A Panel Data Model with Time-variant Heterogeneity: A Bayesian Treatment with an Application to the Translog Distance Function*

Junrong Liu†

This Draft

October, 2013

Abstract

This paper proposes a Bayesian approach to estimate a panel data model with unobserved heterogeneous individual effects. Minimal assumptions have been imposed on the effect terms. They are allowed to change over time as well as to have any functional form since no functional form is imposed on their prior distribution. Bayesian inference techniques and MCMC methods are applied to implement the model. Monte Carlo experiments are performed to examine the finite-sample performance of this approach and have shown that the method proposed is comparable to the recently proposed estimator of Kneip, Sickles and Song (2012) (KSS) and dominates a variety of estimators that rely on parametric assumptions. In order to illustrate the new method, the Bayesian estimator 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. It is also shown that the monotonicity and curvature constraints implied by economic theory are effectively and straightforwardly imposed using this Bayesian approach.

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

JEL Classification: C23; C11; G21; D24

*This research is developed from Liu, Sickles and Tsionas (2013) working paper, which proposed a Bayesian approach in a panel data model with effects specified in a factor model setting. This paper considers the effects as random function of time, and shows how to use a Bayesian approach to impose functional form constraints in estimating a distance function. I really appreciate all the advisories and suggestions from my advisor Dr. Robin Sickles. Earlier versions of this paper were presented at Texas Econometrics Camp XVIII and Rice University Econometrics Workshop and Brown Bag Seminars. I am also grateful to all the participants for all their comments.

† Department of Economics, Rice University, 6100 Main Street, Houston, TX, 77005; Email: jl29@rice.edu

1. Introduction

The use of panel data methods has proven fruitful to productivity researchers. Many parts of the productivity decomposition exercise rely on panel data and with newer and better panel data techniques productivity researcher are able to make for more robust and general inferences. For an extensive discussion of such estimators and their role in identifying and estimating productivity, see Sickles, Hao and Shang (2013). Early literatures primarily rely on parametric functional form assumptions on the heterogeneous individual effects. For example, Schmidt and Sickles (1984) specified a half-normal distribution for the effects, Greene (1990) extended the specification to a two-parameter gamma distribution. In order to model the unobserved heterogeneous individual effects of a more general form and to avoid the misspecification problem, the work in Park and Simar (1994) and the series of papers of Park, Sickles and Simar (1998, 2003, 2007) focused on semiparametric efficient panel data estimators under a variety of specifications, including a model in which misspecified dynamics may cause the idiosyncratic error to follow autoregressive patterns of various forms and the dynamic panel data model.

Of course such misspecified dynamics may also be due to individual specific heterogeneous dynamic effects; hence, allowing for time variation of the effects is also a popular trend. Cornwell, Schmidt and Sickles (1990) extended the basic fixed effects and random effects models to allow for time-varying efficiency using a quadratic function of time for each cross-sectional unit. Battese and Coelli (1992) considered an exponential specification of time-varying firm effects for panel data using a stochastic frontier production function. However, both of these time-varying estimators have imposed specific parametric assumptions on the time effects. Kneip et al. (2012), on the other hand, treated the time-varying heterogeneity nonparametrically. Their approach is based on a factor model, where the effects are represented by linear combinations of a small number of unknown basis functions with cross-sectional heterogeneous coefficients, i.e. factors and their corresponding factor loadings.

In this paper, a panel data model with unobserved heterogeneous time-varying effects is considered, with individual effects treated as random functions of time. A Bayesian framework is employed to estimate the panel data model. There are several advantages of the Bayesian approach. First, following the Bayesian perspective of random coefficient models (Swamy, 1970), the model will not subjectively assume a common functional form for all the individuals. Swamy and Tavlas (1995) point out that subjective processes may vary among individuals and fixed parametric values of the parameters that describe this functional relationship may not be well-defined. Secondly, 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 theories for the inferences (Koop and Poirier, 2004). Moreover, it becomes feasible to impose the monotonicity and other curvature properties implied by economic theory to estimate typical production, cost, or distance functions within a Bayesian framework without resorting to linear and non-linear programming methods (O'Donnell and Coelli, 2005). For example, the Metropolis-Hastings algorithm can be used to impose the monotonicity and curvature properties consistent with economic theory. It is very intuitive and efficient and gives Bayesian approaches much to distinguish them as compared to frequentist approaches.

Bayesian approaches to integrate panel data methods and stochastic frontier analysis were first suggested by Van den Broeck, Koop, Osiewalski and Steel (1994), who consider a Bayesian approach under the composed error model. Koop, Osiewalski and Steel (1997) has used Bayesian methods for both fixed and random effect models; 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. This assumption can be inappropriate in many settings; for example, in the stochastic frontier analysis, the firms' 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. Our paper, like most papers in the literature, 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. 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 is 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 with minimal smoothness assumption. It will be shown in Section 4 that the Bayesian estimator proposed here consistently outperforms some representative parametric as well as nonparametric estimators under various scenarios of data generating processes.

One of the main criticisms of Bayesian methods lies in its incorporation of the subject prior information. However, in this paper, the only prior information is with regard to the smoothness parameter of the heterogeneous effect.

The rest of this paper is organized as follows. Section 2 describes the basic model setup and parameter priors. The Bayesian inference procedures are explained in section 3, followed by section 4, 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 are presented in Section 5. The multi-output/multi-input translog output distance function is adopted and the monotonicity, quasi-convexity in inputs, convexity in outputs properties are discussed and imposed using the Bayesian estimator. The new estimator with the Bayesian method to imposing curvature properties is applied to the U.S. banking industry. The largest U.S. commercial banks are evaluated for their efficiencies of providing intermediation services. Section 6 provides concluding remarks.

2. The Model

The 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 $(Y_{it}, X_{it}),\ i=1,\dots,n;\ t=1,\dots,T$, where the index i denotes the ith individual units, and the index t denotes the t th time period.

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

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

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 reasonable in many applications; in particular this independence assumption can be validated in the scenario where the technical efficiency levels in different firms rely mainly on their own heterogeneous factors such as size, CEO's managerial skills and operational structure.

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

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

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

When applying our model in 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 no less important than the estimation of the slope parameters.

The difference of our model from those in the literature is that no specific parametric form for the unobserved heterogeneous individual effects is imposed, which follows the assumption in Kneip et al. (2012). However, we will not resort to the traditional nonparametric regression techniques to estimate the model as they did in that paper; instead a Markov Chain Monte Carlo method 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 effects that are individual-specific and time-varying as well. Moreover, it does not rely on the conjugate prior formulation for the time varying individual-effects since the conjugate prior assumption has been criticized to be too subjectively informative.

