Bayesian Analysis of Two-Piece Normal Regression Models

Teruo Nakatsuma

Department of Economics, Keio University

Abstract

This paper develops a Markov chain Monte Carlo method for Bayesian inferences about a linear regression model with a two-piece-normal-distributed error term, or two-piece normal regression model for short. The two-piece normal distribution is a generalization of the normal distribution. Its advantage over the normal distribution is that it is asymmetric and thick-tailed, but more tractable than gamma-type distributions. We describe how to analyze the two-piece normal regression model with a Markov chain Monte Carlo technique and show a few examples of applications to a stochastic frontier model for the illustration of this technique.

Key words: Two-piece normal distribution; Regression model; Markov chain Monte Carlo; Gibbs sampler; Hit-and-Run algorithm; Stochastic frontier model.

1 Introduction

According to Kimber (1985), the two-piece normal (TPN) distribution or jointed half-Gaussian distribution has been used in ion-implantation research and other related area. The pdf of a TPN distribution is

$$f(\epsilon) = \begin{cases} \dfrac{2}{\sigma_1 + \sigma_2}\phi\left(\dfrac{\epsilon - \alpha}{\sigma_1}\right), & \text{if } \epsilon \le \alpha; \\ \\ \dfrac{2}{\sigma_1 + \sigma_2}\phi\left(\dfrac{\epsilon - \alpha}{\sigma_2}\right), & \text{if } \epsilon > \alpha, \end{cases} \tag{1}$$

where $\phi(\cdot)$ is the pdf of the standard normal distribution. In Figure 1, several TPN distributions with different σ_2 (σ_1 is set to be unity) are plotted. Statistical properties of the TPN distribution were studied by John (1982) and Kimber (1985). Some of them are summarized in Appendix A. One important property of the TPN distribution is that it is asymmetric when $\sigma_1 \neq \sigma_2$; it is skewed to the left if $\sigma_1 > \sigma_2$ and skewed to the right if $\sigma_1 < \sigma_2$. The TPN distribution is also thick-tailed, i.e., its kurtosis is more than 3. Since the TPN distribution becomes a normal distribution when $\sigma_1 = \sigma_2$, it is regarded as a generalization of the normal distribution.

Although the TPN distribution is rarely applied in econometrics, a similar jointed half normal distribution was considered by Aigner et al. (1976):

$$\epsilon = \begin{cases} \dfrac{\sigma z}{\sqrt{\theta}}, & \text{if } z \leq 0; \\ \\ \dfrac{\sigma z}{\sqrt{1-\theta}}, & \text{if } z > 0, \end{cases} \tag{2}$$

where $z \sim$ i.i.d.$\mathcal{N}(0,1)$ and $0 < \theta < 1$. This distribution (2) is not symmetric around zero, but $\Pr(\epsilon \leq 0) = \Pr(\epsilon > 0) = 1/2$ by construction. Thus on average positive observations are realized as often as non-positive observations. Furthermore the pdf of (2) is discontinuous at zero. For the TPN distribution, on the other hand, when $\alpha = 0$, $\Pr(\epsilon \leq 0) = \sigma_1/(\sigma_1 + \sigma_2)$ and $\Pr(\epsilon > 0) = \sigma_2/(\sigma_1 + \sigma_2)$. Hence it is possible that positive observations are more likely or less likely to be realized than non-positive ones for the TPN distribution. Moreover the pdf of the TPN distribution is continuous everywhere and first-order differentiable at α.

Originally, Aigner et al. (1976) proposed the distribution (2) to estimate a stochastic frontier model. A stochastic frontier model appears in empirical studies of individual firms with micro data. In an economic analysis of individual firms, we often encounter functions which represent optimal values related to activities of firms. For example, a firm's cost function represents the minimum level of costs the firm can attain given the level of an output, factor prices, and technology of production. If a cost function is given as $f(\boldsymbol{x}, \boldsymbol{\beta})$ where $\boldsymbol{\beta}$ is a vector of parameters and $\boldsymbol{x}$ is a vector of factor prices and output level, any observed level of costs y must be greater than or equal to the theoretical minimum level of costs $f(\boldsymbol{x}, \boldsymbol{\beta})$, i.e., $y \geq f(\boldsymbol{x}, \boldsymbol{\beta})$. This inequality implies that the error term in a regression model of the cost function must be non-negative. Furthermore it is possible that an observed level of costs y is affected by unobservable idiosyncratic factors which may have either positive or negative impact on the level of costs. Hence we have a regression model of the cost function $y = f(\boldsymbol{x}, \boldsymbol{\beta}) + u + v$ where the first error term u must be non-negative while the second error term v can be either positive or negative. This model is often called a stochastic frontier model.

Application of a stochastic frontier model is not limited in a cost analysis of firms. If we regard $f(\boldsymbol{x}, \boldsymbol{\beta})$ as a firm's production function, $f(\boldsymbol{x}, \boldsymbol{\beta})$ gives the maximum level of output given quantities of inputs $\boldsymbol{x}$. This is another type of a stochastic frontier model. For a stochastic frontier model of production, the first error term u must be non-positive so that it would represent some inefficiency in the firm's production process. We can also apply a stochastic frontier model to estimate a firm's profit function. The model specification of a stochastic frontier model of profit is basically the same as a stochastic frontier model of production.

In many applications of stochastic frontier models, the frontier function, $f(\boldsymbol{x}, \boldsymbol{\beta})$, is supposed to be linear, i.e., $f(\boldsymbol{x}, \boldsymbol{\beta}) = \boldsymbol{x}'\boldsymbol{\beta}$. The second error term v is supposed to be normal in most cases. The distribution of u is supposed to a half-normal distribution or exponential distribution [Aigner et al. (1977)], truncated normal distribution [Stevenson (1980)], gamma distribution [Greene (1990)], and so forth. If we combine two error terms u and v into one error term $\epsilon = u + v$, the distribution of ϵ is not symmetric around zero. ϵ must be skewed to the right in a cost function while it must be skewed to the left in a production or profit function. Aigner et al. (1976) argued that the distribution (2) could be suitable for such ϵ. In this paper, instead of (2), we propose to use the TPN distribution for ϵ to estimate a stochastic frontier model.

A difficult part of estimating a stochastic frontier model with a TPN-distributed error term is that the likelihood function of the model is neither continuous nor globally concave with respect to the regression coefficients $\boldsymbol{\beta}$. This discontinuity prevents us from using a gradient method to compute the maximum likelihood estimates. In this paper, we consider a Bayesian approach to the TPN distribution instead. In a Bayesian approach, we need to evaluate multiple integrals in calculating posterior statistics such as means, medians, and variances of parameters, computing credible or highest posterior density intervals, and marginalizing the joint posterior distribution. For the TPN distribution, closed-form expressions of those integrals are not available and we must evaluate them numerically. To do so, we apply a Markov chain Monte Carlo (MCMC) method which is a Monte Carlo integration method coupled with a Markov chain sampling from the posterior distribution. Research on the MCMC method has been rapidly expanding for recent years and various MCMC algorithms have been developed in the literature. See Robert and Casella (1999) and Chen et al. (2000) among others for full details on the MCMC. In out study, we found that a Gibbs sampler coupled with a Hit-and-Run (H&R) algorithm worked fine for the model of the TPN distribution.

There is another Bayesian approach to the stochastic frontier model in which one attempts to estimate u for individual firms. Broeck et al. (1994) applies a importance sampling method to estimate u while Koop et al. (1997) uses a Gibbs sampler for the same estimation. Their approach and ours are complementary to each other. In their approach individual u can be estimated, which is the major advantage of theirs, but it is assumed that all firms are inefficient at some degree and the function $f(\boldsymbol{x}, \boldsymbol{\beta})$ is indeed a frontier. In our approach, on the other hand, we do not necessarily assume a priori that $f(\boldsymbol{x}, \boldsymbol{\beta})$ is a frontier. Instead, we can test whether $f(\boldsymbol{x}, \boldsymbol{\beta})$ can be interpreted as a frontier or not by examining the shape of the distribution of the error term.

Organization of the paper is as follows. In Section 2, we introduce a linear regression model with a TPN-distributed error term. In Section 3, we explain outlines of a Bayesian analysis of the model and describe a Markov chain Monte Carlo method for to the model. In Section 4, a numerical example with simulated data, and applications to a cost function of the electric utility industry and a production function of the transportation equipment industry are presented. Concluding remarks are given in Section 5.

2 The Model

Reparameterizing $(\sigma_1, \sigma_2) \to (\omega, \gamma)$ as $\gamma = \sigma_1/(\sigma_1 + \sigma_2)$ and $\omega = \sigma_1 + \sigma_2$, the pdf of a TPN distribution is rewritten as

$$f(\epsilon) = \begin{cases} \dfrac{2}{\omega}\phi\left(\dfrac{\epsilon - \alpha}{\gamma\omega}\right), & \text{if } \epsilon \leq \alpha, \\ \\ \dfrac{2}{\omega}\phi\left(\dfrac{\epsilon - \alpha}{(1-\gamma)\omega}\right), & \text{if } \epsilon > \alpha. \end{cases} \tag{3}$$

In this paper, $\mathcal{TPN}(\alpha, \omega, \gamma)$ denotes a TPN distribution (3). The threshold value α is regarded as a location parameter since with a transformation $\epsilon \to \epsilon + d$, it is transformed as $\alpha \to \alpha + d$. With a transformation $\epsilon \to c\epsilon$, ω is transformed as $\omega \to c\omega$. Thus ω is regarded as a scale parameter. Therefore if ϵ is transformed as $\epsilon \to c\epsilon + d$, the distribution $TPN(\alpha, \omega, \gamma)$ is transformed into $\mathcal{TPN}(c\alpha + d, c\omega, \gamma)$. In particular, if the transformation is $\epsilon \to (\epsilon - \alpha)/\omega$, the new distribution is $\mathcal{TPN}(0, 1, \gamma)$. Note that γ is invariant under this transformation. This property will turn out to be useful later when we analyze a model of the TPN distribution. γ, the ratio of σ_1 to the sum of σ_1 and σ_2, is regarded as a "shape" parameter. A TPN distribution is skewed to the left when $\gamma > 1/2$ while it is skewed to the right when $\gamma < 1/2$. If $\gamma = 1/2$, the distribution is symmetric around α and becomes a normal distribution $\mathcal{N}(\alpha, \omega^2/4)$. γ also gives the probability that ϵ is less than or equal to α. Thus values less than or equal to (more than) α are more likely to be observed if $\gamma > 1/2$ ($\gamma < 1/2$). This property is an advantage over the asymmetric distribution (2) proposed by Aigner et al. (1976) whose median is always equal to the threshold value.

