Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
`http://siba-ese.unisalento.it/index.php/ejasa/index`
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v16n3p722

Using t-distribution for Robust Hierarchical Bayesian Small Area Estimation under Measurement Error in Covariates

By Zarei and Arima

15 December 2023

Electronic Journal of Applied Statistical Analysis
Vol. 16, Issue 3, 2023, 722-739
DOI: 10.1285/i20705948v16n3p722

Using t-distribution for Robust Hierarchical Bayesian Small Area Estimation under Measurement Error in Covariates

Shaho Zarei*[a] and Serena Arima[b]

[a]*Department of Statistics, Faculty of Science, University of Kurdistan, Sanandaj, Iran*
[b]*Departiment of history, society and human studies, University of Salento, Lecce, Italy*

15 December 2023

Small area estimation often suffers from imprecise direct estimators due to small sample sizes. One method for giving direct estimators more strength is to use models. Models employ area effects and include supplementary information from extra sources as covariates to increase the accuracy of direct estimators. The valid covariates are the basis of the small area estimation. Therefore, measurement error (ME) in covariates can produce contradictory results, i.e., even reduce the precision of direct estimators. The measurement error is usually assumed normally distributed with a known mean and variance in most cases. However, in real problem, there might be situations in which the normality assumption of MEs does not hold. In addition, the assumption of known ME variance is restricted. To address these issues and obtain a more robust model, we propose modeling ME using a t-distribution with known and unknown degrees of freedom. Model parameters are estimated using a fully Bayesian framework based on MCMC methods. We validate our proposed model using simulated data and apply it to well-known crop data and the cost and income of households living in Kurdistan province of Iran. The results of the proposed model are promising and, especially in presence of outlying observations, the proposed approach performs better than competing ones.

keywords: Small area estimation, MCMC methods, Area-level model, Measurement error, Hierarchical Bayesian modelling.

*Corresponding author: sh.zarei@uok.ac.ir

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

1 Introduction

For survey statisticians, small area statistics are becoming an increasingly significant subject. The main problem is that direct design-based estimates for small areas usually suffer from large standard errors due to small sample sizes. Consequently, model-based estimation of small area parameters has received a lot of attention in statistical literature aiming at extracting information from sources other than the survey data (see Molina and Rao (2015) for a complete review). Literature on this topic is continuously evolving and several aspects have been investigated. To improve the precision of the small area estimates, model-based methods can be used to borrow strength, by relating the survey and auxiliary data through use of linking models, and by introducing area-specific random effects and covariates. Covariates play a key role in these models: they can be obtained from different data sources such as administrative records or censuses. A problem arises when we suspect that covariates may be affected by measurement error (ME). As well known, when covariates are measured with error, small area estimates may be lead to biased model parameter estimates and they may result in a loss of power for detecting interesting relationships among variables.

In Ybarra and Lohr (2008) the authors suggested a suitable modification of area-level model to the estimates of small area effects in a Fay-Herriot (FH) model where some of the covariates are measured with error. They explain that "when the auxiliary information used in the model is measured with error, using a small area estimator such as the Fay-Herriot estimator while ignoring measurement error may be worse than simply using the direct estimator". Adopting a Bayesian approach, Arima *et al.* (2015) rewrite the measurement error model as a hierarchical model. They use improper non-informative priors on the model parameters and show, under a mild condition, that the joint posterior distribution is proper and the marginal posterior distributions of the model parameters have finite variances. In that framework, the measurement error is modelled as a Normal noise. However, there might be situations that ME are not normally distributed. For example, people tend to over declare or under declare their income depending on specific situations. Moreover, mismeasured covariates can show extreme or outlying values. One way to model outliers in small area estimation is to use t-distribution. As an example, in Bell and Huang (2006), the authors modelled the random effects through a t-distribution with known degrees of freedom (df) to deal with outliers in the data. They explained that the use of a t-distribution with few degrees of freedom can decrease the impact of outliers on the estimation process. Furthermore, Ghosh *et al.* (2018) used modified Jeffry prior for random effects via t-distribution with unknown df. In ME context, Hariyanto *et al.* (2020) in unit-level model used t-distribution with known df to deal with outliers in ME covariates.

To overcome the effects of outlier measurement errors and to make the approach robust with respect violations of the normality assumption of the measurement error, we propose to model the ME in auxiliary variables with t-distribution in area-level model. As highlighted in the simulation study, our proposal makes the small area estimates more robust with respect to violations of the normality assumption as well as to the presence of outlying ME observations.

The paper is organized as follows. In Section 2 we explain our proposal and the computational issues related to the model estimation are outlined in Section 3 and Section 4 for ME with t-distribution with known and unknown df, respectively. A simulation study, described in Section 5, is designed to compare the different measurement error models, highlighting the competitiveness of the proposed model. In Section 6, the well known crop areas (LANDSAT) data as in Battese *et al.* (1998) and the cost and income of households living in Kurdistan province of Iran are analysed. The paper concludes with some discussions in Section 7.

2 The FH model under t-distributed measurement error

The Fay-Herriot model is defined as

$$y_i = \boldsymbol{X}_i^T \boldsymbol{\beta} + v_i + e_i, \qquad i = 1, \dots, m, \tag{1}$$

where y_i is the sample value of variable of interest and the direct estimate of ith small area parameter, $\boldsymbol{X}_i = (X_{i1}, \dots, X_{id})^T$ is the d-dimensional vector of covariates, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)^T$ is a d-dimensional vector of regression coefficients and m is the number of small areas. The sampling error vector $\boldsymbol{e} = (e_1, \dots, e_m)^T$ and the random effect vector $\boldsymbol{v} = (v_1, \dots, v_m)^T$ are assumed to be independent. In (1), it is assumed that v_is are independent and identically distributed (i.i.d) from $N(0, \sigma_v^2)$ distribution, and e_is are independent with $e_i \sim N(0, \psi_i)$ distribution, where ψ_is are all assumed to be known.
When covariates $\boldsymbol{X}_i$ are affected by ME, we assume that they cannot observed directly but we observe a surrogate, $\boldsymbol{W}_i$. In other words, ME models assume that $\boldsymbol{W}_i$ is an estimator for $\boldsymbol{X}_i$ and

$$\boldsymbol{W}_i = \boldsymbol{X}_i + \boldsymbol{\eta}_i$$

or $W_{ij} = X_{ij} + \eta_{ij}$, $j = 1, \dots, d$, where $\boldsymbol{\eta}_i = (\eta_{i1}, \dots, \eta_{id})^T$ is a vector of measurement errors and independent of $\boldsymbol{X}_i$.

ME is typically modelled with a normal distribution (Arima *et al.*, 2015; Goo and Kim , 2013). However, there might be situations that the normality assumption does not hold especially when the observed data are contaminated with outliers. To overcome this issue and to make small area estimates more robust, we propose to assume that ME in auxiliary variables are measured with t-distribution with known and unknown k_i degrees of freedom, i.e., $\eta_{ij} \sim t_{k_i}$ for $j = 1, \dots, d$. We call this model t-ME Fay-Herriot (t-MEFH) model.

