Bayesian Treatments to Panel Data Models

with an Application to Models of Productivity

Junrong Liu

Department of Economics/Rice University

Robin Sickles

Department of Economics/Rice University

E. G. Tsionas

University of Athens

This Draft

November 25, 2013

Abstract

This paper considers two models for uncovering information about technical change in large heterogeneous panels. The first is a panel data model with nonparametric time effects. Second, we consider a panel data model with common factors whose number is unknown and their effects are firm-specific. This paper proposes a Bayesian approach to estimate the two models. Bayesian inference techniques organized around MCMC are applied to implement the models. Monte Carlo experiments are performed to examine the finite-sample performance of this approach, which dominates a variety of estimators that rely on parametric assumptions. In order to illustrate the new method, the Bayesian approach has been applied to the analysis of efficiency trends in the U.S. largest banks using a dataset based on the Call Report data from FDIC over the period from 1990 to 2009.

Keywords: Panel data, time-varying heterogeneity, Bayesian econometrics, banking studies, productivity

JEL Classification: C23; C11; G21; D24

1. Introduction

In this paper, we consider two panel data models with unobserved heterogeneous time-varying effects, one with individual effects treated as random functions of time, the other with common factors whose number is unknown and their effects are firm-specific. This paper has two distinctive features and can be considered as a generalization of traditional panel data models. Firstly, the individual effects that are assumed to be heterogeneous across units as well as to be time varying are treated nonparametrically, following the spirit of the model from Bai (2009) and Kneip et al. (2012), and Ahn et al. (2013). For an extended discussion of these and other models used in panel work in the productivity field see Sickles, Hsiao, and Shang (2013).

The other aspect of our generalization is that we propose a Bayesian framework to estimate the two panel data models. There are several advantages of the Bayesian approach. First, following the Bayesian perspective of random coefficient models, (Swamy (1970); Swamy and Tavlas (1995) the model in this paper will not subjectively assume a common functional form for all the individuals as the subjective processes may vary among individuals and fixed parametric values of the parameters that describe this functional relationship may not be well-defined. Moreover, a Bayesian approach may circumvent the theoretically complex as well as the computationally intense nature of nonparametric or semiparametric regression techniques (Yatchew, 1998) and the need to rely on asymptotic theory for inference (Koop and Poirier, 2004).

The literature incorporating a Bayesian approach to panel data models with applications in stochastic frontier analysis has been growing in the last two decades. The approach was first suggested by Van den Broeck et al. (1994), which considers a Bayesian approach under the composed error model. Koop et al. (1997) has established a Bayesian setting where the fixed and random effect models are defined; they also applied Gibbs Sampling to analyze their model. Bayesian numerical integration methods are described in Osiewalski and Steel (1998) and used to fully perform the Bayesian analysis of the stochastic frontier model using both cross-sectional data and panel data. However, the individual effects are assumed to be time-invariant in the papers listed above, which is inappropriate in many settings; for example, in the stochastic frontier analysis, the technical inefficiency levels typically adjust over time. In order to address the temporal behavior of individual technical efficiency effects, Tsionas (2006) considers a dynamic stochastic frontier model using Bayesian inference, where the inefficiency levels are assumed to evolve log-linearly. In the same spirit as Tsionas (2006), our paper will also use the Bayesian integration method and a Markov chain based sampler or Gibbs sampler to provide slope parameter and heterogeneous individual effects inferences. By drawing sequentially from a series of conditional posteriors, a sequence of random samples can be obtained, which will converge to a draw from the joint posterior distribution. Additionally, a desirable characteristic of the Bayesian analysis in this paper is that no conjugate priors are imposed for the individual effects; i.e. we do not require effects to follow a normal prior distribution to ensure that the posterior are in the same family of the prior as with the conjugate prior assumption imposed in classical Bayesian linear regression model. The prior assumption is only imposed on the first-order or second-order difference of the individual effects; therefore, this approach can be applied to more general cases. One of the primary differences of this paper from Tsionas (2006) is that no parametric form is assumed on the evolvement of the effects; or the effects are treated nonparametrically. It will be shown in Section 4 that the Bayesian estimators proposed here consistently outperforms some representative parametric as well as nonparametric estimators under various scenarios of data generating processes.

The rest of this paper is organized as follows. Section 2 describes the first model setup and parameter priors. The Bayesian inference procedures are explained in section 3. Section 4 introduces the second model and the corresponding Bayesian inferences, followed by section 5, which presents our Monte Carlo simulations results. The estimation of the translog distance function is briefly discussed and the empirical application results of the Bayesian estimation of the multi-output/multi-input technology employed by the U.S. banking industry in providing intermediation services are presented in section 6. Section 7 provides concluding remarks.

2. Model 1: A Panel Data Model with Nonparametric Time Effects

The first model in this paper is based on a balanced design with T observations for *n* individual units. Thus, the observations in the panel can be represented in the form $(y_{it}, x_{it}), i = 1,...,n; t = 1,...,T$, where the index *i* denotes the *i*th individual units, and the index *t* denotes the *t*th time period.

A panel data model with heterogeneous time-varying effects is expressed as

$$Y_{it} = X_{it}'\beta + \varphi_i(t) + v_{it},\ i = 1,...,n;\ t = 1,...,T \tag{2.1}$$

where Y_{it} is the response variable, X_{it} is a $p \times 1$ vector of explanatory variable, β is a $p \times 1$ vector of parameters, and the unit specific function of time $\varphi_i(t)$ is a nonconstant and unknown individual effect. We make the standard assumption for the measurement error that $v_{it} \sim NID(0, \sigma^2)$.

The model can also be written in the form below,

$$Y_{it} = X_{it}'\beta + \gamma_{it} + v_{it} \tag{2.2}$$

where γ_{it} is the time-varying heterogeneity and assumed to be independent across units. This assumption is quite reasonable in many applications, particularly in production/cost stochastic frontier models where the effects are measuring the technical efficiency levels. A firm's efficiency level primarily relies on its own factors such as its executives' managerial skills, firm size and operational structure etc. and should thus be heterogeneous across firms. These factors are usually changing over time and so does the firm's efficiency level; therefore, it makes more sense to assume the efficiency level to be time varying too.

For the *ith* individual, the vector-form expression is presented as:

$$Y_i = X_i\beta + \gamma_i + v_i,\ i = 1,...,n \tag{2.3}$$

where Y_i, X_i and γ_i are vectors of *T* dimension.

When applying our model into the field of stochastic frontier analysis, the estimation of and inference on the individual effects $\varphi_i(t)$ or γ_{it} , which represent the time-varying technical efficiency levels, will be as important as that of the slope parameters.

The difference of our model from those using Bayesian approaches in the literature is that no specific parametric form for the prior of the unobserved heterogeneous individual effects is imposed. Instead of resorting to the classical nonparametric regression techniques as they did in (Kneip et al., 2012), a Markov Chain Monte Carlo algorithm is implemented in the Bayesian inference to estimate the model. This paper can be considered as a generalization of Koop and Poirier (2004) to the case of panel data including both individual-specific and time-varying effects. Moreover, it does not rely on the conjugate prior formulation for the time varying individual-effects, which can be too restrictive and undesirable.

A Bayesian analysis of the panel data model set up above requires a specification of the prior distributions over the parameters (γ, β, σ) and computation on the posterior using Bayesian learning process:

$$p\left(\beta,\gamma,\sigma \mid Y,X,\omega\right) \propto p\left(\beta,\sigma,\gamma\right) \bullet l(Y,X;\beta,\gamma,\sigma) \tag{2.4}$$

The prior of the individual effect γ_i as expressed below is not strictly assumed to follow a normal distribution; instead, it is only assumed that the first-order or second-order difference of γ_i follows a normal prior.

$$p\left(\gamma\right) \propto \prod_{i=1}^{n} \exp\left(-\frac{\gamma_i' Q \gamma_i}{2\omega^2}\right) = \exp\left(-\frac{1}{2\omega^2}\gamma'\left(I_n \otimes Q\right)\gamma\right) \tag{2.5}$$

where $Q = D'D$, and D is the $(T-1)\times T$ matrix whose elements are $D_{tt} = 1$, for t =1,…,T–1; $D_{t-1,t} = -1$ for all t = 2,…,T and zero otherwise. The information implied by this prior is that $\gamma_{i,t} - \gamma_{i,t-1} \sim N(0,\omega^2)$, or $D\gamma_i \overset{IID}{\sim} N\left(0,\omega^2 I_{T-1}\right)$. ω is a smoothness parameter which stands for the degree of smoothness. ω can be considered as a hyperparameter, or it can be assumed to have its own prior, which is explained in next session. Given the continuity and first-order differentiability of $\varphi_i(t)$, this assumption says that the first derivative of the time-varying function $\varphi_i(t)$ in Eq.(2.1) is a smooth function of time. The second-order differentiability assumption can be an alternative, which is implied by $\gamma_{it} - 2\gamma_{i,t-1} + \gamma_{i,t-2} \sim N\left(0,\omega^2\right)$ or $D^{(2)}\gamma_i \overset{IID}{\sim} N\left(0,\omega^2 I_{T-1}\right)$ and $Q = D^{(2)\prime}D^{(2)}$.

A noninformative prior distribution is assumed here for the joint prior distribution of the slope parameter β and the unknown variance term σ^2 in Eq.(2.6).

$$p(\beta,\sigma) \propto \sigma^{-1} \tag{2.6}$$

or equivalently the prior distribution is uniform on $(\beta, \log\sigma)$.

Therefore, with the assumptions on the priors above, we have adopted the following form for the joint prior:

$$p\left(\beta,\sigma,\gamma\right) \propto \sigma^{-1}\prod_{i=1}^{n}\exp\left(-\frac{\gamma_i' Q\gamma_i}{2\omega^2}\right) = \sigma^{-1}\exp\left(-\frac{1}{2\omega^2}\gamma'\left(I_n \otimes Q\right)\gamma\right) \tag{2.7}$$

After a specific dataset is applied, the likelihood function under this model is the following expression,

$$l(Y,X;\beta,\gamma,\sigma) \propto \sigma^{-NT}\exp\{-\frac{1}{2\sigma^2}(Y-X\beta-\gamma)'(Y-X\beta-\gamma)\} \tag{2.8}$$

The likelihood is formed by the product of *NT* independent disturbance terms which follow normal distribution $N(0, \sigma^2)$.

With Bayes' Theorem applied, the probability density function is updated utilizing the information from the dataset, thus the joint posterior distribution is derived in Eq.(2.9).

$$\begin{aligned} p\left(\beta,\gamma,\sigma \mid Y,X,\omega\right) \propto \sigma^{-(nT+1)}\exp\{-\frac{1}{2\sigma^2}\left(Y-X\beta-\gamma\right)'\left(Y-X\beta-\gamma\right)\} \\ \times \exp\{-\frac{1}{2\omega^2}\gamma'\left(I_n \otimes Q\right)\gamma\} \end{aligned} \tag{2.9}$$

To proceed with further inference, we need to solve the posterior above in Eq.(2.9) analytically; however, this posterior is not of standard form, and taking draws directly from it is problematic. Therefore, Markov Chain Monte Carlo techniques are considered to implement the inference for the model. Specifically, Gibbs sampling will be used to perform the Bayesian inference. The Gibbs sampler is commonly used under Bayesian inference because of the desirable results that iterative sampling from the conditional distributions will lead to a sequence of random variables converging to the joint distribution. A general discussion on the use of Gibbs sampling is provided in Gelfand and Smith (1990), in which Gibbs sampler is also compared with alternative sampling-based algorithms. For more detailed discussion on Gibbs sampling, one can refer to Gelman et al. (2003). Gibbs sampling can be well-adapted to sampling the posterior distributions here since a collection of conditional posterior distributions are easily derived.

