On Estimating the Mixed Effects Model

Alois Kneip
Department of Economics
University of Mainz

Robin C. Sickles[1]
Department of Economics
Rice University

Wonho Song
Department of Economics
Rice University

Abstract

This paper introduces a new estimation method for time-varying individual effects in a panel data model. An important application is the estimation of time-varying technical inefficiencies of individual firms using the fixed effects model. Most models of the stochastic frontier production function require rather strong assumptions about the distribution of technical inefficiency (e.g., half-normal) and random noise (e.g., normal), and/or impose explicit restrictions on the temporal pattern of technical inefficiency. These assumptions, however, are not easily justifiable, and thus it is not clear how robust one's results are to these assumptions. This paper drops the assumption of a prespecified model of inefficiency, and provides a semiparametric method for estimation of the time-varying effects. The methods proposed in the paper are related to principal component analysis, and estimate the time-varying effects using a small number of common functions calculated from the data. Finite sample performance of the estimators is examined via Monte Carlo simulations. We apply our methods to the analysis of technical efficiency of the U.S. banking industry.

This version: August 2004.
JEL Classification: C13, C14, C23, G21.
Key words and phrases: mixed effects model, stochastic frontier production function model, principal component analysis, and bank efficiency.

[1]Corresponding author: Department of Economics - MS 22, Rice University, 6100 S. Main Street, Houston, TX 77005-1892, Tel: 713-348-3322, Fax: 713-348-5278, E-mail: rsickles@rice.edu.

1. Introduction

Substantial research interest has focused on the topic of assessing a firm's level of technical efficiency using panel data. Two panel data methodologies that have been widely used in the empirical literature are the random effects and the fixed effects stochastic frontier production function model. The random effects (or composed error) stochastic panel frontier production function is characterized by an error term that has two components, a non-negative error term (e.g., half normal) to account for technical inefficiency and a symmetric error term (e.g., normal) to account for other random noise. The parameters of the model are often estimated using the method of maximum likelihood (Aigner, Lovell, and Schmidt, 1977, and Meeusen and van den Broeck, 1977). Schmidt and Sickles (1984) considered feasible generalized least squares (GLS) estimators for the random effects (RE) stochastic panel frontier production function as an alternative to maximum likelihood estimation (MLE) and also introduced the fixed effects (FE) stochastic panel frontier estimator. These RE and FE estimators do not require strong distributional assumptions about technical inefficiency or random noise. Moreover, the FE estimator does not require the assumption of independence between technical inefficiency and the explanatory variables (inputs) for parameter consistency. Although the MLE estimator for the stochastic panel frontier does not in principle require inefficiency and the regressors to be independent, this is the specification that is used in most likelihood based estimators for the stochastic panel frontier model.

All of these models have one thing in common: technical inefficiency is assumed to be time invariant. Although these models contributed to the literature substantially, it is obvious that in many applications the assumption of time-invariant technical inefficiency is too restrictive. For example, during the period of U. S. financial deregulation in the 1980's, financial intermediaries experienced an abrupt and rapid change in business practices and in the competitive environments and were forced to adjust their efficiency levels to a changing best-practice benchmark as the industry was deregulated. To accommodate time-variant inefficiency terms, Battese and Coelli (1992) extended previous MLE-based stochastic frontier production function models by allowing inefficiency terms to be an exponential function of time. However, their approach has limitations in that each firm has the same pattern of variation and thus the rankings of efficiencies across firms do not change over the sample period. On the other hand, Cornwell, Schmidt, and Sickles (1990) extended the traditional panel data model to allow for time-varying efficiency and in their empirical application of efficiency change in the U. S. airline industry after deregulation used a different quadratic function of time for each firm.

In this paper, we further extend the fixed effects model in such a way that we do not impose any explicit restrictions on the temporal pattern of individual effects. Thus, our model is more general than, say fitting quadratic functions of time, and can be used for virtually any pattern of efficiency change. This generality is accomplished by approximating the effect terms nonparametrically utilizing a suggestion in Kneip (1994). The basic idea is related to principal component 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. This approach provides the most general framework available for time-varying effects, other than using time dummy variables to construct the productivity index (Baltagi and Griffin, 1988). Asymptotic distributions of the new estimator are also derived. Simulation experiments indicate that in finite samples our method works much better than other well known time-varying effects estimators. As an illustration, the method is applied to the analysis of technical efficiency in the U.S. banking industry.

The remainder of the paper is organized as follows. Section 2 introduces our new estimator for arbitrary time-varying mixed effects, derives its asymptotic distribution, and provides other analytical results for optimal choice for the number of principal components and smoothing parameters, and for a Durbin-Watson type specification tests. The finite sample performance of our new estimator is evaluated using Monte Carlo simulations in section 3. In section 4 we use the new estimator to analyze the technical efficiency of banks in the U. S. banking system. Concluding remarks follow in section 5. The mathematical proofs are collected in Appendix.

2. Model

We will assume panel data based on a balanced design with T equally spaced repeated measurements per individual. The resulting observations of n individuals can then be represented in the form (Y_{it}, X_{it}), where $t = 1, \ldots T$ and $i = 1, \ldots, n$.

We consider the model

$$Y_{it} = \sum_{j=1}^{p} \beta_j X_{itj} + w(t) + v_i(t) + \epsilon_{it}, \quad i = 1, \ldots, n, t = 1, \ldots, T \tag{1}$$

where the index i denotes individual units (e.g. firms,households, etc.) and the index t denotes time periods. The functions $v_i(t)$ represent individual effects. In the context of stochastic frontier analysis thus quantify efficiencies of individual firms. The function $w(t)$ quantifies a general mean process. Identifiability is ensured by requiring that $\sum_i v_i(t) = 0$. In stochastic frontier analysis, Y and X will usually refer to the logarithms of the original output and input variables. Note that the model does not contain a general constant β_0. Of course, (**??**) could be rewritten in the form

$$Y_{it} = \beta_0 + \sum_{j=1}^{p} \beta_j X_{itj} + w^*(t) + v_i(t) + \epsilon_{it}$$

with $\beta_0 = \frac{1}{T} \sum_t w(t)$ and $w^*(t) = w(t) - \beta_0$ However, the form (**??**) avoids problems of identifiability and is easier to analyze.

We additionally assume that for some fixed $L \in \{0, 1, 2, \dots\}$ the $v_i(t)$ can be decomposed as in

$$v_i(t) = \sum_{r=1}^{L} \theta_{ir} g_r(t). \tag{2}$$

We will additionally assume that

(a) $\sum_i \theta_{i1}^2 \geq \sum_i \theta_{i2}^2 \geq \dots$

(b) $\sum_i \theta_{ir}\theta_{is} = 0$ for $r \neq s$.

(c) $\frac{1}{T}\sum_{t=1}^{T} g_r(t)^2 = 1$ and $\sum_{t=1}^{T} g_r(t)g_s(t) = 0$ for all $r, s \in \{1, \dots, L\}$, $r \neq s$.

Conditions (a) - (c) do not impose any restrictions, and they introduce a suitable normalization which ensures identifiability of the components up to sign changes (instead of θ_{ir}, g_r one may also use $-\theta_{ir}, -g_r$).

Some simple algebra [compare, e.g., with Kneip (1994)] now shows that, if the v_i were known, the components g_r could be determined from the eigenvectors of the empirical covariance matrix Σ_n of $v_1 = (v_1(1), \dots, v_1(T))', \dots, v_n = (v_n(1), \dots, v_n(T))'$:

$$\Sigma_n = \frac{1}{n}\sum_i v_i {v_i}' \tag{3}$$

and use $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_T$ as well as $\gamma_1, \gamma_2, \dots, \gamma_T$ to denote the resulting eigenvalues and orthonormal eigenvectors of Σ_n. Then

$$\lambda_r = \frac{T}{n}\sum_i \theta_{ir}^2 \quad \text{for all} \quad r = 1, 2, \dots, L, \tag{4}$$

$$g_r(t) = \sqrt{T} \cdot \gamma_{rt} \quad \text{for all} \quad r = 1, \dots, L, \ t = 1, \dots, T. \tag{5}$$

Furthermore, it is easily checked that for all $l \leq L$

$$\sum_{i,t}(v_i(t) - \sum_{r=1}^{l} \theta_{ir} g_r(t))^2 = \min_{\tilde{g}_1, \dots, \tilde{g}_l} \sum_i \min_{\vartheta_{i1}, \dots, \vartheta_{il}} (v_i(t) - \sum_{r=1}^{l} \vartheta_{ir} \tilde{g}_r(t))^2. \tag{6}$$

In other words, $v_i(t) \approx \sum_{r=1}^{l} \theta_{ir} g_r(t)$ provides the best possible approximation of the effects v_i in terms of an l-dimensional linear model.

Obviously, Σ_n and, hence, also the components g_r depend on the observed sample and its sample size n. This does not constitute a serious drawback. In fact, in model (**??**) only the L dimensional linear space spanned by $g_1, \dots, g_L$ is identifiable. There are infinitely many possible choices of basis functions, and by using conditions (a) - (c) we select a particularly well-interpretable basis which satisfies (**??**) - (**??**).

We want to emphasize that g_r and θ_{ir} stabilize as n increases. As $n \to \infty$ (with T fixed) the empirical covariance matrix Σ_n converges in probability to the population covariance matrix $\Sigma = \mathbf{E}(\Sigma_n)$. Consequently, $\frac{T}{n}\sum_i \theta_{ir}^2$ and g_r converge in probability to λ_r^∞ and $\sqrt{T}\cdot\gamma_r^\infty$, where $\lambda_1^\infty, \lambda_2^\infty, \ldots$ as well as $\gamma_1^\infty, \gamma_2^\infty, \ldots$ denote eigenvalues and eigenvectors of Σ.

2. Estimation

In practice, $v_1, \ldots, v_n$ are unknown and all components of model (**??**) thus have to be estimated from the data.The idea of our estimation prosedure is easily described: In a first step partial spline methods as introduced by Speckman (1988) are used to determine estimates $\hat{\beta}_j$ and $\hat{v}_i$. The mean function w is estimated nonparametrically, and then estimates $\hat{g}_r$ are determined from the empirical covariance matrix $\hat{\Sigma}_n$ of $\hat{v}_1, \ldots, \hat{v}_n$.

Let us first introduce some additional notations. Let $\bar{Y}_t = \frac{1}{n}\sum_i Y_{it}$, $\bar{Y} = (\bar{Y}_1, \ldots, \bar{Y}_T)'$, $Y_i = (Y_{i1} \ldots, Y_{iT})'$ and $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{iT})$. Furthermore, let $X_{ij} = (X_{i1j}, \ldots, X_{iTj})'$, $\bar{X}_{tj} = \frac{1}{n}\sum_i X_{itj}$, and $\bar{X}_j = (\bar{X}_{1j}, \ldots, \bar{X}_{Tj})'$. We will use X_i and $\bar{X}$ to denote the $T \times p$ matrices with elements X_{itj} and $\bar{X}_{tj}$.

Step 1: Determine estimates $\hat{\beta}_1, \ldots, \hat{\beta}_p$ and $\hat{v}_i(t)$ by minimzing

$$\sum_i \frac{1}{T}\sum_t (y_{it} - \bar{y}_t - \sum_{j=1}^p \beta_j(x_{itj} - \bar{x}_{tj}) - v_i(t))^2 + \sum_i \kappa \frac{1}{T}\int_1^T (v_i^{(m)}(s))^2 ds \quad (7)$$

over all m-times continuously differentiable functions $v_1, \ldots, v_n$ on $[1, T]$. Here, $\kappa > 0$ is a preselected smoothing parameter and $v_i^{(m)}$ denotes the m-th derivative of v_i.

Spline theory implies that any solution $\hat{v}_i$, $i = 1, \ldots, n$ of (**??**) possess an expansion $\hat{v}_i(t) = \sum_j \hat{\zeta}_{ji} z_j(t)$ in terms of a natural spline basis $z_1, \ldots, z_T$ of order $2m$. In practice, one will often choose $m = 2$ which leads to cubic smoothing splines.

If Z and A denote $T \times T$ matrices with elements $z_j(t)$ and $\int_1^T z_j^{(m)}(s) z_j^{(m)}(t)$, the above minimization problem can be reformulated in matrix notation: Determine $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)'$ and $\hat{\zeta}_i = (\hat{\zeta}_{1i}, \ldots, \hat{\zeta}_{Ti})'$ by minimizing

$$\sum_i \left(\|Y_i - \bar{Y} - (X_i - \bar{X})\beta - Z\zeta_i\|_2^2 + \kappa \zeta_i' A \zeta_i \right), \quad (8)$$

where $\|\cdot\|$ denotes the usual Euclidean norm in $\mathbb{R}^T$, $\|a\| = \sqrt{a'a}$.

It is easily seen that with

$$\mathcal{Z}_\kappa = Z(Z'Z + \kappa A)^{-1} Z'$$

the solutions are given by

$$\hat{\beta} = \left(\sum_i (X_i - \bar{X})'(I - \mathcal{Z}_\kappa)(X_i - \bar{X}) \right)^{-1} \sum_i (X_i - \bar{X})'(I - \mathcal{Z}_\kappa)(Y_i - \bar{Y}) \tag{9}$$

as well as as well as

$$\hat{\zeta}_i = (Z'Z + \kappa A)^{-1} Z'(Y_i - \bar{Y} - (X_i - \bar{X})\hat{\beta}).$$

Therefore,

$$\hat{v}_i = Z\hat{\zeta}_i = \mathcal{Z}_\kappa (Y_i - \bar{Y} - (X_i - \bar{X})\hat{\beta}) \tag{10}$$

estimates $v_i = (v_i(1), \ldots, v_i(T))'$.

