Seemingly Unrelated Regression with Measurement Error: Estimation via Markov chain Monte Carlo and Mean Field Variational Bayes Approximation

Georges Bresson

Department of Economics, Université Paris II, France

Anoop Chaturvedi

Department of Statistics, University of Allahabad, India.

Mohammad Arshad Rahman*

Department of Economic Sciences, Indian Institute of Technology Kanpur, India.

Shalabh

Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, India.

Abstract

Linear regression with measurement error in the covariates is a heavily studied topic, however, the statistics/econometrics literature is almost silent to estimating a multi-equation model with measurement error. This paper considers a seemingly unrelated regression model with measurement error in the covariates and introduces two novel estimation methods: a pure Bayesian algorithm (based on Markov chain Monte Carlo techniques) and its mean field variational Bayes (MFVB) approximation. The MFVB method has the added advantage of being computationally fast and can handle big data. An issue pertinent to measurement error models is parameter identification, and this is resolved by employing a prior distribution on the measurement error variance. The methods are shown to perform well in multiple simulation studies, where we analyze the impact on posterior estimates for different values of reliability ratio or variance of the true unobserved quantity used in the data generating process. The paper further implements the proposed algorithms in an application drawn from the health literature and shows that modeling measurement error in the data can improve model fitting.

Keywords: Classical measurement error, Markov chain Monte Carlo (MCMC), mean field variational Bayes, reliability ratio, seemingly unrelated regression, systolic blood pressure.

1. Introduction

The seemingly unrelated regression (SUR) model consists of a system of linear multiple regression equations such that each equation has a different continuous dependent variable with a potentially different set of exogenous explanatory variables (covariates) and the errors are correlated across equations (Zellner, 1962). When the conditions of the SUR model apply, estimators obtained from SUR are more efficient relative to ordinary least squares estimators. The optimality feature and other theoretical properties of the SUR estimator within the frequentist framework are well studied in Srivastava and Dwivedi (1979), Srivastava and Giles (1987) and Fiebig (2001). The Bayesian approach to estimating SUR model was introduced in Zellner (1971), where the author analytically derived the conditional posterior densities of the parameters. Given the conditional posteriors, the model can then be estimated using a Markov chain Monte Carlo (MCMC) technique, known as Gibbs sampling (Geman and Geman, 1984; Casella and George, 1992). Since the introduction in Zellner (1971), the literature on Bayesian analysis of SUR has grown considerably in various directions, including estimation *via* MCMC (Percy, 1992; Griffiths and Chotikapanich, 1997; Griffiths and Valenzuela, 2006) and direct Monte Carlo approach (Zellner and Ando, 2010; Ando and Zellner, 2010), prediction in SUR model (Percy, 1992) and several model extensions that include restricted SUR (Steel, 1992), SUR with serially correlated errors and time varying parameters (Chib and Greenberg, 1995) and semiparametric inference in SUR model (Koop et al., 2005).

The existing literature on SUR models including the quoted articles have worked based on the assumption that the covariates are measured correctly. Nonetheless, in practice there can emerge situations where one or more of the covariates are recorded with error, thus giving rise to SUR with measurement error (hereafter SURME). Modeling measurement error within a SUR structure or more generally in a multi-equation system has largely gone unnoticed in the literature (both frequentist and Bayesian), the only exception is Carroll et al. (2006a,b) explained in the next paragraph. In contrast, there has been considerable work on single equation models with measurement error. Within a linear regression framework, it is well known that measurement error in the data leads to bias and inconsistency in ordinary least squares (OLS) estimator (see for instance Cheng and Van Ness (1999), Fuller (1987), Wansbeek and Meijer (2000), Rao et al. (2008) and Hu and Wansbeek (2017)). To achieve consistency of OLS estimator, side assumptions are required such as known measurement error variance or known *reliability ratio*.[2] However, consistent estimator of regression parameters without the side assumptions can be constructed when measurement errors have replicated observations (Shalabh, 2003). Measurement error in nonlinear models is discussed in Carroll et al. (2006b) along with the Bayesian analysis of linear and non-linear measurement error models.

Within the multi-equation framework, Carroll et al. (2006b) consider a combination of linear mixed measurement error model and SUR model to understand the properties of measurement error in food frequency questionnaire data for protein and energy. They adopt the frequentist estimation approach and use a nearby adaptive method based on weighted Akaike information criterion (AIC) to select the best fitting model, a form of model averaging which is popular in the Bayesian literature. Carroll et al. (2006b) find that a fully parameterized model in which measurement errors in the two nutrients are modeled jointly, offers no gain in efficiency compared

---

[2]If w and z are two random variables such that $w = z + u$ and the error u is independent of z, then the reliability ratio R_z is defined as the true variance divided by the total variance, *i.e.*, $R_z = Var(z)/(Var(z) + Var(u))$. By definition $0 \leq R_z \leq 1$.

to fitting each model separately. However, when some parameters are set to zero resulting in a reduced model, considerable gains in efficiency is attained. We may adopt the frequentist approach to estimating SURME model with *structural* measurement error, but it is fraught with difficulty because the number of parameters become larger than the number of normal equations derived from the likelihood function. In such cases, side assumptions can be used to identify the model as done in linear regression, but even then deriving the maximum likelihood estimators for SURME model is a challenging task. Besides, ignoring measurement error in the data can lead to a poor model fit.

In this paper, we introduce two novel methods—a pure Bayesian algorithm and a mean field variational Bayes (MFVB) technique—to estimate the SURME model where each equation can potentially have a different covariate that is measured with error. Both the approach employs a classical structural form of measurement error and the link between the covariate measured with error and the other covariates (with no measurement error) is modeled through an exposure equation. Identification of parameters is achieved by placing a prior distribution on the measurement error variance. The pure Bayesian approach is analytically simpler and produces tractable conditional distributions which enables the use of Gibbs sampling. However, the MCMC draws of the parameters corresponding to the covariate measured with error tend to be highly correlated. To reduce autocorrelation in MCMC draws, one may consider *thinning* i.e., use every l-th draw in estimating the parameter. Thinning is debatable and while some authors such as Owen (2017) recommend thinning, others such as Link and Eaton (2012) advise against the use of thinning. So, we explore other methods and come up with a more elegant solution to the problem of high autocorrelation i.e., the MFVB approach to estimating SURME model.

We illustrate both the techniques in multiple simulation studies and compare the results to a standard SUR model, where we ignore or do not model the measurement error. In the first set of simulation studies, data are generated from a SURME model using different values of the variance of the true unobserved variable, while holding the reliability ratio fixed. In the second set of simulations, data are generated using different values of reliability ratio, holding the variance of the true unobserved variable at a fixed value. The results suggest that both the proposed methods perform well and correctly highlight the importance of modeling measurement error within the SUR structure when variables are measured with error. In addition, the SURME model is implemented in an application drawn from the health literature and estimated using the two proposed methods. Specifically, weight and high density lipoprotein (*HDL*) are jointly modeled as a function of several covariates and blood pressure, which is common to both equations and considered to have measurement error. Blood pressure is modeled as a function of the covariates in the exposure equation. Model selection exemplify the practical utility of the SURME model compared to the standard SUR model.

The remainder of the paper is organized as follows. Section 2 presents the SURME model, derives the joint posterior density and proposes a Gibbs sampling algorithm to estimate the model. Section 3 develops the MFVB approximation of the MCMC algorithm. Section 4 demonstrates the two algorithms in several Monte Carlo simulation exercises and Section 5 presents an application drawn from health literature. Section 6 concludes.

2. The SURME Model and Estimation via Gibbs sampling

The seemingly unrelated regression with measurement error (SURME) model incorporates measurement error for covariates in the SUR model and can be expressed in terms of the following

equations,

$$y_{mi} = x'_{mi}\beta_m + z_{mi}\gamma_m + \varepsilon_{mi}, \qquad m = 1, ..., M; \ i = 1, ..., N, \tag{1}$$

where the response y_{mi} is a scalar, x'_{mi} is $(1 \times k_m)$ vector of covariates, z_{mi} is a true unobserved scalar covariate that is prone to measurement error, and the subscripts m and i denote the equation number and individual/observation, respectively. Stacking the equations for each i, we can write model (1) as follows,

$$y_i = X_i\beta + Z_i\gamma + \varepsilon_i, \qquad i = 1, .., N, \tag{2}$$

where $y_i = (y_{1i}, ..., y_{Mi})'$ and $\gamma = (\gamma_1, ..., \gamma_M)'$ are vectors of dimension $(M \times 1)$, and $\beta = (\beta_1, ..., \beta_M)'$ is of dimension $(K \times 1)$, where $K = k_1 + \cdots + k_M$. The matrices,

$$X_i = \begin{pmatrix} x'_{1i} & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & x'_{Mi} \end{pmatrix} \quad \text{and} \quad Z_i = \begin{pmatrix} z_{1i} & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & z_{Mi} \end{pmatrix},$$

are of dimension $(M \times K)$ and $(M \times M)$, respectively. In addition, the error ε_i is assumed to be independently and identically distributed (*i.i.d.*) as a normal distribution i.e., $\varepsilon_i \sim N(0, \Sigma_\varepsilon)$ for $i = 1, \cdots, N$, where the covariance,

$$\Sigma_\varepsilon = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1M} \\ \vdots & \ddots & \vdots \\ \sigma_{M1} & \cdots & \sigma_{MM} \end{pmatrix},$$

is a symmetric matrix that permits nonzero correlation across equations (or first subscript) for any given individual (or second subscript) and ties each independent regression into a system of equations, hence the phrase seemingly unrelated regression. Measurement error in reference to model (2) arises because Z_i is not observed, instead we observe W_i which is a sum of the true unobserved quantity Z_i and a measurement error term u_i. This definition implies a *classical measurement error* (Fuller, 1987). Additionally, we assume that the true unobserved quantity Z_i follows a distribution, so that the measurement error model is of the *structural form*. This can be represented as follows,

$$\widetilde{W}_i = \widetilde{Z}_i + \widetilde{u}_i, \qquad \widetilde{u}_i \sim N_M\left(0, \sigma_u^2 I_M\right), \qquad \textit{classical structural form}, \tag{3}$$

where for algebraic simplification, we use the notations $\widetilde{W}_i = (w_{1i}, ..., w_{Mi})'$, $\widetilde{Z}_i = (z_{1i}, ..., z_{Mi})'$, $\widetilde{u}_i = (u_{1i}, ..., u_{Mi})'$, then $W_i = diag(\widetilde{W}_i)$, $Z_i = diag(\widetilde{Z}_i)$, $u_i = diag(\widetilde{u}_i)$ are $(M \times M)$ diagonal matrices and I_M is a $(M \times M)$ identity matrix.

An interesting addition to equation (3) is to relate the primary explanatory variable of interest (here Z_i) to other covariates (X_i), giving rise to the *exposure model*. The term "exposure model" comes from epidemiology, where the primary explanatory variable is affected by exposure to "toxicants" or "risk factors". Therefore, the potential links between the latent variable Z and the other covariates X can be expressed as follows,

$$\widetilde{Z}_i = X_i\omega + \widetilde{\varepsilon}_{z,i}, \qquad \widetilde{\varepsilon}_{z,i} \sim N_M\left(0, \sigma_Z^2 I_M\right), \qquad \textit{exposure model}. \tag{4}$$

The three equations (2), (3) and (4) together define our SURME model and the resulting likelihood

4

is derived as follows,

$$f(y, W, Z | X, \Delta) = \prod_{i=1}^{N} f(y_i, W_i, Z_i | X, \Delta)$$

$$= \prod_{i=1}^{N} \left\{ f(y_i | W_i, Z_i, X, \Delta) \times f(W_i | Z_i, X, \Delta) \times f(Z_i | X, \Delta) \right\}$$

$$= \prod_{i=1}^{N} \left\{ f(y_i | Z_i, X, \Delta) \times f(W_i | Z_i, X, \Delta) \times f(Z_i | X, \Delta) \right\} \tag{5}$$

$$= \prod_{i=1}^{N} \left\{ (2\pi)^{-M/2} |\Sigma_\varepsilon|^{-1/2} \exp \left[-\frac{1}{2}(y_i - X_i\beta - Z_i\gamma)' \Sigma_\varepsilon^{-1}(y_i - X_i\beta - Z_i\gamma) \right] \right.$$

$$\times (2\pi)^{-M/2} (\sigma_u^2)^{-M/2} \exp \left[-\frac{1}{2\sigma_u^2}(\widetilde{W}_i - \widetilde{Z}_i)'(\widetilde{W}_i - \widetilde{Z}_i) \right]$$

$$\left. \times (2\pi)^{-M/2} (\sigma_Z^2)^{-M/2} \exp \left[-\frac{1}{2\sigma_Z^2}(\widetilde{Z}_i - X_i\omega)'(\widetilde{Z}_i - X_i\omega) \right] \right\},$$

where $\Delta \equiv (\beta, \gamma, \Sigma_\varepsilon, \omega, \sigma_Z^2, \sigma_u^2)$ and as mentioned earlier, $\widetilde{W}_i$ and $\widetilde{Z}_i$ are column vectors that contain the diagonal elements of the matrices W_i and Z_i, respectively.

