \Sigma_{11}^{-1/2} (\mathbf{X}_t - \mu_x \mathbf{1}_p) = \frac{\rho p}{\phi_\delta + p\phi_x} (\bar{X}_t - \mu_x) / \sqrt{1 - \frac{\rho^2 p}{\phi_\delta + p\phi_x}}$, with $\rho = \lambda_x \phi_x / \sqrt{1 + \lambda^\top \lambda}$, then $\mathbf{Y}_t | \mathbf{X}_t, u_t \sim \text{ESN}_q(\boldsymbol{\mu}_{2\cdot 1,t}, \kappa(u_t) \boldsymbol{\Sigma}_{22\cdot 1}; \lambda_{0,t}, \boldsymbol{\lambda}_{1,t})$, where

$$\begin{aligned} \boldsymbol{\mu}_{2\cdot 1,t} &= \left[\alpha + \beta \mu_x + \frac{p\phi_x}{\phi_\delta + p\phi_x} \beta (\bar{X}_t - \mu_x) \right] \mathbf{1}_q, \\ \boldsymbol{\Sigma}_{22\cdot 1} &= \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} = \phi_\varepsilon \mathbf{I}_q + \left(\phi_\varepsilon + \frac{\phi_\delta \phi_x \beta^2}{\phi_\delta + p\phi_x} \right) \mathbf{1}_q \mathbf{1}_q^\top, \\ \lambda_{0,t} &= \kappa^{-1/2}(u_t) \frac{\rho p}{\phi_\delta + p\phi_x} (\bar{X}_t - \mu_x) / \sqrt{1 - \rho^2 \mathbf{b} \boldsymbol{\Sigma}^{-1} \mathbf{b}}, \\ \boldsymbol{\lambda}_{1,t} &= \frac{\rho \beta \phi_\delta}{\phi_\delta + p\phi_x} \boldsymbol{\Sigma}_{22\cdot 1}^{-1/2} \mathbf{1}_q / \sqrt{1 - \rho^2 \mathbf{b} \boldsymbol{\Sigma}^{-1} \mathbf{b}}. \end{aligned}$$

Accordingly, we get the conditional expectation of \mathbf{Y}_t given \mathbf{X}_t and u_t as

$$\begin{aligned} E[\mathbf{Y}_t | \mathbf{X}_t, u_t] &= \boldsymbol{\mu}_{2\cdot 1,t} + \kappa^{1/2}(u_t) \boldsymbol{\Sigma}_{22\cdot 1}^{1/2} \frac{\boldsymbol{\lambda}_{1,t}}{1 + \boldsymbol{\lambda}_{1,t}^\top \boldsymbol{\lambda}_{1,t}} W_\Phi \left(\kappa^{-1/2}(u_t) A_{1,t} \right) \\ &= \boldsymbol{\mu}_{2\cdot 1,t} + \kappa^{1/2}(u_t) W_\Phi \left(\kappa^{-1/2}(u_t) A_{1,t} \right) \frac{\rho \beta \phi_\delta}{\phi_\delta + p\phi_x} \mathbf{1}_q / \sqrt{1 - \frac{\rho^2 p}{\phi_\delta + p\phi_x}}, \end{aligned}$$

Then, the prediction of \mathbf{Y}_t is given by

$$\begin{aligned} \hat{\mathbf{Y}}_t &= E[\mathbf{Y}_t | \mathbf{X}_t] = E\{E[\mathbf{Y}_t | \mathbf{X}_t, u_t] | \mathbf{X}_t\} \\ &= \boldsymbol{\mu}_{2\cdot 1,t} + \tilde{\eta}_{-1,t} \frac{\rho \beta \phi_\delta}{\phi_\delta + p\phi_x} \mathbf{1}_q / \sqrt{1 - \frac{\rho^2 p}{\phi_\delta + p\phi_x}}, \end{aligned} \tag{7}$$

where $\tilde{\eta}_{-1,t} = E[\kappa^{1/2}(U_t) W_\Phi(\kappa^{-1/2}(U_t) A_{1,t}) | \mathbf{X}_t]$.

4 Simulation Studies

In this section, we will report the results of two simulation studies. The first one is to confirm the effectiveness and accuracy of the MLEs under SMSN distributions. The second one is to investigate the robustness of SMSN-RMEM when there has outliers. In each simulation study, we consider three skew distributions, including SN, ST with $\nu = 4$ and SCN with $\nu = 0.2, \gamma = 0.3$, in RMEM with or without equation error, respectively. Under the circumstances of no equation error, we denote the SN, ST and SCN distributions by SN0, ST0 and SCN0 correspondingly. The regression parameters α and β will be regarded as parameters of interest in both simulation studies.

4.1 The First Simulation Study

In this simulation study, the parameters of SMSN-RMEM model are set as $\mu_x = 1.5, \alpha = 2, \beta = 1, \lambda_x = 0.5, \phi_x = 1, \phi_\delta = 1$ and $\phi_e = 0.5$. For comparison, we set $\phi_e = 0.5, 1$ and 1.5 respectively, which shows that the matching degree between the true covariate and response tends from strong to weak. The replicated numbers of the observations are chosen as $p = 4$ and $q = 3$. We generate data from model (2) with sample size $n = 50$ or 100 under ST distribution ($\nu = 4$). Based on the sample data, we compute the MLEs of θ using EM algorithm under SN, ST and SCN distributions with or without equation error, respectively. After 1000 times repetition, we calculate the sample bias (BIAS) and the standard deviation (SD) as assessments for the estimates.

Table 1 displays the estimations of α and β based on six type SMSN-RMEMs (SN-RMEM, ST-RMEM, SCN-RMEM, SN0-RMEM, ST0-RMEM, SCN0-RMEM).