Hereafter, following Arima *et al.* (2015), we divide the covariate vector in two parts: with and without measurement error. So the proposed model with t-distribution measurement error in covariates is defined as:

$$y_i = \boldsymbol{X}_i^T \boldsymbol{\beta} + \boldsymbol{z}_i^T \boldsymbol{\tau} + v_i + e_i,$$

$$\boldsymbol{W}_i = \boldsymbol{X}_i + \boldsymbol{\eta}_i.$$

where $\boldsymbol{z}_i = (z_1, \dots, z_q)^T$ is vector of covariates without ME and $\boldsymbol{\tau} = (\tau_1, \dots, \tau_q)^T$ is a q-dimensional vector of regression coefficients. When one deals with one covariate measured with error, $d = 1$, the distribution of the observed covariate w_i conditional to the true covariate X_i is a univariate t-distribution with k_i degrees of freedom, namely

$$\frac{\Gamma(\frac{k_i+1}{2})}{\Gamma(\frac{k_i}{2})\sqrt{k_i\pi}}\big(1+\frac{(w_i-X_i)^2}{k_i}\big)^{-\frac{k_i+1}{2}}.$$

Notice that X_i acts as location parameter and the scale parameter is fixed equal to 1. In the proposed approach, we assume that the degrees of freedom are a-priori known. This assumption is necessary for identifying the model as well as the knowledge of the variance of the measurement error in the FH model. However, when the degrees of freedom are unknown, one can fit different models with different degrees of freedom and select the best one according to model selection criteria.

3 Parameter estimation in t-MEFH: known df

The following hierarchical model can be employed to represent the model.

Stage 1. $y_i|\theta_i \overset{ind}{\sim} N(\theta_i, \psi_i)$ where ψ_i is known.

$$\boldsymbol{W}_i|\boldsymbol{X}_i, k_i \overset{ind}{\sim} t(\boldsymbol{X}_i, k_i)$$

Stage 2. $\theta_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_v^2 \overset{ind}{\sim} N(\boldsymbol{X}_i^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{\tau}, \sigma_v^2)$

Stage 3. $\pi(X_1, \ldots, X_m, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_v^2) = 1$

where $t(\boldsymbol{X}_i, k_i)$ denote the joint distribution of $t(X_{ij}, k_i)$ for $j = 1, \ldots, d$. From a computational point of view, it is convenient to re-parametrize the model using the scale mixtures of normals property of t-distribution distribution, that is

Stage 1. $y_i|\theta_i \overset{ind}{\sim} N(\theta_i, \psi_i)$

$$\boldsymbol{W}_i|\lambda_i, \boldsymbol{X}_i \overset{ind}{\sim} N_d(\boldsymbol{X}_i, \lambda_i I)$$

$$\lambda_i|k_i \overset{ind}{\sim} IG(\tfrac{k_i}{2}, \tfrac{k_i}{2})$$

Stage 2. $\theta_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_v^2 \overset{ind}{\sim} N(\boldsymbol{X}_i^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{\tau}, \sigma_v^2)$

Stage 3. $\pi(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_v^2) = 1$

According to the Bayes' theorem, the joint posterior distribution is given by

$$\begin{aligned}&\pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_v^2, \boldsymbol{\lambda}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_m|\boldsymbol{y}, \boldsymbol{W}_1, \ldots, \boldsymbol{W}_m)\\ &\quad= \pi(\theta_1, \ldots, \theta_m, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_v^2, \boldsymbol{\lambda}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_m|y_1, \ldots, y_m, \boldsymbol{W}_1, \ldots, \boldsymbol{W}_m)\\ &\quad\propto \pi(y_i|\theta_i)\pi(\theta_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_v^2)\pi(\boldsymbol{W}_i|\lambda_i, \boldsymbol{X}_i)\pi(\lambda_i|k_i)\pi(\boldsymbol{\beta})\pi(\boldsymbol{\tau})\pi(\sigma_v^2)\end{aligned} \tag{2}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^T$ is vector of the small area means, $\boldsymbol{y} = (y_1, \ldots, y_m)^T$ is the observations vector, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)^T$ is the vector of latent variables. Due to the conditional structure of the (2) model, the joint posterior distribution is given by

$$\frac{1}{\sigma_v^m} \prod_{i=1}^{m} \Bigg[\exp\left\{ -\frac{(y_i-\theta_i)^2}{2\psi_i} - \frac{\left(\theta_i - \boldsymbol{X}_i^T\boldsymbol{\beta} - \boldsymbol{z}_i^T\boldsymbol{\tau}\right)^2}{2\sigma_v^2} - \frac{(\boldsymbol{W}_i - \boldsymbol{X}_i)^T(\boldsymbol{W}_i - \boldsymbol{X}_i)}{2\lambda_i} \right\} \times \left(\frac{1}{\lambda_i}\right)^{\frac{k_i}{2}+1} e^{\frac{-k_i}{2\lambda_i}} \Bigg] \tag{3}$$

Given the complexity of the expression (3), posterior distributions cannot be obtained in closed-form and Monte Carlo Markov Chain (MCMC) algorithm should be involved to obtain samples from them.

3.1 Computational details

We use the standard MCMC framework to estimate models parameters. To this end, the full conditional distributions for $\boldsymbol{X}_i$, $\boldsymbol{\beta}$, $\boldsymbol{\tau}$, σ_v^2, λ_i and θ_i $(i = 1, \ldots, m)$ are calculated. We assume that $d+q < m$, $\boldsymbol{X} = X_{m\times d} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)^T$ and $\boldsymbol{Z} = Z_{m\times q} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m)^T$ have rank d and q, respectively.

Parameters θ_i, $\boldsymbol{X}_i$, $\boldsymbol{\beta}$, $\boldsymbol{\tau}$, σ_v^2 and λ_i are updated through Gibbs sampling algorithm according to the following full conditional distributions (see Appendix):

1. $\theta_i|\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_v^2, \theta, \boldsymbol{X}, y, \lambda_i, \boldsymbol{W} \sim N\big(\frac{\psi_i^{-1}y_i + \sigma_v^{-2}(\beta\boldsymbol{X}_i^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{\tau})}{\psi_i^{-1} + \sigma_v^{-2}}, (\psi_i^{-1} + \sigma_v^{-2})^{-1}\big)$
2. $\boldsymbol{X}_i|\boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_v^2, \theta, y, \lambda_i, \boldsymbol{W} \sim N\big(\boldsymbol{W}_i + \frac{y_i - \boldsymbol{W}_i^T\boldsymbol{\beta} - \boldsymbol{z}_i^T\boldsymbol{\tau}}{\psi_i + \sigma_v^2 + \boldsymbol{\beta}^T\lambda_i\boldsymbol{\beta}}\lambda_i\boldsymbol{\beta}, \lambda_i - \frac{\lambda_i\boldsymbol{\beta}\boldsymbol{\beta}'\lambda_i}{\psi_i + \sigma_v^2 + \boldsymbol{\beta}'\lambda_i\boldsymbol{\beta}}\big)$
3. $\boldsymbol{\beta}|\boldsymbol{\tau}, \sigma_v^2, \theta, \boldsymbol{X}, y, \lambda_i, \boldsymbol{W} \sim N\big((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\theta - \boldsymbol{Z}\boldsymbol{\tau}), \sigma_v^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\big)$
4. $\boldsymbol{\tau}|\boldsymbol{\beta}, \sigma_v^2, \theta, \boldsymbol{X}, y, \lambda_i, \boldsymbol{W} \sim N\big((\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'(\theta - \boldsymbol{X}\boldsymbol{\beta}), \sigma_v^2(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\big)$
5. $\sigma_v^2|\boldsymbol{\beta}, \boldsymbol{\tau}, \theta, \boldsymbol{X}, y, \lambda_i, \boldsymbol{W} \sim IG\big(\frac{1}{2}(m-2), \frac{1}{2}\Sigma_{i=1}^m(\theta_i - \boldsymbol{X}_i'\boldsymbol{\beta} - z_i'\boldsymbol{\tau})^2\big)$
6. $\lambda_i|rest \sim IG(\frac{1}{2}(k_i+1), \frac{k_i}{2} + \frac{(\boldsymbol{W}_i - \boldsymbol{X}_i)^2}{2})$