In our study, we will focus our attention on the case that $f(\boldsymbol{x}, \boldsymbol{\beta})$ is linear. So let us consider a linear regression model with a TPN-distributed error term:

$$y_i = \alpha + \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i, \quad \epsilon_i \sim \text{i.i.d. } \mathcal{TPN}(0, \omega, \gamma), \quad (i = 1, \ldots, n) \tag{4}$$

where y_i is a scalar value of a dependent variable, $\boldsymbol{x}_i$ is a $k \times 1$ vector of independent variables excluding the constant term, and $\boldsymbol{\beta}$ is a $k \times 1$ vector of regression coefficients. In this paper, we refer (4) as a two-piece normal regression model (TPNRM).

To understand the nature of the TPNRM, let us consider the conditional distribution of y_i given $\boldsymbol{x}_i$. The conditional distribution of y_i is also a TPN distribution and its pdf is given as

$$f(y_i|\boldsymbol{x}_i) = \begin{cases} \dfrac{2}{\omega}\phi\left(\dfrac{y_i - \alpha - \boldsymbol{x}_i'\boldsymbol{\beta}}{\gamma\omega}\right), & \text{if } y_i \leq \alpha + \boldsymbol{x}_i'\boldsymbol{\beta}, \\ \\ \dfrac{2}{\omega}\phi\left(\dfrac{y_i - \alpha - \boldsymbol{x}_i'\boldsymbol{\beta}}{(1-\gamma)\omega}\right), & \text{if } y_i > \alpha + \boldsymbol{x}_i'\boldsymbol{\beta}. \end{cases} \tag{5}$$

One important feature of the TPNRM is that, unlike the classical regression model, the regression function $\alpha + \boldsymbol{x}_i'\boldsymbol{\beta}$ gives the mode of the conditional distribution (5), instead of the conditional expected value $\mathrm{E}(y_i|\boldsymbol{x}_i)$. Therefore the ordinary least squares estimator (OLSE) of $\boldsymbol{\beta}$ is biased and not consistent.

Another notable feature of the TPNRM is that the shape parameter γ determines whether our interpretation of the TPNRM is suitable or not. To explain this, let us consider a simple regression case in which the regression function is a straight line. If $\gamma < 1/2$, values above the regression line are more likely to be observed those below the regression line. This is illustrated in Figure 2. In Figure 2, we plot data generated by a simple regression model with a TPN-distributed error term:

$$y_i = \alpha + \beta x_i + \epsilon_i, \quad \epsilon_i \sim \text{i.i.d. } \mathcal{TPN}(0, \omega, \gamma).$$

We set $\alpha = \beta = 1$ and $\gamma = 0.2$. We set $\omega = 2/\sqrt{\gamma(1-\gamma) + (1 - 2\pi^{-1})(1-2\gamma)^2} \approx 3.7087$ to make $\mathrm{Var}(\epsilon_i)$ equal to 4. x_i is generated from $\mathcal{U}(-\sqrt{3}, \sqrt{3})$, and the number of observation is $n = 200$. The straight line in Figure 2 represents the regression line $\alpha + \beta x$. Obviously, the majority of observations are located above the regression line. In such a case, it may be acceptable to interpret the TPNRM as a cost function since the distribution of the error term in a stochastic cost frontier model is supposed to be skewed to the right and as a result more residuals are expected to be located above the regression line. If $\gamma > 1/2$, on the other hand, values below the regression line are more likely to be observed. Since a left-skewed distribution of the error term is not consistent with the nature of a stochastic cost frontier model, it is not reasonable to assume that the regression line represent a cost function when $\gamma > 1/2$. However, when we estimate a stochastic production frontier model, γ must be more than 1/2 in order to interpret the regression line as a production function. Therefore whether γ is more than 1/2 or less than 1/2 is crucial in an analysis of a stochastic frontier model, and we may use this feature to check validity of model specification of the TPNRM. For example, if we try to estimate (4) as a cost function but end up with the estimate of γ such that $\gamma > 1/2$, we might as well suspect that something is wrong with our model specification of the cost function.

The third important feature of the TPNRM is that the classical normal regression model is a special case of the TPNRM. Suppose that $\gamma = 1/2$. Then (5) equals

$$f(y_i|\boldsymbol{x}_i) = \frac{2}{\omega}\phi\left(\frac{y_i - \alpha - \boldsymbol{x}_i'\boldsymbol{\beta}}{\omega/2}\right), \tag{6}$$

and the TPNRM (4) becomes

$$y_i = \alpha + \boldsymbol{x}_i'\boldsymbol{\beta} + \frac{\omega}{2}z_i, \quad z_i \sim \text{i.i.d.}\, \mathcal{N}(0,1). \tag{7}$$

This is the classical normal regression model. Since the classical normal regression model is nested in the TPNRM, we may test assumption of the classical normal regression model against the TPNRM by checking how γ is close to 1/2.

Since the OLSE of the TPNRM is biased and not consistent, we may estimate the model with the maximum likelihood estimator (MLE). The likelihood function for the TPNRM is given as

$$p(\boldsymbol{y}|\alpha,\boldsymbol{\beta},\omega,\gamma,\boldsymbol{X}) = \frac{2^n}{\omega^n}\prod_{i=1}^{n}\phi\left(\frac{y_i - \alpha - \boldsymbol{x}_i'\boldsymbol{\beta}}{\gamma\omega}\right)^{d_i}\phi\left(\frac{y_i - \alpha - \boldsymbol{x}_i'\boldsymbol{\beta}}{(1-\gamma)\omega}\right)^{1-d_i}, \tag{8}$$

where $\boldsymbol{y} = [y_1,\dots,y_n]'$, $\boldsymbol{X} = [\boldsymbol{x}_1,\dots,\boldsymbol{x}_n]'$, and d_i $(i = 1,\dots,n)$ is a dummy variable defined as

$$d_i = \begin{cases} 1, & \text{if } y_i \leq \alpha + \boldsymbol{x}_i'\boldsymbol{\beta}, \\ 0, & \text{if } y_i > \alpha + \boldsymbol{x}_i'\boldsymbol{\beta}. \end{cases}$$

Although the likelihood function (8) is continuous with respect to ω and γ, it is not continuous with respect to α or $\boldsymbol{\beta}$. This makes it difficult to maximize the likelihood function (8) by a gradient method. Therefore we propose to analyze the TPNRM in a Bayesian approach instead.

3 Bayesian Inference

Before we proceed to a Bayesian inference about the TPNRM, we briefly describe how to conduct a Bayesian inference in a general setting. Let $\boldsymbol{\theta}$ a vector of unknown parameters and $\boldsymbol{y}$

observed data. We suppose that $\boldsymbol{y}$ is generated from the joint distribution $p(\boldsymbol{y}|\boldsymbol{\theta})$ which is the likelihood function in terms of $\boldsymbol{\theta}$. We also suppose that prior information about the parameters $\boldsymbol{\theta}$ is summarized in the form of the prior distribution $p(\boldsymbol{\theta})$. Then we construct the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ from the prior distribution $p(\boldsymbol{\theta})$ and the likelihood function $p(\boldsymbol{y}|\boldsymbol{\theta})$ by the Bayes theorem:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_{\Omega_\theta} p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}), \tag{9}$$

where $\Omega_{\boldsymbol{\theta}}$ is the domain of $\boldsymbol{\theta}$. The posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ is the conditional distribution of unknown parameters $\boldsymbol{\theta}$ given observed data $\boldsymbol{y}$. In a Bayesian approach, all inferences about the parameters $\boldsymbol{\theta}$ are based on the posterior distribution. If we want to know a "point estimate" of $\boldsymbol{\theta}$, we may evaluate the expected value of the posterior distribution $\mathrm{E}(\boldsymbol{\theta}|\boldsymbol{y}) = \int_{\Omega_\theta} \boldsymbol{\theta} p(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}$. If our purpose of the analysis is to determine the probability $\Pr(\boldsymbol{\theta} \in \mathcal{R}_{\boldsymbol{\theta}})$ where $\mathcal{R}_{\boldsymbol{\theta}}$ is some region of $\boldsymbol{\theta}$, the probability is computed as $\Pr(\boldsymbol{\theta} \in \mathcal{R}_{\boldsymbol{\theta}}) = \int_{\mathcal{R}_\theta} p(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}$.

