ORIGINAL RESEARCH

Measurement error in linear regression models with fat tails and skewed errors

Mahmoud Torabi[a], Malay Ghosh[b], Jiyoun Myung[c] and Mark Steel[d]

[a]University of Manitoba, Canada; [b]University of Florida, USA; [c] California State University – East Bay, USA; [d]University of Warwick, UK

ARTICLE HISTORY
Compiled October 29, 2021

ABSTRACT
Linear regression models which account for skewed error distributions with fat tails have been previously studied. These two important features, skewness and fat tails, are often observed in real data analyses. Covariates measured with an error also happen frequently in the observational data set-up. As a motivating example, wind speed as a covariate is usually used, among other covariates, to estimate the particulate matter (PM) which is one of the most critical air pollutants and has a major impact on human health and on the environment. However, the wind speed is measured with error and the distribution of PM is neither symmetric nor normally distributed (see Section 4 for more details). Ignoring the issue of measurement error in covariates may produce bias in model parameters estimate and lead to wrong conclusions. In this paper, we propose an approach to study properly linear regression models where the covariates are measured with error and the error distribution ~~of model~~ is skewed with fat tails. We use a hierarchical Bayesian approach for inference, addressing also sensitivity of the results to priors. Performance of the proposed approach is evaluated through a simulation study and also by a real data application (PM in Canada).

KEYWORDS
Bayesian inference; skewness; student t distribution

1. Introduction

Linear regression models have ~~has~~ been subject to extensive research in the literature. In many ~~of its~~ real applications, the response variable does not follow nice properties of symmetry and/or normality. ~~Asymmetry in the distribution of the response may be caused by skewness.~~ The presence of skewness arises in many studies including problems in linear regression, estimation, and prediction. In such a setting, an initial strategy is to transfer the data. ~~make data transformation.~~ However, an appropriate transformation may not exist or may be difficult to find. Also, this approach can raise interpretation issues. To address the issues associated with the transformation method, considerable attention has been devoted in the literature to introducing ~~introduce~~ a more suitable theoretical strategy based on the skew-normal (SN) distribution (Azzalini 2013). There is a large number of such distributions ~~sometime~~ making it hard to decide which class of SN model needs to be used. With this in mind, Arellano-Valle

CONTACT Mahmoud Torabi. Email: Mahmoud.Torabi@umanitoba.ca

et al. (2006) introduced a SN model and named it unified skew-normal which also includes the normal density and has very similar properties as the normal density.

On the other hand, in order also to circumvent the problem of departure from normality in the context of linear regression model, there is limited research to address two pervasive features of empirical data (skewness and fat tails) in statistical modeling and inference. Under a classical perspective, there are some works to modeling skewness and heavy-tail in linear regression models (Zeller, Lachos, and Vilca-Labra 2011; Fernandez, Lachos, and Bolfarine 2015; Ferreira and Arellano-Valle 2018). In addition, Buckle (1995) provided a Bayesian analysis using stable laws. In addition, There is also some frequentist research in that direction: the use of distributions by Badrinath and Chatterjee (1991) through matching percentiles; partially adaptive estimation of generalized beta distributions of the second kind of McDonald and Nelson (1993); and approximate maximum likelihood estimation (MLE) of generalized exponential distributions of Lye and Martin (1993)., Fernandez and Steel (1998) and Juarez and Steel (2010) argued that all of these solutions seem quite complicated to implement numerically and seem to lack the flexibility and ease of interpretation that an applied statistician would typically require. To address these difficulties, they introduced a class of sampling models that can simultaneously account for both skewness and fat tails, and conducted Bayesian inference in the context of a regression model with unknown scale. In the following, we briefly explain the idea of Fernandez and Steel (1998) work.

Consider the following linear regression model:

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + e_i, \qquad i = 1, ..., n, \tag{1}$$

where y_i is the response variable, $\boldsymbol{x}_i$ and $\boldsymbol{\beta}$ are $p-$dimensional covariates with corresponding regression coefficients, and e_i is the residual term of the ith observation. Fernandez and Steel (1998) generated a "skewed Student" distribution to deal with skewed error distributions having fat tails in the context of linear regression model (1). Because of the particular skewness-generating mechanism used, this class of skewed distributions is also sometimes referred to as "two-piece Student" as in Fernandez and Steel (1998) and Rubio and Steel (2020). They first introduced skewness into symmetric distributions and then used the Student distribution to deal with the fat tails. Assume that $f(s) = f(|s|)$ which is decreasing in $|s|$, and $f(\cdot)$ is a univariate pdf which is unimodal and symmetric around 0. The skewed distribution of the residual term e_i in (1) is then generated as follows:

$$g(e_i \mid \gamma) = \frac{2}{\gamma + \frac{1}{\gamma}} \left\{ f\left(\frac{e_i}{\gamma}\right) I_{[0,\infty)}(e_i) + f(\gamma e_i)\, I_{(-\infty,0)}(e_i) \right\}, \tag{2}$$

where $\gamma \in (0, \infty)$ is the parameter which determines the allocation of mass to each side of the mode (Fernandez and Steel 1998). Note that (2) has the unique mode at 0 but loses symmetry whenever $\gamma \neq 1$. They then chose Student's t-distribution for $f(\cdot)$ in the skewed distribution (2) to manage the tail behavior. The skewed Student distribution of the response variable y_i in (1) is then defined as

$$f(y_i|\mu_i, \sigma, \nu, \gamma) = \frac{2}{\gamma + \frac{1}{\gamma}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\sigma} \left[1 + \frac{(y_i - \mu_i)^2}{\nu\sigma^2}\left(\frac{1}{\gamma^2} I_{[y_i \geq \mu_i]} + \gamma^2 I_{[y_i < \mu_i]}\right)\right]^{-\frac{\nu+1}{2}}, \tag{3}$$

with location $\mu_i = \boldsymbol{x}_i^T \boldsymbol{\beta}$, scale σ, skewness parameter γ, and degrees of freedom ν. The thickness of the tails is controlled by $\nu \in \mathbb{R}_+$. This distribution displays both flexible tails and possible skewness with two clearly interpretable ~~respective~~ parameters ν and γ. Fernandez and Steel (1998) proposed a Bayesian method for inference with these distributions. ~~analyzing data for skewed error distributions with fat tails.~~

It is important to note that in the model (3) it is assumed that the covariates $\boldsymbol{x}_i$ are perfectly measured for the validity of inferential methods, but for various reasons such as the measurement techniques or instruments used, uncertainty is inherent in observational data and so these data are susceptible to measurement error in the covariates of interest (Fuller 1987; Arellano-Valle et al. 2005; Rodrigues and Bolfarine 2007; Cabral, Lachos, and Zeller 2014). Hence, observational data are prone to be not perfect and results may be seriously biased if one ignores this issue. Covariate measurement error is a ~~common~~ typical aspect of cross-sectional and/or longitudinal studies (Fuller 1987; Carrol et al. 2006). Recently, Arellano-Valle et al. (2020) proposed a model to cover measurement error in a covariate with an SN distribution using Expectation-Maximization technique for the inference. As a motivating example, researchers have used linear regression models for the estimation/prediction of particulate matter (PM) concentrations using data of meteorological parameters including wind speed (WS) and temperature (Roy and Adhikari 2009). However, it appears that measuring WS is associated with an error and the distribution of PM is neither symmetric nor normally distributed (see Section 4).

The aim of this paper is to provide a proper modeling framework for fat tails and skewed errors wherein the covariates of interest cannot be observed precisely. In other words, the proposed model provides flexibility in capturing the effects of skewness and heavy tail behavior of the data and simultaneously facilitates analyzing by ~~representing and~~ taking full~~er~~ account of the susceptibility of measurement error in covariates. As measurement error (ME) is commonly due solely to instrument or laboratory-analysis error in our PM data, the classical measurement error model appears appropriate for this situation as we would expect the surrogate measure to be symmetrically ~~randomly~~ distributed around the true value. We will provide a hierarchical Bayesian approach to study the proposed model.

The rest of the paper is organized as follows. In Section 2, a linear model for non-Gaussian data with ME covariates is proposed using the Bayesian approach. Performance of the proposed approach is evaluated through a simulation study (Section 3) and also by a real data application (PM in Canada) (Section 4). Concluding remarks are given in Section 5. Technical details are deferred to the Appendix.

2. Measurement error model with skewed student errors

Recall the linear regression model (1) where y_i is the *observed* response variable but now $\boldsymbol{x}_i$ is the $p-$dimensional *unobserved* but fixed true covariate ~~variable~~ corresponding to the observed explanatory variable $\boldsymbol{w}_i$. It means that in reality we do not observe $\boldsymbol{x}_i$, but instead $\boldsymbol{w}_i$ are observed. The basic measurement error model is then given by

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + e_i, \quad \boldsymbol{w}_i = \boldsymbol{x}_i + \boldsymbol{u}_i, \quad i = 1, \ldots, n, \tag{4}$$

where the residual term e_i and the measurement error $\boldsymbol{u}_i$ are assumed to be independent. The normality assumptions are usually made for e_i and $\boldsymbol{u}_i$, namely $e_i \overset{iid}{\sim} N(0, \sigma^2)$

and $\boldsymbol{u}_i \overset{iid}{\sim} N_p(\mathbf{0}, \tau^2 \boldsymbol{I_p})$ with known τ^2. Note that we assume τ^2 is known to make sure our model is identifiable. This is also a reasonable assumption since we know something about the measurement error mechanism.

We assume that the measurement errors occur at random. This implies that the observed measurement $\boldsymbol{w}_i$ is modeled as a random variable whose mean is equal to the fixed true (unobserved) auxiliary variable $\boldsymbol{x}_i$ and its variance is fixed equal to τ^2.

