The Canadian Journal of Statistics
Vol. 00, No. 00, 2023, Pages 000–000
La revue canadienne de statistique

Bayesian instrumental variable estimation in linear measurement error models

Qi WANG[1], Lichun WANG[1*], and Liqun WANG[2]

[1]*Department of Statistics, Beijing Jiaotong University, Beijing, 100044, China*
[2]*Department of Statistics, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2*

Key words and phrases: Bayes estimator; instrumental variable; linear Bayes estimator; measurement error models; Metropolis–Hastings sampling.

MSC 2020: Primary 62F10; secondary 62F15.

Abstract: In this article, we study the problem of parameter estimation for measurement error models by combining the Bayes method with the instrumental variable approach, deriving the posterior distribution of parameters under different priors with known and unknown variance parameters, respectively, and calculating the Bayes estimator (BE) of the parameters under quadratic loss. However, it is difficult to obtain an explicit expression for BE because of the complex multiple integrals involved. Therefore, we adopt the linear Bayes method, which does not specify the form of the prior and avoids these complicated integral calculations, to obtain an expression for the linear Bayes estimator (LBE) for different priors. We prove that this LBE is superior to the two-stage least squares estimator under the mean squared error matrix criterion. Numerical simulations show that our LBE is very close to the real parameter whether the variance parameters are known or unknown, and it gradually approaches BE as the sample size increases. Our results indicate that this instrumental variable approach is valid for measurement error models.

Résumé: Les auteurs de ce travail abordent le problème d'estimation de paramètres de modèles d'erreur de mesure en combinant la méthode de Bayes avec une approche de variables instrumentales. Plus précisément, ils examinent les distributions a posteriori des paramètres en fonction de différentes hypothèses a priori, tout en prenant en compte les paramètres de variance connus et inconnus, et proposent une méthode pour calculer l'estimateur de Bayes (EB) des paramètres lorsque la fonction de perte est quadratique. Par ailleurs, étant donné qu'il est difficile d'exprimer explicitement l'estimateur de Bayes, et ce en raison des intégrales multiples complexes impliquées, les auteurs proposent de contourner ce problème en utilisant la méthode de Bayes linéaire. Cette dernière permet, en effet, d'obtenir une expression explicite de l'estimateur sous différentes lois a priori. Ils montrent ensuite que l'estimateur ainsi obtenu est plus performant que l'estimateur des moindres carrés à deux degrés basé sur le critère de la matrice des erreurs quadratiques. Des simulations numériques permettent de conclure que l'estimateur de Bayes linéaire est très proche du paramètre cible, peu importe que les paramètres de variance soient connus ou inconnus, et que plus la taille de l'échantillon augmente, plus l'estimateur de Bayes linéaire se rapproche de l'estimateur de Bayes. Enfin, les résultats du présent travail indiquent que cette approche de variables instrumentales est valide pour les modèles d'erreur de mesure.

1. INTRODUCTION

In statistics, the regression model is an important tool to study the correlation between variables. In order to simplify a study, researchers usually assume that there is no measurement error in the prediction variables. However, in the regression analysis of real data, it is common that

* *Corresponding author: wlc@amss.ac.cn*
[Correction added on 2 May 2023, after first online publication: Acknowledgements have been updated.]

some predictor variables are not directly observable or are measured with substantial error. Measurement error exists in many fields such as medicine, econometrics, and biostatistics. In general, measurement error is unknown and therefore ignored in research. It is well known that some naive estimation methods that ignore such measurement error will lead to biased and inconsistent estimators of unknown parameters. However, estimation procedures that account for measurement error adjust for this deficiency in traditional and naive regression analysis and hence are preferable. The study of the problem of prediction variables with measurement error has a long history, dating back to Adock (1877) and Cochran (1968). A comprehensive introduction to the problem may be found in Fuller (1987).

One of the commonly used methods of dealing with measurement error is the instrumental variable (IV) approach (e.g., Abarin et al., 2014; Abarin & Wang, 2012; Buzas & Stefanski, 1996; Carroll & Stefanski, 1994; Fuller, 1987; Guan et al., 2019; Guan & Wang, 2017; Wang, 2021; Wang & Hsiao, 2007, 2011; Xu, Ma & Wang, 2015). In practice, any variable that is correlated with the error-prone covariates but is independent of the measurement error and model error can serve as a valid IV, e.g., a second independent measurement, or repeated measurements at different time points in longitudinal studies. The IV approach was first studied by Wright (1928) and Reiersol (1945). Subsequently, Basmann (1957) and Zellner & Theil (1962) used this approach to construct the two-stage and three-stage least squares estimators, respectively, which popularized the IV approach in econometrics and statistics. Further, Hansen (1982) used the IV approach to develop the estimator of generalized moments.

As an alternative, some authors considered using model averaging for the two-stage least squares (TSLS) estimator; such methods have a Bayes flavour since they allow multiple models with different weights to be considered (e.g., Liu, 2019; Martins & Gabriel, 2014; Seng & Li, 2021). The Bayes method has also been used to study measurement error problems in linear models. For example, see Wang & Wei (2010), Vidal & Bolfarini (2011) and Li, Qiu & Ke (2020). To our knowledge, the Bayes method has also been combined with the IV approach (e.g., Conley et al., 2008; Hahn, He & Lopes, 2018; Kleibergen & Zivot, 2003; Kozumi, 2001; Lopes & Polson, 2014). However, the resulting estimator is usually a nonlinear function of the sample, which often involves complex multiple integrals and does not lead to useful explicit expressions. As an approximation of Bayes estimators (BEs), linear Bayes estimators (LBEs) not only avoid some complicated integral calculations and furnish explicit solutions but also do not depend on the specific prior form, while ensuring the accuracy of the resulting estimators. The linear Bayes method was proposed by Hartigan (1969). Rao (1973), Lamotte (1978), Heiligers (1993), Samaniego & Vestrup (1999), Pensky & Ni (2000), Wang & Singh (2014) and Jiang, Wang & Wang (2021) used this approach for different models and scenarios.

However, to the best of our knowledge, a combination of the linear Bayes method with the IV approach has not been studied in the literature. In this article we attempt to fill this gap. Specifically, we employ the Bayes and linear Bayes methods based on the traditional estimation method involving an IV to obtain more efficient estimators for the regression coefficients and variance parameters in a linear model. We also study the finite-sample properties of our proposed estimators through Monte Carlo simulations and compare their performance with that of the corresponding traditional and naive least squares estimators.

The rest of the article is organized as follows. In Section 2 we introduce the measurement error models and the TSLS estimator. In Section 3, using Bayes' theorem, we find the posterior distribution of the parameters in the measurement error models with known and unknown variance parameter, respectively; we also derive the BE for the parameters under quadratic loss. In Section 4, we define the LBE for the parameters and establish its superiority with respect to the TSLS estimator. Numerical comparisons between LBE, BE, and TSLS under different prior distributions may be found in Section 5. Section 6 is devoted to two data case studies. Section 7 outlines our conclusions. Various proofs and additional tables have been relegated to the Appendix.

2. THE MODEL

We consider the following linear regression model:

$$Y_i = \beta_0 + \beta_I^\top \tilde{X}_i + \epsilon_i, \quad i \in \{1, \dots, n\}, \tag{1}$$

where Y_i is the response variable, $\tilde{X}_i \in \mathbb{R}^{p-1}$ $(p > 1)$ is the vector of predictor variables, $\beta_I \in \mathbb{R}^{p-1}$ is the vector of regression parameters, and ϵ_i is the model random error. Further, let $Y = (Y_1, \dots, Y_n)^\top$, $\tilde{X} = (\tilde{X}_1, \dots, \tilde{X}_n)^\top$ and $\epsilon = (\epsilon_1, \dots, \epsilon_n)^\top$. Then, Equation (1) can be written as

$$Y = \beta_0 \mathbf{1}_n + \tilde{X}\beta_I + \epsilon = X\beta + \epsilon, \tag{2}$$

where $\mathbf{1}_n = (1, \dots, 1)^\top$, $X = (\mathbf{1}_n, \tilde{X})$, and $\beta = (\beta_0, \beta_I^\top)^\top$. In addition, we assume that the rank of X is p and $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2 I_n)$, where I_n is the n-dimensional identity matrix.

Suppose that $\tilde{X}$ is unobservable and that, instead, we observe

$$W = \tilde{X} + u, \tag{3}$$

where $W = (W_1, \dots, W_{p-1})$ and u is an $n \times (p-1)$ matrix of random measurement errors, which is assumed to follow a matrix normal distribution, denoted by $u \sim \mathcal{MN}\left(0, I_n, \sigma_u^2 I_{p-1}\right)$.

Remark 1. If a random matrix $X_0 \in \mathbb{R}^{m\times n}$ follows a matrix normal distribution $\mathcal{MN}(M, N, V)$, then its density is

$$f(X_0|M, N, V) = (2\pi)^{-mn/2}|V|^{-m/2}|N|^{-n/2} \times \exp\left\{-\frac{1}{2}\mathrm{tr}\left[V^{-1}(X_0 - M)^\top N^{-1}(X_0 - M)\right]\right\},$$

where $M \in \mathbb{R}^{m\times n}$, $N \in \mathbb{R}^{m\times m}$, and $V \in \mathbb{R}^{n\times n}$ are positive definite matrices. A matrix normal distribution can be vectorized to obtain a multivariate normal distribution, i.e.,

$$X_0 \sim \mathcal{MN}(M, N, V) \Leftrightarrow \mathrm{Vec}(X_0) \sim \mathcal{N}\left(\mathrm{Vec}(M), V \otimes N\right),$$

where $\mathrm{Vec}(\cdot)$ denotes the vectorization of the matrix argument, and $\otimes$ is the Kronecker product operation.

The IV is simply a variable that is correlated with $\tilde{X}$ but uncorrelated with ϵ and u. We assume that $\tilde{Z}$ is the IV and is related to $\tilde{X}$ through

$$\tilde{X} = \mathbf{1}_n\alpha_0^\top + \tilde{Z}\alpha_I + e = Z\alpha + e, \tag{4}$$

where $Z = (\mathbf{1}_n, \tilde{Z})$ is an $n \times q$ matrix of rank q, $q \geq p - 1$, $\alpha_0 \in \mathbb{R}^{p-1}$ denotes the intercept, and $\alpha = (\alpha_0, \alpha_I^\top)^\top$ is a $q \times (p-1)$ parameter matrix with rank $p - 1$. $e = (e_1, \dots, e_{p-1})$ is random error, with e_j $(j \in \{1, \dots, p-1\})$ being n-dimensional error vectors. We assume that e has a matrix normal distribution, i.e., $e \sim \mathcal{MN}(0, I_n, \sigma_e^2 I_{p-1})$.

Thus, incorporating Equation (4) into Equation (3) yields

$$W = \mathbf{1}_n\alpha_0^\top + \tilde{Z}\alpha_I + \upsilon_1 = Z\alpha + \upsilon_1, \tag{5}$$

where $\upsilon_1 = e + u \sim \mathcal{MN}\left(0, I_n, \tau_1^2 I_{p-1}\right)$ with $\tau_1^2 = \sigma_e^2 + \sigma_u^2$.

Substituting Equation (4) in Equation (2) yields

$$Y = \gamma_0 \mathbf{1}_n + \tilde{Z}\gamma_I + \upsilon_2 = Z\gamma + \upsilon_2, \tag{6}$$

where $\upsilon_2 = e\beta_I + \epsilon \sim \mathcal{N}(0, \tau_2^2 I_n)$ with $\tau_2^2 = \sigma_\epsilon^2 + \beta_I^\top \beta_I \sigma_e^2$, and $\gamma = (\gamma_0, \gamma_I)^\top$ with $\gamma_0 = \beta_0 + \alpha_0^\top \beta_I$, and $\gamma_I = \alpha_I \beta_I$, i.e., $\gamma = A\beta$ with

$$A = \begin{pmatrix} 1 & \alpha_0^\top \\ 0 & \alpha_I \end{pmatrix}.$$

To estimate the model parameters using the IV, the usual approach is to use the TSLS method. First, using least squares, obtain estimates of the parameters in Equation (5). Next, substitute these estimates into Equation (6). Finally, derive the required estimates using least squares a second time.

The TSLS estimator for α, say $\hat{\alpha}_{TSLS}$, is $\hat{\alpha}_{TSLS} = (Z^\top Z)^{-1} Z^\top W$. To simplify this expression, let $\tilde{W} = (\mathbf{1}_n, W)$ and $\tilde{\upsilon}_1 = (0_{n\times 1}, \upsilon_1)$; then Equation (5) can be written as $\tilde{W} = ZA + \tilde{\upsilon}_1$ and, therefore, the TSLS for A is

$$\hat{A}_{TSLS} = (Z^\top Z)^{-1} Z^\top \tilde{W} = \begin{pmatrix} 1 & \hat{\alpha}_{0TSLS}^\top \\ 0 & \hat{\alpha}_{ITSLS} \end{pmatrix}.$$

It follows that the TSLS for β is

$$\begin{aligned} \hat{\beta}_{TSLS} &= \left(\hat{A}_{TSLS}^\top Z^\top Z \hat{A}_{TSLS}\right)^{-1} \hat{A}_{TSLS}^\top Z^\top Y \\ &= \left(\tilde{W}^\top Z(Z^\top Z)^{-1} Z^\top Z (Z^\top Z)^{-1} Z^\top \tilde{W}\right)^{-1} \tilde{W}^\top Z (Z^\top Z)^{-1} Z^\top Y \\ &= \left(\tilde{W}^\top P_Z \tilde{W}\right)^{-1} \tilde{W}^\top P_Z Y, \end{aligned}$$

where $P_Z = Z(Z^\top Z)^{-1} Z^\top$.

Because of the limitation of the problem, we involve only the first-order moments of Y and W in the calculation. Therefore, σ_ϵ^2, σ_u^2, and σ_e^2 are clearly not identifiable, but τ_1^2 and τ_2^2 can be estimated. The TSLS estimators for τ_1^2 and τ_2^2, say $\hat{\tau}_{1_{TSLS}}^2$ and $\hat{\tau}_{2_{TSLS}}^2$, are given by

$$\hat{\tau}_{1_{TSLS}}^2 = \frac{\|\mathrm{Vec}(W) - (I_{p-1} \otimes Z)\mathrm{Vec}(\hat{\alpha}_{TSLS})\|^2}{(n-q)(p-1)}$$

and

$$\hat{\tau}_{2_{TSLS}}^2 = \frac{\|Y - Z\hat{A}_{TSLS}\hat{\beta}_{TSLS}\|^2}{n-q},$$

where $||\cdot||$ denotes the Euclidean norm.

However, many studies show that the least squares estimator is not always optimal in situations that commonly occur in practice. In the remainder of this article, we combine the linear Bayes method with the IV approach to estimate the parameters of the measurement error model, and investigate the statistical properties of this method of estimation.

3. THE BAYES ESTIMATOR

The Bayes method combines population information, sample information, and prior information to estimate parameters. From the Bayes viewpoint, the unknown parameters are random variables. In most cases, past experience about the parameters $\left(\alpha, \tau_1^2\right)$ and $\left(\beta, \tau_2^2\right)$ is available. Let $\pi\left(\beta, \tau_2^2\right)$ be the joint prior of β and τ_2^2, and let the loss function be given by

$$L_0(\hat{\phi}, \phi) = (\hat{\phi} - \phi)^\top D_0(\hat{\phi} - \phi), \tag{7}$$

where D_0 is a positive definite matrix and $\hat{\phi}$ denotes any estimator for $\phi = (\beta^\top, \tau_2^2)^\top$. Then by Bayes' theorem, the BEs of β and τ_2^2 are

$$\hat{\beta}_B = \iint \beta f\left(\beta, \tau_2^2 \mid Y\right) d\beta d\tau_2^2$$

and

$$\hat{\tau}_{2_B}^2 = \iint \tau_2^2 f\left(\beta, \tau_2^2 \mid Y\right) d\beta d\tau_2^2,$$

where $f(\beta, \tau_2^2 \mid Y)$ is the joint posterior density of β and τ_2^2 given Y. The BEs of α and τ_1^2 can be obtained in a similar way. Because of the issue concerning the identifiability of the variance parameters, we investigate the BEs of various parameters under different priors with known and unknown variance parameters.