The Gibbs sampling algorithm used in this paper generates a sequence of random samples from the conditional posterior distributions of each parameter, in turn conditional on the current values of the other parameters, and it thus generate a sequence of samples that constitute a Markov Chain, where the stationary distribution of that Markov chain is just the desired joint distribution of all the parameters.

In order to derive the conditional posterior distributions of β, γ and σ, rewrite the likelihood function in Eq.(2.8) to the following form.

$$\begin{aligned} & p(Y \mid \beta,\gamma,\sigma) \propto \sigma^{-NT} \exp\{-\frac{1}{2\sigma^2}(Y - X\beta - \gamma)'(Y - X\beta - \gamma)\} \\ & = \sigma^{-NT} \exp\{-\frac{1}{2\sigma^2}[(Y - X\widehat{\beta} - \gamma)'(Y - X\widehat{\beta} - \gamma) + (\beta - \widehat{\beta})'(X'X)(\beta - \widehat{\beta})]\} \end{aligned} \tag{2.10}$$

where $\hat{\beta} = \left(X'X\right)^{-1} X'\left(Y - \gamma\right)$.

The joint posterior can thus be rewritten in the form below:

$$\begin{aligned} & p\left(\beta,\gamma,\sigma \mid Y,X,\omega\right) \propto \sigma^{-\left(nT+1\right)} \exp\{-\frac{1}{2\omega^2}\gamma'\left(I_n \otimes Q\right)\gamma\} \\ & \times\exp\{-\frac{1}{2\sigma^2}[(Y - X\widehat{\beta} - \gamma)'(Y - X\widehat{\beta} - \gamma) + (\beta - \widehat{\beta})'(X'X)(\beta - \widehat{\beta})]\} \end{aligned} \tag{2.11}$$

Thus, the conditional distribution of β follows the multivariate normal distributions with mean $\hat{\beta}$ and covariance matrix $\sigma^2\left(X'X\right)^{-1}$ since the following distribution is derived from Eq.(2.11).

$$p\left(\beta \mid Y,X,\gamma,\sigma,\omega\right) \propto \exp\{-\frac{1}{2\sigma^2}(\beta - \widehat{\beta})'(X'X)(\beta - \widehat{\beta})\} \tag{2.12}$$

$$\beta \mid \sigma,\gamma,\omega,Y,X \propto f_k\left(\beta \mid \hat{\beta},\ \sigma^2\left(X'X\right)^{-1}\right) \tag{2.13}$$

In order to derive the conditional distribution of the individual effect γ_i, rewrite the joint posterior distribution in the following way:

$$\begin{aligned} & p\left(\beta,\gamma,\sigma \mid Y,X,\omega\right) \propto \\ & \sigma^{-\left(nT+1\right)} \exp\{-\frac{1}{2\sigma^2}\left(\gamma - Y + X\beta\right)'\left(\gamma - Y + X\beta\right) - \frac{1}{2\omega^2}\gamma'\left(I_n \otimes Q\right)\gamma\} \propto \\ & \sigma^{-\left(nT+1\right)} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\gamma_i - Y_i + X_i\beta\right)'\left(\gamma_i - Y_i + X_i\beta\right) - \frac{1}{2\omega^2}\sum_{i=1}^{n}\gamma_i'Q\gamma_i\} \end{aligned} \tag{2.14}$$

Therefore, under the assumption that the effect γ_i 's are independent across individuals, the conditional posterior distribution of $\gamma_i \mid \beta,\sigma,\omega,\{\gamma_j, j \neq i\},Y,X$ is the same as that of $\gamma_i \mid \beta,\sigma,\omega,Y,X$, and it is distributed as a multivariate normal with mean $\hat{\gamma}_i$ and covariance matrix *V* as displayed in Eq.(2.15). The detailed derivation is presented in Appendix A.

$$\gamma_i \mid \beta,\sigma,\omega,\{\gamma_j, j \neq i\},Y,X \sim \gamma_i \mid \beta,\sigma,\omega,Y,X \propto f_T\left(\gamma_i \mid \hat{\gamma}_i,\ \sigma^2\omega^2 V\right) \qquad (2.15)$$

where $\hat{\gamma}_i = \omega^2 V(y_i - X_i\beta)$ and $V = \left(\sigma^2 Q + \omega^2 I_T\right)^{-1}$ for $i = 1,...,n$.

Writing the conditional posterior distribution in the form given by Eq.(2.16), it is clear that the sum of the squared residuals over the unobserved variance $\left(Y - X\beta - \gamma\right)'\left(Y - X\beta - \gamma\right)/\sigma^2$ has the a Chi-squared distribution with *nT* degree of freedom as shown in Eq.(2.17).

$$p\left(\sigma^2 \mid \beta,\gamma,Y,X,\omega\right) \propto (\sigma^{-2})^{nT/2-1}\exp\{-\frac{1}{2\sigma^2}\left(Y - X\beta - \gamma\right)'\left(Y - X\beta - \gamma\right)\} \qquad (2.16)$$

$$\frac{\left(Y - X\beta - \gamma\right)'\left(Y - X\beta - \gamma\right)}{\sigma^2} \mid \beta,\gamma,\omega,Y,X \sim \chi^2_{nT} \qquad (2.17)$$

If the smoothing parameter ω is also assumed to follow its own prior instead of being treated as constantly, its conditional posterior distribution can also be derived. Supposed $\frac{\overline{q}}{\omega^2} \sim \chi^2_{\overline{n}}$, where $\overline{n},\ \overline{q} \geq 0$ hyperparameters, the conditional posterior distribution of ω^2 is derived as:

$$\frac{\overline{q} + \sum_{i=1}^{n}\gamma_i' Q\gamma_i}{\omega^2} \mid \beta,\sigma,\gamma,Y,X \sim \frac{\overline{q} + \sum_{i=1}^{n}\gamma_i' Q\gamma_i}{\omega^2} \mid \gamma,Y,X \sim \chi^2_{\overline{n}+n} \qquad (2.18)$$

Obviously, the hyperparameters $\overline{n}$ and $\overline{q}$ control the prior degree of smoothness that is imposed upon the γ_{it} s. Generally, small values of the prior "sum of squares" $\overline{q}/\overline{n}$ correspond to smaller values of ω and thus a higher degree of smoothness.

Alternatively, we can choose the smoothing parameter ω with cross validation under a Bayesian context, which is similar to that in a classical nonparametric regression. The basic idea of the cross validation method is to leave the data points out one at a time and to choose the value of the smoothing parameter, under which the missing value points are best predicted, by the remainder of the data points.

Let $\theta = [\beta',\sigma,\gamma']'$. The posterior distribution for a specific value of the smoothing parameter is $p(\theta \mid Y,\omega) \propto L(Y;\theta)\, p(\theta \mid \omega)$. If we omit the block of time observations for unit i, we have the posterior $p(\theta_{-i} \mid Y_{-i},\omega) \propto L(Y_{-i};\theta_{-i})\, p(\theta_{-i} \mid \omega)$. Suppose now we have a set of draws $\left\{\theta_{-i,\omega}^{(s)}, s = 1,\ldots,S\right\}$ from $\theta_{-i} \mid Y_{-i},\omega$. It is easy to compute the posterior means $\overline{\theta}_{-i,\omega} = S^{-1}\sum_{s=1}^{S}\theta_{-i,\omega}^{(s)}$ and, as a result, the cross validation statistic is

$$CV(\omega) = (nT)^{-1}\sum_{i=1}^{n}(y_i - X_i\overline{\beta}_{-i,\omega} - \overline{\gamma}_{-i,\omega})'(y_i - X_i\overline{\beta}_{-i,\omega} - \overline{\gamma}_{-i,\omega}) \quad (2.19)$$

The problem is that we do not have draws from $\theta_{-i} \mid Y_{-i},\omega$ but only from $\theta \mid Y,\omega$. However, the posteriors $p(\theta_{-i} \mid Y_{-i},\omega)$ and $p(\theta \mid Y,\omega)$ should be fairly close. Therefore, to produce such draws we use the method of sampling importance resampling (SIR): if a sample $\left\{\theta^{(s)}, s = 1,\ldots,S\right\}$ from a distribution with kernel density $g(\theta)$ is available and if the existing sample is resampled with probabilities $W_s = \dfrac{f(\theta^{(s)}) \,/\, g\left(\theta^{(s)}\right)}{\sum_{r=1}^{S} f(\theta^{(r)}) \,/\, g\left(\theta^{(r)}\right)}$, for $s = 1,\ldots,S$, then it can be transformed to a distribution with kernel $f(\theta)$, In our context, the existing sample from $p(\theta \mid Y,\omega)$ is transformed to an approximate sample from $p(\theta_{-i} \mid Y_{-i},\omega)$ using $w_s = \sigma^{(s)T} \exp\left[\dfrac{1}{2\sigma^{(s)2}}\left(y_i - X_i\beta^{(s)} - \gamma_i^{(s)}\right)'\left(y_i - X_i\beta^{(s)} - \gamma_i^{(s)}\right) + \dfrac{1}{2\omega^2}\gamma_i^{(s)\prime}Q\gamma_i^{(s)}\right]$, and $W_s = w_s \,/\sum_{r=1}^{S} w_r$. The size of the resample is set to 20% of the original sample. For each specific value of ω, the posteriors $p(\theta_{-i} \mid Y_{-i},\omega)$ are simulated using SIR for each $i = 1,\ldots,n$, and the value of ω that yields the minimum of $CV(\omega)$ is determined.

A useful byproduct of this approach is that it yields samples $\left\{\beta_{-i}^{(s)}, \sigma_{-i}^{(s)}, \gamma_{-i}^{(s)}\right\}$, which represent all the parameters except one individual block *i*. These samples and the posteriors approximated can be useful when sensitivity analysis with respect to the observations is necessary.

This paper uses a Gibbs sampler to draw observations from the conditional posteriors from Eq.(2.13) to Eq.(2.17) with data augmentation. Draws from these conditional posteriors will eventually converge to the joint posterior in Eq.(2.9). Since the conditional posterior distribution of β follows the multivariate normal distribution displayed in Eq.(2.13), it will be straightforward to sample from it.

For the individual effects γ_i, sampling is also straightforward since its conditional posterior follows a multivariate normal distribution with mean vector $\hat{\gamma}_i$ and covariance matrix $\sigma^2\omega^2 V$ as expressed in Eq.(2.15).