We propose a measurement error model with a skewed Student distribution to account for fat tails and skewed errors, instead of assuming the normality of error terms. For $i = 1, \ldots, n$, we assume n independent samples from a skewed Student distribution (3) with $\mu_i = \boldsymbol{x}_i^T \boldsymbol{\beta}$, scale parameter σ, a scalar parameter $\gamma \in (0, \infty)$, and $\nu \in \mathbb{R}_+$.

$$f(y_i|\boldsymbol{x}_i, \boldsymbol{\beta}, \sigma^2, \gamma, \nu) = \frac{2}{(\gamma + \frac{1}{\gamma})} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\sigma} \left[1 + \frac{(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2}{\nu\sigma^2} \left(\frac{1}{\gamma^2} I_{[y_i \geq \boldsymbol{x}_i^T\boldsymbol{\beta}]} + \gamma^2 I_{[y_i < \boldsymbol{x}_i^T\boldsymbol{\beta}]}\right)\right]^{-\frac{\nu+1}{2}},$$

~~where $\nu \in \mathbb{R}_+$.~~ Also, we assume that

$$\boldsymbol{w}_i \mid \boldsymbol{x}_i \overset{ind}{\sim} N_p(\boldsymbol{x}_i, \tau^2 \boldsymbol{I_p}), \quad \tau^2 \text{ known ~~positive definite~~}, i = 1, \ldots, n.$$

To do a Bayesian analysis, we assign ~~the~~ prior distributions for all unknown parameters to find the posterior distribution of the parameters given the data. We take the following prior distribution:

$$\pi(\boldsymbol{\beta}, \sigma^2, \gamma, \nu) \propto \left[(\sigma^2)^{-\frac{a_1}{2}-1} \exp\left(-\frac{a_2}{2\sigma^2}\right)\right] P_\gamma P_\nu, \tag{5}$$

where P_γ and P_ν are proper distributions, and a_1, a_2 are positive real numbers. Note that we use non-informative priors for $\boldsymbol{\beta}$ and σ^2 independently as follows: $\pi(\boldsymbol{\beta}) \propto 1$ and $\pi(\sigma^2) \sim \text{IG}\left(\frac{a_1}{2}, \frac{a_2}{2}\right)$, where IG stands for the inverse gamma distribution.

Before we investigate the propriety of the posterior, it is convenient to introduce independent latent variables $\lambda_1, \ldots, \lambda_n$ in (3) as follows:

$$f(y_i|\boldsymbol{x}_i, \lambda_i, \boldsymbol{\beta}, \sigma^2, \gamma, \nu) = \frac{2}{(\gamma + \frac{1}{\gamma})} \frac{\sqrt{\lambda_i}}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\lambda_i (y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2}{2\sigma^2} \left(\frac{1}{\gamma^2} I_{[y_i \geq \boldsymbol{x}_i^T\boldsymbol{\beta}]} + \gamma^2 I_{[y_i < \boldsymbol{x}_i^T\boldsymbol{\beta}]}\right)\right]$$

and $f(\lambda_i \mid \nu) = \left(\frac{\nu}{2}\right)^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)^{-1} \lambda_i^{\frac{\nu}{2}-1} \exp\left(-\frac{\nu}{2}\lambda_i\right), i = 1, \ldots, n.$

This enables us to represent student's t-distribution as scale mixtures of normal distributions. Let $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ and $\boldsymbol{w} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n)^T$. By combining the likelihood of the model and the prior (5), the posterior distribution is

$$\begin{aligned}&\pi(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \lambda_1, \ldots, \lambda_n, \boldsymbol{\beta}, \sigma^2, \gamma, \nu \mid \boldsymbol{y}, \boldsymbol{w}) \\ &\qquad \propto \left(\gamma + \frac{1}{\gamma}\right)^{-n} (\sigma^2)^{-\frac{n}{2}} \prod_{i=1}^{n} \left[\lambda_i^{\frac{\nu+1}{2}-1} \exp\left(-\frac{\nu}{2}\lambda_i\right)\right] \left(\frac{\nu}{2}\right)^{\frac{n\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)^{-n}\end{aligned}$$

$$\times \prod_{i=1}^{n} \left\{ \exp\left[-\frac{\lambda_i (y_i - \boldsymbol{x}_i^T \boldsymbol{\beta})^2}{2\sigma^2} \left(\frac{1}{\gamma^2} I_{[y_i \geq \boldsymbol{x}_i^T \boldsymbol{\beta}]} + \gamma^2 I_{[y_i < \boldsymbol{x}_i^T \boldsymbol{\beta}]} \right) \right] \right\}$$
$$\times (\tau^2)^{-\frac{np}{2}} \exp\left[-\frac{1}{2\tau^2} \sum_{i=1}^{n} \|\boldsymbol{w}_i - \boldsymbol{x}_i\|^2 \right] (\sigma^2)^{-\frac{a_1}{2}-1} \exp\left(-\frac{a_2}{2\sigma^2} \right) P_\gamma P_\nu. \tag{6}$$

We first state our main result, which provides a sufficient condition for the propriety of the posterior distribution that results from the aforementioned hierarchical Bayesian model (6).

Theorem 1. The posterior density $\pi(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \lambda_1, \ldots, \lambda_n, \boldsymbol{\beta}, \sigma^2, \gamma, \nu \mid \boldsymbol{y}, \boldsymbol{w})$ is proper if $n \geq p$. and also all the model parameters $(\lambda_1, \ldots, \lambda_n, \boldsymbol{\beta}, \sigma^2, \gamma, \nu)$ have finite variances.

The proof of Theorem 1 is deferred to the Appendix A.

We obtain hierarchical Bayes predictors of model parameters using Gibbs sampling and Metropolis-Hastings. Full conditional distributions to implement the Gibbs samplers and Metropolis-Hastings are provided in the Appendix B.

In terms of priors used in the succeeding sections (simulation study and data application), we used non-informative prior for $\boldsymbol{\beta}$ as $\pi(\boldsymbol{\beta}) \propto 1$. For the inverse of variance component σ^2, we used the gamma distribution prior with shape and rate parameters 0.5. For the parameter ν, we used the exponential distribution prior with mean 10, and for the skewness parameter γ, we used the gamma distribution prior with shape and rate parameters 0.5 for γ^2. Note that the all model parameters in the simulation study and data application converged successfully.

3. Simulation study

In this section, we carry out a simulation study to evaluate the performance of the proposed model. We consider the following linear model:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + e_i, \quad i = 1, ..., n, \\ w_i &= x_i + u_i, \end{aligned} \tag{7}$$

where $n = 50, e_i$ has a skewed t-distribution with parameters $(0, \sigma^2, \gamma, \nu)$, and u_i has a Normal distribution with mean 0 and variance τ^2. We generate x_i from a Normal distribution with mean 5 and variance 9 and keep them fixed during the simulation study. Then, y_i given x_i is generated from a skewed t-distribution with parameters $(\beta_0 + \beta_1 x_i, \sigma^2, \gamma, \nu)$, and w_i given x_i is generated from a Normal distribution with mean x_i and variance τ^2. We choose the true parameters as $\beta_0 = 1, \beta_1 = 3, \sigma^2 = 1$, and different sets of values for τ^2 $(= 1, 3, 10), \gamma$ $(= 0.7, 1.6)$ and ν $(= 3, 5, 10, 20)$. We run $R = 1,000$ simulations to estimate the model parameters and to also provide the corresponding posterior variances.

We also consider different models to evaluate their performances in this set-up. In particular, we consider the following models: skewed t-distribution, skewed Normal, t-distribution, and Normal distribution, noting that the last three models are special cases of skewed t-distribution (3). The bias of parameters estimate and their corresponding mean squared error (MSE) defined as MSE $= bias^2 + var$ are calculated to evaluate the bias and relative efficiency of the model parameters estimate using different models. We may note that the *bias* is based on the average of posterior means of $R = 1,000$ replications subtracted from the true value, and *var* is the average of posterior variances of $R = 1,000$ replications. The results are shown in Tables 1 to 3. As shown in the Tables, the estimate of model parameters in the case of skewed-t distribution are fairly unbiased and the corresponding MSEs are more efficient than other models in most parameters. We observe that with increasing ν, for a given skewness parameter γ, the MSEs of parameter estimators of σ^2 and ν increase for the skewed-t model while the MSE of parameter estimator of σ^2 decreases for the skewed Normal model, although the parameter σ^2 is not directly comparable for the skewed-t and skewed Normal models. Note that the MSE value of $\hat{\sigma}^2$ for the skewed Normal is much larger than the corresponding value of the skewed-t which shows that the skewed Normal model can not properly capture ~~the~~ variation of the dispersion parameter when the data are generated from the skewed-t model. We also observe even worse behavior for the Normal model. Obviously, with increasing the skewness parameter γ, we observe larger MSE of parameter estimator of γ in the cases of skewed-t and skewed Normal models. We also observe that with increasing measurement error variance τ^2, the MSE of parameter estimator of σ^2 increases for ~~the~~ all models.

Also to evaluate the impact of the measurement error on different models, we consider a similar simulation set-up as above, but ignoring measurement error variance. In particular, we generate the data from skewed-t model with the same set-up as above (including the measurement error variances), but fit ~~the~~ models ~~with~~ ignoring the measurement error variance, to be referred to as ~~called~~ the naive models. The results are shown in Tables C1 to C3 in Appendix C. As expected, ignoring the measurement error variance has a big impact on the performance of model parameters for ~~the~~ all models in terms of bias and corresponding MSEs. Especially, we have large bias and MSE for the parameter estimate $\hat{\sigma}^2$ and in particular when the measurement error variance τ^2 increases.

We also conduct ~~employ~~ a sensitivity analysis for the skewed-t model by considering different hyperparameter values for ~~the~~ all priors and observe that the results are not sensitive to the priors (not shown here).

4. PM data application in Canada

In this section, we analyze the PM data in Canada using our proposed approach. Recently, there has been an increasing attention to air population around the world due to its impact on human health and on the environment. Forecasting of the airborne $PM_{2.5}$ (particulate matter with an aerodynamic diameter of less than 2.5 μm) concentrations is of particular interest due to its well known adverse health impact to humans (Paschalidou et al. 2011). It is well known that potential risk factors for the $PM_{2.5}$ are meteorological data including wind speed (WS) and temperature (Roy and Adhikari 2009).

We use the PM data from *Environment and Climate Change Canada* website (https://www.canada.ca/en/environment-climate-change.html) in 2012. We use the annual maximum daily $PM_{2.5}$ concentration as our outcome from 21 major cities in Canada. It is well known that the potential risk factors of $PM_{2.5}$ are WS and tem-

Table 1. Bias and corresponding mean squared error (MSE) of the estimates parameters for different models (skewed t, skewed Normal, t, and Normal) in the case of $\tau^2 = 1$ based on 1000 simulated datasets.