Let us derive the posterior distribution for the TPNRM. First, we rewrite the likelihood function as

$$p(\boldsymbol{y}|\alpha, \boldsymbol{\beta}, \omega, \gamma, \boldsymbol{X}) \propto \omega^{-n} \exp\left[-\frac{1}{2\omega^2}\left\{\frac{\sum_{d_i=1} e_i^2}{\gamma^2} + \frac{\sum_{d_i=0} e_i^2}{(1-\gamma)^2}\right\}\right], \tag{10}$$

where $e_i = y_i - \alpha - \boldsymbol{x}_i'\boldsymbol{\beta}$. As the prior distribution, we consider the following conjugate-type prior:

$$\begin{aligned} p(\alpha, \boldsymbol{\beta}, \omega, \gamma) &\propto p(\alpha|\omega)p(\boldsymbol{\beta}|\omega)p(\omega)p(\gamma) \\ &\propto \exp\left[-\frac{\tau_\alpha(\alpha-\mu_\alpha)^2}{2\omega^2}\right] \times \exp\left[-\frac{1}{2\omega^2}(\boldsymbol{\beta}-\boldsymbol{\mu}_{\boldsymbol{\beta}})'\boldsymbol{T}_{\boldsymbol{\beta}}(\boldsymbol{\beta}-\boldsymbol{\mu}_{\boldsymbol{\beta}})\right] \\ &\quad \times \omega^{-(a_\omega+1)} \exp\left(-\frac{b_\omega}{2\omega^2}\right) \mathbf{1}_{(\omega>0)} \times \mathbf{1}_{(0<\gamma<1)}, \end{aligned} \tag{11}$$

where $(\mu_\alpha, \tau_\alpha, \boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{T}_{\boldsymbol{\beta}}, a_\omega, b_\omega)$ are hyper-parameters. In principle, researchers can choose any values for the hyper-parameters as long as they represent their prior information on parameters in the model. Sometimes we want to make the prior distribution "less informative" in the sense that information from data $\boldsymbol{y}$ dominates the prior information. In such a case, we may set $\tau_\alpha \to 0$, $\boldsymbol{T}_{\boldsymbol{\beta}} \to 0$, $a_\omega \to 0$, and $b_\omega \to 0$, which leads to

$$p(\alpha, \boldsymbol{\beta}, \omega, \gamma) \propto \omega^{-1}\mathbf{1}_{(\omega>0)}\mathbf{1}_{(0<\gamma<1)}. \tag{12}$$

This is a kind of non-informative prior, and often used in a Bayesian inference. Since this non-informative prior is improper (the integration is not equal to unity), the posterior distribution would be improper, too. For this reason, we stick to the proper prior (11) in out study. By the Bayes theorem the posterior distribution is given as

$$\begin{aligned} p(\alpha,\boldsymbol{\beta},\omega,\gamma|D) &\propto p(\boldsymbol{y}|\alpha,\boldsymbol{\beta},\omega,\gamma,\boldsymbol{X})p(\alpha,\boldsymbol{\beta},\omega,\gamma) \\ &\propto \omega^{-(n+a_\omega+1)}\exp\left(-\frac{\bar{S}^2+b_\omega}{2\omega^2}\right)\mathbf{1}_{(\omega>0)}\mathbf{1}_{(0<\gamma<1)}, \end{aligned} \tag{13}$$

where $D=(\boldsymbol{y},\boldsymbol{X})$ and

$$\bar{S}^2=\frac{\sum_{d_i=1}e_i^2}{\gamma^2}+\frac{\sum_{d_i=0}e_i^2}{(1-\gamma)^2}+\tau_\alpha(\alpha-\mu_\alpha)^2+(\boldsymbol{\beta}-\boldsymbol{\mu}_{\boldsymbol{\beta}})'\boldsymbol{T}_{\boldsymbol{\beta}}(\boldsymbol{\beta}-\boldsymbol{\mu}_{\boldsymbol{\beta}}).$$

In a Bayesian inference, we are required to evaluate multiple integrals such as $\int_{\Omega_{\boldsymbol{\theta}}}\boldsymbol{\theta}p(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}$ and $\int_{\mathcal{R}_{\boldsymbol{\theta}}}p(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}$. However, there are no closed-form expressions of these integrals available for the posterior distribution (13). Instead, we try to evaluate them by a Monte Carlo integration method. Let $f(\theta)$ a function we want to integrate. For example, to evaluate the expected value of the posterior distribution, we set $f(\boldsymbol{\theta})=\theta$. Then a Monte Carlo integration method approximates an integral of $f(\boldsymbol{\theta})$ with its sample analog, i.e.,

$$\int_{\Omega_{\boldsymbol{\theta}}}f(\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}\approx\frac{1}{M}\sum_{r=1}^{M}f(\boldsymbol{\theta}^{(r)}), \tag{14}$$

where $\boldsymbol{\theta}^{(r)}$ $(r=1,\ldots,M)$ is a sample of $\boldsymbol{\theta}$ generated from the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$. Under regularity conditions, the sample analog will converges to the true integral by the law of large numbers as $M\to\infty$. To compute the sample analog, we need to generate a sample of parameters from the posterior distribution. In our study, we apply a Markov chain sampling method, in particular a Gibbs sampler, to generate them.

To begin with, we explain a Gibbs sampler for a general distribution. Suppose that $\boldsymbol{\theta}$ is a vector of m random variables and $g(\boldsymbol{\theta})$ is the joint pdf of $\boldsymbol{\theta}$. This is the target distribution from which we want to draw $\boldsymbol{\theta}$. $g(\theta_j|\boldsymbol{\theta}_{-j})$ denotes the conditional distribution of θ_j given a set of the other variables $\boldsymbol{\theta}_{-j}=(\theta_1,\ldots,\theta_{j-1},\theta_{j+1},\ldots,\theta_m)$. Then a Gibbs sampler is implemented as follows:

Step 1. set the starting values, $\theta_1^{(0)},\ldots,\theta_m^{(0)}$.

Step 2. draw θ_j from its conditional distribution for $r=1,2,3,\ldots$

$$\begin{aligned} \theta_1^{(r)} &\leftarrow g(\theta_1|\theta_2^{(r-1)},\theta_3^{(r-1)}\ldots,\theta_m^{(r-1)}), \\ \theta_2^{(r)} &\leftarrow g(\theta_2|\theta_1^{(r)},\theta_3^{(r-1)},\ldots,\theta_m^{(r-1)}), \\ &\vdots \\ \theta_m^{(r)} &\leftarrow g(\theta_m|\theta_1^{(r)},\theta_2^{(r)}\ldots,\theta_{m-1}^{(r)}). \end{aligned}$$

Step 3. repeat **Step 2.** until the draws become stable.

Under regularity conditions, the joint distribution of $(\theta_1^{(r)}, \ldots, \theta_m^{(r)})$ will converge to $g(\theta)$.

In the case of the TPNRM, the parameter vector $\boldsymbol{\theta}$ is $[\alpha, \boldsymbol{\beta}', \omega, \gamma]'$ and the target distribution is the posterior distribution $p(\alpha, \boldsymbol{\beta}, \omega, \gamma | D)$. In our study, the order of drawings is $\omega \to \gamma \to \alpha \to \boldsymbol{\beta}$. Thus the Gibbs sampler for the TPNRM are given as follows:

Step 1. set the starting values, $\alpha^{(0)}$, $\boldsymbol{\beta}^{(0)}$, $\omega^{(0)}$, and $\gamma^{(0)}$.

Step 2. draw parameters from their conditional posterior distributions for $r = 1, 2, 3, \ldots$

$$\begin{aligned}
\omega^{(r)} &\leftarrow p(\omega | \alpha^{(r-1)}, \boldsymbol{\beta}^{(r-1)}, \gamma^{(r-1)}, D), \\
\gamma^{(r)} &\leftarrow p(\gamma | \alpha^{(r-1)}, \boldsymbol{\beta}^{(r-1)}, \omega^{(r)}, D), \\
\alpha^{(r)} &\leftarrow p(\alpha | \boldsymbol{\beta}^{(r-1)}, \omega^{(r)}, \gamma^{(r)}, D), \\
\boldsymbol{\beta}^{(r)} &\leftarrow p(\boldsymbol{\beta} | \alpha^{(r)}, \omega^{(r)}, \gamma^{(r)}, D).
\end{aligned}$$

Step 3. repeat **Step 2.** until the draws become stable.

In Appendix B, we discuss how to draw parameters from their conditional posterior distributions in details.

One thing we must mention here is that we cannot simultaneously draw β from its conditional posterior distribution. We may draw β_1 to β_k once at a time, but it is well-known in the literature that drawing regression coefficients one by one from their conditional posterior distributions tends to make a Gibbs sampler converge slower and it is preferable to avoid such a approach if possible. Instead, we apply a Hit-and-Run (H&R) algorithm to draw $\boldsymbol{\beta}$. See Chen et al. (2000) among others for more information about the H&R algorithm. A H&R algorithm for $\boldsymbol{\beta}$ is as follows:

Step 1. set the starting value of $\boldsymbol{\beta}$, $\boldsymbol{\beta}^{(0)}$.

Step 2. generate a $k \times 1$ vector $\boldsymbol{\xi}^{(r)} = [\xi_1^{(r)}, \ldots, \xi_k^{(r)}]'$ from a distribution on the surface of k-dimensional unit sphere. For example, we may generate $\boldsymbol{\xi}^{(r)}$ by

$$\xi_j^{(r)} = z_j \left(\sum_{h=1}^{k} z_h^2 \right)^{-1/2}, \quad z_j \sim \text{i.i.d.}\, \mathcal{N}(0,1), \quad (j, h = 1, \ldots, k). \tag{15}$$

Step 3. generate a random scale $\lambda^{(r)}$ from a distribution $g(\lambda | \boldsymbol{\beta}^{(r)}, \boldsymbol{\xi}^{(r)})$.

Step 4. set $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(r)} + \lambda^{(r)} \boldsymbol{\xi}^{(r)}$.