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:

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

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

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

where $Q = D'D$, and D is the $\left(T-1\right)\times T$ 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 and given beforehand, 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-derivative smoothness 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 $D^{\left(2\right)}\gamma_i \overset{IID}{\sim} N\left(0,\omega^2 I_{T-1}\right)$ 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).

$$p(\beta,\sigma^2) \propto \sigma^{-2} \tag{2.6}$$

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:

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

According to the model setup in Eq.(2.1) and 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 with the information from the dataset, thus the joint posterior distribution is derived in Eq.(2.9).

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

3. Bayesian Inference

To proceed with further inference, we need to solve the posterior distribution above in Eq.(2.9) analytically; however, this posterior is not of standard form, and taking draws directly from it would be problematic. Therefore, Markov Chain Monte Carlo techniques are considered to implement the inference for the model. Specifically, Gibbs sampling (also called alternating conditional sampling) will be used to perform the Bayesian inference. The Gibbs sampler is commonly used under Bayesian inference for multi-dimensional problems because of the desirable results that iterative sampling from the conditional distributions of the sub-vectors of the parameter vector will lead to a sequence of random draws 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, Carlin, Stern and Rubin (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 sub-vector of the parameters in turn conditional on the most recently updated values of all the others, which are the current iteration values for components already updated and the previous iteration values for others. For example, in iteration t, $\beta^{(t)}$ is drawn from $p(\beta \mid \sigma = \sigma^{(t-1)}, \gamma = \gamma^{(t-1)})$, $\sigma^{(t)}$ is drawn from $p(\sigma \mid \beta = \beta^{(t)}, \gamma = \gamma^{(t-1)})$, and $\gamma^{(t)}$ is drawn from $p(\gamma \mid \beta = \beta^{(t)}, \sigma = \sigma^{(t)})$. This sampling process generates 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.

$$\begin{aligned} & p(Y \mid \beta,\gamma,\sigma) \propto \sigma^{-NT} \exp\{-\frac{1}{2\sigma^2}(Y - X\beta - \gamma)'(Y - X\beta - \gamma)\} \\ & = \sigma^{-NT} \exp\{-\frac{1}{2\sigma^2}[(Y - X\beta - \gamma)'(Y - X\beta - \gamma) + (\beta - \beta)'(X'X)(\beta - \beta)]\} \end{aligned} \tag{3.1}$$

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

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

$$\begin{aligned} & p(\beta,\gamma,\sigma \mid Y,X,\omega) \propto \sigma^{-(nT+1)} \exp\{-\frac{1}{2\omega^2}\gamma'(I_n \otimes Q)\gamma\} \\ & \times \exp\{-\frac{1}{2\sigma^2}[(Y - X\beta - \gamma)'(Y - X\beta - \gamma) + (\beta - \beta)'(X'X)(\beta - \beta)]\} \end{aligned} \tag{3.2}$$

Thus, the conditional distribution of β follows the multivariate normal distributions with mean $\hat{\beta}$ and covariance matrix $\sigma^2 (X'X)^{-1}$ since the following distribution is derived from Eq.(3.2).

$$p(\beta \mid Y,X,\gamma,\sigma,\omega) \propto \exp\{-\frac{1}{2\sigma^2}(\beta - \beta)'(X'X)(\beta - \beta)\} \tag{3.3}$$

$$\beta \mid \sigma,\gamma,\omega,Y,X \propto f_k(\beta \mid \hat{\beta}, \sigma^2 (X'X)^{-1}) \tag{3.4}$$

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

$$
\begin{aligned}
& p(\beta,\gamma,\sigma \mid Y,X,\omega) \propto \\
& \sigma^{-(nT+1)} \exp\{-\frac{1}{2\sigma^2}(\gamma - Y + X\beta)'(\gamma - Y + X\beta) - \frac{1}{2\omega^2}\gamma'(I_n \otimes Q)\gamma\} \propto \\
& \sigma^{-(nT+1)} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\gamma_i - Y_i + X_i\beta)'(\gamma_i - Y_i + X_i\beta) - \frac{1}{2\omega^2}\sum_{i=1}^{n}\gamma_i{}'Q\gamma_i\}
\end{aligned}
\tag{3.5}
$$

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.(3.6). The detailed derivation is presented in Appendix A.

$$
\gamma_i \mid \beta,\sigma,\omega,Y,X \propto f_T(\gamma_i \mid \hat{\gamma}_i, \sigma^2\omega^2 V) \tag{3.6}
$$

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

Writing the conditional posterior distribution in the form given by Eq.(3.7), it is clear that the sum of the squared residuals over the unobserved variance $(Y - X\beta - \gamma)'(Y - X\beta - \gamma)/\sigma^2$ has the a Chi-squared distribution with *nT* degree of freedom as shown in Eq.(3.8).

$$
p(\sigma^2 \mid \beta,\gamma,Y,X,\omega) \propto (\sigma^{-2})^{nT/2-1}\exp\{-\frac{1}{2\sigma^2}(Y - X\beta - \gamma)'(Y - X\beta - \gamma)\} \tag{3.7}
$$

$$
\frac{(Y - X\beta - \gamma)'(Y - X\beta - \gamma)}{\sigma^2} \mid \beta,\gamma,\omega,Y,X \sim \chi^2_{nT} \tag{3.8}
$$

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 $\frac{\bar{q}}{\omega^2} \sim \chi^2_{\bar{n}}$, where $\bar{n}, \bar{q} \geq 0$ hyperparameters, the conditional posterior distribution of ω^2 is derived as:

$$
\frac{\bar{q} + \sum_{i=1}^{n}\gamma_i' Q\gamma_i}{\omega^2} \mid \beta,\sigma,\gamma,Y,X \sim \frac{\bar{q} + \sum_{i=1}^{n}\gamma_i' Q\gamma_i}{\omega^2} \mid \gamma,Y,X \sim \chi^2_{\bar{n}+n} \tag{3.9}
$$

Obviously, the hyperparameters $\bar{n}$ and $\bar{q}$ control the prior degree of smoothness that is imposed upon the γ_{it} s. Generally, small values of the prior "sum of squares" $\bar{q}/\bar{n}$

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 $\{\theta_{-i,\omega}^{(s)}, s = 1, \ldots, S\}$ from $(\theta_{-i} \mid Y_{-i}, \omega)$. It is easy to compute the posterior means $\bar{\theta}_{-i,\omega} = S^{-1} \sum_{s=1}^{S} \theta_{-i,\omega}^{(s)}$ and, as a result, the cross validation statistic is

$$CV(\omega) = (nT)^{-1} \sum_{i=1}^{n} (y_i - X_i \bar{\beta}_{-i,\omega} - \bar{\gamma}_{-i,\omega})'(y_i - X_i \bar{\beta}_{-i,\omega} - \bar{\gamma}_{-i,\omega}) \qquad (3.10)$$

