A New Panel Data Treatment for Heterogeneity in Time Trends*

Alois Kneip
Department of Economics
University of Bonn
Bonn, Germany

Robin C. Sickles†
Department of Economics
Rice University
Houston, Texas

Wonho Song
Department of International Macroeconomics and Finance
Korea Institute for International Economic Policy
Seoul, Korea

December 2007

Abstract

Our paper introduces a new estimation method for arbitrary temporal heterogeneity in panel data models. The paper provides a semiparametric method for estimating general patterns of cross-sectional specific time trends. The methods proposed in the paper are related to principal component analysis and estimate the time-varying trend effects using a small number of common functions calculated from the data. An important application for the new estimator is in the estimation of time-varying technical efficiency considered in the stochastic frontier literature. Finite sample performance of the estimators is examined via Monte Carlo simulations. We apply our methods to the analysis of productivity trends in the U.S. banking industry.

*Earlier versions of this paper under the title "On Estimating the Mixed Effects Model" were presented at North American Productivity Workshop, Toronto, June 2004, COMPSTAT 2004, Prague, August 2004, and the European Workshop of Efficiency and Productivity IX, Brussels, July 2005. The paper was also given at the 2005 Econometric Society World Congress, University College, London and Michigan State, Rice University, and Syracuse University Econometrics Workshops. The authors thank participants at those conferences and workshops, particularly Peter Schmidt, Mahmoud El-Gamal, Yoosoon Chang, Joon Park, and Leopold Simar for constructive criticisms and insights. The usual caveat applies.

†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.

JEL Classification: C13, C14, C22, C23, D24, G21.
Key words and phrases: Time trends, panel models, principal component analysis, smoothing splines, banking efficiency.

1 Introduction

Substantial research interest has focused on controlling for unobserved heterogeneity in panel models. Recent work by Park and Simar and Park, Sickles, and Simar (1994, 1998, 2003, 2005) has focused on semi-parametric efficient panel data estimators for the standard fixed and random effects models with various specifications, including autoregressive errors and dynamic models. As the specifications of unobserved heterogeneity become more and more general, in particular allowing for temporal variation in the unobserved effects, and as trend stationarity of individual cross-sections comes under closer scrutiny, the proper specification of time effects becomes no less important than the specification of a difference or trend stationary time series (Nelson and Plosser, 1982; Maddala and Kim, 1998; Kao and Chiang, 2000; Baltagi, Egger, and Pfaffermayr, 2003; Mark and Sul, 2003, Chang, 2004).

In this paper, we extend the random and fixed effects model in such a way that we do not impose any explicit restrictions on the temporal pattern of individual effects. They are considered as random functions of time, representing a sample of smooth individual time trends. A detailed modelling and analysis of the general structure of these trends is the central point of our methodology. This goal is particularly important in our application to stochastic frontier analysis, where individual effects allow to access time-varying technical efficiencies of banks in the U. S. banking system.

The basic qualitative assumption is a fairly smooth, slowly varying local behavior of trends, although they may possess pronounced temporal patterns on the long-run. We formalize this idea and show that our model can be used for virtually any smooth pattern of temporal and cross-sectional changes in unobserved heterogeneity (time trends) and allows for the possibility that parameter heterogeneity is due to variables other than the constant term. This generality is accomplished by approximating the effect terms nonparametrically. The approach is based on a factor model, where time-varying individual effects are represented by linear combinations of a small number of unknown basis functions, with coefficients varying across cross-sectional units. Fixed effects, basis functions and corresponding coefficients are estimated from the data using methods related to principal component analysis coupled with smoothing spline techniques. Asymptotic distributions of the new estimators are derived, and rank tests are applied to determine the dimensionality of the factor model. Furthermore, goodness-of-fit tests of pre-specified parametric models are elaborated. Simulation experiments indicate that in finite samples our method works much better than other well known time-varying effects estimators. As an illustration, the effects are interpreted in the context of a stochastic frontier production function (Aigner, Lovell, and Schmidt, 1977) and our method is applied to the analysis of time-varying technical efficiency in the U.S. banking industry.