3.1. Known Variance Parameters

In this case, only the regression parameters α and β need to be estimated. Let the prior distribution of α be a matrix normal distribution $\mathcal{MN}(M, N, V)$ and the prior of β be a p-dimensional normal distribution $\mathcal{N}(\mu, \Sigma)$, i.e.,

$$\begin{aligned}\pi(\alpha) &= (2\pi)^{-(pq-q)/2}|V|^{-q/2}|N|^{-(p-1)/2} \\ &\quad \times \exp\left[-\frac{1}{2}\mathrm{tr}\left\{V^{-1}(\alpha - M)^\top N^{-1}(\alpha - M)\right\}\right], \\ \pi(\beta) &= (2\pi)^{-p/2}|\Sigma|^{-1/2}\exp\left\{-\frac{1}{2}(\beta - \mu)^\top\Sigma^{-1}(\beta - \mu)\right\}.\end{aligned}$$

Note that $v_1 \sim \mathcal{MN}\left(0, I_n, \tau_1^2 I_{p-1}\right)$ and $W \sim \mathcal{MN}\left(Z\alpha, I_n, \tau_1^2 I_{p-1}\right)$ accordingly. Then the likelihood function of α is given, up to a proportionality constant, by

$$\begin{aligned}f(W \mid \alpha) \propto \exp&\left[-\frac{\mathrm{tr}\left\{(W - Z\hat{\alpha}_{TSLS})^\top(W - Z\hat{\alpha}_{TSLS})\right\}}{2\tau_1^2}\right] \\ &\times \exp\left[-\frac{\mathrm{tr}\left\{(\hat{\alpha}_{TSLS} - \alpha)^\top Z^\top Z(\hat{\alpha}_{TSLS} - \alpha)\right\}}{2\tau_1^2}\right].\end{aligned}$$

According to Bayes' theorem, the posterior density of α, say $f(\alpha|W)$, is proportional to $\pi(\alpha)f(W|\alpha)$. Hence, the BE $\hat{\alpha}_B$ can be obtained by evaluating some multiple integrals, and

$\hat{A}_B$ can be obtained accordingly. Substituting $\hat{A}_B$ into Equation (6) yields the following new model:

$$Y = Z\hat{A}_B\beta + \upsilon_2,$$

where $\upsilon_2 \sim \mathcal{N}(0, \tau_2^2 I_n)$. Thus, $Y \sim \mathcal{N}(Z\hat{A}_B\beta, \tau_2^2 I_n)$. Again, using Bayes' theorem, we derive the posterior density for β via similar calculations.

3.2. Unknown Variance Parameters

In the case of unknown variance parameters, we outline a method for obtaining the BEs of the parameters when the priors are normal-gamma and normal-uniform, respectively.

Case 1: Normal-gamma priors

Suppose that the prior distributions of α and β are normal, and the priors for σ_ϵ^2, σ_u^2, and σ_e^2 are gamma distributions, i.e.,

$$\alpha \sim \mathcal{MN}(M, N, V), \quad \beta \sim \mathcal{N}(\mu, \Sigma),$$
$$\sigma_\epsilon^2 \sim \mathcal{G}(\lambda_\epsilon, t_\epsilon), \quad \sigma_u^2 \sim \mathcal{G}(\lambda_u, t_u), \quad \sigma_e^2 \sim \mathcal{G}(\lambda_e, t_e).$$

By the identifiability of σ_ϵ^2, σ_u^2, and σ_e^2, we need to compute the prior distribution of τ_1^2 and τ_2^2 via the distribution of the sum of independent gamma variables (see Moschopoulos, 1985). At the same time, since τ_2^2 and β are correlated, the conditional density should be considered when calculating the joint prior distribution. We have

$$\pi_1(\tau_1^2) = C_1 \sum_{k=0}^{\infty} \delta_{1_k} \frac{(\tau_1^2)^{\lambda_u+\lambda_e+k-1} \exp(-\tau_1^2/\varphi_1)}{\Gamma(\lambda_u+\lambda_e+k)\varphi_1^{\lambda_u+\lambda_e+k}},$$
$$\pi_1(\tau_2^2 \mid \beta) = C_2 \sum_{k=0}^{\infty} \delta_{2_k} \frac{(\tau_2^2)^{\lambda_\epsilon+\lambda_e+k-1} \exp(-\tau_2^2/\varphi_2)}{\Gamma(\lambda_\epsilon+\lambda_e+k)\varphi_2^{\lambda_\epsilon+\lambda_e+k}},$$

where $\varphi_1 = \min\{t_u, t_e\}$, $\varphi_2 = \min\{t_\epsilon, \beta_I^\top \beta_I t_e\}$, and

$$C_1 = \left(\frac{\varphi_1}{t_u}\right)^{\lambda_u} \left(\frac{\varphi_1}{t_e}\right)^{\lambda_e}, \quad C_2 = \left(\frac{\varphi_2}{t_\epsilon}\right)^{\lambda_\epsilon} \left(\frac{\varphi_2}{\beta_I^\top \beta_I t_e}\right)^{\lambda_e}$$

with

$$\delta_{1_{k+1}} = \frac{1}{k+1} \sum_{i=1}^{k+1} \left\{ \lambda_u \left(1 - \frac{\varphi_1}{t_u}\right)^i + \lambda_e \left(1 - \frac{\varphi_1}{t_e}\right)^i \right\} \delta_{1_{k+1-i}}$$

and

$$\delta_{2_{k+1}} = \frac{1}{k+1} \sum_{i=1}^{k+1} \left\{ \lambda_\epsilon \left(1 - \frac{\varphi_2}{t_\epsilon}\right)^i + \lambda_e \left(1 - \frac{\varphi_2}{\beta_I^\top \beta_I t_e}\right)^i \right\} \delta_{2_{k+1-i}},$$

where $k \in \{0, 1, 2, \dots\}$ and $\delta_{1_0} = \delta_{2_0} = 0$.

Then the joint prior of α and τ_1^2 is $\pi_1(\alpha, \tau_1^2) = \pi(\alpha)\pi_1(\tau_1^2)$. Note that $W \sim \mathcal{MN}(Z\alpha, I_n, \tau_1^2 I_{p-1})$; hence we can obtain the joint posterior of α and τ_1^2. Inserting $\hat{A}_{B1}$ into Equation (6), where $\hat{A}_{B1}$ can be calculated by $\hat{\alpha}_{B1}$, we have

$$Y = Z\hat{A}_{B1}\beta + \upsilon_2,$$

which yields $Y \sim \mathcal{N}(Z\hat{A}_{B1}\beta, \tau_2^2 I_n)$. Given that the joint prior of β and τ_2^2 can be derived via $\pi_1(\beta, \tau_2^2) = \pi(\beta)\pi_1(\tau_2^2|\beta)$, by Bayes' theorem, the joint posterior of β and τ_2^2 can be obtained subsequently.

Case 2: Normal-uniform priors

Let the prior distributions of α and β be normal distributions, and the prior distributions of σ_ϵ^2, σ_u^2, and σ_e^2 be uniform distributions, i.e.,

$$\alpha \sim \mathcal{MN}(M, N, V), \quad \beta \sim \mathcal{N}(\mu, \Sigma),$$

$$\sigma_\epsilon^2 \sim \mathcal{U}(a_\epsilon, b_\epsilon), \quad \sigma_u^2 \sim \mathcal{U}(a_u, b_u), \quad \sigma_e^2 \sim \mathcal{U}(a_e, b_e).$$

We need to obtain the prior distribution of the sum of τ_1^2 and τ_2^2 based on two non-identically-distributed uniform random variables (see Bradley & Gupta, 2002), which are

$$\pi_2(\tau_1^2) = \frac{1}{2^3 l_u l_e} \sum_{\varepsilon_1, \varepsilon_2 \in \{-1,1\}} \varepsilon_1 \varepsilon_2 (c_1) \text{sign}(c_1),$$

$$\pi_2(\tau_2^2 \mid \beta) = \frac{1}{2^3 \beta_I^\top \beta_I l_\epsilon l_e} \sum_{\varepsilon_1, \varepsilon_2 \in \{-1,1\}} \varepsilon_1 \varepsilon_2 (c_2) \text{sign}(c_2),$$

where $c_1 = \tau_1^2 + \varepsilon_1 l_u + \varepsilon_2 l_e - h_u - h_e$, $c_2 = \tau_2^2 + \varepsilon_1 l_\epsilon + \varepsilon_2 \beta_I^\top \beta_I l_e - h_\epsilon - \beta_I^\top \beta_I h_e$, and

$$\begin{aligned} a_\epsilon &= h_\epsilon - l_\epsilon, b_\epsilon = h_\epsilon + l_\epsilon; \\ a_u &= h_u - l_u, b_u = h_u + l_u; \\ a_e &= h_e - l_e, b_e = h_e + l_e; \end{aligned} \qquad \text{sign}(y) = \begin{cases} 1 & \text{if } y > 0; \\ 0 & \text{if } y = 0; \\ -1 & \text{if } y < 0. \end{cases}$$

According to Bayes' theorem, the joint posterior density of α and τ_1^2 can be calculated. Note that $\hat{A}_{B2}$ can be obtained from $\hat{\alpha}_{B2}$ by substituting $\hat{\alpha}_{B2}$ into Equation (6) and eventually deriving the joint posterior of β and τ_2^2.

So far, we have obtained the joint posteriors of parameters under different priors, and we need to obtain the BEs via some multiple integral calculations. However, they are challenging to evaluate. Simulation-based methods such as the Gibbs sampling procedure and the Metropolis–Hastings method represent one approach to solving such problems. The Metropolis–Hastings algorithm (see Hastings, 1970; Metropolis et al., 1953) constructs an aperiodic and irreducible Markov chain so that its stationary distribution is equal to the target distribution, which is just the posterior distribution in Bayes inference. The idea is to find a simple distribution that approximates the posterior distribution $f(\theta|y)$, called the proposal density and denoted as $g(\theta)$, and then selects an initial value θ_0 for the parameter θ. The specific steps involved are the following:

Step 1: Simulate a candidate sample θ^* from the proposal density $g(\theta^*|\theta_{i-1})$.

Step 2: Calculate the acceptance probability

$$p = \min \left\{ \frac{f(\theta = \theta^* | y) g(\theta_{i-1} | \theta^*)}{f(\theta = \theta_{i-1} | y) g(\theta^* | \theta_{i-1})}, 1 \right\},$$

where θ_{i-1} represents the $(i-1)$th value of the parameter sequence.
Step 3: Accept $\theta_i = \theta^*$ with probability p, and accept $\theta_i = \theta_{i-1}$ with probability $1 - p$.
Step 4: The posterior samples $\theta_1, \theta_2, \ldots, \theta_n$ can be obtained by repeating Steps 1–3 n times.

In Section 5 we obtain the BEs of the various parameters under different priors using suitable versions of this algorithm.

4. THE LINEAR BAYES ESTIMATOR

The BEs that we identified in Section 3 can produce numerical solutions via simulations, and, to a certain extent, the accuracy of the results can be guaranteed. However, we cannot derive explicit expressions for those BEs. Hence, in this case BEs are somewhat intricate and not easy to use. In addition, parameter estimation via the Bayes method needs to specify some priors, which can lead to practical difficulties. Therefore, this section introduces the linear Bayes method, which does not rely on the specific form of the prior but depends solely on its moments, and can result in a linear approximation of Bayes estimation. In what follows, we combine this linear approximation with the use of an IV to estimate the parameters of the measurement error models.

4.1. The Proposed LBE

Denote $T = \left(\mathrm{Vec}(\hat{\alpha})^\top, \hat{\gamma}^\top, (\hat{\tau}^2)^\top\right)^\top$, $\mathrm{Vec}(\hat{\alpha}) = \left(I_{p-1} \otimes (Z^\top Z)^{-1} Z^\top\right) \mathrm{Vec}(W)$, $\hat{\gamma} = (Z^\top Z)^{-1} Z^\top Y$, $\hat{\tau}^2 = \left(\hat{\tau}_1^2, \hat{\tau}_2^2\right)^\top$, and

$$\hat{\tau}_1^2 = \frac{\|\mathrm{Vec}(W) - (I_{p-1} \otimes Z)\mathrm{Vec}(\hat{\alpha})\|^2}{(n-q)(p-1)}, \quad \hat{\tau}_2^2 = \frac{\|Y - Z\hat{\gamma}\|^2}{n-q}.$$

Set $\theta = \left(\mathrm{Vec}(\alpha)^\top, \gamma^\top, (\tau^2)^\top\right)^\top$. We assume that the prior $G(\theta)$ belongs to the distribution family:

$$G = \left\{ G(\theta) : \mathrm{E}\left[\|\mathrm{Vec}(\alpha)\|^2 + \|\gamma\|^2 + \|\tau^2\|^2\right] < \infty \right\},$$

which includes many common distributions such as the normal, uniform, and gamma.

Define the LBE of θ, say $\hat{\theta}_{LB}$, to be of the form $BT + b$ satisfying

$$\mathrm{E}_{(T,\theta)}(\hat{\theta}_{LB} - \theta) = 0, \quad R(\hat{\theta}_{LB}, \theta) = \min_{B,b} \mathrm{E}_{(T,\theta)} L(BT + b, \theta),$$

where $\mathrm{E}_{(T,\theta)}$ denotes the expectation of the joint distribution of T and θ, and the loss function is given by

$$L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^\top D (\hat{\theta} - \theta),$$

where D is a positive definite matrix. As in the case of the loss identified in Equation (7), the matrix D has no effect on the LBE.

Theorem 1. *If $n \geq p$, then*

$$\hat{\theta}_{LB} = T - KH(T - \mathrm{E}\theta), \tag{8}$$

with $B = I - KH$ *and* $b = KH\mathrm{E}(\theta)$, *where* $H = [K + \mathrm{Cov}(\theta)]^{-1}$ *and*

$$K = \mathrm{E}[\mathrm{Cov}(T|\theta)]$$
$$= \begin{pmatrix} (I_{p-1} \otimes (Z^\top Z)^{-1})E(\tau_1^2) & E(\beta_I \otimes (Z^\top Z)^{-1}\sigma_e^2) & \multicolumn{2}{c}{0_{(p-1)q\times 2}} \\ E(\beta_I^\top \otimes (Z^\top Z)^{-1}\sigma_e^2) & (Z^\top Z)^{-1}E(\tau_2^2) & \multicolumn{2}{c}{0_{q\times 2}} \\ 0_{2\times(p-1)q} & 0_{2\times q} & \begin{matrix} \frac{2E(\tau_1^4)}{(n-q)(p-1)} & \frac{2E(\beta_I^\top\beta_I\sigma_e^4)}{(n-q)(p-1)} \\ \frac{2E(\beta_I^\top\beta_I\sigma_e^4)}{(n-q)(p-1)} & \frac{2E(\tau_2^4)}{n-q} \end{matrix} \end{pmatrix}.$$

Since $\gamma = A\beta$, the LBE of β, say $\hat{\beta}_{LB}$, is

$$\hat{\beta}_{LB} = (\hat{A}_{LB}^\top \hat{A}_{LB})^{-1}\hat{A}_{LB}^\top \hat{\gamma}_{LB}, \tag{9}$$

where $\hat{A}_{LB}$ and $\hat{\gamma}_{LB}$ can be obtained from $\hat{\theta}_{LB}$.