Finally, to draw samples from the conditional posterior distribution function for the unobserved variance of the measurement error σ term, we have two simple steps. Firstly, we can draw samples directly from that of $\left(Y - X\beta - \gamma\right)'\left(Y - X\beta - \gamma\right) / \sigma^2$, which is shown in Eq.(2.16) to follow a chi-squared distribution with degree of freedom *nT*. Secondly, assign the values of $\left(Y - X\beta - \gamma\right)'\left(Y - X\beta - \gamma\right) / Chi - rnd$ to σ^2, where $Chi - rnd$ is the generated random variable that follows χ^2_{nT} in the first step.

3. Model 2: A Panel Data Model with Factors

We consider another panel data model, where the effects are treated as linear combination of unknown basis functions or factors.

$$y_{it} = x_{it}'\beta + \phi_t'\gamma_i + v_{it} = x_{it}'\beta + \sum_{g=1}^{G} \phi_{tg}\gamma_{ig} + v_{it} \qquad (3.1)$$

where ϕ_t is a $G \times 1$ vector of common factors, and γ_i is a $G \times 1$ vector of individual-specific factor loadings. For the i th individual we have

$$\underset{(T\times 1)}{y_i} = \underset{(T\times k)}{X_i}\underset{(k\times 1)}{\beta} + \underset{(T\times G)}{\Phi}\underset{(G\times 1)}{\gamma_i} + \underset{(T\times 1)}{v_i}, \quad i = 1,2,\ldots,n \qquad (3.2)$$

For the t th time period we have

$$\underset{(n\times 1)}{y_t} = \underset{(n\times k)}{X_t}\underset{(k\times 1)}{\beta} + \underset{(n\times G)}{\Gamma}\underset{(G\times 1)}{\phi_t} + \underset{(n\times 1)}{v_t}, \quad t = 1,2,\ldots T \qquad (3.3)$$

for each $t = 1,\ldots,T$, where $\Phi = \begin{bmatrix} \phi_1' \\ \vdots \\ \phi_T' \end{bmatrix}$, and $\Gamma = \begin{bmatrix} \gamma_1' \\ \vdots \\ \gamma_n' \end{bmatrix}$.

If we set $\phi_{1t} = 1$, then γ_{it} acts as an individual-specific intercept. Effectively, the first column of Φ contains ones. The model for all observations can be written as $Y = X\beta + (I_n \otimes \Phi)\gamma + v = X\beta + (I_T \otimes \Gamma)\phi$, where $\gamma = vec(\Gamma)$, and $\phi = vec(\Phi)$.

This model setting follows that in Kneip et al. (2012), and it satisfies the structural assumption, which is Assumption 1 from that paper.

Assumption 1: For some fixed $G \in \{0,1,2,\ldots\} < T$, there exists a G-dim space L_T, such that

$\varphi_i(t) = \phi_t{}'\gamma_i$ holds for with probability 1.

We define the priors similarly to Model 1. Regarding the slope parameter β and variance of the noise term σ, we still assume a noninformative prior $p(\beta,\sigma) \propto \sigma^{-1}$. For the common factors it is reasonable to assume

$$p\left(\phi_1,\phi_2,...,\phi_T\right) \propto \exp\left(-\frac{\sum_{t=1}^{T}(\phi_t-\phi_{t-1})'(\phi_t-\phi_{t-1})}{2\omega^2}\right) = \exp\left(-\frac{1}{2\omega^2}tr\Phi' Q\Phi\right) \quad (3.4)$$

This assumption is consistent with the idea that common factors evolve "smoothly" over time, and the degree of smoothness is controlled by the parameter ω and $\phi_0 = 0$. For the loadings we can assume $\gamma_i \overset{IID}{\sim} N_G(\overline{\gamma},\Sigma)$. An alternative is to avoid the proliferation of factors by constraining stochastically the loadings to approach zero in the following sense: If $\underset{(n\times G)}{\Gamma} = \left[\gamma_{(1)},...,\gamma_{(G)}\right]$, then $\gamma_{(1)} \sim N_n\left(\overline{\gamma},\psi^2 I_n\right)$, $\gamma_{(g)} \sim N_n\left(\alpha^g\overline{\gamma},\ \lambda^g\psi^2\right)$, for $g = 1,...,G$, and α,λ are parameters between zero and one.

The posterior kernel distribution is

$$p(\beta,\sigma,\phi,\gamma \mid Y,X) \propto \sigma^{-(nT+1)}\exp\left[-\frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it}-x_{it}'\beta-\phi_t'\gamma_i\right)^2}{2\sigma^2}-\frac{\sum_{t=2}^{T}(\phi_t-\phi_{t-1})'(\phi_t-\phi_{t-1})}{2\omega^2}\right]\prod_{i=1}^{n}p(\gamma_i \mid \zeta) \quad (3.5)$$

where ζ denotes any hyperparameters that are present in the prior of γ_i s. When $\gamma_i \overset{IID}{\sim} N_G(\overline{\gamma},\Sigma)$ we have

$$\begin{gathered} p\left(\beta,\sigma,\phi,\gamma,\overline{\gamma},\Sigma \mid Y,X\right) \propto \\ \sigma^{-\left(nT+1\right)}\exp\left[-\frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it}-x_{it}'\beta-\phi_t'\gamma_i\right)^2}{2\sigma^2}-\frac{\sum_{t=1}^{T}(\phi_t-\phi_{t-1})'(\phi_t-\phi_{t-1})}{2\omega^2}\right] \\ \left|\Sigma\right|^{-n/2}\exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\gamma_i-\overline{\gamma}\right)'\Sigma^{-1}\left(\gamma_i-\overline{\gamma}\right)\right]p\left(\overline{\gamma},\Sigma\right) \end{gathered} \quad (3.6)$$

where $p(\overline{\gamma},\Sigma)$ denotes the prior on the hyperparameters. A reasonable choice is $p(\overline{\gamma} \mid \Sigma) \propto$ const., and $p(\Sigma) \propto |\Sigma|^{-(\overline{\nu}+1)/2}\exp\left(-\frac{1}{2}\overline{A}\Sigma^{-1}\right)$, which leads to:

$$\begin{gathered} p\left(\beta,\sigma,\phi,\gamma,\overline{\gamma},\Sigma \mid Y,X\right) \propto \\ \sigma^{-\left(nT+1\right)}\left|\Sigma\right|^{-\left(n+\overline{\nu}+1\right)/2}\exp\left[-\frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it}-x_{it}'\beta-\phi_t'\gamma_i\right)^2}{2\sigma^2}-\frac{\sum_{t=1}^{T}(\phi_t-\phi_{t-1})'(\phi_t-\phi_{t-1})}{2\omega^2}-\frac{1}{2}tr\left(A\Sigma^{-1}\right)\right] \end{gathered} \quad (3.7)$$

where $A = \overline{A} + \sum_{i=1}^{n} (\gamma_i - \overline{\gamma})(\gamma_i - \overline{\gamma})'$.

In order to proceed with the Bayesian inference, we also use the Gibbs Sampling algorithm. In this scenario, the implementation of Gibbs sampling is quite straightforward since we can derive analytically the following marginal posteriors for the parameters we are interested in.

$$\beta \mid \sigma, \phi, \gamma, \overline{\gamma}, \Sigma, Y, X \sim N_k\left(\overline{\beta}, \sigma^2 \left(X'X\right)^{-1}\right), \text{ where } \overline{\beta} = \left(X'X\right)^{-1} X'\left(Y - \left(I_n \otimes \Phi\right)\gamma\right) \quad (3.8)$$

$$\frac{\left(Y - X\beta - \left(I_n \otimes \Phi\right)\gamma\right)' \left(Y - X\beta - \left(I_n \otimes \Phi\right)\gamma\right)}{\sigma^2} \mid \beta, \gamma, \phi, \overline{\gamma}, \Sigma \sim \chi^2_{nT} \quad (3.9)$$

$$\overline{\gamma} \mid \beta, \sigma, \phi, \gamma, \overline{\gamma}, \Sigma, Y, X \sim \overline{\gamma} \mid \gamma, \Sigma, Y, X \sim N_G\left(n^{-1}\sum_{i=1}^{n} \gamma_i,\ n^{-1}\Sigma\right) \quad (3.10)$$

$$p\left(\Sigma \mid \beta, \sigma, \phi, \gamma, \overline{\gamma}, Y, X\right) \propto \left|\Sigma\right|^{-\left(n+\overline{\nu}+1\right)/2} \exp\left[-\frac{1}{2} tr\left(A\Sigma^{-1}\right)\right] \quad (3.11)$$

$$\gamma_i \mid \beta, \sigma, \overline{\gamma}, \Sigma, Y, X \sim N_G\left(\hat{\gamma}_i,\ \sigma^2\left(\Phi'\Phi + \sigma^2\Sigma^{-1}\right)^{-1}\right) \quad (3.12)$$

where $\hat{\gamma}_i = \left(\Phi'\Phi + \sigma^2\Sigma^{-1}\right)^{-1}\left(\Phi' e_i + \sigma^2\Sigma^{-1}\overline{\gamma}\right)$, $e_i = y_i - X_i\beta$, for each $i = 1, \ldots, n$.

$$\phi_t \mid \beta, \sigma, \gamma, \overline{\gamma}, \Sigma, Y, X, \left\{\phi_\tau, \tau \neq t\right\} \sim N_G\left(\hat{\phi}_t,\ \sigma^2\omega^2\left(\omega^2\Gamma'\Gamma + 2\sigma^2 I_G\right)^{-1}\right) \quad (3.13)$$

where $\hat{\phi}_t = \left(\omega^2\Gamma'\Gamma + 2\sigma^2 I_G\right)^{-1}\left(\omega^2\Gamma' e_t + \sigma^2\left(\phi_{t-1} + \phi_{t+1}\right)\right)$, $e_t = y_t - X_t\beta$, for each $t = 1, \ldots, T$.

Using a Gibbs sampler, we can draw observations from the marginal posteriors from Eq.(3.8) to Eq.(3.13) with data augmentation. Draws from these conditional posteriors will eventually converge to the joint posterior in Eq.(3.7). Since the conditional posterior distribution of β follows the multivariate normal distribution displayed in Eq.(3.8), it will be straightforward to sample from it. To draw samples from the conditional posterior distribution function for the unobserved variance of the measurement error σ term, we can firstly draw samples directly from that of $\left(Y - X\beta - \gamma\right)'\left(Y - X\beta - \gamma\right) / \sigma^2$, which is shown in Eq.(3.9) to follow a chi-squared distribution with degree of freedom *nT*, then assign the values of $\left(Y - X\beta - \gamma\right)'\left(Y - X\beta - \gamma\right) / Chi - rnd$ to σ^2, where $Chi - rnd$ is the generated random variable that follows χ^2_{nT} in the first step.

For the mean parameter $\bar{\gamma}$, sampling is also straightforward since its conditional posterior follows a multivariate normal distribution; the variance matrix Σ follows an inverted Wishart distribution and can thus be drawn directly.

For the unknown common factors γ_i, and the corresponding factor loadings ϕ_t, we can draw directly from multivariate normal distribution following Eq.(3.12) and Eq.(3.13) respectively.