Remark 1

- An obvious problem is the choice of κ. A straightforward approach then is to use (generalized) cross-validation procedures in order to estimate an optimal smoothing parameter $\hat{\kappa}_{opt}$. Note, however, that the goal is not to obtain optimal estimates of the $v_i(t)$ but to approximate the functions g_r in (**??**). Estimating g in the subsequent steps of the algorithm involves a specific way of averaging over individual data which substantially reduces variability. In order to reduce bias, a small degree of undersmoothing, i.e. choosing $\kappa < \hat{\kappa}_{opt}$, will usually be advantageous.

- Our setup is based on assuming a balanced design. However, in practice one will often have to deal with the situation that there are missing observations for some individuals. In principle, the above estimation procedure can easily be adapted to this case. If for an individual k observations are missing, then only the remaining $T - k$ are used for minimizing (**??**). Estimates of $\hat{v}_i(t)$ at all $t = 1, \ldots, T$ are then obtained by spline interpolation.

- In any case, $\mathcal{Z}_\kappa$ is a positive semi-definite, symmetric matrix. All eigenvalues of $\mathcal{Z}_\kappa$ take values between 0 and 1. Moreover, $tr(\mathcal{Z}_\kappa^2) \leq tr(\mathcal{Z}_\kappa) \leq T$.

Step 2: Estimate $w = (w(1), \ldots, w(T))'$ by by minimizing

$$\frac{1}{T} \sum_t \left(\bar{Y}_t - \sum_{j=1}^{p} \hat{\beta}_j \bar{X}_{tj} - w(t) \right)^2 + \kappa^* \frac{1}{T} \int_1^T (w^{(m)}(s))^2 ds.$$

In principle, a smoothing parameter $\kappa^* \neq \kappa$ may be chosen in this step.

Step 3: Determine the empirical covariance matrix $\hat{\Sigma}_n$ of $\hat{v}_1 = (\hat{v}_1(1), \hat{v}_1(2), \ldots, \hat{v}_1(T))', \ldots, \hat{v}_n = (\hat{v}_n(1), \hat{v}_n(2), \ldots, \hat{v}_n(T))'$ by

$$\hat{\Sigma}_n = \frac{1}{n} \sum_i \hat{v}_i \hat{v}_i'$$

and calculate its eigenvalues $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \hat{\lambda}_T$ and the corresponding eigenvectors $\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_T$.

Step 4: Set $\hat{g}_r(t) = \sqrt{T} \cdot \hat{\gamma}_{rt}$, $r = 1, 2, \ldots, L$, $t = 1, \ldots, T$, and for all $i = 1, \ldots, n$ determine $\hat{\theta}_{1i}, \ldots, \hat{\theta}_{Li}$ by minimizing

$$\sum_t (Y_{it} - \bar{Y}_t - (X_i - \bar{X})\hat{\beta} - \sum_{r=1}^{L} \vartheta_{ri} \hat{g}_r(t))^2 \tag{11}$$

with respect to $\vartheta_{1i}, \ldots, \vartheta_{Li}$.

3. Asymptotic Theory

We now consider properties of our estimators. We assume an i.i.d. sample of individual firms and analyze the asymptotic behavior as $n, T \to \infty$. $h \equiv h(n,T) \leq T$ and $\kappa \equiv \kappa(n,T)$ may either remain fixed or may increase with n. Model (**??**) is assumed to posses a fixed dimension L for all n, T. The following assumption then provides the basis of our analysis. We will write $\lambda_{min}(A)$ and $\lambda_{max}(A)$ to denote the minimal and maximal eigenvalues of a symmetric matrix A.

Assumptions

1) For some fixed $L \in \mathbb{N}$ there exists an L-dimensional subspace $\mathcal{L}_T$ of $\mathbb{R}^T$ such that $v_i \in \mathcal{L}_T$ a.e. for all sufficiently large T. Furthermore, $\mathcal{L}_T$ is independent of X_{it}.

2) There exists a monotonically increasing function $c(T)$ of T such that as $n, T \to \infty$

   – $\mathbf{E}(\frac{1}{T}\sum_{t=1}^{T} w(t)^2) = O(c(T))$, $\quad \mathbf{E}(\frac{1}{T}\sum_{t=1}^{T} \bar{X}_{it,j}^2) = O(c(T))$,
   – $\mathbf{E}(\frac{1}{T}\sum_{t=1}^{T} v_i(t)^2) = O(c(T))$,
   – $\frac{1}{n}\sum_i \theta_{ir}^2 = O_P(c(T))$, $\quad (\frac{1}{n}\sum_i \theta_{ir}^2)^{-1} = O_P(\frac{1}{c(T)})$,
   and $|\frac{1}{n}\sum_i \theta_{ir}^2 - \frac{1}{n}\sum_i \theta_{is}^2|^{-1} = O_P(\frac{1}{c(T)})$

   hold for all $r, s = 1, \ldots, L$, $r \neq s$, $j = 1, \ldots, p$.

3) There exists a monotonically increasing function $d(T)$ of T such that as $n, T \to \infty$ $\mathbf{E}(\frac{1}{T}\sum_{t=1}^{T} w(t)^2) = O(d(T))$, $\quad \frac{1}{\mathbf{E}(\frac{1}{T}\sum_{t=1}^{T} w(t)^2)} = O(1/d(T))$, $\frac{\mathbf{E}(\frac{1}{T}\sum_{t=1}^{T} \bar{X}_{it,j}^2)}{\mathbf{E}(\frac{1}{T}\sum_{t=1}^{T} w(t)^2)} = O(1)$, hold for all $r, s = 1, \ldots, L$, $r \neq s$, $j = 1, \ldots, p$.

4) As $n, T \to \infty$ the smoothing parameters $\kappa \equiv \kappa_{n,T} > 0, \kappa^* \equiv \kappa^*_{n,T} > 0$ are non-decreasing functions of n, T. Smoothness of v_i, w and selection of smoothing parameters $\kappa \equiv \kappa_{n,T}, \kappa^* \equiv \kappa^*_{n,T}$ are such that the smoothing biases

$$b_w(n,T) = \sqrt{T^{-1}\mathbf{E}\|(I - \mathcal{Z}_\kappa)w\|_2^2}, \quad b_v(n,T) = \sqrt{T^{-1}\mathbf{E}(\|(I - \mathcal{Z}_\kappa)v_i\|_2^2)}$$

satisfy

$$b_v(n,T) = O(1), \quad \frac{b_v(n,t)}{c(T)^{1/2}} = o(1), \quad b_w(n,T) = O(1)$$

as $n, T \to \infty$.

5) $\mathbf{E}(\frac{1}{T}\sum_{t=1}^{T} X_{it,j}^2) = O(1)$ holds for all $j = 1, \dots, p$ as $n, T \to \infty$. Furthermore,

$$\lambda_{max}\left([\sum_i (X_i - \bar{X})'(I - \mathcal{Z}_\kappa)(X_i - \bar{X})]^{-1}\right) = O_p(\frac{1}{nT}) \tag{12}$$

and there exists a fixed constant $D < \infty$ such that for all $j = 1, \dots, p$ and all vectors $a \in \mathbb{R}^T$

$$a'(I - \mathcal{Z}_{h,\kappa}) \cdot \mathbf{E}\left((X_{ij} - \bar{X})(X_{ij} - \bar{X})'\right)(I - \mathcal{Z}_\kappa)a \leq D \cdot \|(I - \mathcal{Z}_\kappa)a\|^2. \tag{13}$$

holds for all sufficiently large n, T.

6) The error terms ϵ_{it} are i.i.d. with $\mathbf{E}(\epsilon_{it}) = 0$, $var(\epsilon_{it}) = \sigma^2 > 0$, and $\mathbf{E}(\epsilon_{it}^8) < \infty$. Moreover, ϵ_{it} is independent from $v_i(s)$ and $X_{is,j}$ for all t, s, j.

Subsequent theoretical results rely on asymptotic arguments based on Assumptions 1) -6). It is therefore important to understand these assumptions correctly.

First note that Assumptions 1) and 2) formalize our model introduced in the proceeding sections. However, a crucial point is Assumption 4) which quantifies our requirement of "smooth functions" v_i.

Spline Theory provides a basis to understand the impact of Assumption 4 (see, for example, de Boor 1978, or Eubank 1999). We will concentrate on cubic smoothing splines ($m = 4$). Let $\tilde{v}_i(t)$ denote the corresponding spline interpolant of $v_(1), \dots, v_i(T)$, i.e. $\tilde{v}_i$ is a spline function with $\tilde{v}_i(t) = v_i(t)$ for $t = 1, \dots, T$. By definition, the vector $(I - \mathcal{Z}_\kappa)v_i$ is obtained by evaluating the function v minimizing $\frac{1}{T}\sum_t (v_i(t) - v(t))^2 + \kappa\frac{1}{T}\int_1^T v^{(2)}(t)^2 dt$ at $t = 1, \dots, T$. Consequently, $\frac{1}{T}\|(I - \mathcal{Z}_\kappa)v_i\|^2 \leq \kappa\frac{1}{T}\int_1^T \tilde{v}_i^{(2)}(t)^2 dt$. When analyzing properties of $\mathcal{Z}_\kappa$ it turns out that there exists a constant $0 < q < \infty$ such that $tr(\mathcal{Z}_\kappa^2) \leq q \cdot \frac{T}{\kappa^{1/4}}$. Furthermore, in a simple regression model of the form $y_i = v_i(t) + \epsilon_{it}$ the average variance of the resulting estimator will be of order $\sigma^2 tr(\mathcal{Z}_\kappa^2)/T$. As will be seen in the proof of

Theorem 1 below, this generalizes to the variance of the estimators $\hat{v}_i$ to be obtained in the context of our model. These arguments show that for all n, T

$$\frac{1}{T}|(I-\mathcal{Z}_\kappa)v_i\|^2 \leq \kappa\frac{1}{T}\int_1^T \tilde{v}_i^{(2)}(t)^2 dt, \quad tr(\mathcal{Z}_\kappa^2) =\leq q\cdot\frac{T}{\kappa^{1/4}}, \tag{14}$$

$$\frac{1}{T}\sum_t var_\epsilon(\hat{v}_i(t)) = O_P(\sigma^2 tr(\mathcal{Z}_\kappa^2)/T)$$

where var_ϵ denotes conditional variance given v_i, X_{it}. Similar relations can, of course, be obtained with respect to w.

Note that in our it is only required that the above assumptions hold as "$n, T\rightarrow\infty$". Of course, $n\rightarrow\infty$ will correspond to drawing more and more individuals at random, but different asymptotic setups may be used to describe the situation as "$T\rightarrow\infty$. The point is that any asymptotic theory aims to provide first order approximations of a complex finite sample behavior. In practice, one always has to consider the question which asymptotic setup is best suited to approximate the respective finite sample situation.

Situation 1. in the context of nonparametric regression the usual asymptotic setup consists in assuming that the distance between adjacent observational points tends to zero. In other words, in this setup, instead of adding new equidistant period, the time interval in which observations are taken is held fixed but the distance between observations is reduced. For example, for a fixed number of years, T will increase if instead of yearly data we consider monthly or even daily observations. This will clearly be the natural asymptotic setup in application, where t does not represent chronological time, but, for example, measurements at different ages of individuals.

Formally this asymptotic setup can be described as follows. For each individual there are data from T equidistant observations in a *fixed* time interval $[0,1]$. There exist a twice differentiable functions μ as well as i.i.d. twice differentiable random functions $\nu_1,\ldots,\nu$ on $L^2[0,1]$ such that $\mu_i(\frac{t}{T}) = w(t)$ and $\nu_i(\frac{t}{T}) = v_i(t)$ for $t = 1,\ldots,T$.

In this case we, of course, obtain $\frac{1}{T}\sum_t v_i(t) = O(1)$, $\frac{1}{T}\sum_t w(t) = O(1)$ as $T\rightarrow\infty$ and, hence, Assumptions 2) and 3) refer to a constant functions $c(T) = 1$, $d(t) = 1$. Moreover, In this case $v_i^{(2)}(t) = \frac{1}{T^2}\nu_i^{(2)}(t)$, and $\kappa\frac{1}{T}\int_1^T v_i^{(2)}(t)dt = \kappa\frac{1}{T^4}\int_0^1 \nu_i^{(2)}(t)dt + O(1/T)$.

From (**??**) we can infer that an optimal smoothing parameter then satisfies $\frac{\kappa}{7}T^4 = \kappa_T \sim T^{-4/5}$, which means that the smoothing parameter κ in (**??**) has to increase rapidly as $T\rightarrow\infty$. Similar results are to be obtained with respect to w. Assumption 4) then holds with

$$b_v(n,T)^2 = E(\frac{1}{T}\|(I-\mathcal{Z}_\kappa)v_i\|^2 = O(T^{-4/5}), \quad tr(\mathcal{Z}_\kappa^2)/T = O(T^{-4/5}). \tag{15}$$

Similar rates of convergence then can also be derived for $b_w(n,T)$. Also note that in order to satisfy Assumptions 5) we implicitly assume that X_{itj} are generated by *non-smooth* stochastic processes.

From a practical point of view it is important to well interprete this asymptotic setup. Construction of spline smoothers implies that the value of the integral $\frac{1}{T}\int_1^T \tilde{v}_i^{(2)}(t)^2 dt$ in (**??**) is of the same order of magnitude as the average squared second differences $\frac{1}{T}\sum_t (v_i(t+1) - 2v_i(t) + v_i(t-1))^2$. Therefore, for a given finite sample theoretical results based on the above setup will provide a reasonable first order approximation if it can be assumed that the functions v_i are smooth enough such that $\frac{1}{T}\sum_t (v_i(t+1) - 2v_i(t) + v_i(t-1))^2$ is *much smaller* than the error variance σ^2. In this case a fairly large smoothing parameter κ will still result in a small bias while at the same time the average variance of the estimator will be much smaller than σ^2 (due to $tr(\mathcal{Z}_\kappa^2) \ll T$).