4 Parameter estimation in t-MEFH: unknown df

The hierarchical model in this instance is as follows:

Stage 1. $y_i|\theta_i \overset{ind}{\sim} N(\theta_i, \psi_i)$ where ψ_i is known.

$$\boldsymbol{W}_i|\boldsymbol{X}_i, k_i \overset{ind}{\sim} t(\boldsymbol{X}_i, k_i)$$

$$k_i \sim \pi(.)$$

Stage 2. $\theta_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_v^2 \stackrel{ind}{\sim} N(\boldsymbol{X}_i^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{\tau}, \sigma_v^2)$

Stage 3. $\pi(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_v^2) = 1.$

Similar to previous section with respect to the scale mixtures of normals property of t-distribution, we have

Stage 1. $y_i|\theta_i \stackrel{ind}{\sim} N(\theta_i, \psi_i)$

$\boldsymbol{W}_i|\lambda_i, \boldsymbol{X}_i \stackrel{ind}{\sim} N_d(\boldsymbol{X}_i, \lambda_i I)$

$\lambda_i|k_i \stackrel{ind}{\sim} IG(\frac{k_i}{2}, \frac{k_i}{2})$

$\frac{k_i}{2} = s_i \sim E(r_i)$

Stage 2. $\theta_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_v^2 \stackrel{ind}{\sim} N(\boldsymbol{X}_i^T\boldsymbol{\beta} + \boldsymbol{z}_i^T\boldsymbol{\tau}, \sigma_v^2)$

Stage 3. $\pi(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_v^2, r_1, \ldots, r_m) = 1$

Stage 4. $\pi(r_i) \propto IG(a_1, b_1)$

where abbreviation E refer to exponential distribution and we assume that a_1=b_1=1. According to the Bayes' theorem, the joint posterior distribution is given by

$$\begin{aligned}
&\pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_v^2, \boldsymbol{\lambda}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_m, \boldsymbol{S}, \boldsymbol{R}|\boldsymbol{y}, \boldsymbol{W}_1, \ldots, \boldsymbol{W}_m) \\
&\quad = \pi(\theta_1, \ldots, \theta_m, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_v^2, \boldsymbol{\lambda}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_m, \boldsymbol{S}, \boldsymbol{R}|y_1, \ldots, y_m, \boldsymbol{W}_1, \ldots, \boldsymbol{W}_m) \\
&\quad \propto \pi(y_i|\theta_i)\pi(\theta_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\tau}, \sigma_v^2)\pi(\boldsymbol{W}_i|\lambda_i, \boldsymbol{X}_i)\pi(\lambda_i|s_i)\pi(s_i|r_i)\pi(r_i)\pi(\sigma_v^2)\pi(\alpha)\pi(\boldsymbol{\beta})\pi(\boldsymbol{\tau})
\end{aligned} \tag{4}$$

where $\boldsymbol{S} = (s_1, \ldots, s_m)^T$ is the vector of degrees of freedom and $\boldsymbol{R} = (r_1, \ldots, r_m)^T$ is the vector of parameters of s_i for $i = 1, \ldots, m$. Because of the conditional structure of the (4), the joint posterior distribution is given by

$$\begin{aligned}
&\frac{1}{\sigma_v^m}\prod_{i=1}^{m}\Bigg[\left(\frac{1}{\lambda_i}\right)^{\frac{1}{2}}\exp\left\{-\frac{(y_i-\theta_i)^2}{2\psi_i} - \frac{\left(\theta_i - \boldsymbol{X}_i^T\boldsymbol{\beta} - \boldsymbol{z}_i^T\boldsymbol{\tau}\right)^2}{2\sigma_v^2} - \frac{(\boldsymbol{W}_i - \boldsymbol{X}_i)^T(\boldsymbol{W}_i - \boldsymbol{X}_i)}{2\lambda_i}\right\} \\
&\qquad \times \left(\frac{1}{\lambda_i}\right)^{s_i+1} e^{\frac{-s_i}{\lambda_i}} \left(\frac{1}{r_i}\right) e^{\frac{-s_i}{r_i}} \left(\frac{1}{r_i}\right)^{a_1} e^{\frac{-b_1}{r_i}}\Bigg]
\end{aligned} \tag{5}$$

Given the complexity of the expression (5), posterior distributions cannot be obtained in closed-form and Monte Carlo Markov Chain (MCMC) algorithm should be involved to obtain samples from them.

4.1 Computational details

According to assumptions in Subsection 3.1, parameters θ_i, $\boldsymbol{X}_i$, $\boldsymbol{\beta}$, $\boldsymbol{\tau}$, $\boldsymbol{\theta}$, σ_v^2, λ_i and r_i are updated through Gibbs sampling algorithm according to the following full conditional distributions.

