Estimation of Panel Data Stochastic Frontier Models with Nonparametric Time Varying Inefficiencies

Gholamreza Hajargasht
Shiraz University
Shiraz, Iran

D.S. Prasada Rao
Centre for Efficiency and Productivity Analysis
School of Economics
University of Queensland
Brisbane, Australia

Abstract

An assumption frequently made in the stochastic frontier literature with panel data is that the efficiency effects do not change over time. Such an assumption is strong if the number of time periods is large or when the firms work in a competitive environment. A growing literature has emerged on estimation of stochastic frontier models with time-varying efficiencies. These studies have used parametric specifications for modeling temporal patterns in efficiency changes. In this paper, we show how a variety of panel stochastic frontier models with nonparametric time varying features can be estimated using penalized splines. The feasibility and the performance of the nonparametric approach is assessed using a simulation exercise and an empirical application using real data from the banking industry.

Keywords: Stochastic Frontier, Time Varying Panel Data Model, Nonparametric Estimation, Penalized Splines

JEL Codes: C30, D24

1 - We would like to thank Bill Griffiths, Knox Lovell and Chris O’Donnell for their helpful comments on this work. Any remaining errors are our own responsibility.
1. Introduction

One of the assumptions frequently made in the stochastic frontier literature with panel data is that the firm-level efficiency effects do not change over time. Such an assumption is strong specially if the number of time periods is large or firms work in a competitive environment. In response to this shortcoming, a literature has emerged on estimation of stochastic frontier models with time-varying efficiencies in which the central issue is how to model the temporal change in efficiencies. Different parametric specifications have been proposed by different authors [e.g. Cornwell, Schmidt, and Sickles (1990), Kumbhakar (1990) and Battese and Coelli (1992)]. Since there is usually no prior information on the way efficiency of the firms change over time and the fact that changes can be in any direction or form, appealing to flexible specifications seems sensible.

So far few studies have considered estimation of time varying frontier models without restricted parametric assumptions on the way efficiency changes. The purpose of this paper is to provide a general framework for estimation of stochastic frontier models with nonparametric time varying features. We use a recent nonparametric modeling approach known as penalized splines. We show that this approach can be used to estimate a variety of such models. Specifically, we show how to estimate fixed and random effects models in which efficiency changes in a smooth unknown manner. One advantage of the procedure proposed in this paper is that we can use available mixed model software to easily estimate the resulting complex model. We also describe how the framework can be used to estimate other interesting models such as the one proposed by Kneip et al. (2006).

The structure of the paper is as follows: Section 2 briefly reviews the current literature on stochastic frontiers with time-varying efficiencies. Section 3 is devoted to the estimation of a fixed effects model with nonparametric time-varying inefficiencies while Section 4 is devoted to the estimation of the random effects analogue of the model introduced in Section 3. In section 5 we describe how the estimation of the model proposed by Kneip et al (2006) can be simplified using penalized splines. Section 6 reports results from a simulation experiment and results from an empirical application of the methods proposed in the paper using US Banking data for the period 1984 to 1995 are presented in section 7.
2. SFA Models with Time-Varying Inefficiencies

Consider the following panel stochastic frontier model

\[ y_{it} = x_{it} \beta - u_i + v_{it} \]  \hspace{1cm} (2.1)

where \( y_{it} \) represents the log of output, \( x_{it} \) denotes vector of independent variables (e.g. input quantities), \( x_{it} \beta \) represents a technology linear in parameters, \( v_{it} \) represents random noise and \( u_i \) represents the firm-level effect. To interpret \( u_i \) as inefficiency effect it has to be positive. One way to estimate (2.1) is first to estimate the following model as a standard panel data model without imposing any restriction on \( \alpha_i \)

\[ y_{it} = x_{it} \beta + \alpha_i + v_{it} \]

and estimate \( \hat{u}_i \) in a second stage by \( \{ \max_i \{ \hat{\alpha}_i \} - \hat{\alpha}_i \} \) (see e.g. Schmidt and Sickles 1984).

As it was mentioned before, invariance of efficiency effects when the number of time period is large or firms operate in a competitive environment is an unrealistic assumption. A more realistic model can be specified as

\[ y_{it} = \alpha_i(t) + x_{it} \beta + v_{it} \]  \hspace{1cm} (2.2)

in which \( \alpha_i \) changes over time. The central question here is how the change in efficiencies should be modeled. Several different specifications have been proposed in the literature. A number of these models can be considered as a special case of the following general specification

\[ \alpha_i(t) = \alpha(t)u_i \]  \hspace{1cm} (2.3)
in which firms can have different initial inefficiencies but changes in them follows the same pattern over time. A popular model with such a formulation was proposed by Battese and Coelli (1992). They specify \( \alpha(t) \) as

\[
\alpha(t) = \exp[-\gamma (t - T)]
\]  

(2.4)

Compared to invariant model, this model has only one additional parameter. Battese and Coelli (1992) make a truncated normal assumption for \( u_i \) and estimate the model using the maximum likelihood technique. Another such model proposed by Kumbhakar (1990) has the following specification

\[
\alpha_i(t) = \alpha(t)u_i \quad \text{where} \quad \alpha(t) = \frac{1}{[1 + \exp(\gamma t + \delta t^2)]}
\]  