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 $\{\theta^{(s)}, s = 1, \ldots, S\}$ from a distribution with kernel density $g(\theta)$ is available and if the existing sample is resampled with probabilities $W_s = \dfrac{f(\theta^{(s)}) / g(\theta^{(s)})}{\sum_{r=1}^{S} f(\theta^{(r)}) / g(\theta^{(r)})}$, 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 $w_s = \sigma^{(s)T} \exp\left[\dfrac{1}{2\sigma^{(s)2}} (y_i - X_i \beta^{(s)} - \gamma_i^{(s)})'(y_i - X_i \beta^{(s)} - \gamma_i^{(s)}) + \dfrac{1}{2\omega^2} \gamma_i^{(s)\prime} Q \gamma_i^{(s)}\right]$, 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 $\{\beta_{-i}^{(s)}, \sigma_{-i}^{(s)}, \gamma_{-i}^{(s)}\}$, 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.(3.4) to Eq.(3.8) 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.(3.4), 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.(3.6).

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 $Y - X\beta - \gamma\ '\ Y - X\beta - \gamma\ \ / \sigma^2$, which is shown in Eq.(3.7) to follow a chi-squared distribution with degree of freedom *nT*. Secondly, assign the values of $Y - X\beta - \gamma\ '\ Y - X\beta - \gamma\ \ / Chi - rnd$ to σ^2 , where $Chi - rnd$ is the generated random variable that follows χ^2_{nT} in the first step.

4. Monte Carlo Simulations

To illustrate the model and inspect the finite sample performance of the new estimator using the Bayesian estimator with nonparametric individual effects specification (BE 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 independently and identically drawn from *N(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: $\gamma_{it} = \nu_{i1} t/T \cos\left(4\pi t/T\right) + \nu_{i2} t/T \sin\left(4\pi t/T\right)$

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

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

DGP1 specifies a time-varying effect based on a second-order polynomial of the time trend, which is varying across units, DGP1 is considered to model those firms improving their efficiencies over the time smoothly; DGP2 utilizes the effect as a random walk process, modeling the firms whose efficiency levels are experiencing small ups and downs from time to time; 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 35,000 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.

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 lower panel of every table. We will see that throughout the DGPs, the estimation results for the slope coefficients β have no significant difference across different estimators, however the effects term estimation does vary largely across different estimators. For the individual effects γ_{it}, MSE results are displayed at the upper panel of each table. The normalized MSE formula of the individual effects γ_{it} is calculated in(4.1).

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

Table 1 : Monte Carlo Simulations for DGP1

Mean Squared Error for the Individual Effects						
n	T	BC	CSSW	CSSG	KSS	BE
50	20	0.7284	0.0012	0.0012	0.0039	0.0051
	50	0.9371	0.0005	0.0005	0.1255	0.0065
100	20	0.8222	0.0008	0.0008	0.0033	0.0027
	50	0.8245	0.0003	0.0003	0.0220	0.0033
200	20	0.8451	0.0008	0.0008	0.0023	0.0029
	50	0.8823	0.0003	0.0003	0.0021	0.0033

Estimate and Standard Error for the Slope Coefficients										
	T = 20					T = 50				
	BC	CSSW	CSSG	KSS	BE	BC	CSSW	CSSG	KSS	BE
n =50										
EST1	0.5250	0.4961	0.4965	0.4954	0.4981	0.5105	0.4991	0.4992	0.4999	0.4984
SE1	0.0130	0.0033	0.0032	0.0029	0.0039	0.0073	0.0021	0.0020	0.0019	0.0029
EST2	0.4856	0.4949	0.4948	0.4919	0.4917	0.4969	0.5048	0.5047	0.5053	0.5062
SE2	0.0139	0.0035	0.0033	0.0031	0.0042	0.0073	0.0020	0.0020	0.0018	0.0027
n=100										
EST1	0.4973	0.5018	0.5013	0.5045	0.5019	0.4843	0.4999	0.4998	0.4991	0.4994
STD1	0.0099	0.0023	0.0022	0.0021	0.0027	0.0066	0.0014	0.0014	0.0014	0.0019
EST2	0.5047	0.5009	0.5012	0.5022	0.5008	0.4995	0.5001	0.5000	0.4990	0.4994
STD2	0.0098	0.0022	0.0022	0.0020	0.0026	0.0066	0.0014	0.0014	0.0013	0.0019
n=200										
EST1	0.4936	0.5013	0.5015	0.5009	0.5018	0.5000	0.5007	0.5007	0.5000	0.5013
STD1	0.0071	0.0016	0.0016	0.0016	0.0021	0.0042	0.0010	0.0010	0.0010	0.0014
EST2	0.4983	0.5016	0.5020	0.5019	0.5025	0.5027	0.4972	0.4972	0.4969	0.4966
STD2	0.0071	0.0016	0.0016	0.0016	0.0021	0.0042	0.0010	0.0010	0.0010	0.0014

DGP1 is consistent with the assumptions for the time-varying effects in the CSS model, which assumes that the effects are improving gradually along time. 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 BE 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 BE provides more accurate estimation on the individual effects than the KSS estimator. This implies that the performance of the Bayesian Estimator is quite efficient in estimating time-varying effects of the smoothing-curve forms, like the second-order polynomials. It is not surprising that the results of the Bayesian estimator are much better 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	BE
50	20	0.9202	0.1266	0.1266	0.0182	0.0069
	50	0.9052	0.2996	0.2996	0.0238	0.0048
100	20	0.8588	0.4553	0.4553	0.0531	0.0368
	50	0.9884	0.1065	0.1065	0.0046	0.0020
200	20	0.9183	0.6376	0.6375	0.0706	0.0437
	50	0.9526	0.0616	0.0616	0.0028	0.0008

