Journal of the Korean Statistical Society (2002), **31**: *1, pp 77–91*

Bayesian Analysis for a Functional Regression Model with Truncated Errors in Variables

Hea-Jung Kim[1]

ABSTRACT

This paper considers a functional regression model with truncated errors in explanatory variables. We show that the ordinary least squares (OLS) estimators produce bias in regression parameter estimates under misspecified models with ignored errors in the explanatory variable measurements, and then propose methods for analyzing the functional model. Fully parametric frequentist approaches for analyzing the model are intractable and thus Bayesian methods are pursued using a Markov chain Monte Carlo (MCMC) sampling based approach. Necessary theories involved in modeling and computation are provided. Finally, a simulation study is given to illustrate and examine the proposed methods.

Keywords: Functional regression model, biased estimator, Markov chain Monte Carlo, Bayesian estimation, conditional Bayes factor.

1. Introduction

If errors in explanatory variables are known to be present in regression analysis, the effect of the errors on OLS estimators is well known (see Carroll *et al.*, 1995). There are two classes of models that have been considered in describing the effect. A structure model arises when the explanatory variable observations can be considered to be realizations from a random variable, whereas the model is functional if they must be considered to be fixed. Usually it is clear which model class is appropriate according to the context of the data. In cases where data are obtained from carefully controlled changes like experimental procedures in the physical sciences, a functional model will be appropriate for describing the effect. For the structural explanatory variable model, the effect of errors leads to attenuation bias in the OLS estimators and is well investigated by Cochran (1968), Fuller (1987), Schaalje and Butts (1993), Carroll *et al.* (1995) and so on.

[1]Department of Statistics, Dongguk University, Seoul 100-715, Korea

In this paper, we consider functional modeling. We restrict our attention to linear models with, for simplicity, a single explanatory variable, although similar ideas can be applied to multiple linear regression models. Denote the n-dimensional response vector by $\mathbf{Y}$ and the corresponding vector of true explanatory variables by $\mathbf{x}$. Assume

$$\mathbf{Y} \sim N(\beta_0\mathbf{1} + \beta_1\mathbf{x}, \sigma^2 I_n), \text{ and } \mathbf{X} = \mathbf{x} + \theta\varepsilon, \tag{1.1}$$

where $\mathbf{1}$ is an n-vector of ones. Furthermore, assume that the explanatory variables are subject to independent measurement error, with only $\mathbf{X}$ being observed, where each component of ε is independent and identically distributed as $\varepsilon_i \sim TN(0,1)$, $i = 1, \ldots, n$, independent of $\mathbf{Y}$. Here $TN(0,1)$ denotes a singly truncated standard normal distribution with interval space $[0, \infty)$ and θ is an unknown constant.

We will consider the effect on OLS estimator of $\beta = (\beta_0, \beta_1)'$ when using $\mathbf{W} = (\mathbf{1}, \mathbf{X})$ in place of the true but unknown design matrix $\mathbf{w} = (\mathbf{1}, \mathbf{x})$, and then suggest a Bayesian estimation procedure that will take account of the full error structure. Note that Richardson and Wu (1970) and Morton-Jones and Henderson (2000) considered the effect of errors under the functional model with $\theta = 1$ and $\varepsilon \sim N(0, \sigma_\varepsilon^2 I_n)$. Therefore, our model of concern can be viewed as a variant of their model in a sense that it is designed to take account of the truncated measurement error structure in the explanatory variable. An immediate example of the error structure is radiance measurements from satellite-borne infrared sensors. If one is interested in the relationship between sea surface temperatures and associated monthly rainfall during an experimental period, errors in the rainfalls (response variable) are known to be independent. It is also apparent that measurement errors occur in true radiance readings, because the measurements are severely distorted by the presence of clouds in the fields of view of the sensors (see, DePriest 1983 and Azzalini and Valle 1996 for other examples).

2. Bias in OLS Estimator

Consider the effect on OLS estimator of $\beta = (\beta_0, \beta_1)'$ when using $\mathbf{W} = (\mathbf{1}, \mathbf{X})$ in place of the true but unknown design matrix $\mathbf{w} = (\mathbf{1}, \mathbf{x})$. In this case, the estimator of (1.1) is

$$\hat{\beta}^{OLS} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{Y}. \tag{2.1}$$

Let $\mathbf{x}^* = \mathbf{x} + \theta\mu_\varepsilon\mathbf{1}$ and $\varepsilon^* = \theta(\varepsilon - \mu_\varepsilon)1$ so that $E[\varepsilon^*] = 0$, where $\mu_\varepsilon\mathbf{1} = E[\varepsilon]$. Then the OLS estimates of β_0 and β_1 are

$$\hat{\beta}_0^{OLS} = \frac{(\mathbf{x}^* + \varepsilon^*)'\{(\mathbf{x}^* + \varepsilon^*)\mathbf{1}' - \mathbf{1}(\mathbf{x}^* + \varepsilon^*)'\}\mathbf{Y}}{\Delta}, \tag{2.2}$$

$$\begin{aligned} \hat{\beta}_1^{OLS} &= \frac{\mathbf{1}'\mathbf{1}(\mathbf{x}^* + \varepsilon^*)'\mathbf{Y} - \mathbf{1}'(\mathbf{x}^* + \varepsilon^*)\mathbf{1}'\mathbf{Y}}{\Delta} \\ &= (\mathbf{x}^* + \varepsilon^*)'\mathbf{C}_n\mathbf{Y}/\Delta, \end{aligned} \tag{2.3}$$

where $\Delta = (\mathbf{x}^* + \varepsilon^*)'\mathbf{C}_n(\mathbf{x}^* + \varepsilon^*)$ and $\mathbf{C}_n = (n\mathbf{I}_n - \mathbf{1}\mathbf{1}')$, the centering matrix.