Step 5. set

$$\boldsymbol{\beta}^{(r+1)} = \begin{cases} \tilde{\boldsymbol{\beta}} & \text{with probability } a(\tilde{\boldsymbol{\beta}}|\boldsymbol{\beta}^{(r)}); \\ \boldsymbol{\beta}^{(r)} & \text{otherwise.} \end{cases}$$

Obviously, a H&R algorithm is a special case of the Metropolis-Hastings algorithm, and using a H&R algorithm inside a Gibbs sampler makes it a so-called hybrid MCMC or Metropolis-within-Gibbs algorithm. The convergence to the posterior distribution is still guaranteed for a hybrid MCMC algorithm.

One convenient property of the H&R algorithm is that we can set the acceptance probability $a(\tilde{\boldsymbol{\beta}}|\boldsymbol{\beta}^{(r)})$ equal to one if we use

$$g(\lambda|\boldsymbol{\beta}^{(r)}, \boldsymbol{\xi}^{(r)}) = \frac{p(\boldsymbol{\beta}^{(r)} + \lambda\boldsymbol{\xi}^{(r)}|\alpha, \omega, \gamma, D)}{\int_{\Omega_\lambda} p(\boldsymbol{\beta}^{(r)} + \zeta\boldsymbol{\xi}^{(r)}|\alpha, \omega, \gamma, D)d\zeta}, \tag{16}$$

where Ω_λ is the domain of λ given $(\boldsymbol{\beta}^{(r)}, \boldsymbol{\xi}^{(r)})$. We use this distribution of λ (16) in out study. Derivation of (16) is also explained in Appendix B.

4 Applications

4.1 Simulated data

We consider a TPNRM:

$$y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i, \quad \epsilon_i \sim \text{i.i.d. } \mathcal{TPN}(0, \omega, \gamma).$$

We set $\alpha = \beta_1 = \beta_2 = \beta_3 = 1$, $\gamma = 0.2$, and $\omega = 2/\sqrt{\gamma(1-\gamma) + (1-2\pi^{-1})(1-2\gamma)^2} \approx 3.7087$. We generate $x_i = [x_{i1}, x_{i2}, x_{i3}]'$ from

$$\begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \end{bmatrix} \sim \text{i.i.d. } \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.9 & 0.9 \\ 0.9 & 1 & 0.9 \\ 0.9 & 0.9 & 1 \end{bmatrix} \right).$$

The number of observation is $n = 200$. The hyper-parameters are as follows: $\mu_\alpha = 0$, $\tau_\alpha = 0.001$, $\boldsymbol{\mu}_{\boldsymbol{\beta}} = [0, 0, 0]'$, $\boldsymbol{T}_{\boldsymbol{\beta}} = 0.001 \times I_3$, $a_\omega = 0.001$, and $b_\omega = 0.001$. In the Gibbs sampler, we first generate $2,000$ draws and discard them as burn-in. Then we generate $10,000$ draws and use them for a Bayesian inference of the TPNRM. The results are shown in Table 1. The marginal posterior distribution of γ is plotted in Figure 3.

Apparently, the OLSE's of the parameters are biased except for β_2. The marginal posterior distribution of γ is tightly distributed around 0.2, the true value of γ.

4.2 Cost function of the electric utility industry

Greene (1990) provides data for estimation of a cost function of the U.S. electric utility industry. Greene (1990) uses the following specification of a cost function:

$$\ln(\text{Cost}/P_f) = \alpha + \beta_1 \ln Q + \beta_2 \ln^2 Q + \beta_3 \ln(P_\ell/P_f) + \beta_4 \ln(P_\kappa/P_f) + \epsilon, \tag{17}$$

where Q is the output, P_ℓ, P_κ, and P_f are the prices of labor, capital, and fuel respectively. To stabilize the Gibbs sampler, we use standardized data for estimation. Then generated draws of parameters are transformed back into the original parameterization. The Gibbs sampler is implemented in the same manner as the previous example. The results are reported in Table 2. The marginal posterior distribution of γ is plotted in Figure 4.

The OLSE's and posterior means or medians of the parameters are not so much different. The marginal posterior distribution of γ is distributed around 0.5. The probability of $\gamma < 1/2$ is about 62%. Thus we cannot support the assumption that there is some inefficiency in the cost function of the electric utility industry. Of course, this result relies on the model specification. So it may suggest that we might as well reconsider the specification of the cost function.

4.3 Production function of the transportation equipment industry

Zellner and Revankar (1969) provides data for estimation of a production function of the U.S. transportation equipment industry. Greene (2000, p.396) uses the data to estimate a stochastic production frontier model. The production function is Cobb-Douglas[1]:

$$\ln V = \alpha + \beta_1 \ln K + \beta_2 \ln L + \epsilon, \tag{18}$$

where V is the aggregated value added, K is the aggregated capital service flow, and L is the aggregated person-hours worked. The Gibbs sampler is implemented in the same manner. The results are reported in Table 3. The marginal posterior distribution of γ is plotted in Figure 5.

In Table 3, the estimates of the constant term α and the coefficient for capital β_1 is slightly larger in the TPNRM than in the classical normal regression model. The estimated coefficient for labor β_2 is slightly smaller in the TPNRM. This pattern also appears for the half-normal and exponential stochastic frontier models in Table 9.1 of Greene (2000, p.396). The marginal posterior distribution of γ is distributed around 0.5, and the probability of $\gamma > 1/2$ is about 66%.

[1] Zellner and Revankar (1969) proposed an alternative functional form for a production function other than Cobb-Douglas. But we use a Cobb-Douglas function as Greene (2000) did since our purpose is merely to illustrate how the TPN assumption makes differences on the estimates

Thus it is ambiguous whether there are any inefficiencies in the production of transportation equipment or we need to redo the model specification.

5 Concluding Remarks

In this paper, we developed a Markov chain Monte Carlo method for Bayesian inference about a two-piece normal regression model. Our method is based on a Gibbs sampler and uses a Hit-and-Run algorithm to generate regression coefficients. For the purpose of illustration, a cost function of the electric utility industry and a production function of the transportation equipment industry as well as a regression model of simulated data are estimated.

Appendix

A Statistical Properties of a TPN Distribution

John (1982) and Kimber (1985) derived the following properties of a TPN distribution:

1. $\Pr(\epsilon \leq \alpha) = \gamma$ and $\Pr(\epsilon > \alpha) = 1 - \gamma$.
2. $\mathrm{mode}(\epsilon) = \alpha$.
3. $\mathrm{E}[(\epsilon - \alpha)^r] = \pi^{-1/2}\Gamma[(r+1)/2](\sqrt{2}\omega)^r\left[(1-\gamma)^{r+1} - (-\gamma)^{r+1}\right]$.
4. $\mathrm{E}(\epsilon) = \alpha + \sqrt{2\pi^{-1}}\omega(1-2\gamma)$.
5. $\mathrm{Var}(\epsilon) = \omega^2\left[\gamma(1-\gamma) + (1-2\pi^{-1})(1-2\gamma)^2\right]$.
6. $\mathrm{median}(\epsilon) = \alpha + \gamma\omega\Phi^{-1}[1/(4\gamma)]$ if $\gamma \geq 1/2$; $\alpha - (1-\gamma)\omega\Phi^{-1}[1/\{4(1-\gamma)\}]$ otherwise.
7. The TPN distribution is symmetric around α when $\gamma = 1/2$, skewed to the right when $\gamma < 1/2$ and skewed to the left when $\gamma > 1/2$.
8. $\lim_{\gamma\to 1}\sqrt{\beta_1} = -\lim_{\gamma\to 0}\sqrt{\beta_1} = \sqrt{2}(4-\pi)(\pi-2)^{-3/2}$. ($\sqrt{\beta_1}$ is the skewness)
9. $0 \leq \beta_2 - 3 < 8(\pi-3)(\pi-2)^{-2}$. ($\beta_2$ is the kurtosis)

As Kimber (1985) points out, a TPN distribution is regarded as a mixture of two half normal distributions,

$$\begin{aligned} f(u) &= \Pr(\epsilon \leq \alpha) f(\epsilon|\epsilon \leq \alpha) + \Pr(\epsilon > \alpha) f(\epsilon|\epsilon > \alpha) \\ &= \gamma\left[\frac{2}{\gamma\omega}\phi\left(\frac{\epsilon-\alpha}{\gamma\omega}\right)\mathbf{1}_{(\epsilon\leq\alpha)}\right] + (1-\gamma)\left[\frac{2}{(1-\gamma)\omega}\phi\left(\frac{\epsilon-\alpha}{(1-\gamma)\omega}\right)\mathbf{1}_{(\epsilon>\alpha)}\right], \end{aligned}$$

and TPN random numbers can be generated by the following procedure:

Step 1. generate $z \sim \mathcal{N}(0,1)$.

Step 2. generate $u \sim \mathcal{U}(0,1)$.

Step 3. draw ϵ from

$$\epsilon = \begin{cases} \alpha - \gamma\omega|z| & \text{if } u \leq \gamma, \\ \alpha + (1-\gamma)\omega|z| & \text{if } u > \gamma. \end{cases}$$

B Posterior Simulation of Parameters in the TPNRM

B.1 Conditional posterior distribution of ω

Given α, $\boldsymbol{\beta}$, and γ, the conditional posterior density of ω is

$$p(\omega|\alpha,\boldsymbol{\beta},\gamma,D) \propto \omega^{-(n+a_\omega+1)} \exp\left(-\frac{\bar{S}^2+b_\omega}{2\omega^2}\right)\mathbf{1}_{(\omega>0)}. \tag{19}$$

This is the kernel of the square-root inverted gamma distribution,

$$\omega|\alpha,\boldsymbol{\beta},\gamma,D \sim \mathcal{G}a^{-\frac{1}{2}}\left(\frac{n+a_\omega}{2},\frac{\bar{S}^2+b_\omega}{2}\right). \tag{20}$$

The mean and variance of (19) are

$$\mathrm{E}(\omega|\alpha,\boldsymbol{\beta},\gamma,D) = \frac{\Gamma[(n+a_\omega-1)/2]}{\Gamma[(n+a_\omega)/2]}\sqrt{\frac{\bar{S}^2+b_\omega}{2}}, \tag{21}$$

$$\mathrm{Var}(\omega|\alpha,\boldsymbol{\beta},\gamma,D) = \frac{\bar{S}^2+b_\omega}{n+a_\omega-2} - \mathrm{E}(\omega|\alpha,\boldsymbol{\beta},\gamma,D)^2. \tag{22}$$

See Bernardo and Smith (1994, p.119, p.431) for other properties of the square-root inverted gamma distribution.