Before proceeding with estimation, we add a few words on identification issues that typically arise with measurement error models. In linear regression with measurement error, identification of parameters require additional assumptions. Such assumptions can be constant measurement error variance, known reliability ratio or some other conditions as presented in Cheng and Van Ness (1999). The same identification conditions are also applicable to the proposed SURME model under the existing distributional assumptions. Nonetheless, we follow a purely Bayesian approach and employ prior distributions to identify the parameters of the model (see Zellner, 1971, Chap. 5).

The Bayesian estimation method combines the likelihood of the model with suitable prior distributions to obtain the joint posterior distribution. We utilize the following prior distributions:

$$\beta \sim N_K (\beta_0, B_0), \quad \gamma \sim N_M (\gamma_0, G_0), \quad \Sigma_\varepsilon^{-1} \sim W_M (\nu_0, S_0),$$
$$\omega \sim N_K (\omega_0, O_0), \quad \sigma_Z^2 \sim IG (\delta_1, \delta_2), \quad \sigma_u^2 \sim IG (\delta_3, \delta_4), \tag{6}$$

where W_M denotes a Wishart distribution of dimension M and IG denotes an inverse gamma distribution. Here we note that if one is not interested in the exposure equation, it can be dropped from the model. In such a case, $\tilde{Z}_i \sim N(\mu, \sigma_Z^2 I_M)$ and μ can be given a normal prior as $\mu \sim N(\mu_0, \sigma_\mu^2 I_M)$. Coming back to the SURME model, the joint posterior distribution can be obtained

by combining the likelihood (5) with the prior distributions (6) as follows,

$$
\begin{aligned}
p\left(\Delta, Z | y, X, W\right) \propto \prod_{i=1}^{N} & \left\{ |\Sigma_\varepsilon|^{-1/2} \exp\left[-\frac{1}{2} \left(y_i - X_i\beta - Z_i\gamma\right)' \Sigma_\varepsilon^{-1} \left(y_i - X_i\beta - Z_i\gamma\right) \right] \right. \\
& \times \left(\sigma_u^2\right)^{-M/2} \exp\left[-\frac{1}{2\sigma_u^2} \left(\widetilde{W}_i - \widetilde{Z}_i\right)' \left(\widetilde{W}_i - \widetilde{Z}_i\right) \right] \\
& \left. \times \left(\sigma_Z^2\right)^{-M/2} \exp\left[-\frac{1}{2\sigma_Z^2} \left(\widetilde{Z}_i - X_i\omega\right)' \left(\widetilde{Z}_i - X_i\omega\right) \right] \right\} \\
& \times |B_0|^{-1/2} \exp\left[-\frac{1}{2} \left(\beta - \beta_0\right)' B_0^{-1} \left(\beta - \beta_0\right) \right] \\
& \times |G_0|^{-1/2} \exp\left[-\frac{1}{2} \left(\gamma - \gamma_0\right)' G_0^{-1} \left(\gamma - \gamma_0\right) \right] \\
& \times |\Sigma_\varepsilon|^{-\frac{s_0 - M - 1}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}\left(S_0^{-1}\Sigma_\varepsilon^{-1}\right) \right] \\
& \times \left(\sigma_Z^2\right)^{-\delta_1 - 1} \exp\left[-\frac{\delta_2}{\sigma_Z^2} \right] \times \left(\sigma_u^2\right)^{-\delta_3 - 1} \exp\left[-\frac{\delta_4}{\sigma_u^2} \right].
\end{aligned}
\tag{7}
$$

Typical with the Bayesian approach, the joint posterior density (7) is not tractable and the parameters are sampled using MCMC techniques. To this purpose, conditional posterior densities of the parameters are derived (see Appendix A in the supplementary material) and Gibbs sampling is employed to estimate the model as exhibited in Algorithm 1. Note that some of the conditional posteriors are conditioned on a subset of parameters, but these are full conditionals that just do not depend on the full set of parameters. Conditional posteriors that depend on a subset of parameters have also been referred to as reduced conditional posteriors and Gibbs sampling as partially collapsed Gibbs sampling (see Liu, 1994; van Dyk and Park, 2008).

The sampling algorithm, presented in Algorithm 1, shows that β and γ are sampled from an updated multivariate normal distribution. Standard result is obtained for the precision matrix Σ_ε^{-1}, which is sampled from an updated Wishart distribution. All the three parameters $(\beta, \gamma, \Sigma_\varepsilon)$ follow their respective distributions marginally of ω, σ_u^2 and σ_Z^2. The true unobserved quantity Z is drawn from an updated multivariate normal distribution conditional on all the remaining model parameters. Similarly, ω is sampled from an updated multivariate normal distribution conditional on (Z, σ_Z^2). The two variance parameters are drawn from updated inverse gamma distributions with σ_Z^2 conditioned on (Z, ω) and σ_u^2 conditioned on (W, Z). Note that if we drop the exposure equation from the SURME model, Algorithm 1 only requires a slight modification. In this context, μ replaces $X_i\omega$ and is sampled from an updated normal distribution as $\mu | Z, \sigma_Z^2 \sim N_M(\bar{l}, \bar{\Lambda})$, where $\bar{l} = \bar{\Lambda}\left(\sum_{i=1}^{N} \widetilde{Z}_i/\sigma_Z^2 + \frac{\mu_0}{\sigma_\mu^2}\right)$ and $\bar{\Lambda}^{-1} = \left(\frac{N}{\sigma_Z^2} + \frac{1}{\sigma_\mu^2}\right)I_M$ are the posterior mean and posterior precision, respectively.

We note that the model presented in this paper utilizes the structural measurement error model which assumes that Z follows a distribution. Hence, the distribution of Z was introduced as a part of the model. However, in the measurement error literature, there is another form of measurement error known as *functional* form. The functional measurement error model assumes that the true unobserved quantity Z is fixed. In our modeling and estimation framework, we can easily incorporate the functional form of measurement error by modeling the distribution of Z as a part of the subjective prior information (Zellner, 1971). This implies that the joint posterior distribution (6) will remain unchanged and derivations of the conditional posterior distributions

Algorithm 1 (Gibbs sampling for SURME model)

---

1. Sample $\beta|\gamma, \Sigma_\varepsilon, Z, y \sim N_K\left(\overline{\beta}, B_1\right)$, where,

$$B_1^{-1} = \left[\sum_{i=1}^{N} X_i'\Sigma_\varepsilon^{-1}X_i + B_0^{-1}\right], \ \overline{\beta} = B_1\left[\sum_{i=1}^{N} X_i'\Sigma_\varepsilon^{-1}y_i^* + B_0^{-1}\beta_0\right], \text{ and } y_i^* = y_i - Z_i\gamma.$$

2. Sample $\gamma|\beta, \Sigma_\varepsilon, Z, y \sim N_M\left(\overline{\gamma}, G_1\right)$, where,

$$G_1^{-1} = \left[\sum_{i=1}^{N} Z_i'\Sigma_\varepsilon^{-1}Z_i + G_0^{-1}\right], \ \overline{\gamma} = G_1\left[\sum_{i=1}^{N} Z_i'\Sigma_\varepsilon^{-1}\tilde{y}_i + G_0^{-1}\gamma_0\right], \text{ and } \tilde{y}_i = y_i - X_i\beta.$$

3. Sample $\Sigma_\varepsilon^{-1}|\beta, \gamma, Z, y \sim W_M\left(\nu_1, S_1\right)$, where,

$$\nu_1 = \nu_0 + N \text{ and } S_1^{-1} = \left[S_0^{-1} + \sum_{i=1}^{N}(y_i - X_i\beta - Z_i\gamma)(y_i - X_i\beta - Z_i\gamma)'\right].$$

4. Sample $\tilde{Z}_i|\beta, \gamma, \Sigma_\varepsilon, \omega, \sigma_Z^2, \sigma_u^2, W, y \sim N_M\left(M_{1,i}, M_2\right), \forall i = 1, ..., N,$ where,

$$M_2^{-1} = \left[\Psi + \left(\frac{1}{\sigma_Z^2} + \frac{1}{\sigma_u^2}\right)I_M\right], \text{ with } \Psi = \Gamma \odot \Sigma_\varepsilon^{-1}, \ \Gamma = \gamma\gamma',$$

and $M_{1,i} = M_2\left[diag(\gamma)\Sigma_\varepsilon^{-1}(y_i - X_i\beta) + \frac{\widetilde{W}_i}{\sigma_u^2} + \frac{X_i\omega}{\sigma_Z^2}\right]$, where $\odot$ is the dot (or Hadamard) product.

5. Sample $\omega|Z, \sigma_Z^2 \sim N_M\left(\omega_1, \Sigma_\omega\right)$, where,

$$\Sigma_\omega^{-1} = \left[\frac{1}{\sigma_Z^2}\sum_{i=1}^{N} X_i'X_i + O_0^{-1}\right] \text{ and } \omega_1 = \Sigma_\omega\left[\frac{1}{\sigma_Z^2}\sum_{i=1}^{N} X_i'\tilde{Z}_i + O_0^{-1}\omega_0\right].$$

6. Sample $\sigma_Z^2|Z, \omega \sim IG\left(\delta_1^*, \delta_2^*\right)$, where,

$$\delta_1^* = \delta_1 + \frac{N.M}{2} \text{ and } \delta_2^* = \delta_2 + \frac{1}{2}\sum_{i=1}^{N}\left(\tilde{Z}_i - X_i\omega\right)'\left(\tilde{Z}_i - X_i\omega\right).$$

7. Sample $\sigma_u^2|Z, W \sim IG\left(\delta_3^*, \delta_4^*\right)$, where,

$$\delta_3^* = \delta_3 + \frac{N.M}{2} \text{ and } \delta_4^* = \delta_4 + \frac{1}{2}\sum_{i=1}^{N}\left(\tilde{W}_i - \tilde{Z}_i\right)'\left(\tilde{W}_i - \tilde{Z}_i\right).$$

---

will proceed in exactly the same way as described in Appendix A of the supplementary file. To reiterate, the fundamental difference in analyzing SUR model with structural and functional forms of measurement error lies in the interpretation given to the distribution of Z, the true unobserved quantity.

In the MCMC estimation of SURME model, one consideration that arise is that Z and γ are both unknown, and drawing them conditional on each other lead to high autocorrelation in MCMC draws. This occurrence is a general problem and happens when two or more unknown variables/parameters that appear in product form are drawn conditional on each other. To reduce the autocorrelation in MCMC draws (and consequently reduce the inefficiency factors) some au-

thors[3] propose to improve mixing by sampling γ from the marginal distribution and then sampling $Z|\gamma$ or *vice versa*. However, deriving the marginal posterior distribution of γ (or of Z) is not straightforward and the marginalization trick do not improve the results in our modeling context. [4] As a solution to reduce autocorrelation, many researchers have employed *thinning* to improve the mixing of the draws. The thinning of MCMC draws has been criticized by some authors including MacEachern and Berliner (1994), Link and Eaton (2012)), but others such as Geyer (1991) acknowledges that thinning can increase statistical efficiency. In a recent paper, Owen (2017) shows that the usual advice against thinning can be misleading. We employ thinning to improve the mixing properties of the MCMC draws of γ in our simulation studies and application. However, given the controversy around thinning, we explore other methods and come up with the MFVB approximation to estimate SURME model.

3. The mean field variational Bayes (MFVB) approximation

Variational Bayes is an alternative to MCMC methods that provides a locally-optimal, exact analytical solution to an approximation of the posterior distribution. The parameters of the approximate distribution are selected to minimize the Kullback-Leibler divergence (a distance measure) between the approximation and the posterior. The MFVB approximation is a deterministic optimization approach and so is particularly useful for big data sets and/or models with large sparse covariance matrices. Besides, it is similar to Gibbs sampling for conjugate models. Some recent articles on MFVB approach include Bishop (2006), Ormerod and Wand (2010), Pham et al. (2013), Lee and Wand (2016) and Blei et al. (2017).