In considering the bias, we take expectations over $\mathbf{Y}$ first to arrive at

$$E_{\mathbf{Y}}[\hat{\beta}_0^{OLS}] = \beta_0 + \beta_1\frac{(\mathbf{x}^* + \varepsilon^*)'\mathbf{K}_n(\mathbf{x}^* + \varepsilon^*)}{\Delta}, \tag{2.4}$$

$$E_{\mathbf{Y}}[\hat{\beta}_1^{OLS}] = \beta_1\frac{(\mathbf{x}^* + \varepsilon^*)'\mathbf{C}_n\mathbf{x}^*}{\Delta}, \tag{2.5}$$

where $\mathbf{K}_n = (\mathbf{1}'\mathbf{x}^*\mathbf{I}_n - \mathbf{x}^*\mathbf{1}')$.

Then we take expectation over ε^* to get the expressions for the bias. This is not analytically possible because ε^* occurs in both the numerator and denominator of (2.4) and (2.5). Approximate expressions can be obtained, however, through the second order Taylor series expansion. That is

$$\begin{aligned} E_{\varepsilon^*}[f(A,B)] \approx\ & f(E_{\varepsilon^*}[A], E_{\varepsilon^*}[B]) + \frac{1}{2}\mathrm{var}(A)\left(\frac{\partial^2 f}{\partial A^2}\right) + \frac{1}{2}\mathrm{var}(B)\left(\frac{\partial^2 f}{\partial B^2}\right) \\ & + \mathrm{cov}(A,B)\left(\frac{\partial^2 f}{\partial A\partial B}\right), \end{aligned} \tag{2.6}$$

with the derivatives evaluated at the expected values of $A = A(\varepsilon^*)$ and $B = B(\varepsilon^*)$. Here $f(A,B)$ commonly denotes each fraction in the expressions of (2.4) and (2.5).

To evaluate (2.6) the following lemma is useful.

Lemma 2.1. *Suppose ε_i^* are iid random variables with mean zero and finite kth moment μ_k, where $\varepsilon^* = (\varepsilon_1^*, \ldots, \varepsilon_n^*)'$. If we set $A(\varepsilon^*) = \varepsilon^{*'}\mathbf{K}_n\varepsilon^*$ and $B(\varepsilon^*) = \mathbf{x}^{*'}\mathbf{C}_n\mathbf{x}^* + 2\mathbf{x}^{*'}\mathbf{C}_n\varepsilon^* + \varepsilon^{*'}\mathbf{C}_n\varepsilon^*$,*

$$\begin{aligned} var(B(\varepsilon^*)) &= 4n^2V\sigma_\varepsilon^2 + n(n-1)\{(n-1)\mu_4 - (n-3)\sigma_\varepsilon^4\}, \\ cov(A(\varepsilon^*), B(\varepsilon^*)) &= n(n-1)\tilde{x}\{(n-1)\mu_4 - (n-3)\sigma_\varepsilon^4\} - 2n\mu_3V, \end{aligned}$$

where $\sigma_\varepsilon^2 = \theta^2 var(\varepsilon_i)$, $\mathbf{x}^* = (x_1^*, \ldots, x_n^*)'$, $\tilde{x} = \mathbf{1}'\mathbf{x}^*/n$, and $V = \mathbf{x}^{*'}\mathbf{C}_n\mathbf{x}^*/n$.

Proof. With $\varepsilon^{*'}\mathbf{C}_n\varepsilon^* = n\sum_{i=1}^n \varepsilon_i^{*2} - (\sum_{i=1}^n \varepsilon_i^*)^2$, $\mathbf{x}^{*'}\mathbf{C}_n\varepsilon^* = n\sum_{i=1}^n x_i^*\varepsilon_i^* - \sum_{i=1}^n x_i^* \sum_{i=1}^n \varepsilon_i^*$ and $\varepsilon^{*'}\mathbf{K}_n\varepsilon^* = \sum_{i=1}^n x_i^* \sum_{i=1}^n \varepsilon_i^{*2} - \sum_{i=1}^n x_i^*\varepsilon_i^* \sum_{i=1}^n \varepsilon_i^*$, after a straightforward algebra, we see that

$$\begin{aligned} \mathrm{var}(\varepsilon^{*'}\mathbf{C}_n\varepsilon^*) &= n(n-1)\{(n-1)\mu_4 - (n-3)\sigma_\varepsilon^4\}, \\ \mathrm{E}[(\varepsilon^{*'}\mathbf{K}_n\varepsilon^*)(\varepsilon^{*'}\mathbf{C}_n\varepsilon^*)] &= n(n-1)\tilde{x}\{(n-1)\mu_4 + (n^2-2n+3)\sigma_\varepsilon^4\} \\ \mathrm{E}[(x^{*'}\mathbf{C}_n\varepsilon^*)(\varepsilon^{*'}\mathbf{K}_n\varepsilon^*)] &= -nV\mu_3. \end{aligned}$$