1. $\theta_i|\boldsymbol{\beta},\boldsymbol{\tau},\sigma_v^2,\theta,\boldsymbol{X},y,\lambda_i,\boldsymbol{W} \sim N\big(\frac{\psi_i^{-1}y_i+\sigma_v^{-2}(\beta\boldsymbol{X}_i^T\boldsymbol{\beta}+\boldsymbol{z}_i^T\boldsymbol{\tau})}{\psi_i^{-1}+\sigma_v^{-2}},(\psi_i^{-1}+\sigma_v^{-2})^{-1}\big)$
2. $\boldsymbol{X}_i|\boldsymbol{\beta},\boldsymbol{\tau},\sigma_v^2,\theta,y,\lambda_i,\boldsymbol{W} \sim N\big(\boldsymbol{W}_i+\frac{y_i-\boldsymbol{W}_i^T\boldsymbol{\beta}-\boldsymbol{z}_i^T\boldsymbol{\tau}}{\psi_i+\sigma_v^2+\boldsymbol{\beta}^T\lambda_i\boldsymbol{\beta}}\lambda_i\boldsymbol{\beta},\lambda_i-\frac{\lambda_i\boldsymbol{\beta}\boldsymbol{\beta}'\lambda_i}{\psi_i+\sigma_v^2+\boldsymbol{\beta}'\lambda_i\boldsymbol{\beta}}\big)$
3. $\boldsymbol{\beta}|\boldsymbol{\tau},\sigma_v^2,\theta,\boldsymbol{X},y,\lambda_i,\boldsymbol{W} \sim N\big((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\theta-\boldsymbol{Z}\boldsymbol{\tau}),\sigma_v^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\big)$
4. $\boldsymbol{\tau}|\boldsymbol{\beta},\sigma_v^2,\theta,\boldsymbol{X},y,\lambda_i,\boldsymbol{W} \sim N\big((\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'(\theta-\boldsymbol{X}\boldsymbol{\beta}),\sigma_v^2(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\big)$
5. $\sigma_v^2|\boldsymbol{\beta},\boldsymbol{\tau},\theta,\boldsymbol{X},y,\lambda_i,\boldsymbol{W} \sim IG\big(\frac{1}{2}(m-2),\frac{1}{2}\Sigma_{i=1}^m(\theta_i-\boldsymbol{X}_i'\boldsymbol{\beta}-z_i'\boldsymbol{\tau})^2\big)$
6. $\lambda_i|rest \sim IG(s_i+0.5,s_i+\frac{(\boldsymbol{W}_i-\beta\boldsymbol{X}_i)^T(\boldsymbol{W}_i-\beta\boldsymbol{X}_i)}{2})$
7. $\pi(s_i|rest) \propto \left(\frac{1}{\lambda_i}\right)^{s_i+1} e^{-s_i(\frac{1}{\lambda_i}+\frac{1}{r_i})}$
8. $r_i|rest \sim IG(a_1+1,b_1+s_i)$

4.2 Updating s_i

Since $\pi(s_i|rest) \propto \left(\frac{1}{\lambda_i}\right)^{s_i} e^{-s_i(\frac{1}{\lambda_i}+\frac{1}{r_i})}$ does not has a closed-form, following Zarei *et al.* (2021) the s_i is updated using the Metropolis-Hastings algorithm. For each s_i, we choose the uniform distribution centred on the current state of the chain, denoted by *cand*, as proposal distribution, i.e., $q(\lambda_i|\alpha) \sim U(cand-0.3,cand+0.3)$. Since the uniform distribution is symmetric, the acceptance probability in iteration t and for each s_i is

$$\min\big\{1,\frac{\pi({s_i}^{new})\pi(\lambda_i|{s_i}^{new})}{\pi({s_i}^{(t)})\pi(\lambda_i|{s_i}^{(t)})}\big\} \tag{6}$$

5 Simulations

In this section, we investigate the performance of the t-MEFH model in estimating small area means, in controlled simulated settings.

Our data generating setting is similar to Ybarra and Lohr (2008) and Zarei *et al.* (2021). We generate X_i from a $N(5,3^2)$. At each iteration, $\theta_i=1+3x_i+v_i$, $y_i=\theta_i+e_i$ and $W_i=X_i+\eta_i$, where v_i, e_i and η_i are assumed to be independent. Sampling errors e_i simulated from $N(0,\psi_i)$ where ψ_i generated from a gamma distribution with shape 4.5 and scale 2 and $v_i \sim N(0,2^2)$. Similar to Bell and Huang (2006), we consider four simulation scenarios according to different generation scheme of ME, namely:

Scenario 1: $\eta_i \sim t_\infty$ (Standard Gaussian Scenario thus $\sigma_{ME}^2=1$);

Scenario 2: $\eta_i \sim t_8$ (thus $\sigma^2_{ME} = 1.33$);

Scenario 3: $\eta_i \sim t_5$ (thus $\sigma^2_{ME} = 1.67$);

Scenario 4: $\eta_i \sim t_3$ (thus $\sigma^2_{ME} = 3$);

where t_{df} and σ^2_{ME} denote the t-distribution with df degrees of freedom and the measurement error variance, respectively. We assume that degrees of freedom k_i is the same in all small areas. The aforementioned simulation scheme is repeated for $m = 10$ and $m = 50$.

5.1 Performance of t-MEFH in estimating small area means

We compare the estimated small area means θ_i obtained with the following models:

- the proposed model with known df, $\hat{\theta}_{i,TMEK}$;
- the proposed model with unknown df, $\hat{\theta}_{i,TMEU}$;
- a fully Bayesian version of the Fay-Herriot model involving the true covariates, $\hat{\theta}_{i,FHT}$;
- a fully Bayesian version of the Fay-Herriot model not accounting for the measurement error in covariates, $\hat{\theta}_{i,FH}$;
- the Bayesian measurement error model in Arima *et al.* (2015) $\hat{\theta}_{i,ADL}$;
- the EBLUP estimators proposed in Ybarra and Lohr (2008) $\hat{\theta}_{i,YL}$.

Performance in estimating small area means are studied according to the following deviance measures: average absolute deviation (AAD) and empirical mean square error ($EMSE$) or average squared deviation (ASD), defined as follows

$$AAD = \frac{1}{m}\sum_{i=1}^{m}|\hat{\theta}_i - \theta_i|, \qquad ASD = \frac{1}{m}\sum_{i=1}^{m}(\hat{\theta}_i - \theta_i)^2.$$

We also consider the improve percentage (IP) of each method to compare results with direct estimator. This quantity is defined as

$$IP = 100 * \frac{ASD.direct - ASD}{ASD.direct}$$

where

$$ASD.direct = \frac{1}{m}\sum_{i=1}^{m}(y_i - \theta_i)^2.$$

IP is a measure of how much better our revised method is compared to the direct estimator. Therefore, a higher improve percentage indicates a greater improvement in performance, while a lower percentage indicates a smaller improvement. In real data applications, when the true value of θ is unknown, we rely to the deviance information criterion (DIC; Spiegelhalter *et al.* (2002)) for evaluation of goodness of fit of Bayesian approaches and choose the model with smallest DIC as the best performing one. Tables 1 and 2 show the results of simulations in Scenarios 1 to 4, averaged over the 100 datasets, respectively when $m = 10$ and $m = 50$.

Each column in the table corresponds to a simulation Scenario: the first row, corresponding to the Fay-Herriot model involving the true covariates, should be considered as a reference model. As expected, all ME models perform very similarly in the first scenario since the measurement error is normally distributed. On the other hand, decreasing the degrees of freedom, all performance indices show that the proposed approach performs better than the other competing models. The improvement of our proposal is more evident when the number of areas is low, e.g. $m = 10$.

Table 1: Comparison of different models performances when data are simulated under different Scenarios and $m = 10$. Bold numbers highlight the best performance for each setting.