B.2 Conditional posterior distribution of γ

Given α, $\boldsymbol{\beta}$, and ω, the conditional posterior density of γ is

$$p(\gamma|\alpha,\boldsymbol{\beta},\omega,D) \propto \exp\left[-\frac{1}{2\omega^2}\left\{\frac{\sum_{d_i=1} e_i^2}{\gamma^2} + \frac{\sum_{d_i=0} e_i^2}{(1-\gamma)^2}\right\}\right]\mathbf{1}_{(0<\gamma<1)}, \tag{23}$$

which is not a standard distribution. It has the following properties:

1. $\lim_{\gamma\to 0} p(\gamma|\alpha,\boldsymbol{\beta},\omega,D) = 0$ if $\sum_{d_i=1} e_i^2 \neq 0$.

2. $\lim_{\gamma\to 1} p(\gamma|\alpha,\boldsymbol{\beta},\omega,D) = 0$ if $\sum_{d_i=0} e_i^2 \neq 0$.

3. $p(\gamma|\alpha, \boldsymbol{\beta}, \omega, D)$ is unimodal if $\sum_{d_i=1} e_i^2 \neq 0$ and $\sum_{d_i=0} e_i^2 \neq 0$, and

$$\text{mode}(\gamma) = \left[1 + \left(\frac{\sum_{d_i=0} e_i^2}{\sum_{d_i=1} e_i^2}\right)^{1/3}\right]^{-1}.$$

Since the domain of γ is bounded $(0 < \gamma < 1)$, we may apply a griddy Gibbs sampler to drawing ω from its conditional posterior distribution. In our study, the following griddy Gibbs sampler is used to draw γ from its conditional conditional distribution:

Step 1. compute $p(\gamma|\alpha, \boldsymbol{\beta}, \omega, D)$ over the grid, $0 = \gamma_0 < \gamma_1 < \cdots < \gamma_{G-1} < \gamma_G = 1$.

Step 2. compute $F_g = \int_{\gamma_{g-1}}^{\gamma_g} p(\gamma|\alpha, \boldsymbol{\beta}, \omega, D) d\gamma$ $(g = 1, \ldots, G)$ with the trapezoidal method. (We set $F_0 = 0$.)

Step 3. generate $u \sim \mathcal{U}(0, F_G)$ and find g such that $F_{g-1} < u \leq F_g$.

Step 4. invert u into γ by the linear interpolation,

$$\gamma = \gamma_{g-1} + \frac{u - F_{g-1}}{F_g - F_{g-1}}(\gamma_g - \gamma_{g-1}).$$

B.3 Conditional posterior distribution of α

Let us consider $u_i = y_i - x_i'\boldsymbol{\beta}$ $(i = 1, \ldots, n)$ and their order statistics, $u_{(1)} < u_{(2)} < \cdots < u_{(n)}$. Without the loss of generality, we assume that $\{u_{(1)}, \ldots, u_{(n)}\}$ are all distinct. We also assume $u_{(1)} < \alpha \leq u_{(n)}$ since otherwise γ and ω cannot be identified. Within the interval $u_{(h)} < \alpha \leq u_{(h+1)}$ $(h = 1, \ldots, n-1)$, the conditional posterior density of α is continuous and expressed as

$$\begin{aligned} & p(\alpha|u_{(h)} < \alpha \leq u_{(h+1)}, \boldsymbol{\beta}, \omega, \gamma, D) \\ & \quad \propto \exp\left[-\frac{1}{2\omega^2}\left\{\frac{\sum_{i=1}^h (u_{(i)} - \alpha)^2}{\gamma^2} + \frac{\sum_{i=h+1}^n (u_{(i)} - \alpha)^2}{(1-\gamma)^2} + \tau_\alpha(\alpha - \mu_\alpha)^2\right\}\right] \mathbf{1}_{(u_{(h)} < \alpha \leq u_{(h+1)})} \\ & \quad \propto \exp\left[-\frac{(\alpha - \bar{\alpha}_h)^2}{2\bar{\sigma}_{\alpha h}^2}\right] \mathbf{1}_{(u_{(h)} < \alpha \leq u_{(h+1)})}, \end{aligned} \tag{24}$$

where

$$\begin{aligned} \bar{\alpha}_h &= \frac{\gamma^{-2}\sum_{i=1}^h u_{(i)} + (1-\gamma)^{-2}\sum_{i=h+1}^n u_{(i)} + \tau_\alpha \mu_\alpha}{\gamma^{-2} h + (1-\gamma)^{-2}(n-h) + \tau_\alpha}, \\ \bar{\sigma}_{\alpha h}^2 &= \frac{\omega^2}{\gamma^{-2} h + (1-\gamma)^{-2}(n-h) + \tau_\alpha}. \end{aligned}$$

(24) is the kernel of a doubly truncated normal distribution with truncating points $u_{(h)}$ and $u_{(h+1)}$. Therefore the conditional posterior density of α is

$$p(\alpha|\boldsymbol{\beta},\omega,\gamma,D)=\begin{cases}\dfrac{1}{K_\alpha\bar{\sigma}_{\alpha 1}}\phi\left(\dfrac{\alpha-\bar{\alpha}_1}{\bar{\sigma}_{\alpha 1}}\right) & \text{for } u_{(1)}<\alpha\le u_{(2)},\\ \vdots & \vdots\\ \dfrac{1}{K_\alpha\bar{\sigma}_{\alpha,n-1}}\phi\left(\dfrac{\alpha-\bar{\alpha}_{n-1}}{\bar{\sigma}_{\alpha,n-1}}\right) & \text{for } u_{(n-1)}<\alpha\le u_{(n)},\end{cases} \tag{25}$$

where K_α is the normalizing constant and defined as

$$K_\alpha=\sum_{h=1}^{n-1}\left\{\Phi(B_{\alpha h})-\Phi(A_{\alpha h})\right\},\quad A_{\alpha h}=\frac{u_{(h)}-\bar{\alpha}_h}{\bar{\sigma}_{\alpha h}},\quad B_{\alpha h}=\frac{u_{(h+1)}-\bar{\alpha}_h}{\bar{\sigma}_{\alpha h}}.$$

The conditional posterior distribution of α (25) is regarded as a mixture of the doubly truncated normal distributions given in (24), and in general it is discontinuous at each $u_{(h)}$ $(h=2,\dots,n-1)$. The mixture rate for each component in (25) is

$$\psi_{\alpha h}\equiv\Pr(u_{(h)}<\alpha\le u_{(h+1)}|\boldsymbol{\beta},\omega,\gamma,D)=\frac{\Phi(B_{\alpha h})-\Phi(A_{\alpha h})}{K_\alpha}. \tag{26}$$

Therefore we can draw α from its conditional posterior distribution (25) in the following manner:

Step 1. choose an interval $u_{(h)}<\alpha\le u_{(h+1)}$ with probability $\psi_{\alpha h}$ in (26).

Step 2. draw α from (24) corresponding to the chosen interval.

B.4 Hit-and-Run algorithm for $\boldsymbol{\beta}$

We draw $\boldsymbol{\beta}$ by a H&R algorithm. First, let us derive the distribution of λ (16) for the TPNRM. To make mathematical expressions concise, we will suppress the superscript '(r)' in the following derivation. Plugging $\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}+\lambda\boldsymbol{\xi}$ into the posterior density, we have

$$\begin{aligned}&p(\boldsymbol{\beta}+\lambda\boldsymbol{\xi}|\omega,\gamma,D)\\&\quad\propto\exp\left[-\frac{1}{2\omega^2}\left\{\frac{\sum_{d_i=1}\{y_i-\alpha-\boldsymbol{x}_i'(\boldsymbol{\beta}+\lambda\boldsymbol{\xi})\}^2}{\gamma^2}+\frac{\sum_{d_i=0}\{y_i-\alpha-\boldsymbol{x}_i'(\boldsymbol{\beta}+\lambda\boldsymbol{\xi})\}^2}{(1-\gamma)^2}\right\}\right]\\&\quad\times\exp\left[-\frac{1}{2\omega^2}(\boldsymbol{\beta}+\lambda\boldsymbol{\xi}-\boldsymbol{\mu}_\beta)'\boldsymbol{T}_\beta(\boldsymbol{\beta}+\lambda\boldsymbol{\xi}-\boldsymbol{\mu}_\beta)\right]\\&\quad\propto\exp\left[-\frac{1}{2\omega^2}\left\{\frac{\sum_{d_i=1}\{e_i-\lambda\eta_i\}^2}{\gamma^2}+\frac{\sum_{d_i=0}\{e_i-\lambda\eta_i\}^2}{(1-\gamma)^2}+\frac{(\lambda-\mu_\lambda)^2}{\tau_\lambda^2}\right\}\right],\end{aligned} \tag{27}$$

where $\eta_i=\boldsymbol{x}_i'\boldsymbol{\xi}$, $\mu_\lambda=\boldsymbol{\xi}'\boldsymbol{T}_\beta\boldsymbol{\mu}_\beta$, $\tau_\lambda=\boldsymbol{\xi}'\boldsymbol{T}_\beta\boldsymbol{\xi}$, and the dummy variable d_i is determined by

$$d_i=\begin{cases}1, & \text{if } y_i\le\boldsymbol{x}_i'\tilde{\boldsymbol{\beta}}\Leftrightarrow e_i\le\lambda\eta_i,\\0, & \text{if } y_i>\boldsymbol{x}_i'\tilde{\boldsymbol{\beta}}\Leftrightarrow e_i>\lambda\eta_i.\end{cases}$$