These results and equations $E(\varepsilon^{*'}\mathbf{K}_n\varepsilon^*) = n(n-1)\tilde{x}\sigma_\varepsilon^2$, $E(\varepsilon^{*'}\mathbf{C}_n\varepsilon^*) = n(n-1)\sigma_\varepsilon^2$, and $\mathbf{C}_n^2 = n\mathbf{C}_n$ give the theorem. □

The kth moment $\mu_k = E[\varepsilon_i^{*k}] = \theta^k E[(\varepsilon_i - \mu_\varepsilon)^k]$ can be easily obtained from Nakamura (1980) that discussed the moment of positively truncated normal distribution.

Theorem 2.1. *When we use* $\mathbf{W} = (\mathbf{1}, \mathbf{X})$ *in place of the true but unknown design matrix* $\mathbf{w} = (\mathbf{1}, \mathbf{x})$ *for the functional model (1.1), the biases of the OLS estimators are given by*

$$\begin{aligned} bias(\hat{\beta}_0^{OLS}) &= \frac{\beta_1}{1+\tau_f}\left[\tilde{x} + \frac{2\tau_f(\mu_3/\sigma_\varepsilon^2)}{n(n-1)(1+\tau_f)}\right. \\ &\quad \left. + \frac{\tilde{x}\tau_f\{(5n-3)-(n-1)(\mu_4/\sigma_\varepsilon^4)\}}{n(n-1)(1+\tau_f)^2} + O(n^{-3})\right], \qquad (2.7) \\ bias(\hat{\beta}_1^{OLS}) &= \frac{\beta_1}{1+\tau_f}\left[-1 + \frac{2\tau_f^2}{(n-1)(1+\tau_f)^2}\right. \\ &\quad \left. + \frac{\tau_f(\mu_4/\sigma_\varepsilon^4 - 3)}{n(1+\tau_f)^2} + O(n^{-2})\right], \qquad (2.8) \end{aligned}$$

where $\tau_f = V/\{(n-1)\sigma_\varepsilon^2\}$.

Proof. For (2.7), first take expectation over $\mathbf{Y}$ to get (2.4) which is $\beta_0 + \beta_1 A(\varepsilon^*)/B(\varepsilon^*)$, where $A(\varepsilon^*) = \varepsilon^{*'}\mathbf{K}_n\varepsilon^*$ and $B(\varepsilon^*) = (\mathbf{x}^{*'}\mathbf{C}_n\mathbf{x}^* + 2\mathbf{x}^{*'}\mathbf{C}_n\varepsilon^* + \varepsilon^{*'}\mathbf{C}_n\varepsilon^*)$ for $\mathbf{x}^{*'}\mathbf{K}_n\mathbf{x}^* = \mathbf{x}^{*'}\mathbf{K}_n\varepsilon^* = 0$. Since ε^* occurs in both the numerator and denominator of (2.4) it is not possible to obtain explicit values for these quantities. However, applying Lemma 2.1 to the Taylor expansion (2.6) of the function of two variables $f(A,B) = A(\varepsilon^*)/B(\varepsilon^*)$, we have the result. In the approximation, we use $E_{\varepsilon^*}[A(\varepsilon^*)] = n(n-1)\tilde{x}\sigma_\varepsilon^2$ and $E_{\varepsilon^*}[B(\varepsilon^*)] = nV + n(n-1)\sigma_\varepsilon^2$ because the ε_i^* are independent one another and have mean zero (note that $\partial^2 f/\partial A^2 = 0$). For the slope term, similar methods can be used to show (2.8). □

Setting $\theta = 1$ and $\varepsilon \sim N(0, \sigma_\varepsilon^2 \mathbf{I}_n)$, i.e. $\mu_4/\sigma_\varepsilon^4 = 3$ and $\mu_3 = 0$, we see that the biases in the Theorem 2.1 reduce to those of Morton-Jones and Henderson (2000).

3. Likelihood Function

As was assumed if errors in explanatory variables are known to be present in terms of the functional form, then the most appropriate modeling and fitting procedure will take account of the full error structure. In what follows we reparametrize the model (1.1) in terms of β_0, β_1, $\delta = -\beta_1\theta\lambda^{1/2}$ and $\lambda = 1/\sigma^2$. That is

$$\mathbf{Y} = \beta_0 + \beta_1\mathbf{X} + \delta\mathbf{z} + \mathbf{e}, \tag{3.1}$$

where $\mathbf{Y} = (y_1, \ldots, y_n)'$, $\mathbf{X} = (X_1, \ldots, X_n)'$, $\mathbf{e} = (\epsilon_1, \ldots, \epsilon_n)'$, $\mathbf{z} = (z_1, \ldots, z_n)'$. Here $\epsilon_i \overset{iid}{\sim} N(0, 1/\lambda)$ are independent of $z_i \overset{iid}{\sim} TN(0, 1/\lambda)$, a normal $N(0, 1/\lambda)$ distribution truncated to the interval $[0, \infty)$.