Factor models related to our setup have already been extensively studied in the econometric literature. Among others, important contributions are given by the work of Forni and Lippi (1997), Forni and Reichlin (1998), Stock and Watson (2002), Forni et al. (2000), Barnanke and Bovin (2000), or Bai and Ng (2002). Bai (2003, 2005) provides a general inferential theory. Ahn, Lee, and Schmidt (2005) give a generalization of Bai's methodology. Our approach is more general, fully integrating panel and factor models. It allows us to simultaneously estimate fixed effects, common factors (basis functions), and individual factor scores under a wide variety of conditions, including the possible existence of dynamic effects and/or correlations between individual effects and explanatory variables. Different from existing work the asymptotic theory also covers situations where dynamic effects follow non-stationary time series models, as for example random walks.

Another related branch of research is given by the statistical literature on "functional data analysis" which deals with the analysis of multiple smooth curves. For an overview one may consult the book by Ramsay and Silverman (1997). Although most of the work in this direction is descriptive, explicit factor models and corresponding inferential results based on "functional principal component analysis" are given, for example, by Kneip (1994), Ferré (1995), or Kneip and Utikal (2001) for different applications. An essential feature of our approach, taken from this literature, is the use of nonparametric smoothing techniques as an inherent part of the estimation procedure. The asymptotic theory of Section 2.2 indicates that econometric factor models in other contexts may also profit from incorporating smoothing procedures, since compared to standard results one may then achieve dramatically improved rates of convergence when estimating common factors.

The remainder of the paper is organized as follows. Section 2 introduces our new estimator for arbitrary time-varying effects, derives its asymptotic distribution, and provides other analytical results for optimal choice for the number of principal components and smoothing parameters. 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 the Appendix.

2 Model and main results

Panel studies in econometrics provide data from a sample of individual units where each unit is observed repeatedly over time (or age, etc.). Statistical analysis then usually aims to model the variation of some response variable Y. In addition to its dependence on some vector of explanatory variables $X \in \mathbb{R}^p$, the variability of Y between different individual units is of primary interest.

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}), $t = 1, \dots T$, $i = 1, \dots, n$, where the index i denotes individual units (e.g. firms, households, etc.) and the index t denotes time periods.

We consider the model

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

Although we consider non-constant individual effects, we will assume that $u_i(t)$ is varying "slowly" with t, and that $u_1, \dots, u_n$ therefore can be considered as a sample of *smooth* random functions. A precise discussion of the role of smoothness of u will be given in Subsection 2.2.

In our approach "individual effects´´ $u_i(t)$ necessarily play a more important role than in textbook panel data models, where they are sometimes considered as nuisance parameters. Identifiability of (1) requires that all variables X_{itj}, $j = 1, \dots, p$ possess a considerable variation over t. All individual differences are captured by $u_i(t)$, and this includes the effects of additional variables, like e.g. socioeconomic attributes, which characterize individuals but do not change over time. For example, suppose that there are q additional explanatory variables $X_{i,p+1}, \dots, X_{i,p+q}$ which do not change over time. The traditional framework then leads to the model

$$Y_{it} = \sum_{j=1}^{p} \beta_j X_{itj} + \sum_{j=p+1}^{p+q} \beta_j X_{ij} + \tau_i + \epsilon_{it} \tag{2}$$

with constant individual coefficients τ_i. In model (1), $u_i(t)$ then is a constant function with $u_i(t) \equiv \sum_{j=p+1}^{p+q} \beta_j X_{ij} + \tau_i$.

Based on (1), the coefficients β as well as the functions u_i can be estimated by semiparametric techniques. Indeed, in Subsection 2.1 this will be done by using partial spline estimation. However, a completely nonparametric analysis of the individual effects $u_i(t)$ possess a relatively poor degree of accuracy. Furthermore, economic interpretation and a further analysis of effects of socioeconomic characteristics is difficult.

In order to deal with (1) it thus makes sense to try to represent the functions u_i in a more convenient form which can be estimated more efficiently, is easier to interpret, and at the same time does not impose a severe restriction.