(2.5)

Kumbhakar assumes a half normal distribution for \( u_i \) and derives the associated likelihood function for this model. Lee and Schmidt (1993) proposed an alternative formulation in which the \( \alpha_i(t) \)s in equation (2.2) are specified as

\[
\alpha_i(t) = \alpha_iu_i
\]  

(2.6)

where the \( \alpha_i \) is specified as a set of time dummy variables. This model is appropriate for short panels, since it requires estimation of \( T-1 \) additional parameters (we have to make the normalizing assumption that \( u_i = 1 \)). Lee and Schmidt estimated both fixed and random-effects versions of the model (2.6). In the fixed effects case both \( \alpha_i \) and \( u_i \) are considered as fixed terms, and in the random effects case \( u_i \) is treated as a random variable. Lee and Schmidt used a least squares estimator, while a generalized method of moments approach to the estimation of the model has been developed by Ahn, Lee, and Schmidt (2001). If the independence assumption between the noise component and the efficiency effects is plausible, and we are willing to make parametric assumptions on the
distribution of inefficiency (e.g., half-normal) a maximum likelihood approach is possible (see Kumbhakar and Lovell 2000). If efficiencies changes smoothly over time then an alternative interesting model can be specified as

$$\alpha_i(t) = \alpha(t) u_i \tag{2.7}$$

where $\alpha(t)$ is an unknown smooth function of time. To our knowledge such a model has not been studied or estimated in the efficiency literature but it seems quite possible to employ penalized splines to estimate such a model.

Another set of specifications can be written in the general form

$$\alpha_i(t) = \sum_{j=1}^{J} \alpha_j(t) u_{ij} \tag{2.8}$$

where J may not be fixed and estimated as a part of the estimation procedure. Cornwell et al. (1990) (hereafter CSS) were one of the first to propose a parametric model of the above form. They specified a quadratic form for the temporal pattern in efficiencies:

$$\alpha_i(t) = u_{i0} + tu_{i1} + t^2u_{i2} \tag{2.9}$$

This quadratic specification allows technical efficiency to vary through time and in a different manner for each firm. If $u_{i1} = u_{i2} = 0$, this model collapses to the time-invariant technical efficiency model. If $u_{i1} = u_1$ and $u_{i2} = u_2$ it collapses to a fixed-effects model with producer-specific intercepts, $u_{i0}$, and a quadratic term in time common to all producers, $u_{1t} + u_{2t^2}$.

Ahn et al. (2006) and Bai (2005) have considered a model with following specification

$$\alpha_i(t) = \sum_{j=1}^{J} \alpha_{ji} u_{ij} \tag{2.10}$$

in which $\alpha_{ji}$ s are specified as a set of appropriate time dummy variables. As it can be seen this model is a generalization of Lee and Schmidt (1993) model. One justification
for such a specification is that the individual effects have multiple components and each component is time varying in a Lee and Schmidt (1993) manner.

Kneip et al. (2006) proposed a similar specification with the difference that inefficiencies change slowly and smoothly over time. Their model can be written as

\[ \alpha(t) = \sum_{j=1}^{J} \alpha_j(t)u_{ij} \]  

(2.11)

where \( \alpha_j(t) \)'s are smooth unknown functions of time. In their model it is assumed that the technology is of a linear form and the model is estimated utilizing a suggestion in Kneip (1994). They use principal components analysis coupled with smoothing spline techniques, and the time-varying effects are represented using a small number of common functions calculated from the data, with coefficients varying across firms. In Table 1, we provide a general taxonomy of the models discussed in this section.

Table 1. Classification of Time-Varying Panel Data Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Author</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_i(t) = \alpha(t)u_i )</td>
<td>Kumbhakar (1990) Battese and Coelli (1992) Lee Schmidt (1993) &amp; Ahn et al. (2001) not yet estimated</td>
<td>( \alpha(t) = [1 + \exp(\gamma t + \delta t^2)]^{-1} ) ( \alpha(t) = \exp[-\gamma(t - T)] ) ( \alpha_i(t) = \alpha_iu_i ) ( \alpha_i(t) = \alpha(t)u_i )</td>
</tr>
<tr>
<td>( \alpha_i(t) = \sum_{j=1}^{J} \alpha_j(t)u_{ij} )</td>
<td>Cornwell et al. (1990) Ahn et al. (2006) Kneip et al. (2006)</td>
<td>( \alpha_i(t) = u_{i0} + iu_{i1} + i^2u_{i2} ) ( \alpha_i(t) = \sum_{j=1}^{J} \alpha_ju_{ij} ) ( \alpha_i(t) = \sum_{j=1}^{J} \alpha_j(t)u_{ij} )</td>
</tr>
</tbody>
</table>
In this paper, we first estimate efficiencies without imposing any restrictions (apart from smoothness) on the temporal pattern of changes. The advantage is the flexibility of the model to capture almost any pattern; the potential drawback is that we might lose accuracy especially because usually the number of time periods is not very large although our simulation experiment shows that with reasonable number of time periods the model is capable of tracking a variety of interesting patterns. In response to this we also show how the flexible model proposed by Kneip et al. (2006) can be estimated using this approach.