	Skewed t		Skewed Normal		t-distribution		Normal distribution	
True parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
β_0=1	0.17	1.34	0.50	1.98	-0.65	1.47	-0.83	1.84
β_1=3	-0.02	0.03	-0.02	0.03	-0.02	0.03	0.01	0.03
σ^2=1	0.34	0.19	1.04	3.32	0.74	0.72	12.42	272.20
γ=0.7	0.05	0.02	-0.02	0.02	–	-	–	–
ν=3	1.91	4.29	–	–	1.59	3.25	–	–
β_0=1	0.15	1.21	0.32	1.32	-0.59	1.28	-0.70	1.43
β_1=3	-0.02	0.03	-0.02	0.03	-0.02	0.03	0.00	0.03
σ^2=1	0.38	0.20	0.80	0.77	0.78	0.73	10.19	110.35
γ=0.7	0.07	0.02	0.02	0.01	–	-	–	–
ν=5	2.40	6.35	–	–	1.99	4.62	–	–
β_0=1	0.20	1.16	0.31	1.23	-0.53	1.19	-0.63	1.28
β_1=3	-0.02	0.03	-0.02	0.03	-0.02	0.03	0.00	0.03
σ^2=1	0.47	0.28	0.74	0.65	0.91	0.93	9.60	96.94
γ=0.7	0.07	0.02	0.03	0.01	–	-	–	–
ν=10	2.88	9.26	–	–	2.29	6.41	–	–
β_0=1	0.24	1.17	0.32	1.20	-0.51	1.15	-0.60	1.23
β_1=3	-0.02	0.03	-0.02	0.03	-0.02	0.03	0.00	0.03
σ^2=1	0.54	0.37	0.73	0.64	1.05	1.25	9.41	93.10
γ=0.7	0.06	0.02	0.04	0.01	–	-	–	–
ν=20	3.31	13.10	–	–	2.46	8.80	–	–
β_0=1	0.66	1.96	0.62	2.40	1.13	2.36	1.07	2.26
β_1=3	-0.02	0.03	-0.02	0.04	-0.02	0.03	0.00	0.03
σ^2=1	0.42	0.30	1.39	12.41	0.78	0.79	13.46	1360.38
γ=1.6	-0.17	0.13	-0.10	0.18	–	-	–	–
ν=3	1.83	3.99	–	–	1.53	3.08	–	–
β_0=1	0.73	1.89	0.74	2.02	1.02	2.03	0.93	1.81
β_1=3	-0.02	0.03	-0.02	0.03	-0.02	0.03	0.00	0.03
σ^2=1	0.42	0.26	0.94	1.24	0.81	0.78	10.49	119.07
γ=1.6	-0.24	0.14	-0.22	0.14	–	-	–	–
ν=5	2.33	6.03	–	–	1.94	4.45	–	–
β_0=1	0.72	1.80	0.73	1.84	0.95	1.82	0.84	1.61
β_1=3	-0.02	0.03	-0.02	0.03	-0.02	0.03	0.00	0.03
σ^2=1	0.50	0.34	0.82	0.84	0.94	1.01	9.80	101.24
γ=1.6	-0.27	0.15	-0.25	0.14	–	-	–	–
ν=10	2.79	8.90	–	–	2.26	6.34	–	–
β_0=1	0.71	1.81	0.73	1.82	0.92	1.73	0.81	1.54
β_1=3	-0.02	0.03	-0.02	0.03	-0.02	0.03	0.00	0.03
σ^2=1	0.58	0.44	0.80	0.79	1.08	1.30	9.59	96.76
γ=1.6	-0.27	0.15	-0.27	0.15	–	-	–	–
ν=20	3.21	12.80	–	–	2.41	8.68	–	–

Table 2. Bias and corresponding mean squared error (MSE) of the estimates parameters for different models (skewed t, skewed Normal, t, and Normal) in the case of $\tau^2 = 3$ based on 1000 simulated datasets.

	Skewed t		Skewed Normal		t-distribution		Normal distribution	
True parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
β_0=1	0.57	3.44	0.98	4.62	-0.50	2.96	-0.82	3.36
β_1=3	-0.05	0.09	-0.06	0.09	-0.05	0.09	0.01	0.08
σ^2=1	1.02	1.27	2.14	10.17	1.69	3.35	30.58	1099.21
γ=0.7	0.04	0.02	-0.03	0.02	–	-	–	–
ν=3	1.82	3.61	–	–	1.52	2.63	–	–
β_0=1	0.64	3.42	0.90	3.77	-0.42	2.75	-0.70	2.91
β_1=3	-0.05	0.08	-0.05	0.08	-0.05	0.08	0.01	0.08
σ^2=1	1.16	1.58	1.95	4.36	1.56	2.82	28.32	839.19
γ=0.7	0.04	0.01	-0.01	0.01	–	-	–	–
ν=5	2.11	4.75	–	–	-0.39	0.30	–	–
β_0=1	0.75	3.44	0.94	3.85	-0.34	2.62	-0.63	2.76
β_1=3	-0.05	0.08	-0.05	0.08	-0.05	0.08	0.01	0.07
σ^2=1	1.40	2.28	1.91	4.21	2.28	5.81	27.72	802.46
γ=0.7	0.03	0.01	-0.01	0.01	–	-	–	–
ν=10	2.41	6.48	–	–	1.85	4.16	–	–
β_0=1	0.85	3.67	0.99	3.92	-0.33	2.55	-0.60	2.70
β_1=3	-0.05	0.08	-0.05	0.08	-0.05	0.08	0.00	0.07
σ^2=1	1.58	2.87	1.89	4.15	2.60	7.35	27.54	792.22
γ=0.7	0.01	0.01	-0.01	0.01	–	-	–	–
ν=20	2.75	9.19	–	–	1.95	5.54	–	–
β_0=1	1.04	4.58	1.06	4.96	1.36	4.55	1.08	3.78
β_1=3	-0.06	0.09	-0.06	0.09	-0.06	0.09	0.00	0.08
σ^2=1	1.12	1.62	2.54	23.44	1.73	3.52	31.65	2294.30
γ=1.6	-0.25	0.14	-0.22	0.14	–	-	–	–
ν=3	1.76	3.40	–	–	1.49	2.56	–	–
β_0=1	0.98	4.18	1.03	4.50	1.21	3.98	0.93	3.32
β_1=3	-0.05	0.08	-0.05	0.08	-0.05	0.08	0.00	0.08
σ^2=1	1.24	1.86	2.18	6.07	1.93	4.23	28.64	861.35
γ=1.6	-0.27	0.14	-0.26	0.14	–	-	–	–
ν=5	2.05	4.53	–	–	1.70	3.22	–	–
β_0=1	0.94	4.07	1.01	4.20	1.12	3.79	0.85	3.11
β_1=3	-0.05	0.08	-0.06	0.08	-0.05	0.08	0.00	0.08
σ^2=1	1.47	2.57	2.08	5.08	2.33	6.01	27.94	815.60
γ=1.6	-0.28	0.14	-0.28	0.14	–	-	–	–
ν=10	2.29	5.99	–	–	1.84	4.16	–	–
β_0=1	0.92	4.16	0.97	4.23	1.09	3.67	0.82	3.03
β_1=3	-0.05	0.08	-0.05	0.08	-0.05	0.08	0.00	0.08
σ^2=1	1.68	3.33	2.11	5.33	2.60	7.47	27.72	803.01
γ=1.6	-0.27	0.14	-0.28	0.15	–	-	–	–
ν=20	2.55	8.31	–	–	1.94	5.48	–	–

Table 3. Bias and corresponding mean squared error (MSE) of the estimates parameters for different models (skewed t, skewed Normal, t, and Normal) in the case of $\tau^2 = 10$ based on 1000 simulated datasets.

	Skewed t		Skewed Normal		t-distribution		Normal distribution	
True parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
β_0=1	1.68	12.37	2.13	15.51	0.06	9.07	-0.81	8.52
β_1=3	-0.16	0.31	-0.15	0.31	-0.16	0.32	0.01	0.25
σ^2=1	2.35	6.77	4.36	25.56	3.71	16.54	94.13	9373.15
γ=0.7	0.01	0.01	-0.06	0.01	–	-	–	–
ν=3	1.61	2.75	–	–	1.32	1.91	–	–
β_0=1	1.83	13.37	1.70	12.83	0.17	8.47	-0.67	8.05
β_1=3	-0.16	0.31	-0.48	0.28	-0.16	0.31	0.01	0.24
σ^2=1	2.71	8.86	3.58	18.75	4.31	21.68	91.83	8788.53
γ=0.7	0.00	0.01	-0.13	0.01	–	-	–	–
ν=5	1.78	3.35	–	–	1.40	2.17	–	–
β_0=1	2.05	14.16	2.32	15.13	0.21	8.34	-0.61	7.85
β_1=3	-0.17	0.33	-0.16	0.31	-0.16	0.30	0.01	0.24
σ^2=1	3.21	12.29	4.30	21.86	5.14	30.20	91.26	8676.81
γ=0.7	-0.02	0.01	-0.06	0.01	–	-	–	–
ν=10	2.01	4.55	–	–	1.47	2.65	–	–
β_0=1	2.09	14.22	2.17	14.54	0.24	8.36	-0.59	7.85
β_1=3	-0.16	0.32	-0.15	0.29	-0.16	0.30	0.01	0.24
σ^2=1	3.65	15.93	4.25	21.48	5.71	37.14	91.07	8641.32
γ=0.7	-0.03	0.01	-0.05	0.01	–	-	–	–
ν=20	2.18	6.00	–	–	1.53	3.59	–	–
β_0=1	1.55	13.12	1.74	14.31	1.94	12.76	1.09	8.99
β_1=3	-0.15	0.32	-0.17	0.33	-0.17	0.32	0.01	0.25
σ^2=1	2.54	8.00	5.26	76.68	3.72	16.44	95.30	10877.01
γ=1.6	-0.25	0.13	-0.26	0.14	–	-	–	–
ν=3	1.54	2.53	–	–	1.32	1.89	–	–
β_0=1	1.64	13.20	1.65	13.29	1.78	11.91	0.95	8.52
β_1=3	-0.17	0.33	-0.16	0.32	-0.16	0.32	0.01	0.24
σ^2=1	3.01	11.13	4.89	30.81	4.38	22.00	92.23	8869.82
γ=1.6	-0.28	0.14	-0.27	0.14	–	-	–	–
ν=5	1.70	3.13	–	–	1.40	2.21	–	–
β_0=1	1.56	12.66	1.61	13.13	1.78	12.11	0.87	8.31
β_1=3	-0.16	0.31	-0.15	0.31	-0.17	0.32	0.01	0.24
σ^2=1	3.47	14.91	4.76	28.18	5.23	31.85	91.52	8728.72
γ=1.6	-0.27	0.13	-0.29	0.14	–	-	–	–
ν=10	1.85	4.03	–	–	1.49	2.75	–	–
β_0=1	1.50	13.28	1.50	13.44	1.68	11.45	0.83	8.21
β_1=3	-0.15	0.32	-0.14	0.32	-0.16	0.31	0.01	0.24
σ^2=1	4.04	20.32	4.89	32.40	5.85	39.04	91.24	8673.59
γ=1.6	-0.28	0.14	-0.27	0.14	–	-	–	–
ν=20	1.98	5.53	–	–	1.54	3.63	–	–

Table 4. Bayesian mean and 95% credible interval of model parameters estimates; data on particulate matter in Canada.