Up till now, we assume the number of factors is known as G. Now, we need to find what G is under a Bayesian way. We consider models with G=1, 2,…,L. Suppose $p(\theta,\Gamma_G)$ and $L(\theta,\Gamma_G;Y,G)$ denote the prior and likelihood, respectively, of a model with G factors, where θ is the vector of parameters common in all models (like β and σ) and Γ_G denotes a vector of parameters related to the factors and their loadings, φ and γ. The marginal likelihood is $M_G(Y)=\int L(\theta,\Gamma_G;Y,G)\,p(\theta,\Gamma_G)\,d\Gamma_G d\theta$. For models with different number of factors, say G and G' we can consider the Bayes factor in favor of the first model and against the second:

$$BF=\frac{\int L(\theta,\Gamma_G;Y,G)\,p(\theta,\Gamma_G)\,d\Gamma_G d\theta}{\int L(\theta,\Gamma_{G'};Y,G')\,p(\theta,\Gamma_{G'})\,d\Gamma_{G'} d\theta}=\frac{M_G(Y)}{M_{G'}(Y)} \tag{3.14}$$

Essentially, what is required is the ability to generate MCMC draws from models with different numbers of factors and record the draws $\{\theta_G^{(s)}, s=1,\dots,S\}$, for $G=1,2,\dots$ Then from the candidate's formula introduced by Chib (1995) , we have $M_G(Y)=\frac{\mathscr{P}(\theta\,|\,Y,G)}{p(\theta\,|\,Y,G)}$, where the numerator $\mathscr{P}(\theta\,|\,Y,G)=\int L(\theta,\Gamma_G;Y,G)\,p(\theta,\Gamma_G)\,d\Gamma_G$, and the denominator $p(\theta\,|\,Y,G)=\int p(\theta,\Gamma_G\,|\,Y,G)d\Gamma_G$, the normalized posterior in terms of the structural parameters, θ, only. Given a point $\hat{\theta}$, for example, the marginal likelihood can be estimated as:

$$M_G(Y)=\frac{L(\hat{\theta};Y,G)\,p(\hat{\theta})}{p(\hat{\theta}\,|\,Y,G)},$$

where the denominator is estimated using the Laplace approximation:

$$p(\hat{\theta}\,|\,Y,G)=(2\pi)^{-K/2}\,|\mathbf{S}|^{-1/2},$$

where $\mathbf{S} = S^{-1}\sum_{s=1}^{S}\left(\theta^{(s)} - \hat{\theta}\right)\left(\theta^{(s)} - \hat{\theta}\right)'$, and K is the dimensionality of θ.

Computation of the marginal likelihood requires the computation of the integral in the numerator $\mathscr{P}\left(\theta \mid Y, G\right)$ with respect to φ and γ. As this is not available analytically, we adopt the following approach.

$$\mathscr{P}\left(\theta \mid Y, G\right) = \int L\left(\theta, \Gamma_G; Y, G\right) p\left(\theta, \Gamma_G\right) d\Gamma_G = \int \frac{L\left(\theta, \Gamma_G; Y, G\right) p\left(\theta, \Gamma_G\right)}{q\left(\Gamma_G\right)} q\left(\Gamma_G\right) d\Gamma_G \quad (3.15)$$

where $q\left(\Gamma_G\right)$ is a convenient importance sampling density. We factor the importance density as $q\left(\Gamma_G\right) = \prod_{t=1}^{T} q_t^{\phi}\left(\phi_t\right) \prod_{i=1}^{n} q_i^{\gamma}\left(\gamma_i\right)$, where q_t^{ϕ} and q_i^{γ} are univariate densities. The densities are chosen to be univariate Student-*t* with 5 degrees of freedom, with parameters matched to the posterior mean and standard deviation of MCMC draws for φ and γ respectively. Then the integral in (2) is estimated using standard importance sampling. The standard deviations are multiplied by constants h_ϕ and h_γ, which are selected so that the importance weights are as close to uniformity as possible. Specifically we try 100 random pairs in the interval 0.1 to 10 and select the values of h for which the Kolmogorov-Smirnov test is lowest. Of course, acceptance of the uniformity is not possible but the weights so selected are not concentrated around zero with a few outliers. Finally we truncate the weights to their 99.5% confidence interval but in very few instances this has been found necessary as outlying values are only rarely observed. There is some evidence that changing also the degrees of freedom of the Student-*t* provides some improvement but we did not pursue this further as the final results for Bayes factors were not found to differ significantly.

Given marginal likelihoods $M_g(Y)$, $g = 1, \ldots, G$, posterior model probabilities can be estimated as

$$p_g(Y) = \frac{M_g\left(Y\right)}{\sum_{g=1}^{G} M_g\left(Y\right)}, \; g = 1, \ldots, G \quad (3.16)$$

The posterior model probabilities summarize the evidence in favor of a model with a given number of factors.

4. Monte Carlo Simulations

To illustrate the model and inspect the finite sample performance of the new estimators using the Bayesian approach with nonparametric individual effects specification and with the factor model setting (BE1 and BE2 henceforth), Monte Carlo experiments are carried out in this section. The performance of the Bayesian estimator is compared with the parametric time-variant estimator BC, the estimators proposed by (Cornwell et al., 1990)- within estimator (CSSW hereafter) and GLS estimator (CSSG henceforth)- and the (Kneip et al., 2012) estimator utilizing the nonparametric regression techniques (KSS henceforward) based on a factor analysis.

Consider the panel data model(2.2), which can be written in the sum form: $Y_{it} = \sum_{k=1}^{p} \beta_k X_{it}^k + \gamma_{it} + v_{it}$. Samples of size *n = 50, 100,200* with *T = 20, 50* in a model with *p = 2* regressors are simulated. In each sample of the Monte Carlo experiments, the regressors X_{it} are randomly drawn from a standard multivariate normal distribution $N(0, I_p)$. The disturbance term σ^2 is randomly drawn from the $IIDN(0, 0.1^2)$.

The time-varying individual effects are generated by the following DGPs respectively, which includes as many different types of parametric forms such as quadratic function of time trend, random walk, the functional form capturing significant temporal variations.

DGP1: $\gamma_{it} = \theta_{i0} + \theta_{i1}(t/T) + \theta_{i2}(t/T)^2$

DGP2: $\gamma_{it} = \phi_i r_t$

DGP3: $\gamma_{it} = v_{i1} t/T \cos(4\pi t/T) + v_{i2} t/T \sin(4\pi t/T)$

DGP4: $\gamma_{it} = \theta_{i0} + \theta_{i1}(t/T) + \theta_{i2}(t/T)^2 + \phi_i r_t + v_{i1} t/T \cos(4\pi t/T) + v_{i2} t/T \sin(4\pi t/T)$

where $\theta_{ij}(j = 0,1,2)$ is drawn *i.i.d.* from a standard normal distribution N(0,1), $\phi_i \sim i.i.d.\ N(0,1)$, $r_{t+1} = r_t + \delta_t, \delta_t \sim i.i.d.\ N(0,1)$, $v_{ij}(j = 1,2) \sim i.i.d.\ N(0,1)$.

DGP1 specifies a time-varying effect based on a second-order polynomial of the time trend, which is used to model smooth change in efficiency levels; DGP2 utilizes the effect as a random walk process; DGP3 is considered as the case that large temporal variations are modeled; DGP 4 is the general case that integrate all the scenarios in DGP1 through DGP3 in order to provide the evidence that the Bayesian Estimator is of expansive use in different types of parametric forms.

In this paper, Gibbs sampling has been implemented using 55,000 iterations with the first 5,000 samples ignored, the commonly called burn-in periods. The reason for discarding the first several periods is that it may take a while to reach the stationary distribution of the Markov chain, which is the desired joint distribution. Then we consider only every other 10th draw to mitigate the impact of autocorrelation since successive samples from a Markov chain tend to have correlations to some extent and thus are not independent from each other. With regard to the selection of the number of factors, Gibbs samplers for all DGPs rely on MCMC simulation from models with G ranging from one to eight. The true number of factors is 3,2,1, and 6 for the four DGPs respectively.

The simulation results for all the DGPs are displayed as below in Table 1 through Table 4. The BC time-varying estimator along with CSSW, CSSG, and KSS estimators are displayed for a comparison with the Bayesian Estimator proposed in this paper. For the coefficient parameter β in the model, both the estimate and the standard deviation results are presented at the upper panel of every table; for the individual effects γ_{it}, MSE results are displayed at the lower panel of each table. The normalized MSE formula of the individual effects γ_{it} is calculated in (4.1).

$$R(\hat{\gamma}_{it}, \gamma_{it}) = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(\hat{\gamma}_{it} - \gamma_{it})^2}{\sum_{i=1}^{n}\sum_{t=1}^{T}\gamma_{it}^{2}} \tag{4.1}$$

Table 1 : Monte Carlo Simulations for DGP1

Mean Squared Error for the Individual Effects

n	T	BC	CSSW	CSSG	KSS	BE1	BE2
50	20	0.7284	0.0012	0.0012	0.0039	0.0053	0.0671
	50	0.9371	0.0005	0.0005	0.1255	0.0021	0.0323
100	20	0.8222	0.0008	0.0008	0.0033	0.0031	0.0183
	50	0.8245	0.0003	0.0003	0.0220	0.0018	0.0115
200	20	0.8451	0.0008	0.0008	0.0023	0.0027	0.0101
	50	0.8823	0.0003	0.0003	0.0021	0.0011	0.0083

Estimate and Standard Error for the Slope Coefficients

	T = 20						T=50					
	BC	CSSW	CSSG	KSS	BE1	BE2	BC	CSSW	CSSG	KSS	BE1	BE2
n =50												
EST1	0.5250	0.4961	0.4965	0.4954	0.4981	0.5017	0.5105	0.4991	0.4992	0.4999	0.5013	0.4998
SE1	0.0130	0.0033	0.0032	0.0029	0.0057	0.0042	0.0073	0.0021	0.0020	0.0019	0.0035	0.0027
EST2	0.4856	0.4949	0.4948	0.4919	0.4985	0.5020	0.4969	0.5048	0.5047	0.5053	0.5001	0.5002
SE2	0.0139	0.0035	0.0033	0.0031	0.0055	0.0041	0.0073	0.0020	0.0020	0.0018	0.0032	0.0024
n=100												
EST1	0.4973	0.5018	0.5013	0.5045	0.5023	0.5002	0.4843	0.4999	0.4998	0.4991	0.4999	0.5001
STD1	0.0099	0.0023	0.0022	0.0021	0.0032	0.0027	0.0066	0.0014	0.0014	0.0014	0.0023	0.0018
EST2	0.5047	0.5009	0.5012	0.5022	0.5016	0.5001	0.4995	0.5001	0.5000	0.4990	0.5003	0.5001
STD2	0.0098	0.0022	0.0022	0.0020	0.0032	0.0028	0.0066	0.0014	0.0014	0.0013	0.0022	0.0017
n=200												
EST1	0.4936	0.5013	0.5015	0.5009	0.5000	0.4981	0.5000	0.5007	0.5007	0.5000	0.5012	0.5001
STD1	0.0071	0.0016	0.0016	0.0016	0.0027	0.0022	0.0042	0.0010	0.0010	0.0010	0.0019	0.0015
EST2	0.4983	0.5016	0.5020	0.5019	0.5002	0.4993	0.5027	0.4972	0.4972	0.4969	0.5003	0.5004
STD2	0.0071	0.0016	0.0016	0.0016	0.0027	0.0022	0.0042	0.0010	0.0010	0.0010	0.0018	0.0014

DGP1 is consistent with the assumptions for the time-varying effects in the CSS model. Hence, it is expected that the CSSW and CSSG estimators will have better performance compared with other estimators. The conjecture turns out to be true and is proved in Table 1. It is also shown in Table 1 that the performances of the Bayesian estimators are comparable to those of the CSSW, CSSG and KSS estimators in terms of the estimation on individual effects. Under the cases of n = 50, T = 50 and of n = 100, T = 50, the Bayesian estimators provides more accurate estimation on the individual effects than the KSS estimator. This implies that the performance of the Bayesian estimators is quite efficient in estimating time-varying effects of the smoothing-curve forms, like the second-order polynomials. It is not surprising that the mean squared errors of the Bayesian estimators are consistently much lower than those of the BC estimator for all sample sizes.