Suppose, y denotes an observed data vector and θ is a parameter vector defined over the parameter space Θ. Following the Bayes theorem, the posterior distribution can be written as:

$$p\left(\theta|y\right) = \frac{p\left(\theta,y\right)}{p\left(y\right)} = \frac{p\left(y|\theta\right)p\left(\theta\right)}{p\left(y\right)},$$

where $p\left(y\right) = \int_{\Theta} p\left(\theta,y\right)d\theta$ is the marginal likelihood. Let q be an arbitrary density function over Θ. Then, the logarithm of the marginal likelihood is,

$$
\begin{aligned}
\log p\left(y\right) &= \log p(y)\int_{\Theta} q(\theta)d\theta = \int_{\Theta} q(\theta)\log p(y)d\theta \\
&= \int_{\Theta} q(\theta)\log\left\{\frac{p(\theta,y)/q(\theta)}{p(\theta/y)/q(\theta)}\right\}d\theta \\
&= \int_{\Theta} q\left(\theta\right)\log\left\{\frac{p\left(\theta,y\right)}{q\left(\theta\right)}\right\}d\theta + \int_{\Theta} q\left(\theta\right)\log\left\{\frac{q\left(\theta\right)}{p\left(\theta|y\right)}\right\}d\theta \\
&= \log \underline{p}\left(y,q\right) + KL(q,p) \\
&= \log \underline{p}\left(y,q\right) + E_{q(\theta)}\left[\log q\left(\theta\right)\right] - E_{q(\theta)}\left[\log p\left(\theta,y\right)\right] + \log p\left(y\right),
\end{aligned}
\tag{8}
$$

where $\log \underline{p}\left(y,q\right) = E_{q(\theta)}\left[\log\left(\frac{p(\theta,y)}{q(\theta)}\right)\right]$ denotes the lower bound on the marginal log-likelihood and $KL(q,p) = E_{q(\theta)}\left[\log q\left(\theta\right)\right] - E_{q(\theta)}\left[\log p\left(\theta|y\right)\right]$ is the Kullback-Leibler divergence between $q\left(\theta\right)$ and

---

[3]See for instance Jeliazkov (2013) for the case of latent variables in a non parametric VAR specification.

[4]Many thanks to Ivan Jeliazkov and the participants of the UCI seminar for the suggestion to sample γ marginally of Z and then sampling $Z|\gamma$. See appendix E in the supplementary material. However, the several tests we conducted did not improve our initial results with standard Gibbs sampling.

$p(\theta|y)$. Since $\log p(y)$ is a constant, the minimization of $KL(q,p)$ is equivalent to maximizing the scalar quantity $\log \underline{p}(y,q)$, typically known as evidence lower bound (ELBO) or variational lower bound. In practice, the maximization of the ELBO is often preferred to minimization of the KL divergence since it does not require knowledge of the posterior.

The MFVB approximates the posterior distribution $p(\theta|y)$ by the product of the q-densities,

$$q(\theta) = \prod_{j=1}^{P} q_j(\theta_j). \tag{9}$$

Each *optimal q*-density minimizes the Kullback-Leibler divergence and is given by,

$$q_j(\theta_j) \propto \exp\left[E_{q(-\theta_j)}\{\log p(\theta_j|\Omega)\}\right] , \ j = 1, ..., P \tag{10}$$

where $E_{q(-\theta_j)}$ denotes the expectation with respect to $\prod_{k \neq j} q_k(\theta_k)$, $\Omega \equiv \{y, \theta_1, ..., \theta_{j-1}, \theta_{j+1}, ..., \theta_P\}$ is the set containing all random vectors in the model except θ_j, and $p(\theta_j|\Omega)$ are the full conditional distributions of the parameters.

For the SURME model, we now consider a MFVB approximation based on the following factorization:

$$q\left(\beta, \gamma, \Sigma_\varepsilon, \omega, \sigma_Z^2, \sigma_u^2, \tilde{Z}\right) = q(\beta)\, q(\gamma)\, q(\Sigma_\varepsilon)\, q(\omega)\, q\left(\sigma_Z^2\right) q\left(\sigma_u^2\right) \prod_{i=1}^{N} q\left(\tilde{Z}_i\right).$$

These optimal q-densities can be derived, as presented in Appendix B of the supplementary file, to have the following form,

$$
\begin{aligned}
q(\beta) &= f_{N_K}\left(\mu_{q(\beta)}, \Sigma_{q(\beta)}\right) & q(\gamma) &= f_{N_M}\left(\mu_{q(\gamma)}, \Sigma_{q(\gamma)}\right) \\
q\left(\Sigma_\varepsilon^{-1}\right) &= f_{W_M}\left(\nu_1, B_{q(\Sigma)}\right) & q(\omega) &= f_{N_K}\left(\mu_{q(\omega)}, \Sigma_{q(\omega)}\right) \\
q\left(\tilde{Z}_i\right) &= f_{N_M}\left(\mu_{q(\tilde{Z}_i)}, \Sigma_{q(\tilde{Z}_i)}\right) & q\left(\sigma_Z^2\right) &= f_{IG}\left(\delta_1^*, B_{q(\sigma_Z^2)}\right) \\
q\left(\sigma_u^2\right) &= f_{IG}\left(\delta_3^*, B_{q(\sigma_u^2)}\right),
\end{aligned}
\tag{11}
$$

where f denotes the density function of the distribution given in the subscript. The parameters of the optimal densities are updated according to Algorithm 2. When exposure is dropped, $q(\omega)$ is replaced with $q(\mu)$ and the optimal density is an updated normal distribution. Convergence of Algorithm 2 is assessed using the evidence lower bound ℓ on the marginal log-likelihood (see Appendix C in the supplementary material) that is guaranteed to reach a local optima based on the convexity property. This algorithm belongs to the family of coordinate ascent variational inference (CAVI) and iteratively optimizes each factor of the mean field variational density, while holding the remaining fixed (see Bishop, 2006; Blei et al., 2017).

The MFVB technique provides computational advantages compared to MCMC because it is deterministic and does not require a large number of iterations (Pham et al., 2013; Lee and Wand, 2016). Besides, existing works including Bishop (2006), Ormerod and Wand (2010), Faes et al. (2011), Pham et al. (2013), and Lee and Wand (2016) suggest that the accuracy scores of the MFVB approximation, relative to MCMC, generally exceed $95 - 97\%$ and rarely goes below 90%. Given these advantages, the MFVB approach can be gainfully utilized for large data models. However, some authors have reported that covariance matrices from variational approximation may be typically "too small" relative to the sampling distribution of the maximum likelihood estimator.

Algorithm 2 (MFVB algorithm for SURME model)

1. Initialize δ_1^*, δ_3^*, $B_{q(\sigma_Z^2)}$, $B_{q(\sigma_u^2)}$, $\mu_{q(\beta)}$, $\mu_{q(\gamma)}$, $\mu_{q(\omega)}$, $\Sigma_{q(\beta)}$, $B_{q(\Sigma)}$, $\Sigma_{q(\gamma)}$, $\Sigma_{q(\omega)}$, $\mu_{q(\tilde{Z}_i)}$, $\Sigma_{q(\tilde{Z}_i)}$ (for $i = 1, ..., N$).

2. Cycle:

   (a) $\Sigma_{q(\beta)} \leftarrow \left[\sum_{i=1}^N X_i' \left(\nu_1 B_{q(\Sigma)} \right) X_i + B_0^{-1} \right]^{-1}$

   (b) $\mu_{q(\beta)} \leftarrow \Sigma_{q(\beta)} \left[\sum_{i=1}^N X_i' \left(\nu_1 B_{q(\Sigma)} \right) \left(y_i - diag(\mu_{q(\tilde{Z}_i)})\mu_{q(\gamma)} \right) + B_0^{-1}\beta_0 \right]$

   (c) $\Sigma_{q(\gamma)} \leftarrow \left[\sum_{i=1}^N \left(\Sigma_{q(\tilde{Z}_i)} + \mu_{q(\tilde{Z}_i)}\mu_{q(\tilde{Z}_i)}' \right) \odot \left(\nu_1 B_{q(\Sigma)} \right) + G_0^{-1} \right]^{-1}$

   (d) $\mu_{q(\gamma)} \leftarrow \Sigma_{q(\gamma)} \left[\sum_{i=1}^N diag(\mu_{q(\tilde{Z}_i)}) \left(\nu_1 B_{q(\Sigma)} \right) \left(y_i - X_i\mu_{q(\beta)} \right) + G_0^{-1}\gamma_0 \right]$

   (e) $B_{q(\Sigma)} \leftarrow \left[S_0^{-1} + \sum_{i=1}^N \left[\left(y_i - X_i\mu_{q(\beta)} - diag(\mu_{q(\tilde{Z}_i)})\mu_{q(\gamma)} \right) \right. \right.$
   $$\times \left(y_i - X_i\mu_{q(\beta)} - diag(\mu_{q(\tilde{Z}_i)})\mu_{q(\gamma)} \right)' + X_i\Sigma_{q(\beta)}X_i'$$
   $$\left. \left. + \left(\mu_{q(\tilde{Z}_i)}\mu_{q(\tilde{Z}_i)}' \right) \odot \Sigma_{q(\gamma)} + \Sigma_{q(\tilde{Z}_i)} \odot \left(\Sigma_{q(\gamma)} + \mu_{q(\gamma)}\mu_{q(\gamma)}' \right) \right] \right]^{-1}$$

   (f) $B_{q(\sigma_Z^2)} \leftarrow \delta_2 + \frac{1}{2}\sum_{i=1}^N \left\{ \parallel \mu_{q(\tilde{Z}_i)} - X_i\mu_{q(\omega)} \parallel^2 + tr\left[\Sigma_{q(\tilde{Z}_i)} \right] \right\}$

   (g) $B_{q(\sigma_u^2)} \leftarrow \delta_4 + \frac{1}{2}\sum_{i=1}^N \left\{ \parallel \tilde{W}_i - \mu_{q(\tilde{Z}_i)} \parallel^2 + tr\left[\Sigma_{q(\tilde{Z}_i)} \right] \right\}$

   (h) $\Sigma_{q(\omega)} \leftarrow \left[\left(\frac{\delta_1^*}{B_{q(\sigma_Z^2)}} \right) \sum_{i=1}^N X_i'X_i + O_0^{-1} \right]^{-1}$

   (i) $\mu_{q(\omega)} \leftarrow \Sigma_{q(\omega)} \left[\left(\frac{\delta_1^*}{B_{q(\sigma_Z^2)}} \right) \sum_{i=1}^N X_i'\mu_{q(\tilde{Z}_i)} + O_0^{-1}\omega_0 \right]$

   (j) $\Sigma_{q(\tilde{Z}_i)} \leftarrow \left[\left\{ \Sigma_{q(\gamma)} + \mu_{q(\gamma)}\mu_{q(\gamma)}' \right\} \odot \left(\nu_1 B_{q(\Sigma)} \right) + \left(\frac{\delta_1^*}{B_{q(\sigma_Z^2)}} + \frac{\delta_3^*}{B_{q(\sigma_u^2)}} \right) I_M \right]^{-1}$

   (k) $\mu_{q(\tilde{Z}_i)} \leftarrow \Sigma_{q(\tilde{Z}_i)} \left[diag(\mu_{q(\gamma)}) \left(\nu_1 B_{q(\Sigma)} \right) \left(y_i - X_i\mu_{q(\beta)} \right) + \left(\frac{\delta_3^*}{B_{q(\sigma_u^2)}} \right) \widetilde{W}_i \right.$
   $$\left. + \left(\frac{\delta_1^*}{B_{q(\sigma_Z^2)}} \right) X_i\mu_{q(\omega)} \right]$$

   until the increase in the ELBO (ℓ) is negligible $(\approx 10^{-7})$.

In this regard, Blei et al. (2017) opine that underestimation of the variance should be judged in relation to the task at hand. However, evidence from empirical research indicate that variational inference typically do not suffer in accuracy.