Parameter	Bayesian estimation Skewed t				Bayesian estimation Skewed Normal			
	Without ME		With ME		With ME		Without ME	
	Mean	95% Credible Interval	Mean	95% Credible Interval	Mean	95% Credible Interval	Mean	95% Credible Interval
β_0	14.05	(-22.16, 56.74)	12.17	(-20.53, 44.87)	13.65	(-28.45 , 52.94)	14.14	(-27.41 , 55.04)
β_1	0.28	(-2.60, 3.00)	0.36	(-1.83 , 2.72)	0.28	(-2.40 , 3.35)	0.29	(-2.42 , 3.23)
β_2	0.03	(-0.56, 0.54)	0.06	(-0.46 , 0.47)	0.01	(-0.52 , 0.47)	0.01	(-0.55 , 0.52)
σ^2	215.52	(49.52, 574.55)	197.49	(36.64, 507.66)	301.81	(88.96 , 703.47)	311.75	(100.11 , 724.06)
ν	9.73	(1.47, 33.00)	9.73	(1.42 , 33.36)	–	–	–	–
γ	1.85	(0.90, 3.13)	1.92	(0.97 , 3.24)	1.92	(1.00 , 3.16)	1.86	(0.97 , 3.02)

perate, among other covariates such as dust, oil combustion, and wood combustion. However, we only have access to $PM_{2.5}$, WS, and temperature. We obtain mean WS (by averaging over the observed values during the day ~~the corresponding values~~) and temperature while the $PM_{2.5}$ concentration was in the highest level at each location.

In the previous studies, multiple linear regression models have been used for the prediction of $PM_{2.5}$. However, it is known from the literature that the covariate WS is measured with error (Danish Wind Turbine Manufacturers Association). The outcome $PM2.5$ is also neither symmetric nor normally distributed (Figure 1). Figure 1 shows histogram of the maximum $PM_{2.5}$ concentration of 21 sites in Canada. As shown in Figure 1, the histogram suggests ~~induces~~ a non-Gaussian feature which needs further investigation for more precise understanding of this feature.

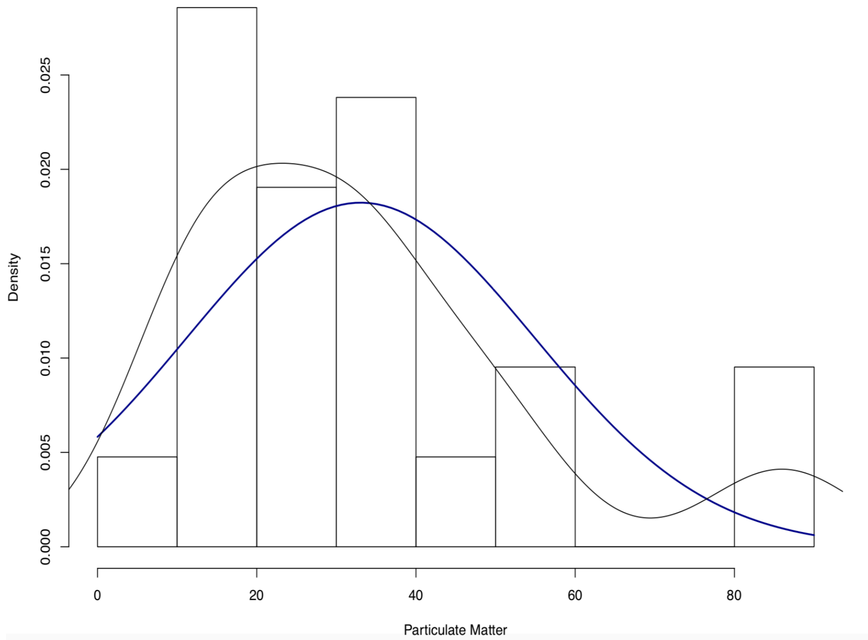

Figure 1. Histogram, normal density (blue), and kernel density (black) estimates of the maximum $PM_{2.5}$ concentration of 21 major cities in Canada in 2012.

The Bayesian analysis of the data, assuming the covariate WS is measured with error, using skewed t- and skewed Normal distribution is provided in Table 4. We also analyze the data using skewed t- and skewed Normal distribution without assuming the measurement error in the covariate WS to evaluate the impact of measurement error in the results. Note that β_1 and β_2 refer to the covariates WS and temperature, respectively.

As shown in Table 4, the parameter β_1 has a smaller credible interval for the skewed t- distribution with the ME compared with the corresponding interval without the ME. Similar behaviour is also observed for the σ^2. The df of t-distribution also explains the difference between the posterior mean of σ^2 in the skewed t- and skewed Normal distribution. In terms of measurement error variation, we assume $\tau^2 = 12.94$ (which is the variance of monthly wind speed values) to avoid identifiability issue in the model. The skewness parameter γ also suggests a large degree of non-symmetry in the observed data. To evaluate the impact of the measurement error in the covariate WS, we compare the skewed t- and skewed Normal distribution with and without measurement error. To evaluate which model performs better to fit the data, we employ the Deviance Information Criterion (DIC). We observe that these values for the skewed t- and skewed Normal, with measurement error in covariate WS, are 419720 and 419780, respectively, which is also another indication that the skewed t distribution fits the data slightly better than the skewed Normal distribution. Note that the DIC values for the skewed t- and skewed Normal distribution in the case of without measurement error in covariate WS are 419799 and 419860, respectively. We also use the Bayes factor (Kass and Raftery 1995) to compare different models. The Bayes factor is the ratio of the predictive probability of the data under the compared models. To compare the ME skewed t- with other models, we calculate the corresponding Bayes factors. In particular, these ratios for the ME skewed Normal, skewed t-, and skewed Normal are 8.4, 17.3, and 24.1. As indicated by Kass and Raftery (1995), the ratios which are bigger than 3 provide strong evidence that the model defined in the numerator of the Bayes factor (ME skewed t-) performs a better fit to ~~much better to fit~~ the data.

5. Concluding remarks

In this paper, we have proposed a Bayesian approach to linear regression models with fat tails and skewed errors in response variable when auxiliary information is measured with error. We have used appropriate prior distributions on the parameters of the model, and we have shown, under a mild condition, that the resulting posterior distribution is proper ~~with finite variances of the model parameters~~. We have compared the proposed model with various versions of the linear regression models with the measurement error in covariates (Normal, t-distribution, and skewed Normal) through a simulation study. The proposed approach has also been ~~also~~ applied to analyze a motivating data application where covariates wind speed (measured with error) and temperature are used for the prediction of particulate matter (which is neither symmetric nor normally distributed) in Canada.

In this paper, we have assumed that the outcomes (response variables) are independent from each other. It would be interesting to study the proposed model in the context of linear mixed models where the outcomes are dependent. We have also assumed that the measurement error variation is normally distributed. One can also study the robustness of the results in terms of misspecification of error distribution. These are some of the topics for future study.

Supplementary materials

The supplementary materials contain R codes and corresponding "readme" files for the simulations and real data application conducted in this paper.

Acknowledgments

Constructive comments and suggestions of a referee, which led to an improved version of this paper, are greatly appreciated. This work was supported by a research grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) to M. Torabi.

References

Arellano-Valle, R.B., S. Ozan, H. Bolfarine, and V. Lachos, *Skew normal measurement error models*, J. Multi. Anal. 96 (2005), pp. 265–281.

Arellano-Valle, R.B., and A. Azzalini, *On the Unification of Families of Skew-normal Distributions*, Scand. J. Stat. 33 (2006), pp. 561–574.

Arellano-Valle, R.B., A. Azzalini, C.S. Ferreira, and K. Santoro, *A two-piece normal measurement error model*, Comput. Stat. & Data Anal. 144 (2020), pp. 1–17.

Azzalini, A.*The Skew-Normal and Related Families*, (Vol. 3), 2013, Cambridge University Press.

Azzalini, A., and A. Dalla Valle, *The multivariate skew-normal distribution*, Biometrika 83 (1996), pp. 715–726.

Badrinath, S.G., and S. Chatterjee, *A data-analytic look at skewness and elongation in common-stock-return distributions*, J. Busin. Econ. Stat. 9 (1991), pp. 223–233.

Buckle, D.J. *Bayesian inference for stable distributions*, J. Amer. Stat. Assoc. 90 (1995), pp. 605–613.

Cabral, C.R.B., V.H. Lachos, and C.B. Zeller, *Multivariate measurement error models using finite mixtures of skew-Student t distributions*, J. Multi. Anal. 124 (2014), pp. 179–198.

Carrol, R.J., D. Ruppert, L.A. Stefanski, and C.M. Crainiceanu, *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd ed., 2006, London: Chapman & Hall.

Chib, S.*Marginal likelihood from the Gibbs output*, J. Amer. Stat. Assoc. 90 (1995), pp. 1313–1321.

Fernandez, C., and M.F.J. Steel, *On Bayesian modeling of fat tails and skewness*, J. Amer. Stat. Assoc. 93 (1998), pp. 359–371.

Ferreira, C.S., V.H. Lachos, and H. Bolfarine, *Inference and diagnostics in skew scale mixtures of normal regression models*, J. Stat. Comput. Simul. 85 (2015), pp. 517–537.