Let $v_i = e_i/\eta_i$ $(i = 1, \ldots, n)$ and its order statistics $v_{(1)} < \cdots < v_{(n)}$. We assume that $\{v_{(1)}, \ldots, v_{(n)}\}$ are all distinct and $\boldsymbol{x}_i'\boldsymbol{\xi} \neq 0$ for all i. We also assume $v_{(0)} = -\infty$ and $v_{(n+1)} = \infty$. Within the interval $v_{(h-1)} < \lambda \leq v_{(h)}$ $(h = 1, \ldots, n+1)$, the density of $\boldsymbol{\beta} + \lambda\boldsymbol{\xi}$ (27) is continuous with respect to λ. Thus if we regard (27) as the density of λ, it will be expressed as

$$\begin{aligned}
& g(\lambda | v_{(h-1)} < \lambda \leq v_{(h)}, \boldsymbol{\beta}, \boldsymbol{\xi}) \\
& \quad \propto \exp\left[-\frac{1}{2\omega^2}\left\{\frac{\sum_{d_i=1}\eta_i^2(v_i-\lambda)^2}{\gamma^2} + \frac{\sum_{d_i=0}\eta_i^2(v_i-\lambda)^2}{(1-\gamma)^2} + \tau_\lambda(\lambda-\mu_\lambda)^2\right\}\right]\mathbf{1}_{(v_{(h-1)}<\lambda\leq v_{(h)})} \\
& \quad \propto \exp\left[-\frac{(\lambda-\bar{\lambda}_h)^2}{2\bar{\sigma}_{\lambda h}^2}\right]\mathbf{1}_{(v_{(h-1)}<\lambda\leq v_{(h)})},
\end{aligned} \tag{28}$$

where

$$\begin{aligned}
\bar{\lambda}_h &= \frac{\gamma^{-2}\sum_{d_i=1}\eta_i^2 v_i + (1-\gamma)^{-2}\sum_{d_i=0}\eta_i^2 v_i + \tau_\lambda\mu_\lambda}{\gamma^{-2}\sum_{d_i=1}\eta_i^2 + (1-\gamma)^{-2}\sum_{d_i=0}\eta_i^2 + \tau_\lambda}, \\
\bar{\sigma}_{\lambda h}^2 &= \frac{\omega^2}{\gamma^{-2}\sum_{d_i=1}\eta_i^2 + (1-\gamma)^{-2}\sum_{d_i=0}\eta_i^2 + \tau_\lambda}.
\end{aligned}$$

(28) is also the kernel of a truncated normal distribution. Therefore the pdf of λ, $g(\lambda|\boldsymbol{\beta}, \boldsymbol{\xi})$, is given as

$$g(\lambda|\boldsymbol{\beta},\boldsymbol{\xi}) = \begin{cases} \dfrac{1}{K_\lambda\bar{\sigma}_{\lambda 1}}\phi\left(\dfrac{\lambda-\bar{\lambda}_1}{\bar{\sigma}_{\lambda 1}}\right) & \text{for } v_{(0)} < \lambda \leq v_{(1)}, \\ \vdots & \vdots \\ \dfrac{1}{K_\lambda\bar{\sigma}_{\lambda,n+1}}\phi\left(\dfrac{\lambda-\bar{\lambda}_{n+1}}{\bar{\sigma}_{\lambda,n+1}}\right) & \text{for } v_{(n)} < \lambda \leq v_{(n+1)}, \end{cases} \tag{29}$$

where K_λ is the normalizing constant and defined as

$$K_\lambda = \sum_{h=1}^{n+1}\{\Phi(B_{\lambda h}) - \Phi(A_{\lambda h})\}, \quad A_{\lambda h} = \frac{v_{(h-1)}-\bar{\lambda}_h}{\bar{\sigma}_{\lambda h}}, \quad B_{\lambda h} = \frac{v_{(h)}-\bar{\lambda}_h}{\bar{\sigma}_{\lambda h}}.$$

$g(\lambda|\boldsymbol{\beta}, \boldsymbol{\xi})$ in (29) is also regarded as a mixture of truncated normal distributions in (28) with the mixture rate,

$$\psi_{\lambda h} \equiv \Pr(v_{(h-1)} < \lambda \leq v_{(h)}|\boldsymbol{\beta},\boldsymbol{\xi}) = \frac{\Phi(B_{\lambda h}) - \Phi(A_{\lambda h})}{K_\lambda}. \tag{30}$$

Hence we can draw λ from its distribution (29) in the same manner as α. In summary, a H&R algorithm for generation of $\boldsymbol{\beta}$ is implemented as follows:

Step 1. generate a $k \times 1$ vector $\boldsymbol{\xi}$ by (15).

Step 2. choose an interval $v_{(h-1)} < \lambda \leq v_{(h)}$ with probability $\psi_{\lambda h}$ in (30).

Step 3. draw λ from (28) corresponding to the chosen interval.

Step 4. replace $\boldsymbol{\beta}$ with $\boldsymbol{\beta} + \lambda\boldsymbol{\xi}$.

References

[1] Aigner, D. J., T. Amemiya, and D. J. Poirier (1976) "On the estimation of production frontiers: Maximum likelihood estimation of the parameters of a discontinuous density function," *International Economic Review*, 17, 377–396.

[2] Aigner, D., K. Lovell, and P. Schmidt (1977) "Formulation and estimation of stochastic frontier models," *Journal of Econometrics*, 6, 21–37.

[3] Bernardo, J. M., and A. F. M. Smith (1994) *Bayesian Theory*, Wiley.

[4] Broeck, J. van den, G. Koop, J. Osiewalski, and M. F. J. Steel (1994) "Stochastic frontier models: A Bayesian perspective," *Journal of Econometrics*, 61, 273–303.

[5] Chen M.-H., Q.-M. Shao, and J. G. Ibrahim (2000) *Monte Carlo Methods in Bayesian Computation*, Springer-Verlag.

[6] Greene, W. H. (1990) "A gamma-distributed stochastic frontier mode," *Journal of Econometrics*, 46, 141–163.

[7] Greene, W. H. (2000) *Econometric Analysis*, 4th ed., Prentice Hall.

[8] John, S. (1982) "The three parameter two-piece normal family of distributions and its fitting," *Communications in Statistics: Theory and Methods*, 11, 879–885.

[9] Kimber, A. C. (1985) "Methods for the two-piece normal distribution," *Communications in Statistics: Theory and Methods*, 14, 235–245.

[10] Koop, G., J. Osiewalski, and M. F. J. Steel (1997) "Bayesian efficiency analysis through individual effects: Hospital cost frontiers," *Journal of Econometrics*, 76, 77–105.

[11] Robert, C. P., and G. Casella (1999) *Monte Carlo Statistical Methods*, Springer-Verlag.

[12] Stevenson, R. E., (1980) "Likelihood functions for generalized stochastic frontier estimation," *Journal of Econometrics*, 13, 57–66.

[13] Zellner, A., and N. Revankar (1969) "Generalized production functions," *Review of Economic Studies*, 36, 241–250.

Table 1: A Simulated TPNRM

	OLSE	Posterior Statistics			
		Mean	Median	2.5%	97.5%
α	2.7608	1.0066	1.0018	0.5632	1.4862
	(0.1381)	(0.2334)			
β_1	0.3108	1.0180	1.0292	0.4314	1.5597
	(0.3534)	(0.2846)			
β_2	0.8752	0.8520	0.8483	0.4207	1.2862
	(0.3463)	(0.2194)			
β_3	1.8103	1.0716	1.0756	0.4548	1.7056
	(0.3703)	(0.3109)			
ω	—	3.5205	3.5131	3.1915	3.8943
		(0.1819)			
γ	—	0.1795	0.1771	0.1081	0.2645
		(0.0398)			
σ	1.9408	1.9193	1.9138	1.7293	2.1375
		(0.1044)			

Notes: Numbers in parentheses are standard errors for the OLSE and posterior standard deviations for the Bayesian estimates. σ is the standard deviation of the error term.

Table 2: Cost Function

		Posterior Statistics			
	OLSE	Mean	Median	2.5%	97.5%
α	-7.2077	-7.2092	-7.2064	-7.8987	-6.4996
	(0.3372)	(0.3531)			
β_1	0.3858	0.3893	0.3876	0.3135	0.4653
	(0.0383)	(0.0397)			
β_2	0.0316	0.0315	0.0316	0.0264	0.0365
	(0.0027)	(0.0026)			
β_3	0.2470	0.2427	0.2424	0.1124	0.3751
	(0.0670)	(0.0674)			
β_4	0.0784	0.0737	0.0735	-0.0498	0.1962
	(0.0617)	(0.0629)			
ω		0.2922	0.2910	0.2575	0.3331
	—	(0.0194)			
γ		0.4746	0.4736	0.3251	0.6352
	—	(0.0806)			
σ	0.1448	0.1470	0.1465	0.1295	0.1678
		(0.0099)			

Notes: Numbers in parentheses are standard errors for the OLSE and posterior standard deviations for the Bayesian estimates. σ is the standard deviation of the error term.