Situation 2. Smoothness can also be formalized in a setup which corresponds to the usual time series asymptotics. Indeed, $w(t), v_i(t)$ may be generated by $I(1)$ or $I(2)$ processes. In this case the asymptotic setup of Situation 1 may not be appropriate since $\frac{1}{T}\sum_t (v_i(t+1) - 2v_i(t) + v_i(t-1))^2$ may be of the same order of magnitude as σ^2. However, reasonable convergence results can still be established due to the fact that $\frac{1}{T}\sum_t (v_i(t+1) - 2v_i(t) + v_i(t-1))^2$ is of a smaller stochastic order of magnitude as $\frac{1}{T}\sum_t v_i(t)^2$.

Let us consider the example of a random walk. Assume that

$$w(t) + v_i(t) = \vartheta_i r_t, \quad \text{with } r_{t+1} = r_t + \delta_t,$$

where $\delta_1, \delta_2, \ldots$ are i.i.d with $E(\delta_t) = 0$, $var(\delta_(t)) = \sigma_\delta^2$, and δ_t is independent of ϑ_i.

Our model then holds with $L = 1$, $w(t) = \bar{\vartheta}_i r_t$, $g_r(t) = \frac{r_t}{\sqrt{T}}$ and $\theta_{1i} = \sqrt{T}(\vartheta_i - \bar{\vartheta}_i)$. Since, $\frac{1}{T}E(\vartheta_i)^2)E(r_t^2) = O(T)$, Assumptions 2) and 3) are then satisfied with $c(T) = d(T) = T$.

On the other hand, averages of squared first or second differences $(r_{t+1} - r_t)^2$ or $(r_{t+2} - 2r_t + r_{t-1})^2$ are bounded in probability which implies that for a cubic spline interpolant $r(t)$ of $r - t$ we obtain $E(\frac{1}{T}\int_1^T r^{(2)}(t)dt) = O(1)$ as $T \to \infty$. It is then easy to show that an optimal smoothing parameter may be chosen as a constant (independent of n and T) such that

$$b_v(n, T) = E(\frac{1}{T}\|(I - \mathcal{Z}_\kappa)v_i\|) = O(1), \quad tr(\mathcal{Z}_\kappa^2)/T = O(1). \tag{16}$$

This, of course implies that there is convergence when considering the difference $v_i - \mathcal{Z}_\kappa v_i$ relative to the size of v_i:

$$E(\frac{1}{\|v_i\|^2}\|(I - \mathcal{Z}_\kappa)v_i\|^2) = O(1/T)$$

Assumption 5) contains regularity conditions which imposes a restriction on the design matrix. It essentially requires that the time paths $\{X_{itj} - \bar{X}_{ij}\}_t$ are "less smooth" than those of $\{v_i(t)\}_t$. In particular, stationary processes generate non-smooth time parts.

When considering the simplest case $p = 1$, Assumption 5) is, for example, fulfilled if the individual processes $\{X_{it}\}_t$ are independent realizations of some $ARMA(q_1, q_2)$ process. Then $\mathbf{E}((X_i - \bar{X})(X_i - \bar{X})')$ corresponds to the autocovariance matrix of this ARMA process, and if $tr(\mathcal{Z}_\kappa^2) = o(T)$ (**??**) as well as (**??**) follow from the well known structure of such autocovariance matrices.

Assumption 5) also holds if $\{X_{it}\}_t$ are generated by $ARMA(q_1, q_2)$ with individually different parameters. For example assume that $X_{it} = \tilde{X}_{it} + \delta_i$, where $\{\tilde{X}_{it}\}_t$ are independent realizations of an $MA(q)$ process and δ_i are independent, zero mean random variables with variance Δ^2. Then

$$\mathbf{E}\left((X_{ij} - \bar{X})(X_{ij} - \bar{X})'\right) = \Gamma + \Delta^2 \cdot \mathbf{1}\mathbf{1}',$$

where Γ is the autocovariance matrix of the underlying $MA(q)$ process. Since by assumption $\mathcal{Z}_\kappa \mathbf{1} = \mathbf{1}$ for $\mathbf{1} = (1, 1, \ldots, 1)'$ we arrive at

$$(I - \mathcal{Z}_\kappa)\mathbf{E}\left((X_{ij} - \bar{X})(X_{ij} - \bar{X})'\right)(I - \mathcal{Z}_\kappa) = (I - \mathcal{Z}_\kappa)\Gamma(I - \mathcal{Z}_\kappa).$$

Since by Assumption 4) there exists a $q < 1$ such that $tr(\mathcal{Z}_\kappa^2) \leq q \cdot T$ relations (**??**) as well as (**??**) are an immediate consequence.

We are now ready to state our main theorem. We will use the notation "$\mathbf{E}_\epsilon$" to denote conditional expectation given v_i and X_i, $i = 1, \ldots n$. Moreover, $\tilde{X}_i = X_i - \bar{X}$, and we will say that v_i and X_i are uncorrelated, if $E(v_i|X_i) = 0$ as well as $E(v_i(s)v_i(t)|X_i) = E(v_i(s)v_i(t))$ for all, s, t.

Theorem 1 Under Assumption A we obtain as $n, T \to \infty$

(a) $\|\beta - \mathbf{E}_\epsilon(\hat{\beta})\| = O_P(b_\beta(n, T))$, where

$$b_\beta(n,T) := \begin{cases} \frac{b_v(n,T))}{\sqrt{Tn}} & \text{if } X_i \text{ and } v_i \text{ are uncorrelated,} \\ \frac{b_v(n,T))}{\sqrt{T}} & \text{else,} \end{cases}$$

and $V_{n,T}^{-1/2}(\hat{\beta} - \mathbf{E}_\epsilon(\hat{\beta})) \sim \mathbf{N}(0, I)$, where

$$\begin{aligned} V_{n,T} &= \sigma^2 \left(\sum_i \tilde{X}_i'(I - \mathcal{Z}_\kappa)\tilde{X}_i\right)^{-1} \left(\sum_i \tilde{X}_i'(I - \mathcal{Z}_\kappa)(I - \frac{1}{n}\mathbf{1}\mathbf{1}')(I - \mathcal{Z}_\kappa)\tilde{X}_i\right) \left(\sum_i \tilde{X}_i'(I - \mathcal{Z}_\kappa)\tilde{X}_i\right)^{-1} \\ &= O_P\left(\frac{1}{nT}\right). \end{aligned}$$

(b) $\frac{\|w - \hat{w}\|}{\|w\|} = O_P\left(\frac{b_w(n,T)}{d(T)^{1/2}} + b_\beta(n,t)) + \sqrt{\frac{tr(\mathcal{Z}_\kappa^2)}{nTd(T)}}\right)$.

(c) For all $r = 1, \ldots, L$

$$T^{-1/2}\|g_r - \hat{g}_r\| = O_P\left(\frac{b_v(n,T)}{c(T)^{1/2}} + \frac{1}{T^2 c(T)^2} + \sqrt{\frac{tr(\mathcal{Z}_\kappa^2)}{nTc(T)}}\right).$$

(d) For all $r = 1, \dots, L$

$$\hat{\theta}_{ri} - \theta_{ri} = O_P\left(\frac{b_v(n,T)^2}{c(T)} + \frac{tr(\mathcal{Z}_\kappa^2)}{nT} + \frac{1}{\sqrt{T}}\right).$$

Furthermore, if $\frac{b_v(n,T)^2}{c(T)} + b_\beta(n,t) + \frac{tr(\mathcal{Z}_\kappa^2)}{nT} = o(T^{-1/2})$, then

$$\sqrt{T}(\hat{\theta}_{1i} - \theta_{1i}, \dots, \hat{\theta}_{Li} - \theta_{Li})' \to_d \mathbf{N}(0, \sigma^2 I).$$

(e) If additionally $tr(\mathcal{Z}_\kappa^2)/n \to 0$ as well as $T \cdot b_\beta(n,T)^2 + \frac{1}{Tc(T)} = o\left(\sqrt{tr(\mathcal{Z}_\kappa^4)/n}\right)$, then

$$\frac{n\sum_{r=L+1}^T \hat{\lambda}_r \; - (n-1)\sigma^2 \cdot tr(\mathcal{Z}_\kappa \hat{\mathcal{P}}_L \mathcal{Z}_\kappa)}{\sqrt{2n\sigma^4 \cdot tr((\mathcal{Z}_\kappa \hat{\mathcal{P}}_L \mathcal{Z}_\kappa)^2)}} \to_d \mathbf{N}(0,1),$$

where $\hat{\mathcal{P}}_L = I - \sum_{r=1}^L \hat{\gamma}_r \hat{\gamma}_r'$.

3.2 Choice of Dimension

Theorem 1(e) may be used to estimate the dimension L. A prerequisite is of course the availability of a reasonable estimator of σ^2. We propose to use

$$\hat{\sigma}^2 := \frac{1}{(n-1) \cdot tr(I - \mathcal{Z}_\kappa)^2} \sum_i \|(I - \mathcal{Z}_\kappa)(Y_i - \bar{Y} - (X_i - \bar{X})\hat{\beta})\|^2. \tag{17}$$

We then use the following procedure to determine an estimate $\hat{L}$ of L:

First select an $\alpha > 0$ (e.g., $\alpha = 1\%$). For $l = 1, 2, \dots$ determine

$$C(l) := \frac{n\sum_{r=l+1}^T \hat{\lambda}_r \; - (n-1)\hat{\sigma}^2 \cdot tr(\mathcal{Z}_\kappa \hat{\mathcal{P}}_l \mathcal{Z}_\kappa)}{\sqrt{2n\hat{\sigma}^4 \cdot tr((\mathcal{Z}_\kappa \hat{\mathcal{P}}_l \mathcal{Z}_\kappa)^2)}}. \tag{18}$$

Choose $\hat{L}$ as the smallest $l = 1, 2, \dots$ such that

$$C(l) \leq z_{1-\alpha},$$

where $z_{1-\alpha}$ is the $1-\alpha$ quantile of a standard normal distribution.

The following theorem provides a theoretical justification for this procedure.

Theorem 2 In addition to the assumptions of Theorem 1, assume that $tr(I - Z_\kappa) \geq D_1 \cdot T$ for some constant $D_1 > 0$ as well as $tr(\mathcal{Z}_\kappa^2)/n \to 0$ and $T \cdot B(n,T)^2 + \frac{1}{T} = o\left(\sqrt{tr(\mathcal{Z}_\kappa^4)/n}\right)$ as $n, T \to \infty$. Then,

$$\liminf_{n,T\to\infty} \mathbf{P}(\hat{L} = L) \geq 1 - \alpha.$$

3.3 Durbin-Watson Type Test

An interesting question is whether in a particular application it is necessary to use our complicated procedure to estimate time varying individual effects, or whether it is simply possible to assume constant effects $\theta_i = v_i(1) = v_i(2) = \cdots = v_i(T)$. This question can be resolved by a Durbin-Watson type test.

The procedure aims to test the null hypothesis $H_0 : \theta_i = v_i(1) = v_i(2) = \cdots = v_i(T)$ against the general alternative H_1: v_i is time varying. Note that under H_0 our model takes the form

$$Y_{it} = \beta_0 + \sum_{j=1}^{p} \beta_j X_{itj} + \theta_i + \epsilon_{it}, \quad i = 1, \ldots, n, t = 1, \ldots, T \tag{19}$$

with $\frac{1}{n}\theta_i = 0$.

Let $\theta_i^* = \beta_0 + \theta_i$. Our test of H_0 then is constructed in the following way.

- Fit the null model to the data by determining parameter estimates $\hat{\beta}_1, \ldots, \hat{\beta}_p$ and $\hat{\theta}_i^*$ via minimizing
$$\sum_i \sum_t (Y_{it} - \sum_{j=1}^{p} \beta_j X_{itj} - \theta_i^*)^2.$$
The solutions are given by
$$\hat{\beta} = \left(\sum_i X_i'(I - \frac{1}{T}\mathbf{1}\mathbf{1}')X_i \right)^{-1} \sum_i X_i'(I - \frac{1}{T}\mathbf{1}\mathbf{1}')Y_i \tag{20}$$
as well as
$$\hat{\theta}_i^* = \frac{1}{T}\mathbf{1}'(Y_i - X_i\hat{\beta}).$$

- Calculate the residuals
$$\hat{\epsilon}_{it} = Y_{it} - \sum_{j=1}^{p} \hat{\beta}_j X_{itj} - \hat{\theta}_i^*.$$

- Determine the test statistic
$$D = \frac{\sum_{i=1}^{n} \sum_{t=2}^{T} (\hat{\epsilon}_{it} - \hat{\epsilon}_{i,t-1})^2}{\sum_{i=1}^{n} \sum_{t=1}^{T} \hat{\epsilon}_{it}^2}.$$

- For a selected significance level $\alpha > 0$, reject H_0 if
$$D \notin \left[2 - z_{1-\alpha/2} \frac{2}{\sqrt{nT}}, 2 + z_{1-\alpha/2} \frac{2}{\sqrt{nT}} \right].$$

The above critical values for the test statistic are justified by the following Theorem 3, which provides an asymptotic approximation of the null distribution of D. Asymptotics is based on $n \to \infty$, while $T \equiv T(n)$ may either be constant or increase with n.