Under the model (3.1), the distribution of y_i is a skewed normal density (cf. Henze, 1986) given by

$$\begin{aligned} &h(y_i|\beta, \lambda, \delta) \\ &= \frac{1}{\pi}\int_0^\infty \lambda \exp\left\{-\frac{\lambda(y_i - \mathbf{X}_i'\beta - \delta z_i)^2}{2}\right\}\exp\{-\lambda z_i^2/2\}dz_i \\ &= 2\lambda(\delta)\phi(\lambda(\delta)(y_i - \mathbf{X}_i'\beta))\Phi(\delta\lambda(\delta)(y_i - \mathbf{X}_i'\beta)), \end{aligned} \tag{3.2}$$

where $\mathbf{X}_i = (1, X_i)'$, $\beta = (\beta_0, \beta_1)'$, $\lambda(\delta) = (\lambda/(1+\delta^2))^{1/2}$, and ϕ and Φ are the standard normal pdf and distribution function (df), respectively.

Lemma 3.1. *Let σ_z^2 and σ_ϵ^2 be the variances of z_i and ϵ_i, and let μ_z^3 be the standardized third moment of z_i: that is, $\mu_z^3 = E\{[z_i - E(z_i)]/\sigma_z\}^3$. Then the standardized third moment μ_y^3 of y_i is given by*

$$\mu_y^3 = \frac{\delta^3\sigma_z^3\mu_z^3}{\sigma_y^3}, \tag{3.3}$$

where $\sigma_y^2 = var(y_i) = \delta^2\sigma_z^2 + \sigma_\epsilon^2$.

Proof. From the model (3.1), we see that

$$E[y_i - E(y_i)]^3 = E[(\delta z_i - \delta E(z_i) + \epsilon_i)^3]\} = E[\delta^3(z_i - E(z_i))^3] = \delta^3\sigma_z^3\mu_z^3.$$

This and the definition of the standardized third moment of y_i give the result. □

Since $\mu_z^3 > 0$ (i.e. the distribution of z_i is skewed to the right as defined in (3.1), (3.3) implies that the marginal distribution of y_i is skewed to the right (or the left) when $\delta > 0$ (or $\delta < 0$). Therefore, we see that the skewness of the distribution of y_i is characterized by δ.

Let $D_{obs} = (n, \mathbf{X}, \mathbf{Y})$. Then the likelihood function of the model (3.1) is given by

$$L(\beta, \delta, \lambda | D_{obs}) \propto \prod_{i=1}^{n} \int_0^\infty \lambda \exp\left\{-\frac{\lambda(y_i - \mathbf{X}_i'\beta - \delta z_i)^2}{2}\right\} \exp\{-\lambda z_i^2/2\} dz_i. \quad (3.4)$$

The likelihood function shows that the distribution of y_i, which allows for flexible modeling of the functional relation of (3.1), induces the problem of evaluating the likelihood function. Recently, developments in simulation based Bayesian and classical methods have given rise to reasonably effective methods for estimating the model (cf. McFadden 1989, Albert and Chip 1993, McCulloch and Rossi 1994, and Stern 1997). Despite these developments, further improvements in fitting of the model are possible, based on Markov chain Monte Carlo methods (cf. Gelfand and Smith 1990, Chib and Greenberg 1996, and Chen and Dey 1998). In general term, Markov chain simulation methods provide a rather attractive framework for dealing with the functional model.

To ease the simulation method, we represent the likelihood (3.4) using auxiliary variables. Let $\mathbf{Z} = (z_1, \ldots, z_n)'$ and let $D = (n, \mathbf{Y}, \mathbf{X}, \mathbf{Z})$ denote the complete data. Then complete-data likelihood function of the parameters (β, δ, λ) can be written as

$$L(\beta, \delta, \lambda | D) \propto \lambda^n \exp\left\{-\frac{\lambda \sum_{i=1}^n [(y_i - \mathbf{X}_i'\beta - \delta z_i)^2 + z_i^2]}{2}\right\}. \quad (3.5)$$

This representation will ease computation involved in the Markov chain simulation method. We demonstrate this idea and the role of the auxiliary variables z_i in the MCMC algorithm in Section 4.

4. Markov Chain Monte Carlo Method

In this section we present prior distributions for the reparametrized model (3.1) and an algorithm to perform posterior computations for the model.

4.1. Prior distributions

First, we choose multivariate normal prior distribution for the regression coefficient vector $\beta = (\beta_0, \beta_1)'$ in the model presented in (3.1). That is

$$\pi(\beta|\theta_0, B_0) \propto \exp\left\{-\frac{1}{2}(\beta-\theta_0)'B_0(\beta-\theta_0)\right\}, \tag{4.1}$$

where B_0 is a precision matrix, θ_0 is a location parameter vector, and both θ_0 and B_0 are specified. Second, we choose

$$\pi(\delta|\delta_0, \tau) \propto \exp\left\{-\frac{\tau(\delta-\delta_0)^2}{2}\right\} \tag{4.2}$$

for the prior distribution of the skewness parameter, where τ and δ_0 are precision and scale parameters, respectively. The hyperparameters B_0, θ_0, τ and δ_0 are chosen to reflect the prior information. Typically, we choose $\theta_0 = 0$, $\delta_0 = 0$, $\tau = 0.01$ and $B_0 = diag\{B_1, B_2\}$ where $B_j, j = 1, 2$, are chosen to be small (e.g., $B_j = 0.01$) so that a vague prior distribution for β is obtained, which ensures that posterior is driven by the data. Finally the conjugate uninformative prior $\pi(\lambda) \propto \lambda^{-1}$ is assumed for λ, although what follows could easily be replicated for any of the family of proper, gamma prior for λ, of which $p(\lambda)$ is a limit.

