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

Hea-Jung Kim
Dongguk University, Department of Statistics
Seoul 100-715, Korea
kim3hj@dongguk.edu

1. Introduction

Denote the n-dimensional response vector by $\mathbf{Y}$ and the corresponding vector of true explanatory variables by $\mathbf{x}$. Assume a functional regression model

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

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 independently distributed as $\varepsilon_i \sim TN(0,1),\ i = 1, \ldots, n$, independent of $\mathbf{Y}$. Here $TN(0,1)$ denotes a standard normal distribution truncated to the interval $[0, \infty)$ 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 (or plus) measurement error structure in the explanatory variable.

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) is

$$\hat{\beta}^{OLS} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{Y}. \tag{2}$$

In considering the bias, we take expectations over $\mathbf{Y}$ first to arrive at the biases of the OLS estimators for the functional model (1) given by

$$\text{bias}(\hat{\beta}_0^{OLS}) \;=\; \frac{\beta_1}{1+\tau_f}\left[\bar{x} + \frac{2\tau_f(\mu_3/\sigma_\varepsilon^2)}{n(n-1)(1+\tau_f)}\right.$$

$$+ \quad \frac{\bar{x}\tau_f\{(5n-3)-(n-1)(\mu_4/\sigma_\varepsilon^4)\}}{n(n-1)(1+\tau_f)^2} + O(n^{-3})\Bigg], \tag{3}$$

$$\text{bias}(\hat{\beta}_1^{OLS}) = \frac{\beta_1}{1+\tau_f}\left[-1+\frac{2\tau_f^2}{(n-1)(1+\tau_f)^2}\right.$$

$$+ \quad \frac{\tau_f(\mu_4/\sigma_\varepsilon^4-3)}{n(1+\tau_f)^2} + O(n^{-2})\Bigg], \tag{4}$$

where $\tau_f = V/\{(n-1)\sigma_\varepsilon^2\}$ and $V = \sum_{i=1}^n (x_i - \bar{x})^2$.

3. Bayesian Analysis

In what follows we reparameterize the model (1) in terms of β_0, β_1, $\delta = -\beta_1\theta\lambda^{1/2}$ and $\lambda = 1/\sigma^2$. That is

$$\mathbf{Y} = \beta_0 + \beta_1\mathbf{X} + \delta\mathbf{z} + \mathbf{e}, \tag{5}$$

where $\mathbf{Y} = (y_1, \ldots, y_n)'$, $\mathbf{X} = (X_1, \ldots, X_n)'$, $\mathbf{e} = (\epsilon_1, \ldots, \epsilon_n)'$ and $\mathbf{z} = (z_i, \ldots, z_n)'$. 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 $[0, \infty)$.

The likelihood function of the model (5) is given by

$$L(\beta, \delta, \lambda | D_{obs})$$

$$= \prod_{i=1}^n \int_0^\infty \frac{2\lambda}{\pi} \exp\left\{-\frac{\lambda(y_i - \mathbf{X}_i'\beta - \delta z_i)^2}{2}\right\} \exp\{-\lambda z_i^2/2\} dz_i. \tag{6}$$

3.1. Prior Distributions

We choose multivariate normal prior distribution for the regression coefficient vector $\beta = (\beta_0, \beta_1)'$ in the model presented in (5). That is $\beta|\theta_0, B_0 \sim N_2(\theta_0, B_0^{-1})$, and $\delta|\delta_0, \tau \sim N(\delta_0, \tau^{-1})$. 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.

3.2. The Gibbs Sampler

Introducing the latent variables $\mathbf{Z} = (z_1, \ldots, z_n)'$.

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

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

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

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

To implement the Gibbs sampler, we start with initial value of Ω and cycle through the conditional distributions in that order.

3.3. Model Comparison

It is of practical interest to compare models formulated by different choices of δ in (5), 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.

With prior probability q, $\delta = 0$; conditional on $\delta \neq 0$ the prior distribution of δ is $N(\delta_0, 1/\tau)$:

$$d\Pi(\delta) = qdH(\delta) + (1-q)(2\pi)^{-1/2}\tau^{1/2}\exp\left\{-\frac{\tau(\delta-\delta_0)^2}{2}\right\}, \tag{7}$$

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

Comparing this marginal likelihoods, we have the conditional Bayes factor in favor of $\delta \neq 0$, versus $\delta = 0$, that is

$$BF^c = \left(\frac{\tau}{C}\right)^{1/2} \exp\left\{\frac{C\hat{\delta}^2 - \tau\delta_0^2}{2}\right\}. \tag{8}$$

To draw δ from its conditional distribution, the conditional posterior probability that $\delta = 0$ is computed from the conditional Bayes factor (12):

$$q^c = q/\{q + (1-q)BF^c\}. \tag{9}$$

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` $u \leq q^c$, `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

$$Pr(M_1|D) = Pr(\delta \neq 0|D) = \frac{\#\text{ of nonzero }\delta}{\#\text{ of iterations}}. \tag{10}$$

REFERENCES

Geweke, J. (1996). Variable Selection and Model Comparison in Regression. *Bayesian Statistics 5.* eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford University Press, 609-620.

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

Richardson, D. H. and Wu, D.-M. (1970). Least squares and grouping method estimators in the errors in variables model, *J. Amer. Statist. Asso.,* **65**, 724-748.