Table 3: Production Function

	OLSE	Posterior Statistics			
		Mean	Median	2.5%	97.5%
α	1.8444	1.9391	1.9260	1.3273	2.6402
	(0.2336)	(0.3254)			
β_1	0.2454	0.2517	0.2525	0.0337	0.4644
	(0.1069)	(0.1073)			
β_2	0.8052	0.7863	0.7848	0.5296	1.0479
	(0.1263)	(0.1331)			
ω		0.4951	0.4858	0.3669	0.6746
	—	(0.0788)			
γ		0.5534	0.5492	0.3314	0.8078
	—	(0.1202)			
σ	0.2357	0.2514	0.2466	0.1852	0.3446
		(0.0410)			

Notes: Numbers in parentheses are standard errors for the OLSE and posterior standard deviations for the Bayesian estimates. σ is the standard deviation of the error term.

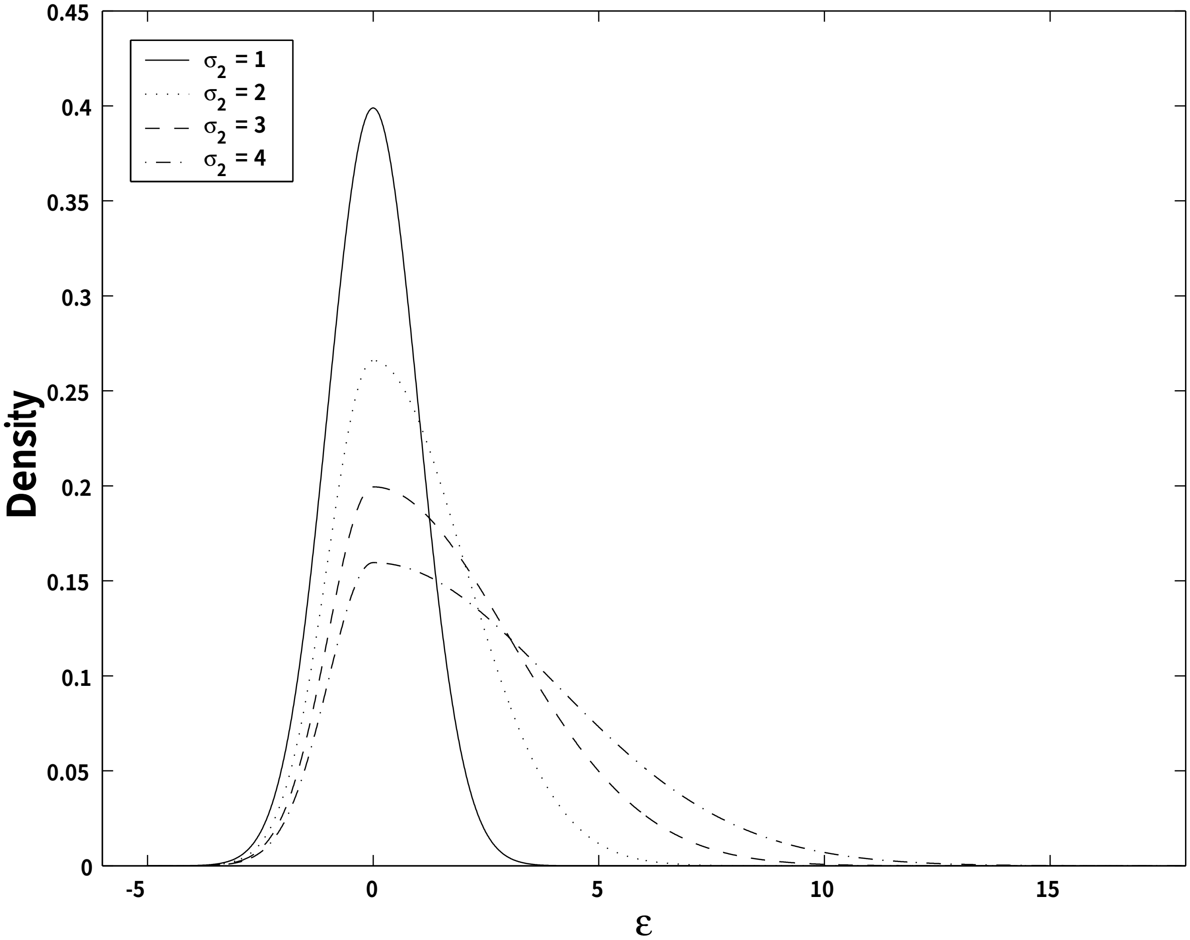

Figure 1: TPN Distributions

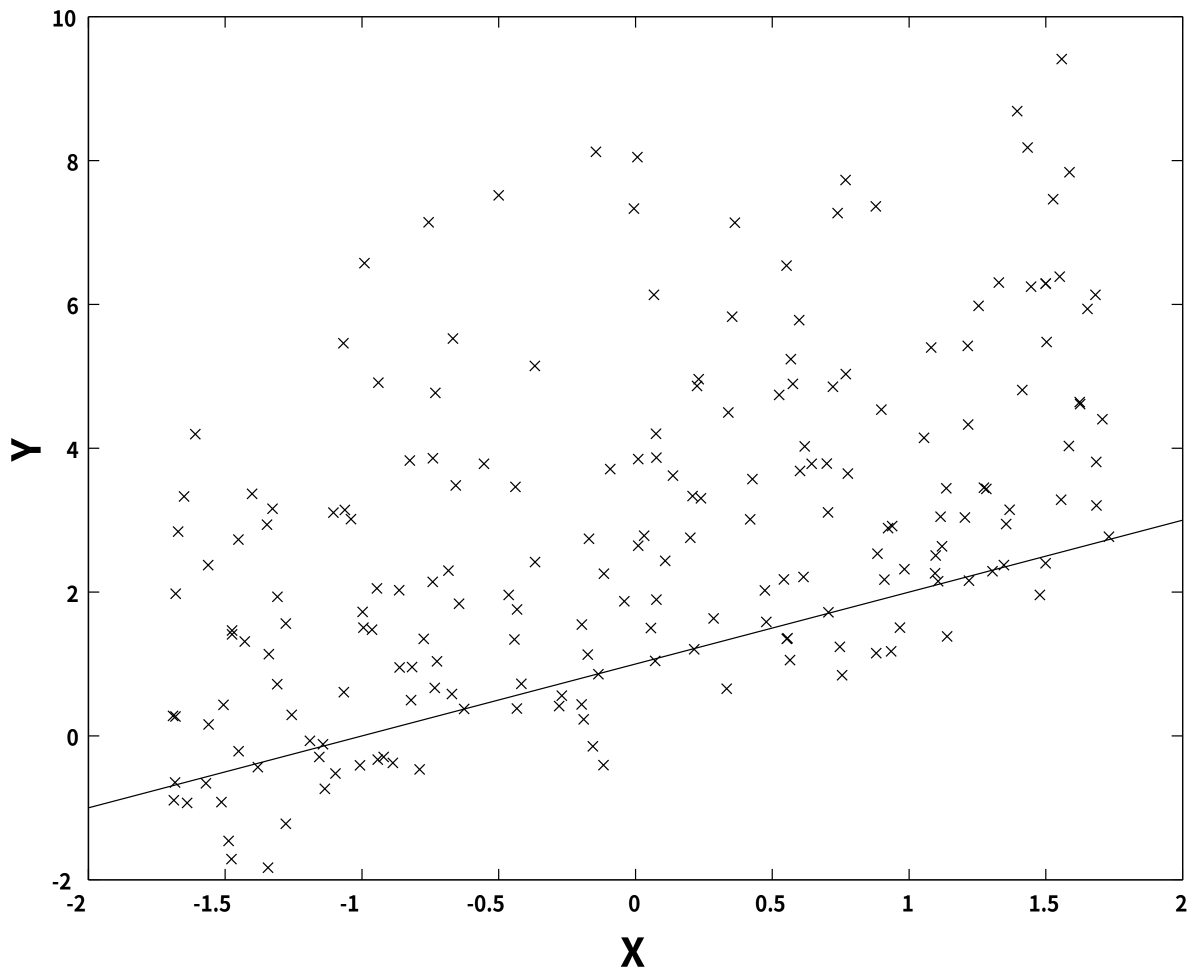

Figure 2: TPN Regression Model: $y_i = 1 + x_i + \epsilon_i$, $\epsilon_i \sim$ TPN(0,3.7087,0.2)