4. Monte Carlo simulation studies

This section examines the performance of the two proposed methods in multiple simulation studies. The first set of simulations (Case I) employ different values of σ_Z^2 to generate the simulated

data. The second set of simulations (Case II) use different values of reliability ratio defined as $R_Z = \sigma_Z^2/(\sigma_Z^2 + \sigma_u^2)$. In both sets of simulations, we use a two equation structure represented as follows,

$$
\begin{aligned}
y_{1i} &= \beta_{11} + x_{1i2}\beta_{12} + x_{1i3}\beta_{13} + z_{1i}\gamma_1 + \varepsilon_{1i}, \\
y_{2i} &= \beta_{21} + x_{2i2}\beta_{22} + x_{2i3}\beta_{23} + z_{2i}\gamma_2 + \varepsilon_{2i},
\end{aligned}
\tag{12}
$$

where the first, second and third subscripts in x_{mij} denote the equation number ($m = 1, 2$), observation ($i = 1, \cdots, N$) and variable number ($j = 2, 3$), respectively. The first variable is common to both the equations (i.e., $x_{1i2} = x_{2i2}$ for all $i = 1, ..., N$) and the remaining covariates are exclusive to the respective equations. Moreover, we assume the error prone covariate $Z_i = diag(z_{1i}, z_{2i})$ for all i is unobserved, but is defined by an exposure model as follows,

$$
\begin{aligned}
z_{1i} &= \omega_{11} + x_{1i2}\omega_{12} + x_{1i3}\omega_{13} + \varepsilon_{z1i}, \\
z_{2i} &= \omega_{21} + x_{2i2}\omega_{22} + x_{2i3}\omega_{23} + \varepsilon_{z2i}.
\end{aligned}
\tag{13}
$$

The unobserved Z_i is related to the observed $W_i = diag(w_{1i}, w_{2i})$ by the equations below,

$$
\begin{aligned}
w_{1i} &= z_{1i} + u_{1i}, \\
w_{2i} &= z_{2i} + u_{2i}.
\end{aligned}
\tag{14}
$$

Note that the estimation of the SURME model solely relies on W and the role of Z is limited to generating values for (W, y).

To proceed with data generation, we assign specific values to the parameters β, γ, ω, Σ_ε, σ_Z^2, σ_u^2 and generate $N = 300$ observations in each simulation study for all the variables in the model. Let $\beta_{11} = 3$, $\beta_{12} = 5$, $\beta_{13} = 4$, $\beta_{21} = 4$, $\beta_{22} = 3.8$, $\beta_{23} = 3$, $\gamma_1 = 4$, $\gamma_2 = 4$, $\omega_{11} = 1.5$, $\omega_{12} = 0.75$, $\omega_{13} = 0.3$, $\omega_{21} = 1.5$, $\omega_{22} = 1.05$, and $\omega_{23} = 0.45$. For all values of i, the error vector $\varepsilon_i = (\varepsilon_{1i}, \varepsilon_{2i})'$ is generated from a bivariate normal distribution $N(0_M, \Sigma_\varepsilon)$, where $\Sigma_\varepsilon = [1\ 0.5; 0.5\ 1]$. Values for the common covariate ($x_{1i2} = x_{2i2}$) are generated from $U(0, 2)$ and values for the exclusive covariates x_{1i3} and x_{2i3} are generated from $U(0, 4)$, where U denotes an uniform distribution. Values for $\widetilde{Z}_i = (z_{1i}, z_{2i})'$ are generated as $\widetilde{Z}_i \sim N(X_i\omega, \sigma_Z^2 I_M)$, and the $\widetilde{W}_i$'s are generated as $\widetilde{W}_i = \widetilde{Z}_i + \widetilde{u}_i$ where $\widetilde{u}_i \sim N(0_M, \sigma_u^2 I_M)$. The above setting remains the same in the following subsections, with change occurring only in values of σ_u^2 (through R_Z) or σ_Z^2.

In Case I, we investigate the performance of the proposed algorithms in two simulation studies where the reliability ratio R_Z is fixed ($R_Z = 0.8$) and σ_Z^2 is gradually decreased. Specifically, two values are considered $\sigma_Z^2 = \{1, 0.0625\}$. The definition of R_Z is used to generate the corresponding values for $\sigma_u^2 = \sigma_Z^2(1 - R_Z)/R_Z$, which leads to a noise-to-true variance ratio $(1 - R_Z)/R_Z$ of 25%. In Case II, we again examine the performance of the proposed algorithms in two simulation studies by keeping σ_Z^2 fixed ($\sigma_Z^2 = 0.0625$) and using two values of reliability ratio $R_Z = \{0.8, 0.5714\}$. The chosen values are similar to those used in Pham et al. (2013) and leads to noise-to-true variance ratios of 25% and 75%, respectively. We could define a noise-to-true variance of 100%, 150% or more, but in those cases we will be dealing more with outliers than with measurement errors.

Bayesian procedures require prior distribution on the parameters of the model. For the SURME model, we stipulate the following priors: $\beta \sim N_K(\beta_0, B_0)$ with $\beta_0 = \iota_K$, $B_0 = I_K$, and ι_M is a $(M \times 1)$ vector of ones; $\gamma \sim N_M(\gamma_0, G_0)$ with $\gamma_0 = \iota_M$, $G_0 = I_M$; $\omega \sim N_K(\omega_0, O_0)$ with $\omega_0 = \iota_K$, $O_0 = I_K$; $\Sigma_\varepsilon^{-1} \sim W_M(\nu_0, S_0)$ with $\nu_0 = 50$ and $S_0 = \nu_0[1\ 0.5; 0.5\ 1]$; $\sigma_Z^2 \sim IG(\delta_1, \delta_2)$ and $\sigma_u^2 \sim IG(\delta_3, \delta_4)$ with $\delta_1 = \delta_2 = \delta_3 = \delta_4 = 1/100$. All these priors are proper yet specify vague information about the parameters mainly for the measurement error u_i and the error prone covariate

Z_i. In addition, the same priors are used in all the simulations to highlight the effect of changing R_Z or σ_Z^2 in estimation of the parameters and consequently on the performance of the algorithms.

The MCMC results are obtained from $50,000$ draws, after a burn-in of $1,000$ draws. We replicate these simulations 100 times and report the means over these 100 replications of the posterior means[5]. We also compare the results with the usual frequentist SUR estimation and the standard Bayesian estimation of SUR model. The Gibbs sampling algorithm for the latter is presented in Appendix D of the supplementary material.

4.1. Case I: Altering σ_Z^2

Amongst the first set of simulation studies labeled Case I, Table 1 presents the results from the frequentist estimation of SUR[6] model for the case $R_Z = 0.8$ and $\sigma_Z^2 = 1$. Results show that estimates are strongly biased mainly for the intercepts β_{11} and β_{21}, and for γ_1 and γ_2. The relative biases ($\hat{\beta}/\beta - 1$) of the intercepts (resp. the γ's) are 38.1% and 28.8%, (resp. -19.8% and -19.6%). The γ's are strongly under-estimated. On the contrary, slope coefficients β_{12}, β_{13} and β_{23} are less contaminated by the measurement error and have a lower dispersion of the estimated coefficients than the intercepts. The relative bias of β_{22} (22.3%) is close (in absolute value) to that of γ's. Elements of the variance-covariance matrix Σ_ε are strongly over-estimated with a relative error of 317% and 314% for the variances and 4.6% for the covariance σ_{12}. It leads to a strong under-evaluation of the coefficient of correlation $\rho_{\varepsilon_1\varepsilon_2} = 0.126$ far from the true value (0.5). The Bayesian estimation of SUR model, presented in Table G1 of the supplementary material, give similar results. The posterior means of the coefficients (resp. posterior standard errors) are similar to the frequentist coefficient estimates (resp. standard errors) of the SUR model. The 95% highest posterior density intervals (HPDI) are also close to the 95% confidence interval of the frequentist estimation. The estimated correlation coefficient $\rho_{\varepsilon_1\varepsilon_2} = 0.125$ is similar to the frequentist estimate. We also report Geweke's convergence diagnostic (CD), which tests for the equality of means of the first and last part of a Markov chain on the basis of samples drawn from the stationary distribution of the chain. In more than 98% of cases, Geweke's CD (under the null hypothesis, $CD \sim N(0,1)$) accepts the null hypothesis at 5% level, which suggests that a sufficiently large number of draws have been taken. Moreover, inefficiency factors (reported in Table G1) are also close to 1, which confirms that the chain is mixing well.

In the upper panel of Table G2 (of the supplementary material), we see how the Bayesian estimation of SURME model improves the results. The intercepts β_{11} and β_{21} are now less biased as compared to that of SUR model in Table G1. Their relative errors are -3.5% and -7.1%, respectively. This is also true for the other slope coefficients β. Moreover, the model neatly corrects the measurement errors and results in better estimates of γ_1 and γ_2, their relative errors being 2.1% and 2.6%, respectively. We also note that the variance-covariance matrix is precisely estimated leading to a correlation coefficient $\rho_{\varepsilon_1\varepsilon_2} = 0.494$. The parameters σ_Z^2 and σ_u^2 are well estimated with small posterior standard deviations and small relative errors (-2.6% and 0.6%, respectively). But inefficiency factors are large indicating strong autocorrelation in MCMC draws, particularly for γ_1 and γ_2 whose inefficiency factors are 8.62 and 10.52, respectively. In more than

---

[5]To save time, we only run $51,000$ draws per replication. Higher number of MCMC draws, such as $100,000$ or $200,000$, exponentially increase the computing time without any increase in precision. As an example for $\sigma_Z^2 = 0.0625$ and $R_Z = 0.909$, the MFVB takes only 11.52 seconds per replication. If $51,000$ (resp. $101,000$ and $201,000$) draws are used, the computing time per replication for the Gibbs sampling of the BSURME model is about 65.64 (resp. 142.07 and 352.29) seconds using a MacBook Pro, 2.8 GHz core i7 with 16Go 1600 MHz DDR3 RAM.

[6]Without any prior information on the measurement error, the SUR model for M equations is the following: $y_i = X_i\beta + W_i\gamma + \varepsilon_i$, $\varepsilon_i \sim N(0, \Sigma_\varepsilon)$, $i = 1, .., N$ where W_i is the covariate with measurement error.

90% of cases, the Geweke's CD confirms that a sufficiently large number of draws have been taken. The improvement obtained with a SURME (as compared to the SUR) is interesting and emphasizes the need to model measurement error. The lower panel of Table G2 (in the supplementary material) presents the results of the exposure equation from the SURME model. They show that the biases are negligible, the posterior standard deviations are small and so are the inefficiency factors.

Overall, the SURME model is well estimated, but the high autocorrelation in MCMC draws of γ needs additional consideration. According to Owen (2017), the problem of high autocorrelation can be dealt with thinning, which itself can be optimized according to the cost of computing the quantities of interest (after advancing the Markov chain) and the speed at which autocorrelations decay. As shown in Table G3 (see the supplementary material), autocorrelations between the successive draws of γ_1 and γ_2, denoted $\rho_\tau(\gamma_1)$ and $\rho_\tau(\gamma_2)$, are close to one and the rate of decay is very slow. For example, $\rho_1(\gamma_1) = 0.98$, $\rho_{10}(\gamma_1) = 0.82$ and $\rho_1(\gamma_2) = 0.98$, $\rho_{10}(\gamma_2) = 0.87$. The autocorrelations of some latent variables Z_i are slightly higher (0.995) than those of the γ's, but not reported for the sake of brevity. The cost of computing of $\tilde{Z}_i$ is on average 2.71 and an autocorrelation of 0.995 leads to an optimal thinning of factor $k = 86$ (see appendix F and Table F1 in the supplementary material). Henceforth, we use a thinning of factor $k = 100$ for all simulations.

We re-estimate the Bayesian SUR and SURME models with a thinning of 100, but only report the results for SURME. The results, presented in the upper panel of Table 2, show that the posterior means and standard deviations are close (or identical) to those of Table G2 (in the supplementary material). Values of the inefficiency factors are small and are all between $(1.004, 1.23)$. Specifically, the reduction in inefficiency factor is tremendous for γ's (*e.g.*, 1.11 *versus* 8.62 for γ_1 and 1.23 *versus* 10.52 for γ_2). The lower panel of Table 2 presents the results for the exposure equation in the SURME model. Once again, the results show that the biases are negligible, the posterior standard deviations are small and the inefficiency factors and Geweke's CD suggest good mixing of the MCMC draws. Specifically, the autocorrelations of γ_1 and γ_2 are now small ($\rho_1(\gamma_1) = 0.15$, $\rho_1(\gamma_2) = 0.27$) and quickly converge towards zero ($\rho_{10}(\gamma_1) = -0.007$, $\rho_{10}(\gamma_2) = -0.003$) confirming a good mixing of Markov chains (see Table G4 in the supplementary material).

To compare models, we employ the deviance information criterion or DIC[7] proposed by Spiegelhalter et al. (2002), and further studied in Celeux et al. (2006) and Spiegelhalter et al. (2014). Following Chan and Grant (2016), we compute the integrated likelihood for the SUR and SURME model with a thinning factor of 1 and 100. This is used to calculate the marginal likelihood, which is then utilized in DIC and the effective number of parameters p_D. For the SUR model, we get negative estimates of p_D which is indicative of either a poor fit between the model and data or a conflict between the prior and data. Different variations on the prior yield negative p_D, so it is more likely due to a poor fit between the SUR model and data. When $p_D < 0$, the DICs are not adequate for evaluating the complexity and the fit of a model (Celeux et al., 2006). On the other hand, for the SURME model we get a positive estimate of p_D, synonymous with better fit (see Appendix E in the supplementary material for further discussion of the method and Table G5 for the results).