Our approach is motivated by ideas from (functional) principal component analysis leading to factor models studied in the statistical and econometric literature [see, e.g. Ramsay and Silverman, 1997, or Bai (2003)]. In our context we consider a version based on the vectors of functional values at the observed time points. Let $w(t) = \frac{1}{n} \sum_i u_i(t)$ denote the sample average function. It is then assumed that for some fixed $L \in \{0, 1, 2, \dots\}$ there exist some basis functions (common factors) $g_1, \dots, g_L$ such that

$$v_i(t) := u_i(t) - w(t) = \sum_{r=1}^{L} \theta_{ir} g_r(t). \tag{3}$$

Together with (1) this leads to the model

$$Y_{it} = \sum_{j=1}^{p} \beta_j X_{itj} + w(t) + \sum_{r=1}^{L} \theta_{ir} g_r(t) + \epsilon_{it}, \quad i = 1, \ldots, n, t = 1, \ldots, T \tag{4}$$

The dimension L as well as $g_1, \ldots, g_L$ and the coefficients (scores) θ_{ir} are unknown and have to be determined from the data. Obviously, different from traditional factor models as analyzed by Bai (2003), (4) additionally incorporates a fixed effect term. This is similar to the approach by Ahn et al. (2005). Note that by (3) only the linear factor space $\mathcal{L}_T$ spanned by $g_1, \ldots, g_L$ is identified but not the particular basis. We will thus additionally rely on the following normalizing conditions:

(a) $\frac{1}{n} \sum_i \theta_{i1}^2 \geq \frac{1}{n} \sum_i \theta_{i2}^2 \geq \cdots \geq \frac{1}{n} \sum_i \theta_{iL}^2 > 0$

(b) $\frac{1}{n} \sum_i \theta_{ir} \theta_{is} = 0$ 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 $r, s \in \{1, \ldots, L\}$, $r \neq s$.

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 $-\theta_{ir}, -g_r$). Note that (a) - (c) lead to orthogonal vectors g_r as well as empirically uncorrelated coefficients θ_{ir}. This ensures that all components can be interpreted separately, since they vary orthogonally to each other, a property which may be very helpful in practice when analyzing and interpreting these components.

It is important to consider (3) more closely. Obviously, g_r denote general functional components (common factors) whose structure provides general information about the common functional structure of the sample $\{v_i\} = \{u_i - w\}$. It will be shown in Section 3 that w and $g_1, \ldots, g_L$ can be estimated more efficiently than the *individual* random functions u_i.

Differences between individuals are captured by the coefficients θ_{ir}. For example, under (2) we have $L = 1$ and $\theta_{i1} = \sum_{j=p+1}^{p+q} \beta_j X_{ij} + \tau_i$. When having estimated θ_{i1}, estimates of $\beta_{p+1}, \ldots, \beta_{p+q}$ can then be obtained from a linear regression of θ_{i1} on $X_{i,p+1}, \ldots, X_{i,p+q}$. This generalizes to more interesting situations with $L \geq 1$ and non-constant functions $g_r(t)$. Effects of socioeconomic or demographic variables which do not change over time may be quantified by regressing the scores θ_{ir} on $X_{i,p+1}, \ldots, X_{i,p+q}$. In many applications such regressions will constitute an important step in econometric analysis. It will allow to access differences between important groups of individuals as well as the evolution of these differences over time as induced by the structure of $g_r(t)$.

When generalizing (2) with respect to possibly time varying effects this can be done either from the point of view of mixed effects models or from the point of view of time series analysis. Parametric mixed effects models are widely used in applications and assume that individual effects can be modelled by linear combinations of smooth, continuously differentiable basis function (e.g. polynomials). For example, in the context of production frontier analysis Cornwell, Schmidt, and Sickles (1990) assume that the u_i can be modelled by quadratic polynomials. In our notation, then $L = 3$ and g_1, g_2, g_3 correspond to a polynomial basis.