4.2. The Superiority of the LBE

After obtaining the preceding expression for the LBEs of the model parameters, we proceed to investigate the superiority of the LBE as a method of estimating those same parameters. There are many criteria for evaluating estimators. In this article, we will use the mean squared error matrix (MSEM) criterion to compare the LBE and TSLS estimators of the model parameters, where

$$\begin{aligned} \mathrm{MSEM}(\hat{\theta}_{LB}) &= \mathrm{E}_{(T,\theta)}\left[\left(\hat{\theta}_{LB} - \theta\right)\left(\hat{\theta}_{LB} - \theta\right)^\top\right] \\ &= \mathrm{E}\left[\mathrm{Cov}\left(\left(\hat{\theta}_{LB} - \theta\right)|\theta\right)\right] + \mathrm{Cov}\left(\mathrm{E}\left[\left(\hat{\theta}_{LB} - \theta\right)|\theta\right]\right). \end{aligned}$$

Denote the TSLS of θ by $\hat{\theta}_{TSLS}$, and set $T = \left(\mathrm{Vec}(\hat{\alpha})^\top, \hat{\gamma}^\top, (\hat{\tau}^2)^\top\right)^\top$; then

$$\hat{\theta}_{TSLS} = \begin{pmatrix} I_{n(p-1)} & 0 & 0 \\ 0 & I_n & 0 \\ 0 & 0 & I_2 \end{pmatrix}\begin{pmatrix} \mathrm{Vec}(\hat{\alpha}) \\ \hat{\gamma} \\ \hat{\tau}^2 \end{pmatrix} = I_{np+2}T = T. \tag{10}$$

Theorem 2. *Let* $\hat{\theta}_{LB}$ and $\hat{\theta}_{TSLS}$ be identified in Equations (8) *and* (10), respectively. If $n \geq q$, *then* $\hat{\theta}_{LB}$ *is superior to* $\hat{\theta}_{TSLS}$ *with respect to the MSEM criterion, i.e.,* $\mathrm{MSEM}(\hat{\theta}_{LB}) \leq \mathrm{MSEM}(\hat{\theta}_{TSLS})$.

5. SIMULATION STUDIES

When $p = 2$, the model specified in Equation (1) becomes a simple linear regression model with $\beta = (\beta_0, \beta_1)^\top$. Set $\beta = (-4, 0.6)^\top$ and $\alpha = (\alpha_0, \alpha_1)^\top = (5, -1)^\top$. Suppose the true values of σ_ϵ^2, σ_u^2, and σ_e^2 are given by the following distributions:

$$\epsilon_i \sim \mathcal{N}(0, 16), \quad u_i \sim \mathcal{N}(0, 16), \quad e_i \sim \mathcal{N}(0, 25).$$

At the same time, we assume that the IV z has a normal distribution, i.e., $z_i \sim \mathcal{N}(0, 25)$. Then we can use the dataset (y, w, z) to estimate the model parameters.

5.1. Known Variance Parameters

In this case, we also assume that the prior distributions of the regression parameters are normal distributions, i.e.,

$$\alpha_0 \sim \mathcal{N}\left(\mu_{01}, \eta_{01}^2\right), \alpha_1 \sim \mathcal{N}\left(\mu_{11}, \eta_{11}^2\right), \beta_0 \sim \mathcal{N}\left(\mu_{02}, \eta_{02}^2\right), \beta_1 \sim \mathcal{N}\left(\mu_{12}, \eta_{12}^2\right).$$

The joint posterior densities of the parameters α_0, α_1 and β_0, β_1 can be evaluated as was outlined previously in Section 3. Finally, we obtain the BEs of these four parameters via Metropolis–Hastings sampling.

Denote $\xi^* = (\alpha_0, \alpha_1, \gamma_0, \gamma_1)^\top$; according to Theorem 1, we know that an expression for the LBE of the parameter vector ξ^* equals

$$\hat{\xi^*}_{LB} = T^* - K^* H^* (T^* - \mathrm{E}\xi^*),$$

where $H^* = [K^* + \mathrm{Cov}(\xi^*)]^{-1}$, and $T^* = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\gamma}_0, \hat{\gamma}_1)^\top$. From Equation (9), we have the LBE of $\xi = (\alpha_0, \alpha_1, \beta_0, \beta_1)^\top$, say $\hat{\xi}_{LB}$. Thus, the simulation value of $\hat{\xi}_{LB}$ can be obtained via

$$\mathrm{E}(\xi^*) = (\mu_{01}, \mu_{11}, \mu_{02} + \mu_{12}\mu_{01}, \mu_{12}\mu_{11})^\top,$$

where

$$K^* = \begin{bmatrix} (Z^\top Z)^{-1}\tau_1^2 & (Z^\top Z)^{-1}\sigma_e^2 \mathrm{E}(\beta_1) \\ (Z^\top Z)^{-1}\sigma_e^2 \mathrm{E}(\beta_1) & (Z^\top Z)^{-1}\left(\mathrm{E}(\beta_1^2)\sigma_e^2 + \sigma_\epsilon^2\right) \end{bmatrix}$$
$$= \begin{bmatrix} (Z^\top Z)^{-1}\tau_1^2 & (Z^\top Z)^{-1}\sigma_e^2 \mu_{12} \\ (Z^\top Z)^{-1}\sigma_e^2 \mu_{12} & (Z^\top Z)^{-1}\left(\left(\mu_{12}^2 + \eta_{12}^2\right)\sigma_e^2 + \sigma_\epsilon^2\right) \end{bmatrix}.$$

Also

$$\mathrm{Cov}(\xi^*) = \begin{bmatrix} \mathrm{Cov}(\alpha_0,\alpha_0) & \mathrm{Cov}(\alpha_0,\alpha_1) & \mathrm{Cov}(\alpha_0,\gamma_0) & \mathrm{Cov}(\alpha_0,\gamma_1) \\ \mathrm{Cov}(\alpha_1,\alpha_0) & \mathrm{Cov}(\alpha_1,\alpha_1) & \mathrm{Cov}(\alpha_1,\gamma_0) & \mathrm{Cov}(\alpha_1,\gamma_1) \\ \mathrm{Cov}(\gamma_0,\alpha_0) & \mathrm{Cov}(\gamma_0,\alpha_1) & \mathrm{Cov}(\gamma_0,\gamma_0) & \mathrm{Cov}(\gamma_0,\gamma_1) \\ \mathrm{Cov}(\gamma_1,\alpha_0) & \mathrm{Cov}(\gamma_1,\alpha_1) & \mathrm{Cov}(\gamma_1,\gamma_0) & \mathrm{Cov}(\gamma_1,\gamma_1) \end{bmatrix}$$
$$= \begin{bmatrix} \eta_{01}^2 & 0 & \mu_{12}\eta_{01}^2 & 0 \\ 0 & \eta_{11}^2 & 0 & \mu_{12}\eta_{11}^2 \\ \mu_{12}\eta_{01}^2 & 0 & \eta_{02}^2 + \eta_{12}^2\eta_{01}^2 + \mu_{01}^2\eta_{12}^2 + \mu_{12}^2\eta_{01}^2 & \mu_{01}\mu_{11}\eta_{12}^2 \\ 0 & \mu_{12}\eta_{11}^2 & \mu_{01}\mu_{11}\eta_{12}^2 & \eta_{12}^2\eta_{11}^2 + \mu_{11}^2\eta_{12}^2 + \mu_{12}^2\eta_{11}^2 \end{bmatrix}.$$

We also calculate the distances between BE, LBE, TSLS, and the true values of the various parameters to evaluate the advantages and disadvantages of the estimators. The formula for the distance between the estimated value and the true value is

$$\|\hat{\xi} - \xi\| = \sqrt{(\hat{\alpha}_0 - \alpha_0)^2 + (\hat{\alpha}_1 - \alpha_1)^2 + (\hat{\beta}_0 - \beta_0)^2 + (\hat{\beta}_1 - \beta_1)^2}. \tag{11}$$

TABLE 1: Distance between the estimated value of regression parameters and the true value under different prior hyperparameters.

n		μ	η	$\|\hat{\xi}_B - \xi\|$	$\|\hat{\xi}_{LB} - \xi\|$	$\|\hat{\xi}_{TSLS} - \xi\|$
50	pr1	$(5, -1, -4, 0.6)^\top$	$(1, 1, 1, 1)^\top$	0.8377	1.0997	1.5593
	pr2	$(5, -1, -4, 0.6)^\top$	$(9, 9, 9, 9)^\top$	1.4133	1.4706	
	pr3	$(5, -1, -4, 0.6)^\top$	$(25, 25, 25, 25)^\top$	1.490	1.5254	
100	pr1	$(5, -1, -4, 0.6)^\top$	$(1, 1, 1, 1)^\top$	0.7185	0.9763	1.0381
	pr2	$(5, -1, -4, 0.6)^\top$	$(9, 9, 9, 9)^\top$	0.8298	0.9851	
	pr3	$(5, -1, -4, 0.6)^\top$	$(25, 25, 25, 25)^\top$	1.0350	1.0218	
500	pr1	$(5, -1, -4, 0.6)^\top$	$(1, 1, 1, 1)^\top$	0.3173	0.3308	0.3462
	pr2	$(5, -1, -4, 0.6)^\top$	$(9, 9, 9, 9)^\top$	0.3436	0.3444	
	pr3	$(5, -1, -4, 0.6)^\top$	$(25, 25, 25, 25)^\top$	0.3440	0.3456	

Note: $\mu = (\mu_{01}, \mu_{11}, \mu_{02}, \mu_{12})^\top$, $\eta = (\eta_{01}^2, \eta_{11}^2, \eta_{02}^2, \eta_{12}^2)^\top$, and *pr* stands for the prior.

For the sample size $n \in \{50, 100, 500\}$, the distances between the estimated value of the regression parameters and the true value under different priors are reported in Table 1. It can be seen that when the prior hyperparameters are the same, as the sample size n increases, the distances between BE, LBE, TSLS and the true values all decrease, and also $\|\hat{\xi}_B - \xi\| < \|\hat{\xi}_{LB} - \xi\| < \|\hat{\xi}_{TSLS} - \xi\|$, i.e., LBE is close to BE and superior to TSLS. When the sample size n is the same, as the prior hyperparameter increases, both $\|\hat{\xi}_B - \xi\|$ and $\|\hat{\xi}_{LB} - \xi\|$ increase but are still less than $\|\hat{\xi}_{TSLS} - \xi\|$. With respect to the concentration of prior information, the distance between $\|\hat{\xi}_{LB} - \xi\|$ and $\|\hat{\xi}_B - \xi\|$ gradually decreases, indicating that LBE appears to provide a good approximation to BE and seems to be relatively robust. This conclusion is further verified by the curve of the distances between the estimated value and the true value of each model parameter.

To reinforce the apparent superiority of LBE, we also calculate the root mean squared error (RMSE) and bias of the estimator for each parameter. For example, the RMSE of $\hat{\beta}_1$ is given by

$$\text{RMSE}(\hat{\beta}_1) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}_{1i} - \beta_1)^2}.$$

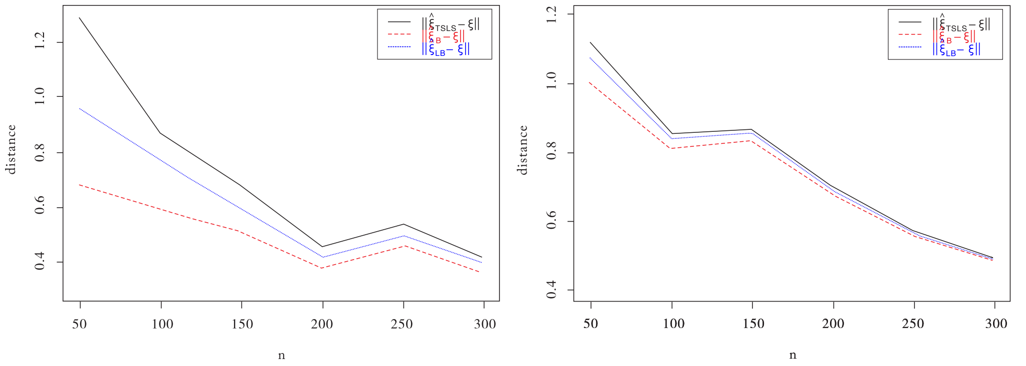
The RMSEs of the various estimators for the remaining model parameters are calculated similarly; the observed results may be found in Tables A1 and A2 in the Appendix. We conclude that both the RMSE and the bias of each estimator are small and that the RMSE values for the BE and LBE parameter estimators increase as the prior hyperparameters increase. More importantly, their RMSE values are smaller than the corresponding values for the TSLS estimators. We also observe that the RMSE values are decreasing functions with respect to increasing sample size. Thus, we conclude that both BE and LBE of the model parameters are superior to their corresponding TSLS competitors, as the results reported in Table 1 already suggest (Figure 1).

5.2. Unknown Variance Parameters

Case 1: Normal-gamma priors

Let the prior distributions of the various model parameters be

$$\alpha_0 \sim \mathcal{N}\left(\mu_{01}, \eta_{01}^2\right), \alpha_1 \sim \mathcal{N}\left(\mu_{11}, \eta_{11}^2\right), \beta_0 \sim \mathcal{N}\left(\mu_{02}, \eta_{02}^2\right), \beta_1 \sim \mathcal{N}\left(\mu_{12}, \eta_{12}^2\right),$$
$$\sigma_\epsilon^2 \sim \mathcal{G}(\lambda_\epsilon, t_\epsilon), \quad \sigma_u^2 \sim \mathcal{G}(\lambda_u, t_u), \quad \sigma_e^2 \sim \mathcal{G}(\lambda_e, t_e).$$



(a) pr 1 (b) pr 2

FIGURE 1: Distance between the estimated value of the regression parameters and the true value as a function of sample size.

