Electronic Journal of Statistics
Vol. 18 (2024) 1495–1516
ISSN: 1935-7524
https://doi.org/10.1214/24-EJS2235

Geometric ergodicity of Gibbs samplers for Bayesian error-in-variable regression

Austin Brown[1]

[1] *Department of Statistical Sciences, University of Toronto, Toronto, ON,*
e-mail: ad.brown@utoronto.ca

Abstract: Multivariate Bayesian error-in-variable (EIV) linear regression is considered to account for additional additive Gaussian error in the features and response. A 3-variable deterministic scan Gibbs sampler is constructed for multivariate EIV regression models using classical and Berkson errors with independent normal and inverse-Wishart priors. These Gibbs samplers are proven to always be geometrically ergodic which ensures a central limit theorem for many time averages from the Markov chains. We demonstrate the strengths and limitations of the Gibbs sampler with simulated data for large data problems, robustness to misspecification and also analyze a real-data example in astrophysics.

MSC2020 subject classifications: Primary 60J22; secondary 62F15.
Keywords and phrases: Bayesian statistics, convergence analysis, error-in-variable models, Gibbs sampling, Markov chain Monte Carlo, measurement error models.

Received April 2023.

Contents

1. Introduction

Many problems in astrophysics (Feigelson and Babu, 1992; Hilbe, de Souza and Ishida, 2017; Kelly, 2012; Stefanski, 2000) and epidemiology (Achic et al., 2018; Buonaccorsi, 2010; Carroll et al., 2006; Clayton et al., 1992) among other

arXiv: 2209.08301

areas of science (Groß, 2016; Pollice et al., 2019; Tang, Li and Tang, 2017) involve error in variables (EIV) which classical linear regression does not take into account. EIV can occur in many situations such as measurement error in data collection (Hilbe, de Souza and Ishida, 2017; Kelly, 2012), discrepancies between the data distribution and the model (Carroll et al., 2006; Buonaccorsi, 2010), or purposeful adversarial attacks against the data (Goodfellow, Shlens and Szegedy, 2015; Szegedy et al., 2014). Not surprisingly, multiple critical issues arise in parameter estimation and statistical inference when ignoring additional errors in the data such as poor predictive performance (Goodfellow, Shlens and Szegedy, 2015), statistical bias (Damgaard, 2020; Kröger, Hoffmann and Pakpahan, 2016; Vidal and Iglesias, 2008), and estimators fail to be consistent (Michalek and Tripathi, 1980).

Bayesian approaches develop a strategy for additional error in the variables by constructing a new model incorporating additional error. We consider multivariate Bayesian EIV linear regression (Charisse Farr et al., 2020; Dellaportas and Stephens, 1995; Fang et al., 2017; Huang, 2010; Mallick and Gelfand, 1996; Muff et al., 2015; Richardson S, 1993; Rodrigues and Bolfarine, 2007; Torabi et al., 2021; Vidal and Arellano-Valle, 2010) accounting for additive Gaussian error in the features (covariates) and response. This model assumes the variability of the additive Gaussian error is known beforehand which has been successful in astrophysics applications in the presence of known instrumentation error (Hilbe, de Souza and Ishida, 2017; Kelly, 2012). Alternative approaches to EIV models attempt to correct existing parameter estimation methods such as least squares or method of moments with weighting and other techniques (Fuller, 1987; Stefanski and Carroll, 1985). Several other strategies for EIV modeling are discussed in more comprehensive treatments on the topic (Buonaccorsi, 2010; Carroll et al., 2006; Fuller, 1987).

We write $x \sim N_d(m, C)$ to mean the d-dimensional normal distribution with mean m and symmetric, positive-definite (SPD) covariance matrix C. We also write $x \sim \mathcal{W}_d^{-1}(\nu, V)$ to be the inverse-Wishart distribution with positive integer degrees of freedom $\nu \geq d$ and scale SPD matrix $V \in \mathbb{R}^{d \times d}$. Let $\text{vec}(A)$ denote the vectorization of a matrix A by stacking the columns. Let $(Y_i, X_i, Z_i)_{i=1}^n$ be independent and identically distributed (i.i.d.) where the response Y_i takes values in $\mathbb{R}^m$ along with features X_i taking values in $\mathbb{R}^p$ and fixed, known features $Z_i \in \mathbb{R}^q$ where m, n, p, q are positive integers. Let $\theta = \text{vec}(\Theta) \in \mathbb{R}^{qm}$, $\beta = \text{vec}(\mathcal{B}) \in \mathbb{R}^{pm}$, and SPD matrix $\Sigma \in \mathbb{R}^{m \times m}$ be unknown regression and covariance parameters respectively. We introduce new parameters $\mathcal{A} = (\mathcal{A}_1, \ldots, \mathcal{A}_n)^T$ with $\mathcal{A}_i \in \mathbb{R}^p$ to model additional error in X_i using classical or Berkson errors (Berkson, 1950). The classical error model specifies $X_i|\mathcal{A}_i$ and the Berkson error model (Berkson, 1950) assumes instead a data-dependent prior on $\mathcal{A}_i|X_i$. When there is additional error in X_i, the EIV linear regression model for $i \in 1, \ldots, n$ is i.i.d. with

$$Y_i|\mathcal{A}_i, \theta, \beta, \Sigma \sim N_m(\Theta^T Z_i + \mathcal{B}^T \mathcal{A}_i, \Sigma) \tag{1a}$$

$$X_i|\mathcal{A}_i \sim N_p(\mathcal{A}_i, V_i) \text{ (Classical)} \quad \text{or} \quad \mathcal{A}_i|X_i \sim N_p(X_i, V_i) \text{ (Berkson)} \tag{1b}$$

where the SPD matrices $V_i \in \mathbb{R}^{p \times p}$ are known. When there is also additional error in the responses Y_i, we assume an i.i.d. hierarchical regression model with

$$Y_i | \mathcal{V}_i \sim N_m(\mathcal{V}_i, U_i) \tag{2a}$$

$$\mathcal{V}_i | \mathcal{A}_i, \theta, \beta, \Sigma \sim N_m(\Theta^T Z_i + \mathcal{B}^T \mathcal{A}_i, \Sigma) \tag{2b}$$

$$X_i | \mathcal{A}_i \sim N_p(\mathcal{A}_i, V_i) \text{ (Classical)} \quad \text{or} \quad \mathcal{A}_i | X_i \sim N_p(X_i, V_i) \text{ (Berkson)} \tag{2c}$$

where $U_i \in \mathbb{R}^{m \times m}$ are known SPD matrices.

We will be interested in the posterior for both models (1) and (2) using independent normal and inverse-Wishart priors on the parameters $(\mathcal{A}, \theta, \beta, \Sigma)$. The independent prior choice is a popular choice in Bayesian regression models with and without measurement error (Carroll et al., 2006; Dellaportas and Stephens, 1995; Ekvall and Jones, 2021; Rajaratnam and Sparks, 2015). For example, in the special case of a univariate response, this classical Bayesian EIV model has been used in astrophysics to study supermassive black hole mass (Harris, Poole and Harris, 2014; Hilbe, de Souza and Ishida, 2017). However, the general multivariate Bayesian EIV regression models (1) and (2) with these priors have not been previously introduced to the best of our knowledge. For the EIV regression models (1) and (2), the independent priors are chosen

$$\Sigma \sim \mathcal{W}_m^{-1}(a_0, B_0) \tag{3a}$$

$$(\theta, \beta)^T \sim N_{m(q+p)}(j_0, J_0) \tag{3b}$$

where $a_0 \geq m$ is a positive integer, $B_0 \in \mathbb{R}^{m \times m}$ is a SPD matrix, $j_0 \in \mathbb{R}^{m(q+p)}$ and SPD matrix $J_0 \in \mathbb{R}^{m(q+p) \times m(q+p)}$. The classical and Berkson error models assume either

$$\mathcal{A}_i \sim N_p(k_i, K_i) \text{ (Classical)} \quad \text{or} \quad \mathcal{A}_i \text{ flat prior (Berkson)} \tag{4}$$

where $k_i \in \mathbb{R}^p$ and $K_i \in \mathbb{R}^{p \times p}$ are SPD matrices. For example, an *exposure model* (Gustafson, 2003) utilized often in epidemiology would assume classical errors with a data-dependent prior on each $\mathcal{A}_i$ depending on Z_i. In the Berkson error model, each $\mathcal{A}_i | X_i$ is already specified and it is natural to assume an improper flat prior on each $\mathcal{A}_i$.

Previous work has proposed Gibbs sampling (Geman and Geman, 1984) to draw samples from the posterior, denoted by Π_n, in Bayesian EIV regression models (Bhadra and Carroll, 2016; Carroll et al., 2006; Dellaportas and Stephens, 1995; Richardson S, 1993). However, trustworthy estimation from a Gibbs sampler requires the Markov chain to converge to the posterior distribution at a sufficiently fast rate. Consider a vector-valued function f with $\int \|f\|^{2+\delta} d\Pi_n < \infty$ for some $\delta \in (0, \infty)$ and denote $\bar{f}_m$ as the time average of m samples from the Gibbs sampler. In order to be confident in the estimator $\bar{f}_m$ in applications an estimate of the Monte Carlo simulation error is essential for constructing standard errors and confidence intervals for each coordinate. A Gibbs sampler is geometrically ergodic if initialized at points, its marginal distribution is converging to Π_n at an exponential rate in total variation. Geometrically ergodic Gibbs samplers provide rich theoretical guarantees which are of practical relevance in applications. These Gibbs samplers satisfy a central limit theorem (Chan and Geyer, 1994; Jones, 2004), that is,

$$\sqrt{m}\left(\bar{f}_m - \int f d\Pi_n\right)$$

is asymptotically normally distributed and under suitable assumptions, the covariance in this normal distribution can be consistently estimated (Vats, Flegal and Jones, 2019a). Further pertinent tools to ensuring reliable estimation such as estimates of the effective sample size, consistent confidence ellipsoids, and consistent confidence intervals for quantile estimation are also available (Doss et al., 2014; Vats, Flegal and Jones, 2019a).