Estimate and Standard Error for the Slope Coefficients										
	T = 20					T = 50				
	BC	CSSW	CSSG	KSS	BE	BC	CSSW	CSSG	KSS	BE
n =50										
EST1	0.4786	0.4857	0.4904	0.5059	0.5050	0.4820	0.4811	0.4938	0.4972	0.4937
SE1	0.0460	0.0308	0.0298	0.0136	0.0237	0.0243	0.0230	0.0227	0.0059	0.0151
EST2	0.4664	0.4414	0.4854	0.4599	0.4534	0.4840	0.4660	0.4848	0.4988	0.4928
SE2	0.0491	0.0326	0.0314	0.0146	0.0253	0.0241	0.0226	0.0225	0.0059	0.0145
n=100										
EST1	0.4854	0.4818	0.4898	0.5065	0.4968	0.5137	0.5360	0.5089	0.4950	0.4990
STD1	0.0200	0.0195	0.0188	0.0075	0.0151	0.0415	0.0257	0.0254	0.0055	0.0119
EST2	0.5005	0.4996	0.5115	0.5101	0.5050	0.4482	0.5283	0.5143	0.5127	0.5231
STD2	0.0198	0.0189	0.0186	0.0073	0.0146	0.0415	0.0256	0.0254	0.0055	0.0120
n=200										
EST1	0.5051	0.4995	0.5015	0.4864	0.4875	0.4274	0.5097	0.4968	0.5018	0.5060
STD1	0.0169	0.0175	0.0171	0.0067	0.0131	0.0527	0.0202	0.0200	0.0032	0.0070
EST2	0.4895	0.4898	0.5147	0.4951	0.4885	0.3996	0.4930	0.5015	0.5042	0.5045
STD2	0.0170	0.0175	0.0171	0.0067	0.0130	0.0531	0.0204	0.0202	0.0033	0.0071

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 much worse performance than in DGP1. The BC estimator is also expected to perform poorly on the estimation of the individual effects. However, the BE and KSS estimator 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 BE 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 estimator does in any panel size combination.

Table 3: Monte Carlo Simulations for DGP3

Mean Squared Error for the Individual Effects						
n	T	BC	CSSW	CSSG	KSS	BE
50	20	3.3477	0.8816	0.8816	0.0130	0.0226
	50	3.3639	0.8469	0.8468	0.0082	0.0195
100	20	3.5102	0.8309	0.8303	0.0123	0.0254
	50	3.7625	0.8357	0.8356	0.0072	0.0178
200	20	3.8433	0.8335	0.8333	0.0121	0.0271
	50	3.8513	0.8393	0.8392	0.0063	0.0163

Estimate and Standard Error for the Slope Coefficients										
	T = 20					T = 50				
	BC	CSSW	CSSG	KSS	BE	BC	CSSW	CSSG	KSS	BE
n =50										
EST1	0.5277	0.5250	0.4994	0.4989	0.5096	0.4868	0.4871	0.4976	0.5005	0.4982
SE1	0.0188	0.0203	0.0197	0.0029	0.0074	0.0122	0.0122	0.0120	0.0018	0.0030
EST2	0.4905	0.4998	0.5062	0.4930	0.4963	0.5259	0.5255	0.5207	0.5052	0.5062
SE2	0.0198	0.0215	0.0207	0.0031	0.0079	0.0121	0.0120	0.0119	0.0018	0.0028
n=100										
EST1	0.4816	0.4768	0.4998	0.5030	0.4957	0.4877	0.4863	0.4972	0.4986	0.4986
STD1	0.0132	0.0139	0.0134	0.0021	0.0054	0.0076	0.0077	0.0076	0.0013	0.0020
EST2	0.4907	0.4816	0.5088	0.5028	0.4936	0.5024	0.5118	0.5089	0.4993	0.5002
STD2	0.0131	0.0135	0.0133	0.0021	0.0052	0.0076	0.0077	0.0076	0.0013	0.0019
n=200										
EST1	0.5120	0.5103	0.5110	0.5016	0.5028	0.4976	0.5012	0.4962	0.4999	0.5016
STD1	0.0088	0.0091	0.0089	0.0016	0.0037	0.0055	0.0054	0.0054	0.0010	0.0015
EST2	0.4885	0.4892	0.5019	0.5029	0.4996	0.4874	0.4883	0.4957	0.4973	0.4958
STD2	0.0088	0.0091	0.0089	0.0016	0.0036	0.0055	0.0055	0.0054	0.0010	0.0015

DGP3 is considered to characterize the significant time variations in individual effects. This DGP can capture some start-up companies efficiency changing trend, which can be largely fluctuate due to their immature management and the high turn-over rate in employees. As we can see from Table 3, the performances of our Bayesian Estimator (BE) in the estimation on both the slope parameters and the individual effects are comparable to those of the KSS estimator. Other estimators, whose effects rely on parametric assumptions of simple functional forms, are to a great extent dominated by the BE.

DGP 4 can be considered a mixed scenario of those from the first three DGPs. It is shown in Table 4 that the BE dominantly outperforms the BC, CSSW, and CSSG estimators in terms of the MSE of the individual effects. Similar to the case in DGP1 and DGP3, the Bayesian Estimator (BE) has comparable performance to the KSS estimator and outperforms it in the cases of the combination n =50, T = 50, and n = 100, T = 50, or when n is not considerably greater than T.

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 estimator (BE) is an excellent candidate 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	BE
50	20	0.8042	0.2161	0.2161	0.0030	0.0125
	50	0.9478	0.2056	0.2056	0.0890	0.0030
100	20	0.8770	0.1382	0.1382	0.0026	0.0101
	50	0.8626	0.1337	0.1337	0.0193	0.0021
200	20	0.8764	0.1301	0.1301	0.0020	0.0094
	50	0.9111	0.1445	0.1445	0.0015	0.0022

Estimate and Standard Error for the Slope Coefficients										
	T = 20					T = 50				
	BC	CSSW	CSSG	KSS	BE	BC	CSSW	CSSG	KSS	BE
n =50										
EST1	0.5521	0.5250	0.5329	0.4995	0.5154	0.5031	0.4871	0.4901	0.4999	0.4977
SE1	0.0233	0.0203	0.0197	0.0030	0.0095	0.0148	0.0122	0.0120	0.0019	0.0027
EST2	0.4788	0.4998	0.5014	0.4907	0.4957	0.5201	0.5255	0.5246	0.5053	0.5062
SE2	0.0248	0.0215	0.0207	0.0031	0.0101	0.0147	0.0120	0.0119	0.0018	0.0026
n=100										
EST1	0.4732	0.4768	0.4713	0.5017	0.4922	0.4720	0.4863	0.4867	0.4985	0.4982
STD1	0.0169	0.0139	0.0134	0.0022	0.0068	0.0103	0.0077	0.0076	0.0014	0.0019
EST2	0.4880	0.4816	0.4836	0.5018	0.4910	0.5077	0.5118	0.5117	0.4998	0.5007
STD2	0.0167	0.0135	0.0133	0.0021	0.0066	0.0103	0.0077	0.0076	0.0014	0.0018
n=200										
EST1	0.5029	0.5103	0.5112	0.5011	0.5044	0.4891	0.5012	0.5012	0.5003	0.5017
STD1	0.0116	0.0091	0.0089	0.0016	0.0045	0.0069	0.0054	0.0054	0.0010	0.0014
EST2	0.4892	0.4892	0.4934	0.5021	0.4982	0.4940	0.4883	0.4886	0.4970	0.4953
STD2	0.0116	0.0091	0.0089	0.0016	0.0045	0.0070	0.0055	0.0054	0.0010	0.0014

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, Jacobs, Liu and Sickles (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, Christensen and Diewert (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 $D_o(Y,X)$ allows us to describe the multi-input/multi-output production technology without specifying a behavioral objective and it should follow the a few properties such as non-decreasing, homogeneous, and convex in Y and non-increasing and quasi-convex in X (Battese, Coelli and Rao, 1998).