The joint posterior density of the parameters can be found using the methods outlined in Section 3.2. Let $\theta^* = \left(\alpha^\top, \gamma^\top, \tau_1^2, \tau_2^2\right)^\top$. It follows from Theorem 1 that the expression for $\hat{\theta}^*_{LB}$ is

$$\hat{\theta}^*{}_{LB} = T - KH(T - \mathrm{E}\theta^*),$$

where $H = [K + \mathrm{Cov}(\theta^*)]^{-1}$ and $T = \left(\hat{\alpha}^\top, \hat{\gamma}^\top, \tau_1^2, \tau_2^2\right)^\top$. Using Equation (8), we have $\hat{\theta}_{LB}$, and therefore

$$\mathrm{E}(\theta) = \left(\mu_{01}, \mu_{11}, \mu_{02} + \mu_{12}\mu_{01}, \mu_{12}\mu_{11}, \lambda_u t_u + \lambda_e t_e, \lambda_\epsilon t_\epsilon \left(\mu_{12}^2 + \eta_{12}^2\right) \lambda_e t_e\right)^\top,$$

$$K = \begin{bmatrix} (Z^\top Z)^{-1}(\lambda_u t_u + \lambda_e t_e) & (Z^\top Z)^{-1}\mu_{12}\lambda_e t_e & 0_{2\times 2} \\ (Z^\top Z)^{-1}\mu_{12}\lambda_e t_e & (Z^\top Z)^{-1}\lambda_\epsilon t_\epsilon\left(\mu_{12}^2 + \eta_{12}^2\right)\lambda_e t_e & 0_{2\times 2} \\ 0_{2\times 2} & 0_{2\times 2} & \begin{matrix} \dfrac{2\mathrm{E}\left(\tau_1^4\right)}{n-2} & \dfrac{2\mathrm{E}\left(\beta_1^2\sigma_e^4\right)}{n-2} \\ \dfrac{2\mathrm{E}\left(\beta_1^2\sigma_e^4\right)}{n-2} & \dfrac{2\mathrm{E}\left(\tau_2^4\right)}{n-2} \end{matrix} \end{bmatrix},$$

where

$$\begin{aligned} \mathrm{E}\left(\tau_1^4\right) &= \lambda_u t_u^2 + \lambda_u^2 t_u^2 + \lambda_e t_e^2 + \lambda_e^2 t_e^2 + 2\lambda_u t_u \lambda_e t_e, \\ \mathrm{E}\left(\tau_2^4\right) &= \lambda_\epsilon t_\epsilon^2 + \lambda_\epsilon^2 t_\epsilon^2 + \left(3\eta_{12}^4 + 6\mu_{12}^2\eta_{12}^2 + \mu_{12}^4\right)\left(\lambda_e t_e^2 + \lambda_e^2 t_e^2\right) \\ &\quad + 2\lambda_\epsilon t_\epsilon \lambda_e t_e \left(\mu_{12}^2 \eta_{12}\right), \\ \mathrm{E}\left(\beta_1^2\sigma_e^4\right) &= \left(\mu_{12}^2 + \eta_{12}^2\right)\left(\lambda_e t_e^2 + \lambda_e^2 t_e^2\right). \end{aligned}$$

TABLE 2: Distance between the estimated value and the true value under different hyperparameters.

n		$(\lambda_\epsilon, t_\epsilon)$	(λ_u, t_u)	(λ_e, t_e)	μ	η	$\lVert\hat{\theta}_B - \theta\rVert$	$\lVert\hat{\theta}_{LB} - \theta\rVert$	$\lVert\hat{\theta}_{TSLS} - \theta\rVert$
50	pr4	$(2, 4)$	$(4, 2)$	$(5, 3)$	$(5, -1, -4, 0.6)^\top$	$(1, 1, 1, 1)^\top$	5.3359	6.6931	8.6157
	pr5	$(2, 6)$	$(4, 4)$	$(5, 5)$	$(5, -1, -4, 0.6)^\top$	$(4, 4, 4, 4)^\top$	5.6937	7.4484	
	pr6	$(2, 8)$	$(4, 6)$	$(5, 7)$	$(5, -1, -4, 0.6)^\top$	$(9, 9, 9, 9)^\top$	6.1944	8.4200	
100	pr4	$(2, 4)$	$(4, 2)$	$(5, 3)$	$(5, -1, -4, 0.6)^\top$	$(1, 1, 1, 1)^\top$	4.9267	5.3671	6.1123
	pr5	$(2, 6)$	$(4, 4)$	$(5, 5)$	$(5, -1, -4, 0.6)^\top$	$(4, 4, 4, 4)^\top$	5.2384	5.9531	
	pr6	$(2, 8)$	$(4, 6)$	$(5, 7)$	$(5, -1, -4, 0.6)^\top$	$(9, 9, 9, 9)^\top$	5.9461	6.1229	
500	pr4	$(2, 4)$	$(4, 2)$	$(5, 3)$	$(5, -1, -4, 0.6)^\top$	$(1, 1, 1, 1)^\top$	1.9980	2.0961	2.1529
	pr5	$(2, 6)$	$(4, 4)$	$(5, 5)$	$(5, -1, -4, 0.6)^\top$	$(4, 4, 4, 4)^\top$	2.0835	2.1556	
	pr6	$(2, 8)$	$(4, 6)$	$(5, 7)$	$(5, -1, -4, 0.6)^\top$	$(9, 9, 9, 9)^\top$	2.1541	2.2879	

Note: $\mu = (\mu_{01}, \mu_{11}, \mu_{02}, \mu_{12})^\top$, $\eta = (\eta_{01}^2, \eta_{11}^2, \eta_{02}^2, \eta_{12}^2)^\top$, and *pr* means a prior.

Also

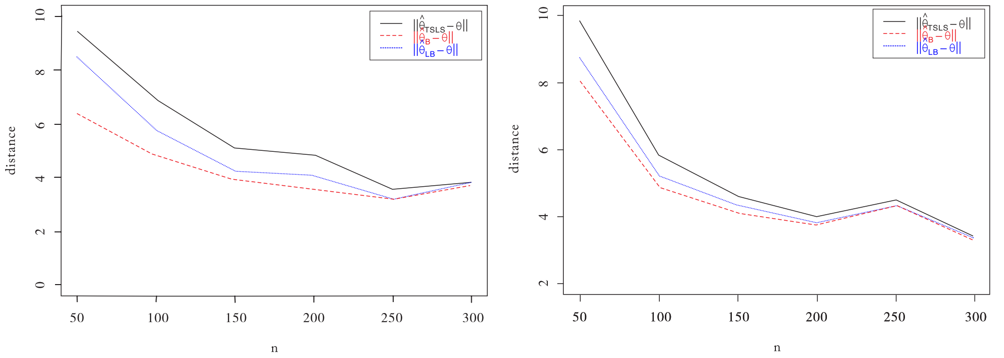
$$\mathrm{Cov}(\theta) = \begin{bmatrix} \eta_{01}^2 & 0 & \mu_{12}\eta_{01}^2 & 0 & 0 & 0 \\ 0 & \eta_{11}^2 & 0 & \mu_{12}\eta_{11}^2 & 0 & 0 \\ \mu_{12}\eta_{01}^2 & 0 & \nu_1 & \mu_{01}\mu_{11}\eta_{12}^2 & 0 & 2\lambda_e t_e \mu_{12}\mu_{01}\eta_{12}^2 \\ 0 & \mu_{12}\eta_{11}^2 & \mu_{01}\mu_{11}\eta_{12}^2 & \nu_2 & 0 & 2\lambda_e t_e \mu_{12}\mu_{11}\eta_{12}^2 \\ 0 & 0 & 0 & 0 & \nu_3 & \nu_5 \\ 0 & 0 & 2\lambda_e t_e \mu_{12}\mu_{01}\eta_{12}^2 & 2\lambda_e t_e \mu_{12}\mu_{11}\eta_{12}^2 & \nu_5 & \nu_4 \end{bmatrix},$$

where
$$\nu_1 = \eta_{02}^2 + \eta_{12}^2\eta_{01}^2 + \mu_{01}^2\eta_{12}^2 + \mu_{12}^2\eta_{01}^2,$$
$$\nu_2 = \eta_{12}^2\eta_{11}^2 + \mu_{11}^2\eta_{12}^2 + \mu_{12}^2\eta_{11}^2,$$
$$\nu_3 = \lambda_u t_u^2 + \lambda_e t_e^2,$$
$$\nu_4 = \lambda_\epsilon t_\epsilon^2 + \left(2\eta_{12}^4 + 4\mu_{12}^2\eta_{12}^2\right)\lambda_e^2 t_e^2 + \left(3\eta_{12}^4 + 6\mu_{12}^2\eta_{12}^2 + \mu_{12}^4\right)\lambda_e t_e^2,$$
$$\nu_5 = \left(\mu_{12}^2 + \eta_{12}^2\right)\lambda_e t_e^2.$$

Let the sample size n be 50, 100, and 500, respectively. The distance between the estimated value and the true value for various choices of the hyperparameters may be found in Table 2.

Based on the observed values reported in Table 2, it seems that regardless of the value n takes, we have $\lVert\hat{\theta}_B - \theta\rVert < \lVert\hat{\theta}_{LB} - \theta\rVert < \lVert\hat{\theta}_{TSLS} - \theta\rVert$. Furthermore, the values of LBE and BE are similar, and both of them are superior to their TSLS competitor. For constant n, with the increase of the prior hyperparameters, both $\lVert\hat{\theta}_B - \theta\rVert$ and $\lVert\hat{\theta}_{LB} - \theta\rVert$ are increasing functions of the prior hyperparameters, but $\lVert\hat{\theta}_{TSLS} - \theta\rVert$ is always greater. As the prior distributions of the model parameters become more concentrated, the distances $\lVert\hat{\theta}_B - \theta\rVert$ and $\lVert\hat{\theta}_{LB} - \theta\rVert$ decrease, indicating that the LBE approximation to the BE improves. Since LBE does not depend on the specific form of the prior distribution but depends only on its moments, the values of the prior hyperparameters have little influence on the LBE. Figure 2 is a plot of the observed results reported in Table 2.

In contrast to the situation where the variance parameters are known, we observe that unknown variances yield larger observed distances between the parameter estimates and the corresponding true values. This observation is reinforced by the RMSEs of the estimators for each parameter when the prior distributions are normal-gamma; see the observed values summarized in Tables A3 and A4 in the Appendix.



(a) pr 1 (a) pr 2

FIGURE 2: Distance between the estimated value and the true value as a function of sample size.

In summary, when the prior distributions are normal-gamma, the estimated values of the regression parameters are very similar and approximately equal to the corresponding estimates obtained when the values of the variance parameters are known. Occasionally, the bias exhibited by either BE or LBE is greater than that of the TSLS alternative. However, because the corresponding RMSE values for the variances are always smaller, we obtain the observed results reported in Table 2.

Case 2: Normal-uniform priors

Let the prior distributions of the unknown model parameters be

$$\alpha_0 \sim \mathcal{N}(\mu_{01}, \eta_{01}^2),\ \alpha_1 \sim \mathcal{N}(\mu_{11}, \eta_{11}^2),\ \beta_0 \sim \mathcal{N}(\mu_{02}, \eta_{02}^2),\ \beta_1 \sim \mathcal{N}(\mu_{12}, \eta_{12}^2),$$
$$\sigma_\epsilon^2 \sim \mathcal{U}(a_\epsilon, b_\epsilon), \quad \sigma_u^2 \sim \mathcal{U}(a_u, b_u), \quad \sigma_e^2 \sim \mathcal{U}(a_e, b_e).$$

The joint posterior densities can be obtained as outlined in Section 3.2. According to Theorem 1, we have

$$\mathrm{E}(\theta) = \left(\mu_{01}, \mu_{11}, \mu_{02} + \mu_{12}\mu_{01}, \mu_{12}\mu_{11}, \frac{a_u + b_u + a_e + b_e}{2}, \frac{a_\epsilon + b_\epsilon + (a_e + b_e)(\mu_{12}^2 + \eta_{12}^2)}{2} \right)^\top,$$

$$K = \begin{bmatrix} (Z^\top Z)^{-1}\dfrac{a_u + b_u + a_e + b_e}{2} & (Z^\top Z)^{-1}\dfrac{\mu_{12}(a_e + b_e)}{2} & 0_{2\times 2} \\ (Z^\top Z)^{-1}\dfrac{\mu_{12}(a_e + b_e)}{2} & (Z^\top Z)^{-1}(\mu_{12}^2 + \eta_{12}^2)\dfrac{a_\epsilon + b_\epsilon + a_e + b_e}{2} & 0_{2\times 2} \\ 0_{2\times 2} & 0_{2\times 2} & \begin{matrix} \dfrac{2\mathrm{E}(\tau_1^4)}{n-2} & \dfrac{2\mathrm{E}(\beta_1^2\sigma_e^4)}{n-2} \\ \dfrac{2\mathrm{E}(\beta_1^2\sigma_e^4)}{n-2} & \dfrac{2\mathrm{E}(\tau_2^4)}{n-2} \end{matrix} \end{bmatrix},$$

where $\mathrm{E}(\tau_1^4) = \left(\dfrac{a_u + b_u + a_e + b_e}{2}\right)^2 + \dfrac{(b_u - a_u)^2}{12} + \dfrac{(b_e - a_e)^2}{12}$,

$$\mathrm{E}(\tau_2^4) = \frac{7a_\epsilon^2 + 7b_\epsilon^2 - 10a_\epsilon b_\epsilon}{12} + \frac{3\eta_{12}^4 + 6\mu_{12}^2\eta_{12}^2}{12},$$
$$\times \frac{(\mu_{12}^4)(7b_e^2 + 7b_e^2 - 10a_e b_e)}{12} + \frac{(\mu_{12}^2 + \eta_{12}^2)(a_\epsilon + b_\epsilon)(b_e - a_e)}{2},$$
$$\mathrm{E}(\beta_1^2\sigma_e^4) = (\mu_{12}^2 + \eta_{12}^2)\left[\frac{(b_e - a_e)^2}{12} + \left(\frac{a_e + b_e}{2}\right)^2\right].$$

Thus

$$\mathrm{Cov}(\theta) = \begin{bmatrix} \eta_{01}^2 & 0 & \mu_{12}\eta_{01}^2 & 0 & 0 & 0 \\ 0 & \eta_{11}^2 & 0 & \mu_{12}\eta_{11}^2 & 0 & 0 \\ \mu_{12}\eta_{01}^2 & 0 & v_1 & \mu_{01}\mu_{11}\eta_{12}^2 & 0 & (a_e+b_e)\mu_{12}\mu_{01}\eta_{12}^2 \\ 0 & \mu_{12}\eta_{11}^2 & \mu_{01}\mu_{11}\eta_{12}^2 & v_2 & 0 & (a_e+b_e)\mu_{12}\mu_{11}\eta_{12}^2 \\ 0 & 0 & 0 & 0 & v_3 & v_5 \\ 0 & 0 & (a_e+b_e)\mu_{12}\mu_{01}\eta_{12}^2 & (a_e+b_e)\mu_{12}\mu_{11}\eta_{12}^2 & v_5 & v_4 \end{bmatrix},$$

where

$$\begin{aligned} v_1 &= \eta_{02}^2 + \eta_{12}^2\eta_{01}^2 + \mu_{01}^2\eta_{12}^2 + \mu_{12}^2\eta_{01}^2, \\ v_2 &= \eta_{12}^2\eta_{11}^2 + \mu_{11}^2\eta_{12}^2 + \mu_{12}^2\eta_{11}^2, \\ v_3 &= \frac{(b_u - a_u)^2}{12} + \frac{(b_e - a_e)^2}{12}, \\ v_4 &= \frac{(b_\epsilon - a_\epsilon)^2}{12} + \left(\frac{1}{2}\eta_{12}^4 + \mu_{12}^2\eta_{12}^2\right)(a_e + b_e)^2 \\ &\quad + \left(3\eta_{12}^4 + 6\mu_{12}^2\eta_{12}^2 + \mu_{12}^4\right)\frac{(b_e - a_e)^2}{12}, \\ v_5 &= \left(\mu_{12}^2 + \eta_{12}^2\right)\frac{(b_e - a_e)^2}{12}. \end{aligned}$$

The values of the normal-uniform prior distribution hyperparameters are indicated in Table 3, where the sample size n is 50, 100, or 500. In the case of the normal-uniform priors, we reach conclusions that are similar to those noted in Case 1, i.e., LBE and BE are close, and the methods of estimation are superior to the TSLS alternative. When we compare the observed results in Table 2 with those reported in Table 3, we notice that the bias of both BE and LBE may depend on the choice of prior distributions. See also the plots in Figure 3, which display the observed results reported in Table 3.

We also calculate RMSE values for each estimator of each model parameter; the observed values may be found in the Appendix in Tables A5 and A6. As in Case 1, the observed RMSE values for both the BE and LBE are smaller than the corresponding values for their TSLS competing method of parameter estimation. The RMSEs of τ^2 are also much larger than those of ξ, which affects the observed values summarized in Table 3.