4.2. The Gibbs sampler

We use Gibbs sampling (e.g., Gelfand and Smith, 1990) to perform the posterior computation. We present the steps needed to perform the Gibbs sampling algorithm for the skewed t link model. To sample from the posterior distribution

$$p(\beta, \delta, \lambda|D_{obs}) \propto \pi(\beta|\theta_0, B_0)\pi(\delta|\delta_0, \tau)\pi(\lambda)L(\beta, \delta, \lambda|D_{obs}), \tag{4.3}$$

we introduce the latent variables $\mathbf{Z} = (z_1, \ldots, z_n)'$. Then the joint posterior distribution for $\Omega = \{\beta, \delta, \mathbf{Z}, \lambda\}$ is given by

$$\begin{aligned} p(\Omega|D_{obs}) \quad &\propto \quad \lambda^{n-1}\exp\left\{-\frac{\lambda\sum_{i=1}^n[(y_i - \mathbf{X}_i'\beta - \delta z_i)^2 + z_i^2]}{2}\right\} \\ &\times\ \pi(\beta|\theta_0, B_0)\pi(\delta|\delta_0, \tau). \end{aligned} \tag{4.4}$$

From (4.3), we can derive following full conditional distributions of λ, z_i, β and δ :

$$z_i|\lambda, \delta, \beta \overset{ind}{\sim} TN_{(z_i \ge 0)}\left(\frac{\delta(y_i - \mathbf{X}_i'\beta)}{1+\delta^2}, \frac{1}{\lambda(1+\delta^2)}\right), \tag{4.5}$$

for $i = 1, \ldots, n$, where $TN_{(z_i \geq 0)}(a, b)$ denotes $N(a, b)$ truncated to the interval $(z_i \geq 0)$.

$$\lambda | \mathbf{Z}, \delta, \beta \sim Gamma\left(n, \frac{2}{\sum_{i=1}^{n}[z_i^2 + (y_i - \mathbf{X}_i'\beta - \delta z_i)^2]}\right). \tag{4.6}$$

Let $\hat{\beta} = B^{-1}(B_0\theta_0 + \lambda \sum_{i=1}^{n}(y_i - \delta z_i)\mathbf{X}_i)$ and $B = B_0 + \lambda \sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i'$. Then

$$\beta | \lambda, \mathbf{Z}, \delta \sim N(\hat{\beta}, B^{-1}). \tag{4.7}$$

Finally, if $\hat{\delta} = C^{-1}(\delta_0 \tau + \lambda \sum_{i=1}^{n} z_i(y_i - \mathbf{X}_i'\beta))$ and $C = \tau + \lambda \sum_{i=1}^{n} z_i^2$. Then full conditional distribution of δ is given by

$$\delta | \lambda, \mathbf{Z}, \beta \sim N(\hat{\delta}, C^{-1}). \tag{4.8}$$

To implement the Gibbs sampler, we start with initial value of Ω and cycle through the conditional distributions (4.5), (4.6), (4.7) and (4.8), in that order.

5. Model Comparison

In Section 4, we proposed a Bayesian estimation of the functional model (3.1) in which the distribution of responses, y_i are determined by δ. Therefore it is of practical interest to compare models formulated by different choices of δ in (3.1), usual regression model (M_0) with $\delta = 0$ and the functional one (M_1) with $\delta \neq 0$. To this end, we propose an algorithm via the conditional Bayes factor approach by Geweke (1996) in order to perform the model comparison. The algorithm can be made by modifying the posterior simulation step of δ in the Gibbs sampler of Section 4. Note that $\delta = 0$ is equivalent to $\theta = 0$ in (1.1) for $\delta = -\beta_1 \theta \lambda^{1/2}$ with $\beta_1 \neq 0$ and $\lambda \neq 0$.

Under the same prior distributions for β and λ, the prior distribution of δ, is modified as follows. With prior probability q, $\delta = 0$; conditional on $\delta \neq 0$ the prior distribution of δ is $N(\delta_0, 1/\tau)$:

$$d\Pi(\delta) = qdH(\delta) + (1 - q)(2\pi)^{-1/2}\tau^{1/2}\exp\left\{-\frac{\tau(\delta - \delta_0)^2}{2}\right\}, \tag{5.1}$$

where $\Pi(\cdot)$ denotes the prior c.d.f. of δ; $H(\delta) = 0$ if $\delta < 0$ and $H(\delta) = 1$ if $\delta \geq 0$.