Theorem 3 Suppose that $H_0 : \theta_i = v_i(1) = v_i(2) = \cdots = v_i(T)$ is true and that Assumptions A(1) as well as A(3) are fulfilled for $Z_\kappa := \frac{1}{T}\mathbf{1}\mathbf{1}'$. Then, as $n \to \infty$

$$\frac{\sqrt{nT}}{2}(D-2) \to_d \mathbf{N}(0,1).$$

4. Simulations

In this section, we investigate the finite sample performances of the new estimator described in Section 2 (hereafter we will call it KSS estimator) through Monte Carlo experiment. Two of the existing time-varying individual effects estimators [Cornwell, Schmidt, and Sickles (1990) and Battese and Coelli (1992)] as well as the fixed and the random effects estimators are also considered to compare with our estimator.

We consider the following panel data model:

$$Y_{it} = \sum_{j=1}^{p} \beta_j X_{itj} + v_{it} + \epsilon_{it}$$

where v_{it} will be discussed later. We simulate samples of size $n = 30, 100, 300$ with $T = 12, 30$ in a model with $p = 2$ regressors. The error process ϵ_{it} is drawn randomly from i.i.d. $\mathbf{N}(0,1)$. The values of true β 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, Sickles, and Simar (2002):

$$X_{it} = RX_{i,t-1} + \eta_{it}, \text{ where } \eta_{it} \sim \mathbf{N}(0, I_2), \tag{21}$$

and

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

To initialize the simulation, we choose $X_{i1} \sim \mathbf{N}(0, (I_2 - R^2)^{-1})$ and generate the samples using (**??**) for $t \geq 2$. Then, the obtained values of X_{it} are shifted around three different means to obtain almost 3 balanced groups of firms from small to large. We fix each group at $\mu_1 = (5,5)'$, $\mu_2 = (7.5, 7.5)'$, and $\mu_3 = (10,10)'$. The idea is to generate a reasonable cloud of points for X.

We generate time-varying individual effects in the following ways:

$$\begin{aligned}
\text{DGP1} &: v_{it} = \theta_{i0} + \theta_{i1}t + \theta_{i2}t^2 \\
\text{DGP2} &: v_{it} = -\exp(-\eta(t-T))u_i \\
\text{DGP3} &: v_{it} = v_{i1}g_{1t} + v_{i2}g_{2t} \\
\text{DGP4} &: v_{it} = -u_i
\end{aligned}$$

where θ_{ij} $(j = 0, 1, 2) \sim \mathbf{N}(0, 1)/10^2$, $\eta = 0.15$, $u_i \sim$ i.i.d. $|\mathbf{N}(0, 1)|$, v_{ij} $(j = 1, 2) \sim \mathbf{N}(0, 1)$, $g_{1t} = \sin(\pi t/4)$ and $g_{2t} = \cos(\pi t/4)$. DGP1 is the model considered in Cornwell, Schmidt and Sickles (1990), and DGP2 in Battese and Coelli (1992). DGP3 is considered here to model effects with large temporal variations. DGP4 is the usual constant effects model. Thus, we may consider DGP3 and 4 as two extreme cases among the possible functional forms of time-varying individual effects.

For the KSS estimator, cubic smoothing splines were used to approximate v_{it} in Step 1, and the smoothing parameter κ was selected by using generalized cross-validation.[2] Most simulation experiments were repeated 1,000 times except the cases for $n = 300$ for which 500 times repetitions were carried out. To measure the performances of the effect and efficiency estimators, we used normalized mean squared error (MSE):

$$R(\widehat{v}, v) = \frac{\sum_{i,t} (\widehat{v}_{it} - v_{it})^2}{\sum_{i,t} v_{it}^2}.$$

For the estimates of technical efficiency, we also considered the Spearman rank order correlation.

Before we present the simulation results, we briefly introduce the other estimators. For Within and GLS estimators, once individual effects v_i are estimated, technical efficiency is calculated as $TE = \exp\{v_i - \max(v_i)\}$ following Schmidt and Sickles (1984). Battese and Coelli (1992) (hereafter BC) employ the maximum likelihood estimation method to estimate the following equation

$$Y_{it} = \beta_0 + \sum_{j=1}^{p} \beta_j X_{itj} + \epsilon_{it} - u_{it}$$

where the time-varying effects terms are defined as $u_{it} = \eta_{it} u_i = \{\exp[-\eta(t - T)]\} u_i$ for $i = 1, \ldots, n$. Technical efficiency is then calculated as $TE_{BC} = \exp(-u_{it})$. Cornwell, Schmidt, and Sickles (1990) (hereafter CSS) approximate time-varying effects by a quadratic function of time. Thus, the CSS estimator is

$$\beta_{CSS} = (X' M_Q X)^{-1} X' M_Q y$$

where $M_Q = I - Q(Q'Q)^{-1}Q'$, $Q = \text{diag}(W_i)$, $i = 1, \ldots, n$, and $W_{it} = [1, t, t^2]$. Technical efficiency is defined as $TE_{CSS} = \exp\{v_{it} - \max(v_{it})\}$. For the KSS estimator, technical efficiency is calculated similarly as for the CSS estimator.

Now we present the simulation results. Tables 1-4 present mean squared errors (MSE) of coefficients, effects, and efficiencies, and the Spearman rank order correlation coefficient of efficiencies for each DGP. Also, average optimal dimensions, L, chosen by $C(l)$ criterion

[2] We let $\kappa = (1 - p)/p$ and chose p among a selected grid of 9 equally spaced values between 0.1 and 0.9 so that generalized cross-validation rule is minimized.

are reported in the last column of second panel in each Table. Note first that optimal dimension, L, is correctly chosen for the KSS estimator in all DGPs.[3] Thus, we can verify the validity of the dimension test $C(l)$ discussed in Section 2.

For DGP1, the performances of the KSS estimator are better than the other estimators by any standards. This is true even when the data is as small as $n = 30$ and $T = 12$. In particular, the KSS estimator outperforms the other estimators in terms of MSE of efficiency. Since the data are generated by DGP1, we may expect that CSS estimator performs well. This is true for $T = 30$. However, if T is small ($T = 12$), the CSS estimator is no better than the other estimators. The performances of Within, GLS, and BC estimators generally get worse as T increases.

For DGP2, note that the data is generated using the model specification of the BC estimator. Even in this situation, overall the performances of the KSS estimator are comparable to or sometimes better than those of the BC estimator. The BC estimator seems to work fine for the estimation of effects and efficiencies. In terms of MSE of coefficients, however, it appears that the BC estimator is not reliable when T is large ($T = 30$). The Within and GLS estimators also suffer from tremendous distortions in their coefficients estimates when T is large.

DGP3 generates effects with large temporal variations. Hence, simple functions of time such as used in the CSS or BC estimators are not sufficient for this type of DGP. However, the KSS estimator does not impose any specific forms on the temporal pattern of effects, and thus it can approximate any shape of time varying effects. We may then expect good performances of the KSS estimator even in this situation, and results in Table 3 confirm such belief. On the other hand, the other estimators suffer from severe distortions in the estimates of effects and efficiencies, although coefficient estimates look reasonably good. In particular, rank correlations of efficiencies are almost zero when T is large.

DGP4 represents the reverse situation so that there is no temporal variation in the effects. Hence, the Within and GLS estimators work very well. Now, our primary question is what are the performances of KSS estimator in this situation. As seen in Table 4, its performances are fairly well and comparable to those of the Within and GLS estimators. Therefore, the KSS estimator may be safely used even when temporal variation is not noticeable.

In summary, simulation experiments show that either if constant effects are assumed when the effects are actually time-variant, or if the temporal patterns of effects are misspecified, parameters as well as effect and efficiency estimates become severely biased. In these cases, large T increases the bias, and large n does not help solve the problem. On the other hand, our estimator performs very well regardless of the assumption on the temporal pattern of effects, and therefore, our estimator is preferred to other existing estimators.

[3] Although DGP1 consists of three different functions, $[1, t, t^2]$, t^2 term is dominating as T gets large. Thus one dimensional model is sufficient to approximate the effects generated by DGP1.

5. Efficiency Analysis of Banking Industry

5.1 Empirical Model

We model the multiple output/multiple input technology using the output distance function. The output distance function, $D(Y,X) \leq 1$, provides a radial measure of technical efficiency by specifying the fraction of aggregated outputs (Y) produced by given inputs (X). An m-output, n-input deterministic distance function can be approximated by

$$\frac{\prod_j^m Y_j^{\gamma_j}}{\prod_k^n X_k^{\beta_k}} \leq 1,$$

where the $\gamma_j' s$ and the $\beta_k' s$ are weights describing the technology of a firm. If it is not possible to increase the index of total output without either decreasing an output or increasing an input, the firm is producing efficiently or the value of the distance function equals 1.

The Cobb-Douglas stochastic distance frontier that we utilize below in our empirical illustration is derived by simply multiplying through by the denominator, approximating the terms using natural logarithms of outputs and inputs, and adding a disturbance term ϵ_{it} to account for statistical noise. We also specify a nonnegative stochastic term u_{it} for the firm specific level of radial technical inefficiency, with variations in time allowed. The Cobb-Douglas stochastic distance frontier is thus

$$0 = \sum_j \gamma_j \ln y_{j,it} - \sum \beta_k \ln x_{k,it} + u_{it} + \epsilon_{it}.$$

Then, we normalize the outputs with respect to the first output and rearrange to get

$$\ln y_J = \sum_j \gamma_j(-\ln \widehat{y}_{j,it}) - \sum \beta_k(-\ln x_{k,it}) - u_{it} + \epsilon_{it},$$

where y_J is the normalizing output and $\widehat{y}_j = y_j/y_J$, $j = 1,\ldots,m$, $j \neq J$. To streamline notations, let $Y_{it} = \ln y_J$, $Y_{it}^* = -\ln \widehat{y}_{j,it}$, $X_{it} = -\ln x_{k,it}$, and $v_{it} = -u_{it}$, in which case we can write the stochastic distance frontier as

$$Y_{it} = Y_{it}^{*\prime}\gamma - X_{it}'\beta + v_{it} + \epsilon_{it}. \tag{22}$$

This model can be viewed as a generic panel data model in which the effects are interpreted as time-varying firm efficiencies, and fits into the class of frontier models developed and extended by Aigner, Lovell, and Schmidt (1977), Meeusen and van den Broeck (1977), Schmidt and Sickles (1984), and Cornwell, Schmidt, and Sickles (1990).

5.2 Data

We use panel data from 1984 through 1995 for U.S. commercial banks in limited branching regulatory environment. The data are taken from the Report of Condition and Income

(Call Report) and the FDIC Summary of Deposits.[4] The data set include 1220 banks or 14,640 total observations. Table 5 provides variables description and gives the means of the samples.

The variables used to estimate the Cobb-Douglas stochastic distance frontier are $Y =$ 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 in this paper, see Adams, Berger, and Sickles (1999).

5.3 Empirical Results

The Hausman-Wu test, which tests the correlation assumptions for regressors and individual effects, was performed. The null hypothesis of no correlation was rejected at the 1% significance level. Thus there is strong evidence against the exogeneity assumption underlying the GLS estimator. Consequently, in the following analysis we do not report the results from the GLS estimator. The assumption is also fatal to the consistency of the BC estimator. However, we will provide estimation results for the BC estimator to compare them with those from the other estimators (Within, CSS, and KSS) which are robust to the existence of correlation between regressors and effects. The Durbin-Watson type test was also performed to test the null hypothesis of constant individual effects. Test statistic is $D = -90.876$, which leads to strong rejection of the null. This implies that individual effects are better approximated by time-varying effects estimators rather than time-invariant effects estimators.

Table 6 displays the results for parameter estimates from the Within, BC, CSS, and KSS estimators. The dimension L is chosen according to the rule described in Section 2 with the maximum dimension set to 8. In calculating efficiency scores from the effects estimators, the effects estimates are truncated at the top and bottom 5% level (see Berger 1993).[5] To calculate technological changes, time trend is included in the estimation of the Within and BC estimators. For KSS, $w(t)$ estimates the time-specific effects. Resulting average technological changes are 1.40%/year from Within, 0.43%/year from BC, and 3.53%/year from the KSS estimator.

Results for the respective estimators do not indicate any significant scale economies. Estimated technical efficiencies range from 50.26% to 65.7% (ignoring the result from the BC estimator). Figure 1 displays the temporal pattern of efficiency changes for time-variant

[4] For a more detailed discussion of data, see the Appendix in Jayasiriya (2000).

[5] This does not apply to the BC estimator because it directly calculates efficiencies. For the time-varying effects estimators, the firms which enter the top and bottom 5% range of effects in any time periods were excluded in calculating average efficiencies. Therefore, in this sense, it is not fair to directly compare the efficiencies from the Within or BC estimators from those from the CSS and KSS estimators.

efficiency estimators. Technical efficiencies calculated from the BC estimator increase during the sample period, while those from the CSS and KSS estimators slightly decrease. The pattern is more clear with the KSS estimator: technical efficiency decreases from 51% in 1984 to 47% in 1995. Therefore, we may conclude that overall technical efficiencies decreased during the sample period.

One may expect that, during the period of deregulation, firms tend to become more efficient due to increased competitive pressures in the industry.[6] According to our results, however, there has not been efficiency improvements in the U.S. banks of limited branching regulatory environment. To shed a light on the observation that efficiencies decreased even under deregulation, we note Adams, Berger, and Sickles's (1999) comment that analysis of banks with varying total asset size might be more appropriate to account for possible heterogeneities across banks. Thus we divided the sample into two groups by bank size. It is known that bank size (in terms of total assets) is highly correlated with the size of a given output (Jayasiriya 2000). So we divided the sample into large and small banks depending on whether the output of a bank is larger than the median output of entire banks.