After taking logs and rearranging terms from Eq. (5.1) we have:

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

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

A Cobb-Douglas specification to the distance function as proposed by (Klein, 1953) can be used as a valid first-order approximation. However, we specify the distance function as translog since the Cobb-Douglas distance function has been criticized for its assumption of separability of outputs and inputs and for incorrect curvature as the production possibility frontier is convex instead of concave. The translog function has more advantages such as the second-order approximation allowing for more flexibility, proper local curvature in the productivity possibility curve, and the separability of outputs and inputs. If the translog technology is 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).

When the translog distance function is applied, it is natural to consider its monotonicity and curvature properties implied by production theory: the output distance is non-decreasing in inputs, non-increasing in outputs, quasi-convex in inputs and convex in outputs.

Elasticities of the output distance function with respect to input and output variables after taking the first derivatives are expressed as the following forms. The notations in this section follow largely from O'Donnell and Coelli (2005).

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

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

Monotonicity implies that D to be non-increasing in x such that

$$f_p = \frac{\partial D_o}{\partial x_p} = \frac{\partial \ln D_o}{\partial \ln x_p}\frac{D_o}{x_p} = s_p \frac{D_o}{x_p} \le 0 \Leftrightarrow \mathrm{s}_p \le 0, \quad p = 1,2,\ldots,n \tag{5.6}$$

For D to be non-decreasing in y, we must have the following conditions:

$$h_j = \frac{\partial D_o}{\partial y_j} = \frac{\partial \ln D_o}{\partial y_j}\frac{D_o}{y_j} = r_j \frac{D_o}{y_j} \geq 0 \Leftrightarrow r_j \geq 0, \quad j = 2,...,m \tag{5.7}$$

The sufficient condition for D to be quasi-convex in x over the non-negative orthant is that all the principal minors for the bordered Hessian matrix must be negative. The bordered Hessian matrix, which includes the first-derivatives and second-derivatives of D, takes the following form:

$$F = \begin{pmatrix} 0 & f_1 & \cdots & f_n \\ f_1 & f_{11} & \cdots & f_{1n} \\ \vdots & \vdots & \cdots & \vdots \\ f_p & f_{1n} & \cdots & f_{nn} \end{pmatrix} \tag{5.8}$$

Each entry of the second-derivatives $f_{pk} = \frac{\partial^2 D_o}{\partial x_p \partial x_k} = (\delta_{pk} + s_p s_k - I\{p = k\} s_p)(D_o / x_p x_k)$

and $I\{p = k\}$ is an indicator function taking the value 1 if p = k and 0 otherwise. The details of the quasi-convexity checking criteria can be found in Lau (1978).

To ensure the distance function D convexity in output y over the non-negative orthant, we must have the following Hessian matrix H is positive semi-definite.

$$H = \begin{pmatrix} h_{11} & \cdots & h_{1m} \\ \vdots & \ddots & \vdots \\ h_{m1} & \cdots & h_{mm} \end{pmatrix} \tag{5.9}$$

Each entry of the Hessian matrix above $h_{kl} = \frac{\partial^2 D_o}{\partial y_k \partial y_l} = (\gamma_{kl} + r_k r_l - I\{k = l\} r_k)(D_o / y_k y_l)$

With the monotonicity and curvature properties imposed, the joint prior described in Eq.(2.9) has become to the following form:

$$\begin{aligned} p \;\; \beta,\gamma,\sigma \mid Y,X,\omega \;\; &\propto \sigma^{-\;nT+1} \exp\{-\frac{1}{2\sigma^2} \;\; Y - X\beta - \gamma \;\; ' \;\; Y - X\beta - \gamma \;\; \} \\ &\times \exp\{-\frac{1}{2\omega^2}\gamma' \;\; I_n \otimes Q \;\; \gamma\} \times I[\beta \in R_j] \end{aligned} \tag{5.10}$$

The corresponding conditional posteriors becomes to:

$$\beta \mid \sigma,\gamma,\omega,Y,X \propto f_k \;\; \beta \mid \hat{\beta}, \; \sigma^2 \;\; X'X \;\; ^{-1} \;\; \cdot I[\beta \in R_j] \tag{5.11}$$

$$\gamma_i \mid \beta,\sigma,\omega, \;\; \gamma_j, j \neq i \;\; ,Y,X \sim \gamma_i \mid \beta,\sigma,\omega,Y,X \propto f_T \;\; \gamma_i \mid \hat{\gamma}_i, \; \sigma^2\omega^2 V \tag{5.12}$$

$$\frac{(Y - X\beta - \gamma)'(Y - X\beta - \gamma)}{\sigma^2} \mid \beta, \gamma, \omega, Y, X \sim \chi^2_{nT} \tag{5.13}$$

The indicator function $I[\beta \in R_j]$ takes the value 0 when the slope parameter β falls in the restricted region or not corresponding to R_j; j = 0, 1. To impose the properties, we need to sampling from the truncated multivariate normal distribution described in Eq.(5.11). A Metropolis-Hastings algorithm has been applied within the Gibbs Sampling in our application to impose the monotonicity and curvature constraints implied by production economics theory. A description of the M-H method can be found in (Gelman et al., 2003). It is discussed in (O'Donnell and Coelli, 2005) that the M-H algorithm is more efficient compared to the rejection algorithm. The detailed algorithm follows mainly from (Griffiths, O'Donnell and Cruz, 2000).

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

$$TE_{it} = \exp\{v_i(t) - \max_{i=1,\ldots n} v_i(t)\} \tag{5.14}$$

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

$$TE_{it} = \{\exp[-\varsigma(t-T)]\}\eta_i \tag{5.15}$$

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 as control variables. The three different types of risks include Credit Risk ("CR"), 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.