We next discuss the results from MFVB approach, which on average takes about 145 cycles to get the maximum of the evidence lower bound l and the algorithm is terminated when the relative increase in the evidence lower bound l is less than 10^{-7}. The results from the MFVB estimation of SURME model are presented in Table 3, which shows that the MFVB approach gives better results

---

[7]Note that there does not exist any model adequacy measure that takes into account measurement error in a multi-equation setup. This is an open area of research and the only related work is Cheng et al. (2014), where they propose a coefficient of determination for linear regression models with measurement error.

compared to Gibbs sampling. All parameters have similar or lower biases, mainly for the intercepts β_{11}, β_{21} and for γ. However, the relative biases of the γ's are now reduced (0.7% and -0.3%) as compared to Bayesian estimation of SURME model. The estimates for σ_Z^2 and σ_u^2 show that the model accurately estimates the variances and their relative biases are small (0.4% and -3.5%, respectively). The standard deviation of all the parameters are smaller compared to those from MCMC estimation leading to slightly narrower 95% credible intervals (as compared to the 95% HPDI)[8]. Additionally, estimates of $\sigma_{mm'}$ are closer to their theoretical values and the estimated correlation coefficient $\rho_{\varepsilon_1\varepsilon_2} = 0.488$ is close to 0.5, the actual value. The lower panel of Table 3 presents the results from the exposure equation which emphasizes the accurate estimation of the ω parameters. The MFVB approximation of both the classical structural form and the exposure model shows that there are definite advantages in adopting the MFVB approach to estimate measurement error models as compared to the pure Bayesian method.

We next decrease the variance σ_Z^2 from $\sigma_Z^2 = 1$ to $\sigma_Z^2 = 0.0625$ leading to $\sigma_u^2 = 0.0156$. The results are presented in Tables G6 to G9 of the supplementary material. Results from the frequentist and Bayesian estimation of SUR model always reveal strong over-estimation of the intercepts, β_{22} and strong under-estimation of the slopes γ of the error prone covariate Z_i. However, over-estimation of the variances σ_{11} and σ_{22} ($\simeq 19\%$ for both) are largely reduced as compared to those of $\sigma_Z^2 = 1$, but leads to a slightly under-estimated correlation coefficient $\rho = 0.42$. When we incorporate the measurement error in the model, $i.e.$, SURME model with a very small variance of σ_Z^2, the Bayesian estimates show a less accurate estimation of the intercepts (increasing the negative relative biases -29.4% and -25.6%), of the γ's (15% and 16.8%) and of all the β's. Moreover, inefficiency factors rise to about 2 indicating a relative loss of efficiency due to slightly correlated samples. To neutralize this effect, we can increase the thinning appropriately[9]. Results for the exposure equation in the SURME model do not seem to be affected by the strong decrease of the variance σ_Z^2. The use of the MFVB approximation significantly attenuates the biases observed with the Bayesian estimation of SURME. The relative biases for the intercepts are now -17% and -11%, and those for the γ's are 8.6% and 7%, respectively. The relative errors for the variances σ_{mm}, $(m = 1, 2)$ reduce to -2.5% approximately and we get an estimated correlation coefficient $\rho = 0.52$. The MFVB approximation accurately estimates parameters of the exposure equation. Once again, the MFVB method reveals its advantages in estimating a SUR model with measurement error although this advantage tends to be attenuated when a very small variance σ_Z^2 occurs.

In summary, for a fixed measurement error of 25%, increasing the variance σ_Z^2 of the error prone covariate Z_i strongly biases the estimated variances σ_{mm} $(m = 1, 2)$ as well as the whole set of coefficients (intercepts and slopes) in the SUR model irrespective of the method of estimation. But, taking into account the measurement errors through SURME model neutralizes the negative effects of the increasing uncertainty on the error prone covariate Z_i and thus strongly reduces, or even eliminates the biases to obtain satisfactory estimates. This conclusion is further reinforced with the use of the MFVB approximation.

---

[8]When calibrating this Monte Carlo study, we found a significant underestimation of the variances of the coefficients γ_1 and γ_2, echoing the previous discussion around the work of Blei *et al.* (2017) (Section 3). After several trials (and to avoid embarking on more complex approaches such as linear response variational Bayes (Giordano et al., 2018) or α-variational inference (Yang et al., 2018), we decided to use the following simple trick to correct this under-valuation: σ_{γ_j} is replaced by $\sigma_{\gamma_j} \times \sqrt{MK/E_{q(\sigma_Z^2)}}$, for $j = 1, .., M$ (see Section B2 of the supplementary material).

[9]We relaunched the simulations for this case with a thinning of 120 and we find inefficiency factors close to 1. The results are available upon request for the sake of brevity.

4.2. Case II: Altering R_Z

We now investigate the performance of the proposed algorithms where $\sigma_Z^2 = 0.0625$ and the reliability ratio R_Z is gradually decreased. Specifically, we consider $R_Z = \{0.8, 0.5714\}$, which leads to $\sigma_u^2 = \sigma_Z^2(1 - R_Z)/R_Z = \{0.0156, 0.0469\}$ and noise-to-true variance ratio $(1 - R_Z)/R_Z$ of $\{25\%, 75\%\}$. In the previous subsection, we have already studied the case where $\sigma_Z^2 = 0.0625$ and $R_Z = 0.8$, therefore the focus is only on the case where the reliability is reduced to 0.5714.

The results presented in Tables G14-G17 of the supplementary material are poorer than those of $R_Z = 0.8$ for both the frequentist and Bayesian estimates of SUR model, with stronger over-estimation of the intercepts (84.3% and 25.6%) and stronger under-estimation of the slopes γ (-42.5%). The relative biases of the intercepts are larger than in the previous cases and the same is true for the γ's and even more obvious for the σ_{mm} $(m = 1, 2)$ (approximately 42%). The estimated correlation coefficient $\rho = 0.38$ is far from the true value 0.5. The Bayesian estimates of SURME model show a significant improvement, reducing the biases for the intercepts (33% and -10.8%) and the γ's (16.2% and 18.4%), but with slightly larger posterior standard deviations. The variance-covariance matrix Σ_ε is well estimated, with small relative biases of -7% and -9% for σ_{mm} $(m = 1, 2)$. The estimated correlation coefficient turns out to be $\rho = 0.53$. Both σ_Z^2 and σ_u^2 are also close to the true values (their relative biases are -15.2% and 15.1%, respectively). Once again, the improvement with the MFVB approximation is more noticeable as we get better results for the parameters with slightly smaller standard deviations. The relative biases for the intercepts are -16.6% and -5.7%, and those of the γ's are 8.5% and 6.5%. For the slope coefficients, the relative biases range between -11% and -2.5%. Both σ_Z^2 and σ_u^2 are also better estimated (their relative biases are respectively -5.6% and 4.4%). This is also true for the σ_{mm} $(m = 1, 2)$ (with small relative biases of -1.9% and -1.6%) leading to an estimated correlation coefficient $\rho = 0.51$.

To summarize, a change in the reliability ratio R_Z — for example, increasing the measurement error from 25% to 75% — strongly biases the whole set of coefficients (intercepts and slopes), including the estimated variances σ_{mm} $(m = 1, 2)$ in the SUR model. This is true both for the frequentist and Bayesian approach. On the other hand, accounting for measurement error through SURME model largely eliminates the negative effects of this alteration and strongly reduces the biases in SUR estimation. Moreover, the use of MFVB approximation improves the results beyond those obtained with the Bayesian estimation of SURME model.[10]

Finally, we present in Table 4 the relative errors of parameters for all the cases[11] of $\sigma_Z^2 = \{1, 0.0625\}$ and $R_Z = \{0.8, 0.5714\}$. At a glance, this table allows us to compare and contrast all of the previously discussed results and another case provided in the supplementary material. To reiterate, the results show that for a fixed measurement error, increasing the variance of the error prone covariate Z_i strongly biases the whole set of coefficients (intercepts and slopes) as well as the estimated variances σ_{mm} $(m = 1, 2)$ in the SUR model, irrespective of the method of estimation. This is also the case when, for a fixed variance σ_Z^2, the reliability ratio R_Z is reduced. Fortunately, taking into account the measurement error through SURME model attenuates or even neutralises the undesirable effects of the increasing uncertainty on the error prone covariate Z_i or reducing the reliability ratio. This conclusion is further reinforced with the use of the MFVB approximation.

---

[10]To get results with the Bayesian estimation of the SURME equivalent to those obtained with the MFVB approximation, it should be necessary to greatly increase the number of MCMC draws resulting in a very important cost in terms of computing time. Going from $51,000$ draws to $201,000$ draws leads to a relative increase in computing time per replication from 16 to 86 times that of MFVB (see note [5]). There is therefore an obvious trade-off against the Bayesian estimation of the SURME and in favor of the MFVB approximation.

[11]For $\sigma_Z^2 = 1$ and $R_Z = 0.5714$, results are given in Tables G13-G18. Last, Table G25 gives a summary of DICs and p_Ds.

5. Application

The statistical literature on modeling measurement error has often drawn applications from health and epidemiology studies where certain variables such as urinary sodium chloride (Liu and Liang, 1992) and blood pressure (Kannel et al., 1986) are treated as measured with error. In this context, Carroll et al. (2006a) utilizes measurement error in systolic blood pressure (SBP) on several occasions to illustrate different kinds of measurement error models and estimation methods. The idea is that long-term SBP is extremely difficult to measure and hence all recorded observations from clinic visits on SBP have measurement error. We draw motivation from Carroll et al. (2006b) and Tao et al. (2011)[12] and present an application of SURME model where the primary objective is to model the measurement error in SBP and explore the possibility of a better model fit relative to a standard SUR model.

The current study utilizes data from the National Health and Nutrition Examination Survey (NHANES) for 2007-2008, a widely used survey designed to assess health and nutritional status of civilians, non-institutionalized adults and children in the United States. NHANES collects data by interviewing individuals at home, who then report to mobile examination centers (MECs) to complete the health examination component of the survey. The MEC's provide a standardized environment for the collection of high quality data, thus favoring dependable statistical estimation and interpretations. The survey is unique in the sense that it combines interviews and physical examinations of the respondents.

The dependent variables in the model are log of weight and high density lipoprotein (HDL), which is also known as 'good cholesterol'. The covariates that are common to both equations and assumed to be measured without error are as follows: age, gender, smoking status, hours of sedentary activities, sleep disorder and low density lipoprotein plus 20 percent of Triglyceride ($LDL20T$). The variable 'height' is only expected to effect weight, not HDL, and therefore only included in the log weight equation. Observed SBP is assumed to have measurement error and transformed as $\ln(SBP-50)$ to avoid scaling problems, as done in Carroll et al. (2006b). The third reading on SBP is used as data and the first two readings are utilized to form priors on relevant parameters. Focussing on adults and removing missing observations on all variables of interest leaves us with a total of $N = 1,001$ observations. Table 5 presents the definition and descriptive statistics of all the variables used in the study.

To estimate the different SUR models with and without measurement error, we utilize the following relatively vague priors on the parameters: $\beta \sim N_{15}(\beta_0, B_0)$ with $\beta_0 = 0_{15}$, $B_0 = 10I_{15}$, $\gamma \sim N_2(\gamma_0, G_0)$ with $\gamma_0 = 0_2$, $G_0 = 10I_2$, $\Sigma_\varepsilon \sim IW_2(\nu_0, S_0)$ with $\nu_0 = 10$ and $S_0 = 10I_2$, $\omega \sim N(\omega_0, O_0)$ with $\omega_0 = 0_{15}$, $O_0 = I_{15}$, $\sigma_Z^2 \sim IG(50, 10)$ and $\sigma_u^2 \sim IG(50, 5)$. The prior distribution for σ_Z^2 is specified such that the prior mean (0.2) is close to the mean difference

---

[12]The Association for the Advancement of Medical Instrumentation (AAMI) and the British Hypertension Society recommend an absolute mean deviation (between oscillometric and invasive measurements of systolic blood pressure) less that 5 mmHg and a standard deviation of less than 8 mmHg. Using 6640 systolic blood pressure measures from 270 participants, Tao et al. (2011) find that large measurement errors of > 10 mmHg, (i.e., oscillometric measurement overestimates the real SBP) in 28.78% of the sample when their SBP values are around 90 mmHg. They also find that when SBP is more than 150 mmHg, most of the measurement errors are negative (i.e., oscillometric measurement underestimates the real SBP) . In their study, Tao et al. (2011) found an absolute mean deviation of 1.98 mmHg but a standard deviation of 14.87 mmHg, so practically doubling the recommended norm for measurement errors. As the authors say (p.288) "*If oscillometric measurement underestimates the real BP around the critical value (90 mmHg), the physician may give a wrong treatment. If the oscillometric measurement overestimates the real BP around 90 mmHg, the error may lead to an under-diagnosis and a delayed treatment response to perioperative hypotension and significantly increases the risk of dying*".

between first and second readings on transformed SBP (0.024). Similarly, the prior distribution for measurement error variance σ_u^2 is stipulated such that prior mean (0.1) is near the mean difference in variance from first and second readings on transformed SBP (0.002). Note that some of the parameters only appear in the measurement error model and the priors are used accordingly.