		Simulation scenarios ($m = 10$)						Simulation scenarios ($m = 10$)			
Criteria	Estimate	t_∞	t_8	t_5	t_3	Criteria	Estimate	t_∞	t_8	t_5	t_3
	$\hat{\theta}_{FHT}$	1.708	1.619	1.523	1.774		$\hat{\theta}_{FHT}$	22.648	17.995	24.642	18.241
	$\hat{\theta}_{FH}$	2.180	2.104	1.998	2.253		$\hat{\theta}_{FH}$	9.215	2.878	5.191	0.362
EMSE	$\hat{\theta}_{TMEK}$	**2.071**	**1.979**	**1.975**	**2.118**	IP	$\hat{\theta}_{TMEK}$	**9.746**	**5.799**	**7.414**	1.328
	$\hat{\theta}_{ADL}$	2.131	2.000	1.993	2.128		$\hat{\theta}_{ADL}$	9.039	5.259	7.154	**1.391**
	$\hat{\theta}_{YL}$	3.296	3.065	2.176	2.214		$\hat{\theta}_{YL}$	1.662	2.125	2.601	0.467
	$\hat{\theta}_{FHT}$	1.018	0.992	0.975	1.012		$\hat{\theta}_{FHT}$	43.004	42.806	42.972	42.680
	$\hat{\theta}_{FH}$	1.113	1.097	1.098	1.138		$\hat{\theta}_{FH}$	44.426	44.327	44.687	45.478
ADD	$\hat{\theta}_{TMEK}$	**1.111**	**1.083**	**1.082**	1.131	DIC	$\hat{\theta}_{TMEK}$	44.312	**44.242**	**44.673**	**44.559**
	$\hat{\theta}_{ADL}$	1.123	1.084	1.100	**1.130**		$\hat{\theta}_{ADL}$	**44.308**	44.262	44.691	44.575
	$\hat{\theta}_{YL}$	1.218	1.371	1.144	1.159		$\hat{\theta}_{YL}$	-	-	-	-

Table 2: Comparison of different models performances when data are simulated under different Scenarios and $m = 50$. Bold numbers highlight the best performance for each setting.

		Simulation scenarios ($m = 50$)						Simulation scenarios ($m = 50$)			
Criteria	Estimate	t_∞	t_8	t_5	t_3	Criteria	Estimate	t_∞	t_8	t_5	t_3
	$\hat{\theta}_{FHT}$	1.512	1.398	1.358	1.462		$\hat{\theta}_{FHT}$	30.557	34.850	33.394	31.952
	$\hat{\theta}_{FH}$	1.848	1.907	2.210	1.890		$\hat{\theta}_{FH}$	15.911	12.508	10.667	11.931
EMSE	$\hat{\theta}_{TMEK}$	**1.839**	1.878	**1.991**	**1.841**	IP	$\hat{\theta}_{TMEK}$	**16.377**	12.885	**11.451**	**12.952**
	$\hat{\theta}_{ADL}$	1.855	**1.869**	1.997	1.889		$\hat{\theta}_{ADL}$	15.470	**13.013**	11.144	10.770
	$\hat{\theta}_{YL}$	1.840	1.911	2.049	1.890		$\hat{\theta}_{YL}$	16.225	12.670	8.453	10.824
	$\hat{\theta}_{FHT}$	0.975	0.932	0.945	0.952		$\hat{\theta}_{FHT}$	208.546	211.611	211.290	210.913
	$\hat{\theta}_{FH}$	1.073	1.095	1.206	1.099		$\hat{\theta}_{FH}$	219.786	222.772	222.835	223.416
ADD	$\hat{\theta}_{TMEK}$	1.074	**1.069**	**1.104**	**1.078**	DIC	$\hat{\theta}_{TMEK}$	**219.187**	222.325	**222.136**	**222.444**
	$\hat{\theta}_{ADL}$	1.078	1.074	1.107	1.098		$\hat{\theta}_{ADL}$	220.105	**222.303**	222.262	222.583
	$\hat{\theta}_{YL}$	**1.072**	1.076	1.116	1.093		$\hat{\theta}_{YL}$	-	-	-	-

Notice that all models, with the exception of the proposed model, require the specification of the true value of measurement error variance. On the other hand, the proposed model requires the specification of the degrees of freedom. Both specifications can be hard in real data analysis. In order to investigate the robustness of all proposed model with respect to model misspecification, we simulate data according to the simulation scheme described above in which $\eta_i \sim t_3$: the proposed model is then estimated by fixing a different number of degrees of freedom, namely 6 and 10; the other models are estimated by fixing σ^2_{ME} equal to a wrong value, namely, 1.5 and 1.25, respectively.

Table 3 reports the performance of the competing models to model misspecification. Each column shows the estimates obtained with different models: in particular, the first column shows the estimates obtained when we fit the proposed approach modelling the measurement error as a t-distribution with 6 degrees of freedom and the competing models fixing the variance of the measurement error equal to 1.5. Table 3 highlights that the proposed t-MEFH model is more robust rather than the competitive ME models and it has the best overall performance under model misspecification. As expected, since the true df is 3, the proposal has better performance for df=6 than df=10.

Table 3: Scenario 5: Comparison of different ME models performances when data are simulated under $\eta_i \sim t_3$, $m = 10$ and $m = 50$ with wrong estimated of σ^2_{ME} and df (namely 6 and 10). Bold numbers highlight the best performance for each setting.

		Simulation scenario 5: $\eta_i \sim t_3$			
Criteria	Estimate	$m = 10$	$m = 50$	$m = 10$	$m = 50$
		(t_6; $\sigma^2_{ME} = 1.5$)	(t_{10}; $\sigma^2_{ME} = 1.25$)	(t_6; $\sigma^2_{ME} = 1.5$)	(t_{10}; $\sigma^2_{ME} = 1.25$)
EMSE	$\hat{\theta}_{TMEK}$	**1.953**	**1.914**	**2.057**	2.179
	$\hat{\theta}_{ADL}$	1.985	1.912	2.084	**2.177**
	$\hat{\theta}_{YL}$	2.262	1.922	2.145	**2.177**
ADD	$\hat{\theta}_{TMEK}$	**1.078**	1.084	**1.105**	1.155
	$\hat{\theta}_{ADL}$	1.085	1.086	**1.105**	**1.150**
	$\hat{\theta}_{YL}$	1.137	**1.083**	1.114	1.154
IP	$\hat{\theta}_{TMEK}$	**5.884**	**8.561**	8.005	8.086
	$\hat{\theta}_{ADL}$	4.767	8.470	**8.150**	**8.450**
	$\hat{\theta}_{YL}$	-2.785	8.461	3.345	7.876
DIC	$\hat{\theta}_{TMEK}$	44.594	**222.755**	44.818	222.811
	$\hat{\theta}_{ADL}$	**44.552**	222.758	**44.781**	**222.781**
	$\hat{\theta}_{YL}$				

To examine the proposal's robustness further, we simulate data according to the simulation scheme described above and simulate the measurement error according to a Pareto distribution $\eta_i \sim \text{Pareto}(5, 3)$. Such a distribution is particularly skewed and it presents several outlying observations making it not suitable for all aforementioned models. The proposed model (with known df) is estimated by fixing the degrees of freedom equal to $3, 6$ and 10; the variance of the measurement error σ^2_{ME} is fixed equal to variance of Pareto$(5, 3)$ i.e., 18.75 (scenario 6).

Table 4: Scenario 6: $m = 50$, $\eta_i \sim$ Pareto$(5, 3)$. Bold and blue numbers highlight the first and second best performance, for each setting after $\hat{\theta}_{FHT}$, respectively. The numbers in parentheses are df of the t-MEFH model.