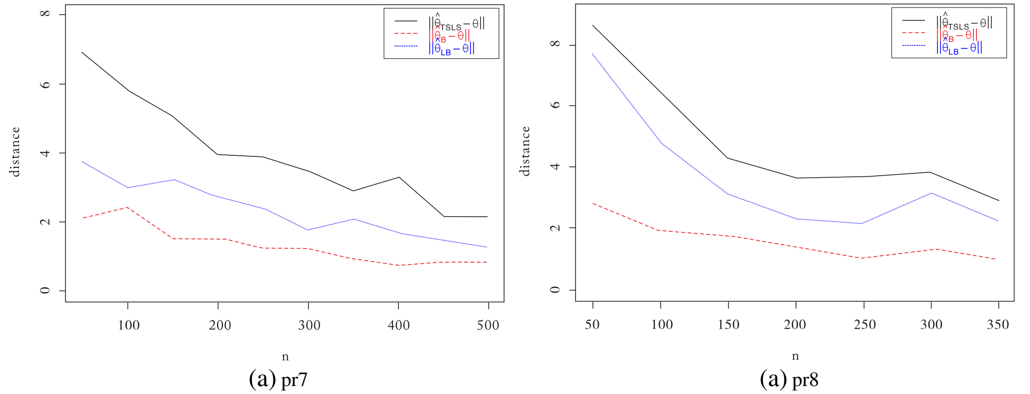
5.3. Numerical Comparisons Between LBE and BE

The LBE method is an approximation to BE but avoids the complex evaluation of multiple integrals and provides an explicit expression to evaluate. We use the distance between LBE and BE, say $\|\hat{\theta}_{LB} - \hat{\theta}_B\|$, to characterize the effectiveness of this approximation. Figure 4 displays plots of this measure for the two cases that we considered in the simulation studies reported in Section 5.2. The plots of the distance as a function of sample size and the two different priors are shown in Figure 4.

TABLE 3: Distance between the estimated value and the true value for different hyperparameters.

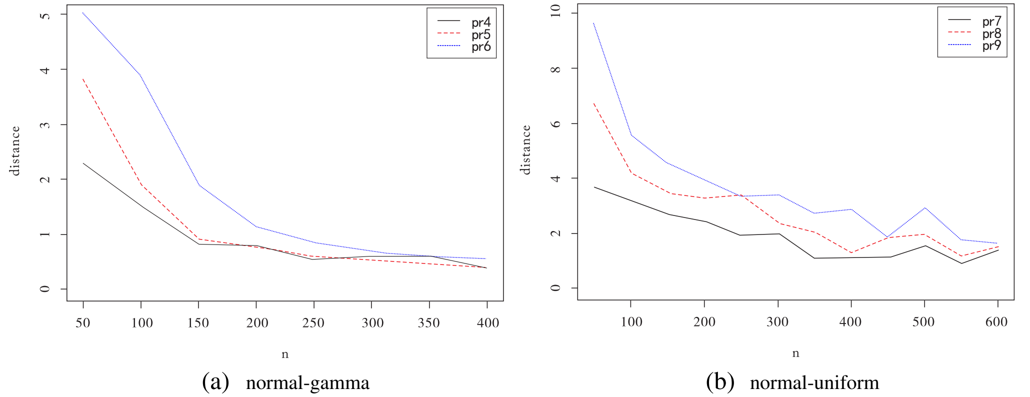
n		(a_ϵ, b_ϵ)	(a_u, b_u)	(a_e, b_e)	μ	η	$\lVert\hat{\theta}_B - \theta\rVert$	$\lVert\hat{\theta}_{LB} - \theta\rVert$	$\lVert\hat{\theta}_{TSLS} - \theta\rVert$
50	pr7	(15.5, 16.5)	(15.5, 16.5)	(24.5, 25.5)	$(5, -1, -4, 0.6)^\top$	$(1, 1, 1, 1)^\top$	2.4048	4.6617	8.2202
	pr8	(15, 17)	(15, 17)	(24, 26)	$(5, -1, -4, 0.6)^\top$	$(4, 4, 4, 4)^\top$	2.5191	5.6126	
	pr9	(14.5, 17.5)	(14.5, 17.5)	(23.5, 26.5)	$(5, -1, -4, 0.6)^\top$	$(9, 9, 9, 9)^\top$	2.6685	7.2204	
100	pr7	(15.5, 16.5)	(15.5, 16.5)	(24.5, 25.5)	$(5, -1, -4, 0.6)^\top$	$(1, 1, 1, 1)^\top$	1.8816	3.7247	6.1539
	pr8	(15, 17)	(15, 17)	(24, 26)	$(5, -1, -4, 0.6)^\top$	$(4, 4, 4, 4)^\top$	2.0183	5.0312	
	pr9	(14.5, 17.5)	(14.5, 17.5)	(23.5, 26.5)	$(5, -1, -4, 0.6)^\top$	$(9, 9, 9, 9)^\top$	2.0338	5.8641	
500	pr7	(15.5, 16.5)	(15.5, 16.5)	(24.5, 25.5)	$(5, -1, -4, 0.6)^\top$	$(1, 1, 1, 1)^\top$	0.9349	1.4809	2.6581
	pr8	(15, 17)	(15, 17)	(24, 26)	$(5, -1, -4, 0.6)^\top$	$(4, 4, 4, 4)^\top$	0.9427	1.8130	
	pr9	(14.5, 17.5)	(14.5, 17.5)	(23.5, 26.5)	$(5, -1, -4, 0.6)^\top$	$(9, 9, 9, 9)^\top$	1.0557	2.2786	

Note: $\mu = (\mu_{01}, \mu_{11}, \mu_{02}, \mu_{12})^\top$, $\eta = \left(\eta_{01}^2, \eta_{11}^2, \eta_{02}^2, \eta_{12}^2\right)^\top$, and pr stands for the prior.



(a) pr7 (a) pr8

FIGURE 3: Distance between the estimated value and the true value as a function of sample size.



(a) normal-gamma (b) normal-uniform

FIGURE 4: Distance between BE and LBE as a function of sample size.

TABLE 4: Effect of IV on the distance between the estimated value and the true value.

n	β_0	β_1	$\|\hat{\beta}^*_B - \beta^*\|$	$\|\hat{\beta}^*_{LB} - \beta^*\|$	$\|\hat{\beta}^*_{TSLS} - \beta^*\|$	$\|\hat{\beta}^+_B - \beta^+\|$	$\|\hat{\beta}^+_{LB} - \beta^+\|$	$\|\hat{\beta}^+_{TSLS} - \beta^+\|$
50	$\mathcal{N}(-4, 1)$	$\mathcal{N}(0.6, 1)$	0.4027	0.8254	0.8779	0.5756	0.8458	0.9127
	$\mathcal{N}(-4, 9)$	$\mathcal{N}(0.6, 9)$	0.7671	0.8553		0.8229	0.8731	
	$\mathcal{N}(-4, 25)$	$\mathcal{N}(0.6, 25)$	0.8277	0.8762		0.8551	0.8911	
100	$\mathcal{N}(-4, 1)$	$\mathcal{N}(0.6, 1)$	0.3459	0.5290	0.5511	0.5954	0.6025	0.7755
	$\mathcal{N}(-4, 9)$	$\mathcal{N}(0.6, 9)$	0.5156	0.5490		0.7229	0.7551	
	$\mathcal{N}(-4, 25)$	$\mathcal{N}(0.6, 25)$	0.5387	0.5503		0.7355	0.7665	
500	$\mathcal{N}(-4, 1)$	$\mathcal{N}(0.6, 1)$	0.2469	0.2736	0.2756	0.4062	0.4254	0.4683
	$\mathcal{N}(-4, 9)$	$\mathcal{N}(0.6, 9)$	0.2725	0.2762		0.4382	0.4632	
	$\mathcal{N}(-4, 25)$	$\mathcal{N}(0.6, 25)$	0.2754	0.2764		0.4411	0.4664	

When the prior distribution is the same, this measure of effectiveness decreases with increasing sample size, i.e., LBE gets closer to BE. Likewise, when the sample size is constant, the effectiveness of the LBE approximation to the BE improves as the prior distribution becomes more informative, i.e., more concentrated. In addition, because LBE does not depend on the specific prior form, when the sample size is large enough, the value of the prior hyperparameter has almost no influence on the effectiveness of the LBE approximation, which means LBE is more robust than BE when the complex multiple integrals are evaluated via the Metropolis–Hastings method.

5.4. Influences of Instrumental Variable on Measurement Error Model

If using an IV is not adopted, then we simply substitute W into the model identified in Equation (1) to obtain (for $p = 2$)

$$y = \beta_0 + \beta_1 W + \epsilon - \beta_1 u. \tag{12}$$

Obviously, the variable W is correlated with the model error. If we were to use least squares to estimate $\beta = (\beta_0, \beta_1)^\top$, the estimation would be biased and also inconsistent. Therefore, the influence of measurement error on the linear model exists and gives rise to increasing bias that is proportional to the variance of the underlying measurement error.

To assess the effect of using an IV in the presence of measurement error, we undertake various comparisons between estimators of our model parameters with and without an IV. We calculate the observed values of the bias in estimating β_0 and β_1 when an IV is, and also is not, used in the method of estimation. The resulting values are reported in Table 4; the symbols β^* and $\beta^\dagger$ indicate that an IV is, or is not, used, respectively.

These values clearly demonstrate that using an IV when measurement error is present affects the estimation results. Intuitively, by using an appropriate IV, the distance between the LBE and the true value is reduced because the IV counteracts the deviation.

In the various simulation studies that we reported above, we only considered the case when $p = 2$, i.e., the model identified in Equation (1) becomes a simple linear regression model. In order to further verify the superiority of our LBE method, we also investigated another linear model. Let $p = q = 3$, and

$$\alpha = (\alpha_0, \alpha_I^\top)^\top = \begin{pmatrix} 5 & -1 \\ 2 & -5 \\ 4 & 1 \end{pmatrix}, \quad \beta = \begin{pmatrix} -4 \\ 0.6 \\ 1 \end{pmatrix}.$$

Suppose the true values of σ_ϵ^2, σ_u^2, and σ_e^2 are given by the following distributions:

$$\epsilon \sim \mathcal{N}(0, 4I_n), \quad u \sim \mathcal{MN}(0, I_n, 4I_{p-1}), \quad e \sim \mathcal{MN}(0, I_n, 9I_{p-1}).$$

We also calculate the distances between the estimated value and the true value for different priors with known and unknown variances (Table A7). The various prior distributions that we adopted are summarized in the Appendix in Table A6; the corresponding numerical results may be found in Tables A8–A10.

The results show that when the variances are known, the distance between the estimated value and the true value is very small, and $\|\hat{\xi}_{LB} - \xi\|s$ are smaller than $\|\hat{\xi}_{TSLS} - \xi\|s$, which shows that LBE works well. In the case of unknown variances, we also notice that LBE is superior to the TSLS alternative. These results demonstrate that LBE also works in a more general linear model setting.

6. TWO CASE STUDIES

The LBE does not rely on the specific form of the prior but depends only on its moments, which reduces the difficulty of determining prior forms in Bayesian estimation. In addition, an explicit expression for the estimator can be derived, which adds to its convenience in practical situations. In this section, we employ two examples to demonstrate the use of our proposed LBE.

6.1. The Problem of Defects of Parts

This first case study concerns defective plastic automotive parts: should they be repaired or discarded? Also, why do they occur? Some engineers think they arise because of the standard deviation of the temperature in the production process, which needs to be minimized. Others argue that it is clearly a matter of production density and that the problem will disappear as the density increases. If so, then the problem can be solved by reducing the production speed, but it will increase the cost. While minimizing the production costs, workers may focus more attention on both the temperature and the density in order to reduce the occurrence of defects.

Siegel (1997) studied the relationship between temperature, density, and the average number of defects per 1000 parts produced for 30 independent production runs. However, there is serious multicollinearity among the three predictive variables. The correlation coefficient between temperature and density is −0.9591. Many investigators have focused their efforts on the relationship between the average number of defects (y) and the temperature (x). However, the temperature recorded typically involves measurement error. We propose to use a suitable measurement error model to analyze these same data while treating the density as an IV. In this article, we use the measurement error models to analyze the above data and denote the density as the IV.

Note that the relationship between the average number of defects and the temperature in the production process is not strictly linear, so we first employ a Box–Cox transformation of the original measurements. Our model for the resulting transformed data is

$$w = \alpha_0 + \alpha_1 z + \upsilon_1, \quad \upsilon_1 \sim \mathcal{N}\left(0, \tau_1^2\right), \tag{13}$$

$$\sqrt{y} = \beta_0 + \beta_1\alpha_0 + \beta_1\alpha_1 z + \upsilon_2, \quad \upsilon_2 \sim \mathcal{N}\left(0, \tau_2^2\right), \tag{14}$$

where y denotes the average number of defects produced per 1000 parts, w is the temperature measured during production, and z denotes the corresponding value of the density.

We calculate the BEs of the parameters, assuming first normal-uniform and then normal-gamma prior distributions, and also derive the corresponding LBEs. We then compare our observed results using these two estimators with the corresponding estimates arising from use of the TSLS estimators. Sample moments derived from the data are used to inform the prior distributions that we use. Thus

$$\hat{\alpha}_1 = \frac{S_{wz}}{S_z} = \frac{\sum_{i=1}^{n}(w_i - \overline{w})(z_i - \overline{z})}{\sum_{i=1}^{n}(z_i - \overline{z})^2}, \ \hat{\alpha}_0 = \overline{w} - \hat{\alpha}_1 \overline{z} = \frac{1}{n}\sum_{i=1}^{n} w_i - \frac{1}{n}\hat{\alpha}_1 \sum_{i=1}^{n} z_i,$$

$$\hat{\beta}_1 = \frac{S_{yz}}{S_{wz}} = \frac{\sum_{i=1}^{n}\left(\sqrt{y}_i - \overline{\sqrt{y}}\right)(z_i - \overline{z})}{\sum_{i=1}^{n}(w_i - \overline{w})(z_i - \overline{z})},$$

$$\hat{\beta}_0 = \overline{\sqrt{y}} - \hat{\beta}_1(\hat{\alpha}_0 + \hat{\alpha}_1 \overline{z}) = \frac{1}{n}\sum_{i=1}^{n}\sqrt{y}_i - \hat{\beta}_1\left(\hat{\alpha}_0 + \frac{\hat{\alpha}_1}{n}\sum_{i=1}^{n} z_i\right),$$

$$\hat{\sigma}_e^2 = \frac{S_w - \hat{\beta}_1 \hat{\alpha}_1^2 S_z}{\hat{\beta}_1}, \quad \hat{\sigma}_u^2 = S_w - \hat{\alpha}_1^2 S_z - \hat{\sigma}_e^2, \quad \hat{\sigma}_\epsilon^2 = S_y - \hat{\beta}_1 S_{wy},$$

where S_w, S_y, and S_{wy} are analogous to the definitions of S_z and S_{wz}. Table A11 in the Appendix summarizes the various choices of prior distributions that we use in these comparisons. Tables A12 and A13 report our observed results in terms of estimated model parameters and the corresponding RMSEs when the chosen priors are normal-uniform and normal-gamma, respectively. To summarize, we find that the model parameter estimates are quite similar. However, the BE method of estimation yields the smallest RMSEs, whereas TSLS estimation gives rise to the largest estimated standard deviations. To assess the model's goodness of fit, we calculate the value of R^2; in each case, the value exceeds 0.9, which suggests that our estimated model provides a satisfactory fit to the data. As for convenience, using the LBE method of estimation is clearly preferable.

6.2. The Problem of Alaskan Earthquakes

More than 30 years ago, Fuller (1987) used the IV approach to investigate the reported magnitude of Alaskan earthquakes that occurred between 1969 and 1978. Three measures of magnitude of an earthquake are the logarithm of the seismogram amplitude of 20-s surface waves, which we denote by y; the logarithm of the seismogram amplitude of longitudinal body waves, which we denote by w; and the logarithm of the maximum seismogram trace amplitude at short distance, which we denote by z, and treat them as the IVs. These observed values are designed to be measures of earthquake magnitude. Strength is a function of such factors as rupture length and stress drop at the fault, both of which increase with strength. A model could be formulated to specify average rupture length and stress drop for a given strength. In addition to variations in fault length and stress drop from averages derived from the strength model, there is a measurement error associated with the observations. The measurement error includes errors made in determining the amplitude of ground motion arising from factors such as the orientation of a limited number of observation stations to the fault plane of the earthquake (Table A14).