Using the 40 by 20 panel data set, the monotonicity and curvature conditions have been checked, and it is found that more than half of the observations have violated the first seven monotonicity constraints, more than 90% have violated the last monotonicity constraint, and all observations in the data set have violated the curvature constraints. Therefore, in the following sub-sections, the estimation results are presented with and without constraints for the Bayesian estimators.

5.3 Empirical Results

The full 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 estimated using BC, CSSW, CSSG, KSS and Bayesian Estimator without imposing constraints (BE0) and Bayesian estimator with constraints (BE1) are reported in Table 5. We can see the elasticity in the input variable Fixed Assets is varying from -0.0448 (KSS) to -0.1411(BE1) across different time-varying estimators; the elasticity estimate in Number of Employees is varying from -0.1518 (BC) to -0.2750 (CSSW); that in Purchase Fund is from -0.0570(BE0) to -0.1088(BC); 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 -0.0321(KSS). 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 fourth 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. Given the results from six different estimators, 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		BE0		BE1	
PFA	-0.1267	(0.0871)	-0.1067	(0.0858)	-0.1243	(0.0871)	-0.0448	(0.0861)	-0.1221	(0.0244)	-0.1411	(0.0105)
NOE	-0.1518	(0.1295)	-0.2750	(0.0997)	-0.2731	(0.1295)	-0.2195	(0.1110)	-0.1520	(0.0281)	-0.1694	(0.0094)
PF	-0.1088	(0.0240)	-0.0571	(0.0245)	-0.0628	(0.0240)	-0.0679	(0.0355)	-0.0570	(0.0057)	-0.0682	(0.0042)
SA	-0.3058	(0.0593)	-0.1026	(0.0667)	-0.1413	(0.0593)	-0.1289	(0.1473)	-0.1700	(0.0150)	-0.2319	(0.0100)
CD	-0.2938	(0.0624)	-0.2422	(0.0630)	-0.2492	(0.0624)	-0.1526	(0.0689)	-0.2363	(0.0134)	-0.2465	(0.0117)
DD	-0.0295	(0.0596)	-0.0055	(0.0526)	-0.0297	(0.0596)	-0.0321	(0.0637)	-0.0259	(0.0139)	-0.0629	(0.0050)
REL	0.6302	(0.0552)	0.6267	(0.0507)	0.6254	(0.0552)	0.5468	(0.0545)	0.6182	(0.0139)	0.6229	(0.0074)
CIL	0.2674	(0.0509)	0.2116	(0.0517)	0.2053	(0.0509)	0.3200	(0.0567)	0.2300	(0.0142)	0.2119	(0.0052)
CL	0.1024	(0.0537)	0.1617	(0.0567)	0.1693	(0.0537)	0.1332	(0.0405)	0.1518	(0.0109)	0.1651	(0.0060)
RTS	1.0165	(0.0343)	0.7891	(0.1155)	0.8804	(0.0343)	0.6459	(0.1445)	0.7634	(0.0266)	-0.9199	(0.0067)

Although the Bayesian estimator 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 decade.

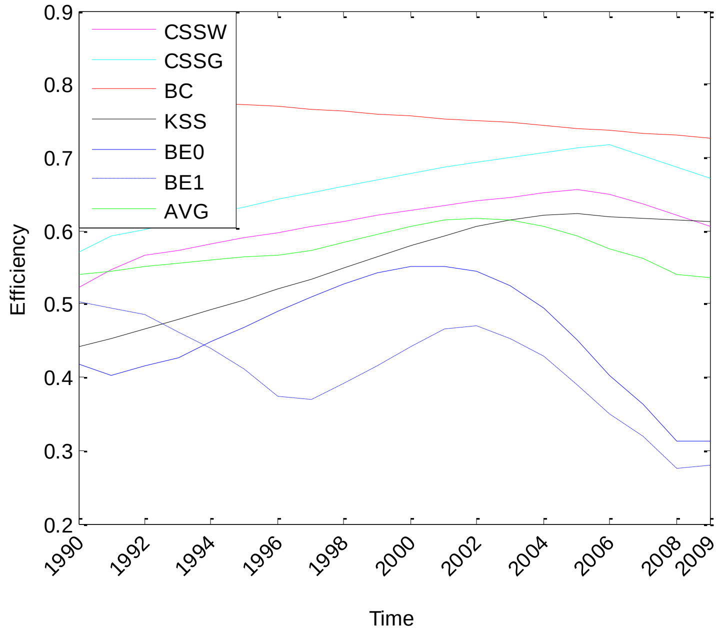

Figure 1: Temporal Pattern of Changes in Average Efficiencies for all Estimators

As we can see from Figure 1, the scale of the average technical efficiency levels in the largest U.S. banks suggested by BC at 0.7576, much 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 estimators with and without constraints, we find that the pattern exhibits more variation. For the Bayesian estimator without constraints (BE0), it displays an initial slowly increasing pattern in the 1990s and a turning point at around 2001. After that, the curve is decreasing more sharply than it is increasing before the turning point until 2008. For the Bayesian estimator with constraints (BE1), it displays more variation along time. In the early 1990s, it shows a mild decreasing pattern from 0.5 to 0.4; it then displays an increase in efficiency from 1996 to 2002; after that, the decreasing pattern is quite similar compared to that shown with BE0.

The decreasing trend in efficiency levels at the beginning of 1990s is probably because of the increased competitiveness in financial industry due to the deregulations since the 1980s. After the banks have adjusted to the competition environment, they may gain some production efficiency back. The decreasing trend in efficiency levels around 2001 is probably due to the Gramm–Leach–Bliley Act (aka the Financial Services Modernization Act of 1999) enacted November, 1999. It repealed part of the Glass–Steagall Act of 1933, which limited commercial bank securities activities and affiliations between commercial banks and securities firms. With the passage of the Gramm–Leach–Bliley Act, commercial banks, investment banks, securities firms, and insurance companies were allowed to consolidate. The decrease in efficiency started after the GLB was enacted perhaps because the financial institutions were taking on more risky activities and less focused on their traditional roles as financial intermediaries in the 2000s 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. This approach does not rely on any parametric assumptions on the prior distribution of the individual effects and that it utilizes Markov Chain Monte Carlo methods including Gibbs sampling and Metropolis-Hastings algorithm to implement Bayesian inference in a panel data setting. Using the Bayesian approach, it is also very intuitive to consider the restricted region of the slope parameters under a specific functional form. In this paper, monotonicity and curvature properties have been imposed when the translog distance function is specified.

The Monte Carlo Simulation experiments show that the new Bayesian estimator has displayed consistently superior performance under various data generating processes, especially compared to the estimators that rely on parametric functional form assumptions.

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 uniformly dominated by the Bayesian estimator.