Criteria	Method	Estimate	Criteria	Method	Estimate
	$\hat{\theta}_{FHT}$	1.258		$\hat{\theta}_{FHT}$	37.167
	$\hat{\theta}_{FH}$	1.942		$\hat{\theta}_{FH}$	5.914
	$\hat{\theta}_{TMEU}$	**1.913**		$\hat{\theta}_{TMEU}$	6.445
	$\hat{\theta}^{(3)}_{TMEK}$	2.063		$\hat{\theta}^{(3)}_{TMEK}$	5.201
EMSE	$\hat{\theta}^{(6)}_{TMEK}$	1.989	IP	$\hat{\theta}^{(6)}_{TMEK}$	**6.693**
	$\hat{\theta}^{(10)}_{TMEK}$	2.064		$\hat{\theta}^{(10)}_{TMEK}$	5.201
	$\hat{\theta}_{ADL}$	1.917		$\hat{\theta}_{ADL}$	6.425
	$\hat{\theta}_{YL}$	2.024		$\hat{\theta}_{YL}$	1.329
	$\hat{\theta}_{FHT}$	0.898		$\hat{\theta}_{FHT}$	210.782
	$\hat{\theta}_{FH}$	1.087		$\hat{\theta}_{FH}$	224.413
	$\hat{\theta}_{TMEU}$	1.084		$\hat{\theta}_{TMEU}$	**223.949**
	$\hat{\theta}^{(3)}_{TMEK}$	1.139		$\hat{\theta}^{(3)}_{TMEK}$	224.019
ADD	$\hat{\theta}^{(6)}_{TMEK}$	**1.079**	DIC	$\hat{\theta}^{(6)}_{TMEK}$	223.966
	$\hat{\theta}^{(10)}_{TMEK}$	1.138		$\hat{\theta}^{(10)}_{TMEK}$	224.178
	$\hat{\theta}_{ADL}$	1.087		$\hat{\theta}_{ADL}$	224.812
	$\hat{\theta}_{YL}$	1.108		$\hat{\theta}_{YL}$	-

Table 4 confirms that the proposed t-MEFH methods (with and without known df) are robust under outlier ME in covariates with respect to model misspecification and they have the best overall performance among the ME models. These results are quite expected since when data are generated from a Pareto distribution, the t-distribution is more robust in accommodating outlying observations. Furthermore, this simulation shows that determining correct df is very important. The estimated value of df, when we assume that df is known, with trail-error method is df $= 7$ and for unknown df case is df $= 7.912$.

To further study of the robustness of the proposal and also comparing between the t-MEFH models (with and without known df) as seventh Scenario, we simulate data according to the simulation scheme described above and for $m \in \{20, 50, 100, 200\}$ where the measurement error has t-distribution with df=3. The results are presented in Table 5.

Table 5 shows under this Scenario as expected the proposal methods generally has the best performance. For $m = 20$, 50 and 200 the $\hat{\theta}_{TMEK}$ is better than $\hat{\theta}_{TMEU}$ and for $m = 100$ vice versa. These results show that the proposal model with unknown df is valuable tool for estimating small area parameters when covariates are measured with outlier MEs. However, estimating the degrees of freedom introduces some error compared to knowing the exact degrees of freedom. Nevertheless, it still performs better than other measurement error methods and does not impose the restrictive assumption of known degrees of freedom.

The simulation study highlights that the proposed approach is competitive with existing approaches resulting more robust in several Scenarios. We should stress that, as shown in Scenario 5, the misspecification of the degrees of freedom in the proposed model as well as the misspecification of the variance of the measurement error in the competing models have important effects on the estimates. However, we acknowledge that in real data analysis we can find the optimal number of degrees of freedom for $\hat{\theta}_{TMEK}$ using standard model choice criteria, such as DIC by trail and error method.

Table 5: Scenario 7: Comparison of different ME models performances when data are simulated under $\eta_i \sim t_3$ for $m = 20,\ 50,\ 100,\ 200$. Bold numbers highlight the best performance for each setting.

		m						m			
Criteria	Estimate	20	50	100	200	Criteria	Estimate	20	50	100	200
	$\hat{\theta}_{TMEU}$	2.366	1.822	**1.857**	1.914		$\hat{\theta}_{TMEU}$	11.003	18.175	**15.485**	13.359
	$\hat{\theta}_{TMEU}$	2.366	1.822	**1.857**	1.914		$\hat{\theta}_{TMEU}$	11.003	18.175	**15.485**	13.359
EMSE	$\hat{\theta}_{TMEK}$	**2.365**	**1.807**	1.863	**1.901**	IP	$\hat{\theta}_{TMEK}$	11.846	**18.826**	15.140	**13.946**
	$\hat{\theta}_{ADL}$	2.366	1.912	1.892	1.935		$\hat{\theta}_{ADL}$	**12.893**	13.877	13.755	12.866
	$\hat{\theta}_{YL}$	2.716	1.930	2.002	2.013		$\hat{\theta}_{YL}$	2.232	13.422	8.555	8.776
	$\hat{\theta}_{TMEU}$	1.143	**1.049**	**1.032**	1.040		$\hat{\theta}_{TMEU}$	**89.369**	221.786	**444.309**	**890.094**
ADD	$\hat{\theta}_{TMEK}$	1.128	1.051	1.035	**1.038**	DIC	$\hat{\theta}_{TMEK}$	89.490	**220.420**	444.917	890.126
	$\hat{\theta}_{ADL}$	**1.125**	1.081	1.041	1.047		$\hat{\theta}_{ADL}$	90.026	222.239	445.330	890.719
	$\hat{\theta}_{YL}$	1.229	1.085	1.064	1.057		$\hat{\theta}_{YL}$	-	-	-	

6 Real data analysis

6.1 The corn and soybean data

In this subsection, we analyse the well known county crop areas data (also known as the corn and soybean data), first analysed by Battese *et al.* (1998). Corn and soybeans areas have been obtained in 37 sample segments from 12 Iowa counties (small areas) by interviewing farm operators. From the analysis of these data in Battese *et al.* (1998), emerged that data corresponding to the Hardin county can be considered outliers, and as a consequence, the authors removed them from the analysis. However, outliers can be a good source of information; furthermore, removing data from any analysis leads to loss of possibly, valuable information about part of the non-sampled units of the population. Similar to Sinha and Rao (2009) and Chakraborty *et al.* (2017), for illustrative purposes, we analyse the full data set for corn. This dataset comprises the number of segments in each county, the reported number of hectares of corn for each sampled segment (direct estimate), and the number of pixels classified by the LANDSAT satellite as corn for each sampled segment (covariate). Furthermore, for corn and soybean data both the sample number of pixels (ME covariate) and the population number of pixels (true covariate) are available. So, we can easily calculate the variance of ME. The original data is unit-level data. To have data in the area-level frame, the sample mean is calculated for each county. The means of the number of pixels of a given crop per segment in sample and population are used as a ME and true auxiliary data, respectively.