Using model equations that are analogous to those found in Equations (13) and (14) in the previous case study, we explored the relationship among x, y, and z. Parameter estimates and the corresponding RMSEs for these Alaskan earthquake data when the prior distributions chosen are normal-uniform and normal-gamma, respectively, may be found in the Appendix in Tables A15 and A16, respectively. Obviously, the RMSEs of BE and LBE are smaller than those generated using the TSLS method of estimation. The RMSEs under the normal-uniform priors are smaller than those under the normal-gamma priors. Overall, we find that our proposed LBE method is both feasible and convenient in analyzing Fuller's Alaskan earthquake data.

7. CONCLUSIONS

We have explored the problem of estimating the parameters in measurement error models and proposed an estimation procedure that combines the Bayes method with use of an IV. In this setting, although TSLS estimation is feasible, we demonstrated how prior information about the model parameters can be incorporated into the measurement error models so that BEs of the model parameters under the assumption of quadratic loss can be derived. However, the resulting calculations involve evaluating complex multiple integrals using the Metropolis–Hastings algorithm. As a result, explicit expressions for the BEs cannot be obtained. To avoid these complex calculations, we adopted the linear Bayes method to derive a linear approximation to the BEs. We proved that the LBE is superior to the TSLS estimator under the mean squared error matrix criterion. Via simulation studies, we showed that LBE is close to its full Bayes counterpart, and both of them are better than the TSLS estimator. Whether the variance parameters in the assumed model are known or unknown, our proposed linear Bayes approximation appears to yield parameter estimates that are very close to the true values. With increasing sample size, as well as more concentrated prior information, our LBE approaches the BE that it approximates. Two case studies provided concrete evidence that our proposed LBE is both feasible and practical in situations involving the presence of measurement error in linear models, and provided a robust alternative to full Bayes estimation of the model parameters, with its associated complexities.

Our proposed linear BE is based on the assumption that the second moment of the prior distribution exists and is solved under the same quadratic loss. In fact, many common distribution families, such as the normal, uniform, and gamma, satisfy this assumption. Note that parameter estimation via the Bayes method needs to specify some priors, which can be problematic in practice. As a linear approximation to BE, LBE does not depend on the specific prior form but relies solely on moments of the prior distribution to generate an explicit expression for the estimated parameter. Therefore, it should be more robust than a full BE, which involves evaluating complex multiple integrals via the Metropolis–Hastings algorithm. However, although linear Bayes estimation is occasionally equal to its full Bayes counterpart, situations can arise when it constitutes simply a local optimum rather than a global optimal solution.

ACKNOWLEDGEMENTS

We are grateful to the Editor and two anonymous reviewers for their valuable comments and suggestions. The research is partly supported by National Natural Science Foundation of China (NO. 11371051) and a grant from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Abarin, T., Li, H., Wang, L. Q., & Briollais, L. (2014). Instrumental variable approach to covariate measurement error in generalized linear models. *Statistics in Biosciences*, 6, 1–18.

Abarin, T. & Wang, L. Q. (2012). Instrumental variable approach to covariate measurement error in generalized linear models. *Annals of the Institute of Statistical Mathematics*, 64, 475–493.

Adock, R. J. (1877). Note on the method of least squares. *The Analyst*, 4, 183–184.

Basmann, R. L. (1957). A generalized classical method of linear estimation of coefficients in a structural system of stochastic equations. *Annals of Mathematical Statistics*, 20, 46–63.

Bradley, D. M. & Gupta, R. C. (2002). On the distribution of the sum of n non-identically distributed uniform random variables. *Annals of the Institute of Statistical Mathematics*, 54, 689–700.

Buzas, J. S. & Stefanski, L. A. (1996). Instrumental variable estimation in generalized measurement error models. *Journal of American Statistical Association*, 91, 999–1006.

Carroll, R. J. & Stefanski, L. A. (1994). Measurement error, instrumental variables and corrections for attenuation with applications to meta-analyses. *Statistics in Medicine*, 13, 1265–1282.

Cochran, W. G. (1968). Errors of measurement in statistics. *Journal of Nankai University (Natural Science Edition)*, 10, 637–666.

Conley, T. G., Hansen, C. B., McCulloch, R. E., & Rossi, P. E. (2008). A semi-parametric Bayesian approach to the instrumental variable problem. *Journal of Econometrics*, 144, 276–305.

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

Guan, J., Cheng, H. J., Bollen, K. A., Thomas, D. R., & Wang, L. Q. (2019). Instrumental variable estimation in ordinal probit models with mismeasured predictors. *Canadian Journal of Statistics*, 47, 653–667.

Guan, J. & Wang, L. Q. (2017). Instrumental variable estimation in linear quantile regression models with measurement error. *Chinese Journal of Applied Probability and Statistics*, 33, 475–486.

Hahn, P. R., He, J., & Lopes, H. (2018). Bayesian factor model shrinkage for linear regression with many instruments. *Journal of Business and Economic Statistics*, 36, 278–287.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50, 1029–1054.

Hartigan, J. A. (1969). Linear Bayesian methods. *Journal of the Royal Statistical Society, Series B: Methodological*, 31, 446–454.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, 97–109.

Heiligers, B. (1993). Linear Bayes and minimax estimation in linear models with partially restricted parameter space. *Journal of Statistical Planning and Inference*, 36, 175–183.

Jiang, J., Wang, L. C., & Wang, L. Q. (2021). Linear approximate Bayes estimator for regression parameter with an inequality constraint. *Communications in Statistics - Theory and Methods*, 51, 1531–1548.

Kleibergen, F. R. & Zivot, E. (2003). Bayesian and classical approaches to instrumental variable regression. *Journal of Econometrics*, 114, 29–72.

Kozumi, H. H. (2001). Bayesian analysis on Engel curves estimation with measurement errors and an instrumental variable. *Journal of Business and Economic Statistics*, 19, 292–298.

Lamotte, L. R. (1978). Bayes linear estimators. *Technometrics*, 3, 281–290.

Li, D., Qiu, M., & Ke, Z. (2020). Bayesian estimation analysis of Bernoulli measurement error model for longitudinal data. *International Journal of Applied Physics and Mathematics*, 10, 160–166.

Liu, X. (2019). Averaging Estimation for Instrumental Variables Quantile Regression. arXiv preprint, arXiv:1910.04245.

Lopes, H. F. & Polson, N. G. (2014). Bayesian instrumental variables: Priors and likelihoods. *Econometric Reviews*, 33, 100–121.

Martins, L. F. & Gabriel, V. J. (2014). Linear instrumental variables model averaging estimation. *Computational Statistics and Data Analysis*, 71, 709–724.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equations of state calculations by fast computing machine. *Journal of Chemical Physics*, 21, 1087–1091.

Moschopoulos, P. G. (1985). The distribution of the sum of independent gamma random variables. *Annals of the Institute of Statistical Mathematics*, 37, 541–544.

Pensky, M. & Ni, P. (2000). Extended linear empirical Bayes estimation. *Communications in Statistics*, 29, 579–592.

Rao, C. R. (1973). *Linear Bayesian Methods. Linear Statistical Inference and its Applications*, 2nd ed., John Wiley & Sons, New York.

Reiersol, O. (1945). Confluence analysis by means of instrumental sets of variables. *Arki for Mathematik Astronomi och Fysik*, 32, 1–119.

Samaniego, F. J. & Vestrup, E. (1999). On improving standard estimators via linear empirical Bayes methods. *Statistics and Probability Letters*, 44, 309–318.

Seng, L. & Li, J. (2021). Structural equation model averaging: Methodology and application. *Journal of Business and Economic Statistics*, 40, 815–828.

Siegel, A. (1997). *Practical Business Statistics*, 3rd ed., Irwin McGraw-Hill, Boston.

Vidal, I. & Bolfarini, H. (2011). Bayesian estimation of regression parameters in elliptical measurement error models. *Statistics and Probability Letters*, 81, 1398–1406.

Wang, L. C. & Singh, R. S. (2014). Linear Bayes estimator for the two-parameter exponential family under type II censoring. *Computational Statistics and Data Analysis*, 71, 633–642.

Wang, L. C. & Wei, L. S. (2010). Empirical Bayes estimation for the scale parameter with errors in variables. *Southeast Asian Bulletin of Mathematics*, 34, 1119–1131.

Wang, L. Q. (2021). Estimation in mixed-effects models with measurement error. In Yi, G. Y., Delaigle, A., & Gustafson, P. (Eds.), *Handbook of Measurement Error Models*, Chapman and Hall/CRC, Boca Raton, 359–378.

Wang, L. Q. & Hsiao, C. (2007). Two-stage estimation of limited dependent variable models with errors-in-variables. *Journal of Econometrics*, 10, 426–438.

Wang, L. Q. & Hsiao, C. (2011). Method of moments estimation and identifiability of semiparametric nonlinear errors-in-variables models. *Journal of Econometrics*, 165, 30–44.

Wright, P. G. (1928). *The Tariff on Animal and Vegetable Oils*, Macmillan, New York.

Xu, K., Ma, Y. Y., & Wang, L. Q. (2015). Instrument assisted regression for errors in variables models with binary response. *Scandinavian Journal of Statistics Theory and Application*, 42, 104–117.

Zellner, A. & Theil, H. (1962). Three-stage least squared: Simultaneous estimation of simultaneous equations. *Econometrica*, 30, 54–78.

APPENDIX

Proof of Theorem 1 . From the constraint $\mathrm{E}_{(T,\theta)}(\hat{\theta}_{LB} - \theta) = 0$, we know $\mathrm{E}_{(T,\theta)}(BT + b - \theta) = 0$. Hence

$$b = \mathrm{E}_{(T,\theta)}(\theta) - B\mathrm{E}_{(T,\theta)}(T) = \mathrm{E}\theta - B\mathrm{E}\theta.$$

According to $L(\hat{\theta}_{LB}, \theta) = (\hat{\theta}_{LB} - \theta)^\top D(\hat{\theta}_{LB} - \theta)$, the Bayes risk can be calculated as

$$\begin{aligned} R(\hat{\theta}_{LB}, \theta) &= \mathrm{E}_{(T,\theta)} L(\hat{\theta}_{LB}, \theta) \\ &= \mathrm{E}_{(T,\theta)}\left[\left(B(T - \mathrm{E}\theta) - (\theta - \mathrm{E}\theta)\right)^\top D\left(B(T - \mathrm{E}\theta) - (\theta - \mathrm{E}\theta)\right)\right] \\ &= \mathrm{E}_{(T,\theta)}\left[\mathrm{tr}\left(D\left(B(T - \mathrm{E}\theta) - (\theta - \mathrm{E}\theta)\right)\left(B(T - \mathrm{E}\theta) - (\theta - \mathrm{E}\theta)\right)^\top\right)\right] \\ &= \mathrm{tr}\left(DB\mathrm{E}_{(T,\theta)}\left((T - \mathrm{E}\theta)(T - \mathrm{E}\theta)^\top\right)B^\top\right) - \mathrm{tr}\left(D\mathrm{Cov}(\theta)B^\top\right) \\ &\quad - \mathrm{tr}\left(DB\mathrm{Cov}(\theta)) - \mathrm{tr}(D\mathrm{Cov}(\theta)\right). \end{aligned}$$

Note that we have

$$\mathrm{E}_{(T,\theta)}[(T - \mathrm{E}\theta)(T - \mathrm{E}\theta)^\top] = \mathrm{E}[\mathrm{Cov}(T|\theta)] + \mathrm{Cov}(\mathrm{E}(T|\theta)) = K + \mathrm{Cov}(\theta).$$

Hence

$$R(\hat{\theta}_{LB}, \theta) = \mathrm{tr}\left(DB(K + \mathrm{Cov}(\theta))B^\top\right) - \mathrm{tr}\left(D\mathrm{Cov}(\theta)B^\top\right) - \mathrm{tr}(DB\mathrm{Cov}(\theta)) - \mathrm{tr}(D\mathrm{Cov}(\theta)).$$

Let $\partial R(\hat{\theta}_{LB}, \theta)/\partial B = 0$; then

$$DB(K + \mathrm{Cov}(\theta)) - D\mathrm{Cov}(\theta) = 0,$$

and

$$B = \mathrm{Cov}(\theta)[K + \mathrm{Cov}(\theta)]^{-1} = I - KH,$$

where $H = [K + \mathrm{Cov}(\theta)]^{-1}$, which yields $b = \mathrm{E}\theta - B\mathrm{E}\theta = KH\mathrm{E}\theta$.

Hence, we can obtain the LBE of θ as

$$\hat{\theta}_{LB} = BT + b = T - KH(T - \mathrm{E}\theta).$$

The specific expression for K is obtained as follows:

$$K = \mathrm{E}[\mathrm{Cov}(T|\theta)] = \mathrm{E}\begin{pmatrix} \mathrm{Cov}(\mathrm{Vec}(\hat{\alpha})|\theta) & \mathrm{Cov}(\mathrm{Vec}(\hat{\alpha}), \hat{\gamma}|\theta) & 0 \\ \mathrm{Cov}(\mathrm{Vec}(\hat{\alpha}), \hat{\gamma}|\theta) & \mathrm{Cov}(\hat{\gamma}|\theta) & \\ 0 & 0 & \mathrm{Cov}(\hat{\tau}^2|\theta) \end{pmatrix}.$$

Note that

$$\begin{pmatrix} \mathrm{Vec}(W) \\ Y \end{pmatrix} \sim \mathcal{N}_{np}\left(\begin{pmatrix} (I_{p-1} \otimes Z)\mathrm{Vec}(\alpha) \\ Z\gamma \end{pmatrix}, \begin{pmatrix} \tau_1^2 I_{n(p-1)} & \left(\beta_I \otimes I_n\right)\sigma_e^2 \\ \left(\beta_I^\top \otimes I_n\right)\sigma_e^2 & \tau_2^2 I_n \end{pmatrix}\right),$$

where

$$\begin{aligned} \mathrm{Cov}(\mathrm{Vec}(\hat{\alpha}) \mid \theta) &= \mathrm{Cov}\left(\left(I_{p-1} \otimes (Z^\top Z)^{-1} Z^\top\right)\mathrm{Vec}(W) \mid \theta\right) \\ &= \left(I_{p-1} \otimes (Z^\top Z)^{-1}\right)\tau_1^2, \\ \mathrm{Cov}(\hat{\gamma} \mid \theta) &= \mathrm{Cov}\left((Z^\top Z)^{-1} Z^\top Y|\theta\right) \\ &= (Z^\top Z)^{-1}\tau_2^2, \\ \text{and }\ \mathrm{Cov}(\mathrm{Vec}(\hat{\alpha}), \hat{\gamma} \mid \theta) &= \mathrm{Cov}\left(\left(I_{p-1} \otimes (Z^\top Z)^{-1} Z^\top\right)\mathrm{Vec}(W), (Z^\top Z)^{-1} Z^\top Y|\theta\right) \\ &= \left(\beta_I \otimes (Z^\top Z)^{-1}\right)\sigma_e^2. \end{aligned}$$

Also, since

$$\mathrm{Cov}(\hat{\tau}^2 \mid \theta) = \begin{pmatrix} \mathrm{Cov}(\hat{\tau}_1^2 \mid \theta) & \mathrm{Cov}(\hat{\tau}_1^2, \hat{\tau}_2^2 \mid \theta) \\ \mathrm{Cov}(\hat{\tau}_1^2, \hat{\tau}_2^2 \mid \theta) & \mathrm{Cov}(\hat{\tau}_2^2 \mid \theta) \end{pmatrix},$$

$$\frac{(n-q)(p-1)\hat{\tau}_1^2}{\tau_1^2} \sim \chi^2_{(n-q)(p-1)} \quad \text{and} \quad \frac{(n-q)\hat{\tau}_2^2}{\tau_2^2} \sim \chi^2_{n-q},$$

we know $\mathrm{Var}\left((n-q)\hat{\tau}_2^2/\tau_2^2\right) = 2(n-q)$. Therefore, $\mathrm{Cov}\left(\hat{\tau}_2^2 \mid \theta\right) = 2\tau_2^4/(n-q)$ and $\mathrm{Cov}\left(\hat{\tau}_1^2 \mid \theta\right) = 2\tau_1^4/[(n-q)(p-1)]$.