Table 1 Performances of estimators under ST-RMEM with or without equation error

Parameter	n	Estimator	$\phi_e = 0.5$		$\phi_e = 1$		$\phi_e = 1.5$	
			BIAS	SD	BIAS	SD	BIAS	SD
α	50	SN	-0.131	0.674	-0.128	0.811	-0.155	0.821
		ST	-0.062	0.430	-0.047	0.506	-0.072	0.570
		SCN	-0.081	0.519	-0.076	0.605	-0.088	0.622
		SN0	-0.972	1.280	-2.093	2.084	-3.239	2.611
		ST0	-0.806	0.572	-1.853	1.121	-2.857	1.585
		SCN0	-0.853	0.699	-1.908	1.246	-2.946	1.739
	100	SN	-0.061	0.445	-0.064	0.552	-0.080	0.546
		ST	-0.026	0.287	-0.027	0.346	-0.030	0.377
		SCN	-0.036	0.327	-0.037	0.396	-0.044	0.417
		SN0	-0.890	0.687	-2.005	1.276	-2.996	1.804
		ST0	-0.800	0.408	-1.790	0.714	-2.800	0.998
		SCN0	-0.833	0.481	-1.864	0.863	-2.826	1.134
β	50	SN	0.067	0.343	0.060	0.411	0.083	0.419
		ST	0.032	0.214	0.021	0.247	0.039	0.287
		SCN	0.041	0.264	0.035	0.302	0.049	0.314
		SN0	0.506	0.667	1.081	1.024	1.690	1.344
		ST0	0.431	0.290	0.989	0.575	1.541	0.838
		SCN0	0.451	0.359	1.006	0.631	1.572	0.916
	100	SN	0.025	0.225	0.026	0.281	0.040	0.279
		ST	0.011	0.141	0.011	0.171	0.014	0.191
		SCN	0.015	0.162	0.014	0.197	0.021	0.214
		SN0	0.453	0.363	1.027	0.666	1.551	0.965
		ST0	0.427	0.207	0.958	0.367	1.496	0.513
		SCN0	0.437	0.249	0.982	0.449	1.491	0.596

As we expected, No matter it is α or β , the BIAS and SD of all estimates get smaller as samples size increasing from 50 to 100. The BIAS and SD under ST distribution are almost always the smallest for all cases, which fully reflects the effectiveness and accuracy of the ML estimates. Moreover, the performance under SCN and ST distribution behaves better than SN distribution, which may attribute to their heavy-tailed features.

Obviously, the estimates with equation error consistently perform better than the estimates without equation error. Especially, when ϕ_e increases from 0.5 to 1.5, the ratios of SD without equation error and with equation error become clearly larger. It states that for the data with skewness or heavy-tailed features, ignoring equation error will bring serious deviation for statistical inference. The equation error plays an important role in expressing the uncertain relationship between the true covariates and the response.

4.2 The Second Simulation Study

In this simulation study, we aim at examining the robustness of skew-heavy tailed models by comparing the performance of estimates in the presence of outliers. We first generate a data set with sample size $n = 100$ from model (2) under SN distribution, and take the first half as training data and the rest as testing data. Then we shift the observed value $X_t^{(i)}$ to $X_t^{(i)} + w\mu_x$ and $Y_{t'}^{(j)}$ to $Y_{t'}^{(j)} + w(\alpha + \beta\mu_x)$ for all i, j , respectively, where the two subject number t and t' are chosen randomly from training data each time. The value of w , from -3 to 3 , indicates the degree of contamination. All the true values of parameters are the same as the first simulation study. After that, we calculate the MLEs of α and β based on six type SMSN-RMEMs (mentioned in the first simulation study) under the shifted data and the non-shifted data, respectively. For the testing data, we use the empirical Bayes method to predict Y_t according to the observed values of X_t under the aforementioned six type SMSN-RMEMs. Based on 1000 simulations, the mean squared error (MSE) values of α , β and the predictions at different w are computed and shown in Fig. 1 for the six models. The smaller MSE value is, the better robustness of model will behave.

Whether it is α or β , MSEs of estimates without assuming equation error are always larger than the ones assuming equation error. It indicates the significant importance of equation error for skew or heavy-tailed data. This finding matches the results from the first simulation study. The MSEs of estimates based on SN, SN0, ST0 and SCN0 models increase quickly along with w , which shows that the influence of the outliers become serious when the absolute value of w increases. On the contrary, the MSEs of estimates based on ST and SCN models almost do not change with w , which reveals the robustness of ST and SCN models. As expected, the MSEs are the smallest at the zero point (i.e., no perturbation in the data) for all the cases. The impacts of positive disturbance are almost as much as negative disturbance. Although the true data follow the SN distribution, the superiority of robustness based on heavy-tailed models is obvious. Similar conclusions for predictions can be drawn from Fig. 1c. That is to say, the robustness of skew-heavy tailed distributions are also efficient for Bayes prediction approach.

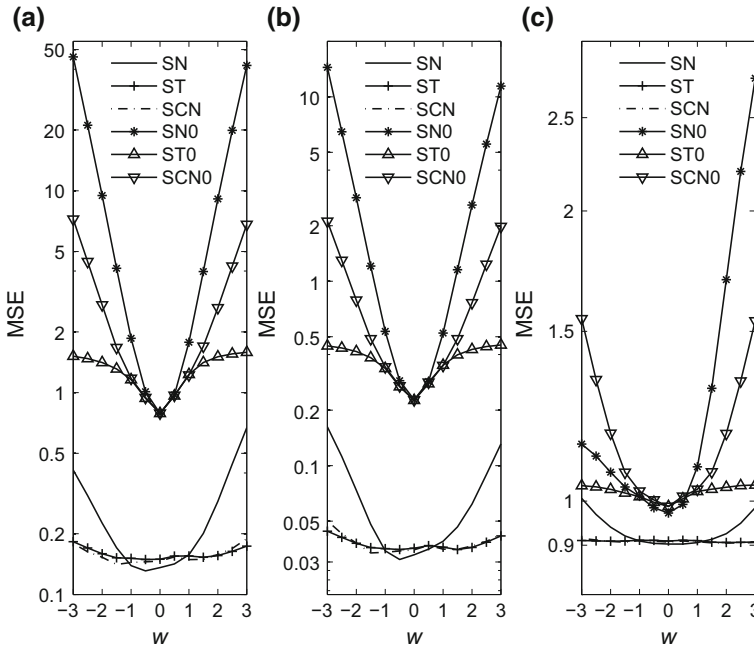


Fig. 1 Performances of estimates in the presence of outliers: MSE versus w , for **a** α ; **b** β ; **c** predictions

To sum up, the MLEs based on skew-heavy tailed models are more appealing since they behave better than the traditional SN ones in terms of robustness and accuracy.