To the best of our knowledge, the rate of convergence for Gibbs sampling in EIV regression models has not been previously investigated. Related approaches have instead proposed variational Bayesian methods (Bresson et al., 2021; Pham, Ormerod and Wand, 2013) and the integrated nested Laplace approximation (INLA) (Håvard Rue, 2009; Muff et al., 2015). We construct a general density which in special cases, is the posterior for the 4 Bayesian EIV regression models (1) and (2) using the independent normal and inverse-Wishart prior choice on the parameters (3) and (4). Our main contribution constructs a 3-variable deterministic scan Gibbs sampler for this general density, and we show it is *always* geometrically ergodic using a drift and minorization condition (Hairer and Mattingly, 2011; Meyn and Tweedie, 2009). Since we develop the 3-variable Gibbs sampler generally, the sampler may have applications beyond the Bayesian EIV models introduced here. The 3-variable Gibbs sampler we construct can be simulated efficiently on a computer without the need for complex Metropolis-Hastings or rejection sampling steps at each iteration. In particular, this analysis provides the first geometrically converging Gibbs sampler applicable to existing Bayesian EIV models used in astrophysics.

The organization of this paper is as follows. In Section 2, we construct a general EIV regression density and construct a 3-variable Gibbs sampler for this density. We show the Gibbs sampler is always geometrically ergodic and apply this to the 4 multivariate Bayesian EIV regression models presented in this introduction. Section 3 studies the algorithm empirically where we demonstrate limitations of the Gibbs sampler with simulated data for large data problems and also the behavior of the Gibbs sampler under model misspecification. Section 4 studies a real-data example in astrophysics to study supermassive black hole mass (Harris, Poole and Harris, 2014; Hilbe, de Souza and Ishida, 2017). Finally in Section 5, we discuss our results and future research directions.

2. General Gibbs Sampler for EIV regression

For positive integers p, define p-norms by $\|\cdot\|_p$ and the Frobenius norm by $\|\cdot\|_F$. Let $\otimes$ denote the Kronecker product. The posteriors for the Bayesian EIV regression models (1) and (2) using independent prior choices (3) and (4) for both classical and Berkson errors share a common general form which we study in this section. The posterior densities for these Bayesian EIV regression models are special cases of the density (5) but will differ depending on the EIV modeling choice illustrated in the subsequent sections. For $i \in 1, \dots, n$, define hyperparameters $a_0 \in (0, \infty)$, SPD matrix $B_0 \in \mathbb{R}^{m \times m}$, $c_0 \in \mathbb{R}^{m(p+q)}$, SPD matrices $C_0 \in \mathbb{R}^{m(p+q) \times m(p+q)}$, $D_i \in \mathbb{R}^{p \times p}$, and $d_i \in \mathbb{R}^p$, $R = (R_1, \dots, R_n)^T \in \mathbb{R}^{n \times m}$, and $M \in \mathbb{R}^{n \times q}$. For $\mathcal{A} \in \mathbb{R}^{n \times p}$, $\theta = \text{vec}(\Theta) \in \mathbb{R}^{mq}$, $\beta = \text{vec}(\mathcal{B}) \in \mathbb{R}^{mp}$, SPD matrix $\Sigma \in \mathbb{R}^{m \times m}$, define the density

$$\pi_n(\mathcal{A}, \theta, \beta, \Sigma) \tag{5a}$$

$$\propto \left(\frac{1}{\det(\Sigma)}\right)^{(n+a_0+1+m)/2} \exp\left[-\frac{1}{2}\text{tr}(\Sigma^{-1}B_0)\right] \tag{5b}$$

$$\times \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(R_i - \Theta^T M_i - \mathcal{B}^T\mathcal{A}_i)^T\Sigma^{-1}(R_i - \Theta^T M_i - \mathcal{B}^T\mathcal{A}_i)\right] \tag{5c}$$

$$\times \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\mathcal{A}_i - d_i)^T D_i^{-1}(\mathcal{A}_i - d_i)\right) \tag{5d}$$

$$\times \exp\left(-\frac{1}{2}((\theta, \beta)^T - c_0)^T C_0^{-1}((\theta, \beta)^T - c_0)\right). \tag{5e}$$

Observe that since (5c) is upper bounded by 1, then this probability density is properly defined. When properly normalized, the distribution corresponding to the densities (5d), (5e), and (5b) have moments of all orders. Since these upper bound π_n, this implies Π_n also has moments of all orders. In particular if we are able to construct a geometrically ergodic Gibbs sampler which has Π_n as its invariant distribution, then the Gibbs sampler will satisfy the Markov chain central limit theorem for a large class of practically relevant functions used in applications.

We will construct a 3-variable deterministic scan Gibbs sampler using a specific update order for the density (5). We also derive the conditional densities for the Gibbs sampler which can be sampled directly. Initialize $(\theta_0, \beta_0, \Sigma_0)$ and $\mathcal{A}_0 = (\mathcal{A}_{1,0}, \dots, \mathcal{A}_{n,0})$ from an initial distribution. For $t \in 1, \dots,$ first generate

$$\Sigma_t | \mathcal{A}_{t-1}, \theta_{t-1}, \beta_{t-1}$$
$$\sim \mathcal{W}_m^{-1}\left(n + a_0, (R - M\Theta_{t-1} - \mathcal{A}_{t-1}\mathcal{B}_{t-1})^T(R - M\Theta_{t-1} - \mathcal{A}_{t-1}\mathcal{B}_{t-1}) + B_0\right)$$

Next, generate $(\theta_t, \beta_t)^T | \mathcal{A}_{t-1}, \Sigma_t \sim N_{m(p+q)}(c_n(\mathcal{A}_{t-1}, \Sigma_t), C_n(\mathcal{A}_{t-1}, \Sigma_t))$ where

$$C_n(\mathcal{A}_{t-1}, \Sigma_t) = \left(\Sigma_t^{-1} \otimes \begin{pmatrix} M & \mathcal{A}_{t-1} \end{pmatrix}^T \begin{pmatrix} M & \mathcal{A}_{t-1} \end{pmatrix} + C_0^{-1}\right)^{-1}$$
$$c_n(\mathcal{A}_{t-1}, \Sigma_t) = C_n(\mathcal{A}_{t-1}, \Sigma_t)\left(\left[\Sigma_t^{-1} \otimes \begin{pmatrix} M & \mathcal{A}_{t-1} \end{pmatrix}^T\right] \text{vec}(R) + C_0^{-1}c_0\right).$$

Finally, generate independently

$$\mathcal{A}_{i,t} | \theta_t, \beta_t, \Sigma_t \sim N_p(d_{n,i}(\theta_t, \beta_t, \Sigma_t), D_{n,i}(\theta_t, \beta_t, \Sigma_t))$$

where

$$D_{n,i}(\beta_t, \Sigma_t) = \left(\mathcal{B}_t \Sigma_t^{-1} \mathcal{B}_t^T + D_i^{-1}\right)^{-1}$$
$$d_{n,i}(\theta_t, \beta_t, \Sigma_t) = D_{n,i}(\beta_t, \Sigma_t)\left[D_i^{-1} d_i + \mathcal{B}_t \Sigma_t^{-1}\left(R_i - \Theta_t^T M_i\right)\right]$$

to obtain $\mathcal{A}_t = (\mathcal{A}_{1,t}, \dots, \mathcal{A}_{n,t})^T$.

For points $(\mathcal{A}, \theta, \beta, \Sigma)$ and $(\mathcal{A}', \theta', \beta', \Sigma')$, the Gibbs sampler has Markov transition density

$$p\left((\mathcal{A}, \theta, \beta, \Sigma), (\mathcal{A}', \theta', \beta', \Sigma')\right) = \pi_n(\mathcal{A}'|\theta', \beta', \Sigma')\pi_n(\theta', \beta'|\mathcal{A}, \Sigma')\pi_n(\Sigma'|\mathcal{A}, \theta, \beta)$$

and Markov transition kernel defined for suitable sets B by

$$P\left((\mathcal{A}, \theta, \beta, \Sigma), B\right) = \int\int\int_B p\left((\mathcal{A}, \theta, \beta, \Sigma), (\mathcal{A}', \theta', \beta', \Sigma')\right) d\mathcal{A}' d\theta' d\beta' d\Sigma'.$$

The Markov kernel at larger iteration times $t \geq 2$ is defined recursively with $P^1 \equiv P$ by

$$P^t\left((\mathcal{A}, \theta, \beta, \Sigma), B\right) = \int P^{t-1}(\cdot, B)\, dP\left((\mathcal{A}, \theta, \beta, \Sigma), \cdot\right).$$

We will use the following drift function defined by

$$V(\mathcal{A}, \theta, \beta) = \frac{1}{2}\sum_{i=1}^{n}(\mathcal{A}_i - d_i)^T D_i^{-1}(\mathcal{A}_i - d_i) + \frac{1}{2}(\theta, \beta) C_0^{-1}(\theta, \beta)^T$$

combined with a minorization condition to show there is a $\rho \in (0,1)$ and $M_0 \in (0, \infty)$ so that for any initialization $\mathcal{A}, \theta, \beta, \Sigma$,

$$\sup_{|\varphi| \leq 1 + M_0 V}\left|\int \varphi dP^t\left((\mathcal{A}, \theta, \beta, \Sigma), \cdot\right) - \int \varphi d\Pi_n\right| \leq M(\mathcal{A}, \theta, \beta)\rho^t \qquad (6)$$

where $M(\mathcal{A}, \theta, \beta) = 2 + M_0 V(\mathcal{A}, \theta, \beta) + M_0 \int V d\Pi_n$ (Hairer and Mattingly, 2011). The condition (6) implies the Gibbs sampler is geometrically ergodic. We now state our main result.