We first look at the results for the Bayesian estimation of SUR model presented in Table 6 from 400,000 draws after a burn-in of 50,000 draws with a thinning factor of 100 (optimized following the approach in Owen (2017)). The posterior estimates show that $\ln(age)$ is not statistically different from zero in both the $\ln(weight)$ and HDL equations. *Male* indicator variable positively affects $\ln(weight)$, but negatively affects HDL. *Height* has a strong positive effect on $\ln(weight)$ and this is typically anticipated for all adults. Smoking daily or some days is negatively associated with $\ln(weight)$. This outcome is not surprising since smoking is well known to reduce appetite. On the contrary, smoking seems to have no significant effect on HDL. Number of hours of sedentary activities is positively associated with $\ln(weight)$ and negatively associated with HDL. The result confirms the generally held belief that being inactive increases weight and is negatively associated with good cholesterol. Sleep disorder is also known to be positively associated with weight gain and this is confirmed in our findings but it has no effect on HDL. *LDL20T* has a positive (negative) effect on $\ln(weight)$ (HDL), which is expected since LDL is commonly referred as 'bad cholesterol' and is associated with weight gain. Transformed SBP has a positive effect on $\ln(weight)$, but a negative effect on HDL which is statistically different from zero at 90% probability level. As the first equation is a log-log specification, the coefficient of $\ln(SBP - 50)$ is an elasticity. Then, a 10% increase in the transformed SBP leads to a 0.967% growth of weight. The second equation is a semi-log specification, so the elasticity of HDL relative to the transformed SBP at the mean of the sample is: $0.1012/1.32 = 0.076$. A 10% increase in the transformed SBP leads to a 0.76% growth of HDL. The estimated correlation coefficients of the residuals between the two equations is $\rho_{\varepsilon_1 \varepsilon_2} = -0.304$. Inefficiency factors are close to 1 suggesting a good mix of the draws and the Geweke's CD ($CD \sim N(0,1)$) confirms that a sufficiently large number of draws have been taken.

The results from the Bayesian estimation of SURME model (which accounts for the measurement error in the covariate SBP) is presented in the upper panel of Table 7. A quick glance shows that the results for the covariates measured without error are similar to those in Table 6, except for the intercept and the male indicator in the $\ln(weight)$ equation. The $\ln(height)$ coefficient in the $\ln(weight)$ equation is now slightly higher but its 95% credible interval overlaps with that of the SUR model. Posterior estimates corresponding to transformed SBP increase in both equations (from 0.097 to 0.141 in the $\ln(weight)$ equation and from 0.101 to 0.152 in the HDL equation) leading to the following elasticities at the mean of the sample: 0.141 and 0.115($= 0.152/1.32$), for weight and HDL, respectively. However, as the posterior standard errors become larger (from 0.027 to 0.044 in the $\ln(weight)$ equation and from 0.056 to 0.092 in the HDL equation)—as in the Monte Carlo study—the 95% HPDI for the coefficient of $\ln(SBP - 50)$ overlap even if the distribution moves to the right when we go from SUR to SURME (see Figure 1). In particular, the 95% HPDI for the coefficient of $\ln(SBP - 50)$ are $[0.042; 0.153]$ and $[-0.008; 0.213]$ in the $\ln(weight)$ and the HDL equations, respectively, in the SUR model, and $[0.055; 0.229]$ and $[-0.024; 0.331]$ in the $\ln(weight)$ and the HDL equations in the SURME model. In the HDL equation, the coefficient of SBP is positive but statistically equivalent to zero. Posterior estimate of measurement error variance is 0.029, which leads to an estimated reliability ratio of about 59.78% and a noise-to-true variance ratio of 67.27%. The posterior variances of the disturbances from $\ln(weight)$ and HDL are close to those of SUR model and lead to an error correlation $\rho_{\varepsilon_1 \varepsilon_2} = -0.25$. Inefficiency factors are close to 1 and Geweke's CD confirms, for most parameters, that a sufficiently large number of draws has been taken.

The lower panel of Table 7 presents the results for the exposure equation from the Bayesian estimation of SURME model. In the first equation, only three variables have a positive effect on $\ln(SBP - 50)$: $\ln(age)$, LDL20T and $\ln(height)$ (only at the 10% probability level). In the second equation, four variables have a positive effect on $\ln(SBP - 50)$: $\ln(age)$, LDL20T, male and smokers (the last two variables are different from zero only at the 10% probability level).

We next estimate the SURME model using the MFVB approach, which takes 1565 cycles to get the maximum of the evidence lower bound l. The results, presented in the upper panel of Table 8, shows that in $\ln(weight)$ equation for the transformed SBP both the coefficient (0.159) and the probability interval ($[0.076; 0.242]$) are similar to those obtained from the Bayesian estimation of SURME model. In the HDL equation, the marginal effect at 0.186 is higher compared to the Bayesian estimate, with a 95% probability interval of $[0.028; 0.345]$. Taking into account measurement errors using MFVB allows to get significantly larger and more accurate elasticities of weight (0.159) and HDL ($0.1414 = (0.186/1.32)$) for the transformed SBP compared to the other method.

Posterior estimate of measurement error variance from the MFVB approach is 0.029, which is similar to the Bayesian estimate, and leads to an estimated reliability ratio of about 60.2% and to a noise-to-true variance ratio of 66.1%. The posterior variances of the disturbances from $\ln(weight)$ and HDL are close to the Bayesian estimates and lead to the same correlation between the errors of the two equations $\rho_{\varepsilon_1 \varepsilon_2} = -0.25$. The lower panel of Table 8 presents results for the exposure model estimated using with MFVB approximation method. In the first equation, four variables have a positive effect on $\ln(SBP - 50)$: $\ln(age)$, smokers, $LDL20T$ and $\ln(height)$. Similarly, in the second equation four variables have a positive effect on $\ln(SBP - 50)$: $\ln(age)$, $male$, $smokers$ and $LDL20T$. The exposure equation in the SURME model allows to define the implicit links between the true systolic blood pressure and the "risk factors" such as age, gender, smokers and "bad cholesterol" ($LDL20T$).

Figure 1 gives the posterior densities of the parameter corresponding to $\ln(SBP - 50)$ in the $\ln(weight)$ equation from the Bayesian estimation of SUR model, SURME model and the MFVB estimation of SURME model. We note a shift of the marginal effect of $\ln(SBP - 50)$ on $\ln(weight)$ to the right of the distribution from a mode established around 0.097 for SUR model to a mode of 0.141 for SURME model but with a wider dispersion. The estimated probability density function (pdf) with MFVB is slightly to the right and centered around the mode (0.159) but with a surface under the curve globally equivalent to that of Bayesian estimation of SURME model. In Figure 2, we observe similar shifts in the posterior density of the parameter corresponding to $\ln(SBP - 50)$ in the HDL equation, when we move from SUR (0.101) to SURME (0.152) or to MFVB (0.187) estimation of SURME model.

We note that similar to the Monte Carlo study, the MFVB approach to estimating the SURME model improves the results compared to those from Gibbs sampling. The results actually lend credibility to the proposed MFVB algorithm since coefficient estimates for variables which do not have measurement error are almost unaltered. However, when measurement error in SBP is ignored as in the SUR model, the posterior estimates are underestimated relative to the MFVB estimates. So, accounting for measurement error potentially corrects or reduces the bias in parameter estimates (see Carroll et al., 2006b).

6. Conclusion

The paper considers a SURME model (seemingly unrelated regression where some covariates have classical measurement error of the structural form) and introduces two novel estimation meth-

ods: a pure Bayesian algorithm based on MCMC and a second algorithm based on mean field variational Bayes approximation. The proposed algorithms use a prior distribution on measurement error variance to resolve identification issues in the model. In the MCMC estimation, Gibbs sampling is employed to sample the parameters from the conditional posterior distributions. While most of the conditional posterior densities have the standard form and are easily derived, the conditional posterior density for the true unobserved quantity associated with covariates having measurement error requires extensive attention to arrive at a manageable form. We also note that the proposed SURME model as explained is based on the structural form of measurement error, but the functional form of measurement error can be easily incorporated by introducing the distribution of the true unobserved quantity as a part of the subjective prior information. The expression for the joint and conditional posteriors will remain unchanged. However, estimating the SURME model using MCMC leads to high autocorrelation in the draws corresponding to the covariate measured with error. While this is easily dealt using *thinning*, the paper also proposes the MFVB approach as an alternative to get around the problem of high autocorrelation.

The proposed estimation algorithms are illustrated in multiple Monte Carlo simulation studies. While, the first set of 2 simulations (labeled Case I) investigate the effect on estimates by varying the variance of the true unobserved variable (for a fixed reliability ratio), the second set of 2 simulations examine the effect for a changing reliability ratio (for a fixed variance of the true unobserved variable). The results from all the simulations show that the Bayesian and MFVB estimation of SURME model reduce the biases to obtain satisfactory estimates as compared to estimates from SUR model. Moreover, the MFVB approach turns out as an excellent alternative to the MCMC and its poor mixing properties in the presence of latent variables. Besides, the MFVB approach has slightly better estimation accuracy and can be advantageous with large data sets.

The proposed models and techniques are also implemented in a health study where the two dependent variables, log of weight and high density lipoprotein (*HDL*), are regressed on a set of covariates measured without error and on systolic blood pressure (*SBP*) known to have measurement error. The model is estimated using the two algorithms and the results obtained reveal that the sign of the estimated coefficients are mostly consistent with what is typically found in the literature. Specifically, *SBP* has a positive effect on both $\ln(weight)$ and *HDL*, measurement error variance is small with an estimated reliability ratio of about 60% and a noise-to-true variance ratio of 66%. To offer a baseline comparison, a SUR model that ignores measurement error in *SBP* is also estimated using Gibbs sampling. Comparing the results across models, we see that posterior estimates for covariates without measurement error are almost identical, but that of *SBP* is lower and hence underestimated both in the weight and the *HDL* equations.

The combination of SUR and measurement error models is attractive and the proposed model can be generalized in several directions. One straightforward extension is the introduction of multiple covariates with measurement error in each SUR equation. However, the challenge here is to keep track of measurement errors arising from different covariates. The proposed SURME model can also be modified by introducing classical measurement error in the response variable or nonclassical measurement error models, where the errors may be correlated with the latent true values. Beyond the SUR models, these Bayesian approaches may be useful for measurement error in simultaneous equation models. We leave these possibilities for future research.

References

Ando, T. and Zellner, A. (2010), "Hierarchical Bayesian Analysis of the Seemingly Unrelated Regression and Simultaneous Equation Models using a Combination of Direct Monte Carlo and Importance Sampling Techniques," *Bayesian Analysis*, 5, 65–95.

Bishop, C. M. (2006), *Pattern Recognition and Machine Learning*, Springer, New York.

Blei, D. M., Kuckelbir, A., and McAuliffe, J. D. (2017), "Variational Inference: A Review for Statisticians," *Journal of the American Statistical Association*, 112, 859–877.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006a), *Measurement Error in Nonlinear Models: A Modern Perspective*, Chapman & Hall, Boca Raton.

Carroll, R. J., Midthune, D., Freedman, L. S., and Kipnis, V. (2006b), "Seemingly Unrelated Measurement Error Models with Application to Nutritional Epidemiology," *Biometrics*, 62, 75–84.

Casella, G. and George, E. I. (1992), "Explaining the Gibbs Sampler," *The American Statistician*, 46, 167–174.

Celeux, G., Forbes, F., Robert, C. P., and Titterington, D. M. (2006), "Deviance Information Criteria for Missing Data Models," *Bayesian Analysis*, 1, 651–674.

Chan, J. and Grant, A. D. (2016), "Fast Computation of the Deviance Information Criterion for Latent Variable Models," *Computational Statistics and Data Analysis*, 100, 847–859.

Cheng, C.-L. and Van Ness, J. W. (1999), *Statistical Regression with Measurement Error*, Arnold Publishers, London.

Cheng, C.-L., Shalabh, and Garg, G. (2014), "Coefficient of Determination for Multiple Measurement Error Models," *Journal of Multivariate Analysis*, 126, 137–152.