Coefficients estimates for large and small banks are not reported here because they are very similar to those for entire banks. Table 7 shows the efficiency estimates for large and small banks. With only large banks, we find that there is an increase in the overall level of technical efficiency. It ranges from 60.5% to 71.6%. Figure 2 displays the time pattern of efficiencies for large banks. Notice that the pattern is very close to that of efficiencies for entire banks, although efficiency levels and magnitude of fluctuations are somewhat different. The average technical efficiency estimates from the KSS estimator decrease from 61.6% in 1984 to 54.0% in 1995. With only small banks, however, we get different results. Overall efficiency level for small banks is lower than that for large banks, ranging from 43.8% to 61.9%. However, efficiency estimates from the KSS estimator now increase from 57.8% in 1984 to 62.2% in 1995. This results provide us with information which is not seen from entire banks, and we may conclude that small banks responded more efficiently to deregulation during the sample period.

6. Conclusion

In this paper we introduced a new approach to estimating time-varying technical efficiency levels for individual firms, without making strong distributional assumptions for technical inefficiency or random noise. We do so by using the fixed effects model allowing for temporal variations in individual effects. More specifically, we estimate the effects using the procedure combining smoothing spline techniques with principal component analysis. In this way, we can approximate virtually any shapes of time-varying effects.

[6] For comprehensive discussions of deregulatory issues and the banking industry's reactions and adjustments to them, see Berger et al. (1995).

Simulation experiments show that previous estimators, which fail to allow for temporal variations in effects terms or misspecify the temporal pattern of variations, suffer from serious distortions in their estimates. On the other hand, our new estimator performs very well regardless of the assumption on the temporal pattern of individual effects. We have used this estimator to analyze the technical efficiency of U.S. banks in limited branching regulatory environment for the period of 1984-1995, and discovered that small banks became more efficient over the years, while large banks suffered efficiency loss.

6. Appendix: Mathematical Proofs

Proof of Theorem 1: It is easily seen that

$$\begin{aligned}
\hat{\beta} &= (\sum_i \tilde{X}_i'(I-\mathcal{Z}_\kappa)\tilde{X}_i)^{-1}\sum_i \tilde{X}_i'(I-\mathcal{Z}_\kappa)(Y_i-\bar{Y}) \\
&= \beta + (\sum_i \tilde{X}_i'(I-\mathcal{Z}_\kappa)\tilde{X}_i)^{-1}\sum_i \tilde{X}_i'(I-\mathcal{Z}_\kappa)v_i \\
&\quad +(\sum_i \tilde{X}_i'(I-\mathcal{Z}_\kappa)\tilde{X}_i)^{-1}\sum_i \tilde{X}_i'(I-\mathcal{Z}_\kappa)(\epsilon_i-\hat{\epsilon}).
\end{aligned}$$

Consequently, $\mathbf{E}_\epsilon(\hat{\beta})-\beta = (\sum_i \tilde{X}_i'(I-\mathcal{Z}_\kappa)\tilde{X}_i)^{-1}\sum_i \tilde{X}_i'(I-\mathcal{Z}_\kappa)v_i$. By Assumption 1) there exists a fixed basis $b_1,\ldots,b_L$ of $\mathcal{L}_T$ which can be chosen independent of X_{it}. Therefore, $v_i=\sum_{r=1}^L \vartheta_{ir}b_r$. Let X_{ij} denote the T-vectors with elements X_{itj}, $t=1,\ldots,T$. In the general case, the $j=1,\ldots,p$ elements of the vectors $\sum_i \tilde{X}_i'(I-\mathcal{Z}_\kappa)v_i$ can thus be bounded by

$$\begin{aligned}
|\sum_i \tilde{X}_{ij}'(I-\mathcal{Z}_\kappa)v_i| &\leq n\sum_{r=1}^L \sqrt{|\frac{1}{n}\sum_i \vartheta_{ir}^2|\cdot|b_r'(I-\mathcal{Z}_\kappa)(\frac{1}{n}\sum_i \tilde{X}_{ij}\tilde{X}_{ij}')(I-\mathcal{Z}_\kappa)b_r|} \\
&= O_P\left(n\sum_{r=1}^L \sqrt{\mathbf{E}(\vartheta_{ir}^2)\cdot|b_r'(I-\mathcal{Z}_\kappa)\mathbf{E}(\tilde{X}_{ij}\tilde{X}_{ij}')(I-\mathcal{Z}_\kappa)b_r|}\right)
\end{aligned}$$

But by Assumptions 4) and 5) we obtain

$$n\sum_{r=1}^L \sqrt{\mathbf{E}(\vartheta_{ir}^2)\cdot|b_r'(I-\mathcal{Z}_\kappa)\mathbf{E}(\tilde{X}_{ij}\tilde{X}_{ij}')(I-\mathcal{Z}_\kappa)b_r|} \leq n\sum_{r=1}^L \sqrt{\mathbf{E}(\vartheta_{ir}^2)\cdot D\cdot\|(I-\mathcal{Z}_\kappa)b_r\|^2} = O(n\sqrt{T}b_v(n,T)).$$

Condition (**??**) of Assumption 5) then leads to $\|\mathbf{E}_\epsilon(\hat{\beta})-\beta\| = O_P((\frac{b_v(n,T)}{T^{1/2}})$. On the other hand, if v_i and X_i are uncorrelated, then

$$\begin{aligned}
|\sum_i \tilde{X}_{ij}'(I-\mathcal{Z}_\kappa)v_i| &= O_P\left(\sqrt{n\cdot\mathbf{E}(\vartheta_{ir}^2)|b_r'(I-\mathcal{Z}_\kappa)\mathbf{E}(\tilde{X}_{ij}\tilde{X}_{ij}'))(I-\mathcal{Z}_\kappa)b_r|}\right) \\
&= O_P(\sqrt{nT\cdot b_v(n,T)^2})
\end{aligned}$$

and $\|\mathbf{E}_\epsilon(\hat{\beta})-\beta\| = O_P((nT)^{-1/2}\cdot b_v(n,T))$. By Assumptions 5) and 6) the assertion on $\hat{\beta}-\mathbf{E}_\epsilon(\hat{\beta}) = (\sum_i \tilde{X}_i'(I-\mathcal{Z}_\kappa)\tilde{X}_i)^{-1}\sum_i \tilde{X}_i'(I-\mathcal{Z}_\kappa)(\epsilon_i-\bar{\epsilon})$ follows from standard arguments.

Consider Assertion (b). Obviously,

$$w - \hat{w} = (I - \mathcal{Z}_\kappa)w - \mathcal{Z}_\kappa\bar{\epsilon} - \mathcal{Z}_\kappa\bar{X}(\beta - \hat{\beta})$$

and $T^{-1/2}\|\mathcal{Z}_\kappa\bar{\epsilon}\| = O_P(\sqrt{tr(\mathcal{Z}_\kappa)/(nT)})$. The assertion then follows from Assumptions 4) and 5) as well as the above results on the convergence of $\|\beta - \hat{\beta}\|$.

In order to prove Assertion (c) first note that

$$\hat{v}_i = v_i + r_i, \quad \text{with } r_i = -(I - \mathcal{Z}_\kappa)v_i + \mathcal{Z}_\kappa\epsilon_i + \mathcal{Z}_\kappa\tilde{X}_i(\beta - \hat{\beta}).$$

Therefore,

$$\hat{\Sigma}_n = \Sigma_n + B, \quad B = \frac{1}{n}\sum_i (v_i r_i' + r_i v_i' + r_i r_i'). \tag{23}$$

Assertion (b) of Lemma A.1 of Kneip and Utikal (2001) implies that for all $r = 1, \ldots, L$

$$\gamma_r - \hat{\gamma}_r = -S_r B \gamma_r + R, \quad \text{with } \|R\| \leq \frac{6\sup_{\|a\|=1} a'B'Ba}{\min_s |\lambda_r - \lambda_s|^2} \tag{24}$$

and with $S_r = \sum_{s \neq r} \frac{1}{\lambda_s - \lambda_r} P_s$, where P_s denotes the projection matrix projecting into the eigenspace corresponding to λ_s.

In order to evaluate the above expression we first have to analyze the stochastic order of magnitude of the different elements of B. Consider the terms appearing in $\frac{1}{n}\sum_i (v_i r_i' + r_i v_i')$. Using Assumptions 1) - 5) some straightforward arguments now lead to

$$\sup_{\|a\|=1} \|\frac{1}{n}\sum_i (I - \mathcal{Z}_\kappa)v_i v_i' a\| \leq \frac{1}{n}\sum_i \sup_{\|a\|=1} |v_i'a| \sqrt{v_i'(I - \mathcal{Z}_\kappa)(I - \mathcal{Z}_\kappa)v_i} = O_P(Tc(T)^{1/2} b_v(n,T)), \tag{25}$$

$$\sup_{\|a\|=1} \|\frac{1}{n}\sum_i v_i v_i'(I - \mathcal{Z}_\kappa)a\| \leq \sup_{\|a\|=1} \frac{1}{n}\sum_i \sqrt{v_i'v_i}\, |v_i'(I - \mathcal{Z}_\kappa)a| = O_P(Tc(T)^{1/2} b_v(n,T)), \tag{26}$$

$$\begin{aligned} \sup_{\|a\|=1} \|\frac{1}{n}\sum_i (\mathcal{Z}_\kappa\tilde{X}_i(\beta - \hat{\beta}))v_i'a\| &\leq \frac{1}{n}\sum_i |v_i'a| \sqrt{(\beta - \hat{\beta})'\tilde{X}_i'\mathcal{Z}_\kappa^2\tilde{X}_i(\beta - \hat{\beta})} \\ &= O_P\left(Tc(T)^{1/2}(b_\beta(n,T) + \frac{1}{\sqrt{nT}})\right). \end{aligned} \tag{27}$$

If X_i and v_i are uncorrelated, then due to $\mathbf{E}(X_{ij}v_i') = 0$ relation (**??**) can be replaced by

$$\sup_{\|a\|=1} \|\frac{1}{n}\sum_i (\mathcal{Z}_\kappa\tilde{X}_i(\beta - \hat{\beta}))v_i'a\| = O_P\left(T\sqrt{\frac{c(T)}{n}}(b_\beta(n,T) + \frac{1}{\sqrt{nT}})\right). \tag{28}$$

By similar arguments

$$\sup_{\|a\|=1} \|\frac{1}{n}\sum_i v_i(\mathcal{Z}_\kappa\tilde{X}_i(\beta - \hat{\beta}))'a\| = \begin{cases} O_P(T\sqrt{\frac{c(T)}{n}}(b_\beta(n,T) + \frac{1}{\sqrt{nT}})) & \text{if } X_i \text{ and } v_i \text{ are uncorrelated} \\ O_P(Tc(T)^{1/2} b_\beta(n,T)) & \text{else} \end{cases} \tag{29}$$

Obviously, $\mathbf{E}_\epsilon(tr((\frac{1}{n}\sum_i v_i\epsilon_i'\mathcal{Z}_\kappa)\cdot(\frac{1}{n}\sum_i \mathcal{Z}_\kappa\epsilon_i v_i'))) = O(\frac{Tc(T)\cdot tr(\mathcal{Z}_\kappa^2)}{n})$, and therefore

$$\sup_{\|a\|=1} \|\frac{1}{n}\sum_i \mathcal{Z}_\kappa\epsilon_i v_i'\gamma_r\|^2 \leq tr((\frac{1}{n}\sum_i v_i\epsilon_i'\mathcal{Z}_\kappa)\cdot(\frac{1}{n}\sum_i \mathcal{Z}_\kappa\epsilon_i v_i')')) = O_P\left(\sqrt{\frac{Tc(T)\cdot tr(\mathcal{Z}_\kappa^2)}{n}}\right), \quad (30)$$

Similarly,

$$\sup_{\|a\|=1} \|\frac{1}{n}\sum_i v_i\epsilon_i'\mathcal{Z}_\kappa\gamma_r\|^2) = O_P\left(\sqrt{\frac{Tc(T)\cdot tr(\mathcal{Z}_\kappa^2)}{n}}\right). \quad (31)$$

For the leading terms appearing in $\frac{1}{n}\sum_i r_i r_i'$ we obtain

$$\sup_{\|a\|=1} \|\frac{1}{n}\sum_i (I-\mathcal{Z}_\kappa)v_i v_i'(I-\mathcal{Z}_\kappa)a\| = O_p(T\cdot b_v(n,T)^2), \quad (32)$$

$$\sup_{\|a\|=1} \|\frac{1}{n}\sum_i (\mathcal{Z}_\kappa\tilde{X}_i(\beta-\hat{\beta}))(\mathcal{Z}_\kappa\tilde{X}_i(\beta-\hat{\beta}))'a\| = O_P\left(T\cdot(b_\beta(n,T)^2+\frac{1}{nT})\right). \quad (33)$$

Obviously,
$\mathbf{E}_\epsilon(tr((\frac{1}{n}\sum_i \mathcal{Z}_\kappa\epsilon_i\epsilon_i'\mathcal{Z}_\kappa - \sigma^2\mathcal{Z}_\kappa^2)\cdot(\frac{1}{n}\sum_i \mathcal{Z}_\kappa\epsilon_i\epsilon_i'\mathcal{Z}_\kappa - \sigma^2\mathcal{Z}_\kappa^2)))) = \frac{1}{n}\mathbf{E}(tr(\mathcal{Z}_\kappa\epsilon_i\epsilon_i'\mathcal{Z}_\kappa\mathcal{Z}_\kappa\epsilon_i\epsilon_i'\mathcal{Z}_\kappa - \sigma^4\mathcal{Z}_\kappa^4)) = O_P(\frac{tr(\mathcal{Z}_\kappa^2)^2}{n})$, and therefore

$$\sup_{\|a\|=1} \|\frac{1}{n}\sum_i (\mathcal{Z}_\kappa\epsilon_i\epsilon_i'\mathcal{Z}_\kappa - \sigma^2\mathcal{Z}_\kappa^2)a\| = O_P\left(\frac{tr(\mathcal{Z}_\kappa)^2}{\sqrt{n}}\right) \quad (34)$$