5 Application

We give an illustrative example using the CSFII dataset. This dataset has conducted the 24-h recall measures, as well as three additional 24-h recall phone interviews of 1827 women who were recorded about their daily diet intake (for example, saturated fat, calories, vitamin, and so on). Carroll et al. (2006) have indicated that saturated fat has great relationship with the risk of breast cancer and other diseases, but the statistical significance for saturated fat disappeared when adding caloric intake into the logistic regression model. Thus it is really necessary to reveal the inner relationship between saturated fat and caloric intake. In this illustrative example, we take the calorie intake/5000 as x and the saturated fat intake/100 as y (Carroll et al. 2006; Lin and Cao 2013). Instead of x and y , the nutrition variables X and Y are calculated by four 24-h recalls, and suppose that they follow model (2) with $p = q = 4$. Acceptable saturated fat daily intake has been considered as an important reference index for health in the survey (le Coutre et al. 2013). There are some differences for daily diet habits for people in different ages, which may be aroused by their different lifestyles or nutrition awareness. In order to compare the difference of adults at different ages, we divide the dataset into two groups. Group 1 is younger than 40 years old, and group 2 is over 40. With the purpose of verifying the existence of skewness and heavy-tails

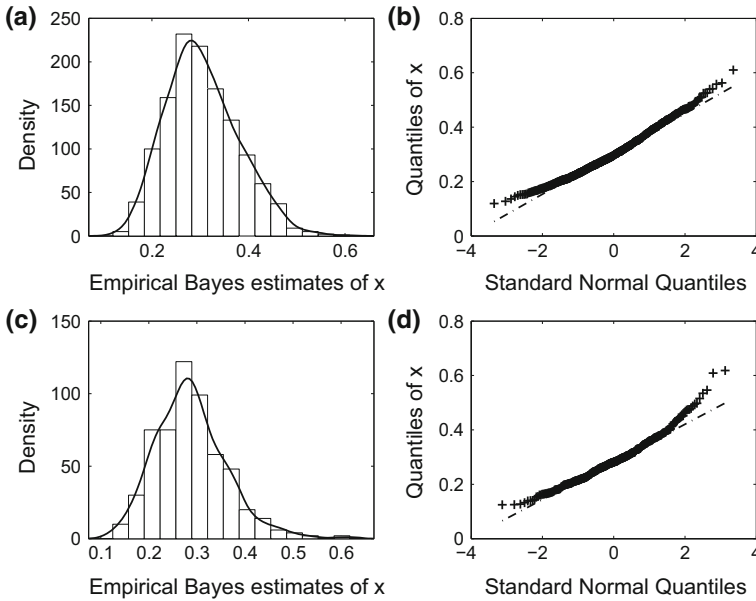


Fig. 2 Histogram and normal Q–Q plot of empirical Bayes estimates of x_t under N_0 model, **a, b** for group 1, and **c, d** for group 2

in the latent covariate x , we fit the data using normal RMEM without equation error. Figure 2 presents the corresponding histogram and Q–Q plot of the empirical Bayes estimates of x_t , both showing that the latent covariate is positively skewed and heavy-tailed in each group. This hints that a normal model may not offer a good fit. We now consider SMSN distributions for x_t and SMN distributions for measurement errors δ_t, ε_t and equation error e_t in the RMEM, i.e., the SMSN-RMEM proposed in this paper. For the sake of comparison, we also consider the SMN distributions for x_t , i.e. the distributions without assuming skewness.

We calculate the MLEs of parameter θ and their standard errors (SEs) and also Akaike information criterion (AIC) (Akaike 1974) values based on RMEM model under above distributions. This estimates are displayed in Table 2 for group 1 and Table 3 for group 2, where ST, ST0, SS, SS0, SCN, SCN0, T, T0, S, S0 and CN, CN0 denote the skew- t , skew-slash, skew-contaminated normal, Student t , slash and contaminated normal distributions with or without equation error correspondingly. The degrees of freedom for different distributions, chosen by the Schwarz (1978) information criterion, are also reported in the tables.

We draw some conclusions from Tables 2 and 3. The AIC values are used for model selection. The smaller the AIC value is, the better the model is. We find that the AIC value under Student- t distribution is always the smallest for each group, whether there are equation error or skewness or not, and it gets the smallest value under ST0 distribution in all situations. Thus, ST0-RMEM is more suitable for the two group data. Moreover, the heavy-tailed models show smaller SEs than the normal model. Meanwhile, we'd better consider skewness in the models for their smaller AIC values.