Ferreira, C.S., and R.B. Arellano-Valle, *Estimation and diagnostic analysis in skew-generalized-normal regression models*, J. Stat. Comput. Simul. 88 (2018), pp. 1039–1059.

Fuller, W.*Measurement Error Models*, 1987, New York: Wiley.

Geweke, J.*Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints*, In Comput. Sci. Stat., eds. E.M. Keramidas and S.M. Kaufman, Fairfax Station, VA: Interface Foundation, 1991, pp. 571–578.

Geweke, J.*Priors for macroeconomic time series and their application*, Discussion paper 64, Federal Reserve Bank of Minneapolis, Institute for Empirical Macroeconomics, 1992.

Juarez, M.A., and M.F.J. Steel, *Non-Gaussian dynamic Bayesian modeling for panel data*, J. Appl. Econom. 25 (2010), pp. 1128–1154.

Kass, R., and A. Raftery, *Bayes factors*, J. Amer. Stat. Assoc. 90 (1995), pp. 773–795.

Li, Y., and S.K. Ghosh, *Efficient sampling methods for truncated multivariate normal and*

student-t distributions subject to linear inequality constraints, J. Stat. Theo. Prac. 9 (2015), pp. 712–732.
Lye, J.N., and V.L. Martin, *Robust estimation, nonnormalities, and generalized exponential distributions*, J. Amer. Stat. Assoc. 88 (1993), pp. 261–267.
McDonald, J.B., and R.D. Nelson, *Beta estimation in the market model: skewness and leptokurtosis*, Comm. Stat. Theo. Meth. 22 (1993), pp. 2843–2862.
Paschalidou, A.K., S. Karakitsios, S. Kleanthous, and P.A. Kassomenos, *Forecasting hourly PM10 concentration in Cyprus through artificial neural networks and multiple regression models: implications to local environmental management*, Envir. Sci. Pollu. Res. 18 (2011), pp. 316–327.
Rodrigues, J., and H. Bolfarine, *Bayesian inference for an extended simple regression measurement error model using skewed priors*, Bayes. Anal. 2 (2007), pp. 349–364.
Roy, S., and G.R. Adhikari, *Seasonal variation in suspended particulate matter vis-a-vis meteorological parameters at Kolar Gold Fields, India*, Inter. J. Envir. Engin. 1 (2009), pp. 432–445.
Rubio, F.J., and M.F.J. Steel, *The family of two-piece distributions*, Significance, 17, (2020), pp. 12–13.
Zeller, C.B., V.H. Lachos, and F.E. Vilca-Labra, *Local influence analysis for regression models with scale mixtures of skew-normal distributions*, J. Appl. Stat. 38 (2011), pp. 348–363.

Appendix A. Proof of Theorem 1

Theorem 1. The posterior density $\pi(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \lambda_1, \ldots, \lambda_n, \boldsymbol{\beta}, \sigma^2, \gamma, \nu \mid \boldsymbol{y}, \boldsymbol{w})$ is proper if $n \geq p$. ~~and also all the model parameters $(\lambda_1, \ldots, \lambda_n, \boldsymbol{\beta}, \sigma^2, \gamma, \nu)$ have finite variances.~~

Proof: Let $h(\gamma) = \max(\gamma^2, \gamma^{-2})$ where $\gamma \in (0, \infty)$. Note that

$$\prod_{i=1}^{n} \left\{ \exp\left[-\frac{\lambda_i (y_i - \boldsymbol{x}_i^T \boldsymbol{\beta})^2}{2\sigma^2} \left(\frac{1}{\gamma^2} I_{[y_i \geq \boldsymbol{x}_i^T \boldsymbol{\beta}]} + \gamma^2 I_{[y_i < \boldsymbol{x}_i^T \boldsymbol{\beta}]} \right) \right] \right\} \leq \exp\left[-\frac{1}{2\sigma^2 h(\gamma)} \sum_{i=1}^{n} \lambda_i (y_i - \boldsymbol{x}_i^T \boldsymbol{\beta})^2 \right], \tag{A1}$$

and also,

$$\begin{aligned} \sum_{i=1}^{n} \lambda_i (y_i - \boldsymbol{x}_i^T \boldsymbol{\beta})^2 &= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T \boldsymbol{\Lambda} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\ &= [\boldsymbol{\beta} - (\boldsymbol{X}^T \boldsymbol{\Lambda} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{\Lambda} \boldsymbol{y}]^T (\boldsymbol{X}^T \boldsymbol{\Lambda} \boldsymbol{X}) [\boldsymbol{\beta} - (\boldsymbol{X}^T \boldsymbol{\Lambda} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{\Lambda} \boldsymbol{y}] \\ &\quad + \boldsymbol{y}^T [\boldsymbol{\Lambda} - \boldsymbol{\Lambda} \boldsymbol{X} (\boldsymbol{X}^T \boldsymbol{\Lambda} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{\Lambda}] \boldsymbol{y}, \end{aligned}$$

where $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$. We assume $n \geq p$ to ensure full column rank of the matrix $\boldsymbol{X}$. First, by integrating out with respect to $\boldsymbol{\beta}$, the right side of (A1) has an upper bound such as

$$\int \exp\left[-\frac{1}{2\sigma^2 h(\gamma)} \sum_{i=1}^{n} \lambda_i (y_i - \boldsymbol{x}_i^T \boldsymbol{\beta})^2 \right] d\boldsymbol{\beta} \leq (2\pi)^{\frac{p}{2}} \left[\sigma^2 h(\gamma) \right]^{\frac{p}{2}} |\boldsymbol{X}^T \boldsymbol{\Lambda} \boldsymbol{X}|^{-\frac{1}{2}}.$$

Note that $\boldsymbol{y}^T [\boldsymbol{\Lambda} - \boldsymbol{\Lambda} \boldsymbol{X} (\boldsymbol{X}^T \boldsymbol{\Lambda} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{\Lambda}] \boldsymbol{y} > 0$ is nonnegative definite. Thus, for a generic constant $K(>0)$, we have

$$\pi(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\lambda_1,\ldots,\lambda_n,\sigma^2,\gamma,\nu\mid\boldsymbol{y},\boldsymbol{w})$$
$$\leq K|\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{X}|^{-\frac{1}{2}}(\sigma^2)^{-\frac{n-p}{2}}h(\gamma)^{\frac{p}{2}}\left(\gamma+\frac{1}{\gamma}\right)^{-n}\prod_{i=1}^{n}\left[\lambda_i^{\frac{\nu+1}{2}-1}\exp\left(-\frac{\nu}{2}\lambda_i\right)\right]\left(\frac{\nu}{2}\right)^{\frac{n\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)^{-n}$$
$$\times\exp\left[-\frac{1}{2\tau^2}\sum_{i=1}^{n}\|\boldsymbol{w}_i-\boldsymbol{x}_i\|^2\right](\sigma^2)^{-\frac{a_1}{2}-1}\exp\left(-\frac{a_2}{2\sigma^2}\right)P_\gamma P_\nu.$$

Next observe that

$$h(\gamma)^{\frac{p}{2}}\left(\gamma+\frac{1}{\gamma}\right)^{-n}\leq h(\gamma)^{\frac{p}{2}}\left(\gamma^2+\frac{1}{\gamma^2}\right)^{-\frac{n}{2}}\leq h(\gamma)^{\frac{p}{2}}h(\gamma)^{-\frac{n}{2}}\leq 1,$$

since $h(\gamma)\geq(\gamma^2\cdot\gamma^{-2})^{\frac{1}{2}}=1$ and $n\geq p$. Hence, for proper P_γ, integrating out with respect to γ,

$$\pi(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\lambda_1,\ldots,\lambda_n,\sigma^2,\nu\mid\boldsymbol{y},\boldsymbol{w})$$
$$\leq K|\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{X}|^{-\frac{1}{2}}(\sigma^2)^{-\frac{n-p}{2}}\prod_{i=1}^{n}\left[\lambda_i^{\frac{\nu+1}{2}-1}\exp\left(-\frac{\nu}{2}\lambda_i\right)\right]\left(\frac{\nu}{2}\right)^{\frac{n\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)^{-n}$$
$$\times\exp\left[-\frac{1}{2\tau^2}\sum_{i=1}^{n}\|\boldsymbol{w}_i-\boldsymbol{x}_i\|^2\right](\sigma^2)^{-\frac{a_1}{2}-1}\exp\left(-\frac{a_2}{2\sigma^2}\right)P_\nu.$$

Noting that $n-p+a_1>0$ and $a_2>0$. The integration with respect to σ^2 yields

$$\pi(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\lambda_1,\ldots,\lambda_n,\nu\mid\boldsymbol{y},\boldsymbol{w})$$
$$\leq K|\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{X}|^{-\frac{1}{2}}\prod_{i=1}^{n}\left[\lambda_i^{\frac{\nu+1}{2}-1}\exp\left(-\frac{\nu}{2}\lambda_i\right)\right]\left(\frac{\nu}{2}\right)^{\frac{n\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)^{-n}$$
$$\times\exp\left[-\frac{1}{2\tau^2}\sum_{i=1}^{n}\|\boldsymbol{w}_i-\boldsymbol{x}_i\|^2\right]P_\nu.$$

Define $\lambda_{(1)}\leq\ldots\leq\lambda_{(n)}$ to be the ordered λ_i and define $\{\lambda_{m_1},\ldots,\lambda_{m_p}\}$ as the set of λ_i's that satisfy $\prod_{i=1}^{p}\lambda_{m_i}=\min\left\{\prod_{i=1}^{p}\lambda_{s_i}:1\leq s_1<\ldots<s_p\leq n\text{ and }\mathrm{Det}(\boldsymbol{x}_{s_1},\ldots,\boldsymbol{x}_{s_p})\neq 0\right\},$ where $\mathrm{Det}(\boldsymbol{x}_{s_1},\ldots,\boldsymbol{x}_{s_p})$ denotes the determinant of the submatrix of $\boldsymbol{X}$. By the Cauchy-