3. A Nonparametric Time-varying Fixed Effects Model

The discussion in the literature review section reveals that one of the concerns in studies using time-varying stochastic frontier models has been the appropriate specification for temporal changes in efficiencies. This is especially problematic when the time dimension of the panel data is large – the efficiency of a firm can change in any manner or direction over a long time interval. Thus, it seems sensible to model the efficiency change over time in a nonparametric manner – then no explicit restrictions need to be imposed on the temporal pattern of individual effects.

Here we consider the following model where the technology is linear and time-varying efficiency component is formulated in a nonparametric manner:

\[
y_{it} = \alpha_i(t) + x_i \beta + v_{it}
\]

(3.1)

where \( \alpha_i(t) \) is a term capturing technical efficiency term. Note that sometimes it is better to write the model as

\[
y_{it} = \alpha_i^* (t) + \phi(t) + x_i \beta + v_{it}
\]

with \( \sum_{i=1}^{N} \alpha_i^* (t) = 0 \) and interpret \( \phi(t) \) as a term capturing technical change or common trend in efficiency change (we can’t distinguish between the two).
Estimation of (3.1) in a penalized spline framework proceeds by writing the model in its regression spline form. Model (3.1) looks like a partially linear model however, an important feature of the model is that $\alpha_i(t)$ are different for different $i$, implying the regression spline must be written as

$$y_{it} = \alpha_{i0} + \alpha_i t + \sum_{k=1}^{K} w_{i,k} (t - \tau_k)_+ + x_{it} \beta + \epsilon_{it}$$  \hspace{1cm} (3.2)$$

where $(\tau_1, \ldots, \tau_{K_T})$ are the set of knots associated with $t$. Equation (3.2) can be written in the following equivalent matrix form

$$y = X_0 \beta_0 + Z w + v$$  \hspace{1cm} (3.3)$$

where

$$X_0 = \begin{bmatrix} [i, t] & 0 & \ldots & 0 & x_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \ldots & [i, t] & x_n \end{bmatrix}, \quad \beta_0 = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}, \quad \alpha_i = \begin{bmatrix} \alpha_{i0} \\ \alpha_i \end{bmatrix}$$

$$Z = \begin{bmatrix} (t - \tau)_+ & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & (t - \tau)_+ \end{bmatrix}, \quad w = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$$  \hspace{1cm} (3.4)$$

where $i$ is a $T \times 1$ vector of ones; $t = (1, 2, \ldots, T)'$; and $(t - \tau)_+$ is a $T \times K_T$ matrix containing the truncated polynomial splines respectively. It has been shown that this model can be estimated under a mixed model framework by assuming (see Ruppert et al. 2003 or Hajargasht 2006)

$$w \sim N(0, K) \quad \text{where} \quad K = \sigma^2_i I_{nK}$$

---

2 For an extensive treatment of penalized splines approach to nonparametric estimation see Ruppert et al. 2003.

3 To keep the presentation as simple as possible we use linear splines. Extensions to polynomial splines are trivial.
The advantage of using a mixed model approach is that we can use the standard mixed model methodology and software to obtain estimates of all the parameters, including the $\alpha$’s and smoothing parameters. Once the parameters are estimated, estimates of the inefficiencies can be obtained from

$$\hat{u}_i(t) = \text{Max}_i \{\hat{\alpha}_i(t)\} - \hat{\alpha}_i(t)$$  \hspace{1cm} (3.5)

4. A Nonparametric Time-varying Random Effects Model

Consider again the following nonparametric time varying stochastic frontier model

$$y_{it} = \alpha_i(t) + \phi(t) + x_{it} \beta + v_{it}$$  \hspace{1cm} (4.1)

where $\alpha_i(t)$ is again modeled nonparametrically. The difference between the model considered here with the model in previous section is that here the $\alpha_i(t)$’s are assumed to be random functions with mean zero. In other words, this is the random effects version of the model introduced in the previous section. $\phi(t)$ is the term capturing technical change or common trend in efficiency change. We show in this section that it is possible to model these random functions as regression splines and place them into a mixed model framework.

A strategy for fitting models with a structure comparable to (4.1) has been developed by Durban et al. (2004) in a different context. We can explain the procedure as follows. The non-random part of the model $\phi(t) + x_{it} \beta$ can be modeled in the familiar regression spline form

$$\phi(t) + x_{it} \beta = \alpha_0 + \alpha_1 t + \sum_{k=1}^{K_i} w_{0,k} (t - \tau_k)^+ + x_{it} \beta$$  \hspace{1cm} (4.2)
For the random part $\alpha_i(t)$ we again write the regression spline equivalent

$$\alpha_i(t) = \alpha_{i0} + \alpha_{i1}t + \sum_{k=1}^{K_i} w_{i,k} (t - \tau_k)_+$$

(4.3)