Table 2 Parameter estimations of group 1 (age ≤ 40) for CSFII data

Model	Degree	AIC	Parameter							
			μ_x	α	β	λ_x	ϕ_x	ϕ_δ	ϕ_ε	ϕ_e
SN	/	-12,993	0.201 (0.006)	-0.050 (0.019)	0.920 (0.056)	3.888 (0.986)	0.018 (0.002)	0.013 (0.0002)	0.015 (0.0003)	0.0010 (0.0005)
ST	$\nu = 4$	-14,549	0.204 (0.006)	-0.047 (0.012)	0.901 (0.041)	2.242 (0.416)	0.012 (0.001)	0.008 (0.0003)	0.009 (0.0003)	0.0004 (0.0003)
SS	$\nu = 1.2$	-14,364	0.205 (0.007)	-0.046 (0.013)	0.896 (0.043)	2.289 (0.553)	0.007 (0.0008)	0.005 (0.0001)	0.005 (0.0002)	0.0003 (0.0002)
SCN	$\nu = 0.4$ $\gamma = 0.2$	-14,341	0.204 (0.006)	-0.045 (0.013)	0.893 (0.043)	2.247 (0.401)	0.008 (0.0009)	0.006 (0.0001)	0.006 (0.0002)	0.0003 (0.0002)
SN0	/	-13,143	0.201 (0.006)	-0.059 (0.014)	0.947 (0.043)	3.723 (0.831)	0.018 (0.002)	0.013 (0.0002)	0.013 (0.0002)	/
ST0	$\nu = 4$	-14,657	0.204 (0.006)	-0.053 (0.010)	0.920 (0.033)	2.214 (0.394)	0.011 (0.001)	0.007 (0.0003)	0.007 (0.0003)	/
SS0	$\nu = 1.2$	-14,473	0.205 (0.007)	-0.051 (0.010)	0.915 (0.034)	2.230 (0.516)	0.006 (0.0008)	0.004 (0.0001)	0.004 (0.0001)	/
SCN0	$\nu = 0.4$ $\gamma = 0.2$	-14,451	0.204 (0.006)	-0.051 (0.010)	0.914 (0.035)	2.163 (0.372)	0.008 (0.0008)	0.005 (0.0001)	0.005 (0.0001)	/
N	/	-12,946	0.305 (0.004)	-0.049 (0.021)	0.915 (0.062)	/	0.007 (0.0005)	0.013 (0.0002)	0.015 (0.0003)	0.0010 (0.0005)
T	$\nu = 5$	-14,376	0.284 (0.003)	-0.046 (0.013)	0.897 (0.045)	/	0.006 (0.0004)	0.008 (0.0003)	0.009 (0.0003)	0.0005 (0.0003)
S	$\nu = 1.3$	-14,289	0.287 (0.003)	-0.045 (0.014)	0.893 (0.045)	/	0.004 (0.0002)	0.005 (0.0001)	0.005 (0.0002)	0.0003 (0.0002)
CN	$\nu = 0.3$ $\gamma = 0.2$	-14,188	0.286 (0.003)	-0.043 (0.014)	0.889 (0.044)	/	0.005 (0.0003)	0.007 (0.0002)	0.007 (0.0002)	0.0004 (0.0002)
N0	/	-13,088	0.305 (0.004)	-0.057 (0.015)	0.941 (0.046)	/	0.007 (0.0005)	0.013 (0.0002)	0.013 (0.0002)	/
T0	$\nu = 5$	-14,478	0.283 (0.003)	-0.052 (0.010)	0.916 (0.035)	/	0.006 (0.0004)	0.008 (0.0003)	0.008 (0.0003)	/
S0	$\nu = 1.2$	-14,392	0.284 (0.003)	-0.050 (0.010)	0.912 (0.034)	/	0.003 (0.0002)	0.004 (0.0001)	0.004 (0.0001)	/
CN0	$\nu = 0.3$ $\gamma = 0.3$	-14,294	0.288 (0.003)	-0.051 (0.010)	0.917 (0.037)	/	0.005 (0.0003)	0.007 (0.0002)	0.007 (0.0002)	/

We now do a test to compare the robustness among each model. We aim at α , β and λ_x as our test targets. Figure 3 shows the index plots of Mahalanobis distance $d_t, t = 1, \dots, n$ (mentioned in Sect. 3.2) under the normal RMEM. In this case, we have $d_t \sim \chi^2(8)$. Here, we adopt the cutoff lines correspond to the quantile $\chi^2_{0.01}(8)$ for the distribution of d_t . As we can see, there are 69 potential outliers in group 1 and 36 points in group 2. Removing these observations, we calculate the change ratios of α ,

Table 3 Parameter estimations of group 2 (age > 40) for CSFII data

Model	Degree	AIC	Parameter							
			μ_x	α	β	λ_x	ϕ_x	ϕ_δ	ϕ_ε	ϕ_e
SN	/	-6397	0.191 (0.009)	-0.070 (0.026)	0.994 (0.078)	3.477 (1.078)	0.016 (0.002)	0.010 (0.0003)	0.014 (0.0004)	0.0011 (0.0007)
ST	$\nu = 4$	-7365	0.195 (0.008)	-0.056 (0.017)	0.927 (0.063)	2.013 (0.535)	0.011 (0.002)	0.006 (0.0003)	0.007 (0.0004)	0.0004 (0.0004)
SS	$\nu = 1.2$	-7311	0.198 (0.011)	-0.056 (0.020)	0.929 (0.068)	1.931 (0.712)	0.006 (0.001)	0.004 (0.0002)	0.004 (0.0002)	0.0002 (0.0002)
SCN	$\nu = 0.3$ $\gamma = 0.2$	-7250	0.200 (0.010)	-0.057 (0.020)	0.940 (0.071)	1.887 (0.540)	0.008 (0.002)	0.005 (0.0002)	0.006 (0.0002)	0.0004 (0.0003)
SN0	/	-6478	0.190 (0.008)	-0.078 (0.020)	1.019 (0.060)	3.604 (1.056)	0.016 (0.002)	0.010 (0.0003)	0.012 (0.0003)	/
ST0	$\nu = 4$	-7425	0.193 (0.008)	-0.060 (0.013)	0.942 (0.049)	2.081 (0.530)	0.010 (0.002)	0.006 (0.0003)	0.006 (0.0003)	/
SS0	$\nu = 1.2$	-7370	0.196 (0.010)	-0.060 (0.014)	0.945 (0.051)	1.987 (0.694)	0.005 (0.001)	0.003 (0.0002)	0.003 (0.0002)	/
SCN0	$\nu = 0.3$ $\gamma = 0.2$	-7310	0.197 (0.009)	-0.063 (0.016)	0.962 (0.055)	1.985 (0.534)	0.008 (0.001)	0.005 (0.0002)	0.006 (0.0002)	/
N	/	-6368	0.287 (0.005)	-0.067 (0.027)	0.983 (0.083)	/	0.007 (0.0006)	0.01 (0.0003)	0.014 (0.0004)	0.0010 (0.0007)
T	$\nu = 4$	-7294	0.266 (0.004)	-0.055 (0.018)	0.922 (0.064)	/	0.006 (0.0006)	0.006 (0.0003)	0.008 (0.0004)	0.0004 (0.0004)
S	$\nu = 1.2$	-7284	0.268 (0.004)	-0.054 (0.020)	0.924 (0.067)	/	0.003 (0.0003)	0.004 (0.0002)	0.004 (0.0002)	0.0002 (0.0002)
CN	$\nu = 0.2$ $\gamma = 0.2$	-7204	0.272 (0.004)	-0.055 (0.020)	0.931 (0.070)	/	0.005 (0.0004)	0.006 (0.0002)	0.007 (0.0003)	0.0004 (0.0004)
N0	/	-6439	0.287 (0.005)	-0.074 (0.020)	1.006 (0.060)	/	0.007 (0.0006)	0.010 (0.0003)	0.013 (0.0003)	/
T0	$\nu = 4$	-7345	0.265 (0.004)	-0.059 (0.013)	0.937 (0.048)	/	0.005 (0.0005)	0.006 (0.0003)	0.006 (0.0003)	/
S0	$\nu = 1.2$	-7336	0.267 (0.004)	-0.058 (0.014)	0.938 (0.050)	/	0.003 (0.0003)	0.003 (0.0002)	0.004 (0.0002)	/
CN0	$\nu = 0.2$ $\gamma = 0.2$	-7256	0.270 (0.004)	-0.061 (0.015)	0.951 (0.053)	/	0.004 (0.0004)	0.006 (0.0002)	0.006 (0.0002)	/