The new estimator is 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, though at a slowing down speed. This can be explained by their tendency to taking on more risky activities at the early 2000s.

Appendix A:

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

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

where $\gamma_i = \omega^2V(Y_i - X_i\beta)$ 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:

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

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

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

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

Appendix B:

The Metropolis-Hastings Algorithm

Step 0: Draw a starting point β^0, such that $p(\beta \mid \sigma, \gamma, \omega, Y, X) > 0$, from a starting distribution $p_o(\beta)$. The starting distribution can be based on a crude estimate that will vary from problem to problem but usually involves in discarding part of the information in the target distribution.

Step 1: For $i = 0, .., M$, sample a candidate β^c using the current value of β^i using the jump distribution $q(\beta^i, \beta^c)$.

Step 2: Use the candidate value β^c to evaluate the monotonicity (R_j = 1), quasi-convexity and convexity (R_j = 2). If any violation appears, set $r(\beta^c, \beta^i) = 0$ and jump to Step 4 directly.

Step 3: Calculate the ratio of the kernel density $g(\beta^c)$ of truncated normal distribution $p(\beta \mid \sigma, \gamma, \omega, Y, X)$, and let $\alpha(\beta^i, \beta^c) = \min(\frac{g(\beta^c)}{g(\beta^c)}, 1)$

Step 4: Generate independent uniform random variable U on the unit interval [0,1], and set $\beta^{i+1} = \begin{cases} \beta^c \ \ if \ \ U < \alpha(\beta^c, \beta^i) \\ \beta^i \ \ if \ \ U \geq \alpha(\beta^c, \beta^i) \end{cases}$.

Step 5: Set $i = i + 1$ and go back to Step 1.

Appendix C:

Table 6: The Estimation for the Slope Parameters

Model	BC	CSSW	CSSG	KSS	BE0	BE1		BC	CSSW	CSSG	KSS	BE0	BE1
CIL	0.267394	0.211625	0.205296	0.320024	0.229974	0.211907	PF*CD	-0.028282	-0.023417	-0.023420	-0.024378	-0.011996	0.002723
	(0.015604)	(0.014842)	(0.004009)	(0.016490)	(0.014168)	(0.005165)		(0.018853)	(0.011323)	(0.007121)	(0.009834)	(0.011182)	(0.005174)
CL	0.102395	0.161658	0.169303	0.133170	0.151814	0.165144	PF*DD	-0.114018	-0.017305	-0.024098	-0.004148	-0.015062	-0.006751
	(0.012878)	(0.012244)	(0.003398)	(0.011736)	(0.010868)	(0.006044)		(0.015688)	(0.009595)	(0.006507)	(0.008484)	(0.008648)	(0.004818)
PFA	-0.126714	-0.106713	-0.124307	-0.044849	-0.122111	-0.141086	SA*CD	-0.141683	-0.033535	-0.059756	-0.067716	-0.055219	-0.034027
	(0.031169)	(0.026743)	(0.008180)	(0.023470)	(0.024393)	(0.010516)		(0.031271)	(0.021438)	(0.012105)	(0.019169)	(0.019877)	(0.013525)
NOE	-0.151782	-0.274994	-0.273066	-0.219497	-0.152019	-0.169378	SA*DD	-0.006703	0.053747	0.061960	0.074933	0.036716	0.010049
	(0.035151)	(0.035071)	(0.009826)	(0.030924)	(0.028075)	(0.009404)		(0.030736)	(0.021559)	(0.011642)	(0.019549)	(0.019763)	(0.010749)
PF	-0.108846	-0.057149	-0.062796	-0.067891	-0.057049	-0.068166	CD*DD	-0.097991	-0.105554	-0.098626	-0.057446	-0.092207	-0.006767
	(0.010370)	(0.006407)	(0.003582)	(0.007493)	(0.005713)	(0.004235)		(0.033377)	(0.020702)	(0.013151)	(0.017910)	(0.018797)	(0.010934)
SA	-0.305845	-0.102552	-0.141275	-0.128912	-0.170044	-0.231892	CIL*CIL	0.239373	0.197944	0.207341	0.189705	0.227465	0.110796
	(0.023115)	(0.017762)	(0.005433)	(0.022026)	(0.014980)	(0.010042)		(0.024646)	(0.018416)	(0.006345)	(0.015932)	(0.017940)	(0.009775)
CD	-0.293822	-0.242206	-0.249235	-0.152578	-0.236288	-0.246547	CL*CL	0.113335	0.045183	0.052787	0.016882	0.042151	0.035818
	(0.019988)	(0.013899)	(0.007184)	(0.014208)	(0.013410)	(0.011690)		(0.013263)	(0.010120)	(0.004141)	(0.009309)	(0.008506)	(0.006473)
DD	-0.029454	-0.005520	-0.029726	-0.032132	-0.025869	-0.062852	CIL*CL	-0.065016	-0.045951	-0.048370	-0.040675	-0.032145	-0.008645
	(0.018062)	(0.014840)	(0.005759)	(0.014345)	(0.013910)	(0.004963)		(0.014523)	(0.011902)	(0.004307)	(0.010305)	(0.010754)	(0.005926)
PFA*PFA	-0.058407	-0.076124	-0.064595	-0.027170	0.054452	0.012992	CIL*PFA	-0.030027	-0.040296	-0.030441	-0.048343	-0.000616	-0.036701
	(0.105551)	(0.081836)	(0.035034)	(0.067810)	(0.079399)	(0.029517)		(0.040094)	(0.029056)	(0.012041)	(0.025079)	(0.030007)	(0.009662)
NOE*NOE	-0.350934	-0.263410	-0.254762	-0.194616	-0.317941	-0.117341	CIL*NOE	0.227956	0.032956	0.037051	0.068312	0.008093	0.008623
	(0.175695)	(0.111222)	(0.049618)	(0.096139)	(0.103994)	(0.039303)		(0.043142)	(0.032479)	(0.016995)	(0.028018)	(0.031391)	(0.013939)
PF*PF	-0.030317	-0.017777	-0.021905	-0.019072	-0.028032	-0.016829	CIL*PF	0.036991	0.066231	0.066914	0.042524	0.044617	0.026456
	(0.009275)	(0.005224)	(0.003626)	(0.004633)	(0.004308)	(0.001667)		(0.011928)	(0.007423)	(0.004722)	(0.006691)	(0.007033)	(0.003318)
SA*SA	0.057266	0.111105	0.088612	0.116956	0.037492	0.006608	CIL*SA	-0.197701	-0.045638	-0.058945	-0.056339	-0.043310	-0.028250
	(0.039891)	(0.031775)	(0.015405)	(0.030207)	(0.026234)	(0.019963)		(0.021737)	(0.015992)	(0.007671)	(0.014339)	(0.016593)	(0.014728)
CD*CD	0.018957	0.063958	0.076793	0.104556	0.065066	0.035884	CIL*CD	0.033169	0.040877	0.040902	0.019851	0.036181	0.018882
	(0.054680)	(0.033776)	(0.020720)	(0.028297)	(0.027536)	(0.021044)		(0.022888)	(0.013701)	(0.009409)	(0.011750)	(0.012349)	(0.010192)
DD*DD	0.008094	0.002895	-0.012089	-0.012085	-0.001555	-0.006233	CIL*DD	-0.106948	-0.048120	-0.056765	-0.016793	-0.042321	-0.012520
	(0.039544)	(0.025435)	(0.014893)	(0.021706)	(0.020225)	(0.014501)		(0.022452)	(0.016474)	(0.008142)	(0.014246)	(0.014006)	(0.010247)
PFA*NOE	-0.112425	0.086399	0.043034	0.070533	0.000583	-0.004156	CL*PFA	0.048747	0.037970	0.032106	0.039504	0.007060	-0.008233
	(0.102859)	(0.079549)	(0.036284)	(0.066471)	(0.071651)	(0.030154)		(0.027867)	(0.020247)	(0.010162)	(0.017601)	(0.021441)	(0.008456)
PFA*PF	-0.023871	0.001925	0.005802	0.014807	0.032413	0.012315	CL*NOE	-0.134762	-0.080639	-0.079836	-0.073912	-0.026040	-0.000316
	(0.024511)	(0.014311)	(0.009119)	(0.012065)	(0.012272)	(0.007988)		(0.033544)	(0.023290)	(0.012057)	(0.020149)	(0.026992)	(0.011126)
PFA*SA	0.181100	0.065775	0.079467	0.056101	0.069803	0.049442	CL*PF	0.024490	-0.023625	-0.022260	-0.016442	-0.018719	-0.014127
	(0.043433)	(0.033537)	(0.015360)	(0.029335)	(0.032844)	(0.012386)		(0.009238)	(0.005687)	(0.003329)	(0.004883)	(0.005182)	(0.003565)
PFA*CD	-0.191012	-0.036207	-0.035895	-0.098737	-0.156555	-0.064113	CL*SA	0.052008	0.062364	0.064640	0.063839	0.044966	0.033255
	(0.053517)	(0.035121)	(0.020674)	(0.029706)	(0.032240)	(0.014627)		(0.014866)	(0.011813)	(0.005851)	(0.010214)	(0.010948)	(0.007686)
PFA*DD	0.079834	-0.017070	-0.019489	-0.060602	-0.022135	-0.008999	CL*CD	0.014655	0.001993	0.006427	0.000114	0.006044	0.003804
	(0.048869)	(0.030405)	(0.016956)	(0.026224)	(0.027307)	(0.016413)		(0.017719)	(0.011504)	(0.006786)	(0.009937)	(0.011207)	(0.008676)
NOE*PF	0.137728	0.098342	0.099405	0.059733	0.049369	0.023694	CL*DD	-0.057972	0.025662	0.021266	0.027790	-0.001980	-0.007019
	(0.031395)	(0.019230)	(0.012957)	(0.016994)	(0.016952)	(0.010686)		(0.015838)	(0.011538)	(0.005404)	(0.009853)	(0.008936)	(0.005510)
NOE*SA	-0.121524	-0.118107	-0.112644	-0.115822	-0.068949	-0.024864	CR	0.217115	0.697573	0.622777	0.661193	0.650641	0.700398
	(0.065179)	(0.049241)	(0.024082)	(0.042867)	(0.042545)	(0.020639)		(0.207247)	(0.113233)	(0.089828)	(0.096325)	(0.074863)	(0.087411)
NOE*CD	0.417943	0.145744	0.148929	0.183621	0.245112	0.059685	LR	1.103688	0.272672	0.303568	0.359501	0.601731	0.535736
	(0.083620)	(0.050044)	(0.029691)	(0.042715)	(0.045040)	(0.017632)		(0.174812)	(0.176180)	(0.057171)	(0.158809)	(0.152083)	(0.146781)
NOE*DD	0.179106	0.083529	0.084223	0.057347	0.098175	0.026154	MR	-0.002988	-0.001070	-0.000878	-0.000466	0.000008	0.000340
	(0.064194)	(0.037821)	(0.022992)	(0.032270)	(0.031018)	(0.019058)		(0.002167)	(0.001070)	(0.000974)	(0.000906)	(0.000728)	(0.000805)
PF*SA	0.031111	-0.034051	-0.028221	-0.022330	-0.012928	-0.007935							
	(0.016180)	(0.010140)	(0.006517)	(0.008749)	(0.009524)	(0.006506)							