We estimate ψ_is with the sample variances S_i^2 divided by the sample sizes n_i, i.e., $\frac{S_i^2}{n_i}$. When computing the sample variances for the first four areas, there is a problem because $n_i < 3$. Following You and Chapman (2006) we limit the number of areas to 8 by using only the counties with a sample size greater than 2. The crop hectares for corn in each counties, i.e., y_i were modelled as a function of the auxiliary data, i.e., x_i for those counties in the form

$$y_i = \beta_0 + \beta_1 x_i + v_i + e_i, \quad i = 1, \ldots, 8,$$

where v_i are assumed to be i.i.d. $N(0, \sigma_v^2)$. The goal is to estimate the small area mean hectares of corn per segment in the ith county. Since the true values of small area means are unknown and for $m = 10$ the Bayesian methods have the best performance, we focus only on Bayesian methods. And DIC is used to compare different models.

In Table 6 we present the values of DIC for all methods. Since for $\hat{\theta}_{TMEK}$, we assume that the degrees of freedom are known: in practice, we fit models assuming $df = 3, ..., 10$ and select the best model according to DIC. Here, the best model is the one in which the degrees of freedom are set equal to 5. Furthermore, the estimated df for $\hat{\theta}_{TMEU}$ is 6.779. Table 6 shows that the proposed models has the best performance when estimating the number of hectares of corn for each sample. Table 7 presents the predicted mean hectares of corn per segment.

Table 6: Values of DIC for different ME models. Bold number highlight the best performance for each scenario.

methods:	$\hat{\theta}_{TMEU}$	$\hat{\theta}_{TMEK}$	$\hat{\theta}_{FHT}$	$\hat{\theta}_{FH}$	$\hat{\theta}_{ADL}$
Estimated DIC :	**74.902**	74.997	76.144	75.024	75.728

6.2 The cost and income of urban households in Kurdistan province of Iran

In this subsection to apply the performance of the proposed model, we analyze Iranian urban household income and expenditure survey data (IUHIE)[1] in the year 2020 for Kurdistan province. The IUHIE general aim is to estimate the average income and expenditure for urban households at provincial and country levels. This survey is done

[1]Available at www.amar.org.ir/english/Statistics-by-Topic/Household-Expenditure-and-Income.

Table 7: Predicted mean hectares of corn per segment in different ME methods.

		Estimated hectares				
County	Sample segments	$\hat{\theta}_{TMEU}$	$\hat{\theta}_{TMEK}$	$\hat{\theta}_{FHT}$	$\hat{\theta}_{FH}$	$\hat{\theta}_{ADL}$
Franklin	3	157.48	157.47	157.40	157.44	157.57
Pocahontas	3	99.72	87.20	105.11	89.91	106.62
Winnebago	3	117.54	116.28	115.59	116.35	116.53
Wright	3	142.63	152.22	129.46	149.21	137.88
Webster	4	114.16	111.28	113.98	112.00	117.53
Hancock	5	112.36	114.04	114.54	111.28	111.89
Kossuth	5	114.79	116.08	112.25	115.72	112.63
Hardin	6	117.18	117..89	124.43	116.40	118.25

monthly. In this survey, the sample size is optimized at the level of the provinces. Therefore, the cities of each province can be small area. Kurdistan province is one of the 31 provinces of Iran and has $m = 10$ cities or counties.

For illustration purposes, we examine households whose head is self-employed and their income is registered with a self-declaration. Since, people usually record and express their income and expenditure more or less than the reality, this can indicate the possibility of measurement error in covariate. We choose the logarithm (in base 10) of household housing cost in each city as response variable ($y_i,\ i = 1, \dots, 10$) and available auxiliary information for this study are the logarithm (in base 10) of the household self-employed income. In Figures 1 and 2, the box plots of these variables are drawn. As it is clear from Figure 1 we can see some outliers in the data.

Table 8: Calculation of DIC for different degrees of freedom of the $\hat{\theta}_{TMEK}$ method, which shows how to choose the appropriate degree of freedom.

		df				
Criteria	Estimate	3	4	5	6	7
DIC	*TD*	−29.65	−30.83	−29.79	−31.72	−24.23

To use the t-MEFH model with known df, first by calculating the DIC index for different degrees of freedom, we determine the optimal degree of freedom by trial and error. Table 8 shows the calculated values for this parameter. According to this, the proposed method $\hat{\theta}_{TMEK}$, in the case of df = 6 has its best value of DIC which is equal to -31.72. For the t-MEFH model with unknown df the estimated df is 5.1 and the value of DIC for this model is -30.91. Comparing the DIC value of the $\hat{\theta}_{TMEK}$ method with df 6 with the DIC values of the $\hat{\theta}_{TMEU}$, $\hat{\theta}_{FHT}$ and $\hat{\theta}_{ADL}$ methods shows that the proposed method with known df is the best and $\hat{\theta}_{TMEU}$ has the second rank. The value of DIC for $\hat{\theta}_{FHT}$ and $\hat{\theta}_{ADL}$ is -5.30 and -4.30, respectively. It should be noted that we estimate ψ_is similar to explained methodology for the corn data.

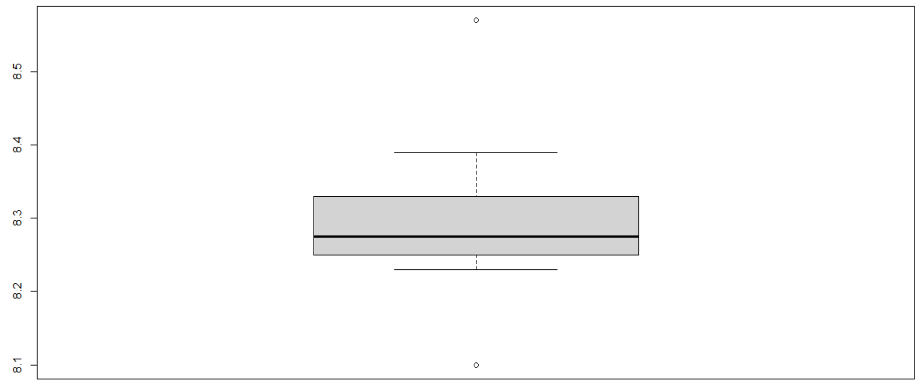

Figure 1: Box plot of the logarithm (in base 10) of households with self-employed head income (covariate) for the cities of Kurdistan province. This graph shows the presence of two outliers in the data. Therefore, there is a possibility of existing of outlier measurement error.

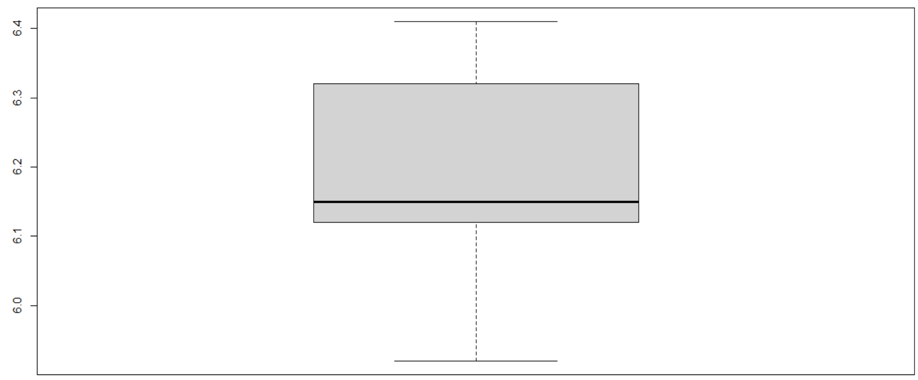

Figure 2: Box plot of the logarithm (in base 10) of housing costs of households with self-employed head (response variable). This plot shows the approximate normal distribution of the logarithm of household housing cost.