Binet Formula, writing $\boldsymbol{G} = \tau^2 \boldsymbol{I}_p$,

$$\begin{aligned}
\left|\boldsymbol{X}^T \boldsymbol{\Lambda} \boldsymbol{X}\right|^{-\frac{1}{2}} &= \left[\sum_{1\le s_1<\ldots<s_p\le n}\left(\prod_{i=1}^{p}\lambda_{s_i}\right)\mathrm{Det}^2(\boldsymbol{x}_{s_1},\ldots,\boldsymbol{x}_{s_p})\right]^{-\frac{1}{2}} \\
&\le \left[\prod_{i=1}^{p}\lambda_{m_i}\right]^{-\frac{1}{2}}\left|\boldsymbol{X}^T\boldsymbol{X}\right|^{-\frac{1}{2}} = \left[\prod_{i=1}^{p}\lambda_{m_i}\right]^{-\frac{1}{2}}\left|\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^T\right|^{-\frac{1}{2}} \\
&= \left[\prod_{i=1}^{p}\lambda_{m_i}\right]^{-\frac{1}{2}}\left|\boldsymbol{G}^{\frac{1}{2}}\sum_{i=1}^{n}\left(\boldsymbol{G}^{-\frac{1}{2}}\boldsymbol{x}_i\right)\left(\boldsymbol{G}^{-\frac{1}{2}}\boldsymbol{x}_i\right)^T\boldsymbol{G}^{\frac{1}{2}}\right|^{-\frac{1}{2}} \\
&= \left[\prod_{i=1}^{p}\lambda_{m_i}\right]^{-\frac{1}{2}}\left|\boldsymbol{G}\right|^{-\frac{1}{2}}\left|\sum_{i=1}^{n}\left(\boldsymbol{G}^{-\frac{1}{2}}\boldsymbol{x}_i\right)\left(\boldsymbol{G}^{-\frac{1}{2}}\boldsymbol{x}_i\right)^T\right|^{-\frac{1}{2}} \\
&\le K\tau^{-1}\left[\prod_{i=1}^{p}\lambda_{m_i}\right]^{-\frac{1}{2}}.
\end{aligned}$$

~~where $\boldsymbol{G} = \tau^2 \boldsymbol{I}_p$.~~ Consider

$$\left[\prod_{i=1}^{p}\lambda_{m_i}\right]^{-\frac{1}{2}}\prod_{i=1}^{n}\left[\lambda_i^{\frac{\nu+1}{2}-1}\exp\left(-\frac{\nu}{2}\lambda_i\right)\right] = \prod_{i\in\{m_1,\ldots,m_p\}}\left[\lambda_i^{\frac{\nu}{2}-1}\exp\left(-\frac{\nu}{2}\lambda_i\right)\right]\prod_{i\in\{m_1,\ldots,m_p\}^c}\left[\lambda_i^{\frac{\nu+1}{2}-1}\exp\left(-\frac{\nu}{2}\lambda_i\right)\right]$$

Now integrating out with respect to $\lambda_1,\ldots,\lambda_n$,

$$\begin{aligned}
&\pi(\nu \mid \boldsymbol{y},\boldsymbol{w}) \\
&\le K\left[\left(\frac{\nu}{2}\right)^{-\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)\right]^p\left[\left(\frac{\nu}{2}\right)^{-\frac{\nu+1}{2}}\Gamma\left(\frac{\nu+1}{2}\right)\right]^{n-p}\left(\frac{\nu}{2}\right)^{\frac{n\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)^{-n}P_\nu \\
&= K\left[\nu^{-\frac{n-p}{2}}\Gamma\left(\frac{\nu+1}{2}\right)^{n-p}\Gamma\left(\frac{\nu}{2}\right)^{-(n-p)}P_\nu\right].
\end{aligned}$$

Note that

$$\begin{aligned}
\Gamma\left(\frac{\nu+1}{2}\right) &= \int_0^\infty x^{\frac{\nu}{2}-\frac{1}{2}}e^{-x}dx = \int_0^\infty x^{\frac{\nu}{4}-\frac{1}{2}}e^{-\frac{x}{2}}x^{\frac{\nu}{4}}e^{-\frac{x}{2}}dx \\
&\le \left[\int_0^\infty x^{\frac{\nu}{2}-1}e^{-x}xdx\right]^{\frac{1}{2}}\left[\int_0^\infty x^{\frac{\nu}{2}}e^{-x}dx\right]^{\frac{1}{2}} = \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{\frac{1}{2}}\left[\Gamma\left(\frac{\nu}{2}+1\right)\right]^{\frac{1}{2}} \\
&= \Gamma\left(\frac{\nu}{2}\right)\left(\frac{\nu}{2}\right)^{\frac{1}{2}}.
\end{aligned}$$

Hence,

$$\nu^{-\frac{n-p}{2}}\Gamma\left(\frac{\nu+1}{2}\right)^{n-p}\Gamma\left(\frac{\nu}{2}\right)^{-(n-p)} \le \nu^{-\frac{n-p}{2}}2^{-\frac{n-p}{2}}\nu^{\frac{n-p}{2}} = 2^{-\frac{n-p}{2}}. \tag{A2}$$

Since P_ν is proper, then, the posterior is proper for $n \ge p$. ~~It is sufficient to note that~~

~~in order to have finite variance for model parameters $(\lambda_1,\dots,\lambda_n,\boldsymbol{\beta},\sigma^2,\gamma,\nu)$, we need $n \geq p$ from (A2). Note that to have finite variance we need to have second moment of the distribution of model parameters bounded; for example, for ν which is obvious from (A2) under $n \geq p$. The same argument will follow similarly for other model parameters.~~ This completes the proof of Theorem 1.

Appendix B. Computational issues

The posterior distribution in (3) cannot be obtained in closed form. The implementation of the Bayesian procedure is greatly facilitated by the Markov chain Monte Carlo technique, in particular the Gibbs sampler and Metropolis-Hastings. These techniques require generating samples from the full conditionals of each of $\lambda_i, \sigma^2, \boldsymbol{x}_i, \boldsymbol{\beta}, \gamma$, and ν given the remaining parameters and the data $(\mathbf{y}, \mathbf{w})$. We note that $\boldsymbol{x} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)^T$ is of full-column rank with probability 1 (with respect to the posterior distribution). The full conditional distributions are specified as follows:

$$
\begin{aligned}
\lambda_i \mid \cdot &\sim \mathrm{G}\left(\frac{\nu+1}{2},\ \frac{1}{2}\left[\nu + \frac{(y_i-\boldsymbol{x}_i^T\boldsymbol{\beta})^2}{\sigma^2\gamma^2} I_{[y_i \geq \boldsymbol{x}_i^T\boldsymbol{\beta}]} + \frac{(y_i-\boldsymbol{x}_i^T\boldsymbol{\beta})^2\gamma^2}{\sigma^2} I_{[y_i < \boldsymbol{x}_i^T\boldsymbol{\beta}]}\right]\right)\\
\sigma^2 \mid \cdot &\sim \mathrm{IG}\left(\frac{n+a_1}{2},\ \frac{1}{2}\left[a_2 + \frac{1}{\gamma^2}\sum_{i=1}^n \lambda_i (y_i-\boldsymbol{x}_i^T\boldsymbol{\beta})^2 I_{[y_i \geq \boldsymbol{x}_i^T\boldsymbol{\beta}]} + \gamma^2 \sum_{i=1}^n \lambda_i (y_i-\boldsymbol{x}_i^T\boldsymbol{\beta})^2 I_{[y_i < \boldsymbol{x}_i^T\boldsymbol{\beta}]}\right]\right)\\
\pi(\boldsymbol{x}_i \mid \cdot) &\propto \left\{\exp\left[-\frac{\lambda_i (y_i-\boldsymbol{x}_i^T\boldsymbol{\beta})^2}{2\sigma^2\gamma^2}\right] I_{[y_i \geq \boldsymbol{x}_i^T\boldsymbol{\beta}]} + \exp\left[-\frac{\lambda_i (y_i-\boldsymbol{x}_i^T\boldsymbol{\beta})^2\gamma^2}{2\sigma^2}\right] I_{[y_i < \boldsymbol{x}_i^T\boldsymbol{\beta}]}\right\}\\
&\quad \times \exp\left(-\frac{1}{2\tau^2}\|\boldsymbol{w}_i - \boldsymbol{x}_i\|^2\right)\\
\pi(\boldsymbol{\beta} \mid \cdot) &\propto \prod_{i=1}^n \left\{\exp\left[-\frac{\lambda_i (y_i-\boldsymbol{x}_i^T\boldsymbol{\beta})^2}{2\sigma^2\gamma^2}\right] I_{[y_i \geq \boldsymbol{x}_i^T\boldsymbol{\beta}]} + \exp\left[-\frac{\lambda_i (y_i-\boldsymbol{x}_i^T\boldsymbol{\beta})^2\gamma^2}{2\sigma^2}\right] I_{[y_i < \boldsymbol{x}_i^T\boldsymbol{\beta}]}\right\}\\
\pi(\gamma \mid \cdot) &\propto \left(\gamma + \frac{1}{\gamma}\right)^{-n} \exp\left\{-\frac{1}{2\sigma^2}\left[\frac{1}{\gamma^2}\sum_{i=1}^n \lambda_i (y_i-\boldsymbol{x}_i^T\boldsymbol{\beta})^2 I_{[y_i \geq \boldsymbol{x}_i^T\boldsymbol{\beta}]} + \gamma^2 \sum_{i=1}^n \lambda_i (y_i-\boldsymbol{x}_i^T\boldsymbol{\beta})^2 I_{[y_i < \boldsymbol{x}_i^T\boldsymbol{\beta}]}\right]\right\} P_\gamma\\
\pi(\nu \mid \cdot) &\propto \left\{\prod_{i=1}^n \left[\lambda_i^{\frac{\nu}{2}} \exp\left(-\frac{\nu}{2}\lambda_i\right)\right]\right\} \left(\frac{\nu}{2}\right)^{\frac{n\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)^{-n} P_\nu
\end{aligned}
$$

We use the Gibbs sampling technique to draw samples from the first three parameters which have standard distributions. To generate samples from the full conditional of $\boldsymbol{\beta}$, we can generate samples from truncated normal random variates by the mixed rejection algorithm of Geweke (1991). The conditional posterior pdf of β_j for $j = 1, \dots, p$ is then

$$
\pi(\beta_j \mid \beta_{(-j)}, \cdot) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n \lambda_i (y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2 \left\{\frac{1}{\gamma^2} I_{[y_i \geq \boldsymbol{x}_i^T\boldsymbol{\beta}]} + \gamma^2 I_{[y_i < \boldsymbol{x}_i^T\boldsymbol{\beta}]}\right\}\right], \quad \text{(B1)}
$$

where the suffix $(-j)$ denote the vector without the $j-$th component (Fernandez and Steel 1998). Define $w_i^j = (y_i - \boldsymbol{x}_i^T\boldsymbol{\beta} + x_{ij}\beta_j)/x_{ij}$. Ordering the observations as $w_{(1)}^j < w_{(2)}^j < \dots < w_{(n)}^j$ and partitioning R, the domain of β_j, into the sets $S_0^j =$