References:

Battese, G.E., T. Coelli, D. Rao, 1998. An introduction to efficiency and productivity analysis. London: Kluwer Academic Publisher.

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: Journal of the Econometric Society, 1393-1414.

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.

Greene, W.H., 1990, A Gamma-distributed stochastic frontier model. Journal of Econometrics 46, 141-163.

Griffiths, W.E., C.J. O'Donnell, A.T. Cruz, 2000, Imposing regularity conditions on a system of cost and factor share equations. Australian Journal of Agricultural and Resource Economics 44, 107-127.

Inanoglu, H., M. Jacobs, J. Liu, R.C. Sickles, 2012, Analyzing Bank Efficiency: Are "too-big-to-fail" Banks Efficient? Working Paper.

Klein, L.R., 1953, A textbook of econometrics. Row, Peterson & Company, Evanston, IL.

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.

Lau, L.J., 1978, Testing and imposing monoticity, convexity and quasi-convexity constraints. Production economics: A dual approach to theory and applications 1, 409-453.

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.

Park, B.U., R.C. Sickles, L. Simar, 1998, Stochastic panel frontiers: A semiparametric approach. Journal of Econometrics 84, 273-301.

Park, B.U., R.C. Sickles, L. Simar, 2003, Semiparametric-efficient estimation of AR (1) panel data models. Journal of Econometrics 117, 279-309.

Park, B.U., R.C. Sickles, L. Simar, 2007, Semiparametric efficient estimation of

dynamic panel data models. Journal of Econometrics 136, 281-301.
Park, B.U., L. Simar, 1994, Efficient semiparametric estimation in a stochastic frontier model. Journal of the American Statistical Association 89, 929-936.
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," Chapter 17 of Oxford Handbook of Panel Data. Oxford University Press, New York.
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, 669-721.