Further, if

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim \mathcal{N}_{2n}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{21} & \Sigma_2 \end{pmatrix}\right),$$

then

$$\mathrm{Cov}\left(y_1^\top Q_1 y_1, y_2^\top Q_2 y_2\right) = 4\mu_1^\top Q_1 \Sigma_{12} Q_2 \mu_2 + 2\mathrm{tr}(Q_1 \Sigma_{12} Q_2 \Sigma_{21}),$$

where Q_1 and Q_2 are $n \times n$ symmetric matrices. Hence

$$\begin{aligned} \mathrm{Cov}\left(\hat{\tau}_1^2, \hat{\tau}_2^2|\theta\right) &= \mathrm{Cov}\left(\frac{\|\mathrm{Vec}(W) - (I_{p-1} \otimes Z)\mathrm{Vec}(\hat{\alpha})\|^2}{(n-q)(p-1)}, \frac{\|Y - Z\hat{\gamma}\|^2}{n-q}\Big|\theta\right) \\ &= \frac{\mathrm{Cov}\left(\mathrm{Vec}(W)^\top (I_{n(p-1)} - (I_{p-1} \otimes P_Z))\mathrm{Vec}(W), Y^\top (I_n - P_Z) Y|\theta\right)}{(n-q)^2(p-1)} \end{aligned}$$

$$= \frac{4\mathrm{Vec}(\alpha)^\top Z^\top \left(I_{n(p-1)} - (I_{p-1} \otimes P_Z)\right)(\beta_I \otimes I_n)\sigma_e^2 (I_n - P_Z) Z\gamma}{(n-q)^2(p-1)}$$
$$+ \frac{2\mathrm{tr}\left(\left(I_{n(p-1)} - (I_{p-1} \otimes P_Z)\right)(\beta_I \otimes I_n)\sigma_e^2 (I_n - P_Z)(\beta_I^\top \otimes I_n)\sigma_e^2\right)}{(n-q)^2(p-1)}$$
$$= \frac{2\beta_I^\top \beta_I \sigma_e^4}{(n-q)(p-1)},$$

where P_Z is a symmetric and idempotent matrix, and accordingly $\mathrm{tr}(I - P_Z) = \mathrm{rk}(I - P_Z) = n - q$.

To sum up, we obtain the matrix K.

This completes the proof of Theorem 1. ■

Proof of Theorem 2 . Given that $\mathrm{E}_{(T,\theta)}(\hat{\theta}_{LB} - \theta) = 0$, we have

$$\begin{aligned} \mathrm{MSEM}(\hat{\theta}_{LB}) &= \mathrm{E}_{(T,\theta)}\left[(\hat{\theta}_{LB} - \theta)(\hat{\theta}_{LB} - \theta)^\top\right] \\ &= \mathrm{E}\left((I - KH)\mathrm{Cov}(T|\theta)(I - KH)^\top\right) + \mathrm{Cov}(KH(\mathrm{E}\theta - \theta)) \\ &= K - 2KHK + KH[K + \mathrm{Cov}(\theta)]HK \\ &= K - KHK. \end{aligned}$$

Furthermore

$$\begin{aligned} \mathrm{MSEM}(\hat{\theta}_{TSLS}) &= \mathrm{E}\left[(\hat{\theta}_{TSLS} - \theta)(\hat{\theta}_{TSLS} - \theta)^\top\right] \\ &= \mathrm{E}\left[\left(\mathrm{E}(T - \theta)(T - \theta)^\top|\theta\right)\right] \\ &= K. \end{aligned}$$

Hence, $\mathrm{MSEM}(\hat{\theta}_{LB}) \le \mathrm{MSEM}(\hat{\theta}_{TSLS})$.

This completes the proof of Theorem 2, establishing that the LBE is superior to the TSLS. ■

SOME ADDITIONAL TABLES

TABLE A1: Bias and RMSE of estimator for the parameter with $n = 50$.

	pr	α_0	α_1	β_0	β_1
B	pr1	0.0291 (0.0395)	−0.0168 (0.0147)	−0.0402 (0.0389)	0.0067 (0.0072)
	pr2	0.0462 (0.0599)	−0.0174 (0.0152)	−0.0805 (0.0628)	0.0120 (0.0088)
	pr3	0.0496 (0.0662)	−0.0172 (0.0153)	−0.0101 (0.0715)	0.0152 (0.0098)
LB	pr1	0.0296 (0.0394)	−0.0171 (0.0147)	−0.1145 (0.0774)	0.0171 (0.0100)
	pr2	0.0448 (0.0599)	−0.0173 (0.0152)	−0.1233 (0.0804)	0.0183 (0.0103)
	pr3	0.0494 (0.0663)	−0.0173 (0.0153)	−0.1246 (0.0809)	0.0185 (0.0103)
TSLS		0.0538 (0.0724)	−0.0173 (0.0153)	−0.1255 (0.0812)	0.0186 (0.0103)

Note: RMSE is inside parentheses and the bias between the estimated value and the true value is outside parentheses.

TABLE A2: Bias and RMSE of estimator for the parameter with $n = 100$.

	pr	α_0	α_1	β_0	β_1
B	pr1	0.0313 (0.0332)	0.0075 (0.0101)	0.0241 (0.0357)	−0.0098 (0.0065)
	pr2	0.0383 (0.0426)	0.0082 (0.0102)	0.0205 (0.0500)	−0.0092 (0.0074)
	pr3	0.0390 (0.0449)	0.0081 (0.0102)	0.0141 (0.0537)	−0.0084 (0.0077)
LB	pr1	0.0296 (0.0332)	0.0077 (0.0100)	0.0192 (0.0555)	−0.0087 (0.0079)
	pr2	0.0379 (0.0426)	0.0081 (0.0102)	0.0151 (0.0572)	−0.0085 (0.0080)
	pr3	0.0401 (0.0449)	0.0082 (0.0102)	0.0163 (0.0574)	−0.0087 (0.0080)
TSLS		0.0419 (0.0470)	0.0082 (0.0103)	0.0133 (0.0576)	−0.0084 (0.0080)

Note: RMSE is inside parentheses and the bias between the estimated value and the true value is outside parentheses.

TABLE A3: Bias and RMSE of the estimator under the normal-gamma priors with $n = 50$.

	pr	α_0	α_1	β_0	β_1	τ_1^2	τ_2^2
B	pr4	0.0381	0.0307	−0.3264	0.0351	−0.2806	−0.3241
		(0.0667)	(0.0304)	(0.0519)	(0.0097)	(0.5299)	(0.3785)
	pr5	0.0401	0.0325	−0.4100	0.0646	−1.0127	−0.3297
		(0.0817)	(0.0309)	(0.0580)	(0.0118)	(0.5466)	(0.4243)
	pr6	0.0502	0.0313	−0.4402	0.0872	−1.4267	0.4135
		(0.0831)	(0.0310)	(0.0623)	(0.0144)	(0.7839)	(0.4372)
LB	pr4	0.0429	0.0302	−0.7196	0.0860	0.7359	0.2580
		(0.0791)	(0.0308)	(0.1112)	(0.0187)	(0.7109)	(0.4260)
	pr5	0.0439	0.0306	−0.7336	0.0878	0.9932	0.4779
		(0.0821)	(0.0309)	(0.1129)	(0.0189)	(0.7861)	(0.4393)
	pr6	0.0441	0.0307	−0.7365	0.0883	−1.3941	0.5653
		(0.0895)	(0.0309)	(0.1133)	(0.0190)	(0.9283)	(0.4431)
TSLS		0.0614	0.0320	−0.7747	0.0954	1.3456	−0.5434
		(0.1167)	(0.0315)	(0.1234)	(0.0201)	(1.0111)	(0.4379)

Note: RMSE is inside parentheses and the bias between the estimated value and the true value is outside parentheses.

TABLE A4: Bias and RMSE of the estimator under the normal-gamma priors with $n = 100$.

	pr	α_0	α_1	β_0	β_1	τ_1^2	τ_2^2
B	pr4	0.0091	0.0678	−0.0662	0.0044	−0.1554	−0.2005
		(0.0318)	(0.0127)	(0.0460)	(0.0053)	(0.2551)	(0.2138)
	pr5	0.0094	−0.0678	−0.1221	−0.0081	−0.4377	−0.2283
		(0.0357)	(0.0129)	(0.0496)	(0.0054)	(0.3280)	(0.2363)
	pr6	0.0102	−0.0680	−0.1586	−0.0218	0.5541	−0.3392
		(0.0381)	(0.0129)	(0.0498)	(0.0056)	(0.4096)	(0.2553)
LB	pr4	0.0971	−0.0676	−0.1140	−0.0170	0.2987	0.2238
		(0.0349)	(0.0128)	(0.0591)	(0.0060)	(0.2385)	(0.2553)
	pr5	0.0982	−0.0677	−0.1149	−0.0172	−0.3353	−0.2779
		(0.0356)	(0.0129)	(0.0593)	(0.0060)	(0.3593)	(0.2598)
	pr6	0.0984	−0.0677	−0.1154	−0.0172	−0.5989	0.4661
		(0.0389)	(0.0129)	(0.0593)	(0.0060)	(0.4238)	(0.2673)
TSLS		−0.0116	−0.0684	−0.1017	0.0175	0.4226	0.3121
		(0.0430)	(0.0130)	(0.0597)	(0.0060)	(0.4277)	(0.2686)

Note: RMSE is inside parentheses and the bias between the estimated value and the true value is outside parentheses.

TABLE A5: RMSE of estimator under the normal-uniform priors with $n = 50$.

	pr	α_0	α_1	β_0	β_1	τ_1^2	τ_2^2
B	pr7	0.0340	0.0004	0.0326	−0.0216	−0.0054	−0.0449
		(0.0583)	(0.0170)	(0.0545)	(0.0088)	(0.0026)	(0.2562)
	pr8	0.0458	0.0007	0.0314	−0.0219	−0.0173	0.0880
		(0.0758)	(0.0174)	(0.0700)	(0.0096)	(0.0090)	(0.2794)
	pr9	0.0493	0.0011	0.0267	−0.0218	−0.0359	0.1175
		(0.0841)	(0.0176)	(0.0779)	(0.0101)	(0.0205)	(0.2935)
LB	pr7	0.0428	0.0005	−0.0752	−0.0096	−0.0050	0.4296
		(0.0584)	(0.0170)	(0.0972)	(0.0145)	(0.0021)	(0.2943)
	pr8	0.0508	0.0008	−0.0715	−0.0081	−0.0181	0.6360
		(0.0756)	(0.0174)	(0.1030)	(0.0149)	(0.0149)	(0.3696)
	pr9	0.0543	0.0012	−0.0681	−0.0076	−0.0139	0.8774
		(0.0839)	(0.0176)	(0.1048)	(0.0150)	(0.0188)	(0.4751)
TSLS		0.0642	0.0021	−0.0190	−0.0062	−0.0063	0.6353
		(0.1075)	(0.0180)	(0.1101)	(0.0152)	(0.8966)	(0.5398)

Note: RMSE is inside parentheses and the bias between estimated value and true value is outside parentheses.

TABLE A6: RMSE of estimator under the normal-uniform priors with $n = 100$.

	pr	α_0	α_1	β_0	β_1	τ_1^2	τ_2^2
B	pr7	0.0116	−0.0098	−0.0122	0.0056	−0.0003	0.4376
		(0.0332)	(0.0102)	(0.0312)	(0.0059)	(0.0022)	(0.1808)
	pr8	0.0138	−0.0102	−0.161	0.0051	−0.0001	0.4463
		(0.0390)	(0.0103)	(0.0371)	(0.0061)	(0.0084)	(0.1876)
	pr9	0.0152	−0.0097	−0.0204	0.0054	0.0025	0.4685
		(0.0413)	(0.0103)	(0.0395)	(0.0062)	(0.0188)	(0.1926)
LB	pr7	0.0184	−0.0099	−0.0398	−0.0010	0.0040	0.7622
		(0.0333)	(0.0102)	(0.0494)	(0.0075)	(0.0022)	(0.2442)
	pr8	0.0182	−0.0101	−0.0342	−0.0007	−0.0160	0.7741
		(0.0390)	(0.0103)	(0.0509)	(0.0075)	(0.0085)	(0.2696)
	pr9	0.0180	−0.0101	−0.0313	−0.0006	0.0352	0.8020
		(0.0414)	(0.0103)	(0.0514)	(0.0075)	(0.0190)	(0.2781)
TSLS		0.0176	−0.0103	0.0038	−0.0003	0.9335	0.8295
		(0.0471)	(0.0104)	(0.0524)	(0.0076)	(0.4477)	(0.2973)

Note: RMSE is inside parentheses and the bias between the estimated value and the true value is outside parentheses.

TABLE A7: The different priors.

pr	α	β	σ_ε^2	σ_u^2	σ_e^2
pr1	$\mathcal{MN}\left(\begin{bmatrix}5&-1\\2&-5\\4&1\end{bmatrix},\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix},\begin{bmatrix}1&0.5\\0.5&1\end{bmatrix}\right)$	$\mathcal{N}\left(\begin{bmatrix}-4\\0.6\\1\end{bmatrix},\begin{bmatrix}1&0.5&0.5\\0.5&1&0.5\\0.5&0.5&1\end{bmatrix}\right)$			
pr2	$\mathcal{MN}\left(\begin{bmatrix}5&-1\\2&-5\\4&1\end{bmatrix},\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix},\begin{bmatrix}4&1\\1&4\end{bmatrix}\right)$	$\mathcal{N}\left(\begin{bmatrix}-4\\0.6\\1\end{bmatrix},\begin{bmatrix}4&1&1\\1&4&1\\1&1&4\end{bmatrix}\right)$			
pr3	$\mathcal{MN}\left(\begin{bmatrix}5&-1\\2&-5\\4&1\end{bmatrix},\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix},\begin{bmatrix}9&1.5\\1.5&9\end{bmatrix}\right)$	$\mathcal{N}\left(\begin{bmatrix}-4\\0.6\\1\end{bmatrix},\begin{bmatrix}9&1.5&1.5\\1.5&9&1.5\\1.5&1.5&9\end{bmatrix}\right)$			
pr4	$\mathcal{MN}\left(\begin{bmatrix}5&-1\\2&-5\\4&1\end{bmatrix},\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix},\begin{bmatrix}1&0.5\\0.5&1\end{bmatrix}\right)$	$\mathcal{N}\left(\begin{bmatrix}-4\\0.6\\1\end{bmatrix},\begin{bmatrix}1&0.5&0.5\\0.5&1&0.5\\0.5&0.5&1\end{bmatrix}\right)$	$\mathcal{G}(2,4)$	$\mathcal{G}(4,2)$	$\mathcal{G}(5,3)$
pr5	$\mathcal{MN}\left(\begin{bmatrix}5&-1\\2&-5\\4&1\end{bmatrix},\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix},\begin{bmatrix}4&1\\1&4\end{bmatrix}\right)$	$\mathcal{N}\left(\begin{bmatrix}-4\\0.6\\1\end{bmatrix},\begin{bmatrix}4&1&1\\1&4&1\\1&1&4\end{bmatrix}\right)$	$\mathcal{G}(2,6)$	$\mathcal{G}(4,4)$	$\mathcal{G}(5,5)$