But we must remember that the functions $\alpha_i(t)$ are random functions with mean zero. Being random with mean zero implies that linear part of the regression should also be random with mean zero. This requirement can be met by assuming $(\alpha_{i0}, \alpha_{i1})' = \alpha_i \sim N(0, \Omega)$ where $\Omega$ is a general covariance matrix (see Durban et al. 2004). With the above information we can rewrite (4.3) in the following mixed model form

$$y = X_0 \beta_0 + Z w + v$$

where

$$X_0 = \begin{bmatrix} i & t & x_1 \\ \vdots & \vdots & \vdots \\ i & t & x_n \end{bmatrix} \quad \beta_0 = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \beta \end{pmatrix}$$

(4.4)

$$Z = \begin{bmatrix} [i, t] & 0 & \cdots & 0 & (t-\tau)_+ & 0 & \cdots & 0 & (t-\tau)_+ \\ 0 & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & [i, t] & 0 & \cdots & 0 & (t-\tau)_+ & (t-\tau)_+ \end{bmatrix} \quad w = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \\ w_0 \\ \vdots \\ w_n \end{pmatrix}$$

$$w \sim N(0, K) \quad K = \begin{bmatrix} \text{Diag}(\Omega) & 0 & 0 \\ 0 & \sigma_{\eta}^2 I_{nK} & 0 \\ 0 & 0 & \sigma_\omega^2 I_K \end{bmatrix}$$

and $i$ is a $T \times 1$ vector of ones; $t = (1,2,\ldots,T)'$; $(t-\tau)_+$ is a $T \times K_T$ and $T \times K$ matrix containing the truncated polynomial splines respectively.
The model (4.4) is in the general form of a linear mixed model with a somewhat complex covariance matrix structure. The first block in the matrix $K$ is not a scalar times an identity matrix. However, available mixed model software (e.g. R, SPLUS or SAS) allows estimating a model with such a complex structure without much difficulty, and this is one of the reasons that we have been emphasizing the mixed model approach in this paper. As usual once the model is estimated, the estimator of inefficiency effects can be obtained as $\hat{u}_i(t) = \max \{\hat{a}_i(t) - \hat{a}_j(t)\}$.

5. KSS Model

So far we have considered models where no restrictions (expect for smoothness) on the temporal pattern of changes on estimated efficiencies are imposed. The advantage is the flexibility of the model to capture any pattern; the drawback is that because the number of time periods is not usually very large a nonparametric estimator may not be accurate. Kneip et al. (2006) have considered estimation of a model of the following form

$$y_{it} = \alpha_i(t) + \phi(t) + x_{it} \beta + v_{it}$$  \hspace{1cm} (5.1)

where $\alpha_i(t)$ follows a time-varying fixed effects model and is specified as

$$\alpha_i(t) = \sum_{j=1}^{J} c_{ji} \alpha_j(t)$$  \hspace{1cm} (5.2)

where $c_j$ s are some parameters and $\alpha_j(t)$ s are invariant functions over firms to be estimated. To identify $\phi(t)$ from $\alpha_i(t)$ the following restriction has to be imposed

$$\sum_{i=1}^{N} \alpha_i(t) = 0$$  \hspace{1cm} (5.3)

Their estimation strategy is based on first estimating of $\alpha_i(t)$ s and $\phi(t)$ nonparametrically and then employing the empirical covariance $\Omega_{N,T}$ of $\alpha_i(t)$ s to estimate $\alpha_j(t)$ s based on the ideas from principal component analysis. Their proposed estimation involves the following four steps:
Step 1: Estimate $\beta$ and $\alpha_i(t)$s from following model by modeling $\alpha_i(t)$s using smoothing spline of the form

$$y_{it} - \overline{y}_t = \alpha_i(t) + (x_{it} - x_i)\beta + \tau_{1it}$$  \hfill (5.4)

Step 2: Having $\widehat{\beta}$, estimate $\phi(t)$ in a second stage by using smoothing spline

$$\overline{y}_t - \overline{x}_i\widehat{\beta} = \phi(t) + \tau_{2it}$$  \hfill (5.5)

Step 3: Determine the empirical covariance

$$\Omega_{N,T} = \frac{1}{N} \sum_{i=1}^{N} \alpha_i\alpha_i'$$  \hfill (5.6)

and calculate its eigenvalues $\lambda_1 \geq 0$, $\lambda_2 \geq 0$, $\ldots$, $\lambda_T \geq 0$ and the corresponding eigenvectors $\gamma_1$, $\gamma_2$, $\ldots$, $\gamma_T$ where $\alpha_i = [\alpha_i(1), \ldots, \alpha_i(T)]$

Step 4: Set $\alpha_j(t) = \sqrt{T} \widehat{\alpha}_{jt}$ for $j=1,\ldots,J$ and $t = 1, \ldots, T$ and for all $i = 1, \ldots, N$ and determine $c_1, \ldots, c_J$ by estimating

$$y_{it} - \overline{y}_t - (x_{it} - x_i)\beta = \sum_{j=1}^{J} c_j \widehat{\alpha}_j(t) + \tau_{3it}$$  \hfill (5.7)

Instead of smoothing spline which has been used in KSS study, we can employ mixed model penalized spline to estimate $\alpha_i(t)$ and $\phi(t)$ directly in one step (which is likely to increase efficiency of the estimator). We can use either a fixed or random effects specification. Here we explain how the estimation can be carried out using the fixed effects model. As before we can write the models as

$$y_{it} = \alpha_i(t) + \phi(t) + x_{it}\beta + v_{it}$$

with the restriction that $\sum_{i=1}^{N} \alpha_i(t) = 0$. This equation can written in the following equivalent matrix form

$$y = X\beta_0 + Z\omega + v$$  \hfill (5.8)

where
With this formulation we can estimate $\phi(t)$ and $\alpha_i(t)$ with identification requirement
\[ \sum_{i=1}^{N} \alpha_i(t) = 0 \] in one step. The estimation proceeds by following steps 3 and 4 of the KSS algorithm.