The conditional distributions of λ, z_i and β, involved in the algorithm are the same as the Gibbs sampler. For the model selection, we apply the conditional

Bayes factor approach for δ. Given $\mathbf{Z}$, β and λ, define $w_i = y_i - \mathbf{X}_i'\beta$, so that the conditional distribution of δ follows from the simplified model

$$w_i = \delta z_i + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} N(0, 1/\lambda), \quad i = 1, \ldots, n. \tag{5.2}$$

The likelihood function kernel is

$$\exp\left\{-\frac{\lambda \sum_{i=1}^{n}(w_i - \delta z_i)^2}{2}\right\}.$$

Conditional on $\delta = 0$ the value of the kernel is

$$\exp\left\{-\frac{\lambda \sum_{i=1}^{n} w_i^2}{2}\right\}. \tag{5.3}$$

Conditional on $\delta \neq 0$ the corresponding kernel density for δ is

$$\left(\frac{\tau}{2\pi}\right)^{1/2} \exp\left\{-\frac{\lambda \sum_{i=1}^{n}(w_i - \delta z_i)^2}{2}\right\} \exp\left\{-\frac{\tau(\delta - \delta_0)^2}{2}\right\}$$
$$= \left(\frac{\tau}{2\pi}\right)^{1/2} \exp\left\{-\frac{\lambda \sum_{i=1}^{n} w_i^2 + \tau\delta_0^2 - C\hat{\delta}^2}{2}\right\} \exp\left\{-\frac{C(\delta - \hat{\delta})^2}{2}\right\}, \tag{5.4}$$

where C and $\hat{\delta}$ denote the same quantities as used for the Gibbs sampler in Section 4.

To calculate the conditional Bayes factor, it is necessary to integrate (5.4) over δ which yields the conditional marginal likelihood

$$\left(\frac{\tau}{C}\right)^{1/2} \exp\left\{-\frac{\lambda \sum_{i=1}^{n} w_i^2 + \tau\delta_0^2 - C\hat{\delta}^2}{2}\right\}. \tag{5.5}$$

Comparing this marginal likelihood to (5.3), we have the conditional Bayes factor in favor of $\delta \neq 0$, versus $\delta = 0$, that is

$$BF^c = \left(\frac{\tau}{C}\right)^{1/2} \exp\left\{\frac{C\hat{\delta}^2 - \tau\delta_0^2}{2}\right\}. \tag{5.6}$$

To draw δ from its conditional distribution, the conditional posterior probability that $\delta = 0$ is computed from the conditional Bayes factor (5.6) :

$$q^c = q/\{q + (1 - q)BF^c\}. \tag{5.7}$$

Based on a comparison of this probability with a drawing from the uniform distribution on [0, 1], the choice $\delta = 0$ or $\delta \neq 0$ is made. Therefore modifying the fourth step of the Gibbs algorithm, we have the following algorithm for the model comparison.

Algorithm for model comparison :

- Independently generate z_i, $i = 1, \ldots, n$ as the Gibbs sampler;
- Generate λ as the Gibbs sampler;
- Generate β as the Gibbs sampler;
- The parameter δ is drawn so that, for δ, compute q^c from BF^c and generate u from $U(0,1)$, if $u \le q^c$, set $\delta = 0$. Else, sample δ from $N(\hat{\delta}, C^{-1})$.

The algorithm proceeds in the usual way. The model comparison could be done in the obvious way, by recording the indicator variables for the model corresponding to the nonzero δ's at the end of each iteration. That is

$$Pr(M_1|D) = Pr(\delta \neq 0|D) = \frac{\#\text{ of nonzero }\delta}{\#\text{ of iterations}}. \tag{5.8}$$

6. Illustrative Example

We now conduct a simulation study for the functional regression model with the truncated errors in explanatory variables. The objectives in this study are (a) to illustrate the numerical accuracy of the Bayesian estimation method and (b) to examine the performance of the model comparison using the conditional Bayes factor approach.

In this study we consider a simulation with the functional model. We generate multiple (100 replicated) simulated datasets from the following model :

$$y_i = \beta_0 + \beta_1 X_i + \delta z_i + \epsilon_i, \quad i = 1, \ldots, n, \tag{6.1}$$

where $\epsilon_i \overset{iid}{\sim} N(0, 1/\lambda)$, $z_i \overset{iid}{\sim} TN(0, 1/\lambda)$.

First, we set $X_i = .5 \times i$, $i = 1, \ldots, n$ ($n = 30, 100, 200$), and then generate n independent response variables, y_i from the model with one covariate X_i, $\beta_0 = 2$, $\beta_1 = 3$, $\delta = -2$ and $\lambda = 1$. For each simulated data of $y_i's$ and $X_i's$, we conduct the proposed Bayesian estimation via the Gibbs sampling method. In the estimation, we considered following values of the hyperparameters, $B_0 = .01I_2$, $\tau = .01$ and

$\theta_0 = \delta_0 = 0$ so that a vague prior distributions for β_0, β_1, λ and δ are obtained, which ensures that the posteriors are driven by the data.

The Gibbs sampler described in Section 4.2 was used to obtain simulated samples $\beta = (\beta_0, \beta_1)'$, δ and λ of each size 10,000 for the model. Many standard diagnostic measures, (see, e.g., Cowles and Carlin, 1996) have been calculated to monitor convergence and those indicated rapid convergence.