Chib, S. and Greenberg, E. (1995), "Hierarchical Analysis of SUR Models with Extension to Correlated Serial Errors and Time-Varying Parameter Models," *Journal of Econometrics*, 68, 339–360.

Faes, C., Ormerod, J. T., and Wand, M. P. (2011), "Variational Bayesian Inference for Parametric and Nonparametric Regression With Missing Data," *Journal of the American Statistical Association*, 106, 959–971.

Fiebig, D. G. (2001), "Seemingly Unrelated Regression," in *A Companion to Theoretical Econometrics*, ed. B. H. Baltagi, pp. 101–121, Blackwell Publishing, Massachusett.

Fuller, W. A. (1987), *Measurement Error Models*, John Wiley & Sons, New York.

Geman, S. and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Geyer, C. J. (1991), "Markov chain Monte Carlo Maximum Likelihood," in *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, ed. E. M. Kemramides, pp. 156–163, Interface Foundation of North America, Fairfax Station, VA, USA.

Giordano, R., Broderick, T., and Jordan, M. I. (2018), "Covariances, Robustness and Variational Bayes," *Journal of Machine Learning Research*, 19, 1–49.

Griffiths, W. E. and Chotikapanich, D. (1997), "Bayesian Methodology for Imposing Inequality Constraints on a Linear Expenditure System with Demographic Factors," *Australian Economic Papers*, 36, 321–341.

Griffiths, W. E. and Valenzuela, M. R. (2006), "Gibbs Samplers for a set of Seemingly Unrelated Regressions," *Australian & New Zealand Journal of Statistics*, 48, 335–351.

Hu, Y. and Wansbeek, T. (2017), "Measurement Error Models: Editor's Introduction," *Journal of Econometrics*, 200, 151–153.

Jeliazkov, I. (2013), "Nonparametric Vector Autoregressions: Specification, Estimation and Inference," *Advances in Econometrics*, 32, 327–359.

Kannel, W. B., Neaton, J. D., Wentworth, D., Thomas, H. E., Stamler, J., Hulley, S. B., and Kjelsberg, M. O. (1986), "Overall and Coronary Heart Disease Mortality Rates in Relation to Major Risk Factors in 325,348 men Screened for the MRFIT," *American Heart Journal*, 112, 825–836.

Koop, G., Poirier, D. J., and Tobias, J. (2005), "Semiparametric Bayesian Inference in Multiple Equation Models," *Journal of Applied Econometrics*, 20, 723–747.

Lee, C. Y. Y. and Wand, M. P. (2016), "Streamlined Mean Field Variational Bayesian Inference in Multiple Equation Models," *Biometrical Journal*, 58, 868–895.

Link, W. A. and Eaton, M. J. (2012), "On Thinning of Chains in MCMC," *Methods in Ecology and Evolution*, 3, 112–115.

Liu, J. S. (1994), "The Collapsed Gibbs Sampler in Bayesian Computations with Applications to a Gene Regulation Problem," *Journal of the American Statistical Association*, 89, 958–966.

Liu, X. and Liang, K.-Y. (1992), "Efficacy of Repeated Measures in Regression Models with Measurement Error," *Biometrics*, 48, 645–654.

MacEachern, S. N. and Berliner, L. M. (1994), "Subsampling the Gibbs Sampler," *The American Statistician*, 48, 188–190.

Ormerod, J. T. and Wand, M. P. (2010), "Explaining Variational Approximations," *The American Statistician*, 64, 140–153.

Owen, A. B. (2017), "Statistically Efficient Thinning of a Markov Chain Sampler," *Journal of Computational and Graphical Statistics*, 26, 738–744.

Percy, D. F. (1992), "Prediction for Seemingly Unrelated Regressions," *Journal of the Royal Statistical Society – Series B*, 54, 243–252.

Pham, T. H., Ormerod, J. T., and Wand, M. P. (2013), "Mean Field Variational Bayesian Inference for Nonparametric Regression with Measurement Error," *Computational Statistics and Data Analysis*, 68, 375–387.

Rao, C. R., Toutenburg, H., Shalabh, and Heumann, C. (2008), *Linear Models and Generalizations: Least Squares and Alternatives*, Springer, Berlin.

Shalabh (2003), "Consistent Estimation of Coefficients in Measurement Error Models with Replicated Observations," *Journal of Multivariate Analysis*, 86, 227–241.

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*, 64, 583–639.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2014), "The Deviance Information Criterion: 12 years on," *Journal of the Royal Statistical Society – Series B*, 76, 485–493.

Srivastava, V. K. and Dwivedi, T. D. (1979), "Estimation of Seemingly Unrelated Regression Equations," *Journal of Econometrics*, 10, 15–32.

Srivastava, V. K. and Giles, D. E. A. (1987), *Seemingly Unrelated Regression Equations Models: Estimation and Inference*, Marcel Dekker, New York.

Steel, M. F. J. (1992), "Posterior Analysis of Restricted Seemingly Unrelated Regression Equation Models: A Recursive Analytical Approach," *Econometric Reviews*, 11, 129–142.

Tao, G., Chen, Y., Wen, C., and Bi, M. (2011), "Statistical Analysis of Blood Pressure Measurement Errors by Oscillometry during Surgical Operations," *Blood Pressure Monitoring*, 16, 285–290.

van Dyk, D. A. and Park, T. (2008), "Partially Collapsed Gibbs Samplers: Theory and Methods," *Journal of the American Statistical Association*, 103, 790–796.

Wansbeek, T. and Meijer, E. (2000), *Measurement Error and Latent Variables in Econometrics*, Noth Holland, Amsterdam.

Yang, Y., Pati, D., and Bhattacharya, A. (2018), "α Variational Inference with Statistical Guarantees," *https://arxiv.org/abs/1710.03266*.

Zellner, A. (1962), "An Efficient Method of Estimating Seemingly Unrelated Regression and Tests for Aggregation Bias," *Journal of the American Statistical Association*, 57, 348–368.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, John Wiley & Sons, New York.

Zellner, A. and Ando, T. (2010), "A Direct Monte Carlo Approach for Bayesian Analysis of the Seemingly Unrelated Regression Model," *Journal of Econometrics*, 159, 33–45.

Table 1: Frequentist estimates of SUR Model. The table presents the true parameter values (TRUE), regression coefficients (COEF), relative error (RE), standard error (SE) and 95% confidence interval (INF, SUP). $N = 300$, $\sigma_Z^2 = 1$, $R_Z = 0.8$, Replications=100.

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ_1	γ_2	σ_{11}	σ_{12}	σ_{22}
TRUE	3	5	4	4	3.8	3	4	4	1	0.5	1
COEF	4.144	5.622	4.246	5.150	4.649	3.353	3.206	3.216	4.171	0.523	4.144
RE	0.381	0.124	0.062	0.288	0.223	0.118	-0.198	-0.196	3.171	0.046	3.144
SE	0.350	0.219	0.106	0.348	0.233	0.112	0.105	0.105			
INF	3.458	5.192	4.038	4.467	4.192	3.134	3.000	3.010			
SUP	4.830	6.051	4.455	5.833	5.105	3.572	3.412	3.422			

Table 2: Bayesian estimation of SURME model. The upper panel presents results for the main equations and the lower panel presents results for the exposure equations. MEAN is posterior mean, RE is relative error, STD is posterior standard deviation, IF is inefficiency factor, and (INF-HPDI, SUP-HPDI) represents 95% HPDI, and CD represents Geweke's convergence diagnostics. $N = 300$, $\sigma_Z^2 = 1$, $R_Z = 0.8$, Draws=51,000, Burnin draws=1,000, thinning=100, Replications=100.

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ_1	γ_2	σ_Z^2	σ_u^2	σ_{11}	σ_{12}	σ_{22}
TRUE	3	5	4	4	3.8	3	4	4	1	0.25	1	0.5	1
MEAN	2.895	4.858	3.986	3.716	3.694	2.999	4.083	4.103	0.974	0.251	1.025	0.504	1.015
RE	-0.035	-0.028	-0.004	-0.071	-0.028	-0.000	0.021	0.026	-0.026	0.005	0.025	0.008	0.015
STD	0.369	0.243	0.119	0.376	0.261	0.125	0.131	0.130	0.071	0.020	0.187	0.135	0.184
IF	1.029	1.028	1.023	1.055	1.066	1.050	1.117	1.232	1.032	1.011	1.011	1.004	1.013
INF-HPDI	2.280	4.450	3.786	3.086	3.257	2.790	3.870	3.893	0.862	0.220	0.752	0.297	0.746
SUP-HPDI	3.494	5.251	4.180	4.320	4.114	3.202	4.301	4.320	1.093	0.284	1.359	0.738	1.347
CD	0.990	0.980	1.000	0.970	0.950	0.980	0.990	0.950	0.960	0.990	0.990	0.960	0.970

	ω_{11}	ω_{12}	ω_{13}	ω_{21}	ω_{22}	ω_{23}
TRUE	1.5	0.75	0.3	1.5	1.05	0.45
MEAN	1.466	0.773	0.307	1.489	1.068	0.443
RE	-0.023	0.031	0.022	-0.007	0.017	-0.016
STD	0.165	0.109	0.055	0.166	0.109	0.055
IF	1.000	1.001	1.001	1.001	1.004	1.000
INF-HPDI	1.194	0.593	0.216	1.214	0.888	0.353
SUP-HPDI	1.735	0.952	0.396	1.758	1.246	0.533
CD	0.990	0.990	0.990	0.970	0.980	0.980

Table 3: MFVB estimation of SURME model. The upper panel presents results for the main equations and the lower panel presents results for the exposure equations. MEAN is posterior mean, RE is relative error, STD is posterior standard deviation, and (INF-CRI, SUP-CRI) represents 95% credible interval. $N = 300$, $\sigma_Z^2 = 1$, $R_Z = 0.8$, Replications=100, Cycles = 145.49, Maximum Elbo $= -3517.205$.

	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ_1	γ_2	σ_Z^2	σ_u^2	σ_{11}	σ_{12}	σ_{22}
TRUE	3	5	4	4	3.8	3	4	4	1	0.25	1	0.5	1
MEAN	2.968	4.908	4.006	3.889	3.823	3.054	4.027	3.990	1.004	0.241	1.089	0.534	1.098
RE	-0.011	-0.018	0.002	-0.028	0.006	0.018	0.007	-0.003	0.004	-0.035	0.089	0.067	0.098
STD	0.149	0.103	0.045	0.149	0.103	0.045	0.065	0.055	0.058	0.014	0.083	0.065	0.084
INF-CRI	2.677	4.707	3.918	3.596	3.621	2.965	3.898	3.882	0.890	0.214	0.926	0.405	0.934
SUP-CRI	3.259	5.110	4.095	4.181	4.026	3.143	4.155	4.097	1.118	0.269	1.251	0.662	1.262

	ω_{11}	ω_{12}	ω_{13}	ω_{21}	ω_{22}	ω_{23}
TRUE	1.5	0.75	0.3	1.5	1.05	0.45
MEAN	1.468	0.771	0.306	1.487	1.066	0.442
RE	-0.021	0.029	0.019	-0.009	0.015	-0.017
STD	0.151	0.099	0.050	0.151	0.099	0.050
INF-CRI	1.172	0.577	0.208	1.191	0.871	0.344
SUP-CRI	1.765	0.966	0.403	1.784	1.261	0.540

Table 4: Summary table showing relative errors of the parameters for all simulations from the frequentist estimation of SUR model, Bayesian estimation of SURME model, and MFVB estimation of SURME model.

		β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	γ_1	γ_2	σ_Z^2	σ_u^2	σ_{11}	σ_{12}	σ_{22}
	SUR	0.381	0.124	0.062	0.288	0.223	0.118	-0.198	-0.196			3.171	0.046	3.144
$(\sigma_Z^2 = 1, R_Z = 0.8)$	SURME	-0.035	-0.028	-0.004	-0.071	-0.028	0.000	0.021	0.026	-0.026	0.005	0.025	0.008	0.015
	MFVB	-0.011	-0.018	0.002	-0.028	0.006	0.018	0.007	-0.003	0.004	-0.035	0.089	0.067	0.098
	SUR	0.386	0.118	0.059	0.282	0.213	0.113	-0.195	-0.190			0.195	0.002	0.187
$(\sigma_Z^2 = 0.0625, R_Z = 0.8)$	SURME	-0.294	-0.096	-0.045	-0.256	-0.186	-0.099	0.150	0.168	-0.123	0.433	-0.079	-0.008	-0.090
	MFVB	-0.170	-0.056	-0.025	-0.111	-0.077	-0.039	0.086	0.070	-0.049	0.174	-0.026	0.011	-0.027
	SUR	0.834	0.263	0.130	0.627	0.481	0.255	-0.427	-0.426			6.796	0.088	6.793
$(\sigma_Z^2 = 1, R_Z = 0.5714)$	SURME	-0.065	-0.078	-0.013	-0.151	-0.070	-0.003	0.053	0.058	-0.081	0.010	0.023	0.005	0.020
	MFVB	-0.001	-0.046	0.004	-0.064	0.004	0.039	0.012	-0.002	0.000	-0.016	0.081	0.069	0.097
	SUR	0.844	0.256	0.128	0.625	0.468	0.251	-0.425	-0.419			0.424	0.000	0.413
$(\sigma_Z^2 = 0.0625, R_Z = 0.5714)$	SURME	-0.317	-0.107	-0.050	-0.283	-0.205	-0.107	0.164	0.184	-0.154	0.154	-0.070	-0.023	-0.087
	MFVB	-0.166	-0.057	-0.025	-0.109	-0.072	-0.035	0.085	0.066	-0.055	0.042	-0.019	0.003	-0.016

Table 5: Health application - variable definitions and data summary.