### 6. A Simulation Experiment

In this section, we compare estimators that were proposed for the fixed and random effects model with some of the existing time-varying individual effects estimators, namely those of Cornwell, Schmidt, and Sickles (1990), Schmidt and Sickles (1984) and Kneip et al. (2006). We do this using an experiment similar to the one conducted by Kneip et al. (2006). We consider the following panel data model:

\[ y_{it} = \alpha_{it} + \beta_1 x_{1it} + \beta_2 x_{2it} + v_{it} \] (6.1)

where our assumptions concerning $\alpha_{it}$ will be discussed below.
To simulate data we follow the conventions first used in Park et al. (2003). Specifically, we simulate samples with $N = 30$ and $100$ and $T = 12$ and $30$. The error process $v_{it}$ is drawn randomly from both i.i.d. $N(0,0.5^2)$ and $N(0,1)$: the values of the true $\beta$s are set equal to $(0.5; 0.5)$. In each Monte Carlo sample, the regressors are generated according to a bivariate VAR model as in Park et al. (2003):

$$x_{it} = Rx_{i,t-1} + \eta_{it} \quad \text{where} \quad \eta \sim N(0, I_2)$$

(6.2)

and

$$R = \begin{pmatrix} 0.4 & 0.05 \\
0.05 & 0.4 \end{pmatrix}$$

To initialize the simulation, we chose $x_{i1} = N(0, [I_2 - R^2]^{-1})$ and generated the samples using (6.2) for $t \geq 2$. Then, the obtained values of $x_{it}$ were shifted around three different means to obtain almost 3 balanced groups of firms from small to large. We fixed each group at $\mu_1 = (5; 5); \mu_2 = (7.5; 7.5); \text{and} \mu_3 = (10; 10)$. The idea is to generate a reasonable cloud of points for $x$.

Following Kneip et al. (2006), we generated time-varying individual effects in the following ways:

DGP1: $\alpha_{it} = \theta_0 + \theta_1 \left( \frac{t}{T} \right) + \theta_2 \left( \frac{t}{T} \right)^2$  

DGP2: $\alpha_{it} = \delta_i u_i$  

DGP3: $\alpha_{it} = \gamma_{i1} \text{Sin} (\pi t / 4) + \gamma_{i2} \text{Cos} (\pi t / 4)$  

DGP4: $\alpha_{it} = u_i$

(6.3)

Where $\theta_j (j = 0,1,2) \sim N(0,1), \; u_i \sim N(0,1), \; \delta_i = \delta_{i-1} + \varepsilon_i, \; \varepsilon_i$ and $\gamma_{ij} \sim N(0,1)$. DGP1 is the model considered in Cornwell, Schmidt and Sickles (1990), and DGP2 is a random walk process. DGP3 is considered to model effects with large temporal variations. DGP4 is the usual constant effects model. Thus, we may consider DGP3 and DGP4 as two extreme cases among the possible functional forms of time-varying individual effects. To
measure the performances of the estimators we use the normalized mean squared error (MSE) of the effects defined by

\[
\frac{\sum_i \sum_t (\hat{\alpha}_{it} - \alpha_{it})^2}{\sum_i \sum_t \alpha_{it}^2}
\]  

(6.4)

All the simulation results are based on 1000 replications and the results for the different data generating processes are reported in Tables 2 to 5 in the appendix. Wand (2003) has suggested the following rule for choosing the number and location of the knots on the sample quantile of unique \( s_x, s \).

\[
K = \text{Max} \left\{ 5, \text{Min} \left( \frac{1}{4} \times \text{number of unique } s_x, 35 \right) \right\}
\]

In our simulation, we used a more conservative rule that is number of time periods divided by two (i.e. 6 and 15) on equidistant intervals. All the estimations were performed using mixed model routines in R. For nonparametric models we used slight modification of the codes from Wand (2003) and Durban et al. (2004).

For DGP1, where the true model is of CSS form, CSS random effects and KSS estimators all have good performances. Our so-called fixed effects estimator is somewhat less efficient than these estimators. As expected, the performance of the invariant model is very poor compared to the others. For DGP2, where the temporal pattern of inefficiency effects is in the form of a random walk, time-invariant and CSS estimators all perform poorly, while the nonparametric estimators and KSS perform very well. For DGP3, where there are large changes in the temporal pattern of inefficiency effects, the CSS, time-invariant and fixed effects estimators perform very poorly. The nonparametric random effects estimator’s performance is modest and KSS performs significantly better than other estimators. For the time invariant case (DPG4), all estimators perform very well.
We may summarize the results from the simulation experiments as follows:

1- Our nonparametric random effects estimator performs very well in all cases except where there huge variations in the way efficiency changes over time (DGP3). Even in that case, its performance might be considered as acceptable.