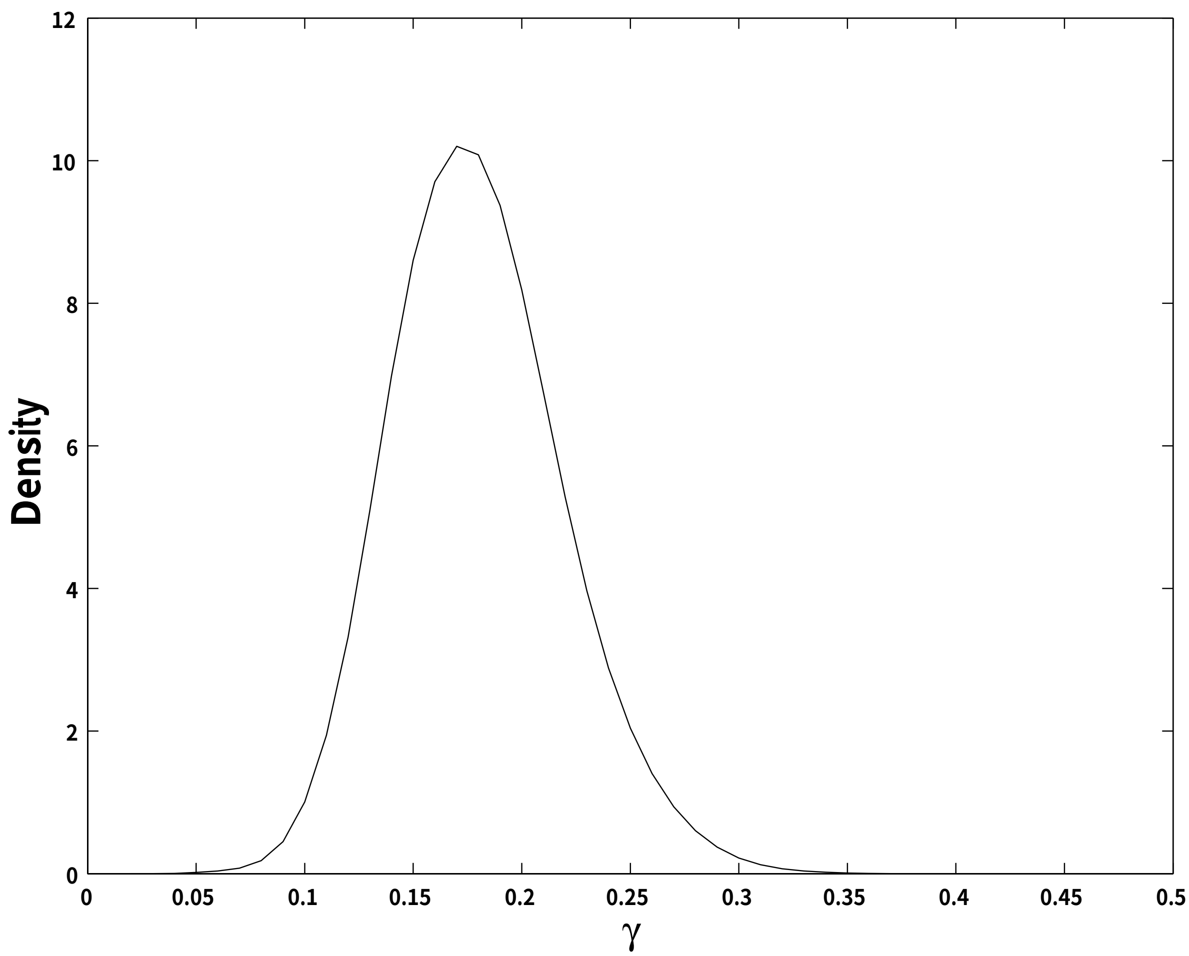

Figure 3: Marginal Posterior Distribution of γ (Simulated Data, $\gamma = 0.2$)

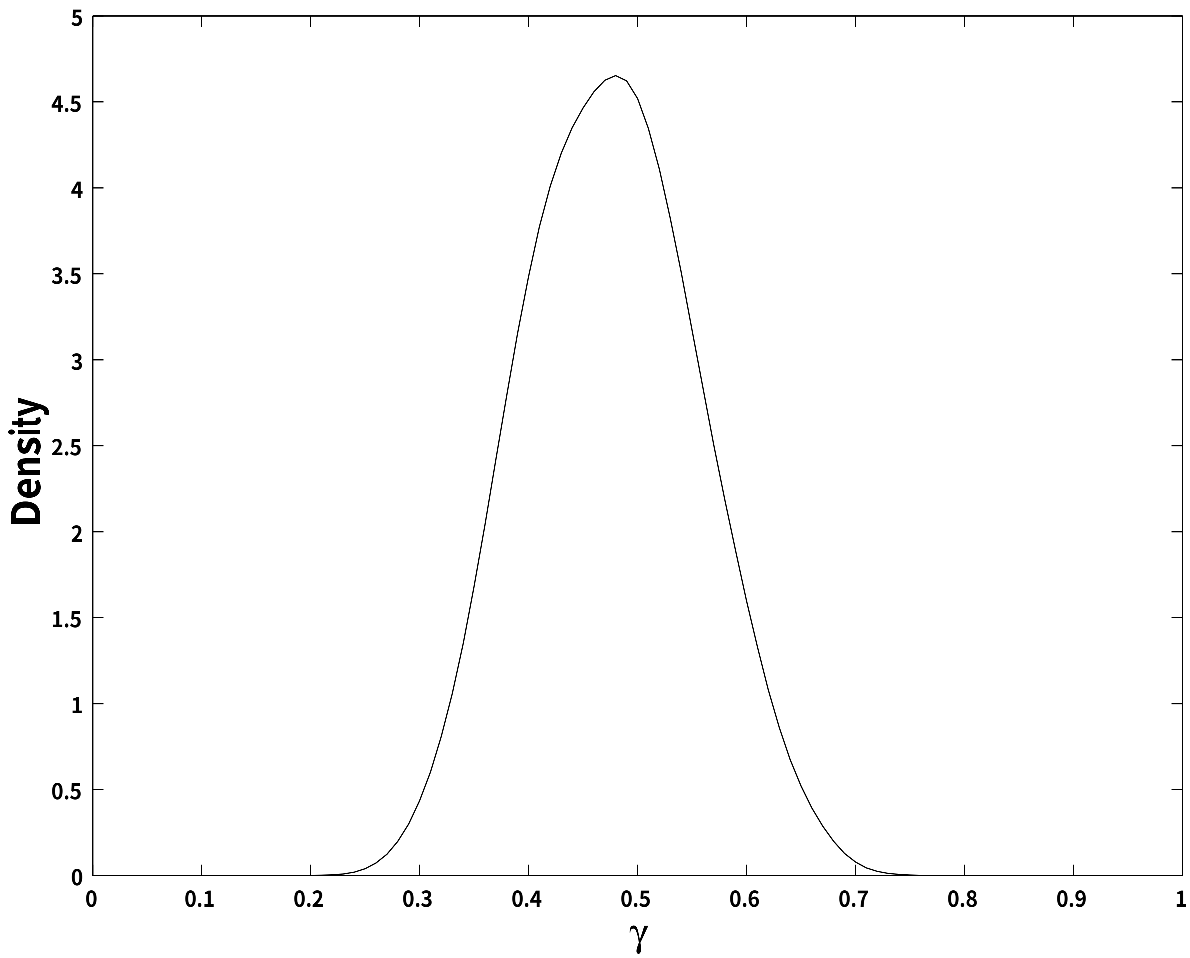

Figure 4: Marginal Posterior Distribution of γ (Cost Function)

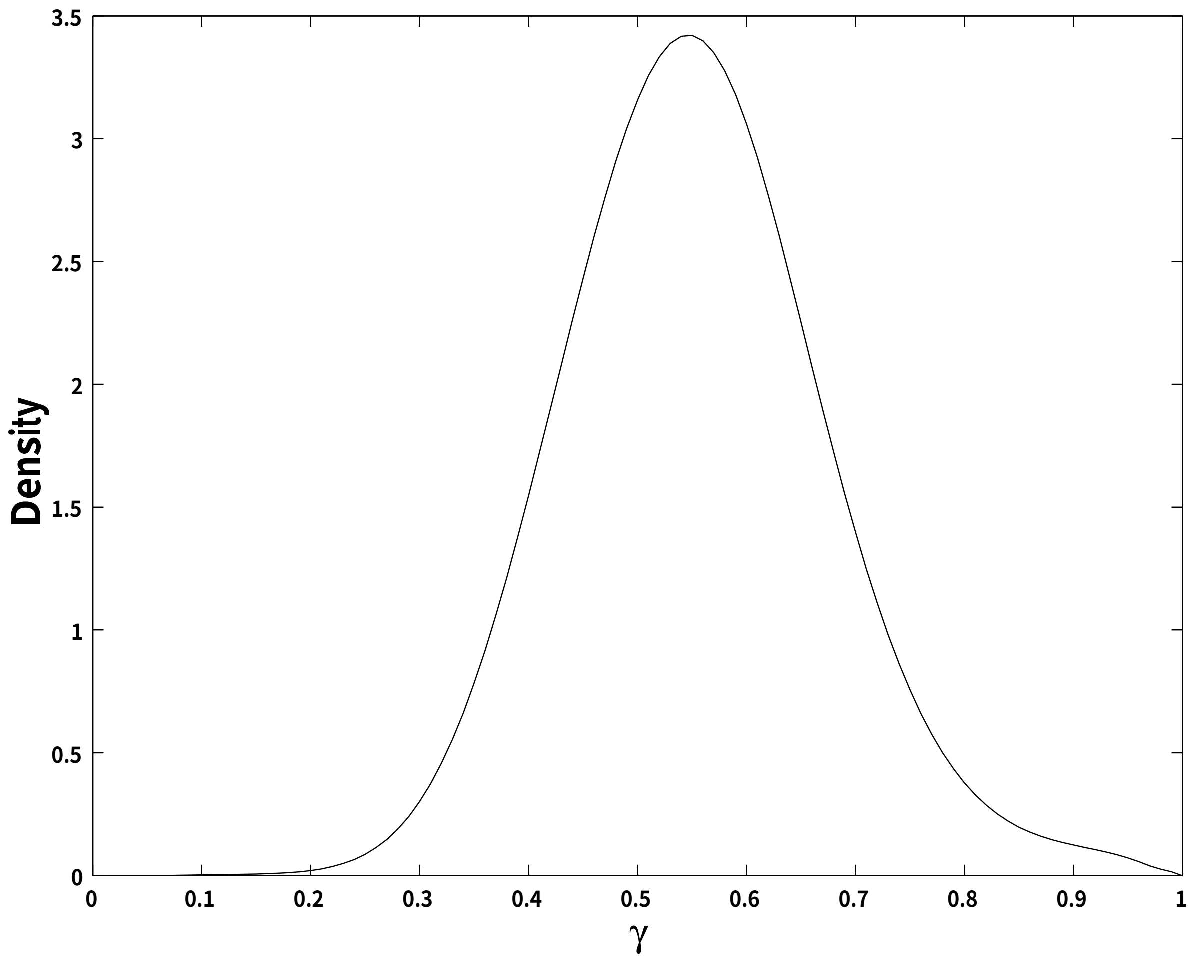

Figure 5: Marginal Posterior Distribution of γ (Production Function)