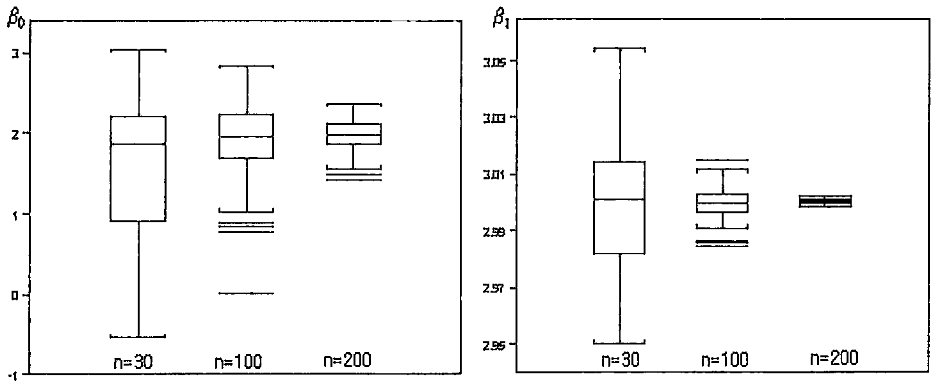

FIGURE 6.1 *Boxplots of the posterior means of the parameters*

Based on the diagnostics we discard first 1000 iterates in each data case and use the subsequent 10,000 iterates to make inference. Boxplots of 100 posterior means (the average of simulated values) of the regression coefficients (β_0 and β_1) in the model (5.8) are displayed in Figure 6.1. These boxplots indicate that the Gibbs sampler has given estimates concentrated on the true values that generated data. Further, the estimation is improved as n become larger.

Second, we independently generate $X_i = .5 \times i$, $i = 1, \ldots, 100$, and then generate 100 independent response variables, y_i from the model (4.6) with $\beta_0 = 2$, $\beta_1 = 3$, $\delta = -2$ and $\lambda = 1$.

For this simulated dataset, we conduct the proposed Bayesian estimation using the same hyperparameters as the foregoing example. The Gibbs sampling results are noted in Table 6.1. From Table 6.1 it is clear that the Gibbs sampling method has accurately produced posterior distribution concentrated on each parameter value that generated data.

TABLE 6.1 *Summaries of the posterior distributions*

Para.	*Mean*	*SD*	*Median*	*Lower (95%)*	*Upper (95%)*
β_0	1.941	0.215	1.935	1.669	2.221
β_1	3.002	0.003	3.002	2.999	3.006
δ	-1.765	0.353	-1.750	-2.234	-1.322
λ	0.999	0.136	0.987	0.835	1.174

Finally, to examine the performance of the algorithm for the model comparison in Section 4, it is applied to multiple (100) simulated data sets of size $n = 100$ from the following model.

$$y_i = 2 + 3X_i + \delta z_i + \epsilon_i, \quad i = 1, \dots, n, \tag{6.2}$$

where $\epsilon_i \overset{iid}{\sim} N(0,1)$, $z_i \overset{iid}{\sim} TN(0,1)$. For illustrative purpose we take $Pr(\delta = 0) = q = 0.5$, as a base prior probability that δ is excluded from the model. To study the relation between the prior and the posterior distribution of δ's in conjunction with the algorithm, we also consider $q = 0.2$ and $q = 0.8$. In order to reflect the vagueness of the prior information about β_0, β_1, δ and λ, we represent our prior distribution through the hyperparameters $B_0 = .01I_2$, $\tau = .01$ and set $\theta = 0$ and $\delta_0 = 0$.

For each given value of $\delta = 0, -.3, -.9, -1.5, -2$ (equivalent to $\theta = 0, .1, .3, .5, .75$), posterior probabilities of alternative states of δ are obtained from the 100 simulated data sets and their mean and standard deviation are presented in Table 6.2. Figure 6.2 is also presented to show the distribution of the probabilities. To obtain the probabilities the method described in Section 4 is used, with $m = 10^4$ iterations of the algorithm beyond 10^3 burn-in iterations. From the table we see that a systematic effect of the value of q on the posterior probabilities is evident. Increases in q, the prior probability that $\delta = 0$, tends to favor the usual linear regression model. Thus giving more informative prior to δ has the potential to effect our posterior inference about δ. From Table 6.2, it can be also observed that, regardless of the particular prior q, the partial Bayes factor method yields the largest posterior probability for the true state ($\delta = 0$ or $\delta \neq 0$).

TABLE 6.2 *Mean and standard deviation (in parenthesis) of posterior probabilities* $Pr(\delta \neq 0|D)$

Prior	*State*	$\delta = 0$ ($\theta = 0$)	$\delta = -0.3$ ($\theta = 0.1$)	$\delta = -0.9$ ($\theta = 0.3$)	$\delta = -1.5$ ($\theta = 0.5$)	$\delta = -2$ ($\theta = 0.75$)
$q = 0.2$	$\delta \neq 0$	0.271	0.402	0.782	0.901	0.998
		(.102)	(.283)	(.294)	(.214)	(.046)
$q = 0.5$	$\delta \neq 0$	0.203	0.381	0.743	0.874	0.923
		(.198)	(.307)	(.355)	(.231)	(.102)
$q = 0.8$	$\delta \neq 0$	0.139	0.320	0.638	0.712	0.825
		(.230)	(.326)	(.362)	(.273)	(.129)