2- For the models considered in this experiment which are all special cases of a KSS model, as expected KSS outperforms the nonparametric estimators although except for the case of DGP3 the difference between KSS and random effects estimator is not significant. Despite the superior performance and flexibility of KSS model we should note that in KSS model, inefficiency is the outcome of several factors initially different across firms but changing in a similar fashion over time. We do not still have enough empirical evidence to know how good of an assumption this is. So in some real data applications, a fully nonparametric specification might be more appropriate.

3- To our surprise, the fixed effects estimator performs poorly in most cases. One possible reason for that could be the large number of dummy variables in the model which may cause the standard mixed model software performs poorly. This needs to be further investigated.

4- Standard estimators such as CSS, invariant model performs poorly when there are significant nonlinearities in the temporal changes in efficiencies.

As we see, in all cases the random effects and KSS estimators produces estimates that are close to the true model. As we do not know the true relationship in practice, we suggest that the proposed estimators be considered as a serious alternative to parametric estimators.

7. An Application to Banking Industry

In this section we apply the proposed time-varying random effects estimator to a real data set. The data has been taken from a series of studies by Sickles and his associates (see Sickles, 2005; and Kneip et al. 2006) on the efficiency of the banking industry. The data consists of a panel from 1984 through 1995 for U.S. commercial banks and it includes
1220 banks over 12 years with the total 14,640 observations\textsuperscript{4}. In this application we have a multiple-output-multiple-input technology and, we model the technology using a distance function. We consider three outputs of banks: real estate loans; commercial and industrial loans and installment loans. On the input side we make use of certificate of deposit; demand deposits; and retail deposits; labor, capital, purchased funds, as the key variables.

We follow Kneip et al. (2006) in using a Cobb-Douglas output distance function. The output distance function \( D(y;x) \leq 1 \) provides a radial measure of technical efficiency by expressing observed outputs as a fraction of the technically feasible maximum outputs (\( y \)) that can be produced by given inputs (\( x \)). The Cobb-Douglas output distance function may be specified as (see e.g. Coelli 1999)

\[
\ln D_{it} = \sum_{j=1}^{m} \gamma_j \ln y_{j, it} - \sum_{k=1}^{K} \beta_k \ln x_{k, it} + v_{it}
\]  

(7.1)

\( D_{it} \) is not directly observable but the homogeneity property of the distance function can be used to write (see e.g. Coelli and Perleman 2000)

\[
\ln y_{m, it} = -\sum_{j} \gamma_j \ln y_{j, it}^* + \sum_{k} \beta_k \ln x_{k, it} - \ln D_{it} + v_{it}
\]  

(7.2)

where \( y_{m} \) is the normalizing output and \( y_{j}^* = y_{j} / y_m \) for \( j = 1, 2, \ldots, m - 1 \). To streamline notation, let \( Y_{j, it}^* = \ln(y_{j, it}^*) \) and \( X_{k, it} = -\ln(x_{k, it}) \). The term \( -\ln D_{it} \) measures inefficiency effects and we model it as \( -\ln D_{it} = \alpha_i(t) \). We also add \( \phi(t) \) to the model to capture technical change or a common trend in efficiency change over time. Thus, our stochastic distance frontier can eventually be written as

\[
Y_{m, it} = \alpha_i(t) + \phi(t) + Y_{it}^* - X_{i}B + v_{it}
\]  

(7.3)

\textsuperscript{4} For more on data see Kneip et al. (2006).
The variables used to estimate the Cobb-Douglas stochastic distance frontier are \( Y_m = \ln \) (real estate loans); \( X = [- \ln \) (certificate of deposit); \( - \ln \) (demand deposit), \( - \ln \) (retail time and savings deposit), -\( \ln \) (labor); \( - \ln \) (capital), and \( - \ln \) (purchased funds)]; \( Y^* = [- \ln \) (commercial and industrial loans/real estate loans), and \( - \ln \) (installment loans/real estate loans)]. For a complete discussion of the approach used here see Adams, Berger, and Sickles (1999).

Table 5 compares the results based on our model [we used 6 knots on equidistant intervals] with the results from other approaches reported from Kneip et al. (2006). The signs and magnitudes of the parameter estimates of the technology appear to be robust to changes in the estimation approach.

Table 5. Estimates of Cobb-Douglas distance function under alternative specification for temporal pattern in efficiencies