β and λ_x under above models and display their values in Tables 4 (for group 1) and 5 (for group 2). The estimated values get slightly smaller after deleting the influential observations in all situations. Specially, both the cases with and without equation error, the change ratios of estimators under heavy-tailed distributions are smaller than those under normal distribution, which also implies that heavy-tailed models fit the data better than normal ones.

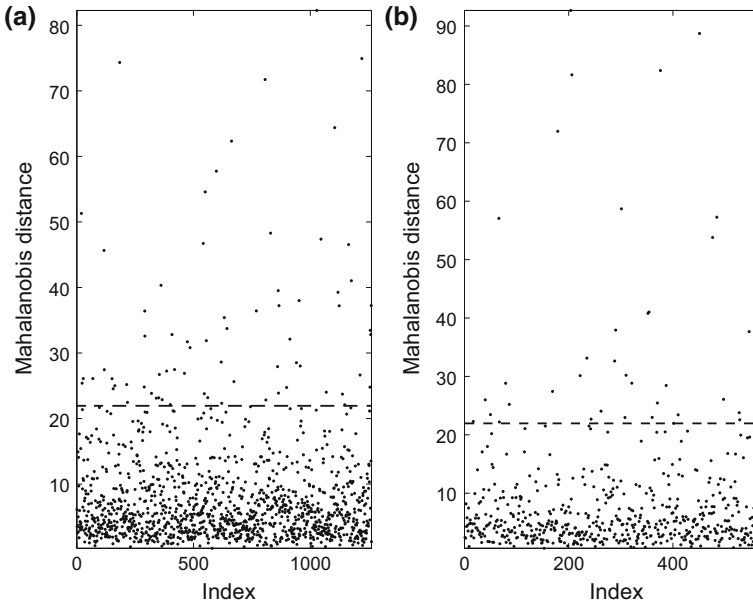


Fig. 3 Performances of the influential points, **a** for group 1 and **b** for group 2

Table 4 Change ratios of α , β , λ_x after deleting outliers in Group 1 (in%)

Model	With equation error			Without equation error		
	α	β	λ_x	α	β	λ_x
SN	-10.52	-2.35	-30.07	-10.43	-2.48	-27.60
ST	-4.00	-0.89	-5.57	-3.96	-0.95	-4.81
SS	-4.17	-0.89	-5.57	-4.09	-0.92	-4.70
SCN	-5.35	-1.29	-15.72	-5.30	-1.32	-14.81
N	-9.61	-2.11	/	-9.52	-2.26	/
T	-3.90	-0.89	/	-4.06	-0.95	/
S	-3.79	-0.83	/	-3.78	-0.87	/
CN	-3.91	-0.91	/	-5.84	-1.40	/

Now, we explain the data by parameters of interest β , λ_x and ϕ_e . From Table 2, The values of ϕ_e seem very small and only slightly changes among different models, indicating that there is inapparent random relationship between calorie and saturated fat intake. The almost same conclusions can be drawn from Table 3 for group 2. The practical significance of parameter β can be described as the positive proportion of saturated fat in calorie. It takes a smaller value in the younger group than the older group. Thus it can be seen that age factor has a certain effect on their diet habits. The younger people may react quickly than the older ones when acquiring new nutrition knowledge. Besides, it should be noted that the estimates of β under heavy-tailed

Table 5 Change ratios of α , β , λ_x after deleting outliers in Group 2 (in%)

Model	With equation error			Without equation error		
	α	β	λ_x	α	β	λ_x
SN	-19.89	-5.96	-37.57	-19.10	-6.04	-36.70
ST	-7.37	-1.99	-7.98	-7.01	-1.96	-7.43
SS	-6.67	-1.80	-8.64	-6.49	-1.79	-7.75
SCN	-13.57	-4.05	-14.67	-17.64	-5.47	-14.27
N	-18.06	-5.37	/	-17.50	-5.47	/
T	-5.85	-1.52	/	-5.60	-1.48	/
S	-5.55	-1.45	/	-5.64	-1.56	/
CN	-8.18	-2.30	/	-7.74	-2.38	/

models are smaller than the normal ones. Skew parameter λ_x is positive, suggesting that the two group data skewed to the right. Moreover, the value of μ_x in group 1 is consistently larger than group 2, that may due to younger's demands on more calories.

6 Conclusions

In this paper, we proposed a RMEM under the SMSN distribution class, called SMSN-RMEM, which is suitable for asymmetric and heavy-tailed data. It includes many special cases, such as non-replicated MEM under SMN distributions (Lachos et al. 2011), RMEM under normal distribution (Lin et al. 2004) or SMN distributions (Lin and Cao 2013). Considering the SMSN-RMEM with equation error, we provided the explicit iterative expressions of MLEs via an EM type algorithm. Additionally, empirical Bayes estimations have been conducted for predicting the true covariate and response. Simulation studies and the application on CSFII data demonstrated the effectiveness and robustness of inferences under the SMSN-RMEM model. The proposed model has good features in terms of robustness against outlying observations, adaptation to general data type and convenience for extensive applications. We expect that the model provides satisfactory results for data in presence of measurement error, skewness and outliers, etc., which are often observed in many areas, particularly in economics, medicine and environment.