Table 9: The average logarithm of housing cost in the cities of Kurdistan province.

County	$\hat{\theta}_{TMEK}$	$\hat{\theta}_{TMEU}$	$\hat{\theta}_{FH}$	$\hat{\theta}_{ADL}$	County	$\hat{\theta}_{TMEK}$	$\hat{\theta}_{TMEU}$	$\hat{\theta}_{FH}$	$\hat{\theta}_{ADL}$
Baneh	6.131	6.089	6.131	6.130	Bijar	6.293	6.291	6.297	6.293
Saghez	6.151	6.154	6.150	6.149	Sanandaj	6.326	6.331	6.330	6.330
Ghorveh	6.152	6.152	6.151	6.150	Marivan	6.121	6.124	6.121	6.122
Divandareh	6.313	6.315	6.310	6.310	Kamyaran	6.399	6.401	6.401	6.402
Sarvabad	5.950	5.951	5.960	6.954	Dehgolan	6.024	6.019	6.011	6.011

In Table 9, the estimated averages of the cost of housing of households with self-employed head for the cities of Kurdistan province as small areas are given. According to obtained results for $\hat{\theta}_{TMEK}$, The most expensive and cheapest cities based on housing costs are Kamyaran and Sarvabad. Furthermore, the city of Sanandaj, which is the center of this province, has the second most expensive housing cost. This means that, on average, each household with self-employed head in Sanandaj spends 2118361 Tomans (the common currency of Iran) per month for housing cost.

7 Conclusion

In this paper, we propose a Bayesian small area mode accounting for the presence of measurement error in covariates. In particular, we propose to model the measurement error as t-distribution with known and unknown degrees of freedom. Simulation studies show that the proposed approaches are an evaluable alternative to the models existing in literature both in terms of predictions as well as in terms of overall fit. One of open problem in ME literature is estimating variance of ME. Because of identifiable problem, often is assumed that this variance is known. We assumed that df is unknown (corresponding to unknown variance). Convergence of our algorithms show that there is no problem related to model identifiable. As previously mentioned, if the degrees of freedom are known in the presence of outlier measurement error, model θ_{TMEK} is more accurate than model θ_{TMEU}. However, knowing the degrees of freedom is a restrictive assumption and is not valid in practice. On the other hand, in the presence of measurement error in the data, model θ_{TMEU} performs better than other available models in controlling measurement error.

We also apply the proposed model to two real data sets the corn and soybean: according to the DIC measure our modeles are better than competitive model even when we use the number of pixels classified by the LANDSAT satellite as corn for each sampled segment obtained by the census from the population. Furthermore, analysing urban household income and expenditure data of Kurdistans' cities show that our modeled are useful to model data under doubt existence of ME. As a future work, one can develop our methodology for situation that MEs have skew t-distribution. All the algorithms have been implemented in the R software and the codes are available upon request from the authors.

Appendix

Comparing (2) with equation (6) in Arima *et al.* (2015), for computing the full conditional distributions, except λ_i, it is only necessary we replace $\boldsymbol{C}_i$ with $\lambda_i I$. So we compute only the full conditional distribution for λ_i.

Posterior distribution of λ_i when $d = 1$: According to the Bayes's rule and Equation (2), since posterior is proportional to the product of the likelihood and prior and using elements that only include λ_i. Therefore,

$$\begin{aligned} \pi(\lambda_i|rest) &\propto \pi(\lambda_i)\pi(W_i|X_i,\lambda_i) \\ &\propto (\frac{1}{\lambda_i})^{a+1} e^{\frac{-b}{\lambda_i}} (\frac{1}{\lambda_i})^{\frac{1}{2}} e^{\frac{-(W_i-X_i)^2}{2\lambda_i}} = (\frac{1}{\lambda_i})^{\frac{1}{2}+a+1} e^{\frac{-1}{\lambda_i}(b+\frac{(W_i-X_i)^2}{2})} \end{aligned}$$

therefore for $a = \frac{k_i}{2}$ and $b = \frac{k_i}{2}$ we have:

$$\lambda_i|rest \sim IG\big(\frac{k_i+1}{2}, \ \frac{k_i}{2} + \frac{(W_i - X_i)^2}{2}\big)$$

References

Arima, S., Datta, G.S. and Liseo, B. (2015). Bayesian estimators for small area models when auxiliary information is measured with error. *Scandinavian Journal of Statistics*, **42**(2), 518–529.

Akinc, D. and Vandebroek, M. (2018). Bayesian estimation of mixed logit models: Selecting an appropriate prior for the covariance matrix. *Journal of choice modelling*, **29**, 133–151.

Battese, G.E., Harter, R.M. and Fuller, W.A. (1988). An error components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, **83**, 28–36.

Bell, W.R. and Huang, E.T. (2006). Using the t-distribution to deal with outliers in small area estimation. *In Proceedings of Statistics Canada Symposium.*

Chakraborty, A., Datta, G.S. and Mandal, A. (1988). Robust hierarchical Bayes small area estimation for nested error regression model. *arXiv preprint arXiv:1702.05832.*

Fay, M.P. and Herriot, R.A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, **74**(366a), 269–277.

Ghosh, M., Myung, J. and Moura, F.A. (2018). Robust Bayesian small area estimation. *Survey Methodology*, **44**(1), 101–116.

Goo, Y.M. and Kim, D.H. (2013). Bayesian small area estimations with measurement errors. *Journal of the Korean Data & Information Science Society*, **24**(4), 885–895.

Hariyanto, S., Notodiputro, K. A., Kurnia, A. and Sadik (2020). Small Area Estimation with Measurement Error in t Distributed Covariate Variable. *International Journal of Advanced Science and Technology*, **10**(4), 1536–1542.

Molina, I. and J.C. Rao (2015). *Small Area Estimation.* John Wiley & Sons.

Sinha, S.K. and Rao, J.N.K. (2009). Robust small area estimation. *Canadian Journal of Statistics*, **37**(3), 381–399.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**(4), 583–639.

You, Y. and Chapman, B. (2006). Small area estimation using area-level models and estimated sampling variances. *Survey Methodology*, **1**(32), 97–103.

Ybarra, L.M. and Lohr, S.L. (2008). Small area estimation when auxiliary information is measured with error. *Biometrika*, **95**(4), 919–931.

Zarei, S. and Arima, S. and Giovanna, J.L. (2021). A new robust Bayesian small area estimation via α-stable model for estimating the proportion of athletic students in California. *Biometrical Journal*, **63**(6), 1309–1324.