<table>
<thead>
<tr>
<th>Variable</th>
<th>Our Model</th>
<th>BC</th>
<th>CSS</th>
<th>KSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>certificate of deposit</td>
<td>-0.0127</td>
<td>-0.0342</td>
<td>-0.023</td>
<td>-0.0025</td>
</tr>
<tr>
<td>demand deposit</td>
<td>-0.1598</td>
<td>-0.0452</td>
<td>-0.1161</td>
<td>-0.0147</td>
</tr>
<tr>
<td>retail deposit</td>
<td>-0.1514</td>
<td>-0.1662</td>
<td>-0.1240</td>
<td>-0.0326</td>
</tr>
<tr>
<td>labor</td>
<td>-0.1998</td>
<td>-0.1490</td>
<td>-0.1547</td>
<td>-0.0799</td>
</tr>
<tr>
<td>capital</td>
<td>-0.0512</td>
<td>-0.0532</td>
<td>-0.0521</td>
<td>-0.0321</td>
</tr>
<tr>
<td>purchased fund</td>
<td>-0.4300</td>
<td>-0.6255</td>
<td>-0.4877</td>
<td>-0.5769</td>
</tr>
<tr>
<td>commercial loan</td>
<td>0.1658</td>
<td>0.1727</td>
<td>0.1566</td>
<td>0.1273</td>
</tr>
<tr>
<td>installment loan</td>
<td>0.3312</td>
<td>0.3142</td>
<td>0.3224</td>
<td>0.3341</td>
</tr>
<tr>
<td>time</td>
<td>-0.0016</td>
<td>0.0043</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The estimated covariance matrix (see equation 6.4) is

\[
K = \begin{bmatrix} \text{Diag}(\Omega) & 0 & 0 \\ 0 & \sigma_i^2 \mathbf{I}_{nK} & 0 \\ 0 & 0 & \sigma_0^2 \mathbf{I}_K \end{bmatrix} = \begin{bmatrix} \text{Diag}.052 & .064 & 0 & 0 \\ .064 & .10 & 2.4 \mathbf{I}_{nK} & 0 \\ 0 & 2.4 \mathbf{I}_{nK} & 0 & .017 \mathbf{I}_K \end{bmatrix} \tag{7.4}
\]
Figures 1-a and 1-b compare the changes in efficiency for the 10 firms out of the first 20 firms over time, estimated from the CSS and semiparametric random effects models respectively. It is apparent that there are significant differences between the estimates, even though the number of time periods in the panel is not very long. We could expect more significant differences in the case of panel data with a longer time dimension. Different conclusion can be inferred from these two sets of estimates. For example from CSS estimates overall difference between the effects between the start of the period and the end of it hasn’t considerably changed while according to nonparametric estimates there has been convergence between the performance of this 10 firms.

Figures 2.a and 2-b compare nonparametric estimates of $\phi(t)$ from model (10.4) with estimates from a CSS model in which $\phi(t)$ has been specified as a quadratic function of time. The economic interpretation of $\phi(t)$ is not very clear – it can be interpreted as technical change or as a common trend in efficiency of all firms. In this example we feel
disinclined to interpret it as technical change because we think it is unlikely that there was technical regress in most years over the period of study. As we see from the figures, there are significant nonlinearities which can not be revealed even by using a quadratic specification.

![Figure 2-a. Graph of $\phi(t)$ over time for our semiparametric model](image1)

![Figure 2-b. Graph of $\phi(t)$ over time for CSS model](image2)

**Conclusion:**

In this paper, time-varying fixed and random effects models were proposed in which the temporal pattern of change in efficiencies was modeled in a nonparametric manner. For estimation purposes, we placed emphasis on using a mixed model penalized spline approach. We also showed how the model proposed by Kneip et al. (2006) can be estimated based on our estimator. There are two advantages for using penalized splines as the nonparametric modeling approach in this context: Firstly, we can use available mixed model routines in software such as R, S-Plus or SAS to easily estimate such complex models. For example we estimated a model with 14460 observations (1220 banks over 12 years) with a few lines of programming in a few minutes. Secondly, it is possible to extend the model to incorporate nonparametrics in the technology part of the model.
To compare the performance of the some of the models proposed here with some of the existing models, a simulation experiment was conducted. The results of the experiment indicate that the proposed random effects estimator and KSS have good performance, no matter what the true temporal pattern is, while the performance of other methods was not robust to changes in the true data generating process.

Finally in Section 7 the random effects model with nonparametric time-varying efficiency effects was used to model a real data set on US banks, and the results were compared with results from a CSS model. The results of the estimation showed that, even in a short panel with only 12 years, a quadratic specification might not be flexible enough to capture changes in individual efficiencies over time. We also found significant nonlinearities in common trends in efficiency or technical change.
References


Table 1- MSE of the Effects for DGP-CSS

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>$\sigma_c$</th>
<th>Invariant</th>
<th>CSS</th>
<th>Nonp-Random</th>
<th>Nonp Fixed</th>
<th>KSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>12</td>
<td>0.5</td>
<td>0.147909</td>
<td>0.041167</td>
<td>0.045594</td>
<td>0.089618</td>
<td>0.045218</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.184046</td>
<td>0.103998</td>
<td>0.106845</td>
<td>0.340252</td>
<td>0.106293</td>
</tr>
<tr>
<td>30</td>
<td>12</td>
<td>0.5</td>
<td>0.139569</td>
<td>0.023499</td>
<td>0.023971</td>
<td>0.037711</td>
<td>0.023979</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.157360</td>
<td>0.058160</td>
<td>0.059123</td>
<td>0.142768</td>
<td>0.059115</td>
</tr>
<tr>
<td>100</td>
<td>12</td>
<td>0.5</td>
<td>0.135024</td>
<td>0.030275</td>
<td>0.030766</td>
<td>0.044714</td>
<td>0.030752</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.170140</td>
<td>0.088102</td>
<td>0.088866</td>
<td>0.187357</td>
<td>0.088826</td>
</tr>
<tr>
<td>30</td>
<td>12</td>
<td>0.5</td>
<td>0.126644</td>
<td>0.014914</td>
<td>0.015151</td>
<td>0.019529</td>
<td>0.015119</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.142368</td>
<td>0.044689</td>
<td>0.045026</td>
<td>0.08257</td>
<td>0.045024</td>
</tr>
</tbody>
</table>
Table 2- MSE of the Effects for DGP-RandomWalk