Table 2: Monte Carlo Simulations for DGP 2

Mean Squared Error for the Individual Effects							
n	T	BC	CSSW	CSSG	KSS	BE1	BE2
50	20	0.9202	0.1266	0.1266	0.0182	0.0071	0.0048
	50	0.9052	0.2996	0.2996	0.0238	0.0053	0.0025
100	20	0.8588	0.4553	0.4553	0.0531	0.0040	0.0037
	50	0.9884	0.1065	0.1065	0.0046	0.0028	0.0013
200	20	0.9183	0.6376	0.6375	0.0706	0.0022	0.0027
	50	0.9526	0.0616	0.0616	0.0028	0.0009	0.0008

Estimate and Standard Error for the Slope Coefficients												
	T = 20						T=50					
	BC	CSSW	CSSG	KSS	BE1	BE2	BC	CSSW	CSSG	KSS	BE1	BE2
n =50												
EST1	0.4786	0.4857	0.4904	0.5059	0.5010	0.4993	0.4820	0.4811	0.4938	0.4972	0.5052	0.4983
SE1	0.0460	0.0308	0.0298	0.0136	0.0262	0.0037	0.0243	0.0230	0.0227	0.0059	0.0177	0.0029
EST2	0.4664	0.4414	0.4854	0.4599	0.5031	0.4992	0.4840	0.4660	0.4848	0.4988	0.5001	0.4999
SE2	0.0491	0.0326	0.0314	0.0146	0.0261	0.0035	0.0241	0.0226	0.0225	0.0059	0.0174	0.0028
n=100												
EST1	0.4854	0.4818	0.4898	0.5065	0.4997	0.5002	0.5137	0.5360	0.5089	0.4950	0.5101	0.4987
STD1	0.0200	0.0195	0.0188	0.0075	0.0163	0.0028	0.0415	0.0257	0.0254	0.0055	0.0128	0.0018
EST2	0.5005	0.4996	0.5115	0.5101	0.4993	0.5001	0.4482	0.5283	0.5143	0.5127	0.5002	0.4992
STD2	0.0198	0.0189	0.0186	0.0073	0.0164	0.0029	0.0415	0.0256	0.0254	0.0055	0.0130	0.0018
n=200												
EST1	0.5051	0.4995	0.5015	0.4864	0.4927	0.5013	0.4274	0.5097	0.4968	0.5018	0.5032	0.5011
STD1	0.0169	0.0175	0.0171	0.0067	0.0120	0.0021	0.0527	0.0202	0.0200	0.0032	0.0078	0.0013
EST2	0.4895	0.4898	0.5147	0.4951	0.4901	0.5020	0.3996	0.4930	0.5015	0.5042	0.5021	0.5031
STD2	0.0170	0.0175	0.0171	0.0067	0.0121	0.0020	0.0531	0.0204	0.0202	0.0033	0.0077	0.0014

DGP2 considers the case where the individual effects are generated by a random walk and can take an arbitrary functional form. Therefore, the CSSW and CSSG estimators, which rely on the assumption that the individual effects are the quadratic function of the time trend, would have worse performance than in DGP1 due to misspecification. The BC estimator is also expected to perform poorly on the estimation of the individual effects. However, the Bayesian and KSS estimators impose no functional forms on the temporal pattern of the individual effects, and thus should be able to approximate arbitrary forms of time-varying effects. The results in Table 2 have confirmed our expectation. It is shown that the Bayesian estimators dominantly outperforms the estimators which rely on functional form assumptions and also have better estimation performance in terms of MSE of individual effects than the KSS estimators do in any size of panel.

DGP3 is considered to characterize the significant time variations in individual effects. As we can see from Table 3, The Bayesian estimators have comparable performance to the KSS estimator and outperforms it in the cases large panels such as when n = 100 and 200. Other estimators, whose effects rely on parametric assumptions of simple functional forms, are to a great extent dominated by the Bayesian estimators.

Table 3: Monte Carlo Simulations for DGP3

Mean Squared Error for the Individual Effects							
n	T	BC	CSSW	CSSG	KSS	BE1	BE2
50	20	3.3477	0.8816	0.8816	0.0130	0.0244	0.0356
	50	3.3639	0.8469	0.8468	0.0082	0.0134	0.0152
100	20	3.5102	0.8309	0.8303	0.0123	0.0116	0.0282
	50	3.7625	0.8357	0.8356	0.0072	0.0028	0.0053
200	20	3.8433	0.8335	0.8333	0.0121	0.0083	0.0116
	50	3.8513	0.8393	0.8392	0.0063	0.0014	0.0019

Estimate and Standard Error for the Slope Coefficients												
	T = 20						T=50					
	BC	CSSW	CSSG	KSS	BE1	BE2	BC	CSSW	CSSG	KSS	BE1	BE2
n =50												
EST1	0.5277	0.5250	0.4994	0.4989	0.5012	0.5002	0.4868	0.4871	0.4976	0.5005	0.4991	0.5001
SE1	0.0188	0.0203	0.0197	0.0029	0.0081	0.0038	0.0122	0.0122	0.0120	0.0018	0.0041	0.0025
EST2	0.4905	0.4998	0.5062	0.4930	0.4981	0.4997	0.5259	0.5255	0.5207	0.5052	0.4994	0.5003
SE2	0.0198	0.0215	0.0207	0.0031	0.0078	0.0035	0.0121	0.0120	0.0119	0.0018	0.0042	0.0023
n=100												
EST1	0.4816	0.4768	0.4998	0.5030	0.4961	0.4992	0.4877	0.4863	0.4972	0.4986	0.4995	0.5002
STD1	0.0132	0.0139	0.0134	0.0021	0.0058	0.0025	0.0076	0.0077	0.0076	0.0013	0.0022	0.0017
EST2	0.4907	0.4816	0.5088	0.5028	0.4971	0.4985	0.5024	0.5118	0.5089	0.4993	0.4990	0.5004
STD2	0.0131	0.0135	0.0133	0.0021	0.0057	0.0024	0.0076	0.0077	0.0076	0.0013	0.0023	0.0018
n=200												
EST1	0.5120	0.5103	0.5110	0.5016	0.5012	0.5011	0.4976	0.5012	0.4962	0.4999	0.4981	0.5052
STD1	0.0088	0.0091	0.0089	0.0016	0.0042	0.0013	0.0055	0.0054	0.0054	0.0010	0.0015	0.0011
EST2	0.4885	0.4892	0.5019	0.5029	0.5015	0.5014	0.4874	0.4883	0.4957	0.4973	0.4992	0.4994
STD2	0.0088	0.0091	0.0089	0.0016	0.0041	0.0012	0.0055	0.0055	0.0054	0.0010	0.0016	0.0012

DGP 4 can be considered a mixed scenario of those from the first three DGPs. It is shown in Table 4 that the Bayesian estimators dominantly outperform the BC, CSSW, and CSSG estimators in terms of the MSE of the individual effects and are comparable to KSS.

Through all the DGPs, although the performance of the slope parameter estimation is reasonably well for all the estimators, those estimators based on simple parametric assumptions on the individual effects are not sufficient to provide sound estimation on the effects. This is undesirable since the individual effects correspond to the technical efficiencies in stochastic frontier analysis and should be drawn on no less attention than the slope parameters. Hence, the Bayesian estimators are excellent candidates among all the estimators in modeling the production or cost frontier.

Table 4: Monte Carlo Simulations for DGP4

Mean Squared Error for the Individual Effects							
n	T	BC	CSSW	CSSG	KSS	BE1	BE2
50	20	0.8042	0.2161	0.2161	0.0030	0.0130	0.0445
	50	0.9478	0.2056	0.2056	0.0890	0.0045	0.0141
100	20	0.8770	0.1382	0.1382	0.0026	0.0112	0.0291
	50	0.8626	0.1337	0.1337	0.0193	0.0028	0.0055
200	20	0.8764	0.1301	0.1301	0.0020	0.0098	0.0108
	50	0.9111	0.1445	0.1445	0.0015	0.0015	0.0021

Estimate and Standard Error for the Slope Coefficients												
	T=20						T = 50					
	BC	CSSW	CSSG	KSS	BE1	BE2	BC	CSSW	CSSG	KSS	BE1	BE2
n =50												
EST1	0.5521	0.5250	0.5329	0.4995	0.4922	0.4951	0.5031	0.4871	0.4901	0.4999	0.5051	0.5010
SE1	0.0233	0.0203	0.0197	0.0030	0.0039	0.0031	0.0148	0.0122	0.0120	0.0019	0.0032	0.0025
EST2	0.4788	0.4998	0.5014	0.4907	0.5011	0.4977	0.5201	0.5255	0.5246	0.5053	0.5001	0.5003
SE2	0.0248	0.0215	0.0207	0.0031	0.0036	0.0032	0.0147	0.0120	0.0119	0.0018	0.0031	0.0028
n=100												
EST1	0.4732	0.4768	0.4713	0.5017	0.5001	0.5052	0.4720	0.4863	0.4867	0.4985	0.5003	0.5001
STD1	0.0169	0.0139	0.0134	0.0022	0.0031	0.0027	0.0103	0.0077	0.0076	0.0014	0.0021	0.0017
EST2	0.4880	0.4816	0.4836	0.5018	0.5000	0.5041	0.5077	0.5118	0.5117	0.4998	0.5002	0.5003
STD2	0.0167	0.0135	0.0133	0.0021	0.0032	0.0025	0.0103	0.0077	0.0076	0.0014	0.0020	0.0015
n=200												
EST1	0.5029	0.5103	0.5112	0.5011	0.5032	0.5001	0.4891	0.5012	0.5012	0.5003	0.4991	0.5002
STD1	0.0116	0.0091	0.0089	0.0016	0.0028	0.0020	0.0069	0.0054	0.0054	0.0010	0.0018	0.0014
EST2	0.4892	0.4892	0.4934	0.5021	0.5013	0.5020	0.4940	0.4883	0.4886	0.4970	0.4987	0.5013
STD2	0.0116	0.0091	0.0089	0.0016	0.0025	0.0022	0.0070	0.0055	0.0054	0.0010	0.0017	0.0013