Theorem 2.1. *The 3-variable deterministic scan Gibbs sampler* $(\mathcal{A}_t, \theta_t, \beta_t, \Sigma_t)_{t=0}^{\infty}$ *for the general density* (5) *is geometrically ergodic.*

Proof. Using a special property of the Gibbs sampler, it will be sufficient to develop a drift and minorization condition based only on the marginal chain $(\mathcal{A}_t, \theta_t, \beta_t)_t$ (Roberts and Rosenthal, 2001, Example 3.6). In particular, we will use the special property of this Gibbs Markov kernel P that for suitable sets B, $P(\cdot, B)$ is a function of only the parameters $(\mathcal{A}, \theta, \beta)$ and does not depend on Σ. We first show a minorization condition. For $\ell \in (0, \infty)$, define the function g_ℓ by

$$g_\ell(\mathcal{A}', \theta', \beta', \Sigma') = \inf_{V(\mathcal{A}, \theta, \beta) \leq \ell} \pi_n(\mathcal{A}'|\theta', \beta', \Sigma')\pi_n(\theta', \beta'|\mathcal{A}, \Sigma')\pi_n(\Sigma'|\mathcal{A}, \theta, \beta)$$

and the constant $Z_{g_\ell} = \int g_\ell(\mathcal{A}', \theta', \beta', \Sigma') d\mathcal{A}' d\theta' d\beta' d\Sigma'$. The drift function V is a continuous, strongly convex function on a closed, convex domain so its sublevel sets are closed and bounded (Nesterov, 2018, Corollary 3.2.2). For fixed θ', β', Σ', the function

$$(\mathcal{A}, \theta, \beta) \mapsto \pi_n(\theta', \beta' | \mathcal{A}, \Sigma') \pi_n(\Sigma' | \mathcal{A}, \theta, \beta)$$

is continuous and achieves its minimum over compact sets. Thus, Z_{g_ℓ} is not 0 and we can define the probability measure for suitable sets B by

$$\nu_\ell(B) = Z_{g_\ell}^{-1} \int_B g_\ell(\mathcal{A}', \theta', \beta', \Sigma') d\mathcal{A}' d\theta' d\beta' d\Sigma'.$$

For any $\ell \in (0, \infty)$ and any suitable set B,

$$\begin{aligned}
&\inf_{\substack{\Sigma \in \mathbb{R}^{m \times m}, \\ V(\mathcal{A}, \theta, \beta) \leq \ell}} P\left((\mathcal{A}, \theta, \beta, \Sigma), B\right) \\
&= \inf_{V(\mathcal{A}, \theta, \beta) \leq \ell} \int_B \pi_n(\mathcal{A}' | \theta', \beta', \Sigma') \pi_n(\theta', \beta' | \mathcal{A}, \Sigma') \pi_n(\Sigma' | \mathcal{A}, \theta, \beta) d\mathcal{A}' d\theta' d\beta' d\Sigma' \\
&\geq \int_B g_\ell(\mathcal{A}', \theta', \beta', \Sigma') d\mathcal{A}' d\theta' d\beta' d\Sigma' \\
&= Z_{g_\ell} \nu_\ell(B).
\end{aligned}$$

This completes the minorization condition.

It remains to show a drift condition. Fix $\mathcal{A}_0, \theta_0, \beta_0$, and fix $i \in 1, \ldots, n$. Since D_i is SPD, let $D_i = D_i^{1/2} D_i^{1/2}$, $D_i^{-1} = D_i^{-1/2} D_i^{-1/2}$ where $D_i^{1/2}, D_i^{-1/2}$ are SPD. Using the identity

$$d_{n,i}(\theta, \beta, \Sigma) = d_i + \left(\mathcal{B} \Sigma^{-1} \mathcal{B}^T + D_i^{-1}\right)^{-1} \mathcal{B} \Sigma^{-1} \left(R_i - \Theta^T M_i - \mathcal{B}^T d_i\right)$$

and taking the expectation with respect to $\mathcal{A}_i | \mathcal{A}_0, \theta_0, \beta_0, \theta, \beta, \Sigma$

$$\mathbb{E}\left[\frac{1}{2} \left\| D_i^{-1/2} (\mathcal{A}_i - d_i) \right\|_2^2 | \mathcal{A}_0, \theta_0, \beta_0, \theta, \beta, \Sigma\right] \tag{7a}$$

$$= \frac{1}{2} \left\| D_i^{-1/2} \left(\mathcal{B} \Sigma^{-1} \mathcal{B}^T + D_i^{-1}\right)^{-1} \mathcal{B} \Sigma^{-1} \left(R_i - \Theta^T M_i - \mathcal{B}^T d_i\right) \right\|_2^2 \tag{7b}$$

$$+ \frac{1}{2} \mathrm{tr}\left[D_i^{-1/2} \left(\mathcal{B} \Sigma^{-1} \mathcal{B}^T + D_i^{-1}\right)^{-1} D_i^{-1/2}\right]. \tag{7c}$$

Using singular value decomposition from (Horn and Johnson, 2012, Theorem 2.6.3), choose matrices $U_i \in \mathbb{R}^{p \times p}$, $V_i \in \mathbb{R}^{m \times m}$ with $U_i^T U_i = U_i U_i^T = I_p$ and $V_i^T V_i = V_i V_i^T = I_m$ and a rectangular diagonal matrix $\Sigma_i \in \mathbb{R}^{p \times m}$ with diagonal nonnegative singular values $(\sigma_{i,k})_k$ so that $D_i^{1/2} \mathcal{B} \Sigma^{-1/2} = U_i \Sigma_i V_i^T$. Then

$$\begin{aligned}
\left(\mathcal{B} \Sigma^{-1} \mathcal{B}^T + D_i^{-1}\right)^{-1} &= D_i^{1/2} \left[(D_i^{1/2} \mathcal{B} \Sigma^{-1/2})(D_i^{1/2} \mathcal{B} \Sigma^{-1/2})^T + I_p\right]^{-1} D_i^{1/2} \\
&= D_i^{1/2} U_i [\Sigma_i \Sigma_i^T + I_p]^{-1} U_i^T D_i^{1/2}.
\end{aligned} \tag{8}$$

Using (8) and properties of the trace

$$\frac{1}{2}\mathrm{tr}\left[D_i^{-1/2}\left(\mathcal{B}\Sigma^{-1}\mathcal{B}^T + D_i^{-1}\right)^{-1} D_i^{-1/2}\right] = \frac{1}{2}\mathrm{tr}\left[\left(\Sigma_i\Sigma_i^T + I_p\right)^{-1} U_i^T U_i\right] \leq \frac{p}{2}.$$

For $x \in [0, \infty)$ and $a \in (0, \infty)$, we have the inequality

$$\frac{x}{x^2 + a} \leq \frac{1}{2\sqrt{a}}. \tag{9}$$

Using inequalities (8) and (9), the matrix norm is sub-multiplicative, and $\|U_i\|_2 = 1$, we have

$$\begin{aligned}
&\left\|D_i^{-1/2}\left(\mathcal{B}\Sigma^{-1}\mathcal{B}^T + D_i^{-1}\right)^{-1}\mathcal{B}\Sigma^{-1/2}\Sigma^{-1/2}\right\|_2^2 \\
&= \left\|U_i(\Sigma_i\Sigma_i^T + I_p)^{-1}U_i^T D_i^{1/2}\mathcal{B}\Sigma^{-1/2}\Sigma^{-1/2}\right\|_2^2 \\
&\leq \left\|(\Sigma_i\Sigma_i^T + I_p)^{-1}\Sigma_i\Sigma^{-1/2}\right\|_2^2 \\
&\leq \left[\frac{\sigma_i}{\sigma_i^2 + 1}\right]^2 \left\|\Sigma^{-1/2}\right\|_2^2 \\
&\leq \frac{\left\|\Sigma^{-1}\right\|_2}{4}.
\end{aligned}$$

Define the matrix $\tilde{X} = (d_1, \ldots, d_n)^T$. Applying these upper bounds to (7) and combining for each $i \in 1, \ldots, n$,

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{2}\sum_{i=1}^n \left\|D_i^{-1/2}(\mathcal{A}_i - d_i)\right\|_2^2 \middle| \mathcal{A}_0, \theta_0, \beta_0, \theta, \beta, \Sigma\right] \\
&\leq \frac{\left\|\Sigma^{-1}\right\|_2}{8}\left\|R - M\Theta - \tilde{X}\mathcal{B}\right\|_F^2 + \frac{pn}{2}.
\end{aligned}$$

By convexity, for every x, y, $\|x - y\|_2^2 \leq 2\|x\|_2^2 + 2\|y\|_2^2$. Since C_0 is SPD, let $C_0 = C_0^{1/2}C_0^{1/2}$, $C_0^{-1} = C_0^{-1/2}C_0^{-1/2}$ where $C_0^{1/2}, C_0^{-1/2}$ are SPD. Using convexity, and the matrix norm is sub-multiplicative, we have

$$\frac{1}{2}\left\|R - M\Theta - \tilde{X}\mathcal{B}\right\|_F^2 \leq \|R\|_F^2 + \left\|\begin{pmatrix} M & \tilde{X}\end{pmatrix} C_0^{1/2}\right\|_2^2 \left\|C_0^{-1/2}(\theta, \beta)^T\right\|_2^2.$$