Assumptions 1) and 2) additionally imply that $\frac{1}{\min_s|\lambda_r-\lambda_s|} = O_P(\frac{1}{T\cdot C(T)})$. When combining (**??**) with (**??**) - (**??**) we thus obtain

$$\begin{aligned}\|S_r B\gamma_r\| &\leq \|\sigma^2 S_r\mathcal{Z}_\kappa^2\gamma_r\| + \frac{1}{\min_s|\lambda_r-\lambda_s|}\|(B-\sigma^2 S_r\mathcal{Z}_\kappa^2)\gamma_r\| \\ &= \|\sigma^2 S_r\mathcal{Z}_\kappa^2\gamma_r\| + O_P\left(\frac{b_v(n,T)}{c(T)^{1/2}} + \sqrt{\frac{tr(\mathcal{Z}_\kappa^2)}{nTc(T)}}\right) \quad (35)\end{aligned}$$

By definition of S_r we have $S_r\gamma_r = 0$. Furthermore, Assumption 3 implies that $\|(I-\mathcal{Z}_\kappa)\gamma_r\| = O_P(\frac{b_v(n,T)}{c(T)^{1/2}})$. Hence,

$$\|\sigma^2 S_r\mathcal{Z}_\kappa^2\gamma_r\| \leq \|\sigma^2 S_r(I-\mathcal{Z}_\kappa)\gamma_r\| + \|\sigma^2 S_r\mathcal{Z}_\kappa(I-\mathcal{Z}_\kappa)\gamma_r\| = O_P(\frac{b_v(n,T)}{Tc(T)^{3/2}}), \quad (36)$$

Let us now consider the remainder term R in (**??**). Note that all eigenvalues of $\mathcal{Z}_\kappa$ are less or equal to 1, and thus $\sup_{\|a\|=1} a'\mathcal{Z}_\kappa^4 a \leq 1$. Relations (**??**) - (**??**) then imply

$$\begin{aligned}\frac{\sup_{\|a\|=1} a'B'Ba}{\min_s|\lambda_r-\lambda_s|^2} &\leq 2\frac{\sup_{\|a\|=1} a'(B-\sigma^2\mathcal{Z}_\kappa^2)'(B-\sigma^2\mathcal{Z}_\kappa^2)a}{\min_s|\lambda_r-\lambda_s|^2} + 2\frac{\sup_{\|a\|=1} a'\mathcal{Z}_\kappa^4 a}{\min_s|\lambda_r-\lambda_s|^2} \\ &= O_P\left(\frac{b_v(n,T)^2}{c(T)} + \frac{1}{T^2c(T)^2} + \frac{tr(\mathcal{Z}_\kappa^2)}{nTc(T)})\right) \quad (37)\end{aligned}$$

By (**??**), (**??**), (**??**) and (**??**) Assertion (c) follows from

$$T^{-1/2}\|g_r - \hat{g}_r\| = \|\gamma_r - \hat{\gamma}_r\| = O_P\left(\frac{b_v(n,T)}{c(T)^{1/2}} + \frac{1}{T^2 c(T)^2} + \sqrt{\frac{tr(\mathcal{Z}_\kappa^2)}{nTc(T)}}\right). \qquad (38)$$

Let us switch to Assertion (d). Definition of $\hat{\theta}_{ir}$ as well as Assertions a) and c) imply that

$$\begin{aligned}\hat{\theta}_{ri} &= \frac{1}{T}\hat{g}_r'(Y_i - \bar{Y} - \tilde{X}_i\hat{\beta}) \\ &= \theta_{ri} + \frac{1}{T}g_r'(\epsilon_i - \bar{\epsilon}) + \frac{1}{T}(\hat{g}_r - g_r)'v_i + O_P(b_\beta(n,T) + o_P(T^{-1/2})\end{aligned}$$

However, one can infer from relations (**??**) - (**??**) that

$$\begin{aligned}\frac{1}{T}(\hat{g}_r - g_r)'v_i = \frac{1}{n\sqrt{T}}\sum_j \gamma_r' v_j v_j'(I - \mathcal{Z}_\kappa)S_r v_i + \frac{1}{n\sqrt{T}}\sum_j \gamma_r'(I - \mathcal{Z}_\kappa)v_j v_j' S_r v_i \\ + O_P\left(\frac{b_v(n,T)^2}{c(T)^{1/2}} + b_\beta(n,T) + \frac{tr(\mathcal{Z}_\kappa^2)}{nT}\right)\end{aligned}$$

However, the well-known properties of $\mathcal{Z}_\kappa$ imply that $\frac{1}{T}g_r'(I - \mathcal{Z}_\kappa)g_r$ is of the same order of magnitude as $\frac{1}{T}g_r'(I - \mathcal{Z}_\kappa)(I - \mathcal{Z}_\kappa)g_r$. Hence,

$$\frac{1}{n\sqrt{T}}\sum_j \gamma_r' v_j v_j'(I - \mathcal{Z}_\kappa)S_r v_i \leq \frac{1}{n}\sum_{s\neq r}\sum_j \frac{|v_i'\gamma_r|}{\sqrt{T}|\lambda_r - \lambda_s|}|v_j'(I - \mathcal{Z}_\kappa)\theta_{si}g_s| = O_P(\frac{b_v(n,T)^2}{c(T)^{1/2}})$$

as well as

$$\frac{1}{n\sqrt{T}}\sum_j \gamma_r'(I - \mathcal{Z}_\kappa)v_j v_j' S_r v_i \leq \frac{1}{n}\sum_j \frac{|v_i' v_j|}{\sqrt{T}\min_s|\lambda_r - \lambda_s|}|v_j'(I - \mathcal{Z}_\kappa)\gamma_r| = O_P(\frac{b_v(n,T)^2}{c(T)^{1/2}}).$$

This implies

$$(\hat{\theta}_{ri} - \theta_{ri}) = \frac{1}{T}g_r'\epsilon_i + O_P\left(\frac{b_v(n,T)^2}{c(T)^{1/2}} + b_\beta(n,T) + \frac{tr(\mathcal{Z}_\kappa^2)}{nT}\right) + o_P\left(T^{-1/2}\right)$$

Since $\frac{1}{T}g_r'g_r = 1$ we immediately obtain $\sqrt{T}\cdot\frac{1}{T}g_r'\epsilon_i \to_d \mathbf{N}(0,\sigma^2)$. The asserted rate of convergence is an immediate consequence. Note that due to $g_r'g_s = 0$ the random variables $g_r'\epsilon_i$ and $g_s'\epsilon_i$ are uncorrelated for $r \neq s$. Hence, if additionally $\frac{b_v(n,T)^2}{c(T)^{1/2}} + b_\beta(n,T) + \frac{tr(\mathcal{Z}_\kappa^2)}{nT} = o(T^{-1/2})$, the assertion on the multivariate distribution of $\sqrt{T}(\hat{\theta}_{1i} - \theta_{1i}, \ldots, \hat{\theta}_{Li} - \theta_{Li})'$ follows from standard arguments.

It remains to prove assertion (e). First note that

$$\hat{v}_i = \mathcal{Z}_\kappa v_i + \tilde{r}_i, \quad \text{with } \tilde{r}_i = \mathcal{Z}_\kappa(\epsilon_i - \bar{\epsilon}) + \mathcal{Z}_\kappa\tilde{X}_i(\beta - \hat{\beta}).$$

Consequently, with $\tilde{\Sigma}_n = \mathcal{Z}_\kappa(\frac{1}{n}\sum_i v_i v_i')\mathcal{Z}_\kappa$ we obtain

$$\hat{\Sigma}_n = \tilde{\Sigma}_n + \tilde{B}, \quad \tilde{B} = \frac{1}{n}\sum_i (\mathcal{Z}_\kappa v_i \tilde{r}_i' + \tilde{r}_i v_i' \mathcal{Z}_\kappa + \tilde{r}_i \tilde{r}_i').$$

$\tilde{\Sigma}_n$ possesses only L nonzero eigenvalues $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_L$ with corresponding eigenvectors $\tilde{\gamma}_1, \ldots, \tilde{\gamma}_L$. Our assumptions and arguments similar to (**??**) - (**??**) then show that $\tilde{\lambda}_r = O(Tc(T))$, $\frac{1}{\min_s |\tilde{\lambda}_r - \tilde{\lambda}_s|} = O_P(\frac{1}{T \cdot C(T)})$, $\|\gamma_r - \tilde{\gamma}_r\| = O_P(b(\kappa))$, and $\|\hat{\gamma}_r - \tilde{\gamma}_r\| = O_P(B(n,T) + \frac{1}{T^2} + \frac{tr(\mathcal{Z}_\kappa^2)}{nT})$ for all $r, s = 1, \ldots, L$, $r \neq s$.

Assertion (a) of Lemma A.1. of Kneip and Utikal (2001) implies that

$$\sum_{r=L+1}^{T} \hat{\lambda}_r = tr(\mathcal{P}_L \tilde{B}) + R^*, \quad \text{with } R^* \leq \frac{6L \sup_{\|a\|=1} a' \tilde{B}' \tilde{B} a}{\min_s |\tilde{\lambda}_r - \tilde{\lambda}_s|} \tag{39}$$

where $\mathcal{P}_L = I - \sum_{r=1}^{L} \tilde{\gamma}_r \tilde{\gamma}_r'$. Using again arguments similar to the proof of Assertion (c) it is easily seen that

$$\frac{6L \sup_{\|a\|=1} a' \tilde{B}' \tilde{B} a}{\min_s |\tilde{\lambda}_r - \tilde{\lambda}_s|} = O_P\left(T \cdot b_\beta(n,T)^2 + \frac{1}{Tc(T)} + \frac{tr(\mathcal{Z}_\kappa^2)}{n}\right). \tag{40}$$

On the other hand,

$$tr(\mathcal{P}_L \tilde{B}) = tr\left(\frac{1}{n}\sum_i \mathcal{P}_L \mathcal{Z}_\kappa \tilde{X}_i (\beta - \hat{\beta})(\beta - \hat{\beta})' \tilde{X}_i' \mathcal{Z}_\kappa\right) + tr\left(\mathcal{P}_L \mathcal{Z}_\kappa (\frac{1}{n}\sum_i (\epsilon_i - \bar{\epsilon})(\epsilon_i - \bar{\epsilon})')\mathcal{Z}_\kappa\right) \tag{41}$$

Some straightforward computations lead to

$$\mathbf{E}\left(tr(\mathcal{P}_L \mathcal{Z}_\kappa (\frac{1}{n}\sum_i (\epsilon_i - \bar{\epsilon})(\epsilon_i - \bar{\epsilon})')\mathcal{Z}_\kappa)\right) = \sigma^2 (1 - \frac{1}{n}) tr(\mathcal{Z}_\kappa \mathcal{P}_L \mathcal{Z}_\kappa),$$

$$\text{Var}\left(tr(\mathcal{P}_L \mathcal{Z}_\kappa (\frac{1}{n}\sum_i (\epsilon_i - \bar{\epsilon})(\epsilon_i - \bar{\epsilon})')\mathcal{Z}_\kappa)\right) = \frac{2\sigma^4}{n} \cdot tr((\mathcal{Z}_\kappa \hat{P}_L \mathcal{Z}_\kappa)^2) \cdot (1 + o_P(1)) = O_P\left(\frac{tr(\mathcal{Z}_\kappa^4)}{n}\right)$$

Since $tr(\frac{1}{n}\sum_i \mathcal{P}_L \mathcal{Z}_\kappa \tilde{X}_i (\beta - \hat{\beta})(\beta - \hat{\beta})' \tilde{X}_i' \mathcal{Z}_\kappa \mathcal{P}_L) = O_P\left(T \cdot b_\beta(n,T)^2 + \frac{1}{n}\right)$ and since by assumption $T \cdot b_\beta(n,T)^2 = o\left(\sqrt{tr(\mathcal{Z}_\kappa^4)/n}\right)$ one may invoke standard arguments to show that

$$\frac{\sum_{r=L+1}^{h} \hat{\lambda}_r \ - \sigma^2 \left(1 - \frac{1}{n}\right) tr(\mathcal{Z}_\kappa \mathcal{P}_L \mathcal{Z}_\kappa)}{\sqrt{\frac{2\sigma^4}{n} \cdot tr((\mathcal{Z}_\kappa \mathcal{P}_L \mathcal{Z}_\kappa)^2)}} \rightarrow_d \mathbf{N}(0,1). \tag{42}$$

By (**??**), Relation (**??**) remains valid when $\mathcal{P}_L$ is replaced by $\hat{P}_L$. This proves assertion (e).
□

Proof of Theorem 2: It follows from arguments similar to those used in the proof of Theorem 1 that

$$\hat{\sigma}^2 = \frac{1}{(n-1)\cdot tr(I-\mathcal{Z}_\kappa)^2}\sum_i (\epsilon_i-\bar{\epsilon})'(I-\mathcal{Z}_\kappa)^2(\epsilon_i-\bar{\epsilon})$$
$$+\frac{1}{(n-1)\cdot tr(I-\mathcal{Z}_\kappa)^2}\sum_i v_i'(I-\mathcal{Z}_\kappa)^2 v_i + O_P\left(b_v(n,T)\cdot(b_\beta(n,T)+\frac{1}{\sqrt{nT}}\right).$$