VARIABLE	DESCRIPTION	MEAN	STD
ln(*weight*)	Logarithm of weight (in kilograms).	4.40	0.24
HDL	High density lipoprotein (mmol/l, millimoles per litre).	1.32	0.45
ln(*age*)	Logarithm of age (in years).	4.00	0.15
Male	Indicator variable for male.	0.57	0.49
Smokers	Indicator variable for individuals who smoke daily or some days.	0.46	0.50
Sedentary	Number of hours of sedentary activities.	5.28	3.37
Sleep disorder	Indicator variable for sleep disorder problem.	0.11	0.32
LDL20T	Low density lipoprotein plus 20 percent of Triglyceride (mmol/l, millimoles per litre).	3.94	1.09
ln*(SBP-50)*	Transformation on systolic blood pressure.	4.28	0.24

Table 6: Health application - Bayesian estimation of SUR model. The table presents the posterior mean (MEAN), standard deviation (STD), inefficiency factor (IF), and Geweke's convergence diagnostics (CD) of the parameters. Draws=450,000, Burnin draws=50,000, thinning=100.

	ln(*weight*)				HDL			
	MEAN	STD	IF	CD	MEAN	STD	IF	CD
Intercept	-3.4956	0.5894	0.9850	0.5770	1.1074	0.3665	0.9995	0.1262
ln(*age*)	-0.0497	0.0443	1.0089	-0.3950	0.1035	0.0880	0.9869	0.4550
Male	0.0406	0.0154	0.9914	-0.8753	-0.2794	0.0264	0.9930	-1.0020
Smokers	-0.1008	0.0132	0.9908	0.5216	0.0424	0.0275	0.9870	1.0594
Sedentary	0.0085	0.0019	0.9821	-0.2897	-0.0076	0.0039	1.0069	-0.4709
Sleep disorder	0.1061	0.0202	1.0197	-0.1130	-0.0272	0.0412	0.9903	0.2408
LDL20T	0.0223	0.0058	1.0098	-0.4155	-0.1137	0.0119	1.0024	-0.1327
ln(*height*)	1.4731	0.1134	0.9949	-0.6200				
ln($SBP - 50$)	0.0967	0.0277	0.9963	0.8784	0.1012	0.0560	0.9873	-0.7712
σ_{11}	0.0399	0.0018	0.9900	-1.0924				
σ_{12}	-0.0250	0.0027	1.0219	0.4914				
σ_{22}	0.1693	0.0075	1	-0.5193				

Table 7: Health application - Bayesian estimation of SURME model. The upper panel presents results for the main equations and the lower panel presents results for the exposure equations. MEAN is posterior mean, STD is posterior standard deviation, IF is inefficiency factor, and CD is Geweke's convergence diagnostics. Draws=450,000, Burnin draws=50,000, thinning=100.

	$log(weight)$				HDL			
	MEAN	STD	IF	CD	MEAN	STD	IF	CD
Intercept	-5.2859	0.8284	1.0132	0.6022	1.1825	0.4332	1.7579	0.7772
$\ln(age)$	-0.0383	0.0521	1.1769	0.7601	0.0329	0.0997	1.4330	1.7639
$Male$	0.0156	0.0187	1.0181	-0.0596	-0.2807	0.0276	1.0156	0.7971
$Smokers$	-0.1009	0.0146	0.9998	0.8830	0.0375	0.0280	1.0110	1.7027
$Sedentary$	0.0080	0.0021	1.0279	0.8739	-0.0075	0.0040	1.0071	-0.1389
$Sleep\ disorder$	0.1066	0.0222	0.9863	0.8386	-0.0294	0.0425	0.9941	-0.0086
$LDL20T$	0.0224	0.0066	1.0159	1.4224	-0.1155	0.0126	1.0008	-0.1091
$\ln(height)$	1.7790	0.1561	1.0112	-0.4465				
$\ln(SBP-50)$	0.1416	0.0443	3.3261	-0.6761	0.1520	0.0921	3.8691	-1.1416
σ_{11}	0.0485	0.0022	0.9933	0.1713				
σ_{12}	-0.0252	0.0030	0.9829	-0.7013				
σ_{22}	0.1778	0.0080	1.0163	-0.3718				
σ_Z^2	0.0440	0.0019	1.0025	0.9275				
σ_u^2	0.0296	0.0016	1.0275	-1.4186				
	$\ln(SBP-50)$				$\ln(SBP-50)$			
	MEAN	STD	IF	CD	MEAN	STD	IF	CD
Intercept	1.1847	0.7160	1.0137	-0.0465	2.4526	0.2377	1.0053	0.8891
$\ln(age)$	0.4156	0.0577	1.0250	0.8542	0.4302	0.0580	1.0051	-0.8192
$Male$	0.0118	0.0201	0.9935	-0.0731	0.0318	0.0172	1.0165	-0.4137
$Smokers$	0.0341	0.0180	1.0020	0.6742	0.0363	0.0179	1.0280	-1.3321
$Sedentary$	-0.0018	0.0026	0.9709	0.1812	-0.0013	0.0025	0.9886	1.3253
$Sleep\ disorder$	0.0087	0.0277	0.9841	-0.2518	0.0094	0.0274	1.0220	0.8595
$LDL20T$	0.0185	0.0079	1.0111	-0.1604	0.0190	0.0080	1.0092	-0.8225
$\ln(height)$	0.2620	0.1405	1.0129	-0.2297				

Table 8: Health application - MFVB estimation of SURME model. Cycles=1565, Maximum Elbo $=-9076.16$. The upper panel presents results for the main equations and the lower panel presents results for the exposure equations. MEAN is posterior mean and STD is posterior standard deviation of the parameters.

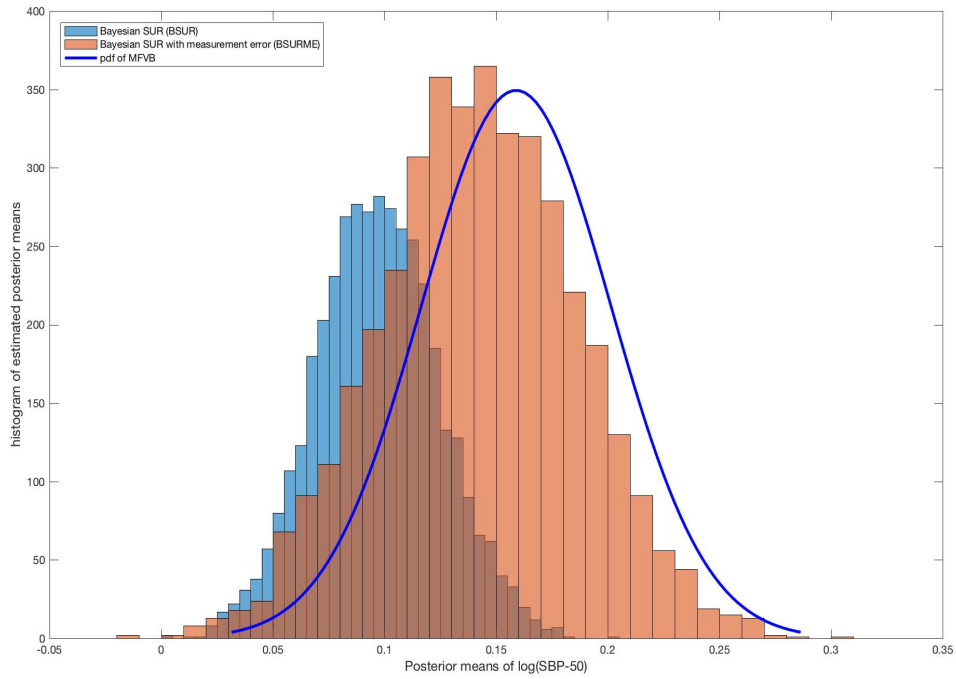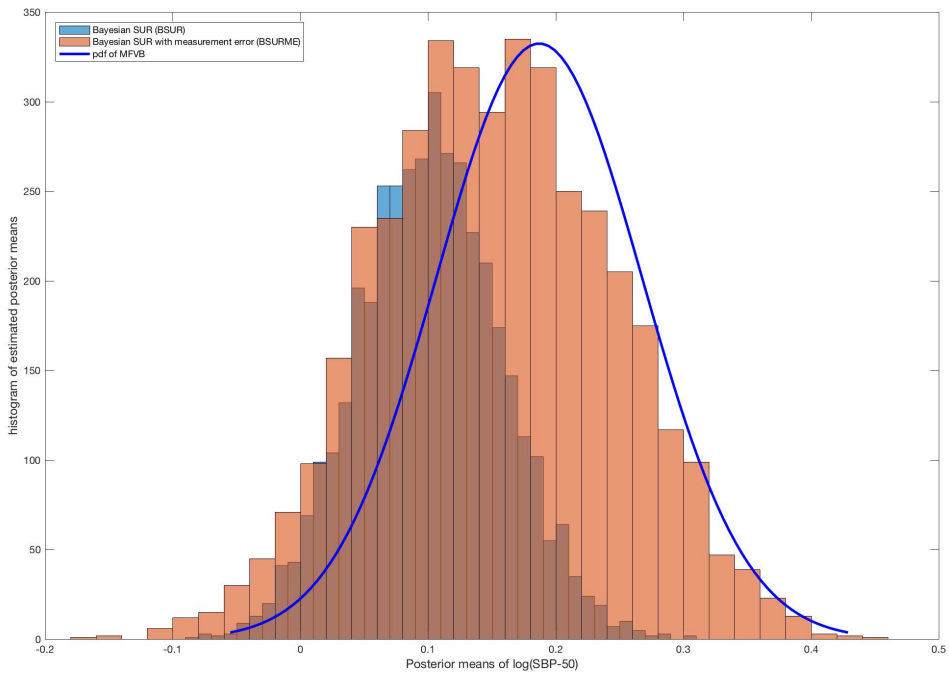|	$\ln(weight)$		HDL	
	MEAN	STD	MEAN	STD
Intercept	-5.3245	0.8069	1.0892	0.3726
$\ln(age)$	-0.0461	0.0477	0.0199	0.0908
$Male$	0.0148	0.0185	-0.2816	0.0272
$Smokers$	-0.1018	0.0144	0.0364	0.0276
$Sedentary$	0.0080	0.0021	-0.0075	0.0040
$Sleep\ disorder$	0.1066	0.0222	-0.0293	0.0425
$LDL20T$	0.0222	0.0064	-0.1160	0.0123
$\ln(height)$	1.7781	0.1537		
$\ln(SBP-50)$	0.1594	0.0422	0.1867	0.0808
σ_{11}	0.0484	0.0022		
σ_{12}	-0.0252	0.0030		
σ_{22}	0.1778	0.0079		
σ_Z^2	0.0443	0.0014		
σ_u^2	0.0293	0.0009		
	$\ln(SBP-50)$		$\ln(SBP-50)$	
	MEAN	STD	MEAN	STD
Intercept	1.1679	0.6340	2.4525	0.1843
$\ln(age)$	0.4153	0.0451	0.4305	0.0449
$Male$	0.0118	0.0165	0.0318	0.0136
$Smokers$	0.0340	0.0138	0.0362	0.0138
$Sedentary$	-0.0018	0.0020	-0.0012	0.0020
$Sleep\ disorder$	0.0090	0.0213	0.0095	0.0213
$LDL20T$	0.0186	0.0061	0.0187	0.0061
$\ln(height)$	0.2655	0.1228		

Figure 1: Posterior means of log(SBP-50) for log(weight) equation.

Figure 2: Posterior means of log(SBP-50) for log(HDL) equation.