Therefore,

$$\mathbb{E}\left[\frac{1}{2}\sum_{i=1}^n \left\|D_i^{-1/2}(\mathcal{A}_i - d_i)\right\|_2^2 \middle| \mathcal{A}_0, \theta_0, \beta_0, \theta, \beta, \Sigma\right] \tag{10a}$$

$$\leq \frac{\left\|\Sigma^{-1}\right\|_2}{8}\left\|R - M\Theta - \tilde{X}\mathcal{B}\right\|_F^2 + \frac{pn}{2} \tag{10b}$$

$$\leq \left\|\Sigma^{-1}\right\|_2 \frac{\|R\|_F^2}{4} + \left\|\Sigma^{-1}\right\|_2 \frac{\left\|\begin{pmatrix} M & \tilde{X} \end{pmatrix} C_0^{1/2}\right\|_2^2}{4} \left\|C_0^{-1/2}(\theta,\beta)^T\right\|_2^2 + \frac{pn}{2}. \quad (10c)$$

Now taking the expectation with respect to $\theta, \beta | \mathcal{A}_0, \theta_0, \beta_0, \Sigma$

$$\begin{aligned} &\mathbb{E}\left[\frac{1}{2}\left\|C_0^{-1/2}(\theta,\beta)^T\right\|_2^2 | \Sigma, \mathcal{A}_0, \theta_0, \beta_0\right] \\ &= \frac{1}{2}\left\|C_0^{-1/2} c_n(\mathcal{A}_0, \Sigma)\right\|_2^2 + \frac{1}{2}\mathrm{tr}(C_0^{-1/2} C_n(\mathcal{A}_0,\Sigma) C_0^{-1/2}). \end{aligned}$$

Using singular value decomposition (Horn and Johnson, 2012, Theorem 2.6.3), choose matrices $U \in \mathbb{R}^{mn \times mn}$, $V \in \mathbb{R}^{m(p+q)\times m(p+q)}$ with $U^T U = UU^T = I_{mn}$ and $V^T V = VV^T = I_{m(p+q)}$ and a rectangular diagonal matrix $\Sigma_{\mathcal{A}_0} \in \mathbb{R}^{mn\times m(p+q)}$ with diagonal nonnegative singular values $(\sigma_{\mathcal{A}_0,k})_k$ so that $\Sigma^{-1/2} \otimes \begin{pmatrix} M & \mathcal{A}_0 \end{pmatrix} C_0^{1/2} = U \Sigma_{\mathcal{A}_0} V^T$. We then have

$$\begin{aligned} C_n(\mathcal{A}_0, \Sigma) &= \left(\Sigma^{-1} \otimes \begin{pmatrix} M & \mathcal{A}_0 \end{pmatrix}^T \begin{pmatrix} M & \mathcal{A}_0 \end{pmatrix} + C_0^{-1}\right)^{-1} \\ &= C_0^{1/2} V \left(\Sigma_{\mathcal{A}_0}^T \Sigma_{\mathcal{A}_0} + I_{m(p+q)}\right)^{-1} V^T C_0^{1/2}. \end{aligned} \quad (11)$$

Using (11) and properties of the trace

$$\begin{aligned} \frac{1}{2}\mathrm{tr}(C_0^{-1/2} C_n(\mathcal{A}_0,\Sigma) C_0^{-1/2}) &\leq \frac{1}{2}\max_k \left[\left(\sigma_{\mathcal{A}_0,k}^2 + 1\right)^{-1}\right] \mathrm{tr}(V^T V) \\ &\leq \frac{m(p+q)}{2}. \end{aligned}$$

Using convexity,

$$\begin{aligned} &\frac{1}{2}\left\|C_0^{-1/2} c_n(\mathcal{A}_0,\Sigma)\right\|_2^2 \\ &= \frac{1}{2}\left\|C_0^{-1/2} C_n(\mathcal{A}_0,\Sigma)\left[\left[\Sigma^{-1} \otimes \begin{pmatrix} M & \mathcal{A}_0 \end{pmatrix}^T\right] \mathrm{vec}(R) + C_0^{-1} c_0\right]\right\|_2^2 \\ &\leq \left\|C_0^{-1/2} C_n(\mathcal{A}_0,\Sigma)\left[\Sigma^{-1/2} \otimes \begin{pmatrix} M & \mathcal{A}_0 \end{pmatrix}^T\right]\left[\Sigma^{-1/2} \otimes I_{mn}\right]\right\|_2^2 \|R\|_F^2 \\ &\quad + \left\|C_0^{-1/2} C_n(\mathcal{A}_0,\Sigma) C_0^{-1} c_0\right\|_2^2. \end{aligned}$$

Using the inequality (9) and the identity (11),

$$\begin{aligned} &\left\|C_0^{-1/2} C_n(\mathcal{A}_0,\Sigma)\left[\Sigma^{-1/2} \otimes \begin{pmatrix} M & \mathcal{A}_0 \end{pmatrix}\right]^T \left[\Sigma^{-1/2} \otimes I_{mn}\right]\right\|_2^2 \\ &= \left\|V\left(\Sigma_{\mathcal{A}_0}^T \Sigma_{\mathcal{A}_0} + I_{m(p+q)}\right)^{-1} V^T \left[\Sigma^{-1/2} \otimes \begin{pmatrix} M & \mathcal{A}_0 \end{pmatrix} C_0^{1/2}\right]^T \left[\Sigma^{-1/2} \otimes I_{mn}\right]\right\|_2^2 \\ &= \left\|V\left(\Sigma_{\mathcal{A}_0}^T \Sigma_{\mathcal{A}_0} + I_{m(p+q)}\right)^{-1} \Sigma_{\mathcal{A}_0}^T U^T\right\|_2^2 \left\|\Sigma^{-1}\right\|_2 \end{aligned}$$

$$\leq \|V\|_2^2 \left\| \left(\Sigma_{\mathcal{A}_0}^T \Sigma_{\mathcal{A}_0} + I_{m(p+q)}\right)^{-1} \Sigma_{\mathcal{A}_0}^T \right\|_2^2 \left\|U^T\right\|_2^2 \left\|\Sigma^{-1}\right\|_2$$
$$\leq \max_k \left(\frac{\sigma_{\mathcal{A}_0,k}}{\sigma_{\mathcal{A}_0,k}^2 + 1} \right)^2 \left\|\Sigma^{-1}\right\|_2$$
$$\leq \frac{\left\|\Sigma^{-1}\right\|_2}{4}.$$

Using (11),

$$\begin{aligned}\left\| C_0^{-1/2} C_n(\mathcal{A}_0, \Sigma) C_0^{-1} c_0 \right\|_2^2 &= \left\| V \left(\Sigma_{\mathcal{A}_0}^T \Sigma_{\mathcal{A}_0} + I_{m(p+q)}\right)^{-1} V^T C_0^{-1/2} c_0 \right\|_2^2 \\ &\leq \|V\|_2^2 \left\| \left(\Sigma_{\mathcal{A}_0}^T \Sigma_{\mathcal{A}_0} + I_{m(p+q)}\right)^{-1} \right\|_2^2 \left\|V^T\right\|_2^2 c_0^T C_0^{-1} c_0 \\ &\leq c_0^T C_0^{-1} c_0.\end{aligned}$$

Combining the upper bounds

$$\mathbb{E}\left[\frac{1}{2} \left\| C_0^{-1/2} (\theta, \beta)^T \right\|_2^2 | \Sigma, \mathcal{A}_0, \theta_0, \beta_0 \right] \leq \frac{\|R\|_F^2}{4} \left\|\Sigma^{-1}\right\|_2 + c_0^T C_0^{-1} c_0 + \frac{m(p+q)}{2}. \tag{12}$$

Now using (10) and (12) and taking the iterated expectation with respect to $\theta, \beta | \mathcal{A}_0, \theta_0, \beta_0, \Sigma$,

$$\mathbb{E}\left[\frac{1}{2} \sum_{i=1}^n \left\| D_i^{-1/2} (\mathcal{A}_i - d_i) \right\|_2^2 | \Sigma, \mathcal{A}_0, \theta_0, \beta_0 \right] \tag{13a}$$
$$\leq \left\|\Sigma^{-1}\right\|_2 \tag{13b}$$
$$\times \left[\frac{\|R\|_F^2}{4} + \frac{\left\| \begin{pmatrix} M & \tilde{X} \end{pmatrix} C_0^{1/2} \right\|_2^2 m(p+q)}{4} + \frac{\left\| \begin{pmatrix} M & \tilde{X} \end{pmatrix} C_0^{1/2} \right\|_2^2}{2} c_0^T C_0^{-1} c_0 \right] \tag{13c}$$
$$+ \left\|\Sigma^{-1}\right\|_2^2 \frac{\left\| \begin{pmatrix} M & \tilde{X} \end{pmatrix} C_0^{1/2} \right\|_2^2 \|R\|_F^2}{8} + \frac{pn}{2}. \tag{13d}$$

Since Σ^{-1} has a Wishart distribution and using properties of the trace,

$$\begin{aligned}\mathbb{E}\left\|\Sigma^{-1}\right\|_2 &\leq \operatorname{tr}\left[\mathbb{E}\left(\Sigma^{-1}\right)\right] \\ &= (n + a_0)\operatorname{tr}\left[\left((R - M\Theta_0 - \mathcal{A}_0\mathcal{B}_0)^T (R - M\Theta_0 - \mathcal{A}_0\mathcal{B}_0) + B_0 \right)^{-1} \right] \\ &\leq (n + a_0) tr[B_0^{-1}].\end{aligned}$$