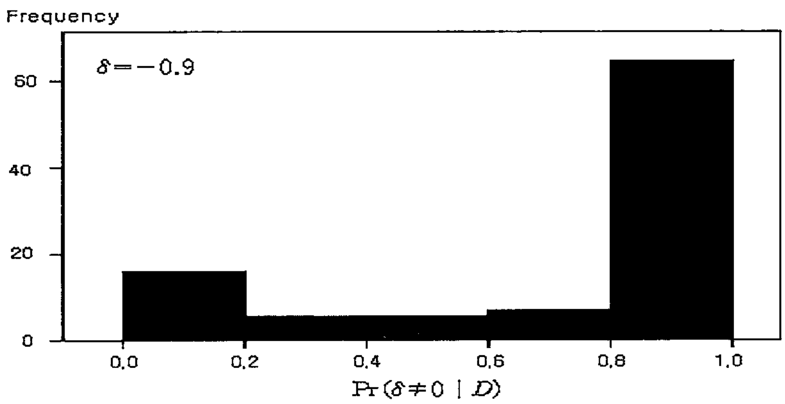

FIGURE 6.2 *Histogram of* $Pr(\delta \neq 0|D)$ *when* $\delta = -0.9$ *with* $q = 0.5$

7. Concluding Remarks

Clearly, an efficient estimation procedure should take into account the full error structure in any data set and should allow for errors in explanatory variables if these are known or suspected to be present. In practice, however, due to lack of testing procedure for the existence of the errors in explanatory variables, investigators often assume that explanatory variables are not subject to error. We have shown that this assumption leads to OLS estimators to be biased.

This paper has presented a Bayesian method for analyzing a functional regression model with truncated error in an explanatory variable. The model is described in terms of a skewed normal distribution by Henze (1986) for its underlying error variables that are consisting of errors from response and explanatory variables. In addition, the paper has established Bayesian techniques for analyzing the functional model from the output of posterior simulation via Markov chain Monte Carlo method. Our illustrative simulation study suggests that (i) the techniques can be applied to data sets of varying complexity and to high dimensional models that are intractable by using a frequentist method; (ii) the goodness of fit of the functional model can be well tested by the suggested conditional Bayes factor method.

Finally, the principle of analyzing the functional model suggested in this paper can be extended to the multiple linear regression model with truncated errors in independent variables, and this extension is now under investigation.

REFERENCES

Albert, J. and Chib, S. (1993). "Bayesian analysis of binary and polytomous response data", *Journal of the American Statistical Association*, **88**, 669-679.

Azzalini, A. and Valle, A. D. (1996). "The multivariate skew-normal distribution", *Biometrika*, **83**, 715-723.

Carroll, R. J., Ruppert, D. and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*, Chapman and Hall, London.

Chen, M.-H. and Dey, D. K. (1998). "Bayesian modeling of correlated binary response responses via scale mixture of multivariate normal link model functions", *The Indian Journal of Statistics*, **A60**, 322-343.

Chib, S. and Greenberg, E. (1996). "Bayesian analysis of multivariate probit models in MCMC", preprint service.

Cochran, W. G. (1968). "Errors of measurement in Statistics", *Journal of the American Statistical Association*, **63**, 637-666.

Cowles, M. K. and Carlin, B. P. (1996). "Markov chain Monte Carlo convergence diagnostics: A comparative review", *Journal of the American Statistical Association*, **91**, 883-904.

DePriest, D. J. (1983). "Using the singly truncated normal distribution to analyze satellite data", *Communications in Statistics-Theory and Methods*, **12**, 263-272.

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

Gelfand, A. E. and Smith, A. F. M. (1990). "Sampling-based approaches to calculating marginal densities", *Journal of the American Statistical Association*, **85**, 398-409.

Geweke, J. (1996). "Variable selection and model comparison in regression", In *Bayesian Statistics 5.* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.), 609-620, Oxford University Press.

Henze, N. (1986). "A probabilistic representation of the skewed-normal distribution", *Scandinavian Journal of Statistics*, **13**, 271-275.

McCulloch, R. E. and Rossi, P. E. (1994). "Exact likelihood analysis of the multinomial probit model", *Journal of Econometrics,* **64**, 207-240.

McFadden, D. (1989). "A method of simulated moments for estimation of discrete response models without numerical integration", *Econometrica*, **57**, 995-1026.

Morton-Jones, T. and Henderson, R. (2000). "Generalized least squares with ignored errors in variables", *Technometrics*, **42**, 366-375.

Nakamura, T. (1980). "On the moment of positively truncated normal distribution", *Journal of the Japan Statistical Society*, **10**, 139-144.

Richardson, D. H. and Wu, D.-M. (1970). "Least squares and grouping method estimators in the errors in variables model", *Journal of the American Statistical Association*, **65**, 724-748.

Schaalje, G. B. and Butts, R. A. (1993). "Some effects of ignoring correlated measurement errors in straight line regression", *Biometrics*, **49**, 1262-1267.

Stern, S. (1997). "Simulation-based estimation", *Journal of Economic Literature*, **35**, 2006-2039.