We assume a SMSN distribution for the true error-prone covariate while the random errors are all assumed following SMN distributions, which is similar as those in Lachos et al. (2010b) and Zeller et al. (2014), etc. Owing to the hierarchical structure of the model, the features of skewness of both covariate and response can be captured under this assumptions. We may also assume that some random errors follow skew distributions, if we have an evidence from the data collection process. However, we need consider carefully the identification of the model. This may be an interest research later. We conducted model selection mainly through AIC criterion. Other criteria, such as BIC and EDC can also be chosen. For the CSFII data, the order of the candidate

models does not change and the advantage of heavy-tailed models is always maintained under different criteria. The computing code is available from the authors upon request. In the future work, it is also necessary to perform some statistic diagnostics and hypothesis tests for the significance of parameters or equation error. Furthermore, the model can be extended to multivariate cases for some statistical studies, or analyzed under a Bayesian paradigm.

Acknowledgements This research was supported by the National Science Foundation of China (Grant No. 11301278), the Natural Science Foundation of Jiangsu Province of China (Grant No. BK2012459), the MOE (Ministry of Education in China) Project of Humanities and Social Sciences (Grant No. 13YJC910001), and Academic Degree Postgraduate innovation projects of Jiangsu province Ordinary University (Grant No. KYLX15-0883).

Appendix

A The PDF of Some SMSN Distributions and the Conditional Moments

The pdf of some important SMSN distributions and the properties about conditional moments are as follows:

(1) The multivariate skew- t distribution $ST_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; \nu)$:

$\kappa(u) = 1/u, U \sim \text{Gamma}(\nu/2, \nu/2)$ with $\nu > 0$. The pdf of \mathbf{Y} is given by

$$f(\mathbf{y}) = 2t_m(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)T\left(\sqrt{\frac{m+\nu}{d+\nu}}A; \nu+m\right),$$

where $d = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})$, $t_m(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$ and $T(\cdot; \nu)$ denote the pdf of m -dimensional Student- t distribution and the cdf of standard univariate t distribution, respectively. The skew-normal distribution is the limiting case when $\nu \rightarrow +\infty$.

The conditional moments take the forms

$$u_r = \frac{f_0(\mathbf{y})}{f(\mathbf{y})} \frac{2^{r+1}\Gamma((\nu+m+2r)/2)(\nu+d)^{-r}}{\Gamma((\nu+m)/2)} T\left(\sqrt{\frac{m+\nu+2r}{d+\nu}}A; \nu+m+2r\right),$$

$$\eta_r = \frac{f_0(\mathbf{y})}{f(\mathbf{y})} \frac{2^{(r+1)/2}\Gamma((\nu+m+r)/2)}{\pi^{1/2}\Gamma((\nu+m)/2)} \frac{(\nu+d)^{(\nu+m)/2}}{(\nu+d+A^2)^{(\nu+m+r)/2}}.$$

where $f_0(\mathbf{y}) = \int_0^\infty \phi_m(\mathbf{y}|\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma})dH(u)$, i.e. the pdf of the class of SMN distribution when $\boldsymbol{\lambda} = \mathbf{0}$.

(2) The multivariate skew-slash distribution $SS_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; \nu)$:

$\kappa(u) = 1/u, U \sim \text{Beta}(\nu, 1)$ with $0 < u < 1$ and $\nu > 0$. The pdf of \mathbf{Y} is given by

$$f(\mathbf{y}) = 2\nu \int_0^1 u^{\nu-1} \phi_m(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})\Phi(u^{1/2}A)du.$$

When $\nu \rightarrow +\infty$, the skew-slash distribution reduces to the skew-normal one.

The conditional moments take the forms

$$\begin{aligned}
 u_r &= \frac{f_0(\mathbf{y})}{f(\mathbf{y})} \frac{2\Gamma((2\nu + m + 2r)/2)}{\Gamma((2\nu + m)/2)} \left(\frac{2}{d}\right)^r \frac{P_1((2\nu + m + 2r)/2, d/2)}{P_1((2\nu + m)/2, d/2)} \\
 &\quad \times E[\Phi(S^{1/2}A)], \\
 \eta_r &= \frac{f_0(\mathbf{y})}{f(\mathbf{y})} \frac{2^{(r+1)/2}\Gamma((2\nu + m + r)/2)}{\pi^{1/2}\Gamma((2\nu + m)/2)} \frac{d^{(2\nu+m)/2}}{(d + A^2)^{(2\nu+m+r)/2}} \\
 &\quad \times \frac{P_1((2\nu + m + r)/2, (d + A^2)/2)}{P_1((2\nu + m)/2, d/2)},
 \end{aligned}$$

where $S \sim \text{Gamma}((2\nu + m + 2r)/2, d/2)I_{(0,1)}$ and $P_x(a, b)$ denotes the cdf of the $\text{Gamma}(a, b)$ distribution evaluated at x .

(3) The multivariate skew-contaminated normal distribution $\text{SCN}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; \nu, \gamma)$:

When $\kappa(u) = 1/u$ and U follows a discrete random probability function $h(u; \nu, \gamma) = \nu I_{(u=\gamma)} + (1 - \nu)I_{(u=1)}$ with given parameter vector $\boldsymbol{\nu} = (\nu, \gamma)^\top$ and $0 < \nu < 1, 0 < \gamma \leq 1$, we get the multivariate skew-contaminated normal distribution with the pdf as

$$f(\mathbf{y}) = 2 \left\{ \nu \phi_m(\mathbf{y}|\boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Sigma}) \Phi(\gamma^{1/2}A) + (1 - \nu)\phi_m(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Phi(A) \right\}.$$

The SN distribution is a special case as $\gamma = 1$.