Similarly, we use the second moment formula of the Wishart (Letac and Massam, 2004) to get the upper bound,

$$\mathbb{E}\left\|\Sigma^{-1}\right\|_2^2 \leq \operatorname{tr}\left[\mathbb{E}\left(\Sigma^{-2}\right)\right]$$

$$= (n+a_0)\text{tr}\left[\left((R - M\Theta_0 - \mathcal{A}_0\mathcal{B}_0)(R - M\Theta_0 - \mathcal{A}_0\mathcal{B}_0)^T + B_0\right)^{-1}\right]^2$$
$$+ (n+a_0)(n+a_0+1)$$
$$\times \text{ tr}\left[\left((R - M\Theta_0 - \mathcal{A}_0\mathcal{B}_0)(R - M\Theta_0 - \mathcal{A}_0\mathcal{B}_0)^T + B_0\right)^{-2}\right]$$
$$\leq (n+a_0)(n+a_0+1)tr[B_0^{-1}]^2.$$

Taking the iterated expectation with respect to $\Sigma|\mathcal{A}_0, \theta_0, \beta_0$ in (12) and (13), there is a constant $L \in (0,\infty)$ so that the drift condition is satisfied with

$$\mathbb{E}\left[V(\mathcal{A}, \theta, \beta)|\mathcal{A}_0, \theta_0, \beta_0\right] \leq L.$$

□

2.1. Bayesian EIV regression with errors in the features

Using Theorem 2.1, we develop geometrically ergodic Gibbs samplers for Bayesian EIV regression with additive Gaussian error in the features. For the remainder, we write the observed data as $Y = (y_1, \ldots, y_n)^T \in \mathbb{R}^{n\times m}$, $X = (x_1, \ldots, x_n)^T \in \mathbb{R}^{n\times p}$, and $Z = (Z_1, \ldots, Z_n)^T \in \mathbb{R}^{n\times q}$. Consider the Bayesian EIV regression (1) with Berkson errors and priors (3) and (4). We will write the posterior density π_n for this Bayesian model as

$$\pi_n(\mathcal{A}, \theta, \beta, \Sigma)$$
$$\propto \left(\frac{1}{\det(\Sigma)}\right)^{(n+a_0+m+1)/2} \exp\left[-\frac{1}{2}\text{tr}[\Sigma^{-1}B_0]\right]$$
$$\times \exp\left[-\frac{1}{2}\sum_{i=1}^n (y_i - \Theta^T Z_i - \mathcal{B}^T\mathcal{A}_i)^T\Sigma^{-1}(y_i - \Theta^T Z_i - \mathcal{B}^T\mathcal{A}_i)\right]$$
$$\times \exp\left(-\frac{1}{2}\sum_{i=1}^n (\mathcal{A}_i - x_i)^T V_i^{-1}(\mathcal{A}_i - x_i)\right)$$
$$\times \exp\left(-\frac{1}{2}((\theta,\beta)^T - j_0)^T J_0^{-1}((\theta,\beta)^T - j_0)\right).$$

This posterior density is a special case of the general density (5) choosing $M \equiv Z$, $R \equiv Y$, $c_0, C_0 \equiv j_0, J_0$, and $d_i, D_i \equiv x_i, V_i$.

We can define a 3-variable deterministic scan Gibbs sampler which generates a Markov chain $(\mathcal{A}_t, \theta_t, \beta_t, \Sigma_t)_{t=0}^{\infty}$ for this posterior density as a special case of the Gibbs sampler constructed in Section 2. Initialize $(\mathcal{A}_0, \theta_0, \beta_0, \Sigma_0)$ and for $t \in 1, \ldots,$

1. Generate $\Sigma_t|\mathcal{A}_{t-1}, \theta_{t-1}, \beta_{t-1} \sim \mathcal{W}_m^{-1}(n+a_0, B_{n,t})$ where

$$B_{n,t} = (Y - Z\Theta_{t-1} - \mathcal{A}_{t-1}\mathcal{B}_{t-1})^T(Y - Z\Theta_{t-1} - \mathcal{A}_{t-1}\mathcal{B}_{t-1}) + B_0$$

2. Generate $(\theta_t, \beta_t)^T | \mathcal{A}_{t-1}, \Sigma_t \sim N_{m(p+q)}(j_{n,t}, J_{n,t})$ where

$$J_{n,t} = \left(\Sigma_t^{-1} \otimes \begin{pmatrix} Z & \mathcal{A}_{t-1} \end{pmatrix}^T \begin{pmatrix} Z & \mathcal{A}_{t-1} \end{pmatrix} + J_0^{-1} \right)^{-1}$$
$$j_{n,t} = J_{n,t} \left[\left[\Sigma_t^{-1} \otimes \begin{pmatrix} Z & \mathcal{A}_{t-1} \end{pmatrix}^T \right] \operatorname{vec}(Y) + J_0^{-1} j_0 \right]$$

3. Generate $\mathcal{A}_{i,t} | \theta_t, \beta_t, \Sigma_t \sim N_p(k_{n,i,t}, K_{n,i,t}), i \in 1, \dots, n$ where

$$K_{n,i,t} = \left(\mathcal{B}_t \Sigma_t^{-1} \mathcal{B}_t^T + V_i^{-1} \right)^{-1}$$
$$k_{n,i,t} = K_{n,i,t} \left[V_i^{-1} x_i + \mathcal{B}_t \Sigma_t^{-1} \left(y_i - \Theta_t^T Z_i \right) \right].$$

Applying Theorem 2.1, we have the following result.

Corollary 2.2. *The 3-variable Gibbs sampler $(\mathcal{A}_t, \theta_t, \beta_t, \Sigma_t)_{t=0}^{\infty}$ for the posterior in Bayesian EIV regression (1) with Berkson errors and priors (3) and (4) is geometrically ergodic.*

Now consider Bayesian EIV regression (1) with additive Gaussian error in X_i using classical errors and priors (3) and (4). The posterior has density

$$\begin{aligned}
&\pi_n(\mathcal{A}, \theta, \beta, \Sigma) \\
&\propto \left(\frac{1}{\det(\Sigma)} \right)^{(n+a_0+m+1)/2} \exp\left[-\frac{1}{2} \operatorname{tr}[\Sigma^{-1} B_0] \right] \\
&\quad \times \exp\left[-\frac{1}{2} \sum_{i=1}^n (y_i - \Theta^T Z_i - \mathcal{B}^T \mathcal{A}_i)^T \Sigma^{-1} (y_i - \Theta^T Z_i - \mathcal{B}^T \mathcal{A}_i) \right] \\
&\quad \times \exp\left(-\frac{1}{2} \sum_{i=1}^n (\mathcal{A}_i - k_i')^T (V_i^{-1} + K_i^{-1}) (\mathcal{A}_i - k_i') \right) \\
&\quad \times \exp\left(-\frac{1}{2} ((\theta, \beta)^T - j_0)^T J_0^{-1} ((\theta, \beta)^T - j_0) \right)
\end{aligned}$$

where $k_i' = (V_i^{-1} + K_i^{-1})^{-1} \left[V_i^{-1} x_i + K_i^{-1} k_i \right]$. The posterior density is also a special case of the general density (5) when $Z \equiv M$, $R \equiv Y$, and $c_0, C_0 \equiv j_0, J_0$, and $d_i, D_i \equiv k_i', (V_i^{-1} + K_i^{-1})^{-1}$.

We define a 3-variable deterministic scan Gibbs sampler similarly. Initialize $(\mathcal{A}_0, \theta_0, \beta_0, \Sigma_0)$ and for $t \in 1, \dots,$

1. Generate $\Sigma_t | \mathcal{A}_{t-1}, \theta_{t-1}, \beta_{t-1} \sim \mathcal{W}_m^{-1}(n + a_0, B_{n,t})$
2. Generate $(\theta_t, \beta_t)^T | \mathcal{A}_{t-1}, \Sigma_t \sim N_{m(p+q)}(j_{n,t}, J_{n,t})$
3. Generate $\mathcal{A}_{i,t} | \theta_t, \beta_t, \Sigma_t \sim N_p(k'_{n,i,t}, K'_{n,i,t}), i \in 1, \dots, n$ where

$$K'_{n,i,t} = \left(\mathcal{B}_t \Sigma_t^{-1} \mathcal{B}_t^T + V_i^{-1} + K_i^{-1} \right)^{-1}$$
$$k'_{n,i,t} = K'_{n,i,t} \left[V_i^{-1} x_i + K_i^{-1} k_i + \mathcal{B}_t \Sigma_t^{-1} \left(y_i - \Theta_t^T Z_i \right) \right].$$

We also have the following as a direct result of Theorem 2.1.