>From a time series point of view "smooth" trends are, however, often described by discrete time stochastic processes. In this context one may, for example, assume that $u_i(t) = \vartheta_i r_t$, where r_t is a random walk. Then, $L = 1$, $w(t) = \bar{\vartheta} r_t$, $g_1(t) = \frac{r_t}{\sqrt{T}}$ and $\theta_{1i} = \sqrt{T}(\vartheta_i - \bar{\vartheta})$. Note that different from mixed effect models, $v_i(t)$ is then only defined at the observation points $t = 1, 2, 3, \ldots$. Furthermore, in mixed effect models $\mathcal{L}_T = span\{g_1, \ldots, g_L\}$ is a fixed function space, while in the random walk example r_t and hence g_1 are random, and hence $\mathcal{L}_T = span\{g_1\}$ is a random subspace of $\mathbb{R}^T$.

Our approach will deal with both situations. Indeed, the general model (3) does not impose any strong restriction on the structure of the functions v_i. It is only assumed that for *some* L relation (3) holds for a "best" possible choice of basis function g_r which are not *a priori* known but are to be estimated from the data.

Our estimation procedure will be based on the fact that under the above normalization $g_1, g_2, \ldots$ are to be obtained as (functional) principal components of the sample
$v_1 = (v_1(1), \ldots, v_1(T))', \ldots, v_n = (v_n(1), \ldots, v_n(T))'$. More precisely, let

$$\Sigma_{n,T} = \frac{1}{n} \sum_i v_i v_i' \tag{5}$$

denote the empirical covariance matrix of $v_1, \ldots, v_n$ (recall that $\sum_i v_i = 0$). We use $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_T$ as well as $\gamma_1, \gamma_2, \ldots, \gamma_T$ to denote the resulting eigenvalues and orthonormal eigenvectors of $\Sigma_{n,T}$. Some simple algebra [compare, e.g., with Rao (1954)] then shows that

$$g_r(t) = \sqrt{T} \cdot \gamma_{rt} \quad \text{for all} \quad r = 1, \ldots, \; t = 1, \ldots, T, \tag{6}$$

$$\theta_{ir} = \frac{1}{T} \sum_t v_i(t) g_r(t) \quad \text{for all} \quad r = 1, 2, \ldots, \; i = 1, \ldots, n, \tag{7}$$

$$\lambda_r = \frac{T}{n} \sum_i \theta_{ir}^2 \quad \text{for all} \quad r = 1, 2, \ldots \tag{8}$$

Furthermore, for all $l = 1, 2, \ldots$

$$\sum_{r=l+1}^{T} \lambda_r = \sum_{i,t} (v_i(t) - \sum_{r=1}^{l} \theta_{ir} g_r(t))^2 = \min_{\tilde{g}_1, \ldots, \tilde{g}_l} \sum_i \min_{\vartheta_{i1}, \ldots, \vartheta_{il}} \sum_t (v_i(t) - \sum_{r=1}^{l} \vartheta_{ir} \tilde{g}_r(t))^2 \tag{9}$$

One can infer from relation (9) that $v_i \approx \sum_{r=1}^{l} \theta_{ir} g_r(t)$ provides the best possible approximation of the functions v_i in terms of an l-dimensional linear model. Model (3) holds for some dimension L if and only if $rank(\Sigma_{n,T}) = L$.

Obviously, $\Sigma_{n,T}$ and, hence, also the components g_r depend on the given values of n and T. A difference to usual factor models as considered by Bai (2003) or Ahn et al. (2005) consists in the fact that common factors are normalized with respect to sample instead of population characteristics. The latter may be achieved by replacing sample averages $\frac{1}{n}\sum_i \theta_{ir}^2, \frac{1}{n}\sum_i \theta_{ir}\theta_{is}$ by population means $\mathbf{E}(\theta_{ir}^2), \mathbf{E}(\theta_{ir}\theta_{is})$ in (a) and (b). However, this alternative normalization runs into problems in the random walk example. Furthermore, the real object of interest in model (3) is the factor space spanned by $g_1, \ldots, g_L$ and not the particular basis. As soon as it is possible to estimate very accurately one particular basis of the factor space, we in turn have a very precise description of this space. In this sense conditions (a) - (c) define a specific set of orthogonal basis functions which can be estimated with a particularly high degree of accuracy (see Subsection 2.2). Of course, suitable rotations of estimated common factors may be applied in subsequent analysis.