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>$\sigma_c$</th>
<th>Invariant</th>
<th>CSS</th>
<th>Nonp-Random</th>
<th>Nonp Fixed</th>
<th>KSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>12</td>
<td>0.5</td>
<td>0.392124</td>
<td>0.126130</td>
<td>0.069497</td>
<td>0.072345</td>
<td>0.06432</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.400354</td>
<td>0.142388</td>
<td>0.103863</td>
<td>0.112345</td>
<td>0.10331</td>
</tr>
<tr>
<td>30</td>
<td>12</td>
<td>0.5</td>
<td>0.381781</td>
<td>0.114267</td>
<td>0.036918</td>
<td>0.036543</td>
<td>0.033245</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.383141</td>
<td>0.121186</td>
<td>0.055597</td>
<td>0.050123</td>
<td>0.049882</td>
</tr>
<tr>
<td>100</td>
<td>12</td>
<td>0.5</td>
<td>0.373331</td>
<td>0.09635</td>
<td>0.038296</td>
<td>0.080070</td>
<td>0.070264</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.381688</td>
<td>0.168483</td>
<td>0.111555</td>
<td>0.127510</td>
<td>0.10277</td>
</tr>
<tr>
<td>30</td>
<td>12</td>
<td>0.5</td>
<td>0.359773</td>
<td>0.096647</td>
<td>0.03761</td>
<td>0.03761</td>
<td>0.023412</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.361617</td>
<td>0.102174</td>
<td>0.05622</td>
<td>0.056217</td>
<td>0.037212</td>
</tr>
</tbody>
</table>
Table 4- MSE of the Effects for DGP-3

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>$\sigma_e$</th>
<th>Invariant</th>
<th>CSS</th>
<th>Nonp-Random</th>
<th>Nonp-Fixed</th>
<th>KSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>12</td>
<td>0.5</td>
<td>0.999492</td>
<td>0.95719</td>
<td>0.125188</td>
<td>0.417851</td>
<td>0.0673801</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.997379</td>
<td>1.193452</td>
<td>0.340994</td>
<td>0.723730</td>
<td>0.278642</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>0.5</td>
<td>0.999999</td>
<td>0.99987</td>
<td>0.102276</td>
<td>0.160267</td>
<td>0.0334489</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>1.01645</td>
<td>1.09453</td>
<td>0.337569</td>
<td>0.552136</td>
<td>0.182969</td>
</tr>
<tr>
<td>100</td>
<td>12</td>
<td>0.5</td>
<td>0.999999</td>
<td>0.98367</td>
<td>0.127022</td>
<td>0.289510</td>
<td>0.0607458</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.999880</td>
<td>1.34325</td>
<td>0.338614</td>
<td>0.632701</td>
<td>0.265133</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>0.5</td>
<td>0.992340</td>
<td>0.95432</td>
<td>0.100302</td>
<td>0.186293</td>
<td>0.029200</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>1.49234</td>
<td>1.09265</td>
<td>0.329730</td>
<td>0.554721</td>
<td>0.164204</td>
</tr>
</tbody>
</table>
Table 5- MSE of the Effects for DGP-Inv

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>$\sigma_c$</th>
<th>Invariant</th>
<th>CSS</th>
<th>Nonp-Random</th>
<th>Nonp Fixed</th>
<th>KSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>12</td>
<td>0.5</td>
<td>0.043887</td>
<td>0.044695</td>
<td>0.047318</td>
<td>0.121977</td>
<td>0.046619</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.108310</td>
<td>0.112370</td>
<td>0.117252</td>
<td>0.487913</td>
<td>0.114584</td>
</tr>
<tr>
<td>30</td>
<td>12</td>
<td>0.5</td>
<td>0.022868</td>
<td>0.023076</td>
<td>0.023494</td>
<td>0.042988</td>
<td>0.023214</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.056085</td>
<td>0.057029</td>
<td>0.058294</td>
<td>0.171953</td>
<td>0.057883</td>
</tr>
<tr>
<td>100</td>
<td>12</td>
<td>0.5</td>
<td>0.025190</td>
<td>0.025399</td>
<td>0.030228</td>
<td>0.07264</td>
<td>0.025303</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.082142</td>
<td>0.082818</td>
<td>0.083864</td>
<td>0.29056</td>
<td>0.0829868</td>
</tr>
<tr>
<td>30</td>
<td>12</td>
<td>0.5</td>
<td>0.011971</td>
<td>0.012055</td>
<td>0.012116</td>
<td>0.032642</td>
<td>0.012040</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.039317</td>
<td>0.039651</td>
<td>0.039957</td>
<td>0.146059</td>
<td>0.039647</td>
</tr>
</tbody>
</table>