$\left(-\infty, w^j_{(1)}\right], S^j_1 = \left(w^j_{(1)}, w^j_{(2)}\right], \ldots, S^j_n = \left(w^j_{(n)}, \infty\right)$, equation (B1) can be rewritten as

$$\pi(\beta_j \mid \beta_{(-j)}, \cdot) \propto \sum_{h=0}^{n} \{p^j_h\}^{-1/2} \exp\left(-\frac{l^j_h}{2\sigma^2}\right) f^1_N\left(\beta_j \middle| \mu^j_h, \frac{\sigma^2}{p^j_h}\right) I_{S^j_h}(\beta_j),$$

with $f^1_N(\cdot|t, v)$ the pdf of a univariate normal distribution with mean t and variance v, $I_S(\cdot)$ is the indicator function of the set S, and

$$\begin{aligned} p^j_h &= \sum_{i=1}^{h} \rho^j_{i1} + \sum_{i=h+1}^{n} \rho^j_{i2}, \\ p^j_h \mu^j_h &= \sum_{i=1}^{h} \rho^j_{i1} w^j_{(i)} + \sum_{i=h+1}^{n} \rho^j_{i2} w^j_{(i)}, \\ l^j_h &= \sum_{i=1}^{h} \rho^j_{i1} \{w^j_{(i)}\}^2 + \sum_{i=h+1}^{n} \rho^j_{i2} \{w^j_{(i)}\}^2 - p^j_h \{\mu^j_h\}^2, \end{aligned}$$

where

$$\begin{aligned} \rho^j_{i1} &= \lambda_i x_{ij} \left\{ \frac{1}{\gamma^2} I_{(-\infty,0)}(x_{ij}) + \gamma^2 I_{(0,\infty)}(x_{ij}) \right\}, \\ \rho^j_{i2} &= \lambda_i x_{ij} \left\{ \gamma^2 I_{(-\infty,0)}(x_{ij}) + \frac{1}{\gamma^2} I_{(0,\infty)}(x_{ij}) \right\}. \end{aligned}$$

Truncated univariate normal random variates are generated through the mixed rejection algorithm proposed by Li and Ghosh (2015).
Next we discuss how to generate samples from the conditional distribution of ν. After choosing P_ν as an exponential distribution,

$$P_\nu = k \exp(-k\nu),$$

the posterior conditional distribution of ν becomes

$$\pi(\nu \mid \cdot) \propto \left(\frac{\nu}{2}\right)^{\frac{n\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)^{-n} \exp\left[-\nu\left\{k + \frac{1}{2}\sum_{i=1}^{n}(\lambda_i - \log\lambda_i)\right\}\right].$$

We draw samples from the posterior conditional distribution by efficient rejection sampling using exponential source density as described in Geweke (1992). The target distribution has kernel density

$$f(\nu \mid n, \eta) = \left(\frac{\nu}{2}\right)^{\frac{n\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)^{-n} \exp(-\nu\eta),$$

where $\eta = k + \frac{1}{2}\sum_{i=1}^{n}(\lambda_i - \log\lambda_i)$ and $\eta \geq \frac{n}{2} + k > \frac{n}{2}$. The sampling distribution is exponential with kernel density function $g(\nu \mid \alpha) = \alpha\text{exp}(-\alpha\nu)$. Consider the function

$$Q \overset{\text{let}}{=} \log\left\{\frac{f(\nu \mid n, \eta)}{g(\nu \mid \alpha)}\right\} = \frac{n\nu}{2}\log\left(\frac{\nu}{2}\right) - n\log\left\{\Gamma\left(\frac{\nu}{2}\right)\right\} + (\alpha - \eta)\nu - \log\alpha,$$

and their partial derivatives are

$$\frac{\partial Q}{\partial \nu} = \frac{n}{2}\log\left(\frac{\nu}{2}\right) + \frac{n}{2} - \frac{n}{2}\Psi\left(\frac{\nu}{2}\right) + (\alpha - \eta), \text{ and } \frac{\partial Q}{\partial \alpha} = \nu - \frac{1}{\alpha},$$

where $\Psi(\cdot)$ being digamma function. Since $\log\left(\frac{\nu}{2}\right) + 1 - \Psi\left(\frac{\nu}{2}\right)$ is monotone decreasing function and $\eta > \frac{n}{2}$, Q has a unique regular maximum defined by solving $\frac{\partial Q}{\partial \nu} = 0$ and $\frac{\partial Q}{\partial \alpha} = 0$. Let ν^* be the solution of

$$\frac{n}{2}\left\{\log\left(\frac{\nu^*}{2}\right) + 1 - \Psi\left(\frac{\nu^*}{2}\right)\right\} + \frac{1}{\nu^*} - \eta = 0.$$

Hence, one can draw from the exponential distribution with density function $\frac{1}{\nu^*}\text{exp}\left(-\frac{\nu}{\nu^*}\right)$ and retains the draw with probability

$$\left[\Gamma\left(\frac{\nu^*}{2}\right)\right]^n \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-n} \left(\frac{\nu^*}{2}\right)^{-\frac{n\nu^*}{2}} \left(\frac{\nu}{2}\right)^{\frac{n\nu}{2}} \text{exp}\left[(1/\nu^* - \eta)(\nu - \nu^*)\right],$$

with the Geweke retention probability

$$\left[\Gamma\left(\frac{\nu^*}{2}\right)\right]^n \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-n} \left(\frac{\nu}{\nu^*}\right)^{\frac{n\nu}{2}} \text{exp}\left[(1/\nu^* - \eta)(\nu - \nu^*)\right].$$

Now turn to how to generate samples from the conditional distribution of γ. With a gamma (a, b) on $\varphi \equiv \gamma^2$, the conditional posterior distribution of φ is

$$\pi(\varphi \mid \cdot) \propto \varphi^{\frac{n}{2}+a-1}(\varphi + 1)^{-n}\text{exp}\left\{-\left(\frac{\vartheta}{\varphi} + \kappa\varphi\right)\right\},$$

where $\vartheta = \frac{1}{2\sigma^2}\sum_{i=1}^n \lambda_i(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2 I_{[y_i \geq \boldsymbol{x}_i^T\boldsymbol{\beta}]} \geq 0$ and $\kappa = b + \frac{1}{2\sigma^2}\sum_{i=1}^n \lambda_i(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2 I_{[y_i < \boldsymbol{x}_i^T\boldsymbol{\beta}]} > 0$.

Using the result of Chib (1995), since $\text{exp}\left\{-\left(\frac{\vartheta}{\varphi} + \kappa\varphi\right)\right\}$ is uniformly bounded by 1, the target density comes from the part $\varphi^{\frac{n}{2}+a-1}(\varphi + 1)^{-n}$. By taking a transformation $u = \frac{\varphi}{1+\varphi}$, we have

$$\pi(u \mid \cdot) \propto u^{\frac{n}{2}+a-1}(1 - u)^{\frac{n}{2}-a-1}\text{exp}\left\{-\left(\frac{\vartheta(1-u)}{u} + \frac{\kappa u}{1-u}\right)\right\},$$

where $\vartheta = \frac{1}{2\sigma^2}\sum_{i=1}^n \lambda_i(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2 I_{[y_i \geq \boldsymbol{x}_i^T\boldsymbol{\beta}]} \geq 0$ and $\kappa = b + \frac{1}{2\sigma^2}\sum_{i=1}^n \lambda_i(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2 I_{[y_i < \boldsymbol{x}_i^T\boldsymbol{\beta}]} > 0$.

Again, $\text{exp}\left\{-\left(\frac{\vartheta(1-u)}{u} + \frac{\kappa u}{1-u}\right)\right\}$ is uniformly bounded by 1, and the target density is

$h(u) \sim \text{Beta}\left(\frac{n}{2}+a, \frac{n}{2}-a\right)$.

Finally, we now turn to generate samples from the conditional distribution of $\boldsymbol{x}_i$. Using the result of Chib (1995), we can set the target density as $N\left(\boldsymbol{w}_i, \tau^2\boldsymbol{I}\right)$ and the uniformly bounded part as $\left\{\exp\left[-\frac{\lambda_i(y_i-\boldsymbol{x}_i^T\boldsymbol{\beta})^2}{2\sigma^2\gamma^2}\right] I_{[y_i\geq\boldsymbol{x}_i^T\boldsymbol{\beta}]} + \exp\left[-\frac{\lambda_i(y_i-\boldsymbol{x}_i^T\boldsymbol{\beta})^2\gamma^2}{2\sigma^2}\right] I_{[y_i<\boldsymbol{x}_i^T\boldsymbol{\beta}]}\right\}$ which is bounded by 1 with the exponential function.

Appendix C. Simulation study–more results

Table C1. Bias and corresponding mean squared error (MSE) of the estimates parameters for different models (skewed t, skewed Normal, t, and Normal) for $\tau^2 = 1$ based on 1000 simulated datasets in the case of naive model (ignoring measurement error variance).