Corollary 2.3. *The 3-variable Gibbs sampler $(\mathcal{A}_t, \theta_t, \beta_t, \Sigma_t)_{t=0}^{\infty}$ for the posterior in Bayesian EIV regression (1) with classical errors and priors (3) and (4) is geometrically ergodic.*

2.2. Bayesian EIV regression with errors in the response and features

Similar to the previous section, we develop geometrically ergodic Gibbs samplers for Bayesian EIV regression with additional additive Gaussian error in the features and response. Consider the Bayesian EIV regression (2) with Berkson errors in X_i and additional error in Y_i along with priors (3) and (4). Let $\mathcal{V} = (\mathcal{V}_1, \dots, \mathcal{V}_n)^T \in \mathbb{R}^{n\times m}$ and $\nu = \text{vec}(\mathcal{V})$, and let $U_0 = \text{blockdiag}(U_i) \in \mathbb{R}^{mn\times mn}$. The Bayesian posterior Π_n has density

$$\begin{aligned}
&\pi_n(\mathcal{A}, \nu, \theta, \beta, \Sigma)\\
&\propto \left(\frac{1}{\det(\Sigma)}\right)^{(n+a_0+m+1)/2} \exp\left[-\frac{1}{2}\text{tr}[\Sigma^{-1}B_0]\right]\\
&\quad\times \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\mathcal{V}_i - \Theta^T Z_i - \mathcal{B}^T\mathcal{A}_i)^T\Sigma^{-1}(\mathcal{V}_i - \Theta^T Z_i - \mathcal{B}^T\mathcal{A}_i)\right]\\
&\quad\times \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\mathcal{V}_i - y_i)^T U_i^{-1}(\mathcal{V}_i - y_i)\right)\\
&\quad\times \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\mathcal{A}_i - x_i)^T V_i^{-1}(\mathcal{A}_i - x_i)\right)\\
&\quad\times \exp\left(-\frac{1}{2}((\theta,\beta)^T - j_0)^T J_0^{-1}((\theta,\beta)^T - j_0)\right).
\end{aligned}$$

This posterior density is a special case of the density (5) when redefining $\tilde{\theta} \equiv (\nu, \theta)^T$, $M \equiv \begin{pmatrix}-I & Z\end{pmatrix}$, $r \equiv 0$, $c_0 = (Y, j_0)^T$,

$$C_0 \equiv \begin{pmatrix} U_0 & 0 \\ 0 & J_0 \end{pmatrix},$$

and $d_i, D_i \equiv x_i, V_i$.

We define a 3-variable deterministic scan Gibbs sampler which generates a Markov chain $(\mathcal{A}_t, \nu_t, \theta_t, \beta_t, \Sigma_t)_{t=0}^{\infty}$ for this posterior density. Initialize $\mathcal{A}_0$, ν_0, θ_0, β_0, Σ_0 and for $t \in 1, \dots,$

1. Generate $\Sigma_t | \mathcal{A}_{t-1}, \nu_{t-1}, \theta_{t-1}, \beta_{t-1} \sim \mathcal{W}_m^{-1}\left(n + a_0, B'_{n,t}\right)$ where

$$B'_{n,t} = (\mathcal{V}_{t-1} - Z\Theta_{t-1} - \mathcal{A}_{t-1}\mathcal{B}_{t-1})^T(\mathcal{V}_{t-1} - Z\Theta_{t-1} - \mathcal{A}_{t-1}\mathcal{B}_{t-1}) + B_0$$

2. Generate $(\nu_t, \theta_t, \beta_t)^T | \mathcal{A}_{t-1}, \Sigma_t \sim N_{p+q}(j'_{n,t}, J'_{n,t})$ where

$$\begin{aligned}
J'_{n,t} &= \left(\Sigma_t^{-1} \otimes \begin{pmatrix}-I & Z & \mathcal{A}_{t-1}\end{pmatrix}^T \begin{pmatrix}-I & Z & \mathcal{A}_{t-1}\end{pmatrix} + \begin{pmatrix} U_0^{-1} & 0 \\ 0 & J_0^{-1} \end{pmatrix}\right)^{-1}\\
j'_{n,t} &= J'_{n,t} \begin{pmatrix} U_0^{-1} & 0 \\ 0 & J_0^{-1} \end{pmatrix} (\text{vec}(Y), j_0)^T
\end{aligned}$$

3. Generate $\mathcal{A}_{i,t}|\nu_t, \theta_t, \beta_t, \Sigma_t \sim N_p(k''_{n,i,t}, K''_{n,i,t}), i \in 1, \ldots, n$ where

$$K''_{n,i,t} = \left(\mathcal{B}_t \Sigma_t^{-1} \mathcal{B}_t^T + V_i^{-1}\right)^{-1}$$
$$k''_{n,i,t} = K''_{n,i,t} \left[V_i^{-1} x_i + \mathcal{B}_t \Sigma_t^{-1} \left(\mathcal{V}_{i,t} - \Theta_t^T Z_i\right)\right].$$

Using Theorem 2.1, we have the following result.

Corollary 2.4. *The 3-variable Gibbs sampler* $(\mathcal{A}_t, \nu_t, \theta_t, \beta_t, \Sigma_t)_{t=0}^{\infty}$ *for Bayesian EIV regression* (2) *with Berkson errors and priors* (3) *and* (4) *is geometrically ergodic.*

Now consider the Bayesian EIV regression (2) with classical errors in X_i and additional error in Y_i with priors (3) and (4). The posterior Π_n for this Bayesian model has density

$$\begin{aligned}
&\pi_n(\mathcal{A}, \nu, \theta, \beta, \Sigma) \\
&\propto \left(\frac{1}{\det(\Sigma)}\right)^{(n+a_0+m+1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}[\Sigma^{-1}B_0]\right] \\
&\quad \times \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\mathcal{V}_i - \Theta^T Z_i - \mathcal{B}^T\mathcal{A}_i)^T\Sigma^{-1}(\mathcal{V}_i - \Theta^T Z_i - \mathcal{B}^T\mathcal{A}_i)\right] \\
&\quad \times \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\mathcal{V}_i - y_i)^T U_i^{-1}(\mathcal{V}_i - y_i)\right) \\
&\quad \times \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(\mathcal{A}_i - k'_i)^T(V_i^{-1} + K_i^{-1})(\mathcal{A}_i - k'_i)\right) \\
&\quad \times \exp\left(-\frac{1}{2}((\theta, \beta)^T - j_0)^T J_0^{-1}((\theta, \beta)^T - j_0)\right).
\end{aligned}$$

This posterior density is also a special case of the density (5) when redefining $\tilde{\theta} \equiv (\nu, \theta)^T$, $M \equiv \begin{pmatrix} -I & Z \end{pmatrix}$, $R \equiv 0$, $c_0 = (y, j_0)^T$,

$$C_0 \equiv \begin{pmatrix} U_0 & 0 \\ 0 & J_0 \end{pmatrix},$$

and $d_i, D_i \equiv k'_i, (V_i^{-1} + K_i^{-1})^{-1}$.

We define a 3-variable deterministic scan Gibbs sampler similarly. Initialize $(\mathcal{A}_0, \nu_0, \theta_0, \beta_0, \Sigma_0)$ and for $t \in 1, \ldots,$

1. Generate $\Sigma_t|\mathcal{A}_{t-1}, \nu_{t-1}, \theta_{t-1}, \beta_{t-1} \sim \mathcal{W}_m^{-1}\left(n + a_0, B'_{n,t}\right)$
2. Generate $(\nu_t, \theta_t, \beta_t)^T|\mathcal{A}_{t-1}, \Sigma_t \sim N_{p+q}(j'_{n,t}, J'_{n,t})$
3. Generate $\mathcal{A}_{i,t}|\nu_t, \theta_t, \beta_t, \Sigma_t \sim N_p(k'''_{n,i,t}, K'''_{n,i,t}), i \in 1, \ldots, n$ where

$$K'''_{n,i,t} = \left(\mathcal{B}_t \Sigma_t \mathcal{B}_t^T + V_i^{-1} + K_i^{-1}\right)^{-1}$$
$$k'''_{n,i,t} = K'''_{n,i,t}\left[V_i^{-1} x_i + K_i^{-1} k_i + \mathcal{B}_t \Sigma_t^{-1}\left(\mathcal{V}_{i,t} - \Theta_t^T Z_i\right)\right].$$

Using Theorem 2.1, we have the following result.

Corollary 2.5. *The 3-variable Gibbs sampler* $(\mathcal{A}_t, \nu_t, \theta_t, \beta_t, \Sigma_t)_{t=0}^{\infty}$ *for Bayesian EIV regression* (2) *with classical errors and priors* (3) *and* (4) *is geometrically ergodic.*

3. Simulations

3.1. Limitations of the Gibbs sampler in large problem sizes

Theoretically, we developed a *qualitative* convergence result for the Gibbs sampler in Bayesian multivariate EIV regression. It is important in practice to understand the relationship between scaling of the problem size and the estimation reliability from the Gibbs sampler. We look at artificially generated data to empirically demonstrate the dependence of the Gibbs sampler when the the dimension of the response m and the dimension of the features p are increasing in configurations $(m, p) = (1, 1), (3, 7)$ in the Bayesian posterior. Artificial data is generated according to the multivariate EIV Berkson linear regression model for $i = 1, \dots, 50$ with

$$\begin{aligned}
&\Sigma \sim \mathcal{W}_m^{-1}(m, 10^{-3} I) \\
&(\theta, \beta)^T \sim N_{m+mp}(0, 10^3 I) \\
&X_i|\mathcal{A}_i \sim N_p(\mathcal{A}_i, .2I) \\
&Y_i|\mathcal{A}_i, \Theta, \mathcal{B}, \Sigma \sim N_m(\Theta^T 1 + \mathcal{B}^T \mathcal{A}_i, \Sigma)
\end{aligned}$$

We simulate $T = 10^5$ MCMC realizations from the Gibbs sampler in each configuration using 10^4 realizations for burn-in and analyze diagnostics for β_t taking values in $\mathbb{R}^{mp}$. We independently replicate the simulation 100 times to reduce variability and plot the median in a solid line along with the .25, and .75 quantiles in a lighter opacity over these independent simulations. As discussed in Section 2, geometric ergodicity guarantees the properly scaled and summed samples from the Gibbs sampler