The conditional moments take the forms

$$\begin{aligned}
 u_r &= \frac{2}{f(\mathbf{y})} \left\{ \nu \gamma^r \phi_m(\mathbf{y}|\boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Sigma}) \Phi(\gamma^{1/2}A) + (1 - \nu)\phi_m(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Phi(A) \right\}, \\
 \eta_r &= \frac{2}{f(\mathbf{y})} \left\{ \nu \gamma^{r/2} \phi_m(\mathbf{y}|\boldsymbol{\mu}, \gamma^{-1}\boldsymbol{\Sigma}) \phi(\gamma^{1/2}A) + (1 - \nu)\phi_m(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \phi(A) \right\}.
 \end{aligned}$$

B The First Derivatives of d_t, A_t and $\log |\boldsymbol{\Sigma}|$ with Respect to $\boldsymbol{\theta}$

By direct calculations, we have the first derivatives of d_t, A_t and $\log |\boldsymbol{\Sigma}|$ as follows: for d_t :

$$\frac{\partial d_t}{\partial \theta_i} = -2(\mathbf{Z}_t - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_i} - (\mathbf{Z}_t - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} (\mathbf{Z}_t - \boldsymbol{\mu}),$$

for A_t :

$$\frac{\partial A_t}{\partial \theta_i} = \left(\frac{\partial \psi}{\partial \theta_i} \mathbf{b}^\top + \psi \frac{\partial \mathbf{b}^\top}{\partial \theta_i} - \psi \mathbf{b}^\top \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \right) \boldsymbol{\Sigma}^{-1} (\mathbf{Z}_t - \boldsymbol{\mu}) - \psi \mathbf{b}^\top \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_i},$$

where $\psi = \frac{\lambda_x \phi_x}{\sqrt{\phi_x + \lambda_x^2 \Lambda_x}}$.

for $\log |\Sigma|$:

$$\begin{aligned} \frac{\partial \log |\Sigma|}{\partial \beta} &= 2q\beta\phi_\delta\phi_x/\tau, \\ \frac{\partial \log |\Sigma|}{\partial \phi_x} &= [p(\phi_\varepsilon + q\phi_e) + q\beta^2\phi_\delta]/\tau, \\ \frac{\partial \log |\Sigma|}{\partial \phi_\delta} &= (p - 1)/\phi_\delta + (\phi_\varepsilon + q\phi_e + q\beta^2\phi_x)/\tau, \\ \frac{\partial \log |\Sigma|}{\partial \phi_e} &= (q - 1)/\phi_e + (\phi_\delta + p\phi_x)/\tau, \\ \frac{\partial \log |\Sigma|}{\partial \phi_e} &= q(\phi_\delta + p\phi_x)/\tau, \end{aligned}$$

where $|\Sigma| = \phi_\delta^{p-1}\phi_e^{q-1}\tau$, $\tau = (\phi_\delta + p\phi_x)(\phi_\varepsilon + q\phi_e) + q\beta^2\phi_\delta\phi_x$.

In addition, we also need to calculate the following derivation:

for μ :

$$\frac{\partial \mu}{\partial \mu_x} = \mathbf{b}, \quad \frac{\partial \mu}{\partial \alpha} = \mathbf{c}, \quad \frac{\partial \mu}{\partial \beta} = \mu_x \mathbf{c}.$$

for Σ :

$$\begin{aligned} \frac{\partial \Sigma}{\partial \beta} &= \phi_x (\mathbf{c}\mathbf{b}^\top + \mathbf{b}\mathbf{c}^\top), \quad \frac{\partial \Sigma}{\partial \phi_x} = \mathbf{b}\mathbf{b}^\top, \quad \frac{\partial \Sigma}{\partial \phi_\delta} = \mathbf{D} \left\{ \left(\mathbf{1}_p^\top, \mathbf{0}_q^\top \right) \right\}, \quad \frac{\partial \Sigma}{\partial \phi_e} \\ &= \mathbf{D}(\mathbf{c}), \quad \frac{\partial \Sigma}{\partial \phi_e} = \mathbf{c}\mathbf{c}^\top. \end{aligned}$$

for \mathbf{b} :

$$\frac{\partial \mathbf{b}}{\partial \beta} = \mathbf{c}.$$

for ψ :

$$\begin{aligned} \frac{\partial \psi}{\partial \beta} &= q\beta c^{-2}\psi^3/(\phi_\varepsilon + q\phi_e), \quad \frac{\partial \psi}{\partial \lambda_x} = \phi_x^2(\phi_x + \lambda_x^2\Lambda_x)^{-3/2}, \\ \frac{\partial \psi}{\partial \phi_x} &= \frac{1}{2}\psi\phi_x^{-1} + \frac{1}{2}\psi^3\phi_x^{-1}c^{-2}\mathbf{b}^\top \Sigma_1^{-1}\mathbf{b}, \quad \frac{\partial \psi}{\partial \phi_\delta} = -\frac{1}{2}p\psi^3c^{-2}\phi_\delta^{-2}, \\ \frac{\partial \psi}{\partial \phi_e} &= -\frac{1}{2}\psi^3c^{-2}q\beta^2(\phi_\varepsilon + q\phi_e)^{-2}, \quad \frac{\partial \psi}{\partial \phi_e} = -\frac{1}{2}q^2\psi^3c^{-2}\beta^2(\phi_\varepsilon + q\phi_e)^{-2}. \end{aligned}$$

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.