Clearly,

$$\mathbf{E}\left(\frac{1}{(n-1)\cdot tr(I-\mathcal{Z}_\kappa)^2}\sum_i (\epsilon_i-\bar{\epsilon})'(I-\mathcal{Z}_\kappa)^2(\epsilon_i-\bar{\epsilon})\right)=\sigma^2$$

and since $tr(I-\mathcal{Z}_\kappa)\geq D_1\cdot T$

$$\text{Var}\left(\frac{1}{(n-1)\cdot tr(I-\mathcal{Z}_\kappa)^2}\sum_i (\epsilon_i-\bar{\epsilon})'(I-\mathcal{Z}_\kappa)^2(\epsilon_i-\bar{\epsilon})\right)=O\left(\frac{1}{nT}\right).$$

Consequently, with

$$0\leq R_{n,T}=\frac{1}{(n-1)\cdot tr(I-\mathcal{Z}_\kappa)^2}\sum_i v_i'(I-\mathcal{Z}_\kappa)^2 v_i = O_p(b_v(n,T)^2) \tag{43}$$

we obtain

$$\hat{\sigma}^2=\sigma^2+R_{n,T}+o_p(1). \tag{44}$$

Let us now consider the behavior of $C(l)$ for $l<L$. We can immediately infer from (**??**) that

$$C(l)=\left[\frac{n\sum_{r=l+1}^{L}\hat{\lambda}_r \; -(n-1)(\sigma^2+R_{n,T})\cdot tr(\mathcal{Z}_\kappa(\hat{\mathcal{P}}_l-\hat{\mathcal{P}}_L)\mathcal{Z}_\kappa)-(n-1)R_{n,T}\cdot tr(\mathcal{Z}_\kappa\hat{\mathcal{P}}_l\mathcal{Z}_\kappa)}{\sqrt{2n\hat{\sigma}^4\cdot tr((\mathcal{Z}_\kappa\hat{\mathcal{P}}_l\mathcal{Z}_\kappa)^2)}}\right. \tag{45}$$

$$\left.+\frac{n\sum_{r=L+1}^{T}\hat{\lambda}_r \; -(n-1)\sigma^2\cdot tr(\mathcal{Z}_\kappa\hat{\mathcal{P}}_L\mathcal{Z}_\kappa)}{\sqrt{2n\hat{\sigma}^4\cdot tr((\mathcal{Z}_\kappa\hat{\mathcal{P}}_l\mathcal{Z}_\kappa)^2)}}\right](1+o_P(1)). \tag{46}$$

By Assumption 2) $n\sum_{r=l+1}^{L}\hat{\lambda}_r=\sum_{r=l+1}^{L}T\sum_i\theta_{ir}^2$ is of order $nTc(T)$, while $(n-1)(\sigma^2+R_{n,T})\cdot tr(\mathcal{Z}_\kappa(\hat{\mathcal{P}}_l-\hat{\mathcal{P}}_L)\mathcal{Z}_\kappa)=O_P(n)$, $(n-1)R_{n,T}\cdot tr(\mathcal{Z}_\kappa\hat{\mathcal{P}}_l\mathcal{Z}_\kappa)=o_P(nTc(T))$, and $\sqrt{2n\hat{\sigma}^4\cdot tr((\mathcal{Z}_\kappa\hat{\mathcal{P}}_l\mathcal{Z}_\kappa)^2)}=O_P((nT)^{1/2})$. Consequently, the term on the right hand side of (**??**) increases as $n,T\to\infty$, while the first term in (**??**) is still bounded in probability. We can thus infer that for $l<L$

$$\mathbf{P}(C(l)>z_{1-\alpha})\to 1 \quad \text{and therefore } \mathbf{P}(\hat{L}\neq l)\to 1 \tag{47}$$

as $n,T\to\infty$.

For $l = L$ we obtain

$$C(l) = \frac{n\sum_{r=L+1}^{T}\hat{\lambda}_r - (n-1)\sigma^2 \cdot tr(\mathcal{Z}_\kappa \hat{\mathcal{P}}_L \mathcal{Z}_\kappa)}{\sqrt{2n\sigma^4 \cdot tr((\mathcal{Z}_\kappa \hat{\mathcal{P}}_L \mathcal{Z}_\kappa)^2)}}$$

$$+ \left(\frac{n\sum_{r=L+1}^{T}\hat{\lambda}_r}{\sqrt{2n(\sigma^4 + R_{n,T}) \cdot tr((\mathcal{Z}_\kappa \hat{\mathcal{P}}_L \mathcal{Z}_\kappa)^2)}} - \frac{n\sum_{r=L+1}^{T}\hat{\lambda}_r}{\sqrt{2n\sigma^4 \cdot tr((\mathcal{Z}_\kappa \hat{\mathcal{P}}_L \mathcal{Z}_\kappa)^2)}} \right)(1 + o_P(1))$$

Since $R_{n,T} \geq 0$ we can infer from Theorem 1(e) that

$$\limsup_{n,T\to\infty} \mathbf{P}(C(L) \geq z_{1-\alpha}) \leq \alpha. \tag{48}$$

The assertion of the theorem now is an immediate consequence of (**??**) and (**??**). □

Proof of Theorem 3: When expanding D we obviously obtain

$$D = 2 - \frac{\sum_{i=1}^{n}\hat{\epsilon}_{i1}^2 + \sum_{i=1}^{n}\hat{\epsilon}_{iT}^2 + 2\sum_{i=1}^{n}\sum_{t=2}^{T}\hat{\epsilon}_{it}\hat{\epsilon}_{i,t-1}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\epsilon}_{it}^2}. \tag{49}$$

It follows from standard arguments of linear regression theory that

$$\mathbf{E}\left(\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\epsilon}_{it}^2\right) = \sigma^2(n(T-1)-p), \quad \mathrm{Var}\left(\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\epsilon}_{it}^2\right) = 2\sigma^4(n(T-1)-p) \tag{50}$$

and under H_0

$$\hat{\beta} = \beta + \left(\sum_i X_i'(I - \frac{1}{T}\mathbf{1}\mathbf{1}')X_i\right)^{-1} (\sum_i X_i'(I - \frac{1}{T}\mathbf{1}\mathbf{1}')\epsilon_i$$

which due to our assumptions implies that $\beta - \hat{\beta} = O_P((nT)^{-1/2})$ and $|\beta - \hat{\beta}|^2 = O_P((nT)^{-1})$ as well as

$$\sum_i\sum_t \epsilon_{it}(X_{i,t-1} - \frac{1}{T}\mathbf{1}'X_i)(\beta - \hat{\beta}) = O_P(1), \tag{51}$$

$$\sum_i\sum_t (X_{it} - \frac{1}{T}\mathbf{1}'X_i)(X_{i,t-1} - \frac{1}{T}\mathbf{1}'X_i)(\beta - \hat{\beta})^2 = O_P(1). \tag{52}$$

By definition,

$$\hat{\epsilon}_i = \epsilon_{it} - \bar{\epsilon}_i + (X_{it} - \frac{1}{T}\mathbf{1}'X_i)(\beta - \hat{\beta}),$$

where $\bar{\epsilon}_i = \frac{1}{T}\sum_t \epsilon_{it}$, and from (**??**) - (**??**) we can thus infer that

$$D = 2 - \frac{\sum_{i=1}^{n}(\epsilon_{i1} - \bar{\epsilon}_i)^2 + \sum_{i=1}^{n}(\epsilon_{iT} - \bar{\epsilon}_i)^2 + 2\sum_{i=1}^{n}\sum_{t=2}^{T}(\epsilon_{it} - \bar{\epsilon}_i)(\epsilon_{i,t-1} - \bar{\epsilon}_i)}{\sigma^2 n(T-1)} + O_P\left(\frac{1}{nT}\right). \tag{53}$$

Some straightforward computations now show that

$$\mathbf{E}\left(\sum_{i=1}^{n}(\epsilon_{i1}-\bar{\epsilon}_i)^2+\sum_{i=1}^{n}(\epsilon_{iT}-\bar{\epsilon}_i)^2+2\sum_{i=1}^{n}\sum_{t=2}^{T}(\epsilon_{it}-\bar{\epsilon}_i)(\epsilon_{i,t-1}-\bar{\epsilon}_i)\right)=0, \quad (54)$$

$$\mathrm{Var}\left(\sum_{i=1}^{n}(\epsilon_{i1}-\bar{\epsilon}_i)^2+\sum_{i=1}^{n}(\epsilon_{iT}-\bar{\epsilon}_i)^2+2\sum_{i=1}^{n}\sum_{t=2}^{T}(\epsilon_{it}-\bar{\epsilon}_i)(\epsilon_{i,t-1}-\bar{\epsilon}_i)\right)=\frac{4\sigma^4 n(T-1)^2}{T}. \quad (55)$$

When additionally applying standard central limit theorems, the assertion of the theorem follows from (**??**) -(**??**). □

7. References

Adams, R.M., A.N. Berger, and R.C. Sickles (1999), "Semiparametric approaches to stochastic panel frontiers with applications in the banking industry," *Journal of Business and Economic Statistics*, 17, 349-358.

Aigner, D.J., C.A.K. Lovell, and P. Schmidt (1977), "Formulation and estimation of stochastic frontier models," *Journal of Econometrics,* 6, 21-37.

Baltagi, B. and J. M. Griffin (1988), "A general index of technical change," *Journal of Political Economy*, 96, 20-41.

Battese, G.E. and 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.

Berger, A.N. (1993), ""Distribution-Free" estimates of efficiency in U.S. banking industry and tests of the standard distributional assumption," *Journal of Productivity Analysis,* 4, 261-292.

Berger, A.N., A.K. Kashyap, and J.M. Scalise (1995), "The transformation of the U.S. banking industry: What a long, strange trip it's been," *Brookings Papers on Economic Activity*, 2, 55–218.

de Boor, C, (1978), *A Practical Guide to Splines*, Springer-Velag, New York.

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

Eubank, R.L. (1999), *Nonparametric Regression and Spline Smoothing*, Marcel Dekker, New York.

Jayasiriya, R. (2000), "Essays on structural modeling using nonparametric and parametric methods with applications in the U.S. banking industry," unpublished Ph.D. dissertation, Rice University.

Kneip, A. (1994), "Nonparametric estimation of common regressors for similar curve data," *Annals of Statistics*, 22, 1386-1427.

Kneip, A. and K.J. Utikal (2001), "Inference for density families using functional principal component analysis," *Journal of American Statistical Association*, 96, 519-532.

Meeusen, W. and J. van den Broeck (1977), "Efficiency estimation from Cobb-Douglas production functions with composed error," *International Economic Review*, 18, 435-444.

Park, B.U., R.C. Sickles, and L. Simar (2002), "Semiparametric efficient estimation of AR(1) panel data models," Manuscript.

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

Speckman, P. (1988), "Kernel smoothing in partial linear models," *Journal of the Royal Statistical Society, Series B*, 50, 413-436.

Table 1. Monte Carlo Simulation Results for DGP1

MSE of Coefficients*

N	T	Within	GLS	BC	CSS	KSS
30	12	0.9107	0.6039	0.4933	0.8863	0.4998
	30	4.5286	4.0001	1.1767	0.2329	0.1462
100	12	0.2635	0.1438	0.1454	0.2504	0.1170
	30	1.2219	1.0068	1.4172	0.0726	0.0410
300	12	0.0801	0.0402	0.0360	0.0790	0.0343
	30	0.3409	0.2848	0.1456	0.0258	0.0151

MSE of Effects

N	T	Within	GLS	CSS	KSS	L
30	12	0.6159	0.5692	0.4675	0.2278	1.1200
	30	0.4476	0.4455	0.0051	0.0037	1.0510
100	12	0.5940	0.5755	0.4438	0.1769	1.0620
	30	0.4539	0.4531	0.0050	0.0100	1.0590
300	12	0.6068	0.5990	0.5504	0.1964	1.0341
	30	0.4379	0.4376	0.0064	0.0025	1.0500

MSE of Efficiencies

N	T	Within	GLS	BC	CSS	KSS
30	12	0.3429	0.3255	0.1485	0.3329	0.0921
	30	0.6967	0.7005	0.8430	0.2069	0.0289
100	12	0.4415	0.4294	0.3817	0.3969	0.0529
	30	0.8305	0.8279	1.1184	0.2790	0.0236
300	12	0.5102	0.5070	0.4574	0.4575	0.0364
	30	0.9401	0.9400	1.6111	0.3470	0.0154

Spearman Rank Correlation of Efficiencies

N	T	Within	GLS	BC	CSS	KSS
30	12	0.5052	0.5004	0.8085	0.7692	0.9806
	30	0.4829	0.4834	0.7533	0.9841	0.9980
100	12	0.3886	0.3886	0.5656	0.7837	0.9923
	30	0.3885	0.3885	0.5900	0.9871	0.9993
300	12	0.3037	0.3037	0.6267	0.7771	0.9924
	30	0.2805	0.2805	0.5469	0.9878	0.9995

Note: * is multiplied by 10^2.