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\beta_t - \int \beta \Pi_n(d\beta) \right] \to N(0, \Sigma_*)$$

as $T \to \infty$ in distribution where Σ_* is a SPD covariance matrix. Figure 1a and Figure 1b plot the largest and smallest eigenvalues of a multivariate batch means estimate to the multivariate standard error matrix ${\Sigma_*}^{1/2}$ in the central limit theorem. The batch means estimate divides the simulation samples into non-overlapping batches of size $B = T^{1/3}$, computes the average over each batch denoted by $\bar{\beta}_B$, and then takes the sample covariance of $\sqrt{B}\bar{\beta}_B$. This simulation shows an increase in the largest eigenvalue in larger size problems which can lead to suggesting more iterations are needed for appropriate estimation in practice. Figure 1c plots the multivariate effective sample size (Vats, Flegal and Jones,

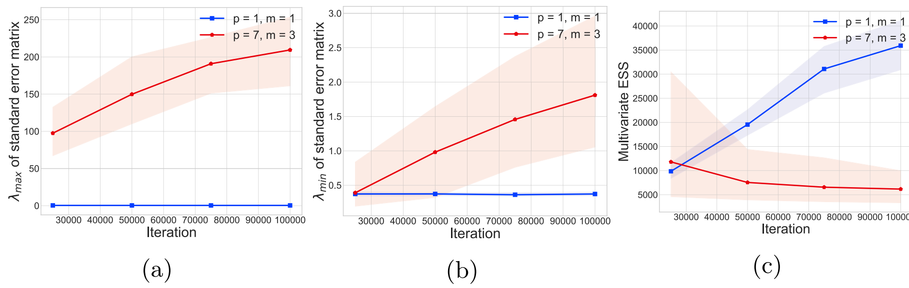


Fig 1. *(a), (b) Largest and smallest eigenvalues of the MCMC standard error matrix targeting the average of β for iterations of the Gibbs sampler and (c) the multivariate effective sample size for iterations of the Gibbs sampler.*

2019b). We can also see a relatively sharp decrease in the estimation of the effective sample size as the problem size increases suggesting the algorithm should be run for many iterations even in moderately sized problems. This simulation demonstrates that even though the algorithm is *always* geometrically ergodic and can scale reasonably well to larger problem sizes, the Gibbs sampler generally requires many more iterations for reliable estimation even in moderately sized problems. Simulation code is available using Python (Van Rossum and Drake Jr, 1995) and (Harris et al., 2020) for matrix calculations at `https://github.com/austindavidbrown/BayesEIV`.

3.2. Robustness to model misspecification

Although the multivariate Bayesian model for EIV accounts for additional error in the features, this error can be misspecified. In particular, the error $X_i|\mathcal{A}_i$ from the model in Section 3.1 may be a multivariate t distribution with heavier tails in practical problems. We are interested to empirically study the robustness of the convergence of the Gibbs sampler to misspecification in this modeling error. Denote $t_d(v, m, V)$ as a multivariate t distribution in dimension d with v degrees of freedom, location vector m, and scale matrix V. With *df* denoting the degrees of freedom, artificial data is generated according to the misspecified multivariate EIV Berkson linear regression model for $i = 1, \dots, 50$ with

$$
\begin{aligned}
&\Sigma \sim \mathcal{W}_3^{-1}(3, 10^{-3} I) \\
&(\theta, \beta)^T \sim N_{3+9}(0, 10^3 I) \\
&X_i|\mathcal{A}_i \sim \mathrm{t}_3(\mathrm{df}, \mathcal{A}_i, .2I) \\
&Y_i|\mathcal{A}_i, \Theta, \mathcal{B}, \Sigma \sim N_3(\Theta^T 1 + \mathcal{B}^T \mathcal{A}_i, \Sigma).
\end{aligned}
$$

We look to compare more dispersed tail behavior with $df = 2$ and less dispersed tail behavior with $df = 10$ to the data from the correctly prescribed model. We replicate the simulation 100 times in the same way as Section 3.1 and analyze diagnostics for β_t. Figure 2a and Figure 2b plot the largest and

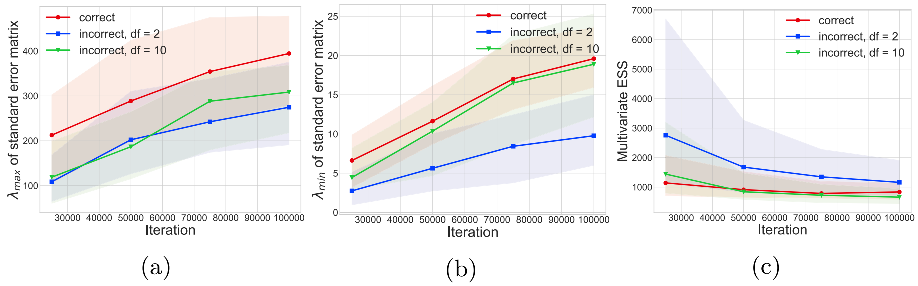


Fig 2. *(a), (b) Largest and smallest eigenvalues of the MCMC standard error matrix (c) the multivariate effective sample size for iterations of the Gibbs sampler.*

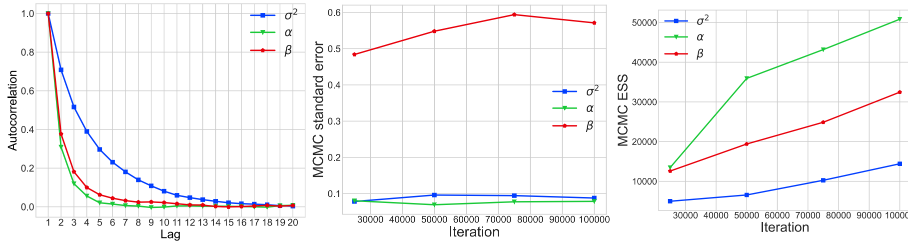
smallest eigenvalues of a batch means estimate to the multivariate standard error matrix and Figure 2c plots the multivariate effective sample size (Vats, Flegal and Jones, 2019b). Although we see some discrepancy in the maximum eigenvalues in Figure 2a, we can see similar behavior in the estimation from the Gibbs sampler based on the both smaller and larger degrees of freedom to the correctly simulated data in the multivaraite effective sample size. The simulation results suggest the Gibbs sampler is reasonably robust to misspecification of the tails in the error distribution of the features for $X_i|\mathcal{A}_i$.

4. Real-data example: measurement error in astrophysics

We look at Bayesian EIV linear regression proposed and analyzed in (Harris, Poole and Harris, 2014; Hilbe, de Souza and Ishida, 2017). The dataset consists of the central galaxy supermassive black hole mass and the stellar bulge velocity dispersion from $n = 46$ different galaxies (Harris, Poole and Harris, 2014). The response Y_i is the logarithm of the observed central black hole mass and the predictor variable X_i is the logarithm of the observed velocity dispersion. The measurement errors are known beforehand and denoted by ϵ_{Y_i} and ϵ_{X_i} for both the response and predictor variables. The EIV linear regression model studied in (Hilbe, de Souza and Ishida, 2017) follows

$$
\begin{aligned}
&\sigma^2 \sim \text{Inverse-gamma}(10^{-3}, 10^{-3}) \\
&\alpha \sim N_1(0, 10^3), \beta \sim N_1(0, 10^3) \\
&\mathcal{A}_i \sim N_1(0, 10^3), X_i|\mathcal{A}_i \sim N_1(\mathcal{A}_i, \epsilon_{X_i}^2) \\
&\mathcal{V}_i|\mathcal{A}_i, \theta, \beta, \sigma^2 \sim N_1(\theta + \mathcal{A}_i\beta, \sigma^2), Y_i|\mathcal{V}_i \sim N_1(\mathcal{V}_i, \epsilon_{y_i}^2).
\end{aligned}
$$

We generate 10^5 MCMC realizations from the Gibbs sampler. Figure 3 plots the autocorrelation, estimates to the standard errors in the central limit theorem, and effective sample sizes from these realizations. The autocorrelations are computed up to lag 20. Overall, we see the Gibbs sampler performs well. However, the standard error and effective sample size plots suggest that even



(a) (b) (c)

Fig 3. *(a) Autocorrelation for each regression parameter (b) MCMC standard errors for the regression parameters (c) MCMC effective sample size plots for each regression parameter.*

though the Gibbs sampler is geometrically ergodic, many iterations are still recommended even in low dimensions. These figures suggest empirical diagnostics for the regression parameter β and σ^2 are reasonable choices as opposed to the other parameters α to determine the reliability of the algorithm in practice.

5. Conclusion and future directions

We showed using a 3-variable deterministic scan Gibbs sampler to sample the posterior in 4 different multivariate Bayesian EIV regression models with additive Gaussian errors and independent priors is always geometrically ergodic. This is of pragmatic importance to practitioners as trustworthy estimation from a Gibbs sampler is dependent on the speed of convergence of the Markov chain. More specifically, time averages from the Markov chains have many practically relevant theoretical guarantees such as a central limit theorem. Secondly, these Gibbs samplers can be simulated efficiently without the need for complex, intermediate Metropolis-Hastings or rejection sampling steps. One drawback, however, is our convergence analysis is qualitative as we do not construct an explicit convergence rate.

There are many future research directions in studying the convergence of Gibbs samplers in EIV models. It appears reasonable that some Gibbs samplers for generalized linear models such as the Pólya-Gamma sampler will also be geometrically ergodic (Choi and Hobert, 2013; Polson, Scott and Windle, 2013; Wang and Roy, 2018). It seems also interesting to look at alternative errors in the variables such as non-Gaussian or non-additive errors.

Acknowledgments

Thanks to Galin L. Jones for helpful guidance and insightful comments for developing this article. Thanks to the anonymous referees for recommendations to improve and extend some results from a previous version of this manuscript.