	Skewed t		Skewed Normal		t-distribution		Normal distribution	
True parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
β_0=1	-0.03	2.48	0.17	3.20	-0.72	1.62	-0.83	1.84
β_1=3	0.00	0.04	0.01	0.04	0.00	0.03	0.01	0.03
σ^2=1	6.57	47.39	9.68	105.19	7.66	63.28	12.45	274.90
γ=0.7	0.20	0.09	0.17	0.09	–	-	–	–
ν=3	3.13	11.35	–	–	3.09	11.22	–	–
β_0=1	-0.12	2.34	-0.06	2.47	-0.66	1.44	-0.70	1.44
β_1=3	0.00	0.03	0.01	0.03	0.00	0.03	0.00	0.03
σ^2=1	6.62	47.74	8.81	84.51	7.64	62.92	10.21	110.71
γ=0.7	0.23	0.10	0.22	0.10	–	-	–	–
ν=5	3.84	17.32	–	–	-3.83	17.39	–	–
β_0=1	-0.15	2.31	-0.12	2.42	-0.62	1.30	-0.63	1.28
β_1=3	0.00	0.03	0.01	0.03	0.00	0.03	0.00	0.03
σ^2=1	6.86	51.04	8.37	75.82	7.88	65.96	9.62	97.42
γ=0.7	0.24	0.11	0.24	0.12	–	-	–	–
ν=10	4.56	27.42	–	–	4.52	27.33	–	–
β_0=1	-0.15	2.30	-0.13	2.38	-0.60	1.24	-0.60	1.23
β_1=3	0.00	0.03	0.01	0.03	0.00	0.03	0.00	0.03
σ^2=1	7.15	55.31	8.23	73.15	8.20	71.14	9.43	93.53
γ=0.7	0.25	0.12	0.25	0.12	–	-	–	–
ν=20	4.99	43.14	–	–	4.86	43.20	–	–
β_0=1	0.33	2.94	0.13	3.58	0.98	2.08	1.07	2.26
β_1=3	0.00	0.04	0.00	0.04	0.00	0.03	0.00	0.04
σ^2=1	6.64	48.36	9.95	115.78	7.90	67.24	13.49	1373.35
γ=1.6	-0.34	0.26	-0.28	0.29	–	-	–	–
ν=3	3.06	10.79	–	–	2.99	10.52	–	–
β_0=1	0.50	3.08	0.45	3.32	0.90	1.82	0.92	1.81
β_1=3	0.00	0.03	0.00	0.03	0.00	0.03	0.00	0.03
σ^2=1	6.61	47.47	8.77	83.57	7.82	65.43	10.51	119.49
γ=1.6	-0.41	0.30	-0.40	0.31	–	-	–	–
ν=5	3.81	17.00	–	–	3.77	16.87	–	–
β_0=1	0.60	3.30	0.54	3.34	0.84	1.64	0.85	1.62
β_1=3	0.00	0.04	0.00	0.03	0.00	0.03	0.00	0.03
σ^2=1	2.24	5.88	8.30	74.84	8.03	68.50	9.82	101.65
γ=1.6	-0.43	0.32	-0.44	0.34	–	-	–	–
ν=10	-8.95	80.14	–	–	4.46	26.86	–	–
β_0=1	0.55	3.15	0.55	3.27	0.81	1.55	0.81	1.54
β_1=3	0.00	0.03	0.00	0.03	0.00	0.03	0.00	0.03
σ^2=1	7.06	53.98	8.14	71.80	8.34	73.67	9.60	97.09
γ=1.6	-0.44	0.34	-0.44	0.34	–	-	–	–
ν=20	4.95	43.34	–	–	4.79	42.71	–	–

Table C2. Bias and corresponding mean squared error (MSE) of the estimates parameters for different models (skewed t, skewed Normal, t, and Normal) for $\tau^2 = 3$ based on 1000 simulated datasets in the case of naive model (ignoring measurement error variance).

	Skewed t		Skewed Normal		t-distribution		Normal distribution	
True parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
β_0=1	0.08	6.45	0.28	7.74	-0.76	3.27	-0.82	3.33
β_1=3	0.01	0.09	0.00	0.09	0.00	0.08	0.01	0.08
σ^2=1	18.70	376.89	26.00	732.32	21.37	483.97	21.63	1102.33
γ=0.7	0.24	0.11	0.23	0.12	–	-	–	–
ν=3	3.38	12.68	–	–	3.38	12.71	–	–
β_0=1	0.05	6.16	0.18	6.76	-0.68	3.01	-0.69	2.91
β_1=3	0.01	0.09	0.00	0.08	0.00	0.08	0.01	0.08
σ^2=1	19.32	400.30	24.94	666.23	21.96	509.67	28.36	841.55
γ=0.7	0.25	0.12	0.25	0.12	–	-	–	–
ν=5	3.96	17.97	–	–	3.96	18.10	–	–
β_0=1	0.09	6.15	0.12	6.55	-0.62	2.84	-0.63	2.75
β_1=3	0.01	0.08	0.01	0.08	0.00	0.08	0.01	0.07
σ^2=1	20.32	442.20	24.42	637.88	23.07	560.89	27.76	805.28
γ=0.7	0.26	0.12	0.26	0.13	–	-	–	–
ν=10	4.55	27.41	–	–	4.54	27.49	–	–
β_0=1	0.10	6.25	0.12	6.52	-0.59	2.74	-0.60	2.70
β_1=3	0.01	0.08	0.01	0.08	0.00	0.08	0.00	0.07
σ^2=1	21.30	485.46	24.30	631.02	24.19	614.77	27.57	794.27
γ=0.7	0.26	0.12	0.26	0.13	–	-	–	–
ν=20	4.94	43.63	–	–	4.82	43.08	–	–
β_0=1	0.50	8.09	0.39	9.29	1.03	3.84	1.08	3.78
β_1=3	0.00	0.09	0.00	0.09	0.00	0.09	0.00	0.08
σ^2=1	18.46	366.52	25.85	727.12	21.66	498.49	31.69	2258.02
γ=1.6	-0.42	0.32	-0.41	0.34	–	-	–	–
ν=3	3.34	12.40	–	–	3.34	12.42	–	–
β_0=1	0.58	8.16	0.56	8.75	0.93	3.44	0.94	3.31
β_1=3	0.00	0.09	0.00	0.09	0.00	0.08	0.00	0.08
σ^2=1	18.90	383.67	24.48	643.85	22.14	518.54	28.67	863.23
γ=1.6	-0.46	0.34	-0.45	0.36	–	-	–	–
ν=5	3.92	17.82	–	–	3.94	17.90	–	–
β_0=1	0.58	8.57	0.56	8.56	0.86	3.20	0.85	3.11
β_1=3	0.00	0.09	0.00	0.08	0.00	0.08	0.00	0.08
σ^2=1	20.16	435.44	23.94	614.62	23.22	568.39	27.99	818.48
γ=1.6	-0.46	0.36	-0.46	0.37	–	-	–	–
ν=10	4.62	27.92	–	–	4.50	27.28	–	–
β_0=1	0.57	7.97	0.55	8.50	0.82	3.08	0.82	3.04
β_1=3	-0.01	0.08	0.00	0.08	0.00	0.08	0.00	0.08
σ^2=1	20.72	461.77	23.77	605.82	24.34	622.45	27.76	805.23
γ=1.6	-0.46	0.35	-0.47	0.37	–	-	–	–
ν=20	4.77	42.52	–	–	4.77	42.93	–	–

Table C3. Bias and corresponding mean squared error (MSE) of the estimates parameters for different models (skewed t, skewed Normal, t, and Normal) for $\tau^2 = 10$ based on 1000 simulated datasets in the case of naive model (ignoring measurement error variance).

	Skewed t		Skewed Normal		t-distribution		Normal distribution	
True parameter	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
β_0=1	0.63	19.10	0.69	22.93	-0.77	8.97	-0.81	8.50
β_1=3	0.02	0.30	0.02	0.27	0.01	0.27	0.01	0.25
σ^2=1	61.58	4012.07	82.27	7259.70	68.78	4997.85	94.23	9394.25
γ=0.7	0.25	0.11	0.25	0.12	–	-	–	–
ν=3	3.53	13.51	–	–	3.49	13.29	–	–
β_0=1	0.51	18.87	0.68	22.02	-0.67	8.45	-0.68	8.05
β_1=3	0.01	0.27	0.01	0.27	0.01	0.26	0.01	0.24
σ^2=1	62.93	4242.94	80.90	6996.57	71.92	5451.09	91.91	8804.12
γ=0.7	0.26	0.12	0.26	0.13	–	-	–	–
ν=5	3.90	17.65	–	–	3.98	18.19	–	–
β_0=1	0.64	20.20	0.73	21.89	-0.61	8.13	-0.61	7.88
β_1=3	0.01	0.27	0.01	0.26	0.01	0.25	0.01	0.24
σ^2=1	67.34	4847.11	80.36	6900.12	76.29	6114.40	91.35	8694.65
γ=0.7	0.26	0.12	0.26	0.13	–	-	–	–
ν=10	4.55	27.52	–	–	4.52	27.38	–	–
β_0=1	0.69	20.75	0.76	22.03	-0.58	7.97	-0.58	7.82
β_1=3	0.01	0.27	0.01	0.26	0.01	0.24	0.01	0.23
σ^2=1	70.81	5353.37	80.18	6867.24	80.22	6744.18	91.14	8655.48
γ=0.7	0.26	0.13	0.26	0.13	–	-	–	–
ν=20	4.90	42.87	–	–	4.77	42.60	–	–
β_0=1	0.54	25.10	0.50	27.79	1.08	9.56	1.10	9.02
β_1=3	0.00	0.29	0.01	0.28	0.00	0.27	0.01	0.25
σ^2=1	59.13	3748.30	80.82	7000.18	69.07	5042.69	95.38	1087.06
γ=1.6	0.14	0.35	-0.46	0.37	–	-	–	–
ν=3	3.45	12.98	–	–	3.48	13.18	–	–
β_0=1	0.50	24.18	0.53	26.70	0.96	8.99	0.94	8.50
β_1=3	0.01	0.28	0.01	0.27	0.00	0.26	0.01	0.24
σ^2=1	62.00	4106.49	79.24	6710.10	72.07	5473.74	92.29	8882.76
γ=1.6	-0.47	0.35	-0.47	0.37	–	-	–	–
ν=5	3.96	18.02	–	–	3.95	18.00	–	–
β_0=1	0.45	24.32	0.47	26.15	0.87	8.54	0.87	8.28
β_1=3	0.01	0.27	0.01	0.27	0.01	0.25	0.01	0.24
σ^2=1	67.87	4636.50	78.78	6630.68	76.46	6142.32	91.60	8742.74
γ=1.6	-0.47	0.36	-0.48	0.37	–	-	–	–
ν=10	4.54	27.35	–	–	4.50	27.23	–	–
β_0=1	0.33	23.27	0.43	26.15	0.83	8.35	0.84	8.22
β_1=3	0.02	0.27	0.01	0.27	0.01	0.25	0.01	0.24
σ^2=1	68.81	5057.00	78.51	6590.02	80.36	6767.46	91.37	8698.90
γ=1.6	-0.47	0.35	-0.47	0.37	–	-	–	–
ν=20	4.68	41.99	–	–	4.79	42.98	–	–