5. Empirical Application: Efficiency Analysis of U.S. Banking Industry.

5.1 Empirical Models:

In this section, the Bayesian approach suggested in this paper will be applied to illustrate the temporal change in the efficiency levels of 40 of the top 50 banks in the U.S. ranked by their book value of assets. We consider only 40 of these banks due to missing observations and other data anomalies. The empirical model is borrowed from Inanoglu et al. (2012), where a suite of econometric models, including time-invariant panel data models, time-variant models as well as the quantile regression methods, are utilized to examine issues of “too big to fail” in the banking industry.. In this paper, we will only compare results across different time-varying stochastic frontier panel estimators such as the CSS Within and GLS estimators, the BC estimator and the KSS estimator and assess the comparability of inferences among them. The estimators we utilize are based on different assumptions on the functional form of the time varying effects and provide various treatments for the unobserved heterogeneity, but they are all based on Eq.(2.1), which characterizes a single output with panel data assuming unobserved individual effects. Here y_{it} is the response variable (e.g. some measure of bank output like loans), η_{it} represents a bank specific effect, x_{it} is a vector of exogenous variables and v_{it} is the error term. We will estimate second order approximations in logs-the translog specification- to a multi-output/multi-input distance function, see Caves et al. (1982). Let the *m* outputs be $Y_{it} = \exp(x_{it})$ and the *n* inputs $X_{is} = \exp(x_{is})$. Then express the *m*-output, *n*-input deterministic distance function $D_O(Y, X)$ as

$$D_O(Y,X) = \frac{\prod_{j=1}^{m} Y_{it}^{\gamma_j}}{\prod_{k=1}^{n} X_{it}^{\delta_k}} \leq 1 \tag{5.1}$$

The output-distance function $D_O(Y, X)$ is non-decreasing, homogeneous, and convex in Y and non-increasing and quasi-convex in X. After taking logs and rearranging terms we have:

$$-y_{1,it} = \eta_{it} + \sum_{j=2}^{m} \gamma_j y^*_{jit} + \sum_{k=1}^{n} \delta_k x_{kit} + v_{it}, i = 1,\dots,N; t = 1,\dots,T \tag{5.2}$$

where $y^*_{jit, j=2,\dots,m} = \ln(Y_{jit} / Y_{1it})$ and the normalization of homogeneity in outputs is applied to satisfy $\sum_{j=1}^{m} \gamma_j = 1$.

We specify the distance function as translog, consistent with the recommendations of Fare et al. (2013). Their functional equation methods yield the translog specification in the case of the Shephard distance function that we utilize below. Using this flexible parametric functional form we can also test if the data are consistent with the curvature properties required of the distance function. Such theoretical underpinnings and empirical testing options are not available for many of the more recent highly parameterized nonparametric specifications utilizing, for example local linear or quadratic approximations, and for that reason we do not pursue such options in our banking productivity illustration.

With the translog technology applied, the distance function will take the following form in Eq.(5.3).

$$
\begin{aligned}
-y_{1it} &= \eta_{it} + \sum_{j=2}^{m} \gamma_j y^*_{jit} + 1/2 \sum_{j=2}^{m} \sum_{l=2}^{m} \gamma_{jl} y^*_{jit} y^*_{lit} + \sum_{k=1}^{n} \delta_k x_{kit} + 1/2 \sum_{k=1}^{n} \sum_{p=1}^{n} \delta_{kp} x_{kit} x_{pit} \\
&+ \sum_{j=2}^{m} \sum_{k=1}^{n} \theta_{jk} y^*_{jit} x_{kit} + v_{it}, \quad i = 1,\ldots,N; t = 1,\ldots,T
\end{aligned} \tag{5.3}
$$

If we denote $X = [x_{NT\times n}, y^*_{NT\times(m-1)}, xx_{NT\times(n\times(n+1)/2)}, y^* y^*_{NT\times((m-1)\times m/2)}, xy^*_{NT\times(m-1)\times n)}]$, model (5.3) can be written in simplicity to the form in Eq.(2.1).

Elasticities of the distance function with respect to input and output variables (O'Donnell and Coelli, 2005) are expressed as

$$s_p = \delta_p + \sum_{k=1}^{n} \delta_{kp} x_k + \sum_{j=2}^{m} \theta_{pj} y^*_j, \quad p = 1, 2, \ldots, n \tag{5.4}$$

$$r_j = \gamma_j + \sum_{l=2}^{m} \gamma_{jl} y^*_j + \sum_{k=1}^{n} \theta_{kj} x_k, \quad j = 2, \ldots, m \tag{5.5}$$

The individual effects are transformed into relative efficiency levels using the standard order statistics argument given in Schmidt and Sickles (1984) as

$$TE_{it} = \exp\{v_i(t) - \max_{i=1,\ldots n} v_i(t)\} \tag{5.6}$$

For the BC estimator, technical efficiency levels can differ but parsimony is achieved by assuming that all firms have the same temporal pattern. The temporal pattern is specified as

$$TE_{it} = \{\exp[-\varsigma(t-T)]\}\eta_i \tag{5.7}$$

where η_i are independent random effects and ς describes the temporal change pattern.

Clearly the levels of efficiency can vary substantially for the methods that use the order statistics (the firm with the largest effect) to benchmark the most efficient firm and thus the relative efficiencies of the remaining firms. Typically, this impact is mitigated by data trimming but with only 40 firms in our study we decided to avoid doing so in presenting the results below. The BC estimator has no such potential drawback. We will consider such trimming approaches as we examine our models and results more fully.

5.2 Data

The dataset analyzed in this paper is a balanced panel of 40 out of the top 50 U. S. commercial banks based on the yearly data of their Book Value of Assets from 1990 through 2009. The panel size is thus 40 by 20. Missing observations and data anomalies reduced the sample from 50 to 40 firms. The data is merged on a pro-forma basis wherein the non-surviving bank's data is represented as part of the surviving bank going back in time. The three output and six input variables used to estimate the translog output orientated distance function are: Real Estate Loans ("REL"), Commercial and Industrial Loans ("CIL"), Consumer Loans ("CL"), Premises & Fixed Assets ("PFA") , Number of Employees ("NOE"), Purchased Funds ("PF"), Savings Accounts ("SA"), Certificates of Deposit ("CD") and Demand Deposits ("DD"). Additionally, three types of risk proxies are considered: Credit Risk ("CR"), which is approximated by the Gross Charge-off Ratio, Liquidity Risk ("LR"), proxied by Liquidity Ratio, and Market Risk ("MR"), proxied by standard deviation of Trading Returns.

5.3 Empirical Results

The estimation results of the first-order and second-order terms are displayed in Table 6 in Appendix B. Since our dataset is geometric mean corrected (each of the data points have been divided by their geometric sample mean), the second-order term in the elasticities expressed in Eq.(5.4) and Eq.(5.5) will diminish to zero when evaluated at the geometric mean of the sample. The elasticities are displayed in Table 5. In order to select the number of factors for KSS and BE2, we set the limit up to 5 and both of the models favor a two-factors. From Table 5, we can see the elasticity in the input variable Fixed Assets is varying from -0.0448 (KSS) to -0.1267(BC) across different time-varying estimators; the elasticity estimate in Number of Employees is varying from -0.0666 (BE2) to -0.2750 (CSSW); that in Purchase Fund is from -0.0570(BE1) to -0.1387(BE2); the elasticity in Saving Account varies from -0.1026(CSSW) to -0.3058(BC); the elasticity in Certificate of Deposit is from -0.1526(KSS) to -0.2938 (BC); and that in Demand Deposit is from -0.0055(CSSW) to -0.0636 (BE2). As it is shown, the results are on the same order in magnitudes and signs of the elasticity estimates across different models, except that for Demand Deposit, where CSSW gives a significantly lower estimate than all the other estimators. The KSS estimator suggests a slightly lower returns-to-scale estimate as shown in the second but last row in Table 5 since KSS tends to give lower estimates on the Fixed Asset and Certificate of Deposit input elasticities than other models, though the estimates are in the same order. In addition, all the estimators suggest decreasing returns to scale except BC. However, the returns-to-scale estimate suggested by BC is 1.0165, which is not significantly different from 1. Alternatively, we can say that there is no evidence of increasing returns to scales based on the estimation results. For the elasticity estimates in output variables, we notice that the estimates are also similar across estimators.

Table 5: Estimation Results
(Evaluated at Sample Mean)

Model	BC	CSSW	CSSG	KSS	BE1	BE2
PFA	-0.1267	-0.1067	-0.1243	-0.0448	-0.1221	-0.0505
NOE	-0.1518	-0.2750	-0.2731	-0.2195	-0.1520	-0.0666
PF	-0.1088	-0.0571	-0.0628	-0.0679	-0.0570	-0.1387
SA	-0.3058	-0.1026	-0.1413	-0.1289	-0.1700	-0.3042
CD	-0.2938	-0.2422	-0.2492	-0.1526	-0.2363	-0.2867
DD	-0.0295	-0.0055	-0.0297	-0.0321	-0.0259	-0.0636
REL	0.6302	0.6267	0.6254	0.5468	0.6182	0.6099
CIL	0.2674	0.2116	0.2053	0.3200	0.2300	0.2630
CL	0.1024	0.1617	0.1693	0.1332	0.1518	0.1271
RTS	1.0165	0.7891	0.8804	0.6459	0.7634	0.9102
Avg.TE	0.7576	0.6094	0.6608	0.5552	0.4584	0.6937

Although the Bayesian estimators proposed in this paper has produced similar estimates for the slopes elasticities, they have variation in the estimation of the temporal pattern of the individual effects as it is displayed in Figure 1. The BC estimator provides higher efficiency estimates through the time period, while all the other estimators tend to give estimates on efficiencies of similar magnitude. In addition, the BC estimator suggests a declining pattern in the average of the technical efficiency levels. This is probably due to the substantial downturns in the economy and the meltdowns of financial institutions during the recent period of the Great Recession. The average of the technical efficiency levels estimated by other time-varying estimators has displayed a turning point in a certain period. Generally, the estimators considered here have indicated a consensus decrease in efficiency of the largest banks over the last two decade.