- Andrews, D. F., & Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society: Series B*, 36(1), 99–102.
- Arellano-Valle, R. B., Bolfarine, H., & Lachos, V. H. (2005). Skew-normal linear mixed models. *Journal of Data Science*, 3, 415–438.
- Azzalini, A., & Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society: Series B*, 61(3), 579–602.
- Bartlett, J. W., De Stavola, B. L., & Frost, C. (2009). Linear mixed models for replication data to efficiently allow for covariate measurement error. *Statistics in Medicine*, 28(25), 3158–3178.
- Basso, R. M., Lachos, V. H., Cabral, C. R., & Ghosh, P. (2010). Robust mixture modeling based on scale mixtures of skew-normal distributions. *Computational Statistics and Data Analysis*, 54(12), 2926–2941.
- Branco, M. D., & Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, 79(1), 99–113.
- Cancho, V. G., Lachos, V. H., & Ortega, E. M. M. (2008). A nonlinear regression model with skew-normal errors. *Statistical Papers*, 52, 571–583.
- Cao, C. Z., Lin, J. G., & Shi, J. Q. (2014). Diagnostics on nonlinear model with scale mixtures of skew-normal and first-order autoregressive errors. *Statistics*, 48(5), 1033–1047.
- Cao, C. Z., Lin, J. G., Shi, J. Q., Wang, W., & Zhang, X. Y. (2015). Multivariate measurement error models for replicated data under heavy-tailed distributions. *Journal of Chemometrics*, 29(8), 457–466.
- Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: A modern perspective* (2nd ed.). Boca Raton: Chapman and Hall.
- Chan, L. K., & Mak, T. K. (1979). Maximum likelihood estimation of a linear structural relationship with replication. *Journal of the Royal Statistical Society: Series B*, 41(2), 263–268.
- Cheng, C. L., & Van Ness, J. W. (1999). *Statistical regression with measurement error*. London: Arnold.
- Cheng, C. L., & Riu, J. (2006). On estimating linear relationships when both variables are subject to heteroscedastic measurement errors. *Technometrics*, 48, 511–519.
- Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39, 1–38.
- Fang, K. T., Kotz, S., & Ng, K. W. (1990). *Symmetrical multivariate and related distributions*. London: Chapman and Hall.
- Fuller, W. A. (1987). *Measurement error models*. New York: Wiley.
- Genton, M. G. (2004). *Skew-elliptical distributions and their applications: A Journey beyond normality*. Boca Raton: Chapman & Hall.
- Giménez, P., & Patat, M. L. (2005). Estimation in comparative calibration models with replicated measurement. *Statistics and Probability Letters*, 71(2), 155–164.
- Gori, L., & Sodini, M. (2011). Nonlinear dynamics in an OLG growth model with young and old age labour supply: The role of public health expenditure. *Computational Economics*, 38, 261–275.
- Harnack, L., Stang, J., & Story, M. (1999). Soft drink consumption among US children and adolescents: Nutritional consequences. *Journal of the American Dietetic Association*, 99(4), 436–441.
- Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*. New York: Springer.
- Isogawa, Y. (1985). Estimating a multivariate linear structural relationship with replication. *Journal of the Royal Statistical Society: Series B*, 47, 211–215.
- Jacobs, H. L., Kahn, H. D., Stralka, K. A., & Phan, D. B. (1998). Estimates of per capita fish consumption in the US based on the continuing survey of food intake by individuals (CSFII). *Risk Analysis*, 18(3), 283–291.
- Jara, A., Quintana, F., & Martin, E. S. (2008). Linear mixed models with skew-elliptical distributions: A Bayesian approach. *Computational Statistics and Data Analysis*, 52(11), 5033–5045.
- Jones, D. Y., Schatzkin, A., Green, S. B., Block, G., Brinton, L. A., Ziegler, R. G., et al. (1987). Dietary fat and breast cancer in the National Health and Nutrition Examination Survey I: Epidemiologic follow-up study. *Journal of the National Cancer Institute*, 79, 465–471.
- Lachos, V. H., Angolini, T., & Abanto-Valle, C. A. (2011). On estimation and local influence analysis for measurement errors models under heavy-tailed distributions. *Statistical Papers*, 52, 567–590.
- Lachos, V. H., Ghosh, P., & Arellano-Valle, R. B. (2010a). Likelihood based inference for skew-normal/independent linear mixed models. *Statistica Sinica*, 20, 303–322.
- Lachos, V. H., Labra, F. V., Bolfarine, H., & Ghosh, P. (2010b). Multivariate measurement error models based on scale mixtures of the skew-normal distribution. *Statistics*, 44(6), 541–556.

- Lange, K. L., & Sinsheimer, J. S. (1993). Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics*, 2, 175–198.
- le Coutre, J., Mattson, M. P., Dillin, A., Friedman, J., & Bistrian, B. (2013). Nutrition and the biology of human aging: Cognitive decline/food intake and caloric restriction. *The Journal of Nutrition, Health and Aging*, 17(8), 717–720.
- Lin, N., Bailey, B. A., He, X. M., & Buttlar, W. G. (2004). Adjustment of measuring devices with linear models. *Technometrics*, 46, 127–134.
- Lin, J. G., & Cao, C. Z. (2013). On estimation of measurement error models with replication under heavy-tailed distributions. *Computational Statistics*, 28(2), 809–829.
- McLachlan, G. L., & Krishnan, T. (1997). *The EM algorithm and extensions*. New York: Wiley.
- Montenegro, L. C., Bolfarine, H., & Lachos, V. H. (2010). Inference for a skew extension of the Grubb's model. *Statistical Papers*, 51, 701–715.
- Osorio, F., Paula, G. A., & Galea, M. (2009). On estimation and influence diagnostics for the Grubb's model under heavy-tailed distributions. *Computational Statistics and Data Analysis*, 53, 1249–1263.
- Reiersol, O. (1950). Identifiability of a linear relation between variables which are subject to errors. *Econometrica*, 18, 375–389.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.
- Sun, S. Z., & Empie, M. W. (2007). Lack of findings for the association between obesity risk and usual sugar-sweetened beverage consumption in adults: A primary analysis of databases of CSFII-1989–1991, CSFII-1994–1998, NHANES III, and combined NHANES 1999–2002. *Food and Chemical Toxicology*, 45(8), 1523–1536.
- Wimmer, G., & Witkovský, V. (2007). Univariate linear calibration via replicated errors-in-variables model. *Journal of Statistical Computation and Simulation*, 77, 213–227.
- Xie, F. C., Wei, B. C., & Lin, J. G. (2008). Homogeneity diagnostics for skew-normal nonlinear regression models. *Statistics and Probability Letters*, 20, 303–322.
- Zeller, C. B., Carvalho, R. R., & Lachos, V. H. (2012). On diagnostics in multivariate measurement error models under asymmetric heavy-tailed distributions. *Statistical Papers*, 53(3), 665–683.
- Zeller, C. B., Lachos, V. H., & 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.
- Zeller, C. B., Lachos, V. H., & Vilca-Labra, F. E. (2014). Influence diagnostics for Grubb's model with asymmetric heavy-tailed distributions. *Statistical Papers*, 55(3), 671–690.