Table 2. Monte Carlo Simulation Results for DGP2

MSE of Coefficients*

N	T	Within	GLS	BC	CSS	KSS
30	12	2.2939	1.6274	0.3427	0.8901	0.4661
	30	161.0314	106.1230	9.6053	5.4253	0.1499
100	12	0.7709	0.6094	0.1149	0.2505	0.1206
	30	53.4336	39.4729	8.1635	1.9065	0.0403
300	12	0.2873	0.1760	0.0339	0.0800	0.0371
	30	18.4371	11.9706	1.3051	0.6689	0.0141

MSE of Effects

N	T	Within	GLS	CSS	KSS	L
30	12	0.3892	0.3753	0.0699	0.1401	1.0720
	30	0.7443	0.7351	0.0202	0.0705	1.0430
100	12	0.4678	0.4642	0.0701	0.2120	1.0350
	30	0.8029	0.8007	0.0217	0.1024	1.0050
300	12	0.4475	0.4452	0.0617	0.1966	1.0260
	30	0.7911	0.7902	0.0213	0.0986	1.0020

MSE of Efficiencies

N	T	Within	GLS	BC	CSS	KSS
30	12	0.2260	0.1951	0.0321	0.2586	0.0786
	30	0.7924	0.7321	0.0096	0.5236	0.0544
100	12	0.2598	0.2473	0.0400	0.2944	0.0787
	30	0.7361	0.7548	0.0091	0.5788	0.0116
300	12	0.2695	0.2618	0.0338	0.3607	0.0916
	30	0.7542	0.7342	0.0213	0.5568	0.0040

Spearman Rank Correlation of Efficiencies

N	T	Within	GLS	BC	CSS	KSS
30	12	0.8941	0.8914	0.9950	0.9716	0.9976
	30	0.6239	0.6293	0.9993	0.8871	0.9946
100	12	0.8283	0.8249	0.9981	0.9784	0.9966
	30	0.5349	0.5342	0.9997	0.8917	0.9999
300	12	0.8448	0.8446	0.9982	0.9726	0.9938
	30	0.5478	0.5479	0.9982	0.8820	1.0000

Note: * is multiplied by 10^2.

Table 3. Monte Carlo Simulation Results for DGP3

MSE of Coefficients*

N	T	Within	GLS	BC	CSS	KSS
30	12	1.6631	0.6852	0.6986	2.7261	0.7099
	30	0.5340	0.2621	0.2779	0.6766	0.1821
100	12	0.4224	0.1597	0.1649	0.6866	0.1290
	30	0.1468	0.0667	0.0715	0.1853	0.0396
300	12	0.1549	0.0606	0.0638	0.2429	0.0378
	30	0.0516	0.0250	0.0281	0.0649	0.0138

MSE of Effects

N	T	Within	GLS	CSS	KSS	L
30	12	1.0897	1.0259	1.1143	0.2710	2.1609
	30	1.0432	1.0240	1.0840	0.1140	2.0483
100	12	1.0602	1.0393	1.0672	0.2351	2.0585
	30	1.0364	1.0294	1.0829	0.0929	2.0102
300	12	1.0424	1.0353	1.0197	0.2081	2.0061
	30	1.0307	1.0285	1.0734	0.0822	2.0021

MSE of Efficiencies

N	T	Within	GLS	BC	CSS	KSS
30	12	2.1298	2.4086	7.9252	1.4860	0.2583
	30	2.2636	2.5640	5.0451	1.6066	0.1031
100	12	2.4655	2.6934	12.8728	1.4582	0.2175
	30	7.1729	7.6171	18.6293	4.2421	0.1109
300	12	3.8455	3.9679	25.7966	1.9365	0.2085
	30	8.9848	9.2055	26.4074	4.8352	0.1122

Spearman Rank Correlation of Efficiencies

N	T	Within	GLS	BC	CSS	KSS
30	12	0.1754	0.1729	0.0408	0.2535	0.9298
	30	0.0597	0.0600	-0.0181	0.0019	0.9842
100	12	0.2050	0.2051	0.1513	0.2674	0.9277
	30	0.0499	0.0498	0.0477	0.0325	0.9731
300	12	0.2131	0.2130	0.0754	0.2615	0.9236
	30	0.0575	0.0574	0.0136	-0.0248	0.9691

Note: * is multiplied by 10^2.

Table 4. Monte Carlo Simulation Results for DGP4

MSE of Coefficients*

N	T	Within	GLS	BC	CSS	KSS
30	12	0.5732	0.3586	0.3734	0.8634	0.6515
	30	0.2023	0.1513	0.1504	0.2319	0.2292
100	12	0.1741	0.1346	0.1260	0.2529	0.1816
	30	0.0571	0.0537	0.0510	0.0695	0.0596
300	12	0.0609	0.0360	0.0364	0.0910	0.0617
	30	0.0218	0.0164	0.0142	0.0258	0.0221

MSE of Effects

N	T	Within	GLS	CSS	KSS	L
30	12	0.4390	0.3500	1.2061	0.5407	1.0250
	30	0.1681	0.1465	0.4526	0.2217	1.0130
100	12	0.2769	0.2631	0.8046	0.2988	1.0300
	30	0.1082	0.1065	0.3145	0.1186	1.0200
300	12	0.2689	0.2614	0.7959	0.2799	1.0250
	30	0.0969	0.0954	0.2871	0.1015	1.0220

MSE of Efficiencies

N	T	Within	GLS	BC	CSS	KSS
30	12	0.1211	0.0993	0.1178	0.2600	0.1344
	30	0.0488	0.0421	0.0416	0.1205	0.0595
100	12	0.1719	0.1622	0.0478	0.3488	0.1778
	30	0.0798	0.0763	0.0252	0.1857	0.0829
300	12	0.2124	0.2075	0.0449	0.4120	0.2157
	30	0.0914	0.0907	0.0231	0.2168	0.0938

Spearman Rank Correlation of Efficiencies

N	T	Within	GLS	BC	CSS	KSS
30	12	0.9964	0.9742	0.9738	0.9481	0.9955
	30	0.9982	0.9804	0.9787	0.9757	0.9977
100	12	0.9989	0.9883	0.9896	0.9106	0.9987
	30	0.9997	0.9946	0.9949	0.9528	0.9996
300	12	0.9997	0.9997	0.9995	0.8946	0.9996
	30	0.9997	0.9995	0.9997	0.9588	0.9997

Note: * is multiplied by 10^2.

Table 5. Data Description: Sample Means

Variable	Definition	Mean
reln	Log of real estate loans	9.334
ciln	Log of commercial and industrial loans	8.215
inln	Log of installment loans	8.424
CD	Log of certificate of deposits	8.126
DD	Log of demand deposits	8.601
OD	Log of retail time and savings deposits	10.614
lab	Log of labor	5.163
cap	Log of capital	6.440
purf	Log of purchased funds	10.721
	Number of observations	14,640

Table 6. Estimation Results for Entire Banks

	Within	BC	CSS	KSS
CD	-0.0410 (0.0036)	-0.0342 (0.0031)	-0.0235 (0.0026)	-0.0025 (0.0015)
DD	-0.0859 (0.0109)	-0.0452 (0.0046)	-0.1161 (0.0096)	-0.0147 (0.0077)
OD	-0.1629 (0.0068)	-0.1662 (0.0063)	-0.1240 (0.0051)	-0.0326 (0.0142)
lab	-0.1643 (0.0116)	-0.1490 (0.0066)	-0.1547 (0.0100)	-0.0799 (0.0064)
cap	-0.0497 (0.0042)	-0.0532 (0.0031)	-0.0521 (0.0041)	-0.0321 (0.0040)
purf	-0.5706 (0.0147)	-0.6255 (0.0041)	-0.4877 (0.0129)	-0.5769 (0.0206)
ciln	0.1722 (0.0033)	0.1727 (0.0031)	0.1566 (0.0028)	0.1273 (0.0022)
inln	0.3239 (0.0042)	0.3142 (0.0039)	0.3224 (0.0041)	0.3341 (0.0036)
time	0.0140 (0.0006)	0.0043 (0.0008)	-	-
Avg TE	0.5026	0.6626	0.6570	0.5079

Table 7. Efficiencies of Banks

	Within	BC	CSS	KSS
Entire	0.5026	0.6626	0.6570	0.5079
Large	0.6050	0.7112	0.7156	0.6002
Small	0.4383	0.5921	0.6189	0.6032

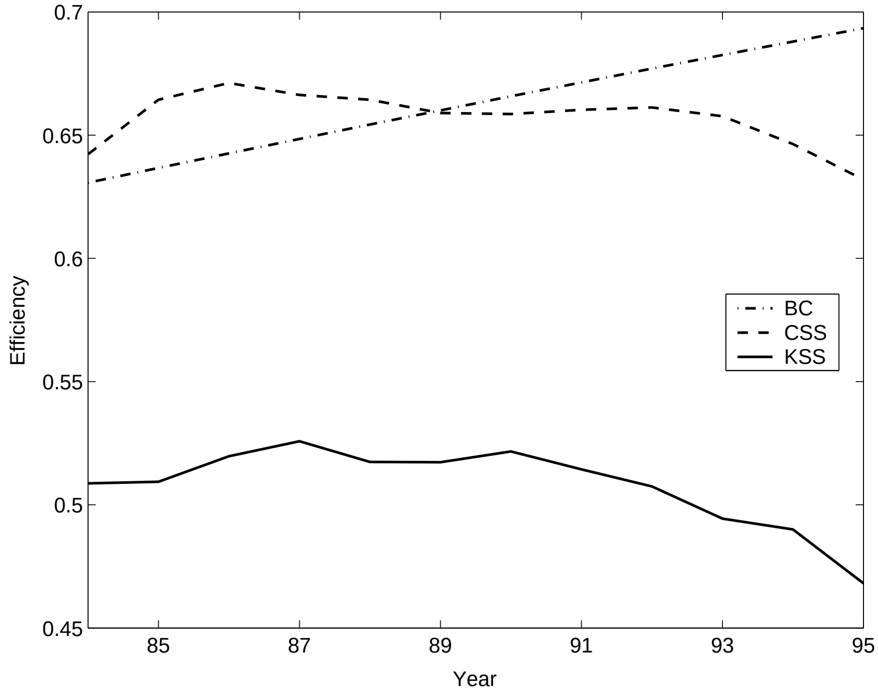

Figure 1. Efficiencies of Entire Banks

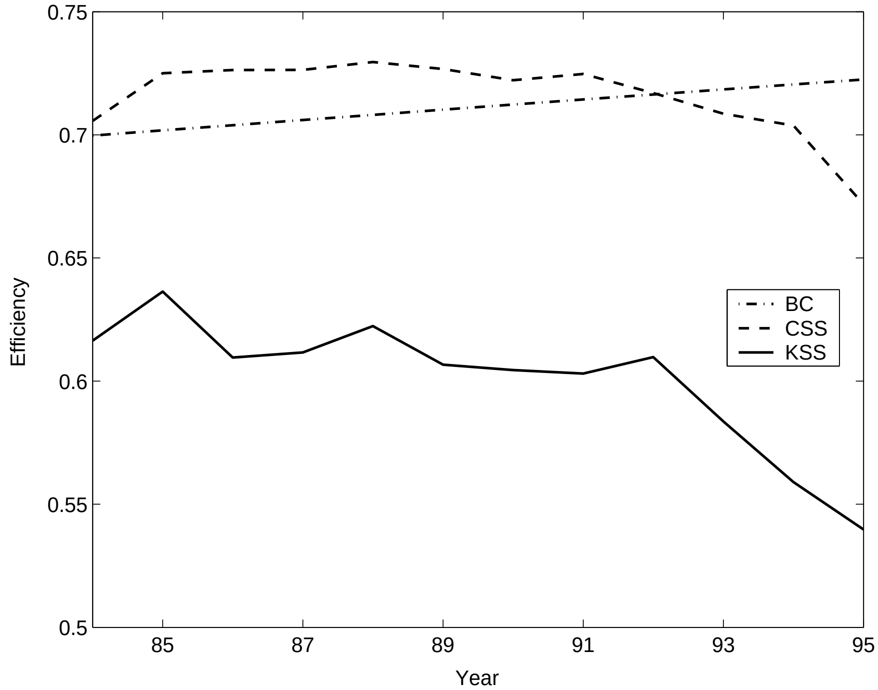

Figure 2. Efficiencies of Large Banks

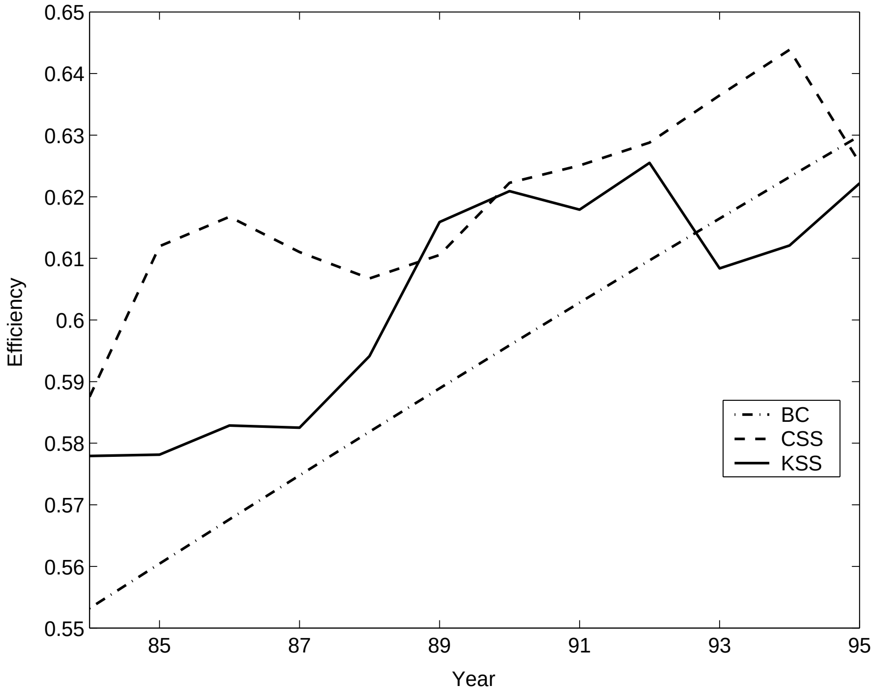

Figure 3. Efficiencies of Small Banks