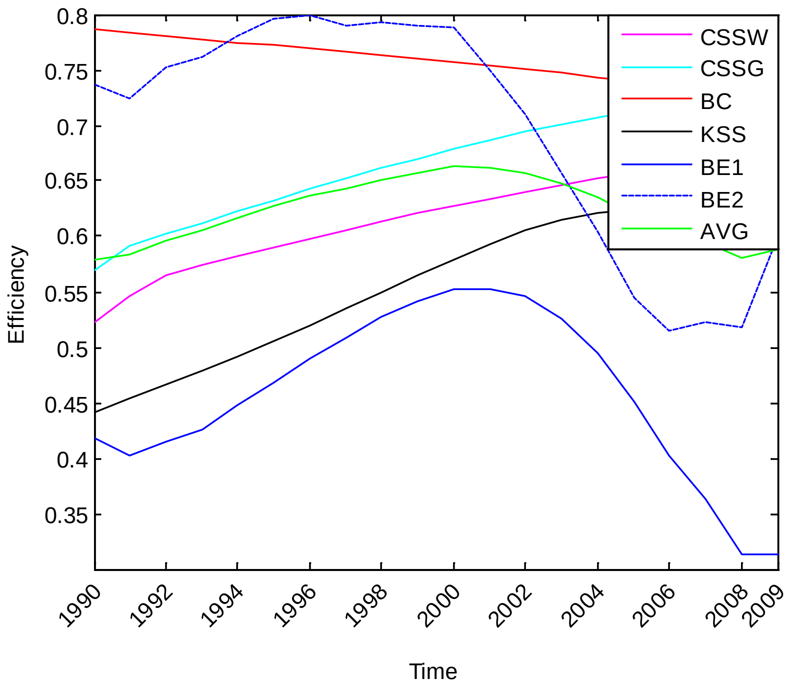

Figure 1: Temporal Pattern of Changes in Average Efficiencies for all Estimators

As we can see from the last row in Table 5 and in Figure 1, the scale of the average technical efficiency levels in the largest U.S. banks suggested by BC is 0.7576, higher than those by the CSSW, CSSG, the KSS and the Bayesian estimators. It ranges from around 0.7267 to 0.7866. The temporal pattern of BC is linearly decreasing, which is consistent with its assumption on the form of the technical efficiencies. The patterns estimated from the CSSW and CSSG estimators have both displayed a turning point at around the year 2005. The KSS estimator provides a similar pattern as CSSW and CSSG but a mild decreasing trend of the technical efficiencies over the recent period. Turning our attention to the estimated temporal pattern of the technical efficiencies using the Bayesian estimator, we notice that the BE1 and BE2 models display similar trends but the efficiency levels suggested by BE2 are consistently higher than those by BE1. They both display an initial slowly increasing pattern in the 1990s, and a sharp kink in the early 2000. After that, the curve is decreasing at a decreasing rate. The increasing trend in efficiency levels at the beginning of 1990s is probably because of the increased competitive pressure in financial industry due to the deregulations since the 1980s. The decreasing trend in efficiency levels started before the Great Recession perhaps because the financial institutions were taking on more risky activities and less focused on their traditional roles as financial intermediaries when the global pool of fixed-income securities increased substantially.

6. Conclusions

This paper has proposed a Bayesian approach to treat time-varying heterogeneity in a panel data model setting. We consider two models: one with nonparametric time effects and the other whose effects are driven by some unknown common factors. In both of the models, we do not impose any parametric assumptions of the individual effects and that we utilize the Gibbs Sampling method to implement the Bayesian inferences.

The Monte Carlo Simulation experiments show that the new Bayesian estimators have displayed consistently superior performance under various data generating processes. On the other hand, the parametric estimators based on some simple functional form assumption on the effects, though allowing for the temporal variations, have the tendency of misspecification on the temporal pattern of the individual effects. Hence, their finite sample performance has been dominated by the Bayesian estimators.

The new Bayesian estimators are applied in analyzing the temporal pattern of the technical efficiencies of the largest 40 U.S. banks over the last two decades (through the 1990 to 2009). It is discovered that the largest banks have experienced a decrease in technical efficiency since early 2000, and a slight recovery after 2008. This can be explained by their tendency to taking on more risky activities at the early 2000s and restrain these risky activities somehow after the Great Recession.

There are several paths for continuing research. An extension of this model is on the prior assumption of the slope parameters; we can consider the restricted region where the slope parameters satisfy reasonable curvature properties under a specific functional form. For example, monotonicity and convexity can be imposed when the translog distance function is specified; thus the assumption on the prior of the slope parameters is reasonable if it restricts the parameter values to a permissible set. The advantage of applying Bayesian method to impose the monotonicity and curvature properties has been elaborated in O'Donnell and Coelli (2005). Our model can also be extended to a panel data discrete choice model, with the effects term modeling the unknown individual heterogeneity.

Appendix A:

1. Detailed Derivation of the conditional posterior distribution of $\gamma_i \mid \beta, \sigma, \omega, Y, X$.

$$
\begin{aligned}
& p\left(\gamma_i \mid \beta, \sigma, \omega, Y, X\right) \propto \sigma^{-(nT+1)} \exp\{-\frac{1}{2\sigma^2}\left(\gamma_i - Y_i + X_i\beta\right)' \left(\gamma_i - Y_i + X_i\beta\right) - \frac{1}{2\omega^2}\gamma_i ' Q\gamma_i\} \\
& \propto \sigma^{-(nT+1)} \exp\{-\frac{1}{2}\gamma_i '(\sigma^{-2}I_T + \omega^{-2}Q)\gamma_i - (Y_i - X_i\beta)'\sigma^{-2}\gamma_i - \gamma_i '\sigma^{-2}(Y_i - X_i\beta) + (Y_i - X_i\beta)'(Y_i - X_i\beta)\} \\
& \propto \sigma^{-(nT+1)} \exp\{-\frac{1}{2}\gamma_i '(\sigma^{-2}I_T + \omega^{-2}Q)\gamma_i - (Y_i - X_i\beta)'\sigma^{-2}\gamma_i - \gamma_i '\sigma^{-2}(Y_i - X_i\beta)\} \\
& \propto \sigma^{-(nT+1)} \exp\{-\frac{1}{2}\gamma_i '(\sigma^2\omega^2 V)^{-1}\gamma_i - (Y_i - X_i\beta)'\omega^2 V(\sigma^2\omega^2 V)^{-1}\gamma_i - \gamma_i '\omega^2 V(\sigma^2\omega^2 V)^{-1}(Y_i - X_i\beta)\} \\
& \propto \sigma^{-(nT+1)} \exp\{-\frac{1}{2}\gamma_i '(\sigma^2\omega^2 V)^{-1}\gamma_i - (Y_i - X_i\beta)'\omega^2 V(\sigma^2\omega^2 V)^{-1}\gamma_i - \gamma_i '\omega^2 V(\sigma^2\omega^2 V)^{-1}(Y_i - X_i\beta) \\
& \qquad + (Y_i - X_i\beta)\omega^2 V '\omega^2 V(Y_i - X_i\beta)\} \\
& \propto \sigma^{-(nT+1)} \exp\{-\frac{1}{2}(\gamma_i - \omega^2 V(Y_i - X_i\beta))'(\sigma^2\omega^2 V)^{-1}(\gamma_i - \omega^2 V(Y_i - X_i\beta))\} \\
& = \sigma^{-(nT+1)} \exp\{-\frac{1}{2}(\gamma_i - \widehat{\gamma_i})'(\sigma^2\omega^2 V)^{-1}(\gamma_i - \widehat{\gamma_i})\}
\end{aligned}
$$

where $\widehat{\gamma_i} = \omega^2 V(Y_i - X_i\beta)$ and $V = (\omega^2 I_T + \sigma^2 Q)^{-1}$

2. Derivations of the posterior distribution of the smoothing parameter ω.

If the smoothing parameter is assumed to follow its the prior distribution: $\frac{\bar{q}}{\omega^2} \sim \chi^2_{\bar{n}}$, or equivalently $p(\omega) \propto (\frac{\bar{q}}{\omega^2})^{\bar{n}/2-1} \exp\{-\frac{\bar{q}}{2\omega^2}\}\omega^{-3} \propto (\frac{\bar{q}}{\omega^2})^{\bar{n}/2+1/2} \exp\{-\frac{\bar{q}}{2\omega^2}\}$

The joint prior will take the form below:

$$
\begin{aligned}
p\left(\beta, \gamma, \sigma, \omega \mid Y, X, \bar{n}, \bar{q}\right) \propto & \ \sigma^{-(nT+1)} \exp\{-\frac{1}{2\sigma^2}\left(Y - X\beta - \gamma\right)' \left(Y - X\beta - \gamma\right)\} \\
& \times \exp\{-\frac{1}{2\omega^2}\gamma'\left(I_n \otimes Q\right)\gamma\} \times (\frac{\bar{q}}{\omega^2})^{\bar{n}/2+1/2} \exp\{-\frac{\bar{q}}{2\omega^2}\}
\end{aligned}
$$

Therefore, the conditional posterior distribution of ω can be derived through the following.

$$
\begin{aligned}
& p\left(\omega \mid \beta, \gamma, \sigma, Y, X, \bar{n}, \bar{q}\right) \propto \exp\{-\frac{1}{2\omega^2}\gamma'\left(I_n \otimes Q\right)\gamma\} \times (\frac{\bar{q}}{\omega^2})^{\bar{n}/2+1/2} \exp\{-\frac{\bar{q}}{2\omega^2}\} \\
& \propto (\frac{\bar{q}}{\omega^2})^{\bar{n}/2+1/2} \exp\{-\frac{\bar{q} + \sum_{i=1}^n \gamma_i ' Q\gamma_i}{2\omega^2}\} \propto (\frac{\bar{q} + \sum_{i=1}^n \gamma_i ' Q\gamma_i}{\omega^2})^{\bar{n}/2+1/2} \exp\{-\frac{\bar{q} + \sum_{i=1}^n \gamma_i ' Q\gamma_i}{2\omega^2}\}
\end{aligned}
$$

Therefore the transformation of the smoothing parameter $\frac{\bar{q} + \sum_{i=1}^n \gamma_i ' Q\gamma_i}{\omega^2}$ follows $\chi^2_{\bar{n}}$.

Appendix B:

Table 6: The Estimation for the Slope Parameters

Model	BC	CSSW	CSSG	KSS	BE1	BE2		BC	CSSW	CSSG	KSS	BE1	BE2
CIL	0.267394	0.211625	0.205296	0.320024	0.229974	0.262958	PF*CD	-0.028282	-0.023417	-0.023420	-0.024378	-0.011996	-0.012073
	(0.015604)	(0.014842)	(0.004009)	(0.016490)	(0.014168)	(0.013784)		(0.018853)	(0.011323)	(0.007121)	(0.009834)	(0.011182)	(0.019844)
CL	0.102395	0.161658	0.169303	0.133170	0.151814	0.127101	PF*DD	-0.114018	-0.017305	-0.024098	-0.004148	-0.015062	-0.101234
	(0.012878)	(0.012244)	(0.003398)	(0.011736)	(0.010868)	(0.010493)		(0.015688)	(0.009595)	(0.006507)	(0.008484)	(0.008648)	(0.017367)
PFA	-0.126714	-0.106713	-0.124307	-0.044849	-0.122111	-0.050466	SA*CD	-0.141683	-0.033535	-0.059756	-0.067716	-0.055219	-0.167241
	(0.031169)	(0.026743)	(0.008180)	(0.023470)	(0.024393)	(0.027912)		(0.031271)	(0.021438)	(0.012105)	(0.019169)	(0.019877)	(0.033330)
NOE	-0.151782	-0.274994	-0.273066	-0.219497	-0.152019	-0.066570	SA*DD	-0.006703	0.053747	0.061960	0.074933	0.036716	0.001257
	(0.035151)	(0.035071)	(0.009826)	(0.030924)	(0.028075)	(0.030319)		(0.030736)	(0.021559)	(0.011642)	(0.019549)	(0.019763)	(0.032234)
PF	-0.108846	-0.057149	-0.062796	-0.067891	-0.057049	-0.138704	CD*DD	-0.097991	-0.105554	-0.098626	-0.057446	-0.092207	-0.119194
	(0.010370)	(0.006407)	(0.003582)	(0.007493)	(0.005713)	(0.010614)		(0.033377)	(0.020702)	(0.013151)	(0.017910)	(0.018797)	(0.036201)
SA	-0.305845	-0.102552	-0.141275	-0.128912	-0.170044	-0.304152	CIL*CIL	0.239373	0.197944	0.207341	0.189705	0.227465	0.287373
	(0.023115)	(0.017762)	(0.005433)	(0.022026)	(0.014980)	(0.016878)		(0.024646)	(0.018416)	(0.006345)	(0.015932)	(0.017940)	(0.019379)
CD	-0.293822	-0.242206	-0.249235	-0.152578	-0.236288	-0.286715	CL*CL	0.113335	0.045183	0.052787	0.016882	0.042151	0.084385
	(0.019988)	(0.013899)	(0.007184)	(0.014208)	(0.013410)	(0.020398)		(0.013263)	(0.010120)	(0.004141)	(0.009309)	(0.008506)	(0.012036)
DD	-0.029454	-0.005520	-0.029726	-0.032132	-0.025869	-0.063642	CIL*CL	-0.065016	-0.045951	-0.048370	-0.040675	-0.032145	-0.058542
	(0.018062)	(0.014840)	(0.005759)	(0.014345)	(0.013910)	(0.016542)		(0.014523)	(0.011902)	(0.004307)	(0.010305)	(0.010754)	(0.012337)
PFA*PFA	-0.058407	-0.076124	-0.064595	-0.027170	0.054452	-0.116818	CIL*PFA	-0.030027	-0.040296	-0.030441	-0.048343	-0.000616	-0.046465
	(0.105551)	(0.081836)	(0.035034)	(0.067810)	(0.079399)	(0.097951)		(0.040094)	(0.029056)	(0.012041)	(0.025079)	(0.030007)	(0.036245)
NOE*NOE	-0.350934	-0.263410	-0.254762	-0.194616	-0.317941	-0.647887	CIL*NOE	0.227956	0.032956	0.037051	0.068312	0.008093	0.245908
	(0.175695)	(0.111222)	(0.049618)	(0.096139)	(0.103994)	(0.170920)		(0.043142)	(0.032479)	(0.016995)	(0.028018)	(0.031391)	(0.046998)
PF*PF	-0.030317	-0.017777	-0.021905	-0.019072	-0.028032	-0.050570	CIL*PF	0.036991	0.066231	0.066914	0.042524	0.044617	0.053115
	(0.009275)	(0.005224)	(0.003626)	(0.004633)	(0.004308)	(0.009989)		(0.011928)	(0.007423)	(0.004722)	(0.006691)	(0.007033)	(0.013064)
SA*SA	0.057266	0.111105	0.088612	0.116956	0.037492	0.051178	CIL*SA	-0.197701	-0.045638	-0.058945	-0.056339	-0.043310	-0.213425
	(0.039891)	(0.031775)	(0.015405)	(0.030207)	(0.026234)	(0.043752)		(0.021737)	(0.015992)	(0.007671)	(0.014339)	(0.016593)	(0.020533)
CD*CD	0.018957	0.063958	0.076793	0.104556	0.065066	0.034873	CIL*CD	0.033169	0.040877	0.040902	0.019851	0.036181	0.021669
	(0.054680)	(0.033776)	(0.020720)	(0.028297)	(0.027536)	(0.057257)		(0.022888)	(0.013701)	(0.009409)	(0.011750)	(0.012349)	(0.025882)
DD*DD	0.008094	0.002895	-0.012089	-0.012085	-0.001555	-0.083644	CIL*DD	-0.106948	-0.048120	-0.056765	-0.016793	-0.042321	-0.137545
	(0.039544)	(0.025435)	(0.014893)	(0.021706)	(0.020225)	(0.043597)		(0.022452)	(0.016474)	(0.008142)	(0.014246)	(0.014006)	(0.021770)
PFA*NOE	-0.112425	0.086399	0.043034	0.070533	0.000583	0.082582	CL*PFA	0.048747	0.037970	0.032106	0.039504	0.007060	0.066231
	(0.102859)	(0.079549)	(0.036284)	(0.066471)	(0.071651)	(0.102329)		(0.027867)	(0.020247)	(0.010162)	(0.017601)	(0.021441)	(0.028297)
PFA*PF	-0.023871	0.001925	0.005802	0.014807	0.032413	0.004723	CL*NOE	-0.134762	-0.080639	-0.079836	-0.073912	-0.026040	-0.121342
	(0.024511)	(0.014311)	(0.009119)	(0.012065)	(0.012272)	(0.025037)		(0.033544)	(0.023290)	(0.012057)	(0.020149)	(0.026992)	(0.034329)
PFA*SA	0.181100	0.065775	0.079467	0.056101	0.069803	0.194046	CL*PF	0.024490	-0.023625	-0.022260	-0.016442	-0.018719	-0.002387
	(0.043433)	(0.033537)	(0.015360)	(0.029335)	(0.032844)	(0.043665)		(0.009238)	(0.005687)	(0.003329)	(0.004883)	(0.005182)	(0.009809)
PFA*CD	-0.191012	-0.036207	-0.035895	-0.098737	-0.156555	-0.235290	CL*SA	0.052008	0.062364	0.064640	0.063839	0.044966	0.064692
	(0.053517)	(0.035121)	(0.020674)	(0.029706)	(0.032240)	(0.055515)		(0.014866)	(0.011813)	(0.005851)	(0.010214)	(0.010948)	(0.015289)
PFA*DD	0.079834	-0.017070	-0.019489	-0.060602	-0.022135	0.000665	CL*CD	0.014655	0.001993	0.006427	0.000114	0.006044	-0.007961
	(0.048869)	(0.030405)	(0.016956)	(0.026224)	(0.027307)	(0.049371)		(0.017719)	(0.011504)	(0.006786)	(0.009937)	(0.011207)	(0.017886)
NOE*PF	0.137728	0.098342	0.099405	0.059733	0.049369	0.116370	CL*DD	-0.057972	0.025662	0.021266	0.027790	-0.001980	-0.044490
	(0.031395)	(0.019230)	(0.012957)	(0.016994)	(0.016952)	(0.035380)		(0.015838)	(0.011538)	(0.005404)	(0.009853)	(0.008936)	(0.016080)
NOE*SA	-0.121524	-0.118107	-0.112644	-0.115822	-0.068949	-0.119951	CR	0.217115	0.697573	0.622777	0.661193	0.650641	0.273590
	(0.065179)	(0.049241)	(0.024082)	(0.042867)	(0.042545)	(0.067800)		(0.207247)	(0.113233)	(0.089828)	(0.096325)	(0.074863)	(0.232643)
NOE*CD	0.417943	0.145744	0.148929	0.183621	0.245112	0.478729	LR	1.103688	0.272672	0.303568	0.359501	0.601731	1.185407
	(0.083620)	(0.050044)	(0.029691)	(0.042715)	(0.045040)	(0.084171)		(0.174812)	(0.176180)	(0.057171)	(0.158809)	(0.152083)	(0.191323)
NOE*DD	0.179106	0.083529	0.084223	0.057347	0.098175	0.329448	MR	-0.002988	-0.001070	-0.000878	-0.000466	0.000008	-0.004905
	(0.064194)	(0.037821)	(0.022992)	(0.032270)	(0.031018)	(0.067150)		(0.002167)	(0.001070)	(0.000974)	(0.000906)	(0.000728)	(0.002496)
PF*SA	0.031111	-0.034051	-0.028221	-0.022330	-0.012928	0.021033							
	(0.016180)	(0.010140)	(0.006517)	(0.008749)	(0.009524)	(0.017828)							

References:

Ahn, S. C., Y. H. Lee, P. Schmidt, 2013, Panel data models with multiple time-varying individual effects." Journal of Econometrics 174, 1-14.

Bai, J., 2009, Panel data models with interactive fixed effects. Econometrica 77, 1229-1279.

Battese, G.E., T.J. Coelli, 1992, Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. Journal of Productivity Analysis 3, 153-169.

Caves, D.W., L.R. Christensen, W.E. Diewert, 1982, The economic theory of index numbers and the measurement of input, output, and productivity. Econometrica 50, 1393-1414.

Chib, S., 1995, Marginal likelihood from the Gibbs output. Journal of the American Statistical Association 90, 1313-1321.

Chambers, R., R. Färe, S. Grosskopf, M. Vardanyan, 2013, Generalized quadratic revenue functions. Journal of Econometrics 173, 11–21.

Cornwell, C., P. Schmidt, R.C. Sickles, 1990, Production frontiers with cross-sectional and time-series variation in efficiency levels. Journal of Econometrics 46, 185-200.

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

Gelman, A., J.B. Carlin, H.S. Stern, D.B. Rubin, 2003, Bayesian data analysis. Chapman & Hall/CRC.

Inanoglu, H., M. Jacobs, J. Liu, R.C. Sickles, 2012, Analyzing bank efficiency: are "too-big-to-fail" banks efficient? Rice University.

Kneip, A., R.C. Sickles, W. Song, 2012, A new panel data treatment for heterogeneity in time trends. Econometric Theory 28, 590-628.

Koop, G., J. Osiewalski, M.F.J. Steel, 1997, Bayesian efficiency analysis through individual effects: hospital cost frontiers. Journal of Econometrics 76, 77-105.

Koop, G., D.J. Poirier, 2004, Bayesian variants of some classical semiparametric regression techniques. Journal of Econometrics 123, 259-282.

O'Donnell, C.J., T.J. Coelli, 2005, A Bayesian approach to imposing curvature on distance functions. Journal of Econometrics 126, 493-523.

Osiewalski, J., M.F.J. Steel, 1998, Numerical tools for the Bayesian analysis of stochastic frontier models. Journal of productivity analysis 10, 103-117.

Schmidt, P., R.C. Sickles, 1984, Production frontiers and panel data. Journal of Business & Economic Statistics 2, 367-374.

Sickles, R. C., J. Hao, C. Shang, 2013, Productivity and panel data. Oxford Handbook of Panel Data, Chapter 17. Badi Baltagi (ed.) New York: Oxford University Press, forthcoming.

Swamy, P., 1970, Efficient inference in a random coefficient regression model. Econometrica: Journal of the Econometric Society, 311-323.

Swamy, P.A.V.B., G.S. Tavlas, 1995, Random coefficient models: theory and applications. Journal of Economic Surveys 9, 165-196.

Tsionas, E.G., 2006, Inference in dynamic stochastic frontier models. Journal of applied econometrics 21, 669-676.
Van den Broeck, J., G. Koop, J. Osiewalski, M.F.J. Steel, 1994, Stochastic frontier models: A Bayesian perspective. Journal of Econometrics 61, 273-303.
Yatchew, A., 1998, Nonparametric regression techniques in economics. Journal of Economic Literature 36, 669-721.