TABLE A7: *Continued*

pr	α	β	σ_ε^2	σ_u^2	σ_e^2
pr6	$\mathcal{MN}\left(\begin{bmatrix}5 & -1\\ 2 & -5\\ 4 & 1\end{bmatrix}, \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}, \begin{bmatrix}9 & 1.5\\ 1.5 & 9\end{bmatrix}\right)$	$\mathcal{N}\left(\begin{bmatrix}-4\\ 0.6\\ 1\end{bmatrix}, \begin{bmatrix}9 & 1.5 & 1.5\\ 1.5 & 9 & 1.5\\ 1.5 & 1.5 & 9\end{bmatrix}\right)$	$\mathcal{G}(2, 8)$	$\mathcal{G}(4, 6)$	$\mathcal{G}(5, 7)$
pr7	$\mathcal{MN}\left(\begin{bmatrix}5 & -1\\ 2 & -5\\ 4 & 1\end{bmatrix}, \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}, \begin{bmatrix}1 & 0.5\\ 0.5 & 1\end{bmatrix}\right)$	$\mathcal{N}\left(\begin{bmatrix}-4\\ 0.6\\ 1\end{bmatrix}, \begin{bmatrix}1 & 0.5 & 0.5\\ 0.5 & 1 & 0.5\\ 0.5 & 0.5 & 1\end{bmatrix}\right)$	$\mathcal{U}$ $(3.5, 4.5)$	$\mathcal{U}$ $(3.5, 4.5)$	$\mathcal{U}$ $(8.5, 9.5)$
pr8	$\mathcal{MN}\left(\begin{bmatrix}5 & -1\\ 2 & -5\\ 4 & 1\end{bmatrix}, \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}, \begin{bmatrix}4 & 1\\ 1 & 4\end{bmatrix}\right)$	$\mathcal{N}\left(\begin{bmatrix}-4\\ 0.6\\ 1\end{bmatrix}, \begin{bmatrix}4 & 1 & 1\\ 1 & 4 & 1\\ 1 & 1 & 4\end{bmatrix}\right)$	$\mathcal{U}$ $(3, 5)$	$\mathcal{U}$ $(3, 5)$	$\mathcal{U}$ $(8, 10)$
pr9	$\mathcal{MN}\left(\begin{bmatrix}5 & -1\\ 2 & -5\\ 4 & 1\end{bmatrix}, \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}, \begin{bmatrix}9 & 1.5\\ 1.5 & 9\end{bmatrix}\right)$	$\mathcal{N}\left(\begin{bmatrix}-4\\ 0.6\\ 1\end{bmatrix}, \begin{bmatrix}9 & 1.5 & 1.5\\ 1.5 & 9 & 1.5\\ 1.5 & 1.5 & 9\end{bmatrix}\right)$	$\mathcal{U}$ $(2.5, 5.5)$	$\mathcal{U}$ $(2.5, 5.5)$	$\mathcal{U}$ $(7.5, 10.5)$

TABLE A8: Distance between the estimated value of the regression parameters and the true value under different prior hyperparameters.

n	pr	$\|\|\hat{\xi}_B - \xi\|\|$	$\|\|\hat{\xi}_{LB} - \xi\|\|$	$\|\|\hat{\xi}_{TSLS} - \xi\|\|$
50	pr1	0.6026	0.7008	0.8132
	pr2	0.6918	0.7461	
	pr3	0.7342	0.7649	
100	pr1	0.7185	0.5300	0.5771
	pr2	0.5509	0.5744	
	pr3	0.5589	0.5935	
500	pr1	0.2634	0.2263	0.2306
	pr2	0.2715	0.2284	
	pr3	0.2741	0.2291	

TABLE A9: Distance between the estimated value and the true value under different prior hyperparameters of the normal-gamma priors.

n	pr	$\|\|\hat{\theta}_B - \theta\|\|$	$\|\|\hat{\theta}_{LB} - \theta\|\|$	$\|\|\hat{\theta}_{TSLS} - \theta\|\|$
50	pr4	5.3541	8.0347	11.9156
	pr5	5.5452	8.8796	
	pr6	6.1941	9.7650	
100	pr4	5.1271	7.4823	8.2348
	pr5	5.2627	7.8595	
	pr6	5.8559	8.4823	
500	pr4	4.5600	5.2223	5.5750
	pr5	4.6954	5.5058	
	pr6	5.5300	5.5966	

TABLE A10: Distance between the estimated value and the true value under different prior hyperparameters of the normal-uniform priors.

n	pr	$\|\|\hat{\theta}_B - \theta\|\|$	$\|\|\hat{\theta}_{LB} - \theta\|\|$	$\|\|\hat{\theta}_{TSLS} - \theta\|\|$
50	pr7	2.4542	5.3854	11.7515
	pr8	3.0428	6.9306	
	pr9	4.0366	6.4823	
100	pr7	2.5470	4.0358	8.3987
	pr8	3.3846	4.4440	
	pr9	4.4821	5.2448	
500	pr7	2.6693	3.2551	5.4527
	pr8	4.3457	4.5635	
	pr9	4.8551	5.5189	

TABLE A11: The different priors for defects of parts.

pr	α_0	α_1	β_0	β_1	σ_ε^2	σ_u^2	σ_e^2
pr1	$\mathcal{N}(6.42, 1)$	$\mathcal{N}((-0.17, 1)$	$\mathcal{N}(-3.80, 1)$	$\mathcal{N}(3.87, 1)$	$\mathcal{U}(0.2, 0.4)$	$\mathcal{U}$ $(0.01, 0.02)$	$\mathcal{U}$ $(0.01, 0.02)$
pr2	$\mathcal{N}(6.42, 4)$	$\mathcal{N}(-0.17, 4)$	$\mathcal{N}(-3.80, 4)$	$\mathcal{N}(3.87, 4)$	$\mathcal{U}(0.1, 0.5)$	$\mathcal{U}$ $(0.005, 0.025)$	$\mathcal{U}$ $(0.005, 0.025)$
pr3	$\mathcal{N}(6.42, 9)$	$\mathcal{N}(-0.17, 9)$	$\mathcal{N}(-3.80, 9)$	$\mathcal{N}(3.87, 9)$	$\mathcal{U}(0, 0.6)$	$\mathcal{U}(0, 0.03)$	$\mathcal{U}(0, 0.03)$
pr4	$\mathcal{N}(6.42, 1)$	$\mathcal{N}(-0.17, 1)$	$\mathcal{N}(-3.80, 1)$	$\mathcal{N}(3.87, 1)$	$\mathcal{G}(1, 3)$	$\mathcal{G}(0.1, 2)$	$\mathcal{G}(0.1, 1)$
pr5	$\mathcal{N}(6.42, 4)$	$\mathcal{N}(-0.17, 4)$	$\mathcal{N}(-3.80, 4)$	$\mathcal{N}(3.87, 4)$	$\mathcal{G}(1, 4)$	$\mathcal{G}(0.1, 3)$	$\mathcal{G}(0.1, 2)$
pr6	$\mathcal{N}(6.42, 9)$	$\mathcal{N}(-0.17, 9)$	$\mathcal{N}(-3.80, 9)$	$\mathcal{N}(3.87, 9)$	$\mathcal{G}(1, 5)$	$\mathcal{G}(0.1, 4)$	$\mathcal{G}(0.1, 3)$

TABLE A12: Estimator and RMSE under the normal-uniform priors for defects of parts.

	pr	α_0	α_1	β_0	β_1	τ_1^2	τ_2^2
B	pr1	6.4168	−0.1667	−3.7593	3.8444	0.0298	0.4960
		(0.2318)	(0.0091)	(0.4528)	(0.2014)	(0.0037)	(0.0681)
	pr2	6.4169	−0.1666	−3.7697	3.8517	0.0303	0.4732
		(0.2429)	(0.0096)	(0.4678)	(0.2068)	(0.0059)	(0.1039)
	pr3	6.4219	−0.1668	−3.7796	3.8555	0.0307	0.4701
		(0.2460)	(0.0096)	(0.4919)	(0.2183)	(0.0069)	(0.1183)
LB	pr1	6.4234	−0.1669	−3.7609	3.8473	0.0294	0.5062
		(0.2364)	(0.0093)	(0.5173)	(0.1077)	(0.0036)	(0.0742)
	pr2	6.4219	−0.1668	−3.7724	3.8527	0.0288	0.4741
		(0.2402)	(0.0094)	(0.5359)	(0.1115)	(0.0085)	(0.1186)
	pr3	6.4181	−0.1667	−3.7863	3.8592	0.0285	0.4541
		(0.2415)	(0.0095)	(0.5649)	(0.1175)	(0.0070)	(0.1488)
TSLS		6.4152	−0.1666	−3.8007	3.8659	0.0282	0.4067
		(0.2442)	(0.0096)	(0.5150)	(0.2116)	(0.0081)	(0.1489)

Note: RMSE is inside parentheses and the estimated value is outside parentheses.

TABLE A13: Estimator and RMSE under the normal-gamma priors for defects of parts.

	pr	α_0	α_1	β_0	β_1	τ_1^2	τ_2^2
B	pr4	6.4146	−0.1665	−3.8019	3.8484	0.0308	0.4257
		(0.2386)	(0.0093)	(0.4657)	(0.2068)	(0.0091)	(0.1137)
	pr5	6.4146	−0.1665	−3.7908	3.8620	0.0309	0.4289
		(0.2347)	(0.0092)	(0.4741)	(0.2112)	(0.0090)	(0.1220)
	pr6	6.4133	−0.1665	−3.7884	3.8615	0.0298	0.4211
		(0.2444)	(0.0096)	(0.4735)	(0.2096)	(0.0079)	(0.1108)
LB	pr4	6.4226	−0.1668	−3.7600	3.8469	0.0317	0.4887
		(0.5965)	(0.0235)	(0.3688)	(0.2615)	(0.0097)	(0.1224)
	pr5	6.4234	−0.1669	−3.7588	3.8463	0.0311	0.4894
		(0.8001)	(0.0315)	(0.4200)	(0.2704)	(0.0094)	(0.1669)
	pr6	6.4231	−0.1669	−3.7607	3.8472	0.0315	0.4953
		(0.9597)	(0.0378)	(0.4822)	(0.2608)	(0.0096)	(0.1702)
TSLS		6.4152	−0.1666	−3.8007	3.8659	0.0282	0.4067
		(0.2442)	(0.0096)	(0.5150)	(0.2116)	(0.0081)	(0.1489)

Note: RMSE is inside parentheses and the estimated value is outside parentheses.

TABLE A14: The different priors for Alaskan earthquakes.

pr	α_0	α_1	β_0	β_1	σ_ε^2	σ_u^2	σ_e^2
pr1	$\mathcal{N}(2.31, 1)$	$\mathcal{N}(0.55, 1)$	$\mathcal{N}(-4.21, 1)$	$\mathcal{N}(1.78, 1)$	$\mathcal{U}$ (0.07, 0.09)	$\mathcal{U}$ (0.03, 0.05)	$\mathcal{U}$ (0.03, 0.05)
pr2	$\mathcal{N}(2.31, 4)$	$\mathcal{N}(0.55, 4)$	$\mathcal{N}(-4.21, 4)$	$\mathcal{N}(1.78, 4)$	$\mathcal{U}$ (0.06, 0.10)	$\mathcal{U}$ (0.02, 0.06)	$\mathcal{U}$ (0.02, 0.06)
pr3	$\mathcal{N}(2.31, 9)$	$\mathcal{N}(0.55, 9)$	$\mathcal{N}(-4.21, 9)$	$\mathcal{N}(1.78, 9)$	$\mathcal{U}$ (0.05, 0.11)	$\mathcal{U}$ (0.01, 0.07)	$\mathcal{U}$ (0.01, 0.07)
pr4	$\mathcal{N}(2.31, 1)$	$\mathcal{N}(0.55, 1)$	$\mathcal{N}(-4.21, 1)$	$\mathcal{N}(1.78, 1)$	$\mathcal{G}(2, 1)$	$\mathcal{G}(1, 2)$	$\mathcal{G}(1, 1)$
pr5	$\mathcal{N}(2.31, 4)$	$\mathcal{N}(0.55, 4)$	$\mathcal{N}(-4.21, 4)$	$\mathcal{N}(1.78, 4)$	$\mathcal{G}(2, 3)$	$\mathcal{G}(1, 3)$	$\mathcal{G}(1, 4)$
pr6	$\mathcal{N}(2.31, 9)$	$\mathcal{N}(0.55, 9)$	$\mathcal{N}(-4.21, 9)$	$\mathcal{N}(1.78, 9)$	$\mathcal{G}(2, 5)$	$\mathcal{G}(1, 4)$	$\mathcal{G}(1, 6)$

TABLE A15: Estimator and RMSE under the normal-uniform priors for Alaskan earthquakes.

	pr	α_0	α_1	β_0	β_1	τ_1^2	τ_2^2
B	pr1	2.3118	0.5529	−4.3129	1.8022	0.0818	0.2128
		(0.2864)	(0.0543)	(0.5907)	(0.1133)	(0.0071)	(0.0226)
	pr2	2.3093	0.5535	−4.2515	1.7904	0.0858	0.2190
		(0.3023)	(0.0571)	(0.6840)	(0.1308)	(0.0114)	(0.0311)
	pr3	2.3081	0.5537	−4.1953	1.7794	0.0879	0.2258
		(0.3074)	(0.0582)	(0.7441)	(0.1426)	(0.0136)	(0.0362)
LB	pr1	2.3603	0.5442	−4.3847	1.8147	0.0869	0.2278
		(0.6578)	(0.1312)	(0.5451)	(0.0968)	(0.2689)	(0.0479)
	pr2	2.3922	0.5381	−4.4933	1.8356	0.0897	0.2303
		(0.9554)	(0.1851)	(0.5583)	(0.1056)	(0.2691)	(0.0654)
	pr3	2.3861	0.5392	−4.4699	1.8313	0.0941	0.2316
		(1.1582)	(0.2223)	(0.5983)	(0.1132)	(0.2693)	(0.0832)
TSLS		2.3114	0.5531	−4.2132	1.7826	0.0871	0.2219
		(2.1171)	(0.4004)	(1.0352)	(1.8547)	(0.7395)	(0.0693)

Note: RMSE is inside parentheses and the estimated value is outside parentheses.

TABLE A16: Estimator and RMSE under the normal-gamma priors for Alaskan earthquakes.

	pr	α_0	α_1	β_0	β_1	τ_1^2	τ_2^2
B	pr4	2.3112	0.5533	−4.0741	1.7551	0.0961	0.2524
		(0.3078)	(0.0581)	(0.8165)	(0.1561)	(0.0184)	(0.0495)
	pr5	2.3275	0.5503	−4.1406	1.7684	0.0959	0.2498
		(0.3233)	(0.0613)	(0.8054)	(0.1542)	(0.0189)	(0.0474)
	pr6	2.3196	0.5515	−4.0873	1.7583	0.0950	0.2477
		(0.3227)	(0.0614)	(0.7789)	(0.1492)	(0.0182)	(0.0485)
LB	pr4	2.3384	0.0.5483	−4.5787	1.8103	0.0971	0.2423
		(0.8197)	(0.1589)	(0.6330)	(0.2295)	(0.1852)	(0.0524)
	pr5	2.3166	0.0.5522	−4.3903	1.8330	0.0974	0.2578
		(1.2264)	(0.2383)	(0.7325)	(0.2325)	(0.1704)	(0.0608)
	pr6	2.3159	0.5523	−4.4100	1.8394	0.0982	0.2567
		(1.2749)	(0.2514)	(0.7815)	(0.2815)	(0.2557)	(0.0698)
TSLS		2.3114	0.5531	−4.2132	1.7826	0.0871	0.2219
		(2.1171)	(0.4004)	(1.0352)	(1.8547)	(0.7395)	(0.0693)

Note: RMSE is inside parentheses and the estimated value is outside parentheses.

Received 16 January 2022
Accepted 26 October 2022