References

Achic, B. G. B., Wang, T., Su, Y., Kipnis, V., Dodd, K. and Carroll, R. J. (2018). Categorizing a continuous predictor subject to measurement error. *Electronic Journal of Statistics* **12** 4032–4056. MR3885744

Berkson, J. (1950). Are There Two Regressions? *Journal of the American Statistical Association* **45** 164–180.

Bhadra, A. and Carroll, R. J. (2016). Exact sampling of the unobserved covariates in Bayesian spline models for measurement error problems. *Statistics and Computing* **26** 827–840. MR3515024

Bresson, G., Chaturvedi, A., Rahman, M. A. and Shalabh (2021). Seemingly unrelated regression with measurement error: estimation via Markov Chain Monte Carlo and mean field variational Bayes approximation. *The International Journal of Biostatistics* **17** 75–97.

Buonaccorsi, J. P. (2010). *Measurement Error: Models, Methods, and Applications*, 1 ed. Chapman and Hall/CRC. MR2682774

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2 ed. Chapman and Hall/CRC. MR2243417

Chan, K. S. and Geyer, C. J. (1994). Discussion: Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics* **22** 1747–1758.

Charisse Farr, A., Mengersen, K., Ruggeri, F., Simpson, D., Wu, P. and Yarlagadda, P. (2020). Combining Opinions for Use in Bayesian Networks: A Measurement Error Approach. *International Statistical Review* **88** 335–353. MR4176179

Choi, H. M. and Hobert, J. P. (2013). The Pólya-Gamma Gibbs sampler for Bayesian logistic regression is uniformly ergodic. *Electronic Journal of Statistics* **7** 2054–2064. MR3091616

Clayton, D. et al. (1992). Models for the analysis of cohort and case-control studies with inaccurately measured exposures. *Statistical models for longitudinal studies of health* 301–331.

Damgaard, C. (2020). Measurement Uncertainty in Ecological and Environmental Models. *Trends in Ecology and Evolution* **35** 871–873.

Dellaportas, P. and Stephens, D. A. (1995). Bayesian Analysis of Errors-in-Variables Regression Models. *Biometrics* **51** 1085–1095.

Doss, C. R., Flegal, J. M., Jones, G. L. and Neath, R. C. (2014). Markov chain Monte Carlo estimation of quantiles. *Electronic Journal of Statistics* **8** 2448–2478. MR3285872

Ekvall, K. O. and Jones, G. L. (2021). Convergence analysis of a collapsed Gibbs sampler for Bayesian vector autoregressions. *Electronic Journal of Statistics* **15** 691–721. MR4202497

Fang, X., Li, B., Alkhatib, H., Zeng, W. and Yao, Y. (2017). Bayesian inference for the Errors-In-Variables model. *Studia Geophysica et Geodaetica* **61** 1573–1626.

Feigelson, E. D. and Babu, G. J. (1992). Linear regression in astronomy. II. *The Astrophysical Journal* **397** 55–67.

FULLER, W. A. (1987). *Measurement Error Models*. John Wiley. MR0898653

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.

GOODFELLOW, I., SHLENS, J. and SZEGEDY, C. (2015). Explaining and Harnessing Adversarial Examples. *International Conference on Learning Representations*.

GROSS, M. (2016). Modeling body height in prehistory using a spatio-temporal Bayesian errors-in variables model. *AStA Advances in Statistical Analysis* **100** 289–311. MR3522056

GUSTAFSON, P. (2003). *Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments*. CRC Press. MR2005104

HAIRER, M. and MATTINGLY, J. C. (2011). Yet Another Look at Harris' Ergodic Theorem for Markov Chains. *Seminar on Stochastic Analysis, Random Fields and Applications VI* **63**. MR2857021

HARRIS, G., POOLE, G. and HARRIS, W. (2014). Globular clusters and supermassive black holes in galaxies: further analysis and a larger sample. *Monthly Notices of the Royal Astronomical Society* **438** 2117–2130.

HARRIS, C. R., MILLMAN, K. J., VAN DER WALT, S. J., GOMMERS, R., VIRTANEN, P., COURNAPEAU, D., WIESER, E., TAYLOR, J., BERG, S., SMITH, N. J., KERN, R., PICUS, M., HOYER, S., VAN KERKWIJK, M. H., BRETT, M., HALDANE, A., FERNÁNDEZ DEL RÍO, J., WIEBE, M., PETERSON, P., GÉRARD-MARCHANT, P., SHEPPARD, K., REDDY, T., WECKESSER, W., ABBASI, H., GOHLKE, C. and OLIPHANT, T. E. (2020). Array programming with NumPy. *Nature* **585** 357–362. `https://doi.org/10.1038/s41586-020-2649-2`.

HILBE, J. M., DE SOUZA, R. S. and ISHIDA, E. E. O. (2017). *Bayesian Models for Astrophysical Data: Using R, JAGS, Python, and Stan*. Cambridge University Press.

HORN, R. A. and JOHNSON, C. R. (2012). *Matrix Analysis*. Cambridge University Press. MR2978290

HUANG, H.-J. (2010). Bayesian Analysis of Errors-in-Variables Growth Curve Models. *PhD Dissertation*. MR2941636

HÅVARD RUE, N. C. SARA MARTINO (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **71** 319–392. MR2649602

JONES, G. L. (2004). On the Markov chain central limit theorem. *Probability surveys* **1** 299–320. MR2068475

KELLY, B. C. (2012). Measurement error models in astronomy. In *Statistical challenges in modern astronomy V* 147–162. Springer. MR3220180

KRÖGER, H., HOFFMANN, R. and PAKPAHAN, E. (2016). Consequences of measurement error for inference in cross-lagged panel design—the example of the reciprocal causal relationship between subjective health and socio-economic status. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*

179 607–628. MR3461597

Letac, G. and Massam, H. (2004). All Invariant Moments of the Wishart Distribution. *Scandinavian Journal of Statistics* **31** 295–318. MR2066255

Mallick, B. K. and Gelfand, A. E. (1996). Semiparametric errors-in-variables models A Bayesian approach. *Journal of Statistical Planning and Inference* **52** 307–321. MR1401026

Meyn, S. P. and Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability*, 2 ed. Cambridge University Press, USA. MR2509253

Michalek, J. E. and Tripathi, R. C. (1980). The Effect of Errors in Diagnosis and Measurement on the Estimation of the Probability of an Event. *Journal of the American Statistical Association* **75** 713–721. MR0590707

Muff, S., Riebler, A., Held, L., Rue, H. and Saner, P. (2015). Bayesian analysis of measurement error models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **64** 231–252. MR3302298

Nesterov, Y. (2018). *Lectures on Convex Optimization*, 2 ed. Springer International Publishing. MR3839649

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. MR3103783

Pollice, A., Jona Lasinio, G., Rossi, R., Amato, M., Kneib, T. and Lang, S. (2019). Bayesian measurement error correction in structured additive distributional regression with an application to the analysis of sensor data on soil–plant variability. *Stochastic Environmental Research and Risk Assessment* **33** 747–763.

Polson, N. G., Scott, J. G. and Windle, J. (2013). Bayesian Inference for Logistic Models Using Pólya–Gamma Latent Variables. *Journal of the American Statistical Association* **108** 1339–1349. MR3174712

Rajaratnam, B. and Sparks, D. (2015). MCMC-Based Inference in the Era of Big Data: A Fundamental Analysis of the Convergence Complexity of High-Dimensional Chains. *preprint arXiv:1508.00947*.

Richardson S, G. W. (1993). A Bayesian approach to measurement error problems in epidemiology using conditional independence models. *American journal of epidemiology* **138** 430–42.

Roberts, G. O. and Rosenthal, J. S. (2001). Markov chains and de-initializing processes. *Scandinavian Journal of Statistics* **28** 489–504. MR1858413

Rodrigues, J. and Bolfarine, H. (2007). Bayesian inference for an extended simple regression measurement error model using skewed priors. *Bayesian Analysis* **61** 349–364. MR2312286

Stefanski, L. A. (2000). Measurement error models. *Journal of the American Statistical Association* **95** 1353–1358. MR1825293

Stefanski, L. A. and Carroll, R. J. (1985). Covariate Measurement Error in Logistic Regression. *The Annals of Statistics* **13** 1335–1351. MR0811496

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I. and Fergus, R. (2014). Intriguing properties of neural networks.

International Conference on Learning Representations.

Tang, N.-S., Li, D.-W. and Tang, A.-M. (2017). Semiparametric Bayesian inference on generalized linear measurement error models. *Statistical Papers* **58** 1091–1113. MR3720953

Torabi, M., Ghosh, M., Myung, J. and Steel, M. (2021). Measurement error in linear regression models with fat tails and skewed errors. *Communications in Statistics – Theory and Methods* **0** 1–20. MR4597945

Van Rossum, G. and Drake Jr, F. L. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

Vats, D., Flegal, J. M. and Jones, G. L. (2019a). Multivariate output analysis for Markov chain Monte Carlo. *Biometrika* **106** 321–337. MR3949306

Vats, D., Flegal, J. M. and Jones, G. L. (2019b). Multivariate output analysis for Markov chain Monte Carlo. *Biometrika* **106** 321–337. `https://doi.org/10.1093/biomet/asz002`. MR3949306

Vidal, I. and Arellano-Valle, R. B. (2010). Bayesian inference for dependent elliptical measurement error models. *Journal of Multivariate Analysis* **101** 2587–2597. MR2719883

Vidal, I. and Iglesias, P. (2008). Comparison between a measurement error model and a linear model without measurement error. *Computational Statistics and Data Analysis* **53** 92–102. MR2528594

Wang, X. and Roy, V. (2018). Geometric ergodicity of Pólya-Gamma Gibbs sampler for Bayesian logistic regression with a flat prior. *Electronic Journal of Statistics* **12** 3295–3311. MR3861283