2.1 Estimation

In practice, $v_1, \ldots, v_n$ are unknown and all components of model (4) thus have to be estimated from the data. The idea of our estimation procedure 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 $\hat{\Sigma}_{n,T}$ 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 minimizing

$$\sum_i \frac{1}{T}\sum_t (Y_{it} - \bar{Y}_t - \sum_{j=1}^{p} \beta_j (X_{itj} - \bar{X}_{tj}) - v_i(t))^2 + \sum_i \kappa \frac{1}{T}\int_1^T (v_i^{(m)}(s))^2 ds \tag{10}$$

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 (10) 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 $2m$ (for a discussion of natural splines and definitions of possible basis functions see, for example, Eubank, 1988). 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 $\{z_j(t)\}_{j,t=1,\ldots,T}$ and $\{\int_1^T z_j^{(m)}(s) z_k^{(m)}(s) ds\}_{j,k=1,\ldots,T}$, 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

$$\sum_i \left(\|Y_i - \bar{Y} - (X_i - \bar{X})\beta - Z\zeta_i\|^2 + \kappa \zeta_i' A \zeta_i \right), \tag{11}$$

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

Note that Z is a regular $T \times T$ matrix. It is then easily seen that with

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

the solutions are given by

$$\hat{\beta} = \left(\sum_i (X_i - \bar{X})'(I - \mathcal{Z}_\kappa)(X_i - \bar{X}) \right)^{-1} \sum_i (X_i - \bar{X})'(I - \mathcal{Z}_\kappa)(Y_i - \bar{Y}) \tag{12}$$

as well as

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

Therefore,

$$\hat{v}_i = Z\hat{\zeta}_i = \mathcal{Z}_\kappa (Y_i - \bar{Y} - (X_i - \bar{X})\hat{\beta}) \tag{13}$$

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

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

Remarks: 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 (3). 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. A possible approach to directly estimate the best possible smoothing parameter for estimating common factors will be discussed at the end of Subsection 2.2.

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 (10). Estimates of $\hat{v}_i(t)$ at all $t = 1, \ldots, T$ are then obtained by spline interpolation.

Step 2: Estimate $w = (w(1), \dots, w(T))'$ 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 $\hat{\Sigma}_{n,T}$ of $\hat{v}_1 = (\hat{v}_1(1), \hat{v}_1(2), \dots, \hat{v}_1(T))', \dots, \hat{v}_n = (\hat{v}_n(1), \hat{v}_n(2), \dots, \hat{v}_n(T))'$ by

$$\hat{\Sigma}_{n,T} = \frac{1}{n}\sum_i \hat{v}_i \hat{v}_i'$$

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

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

$$\sum_t (Y_{it} - \bar{Y}_t - (X_i - \bar{X})\hat{\beta} - \sum_{r=1}^{L} \vartheta_{ri}\hat{g}_r(t))^2 \tag{14}$$

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

As discussed in the preceding section a further step of the analysis may consist in quantifying the influence of socioeconomic or demographic variables by regressing the scores $\hat{\theta}_{ri}$ on $X_{i,p+1}, \dots, X_{i,p+q}$.

2.2 Asymptotic Theory

We now consider properties of our estimators. We assume an i.i.d. sample of individual units and analyze the asymptotic behavior as $n, T \to \infty$. We do not impose any condition on the magnitude of the quotient T/n. The smoothing parameter $\kappa \equiv \kappa(n, T)$ may either remain fixed or may increase with n, T. Model (3) is assumed to possess a fixed dimension L for all n, T.

Before stating further assumptions, let us recall some basic facts of spline theory which provides a basis to understand the impact of these assumptions (see, for example, de Boor 1978, or Eubank 1988). Our analysis will be based the use of cubic smoothing splines ($m = 2$). Let $\tilde{v}_i(t)$ denote the corresponding natural spline interpolant of $v_i(1), \dots, v_i(T)$, i.e. $\tilde{v}_i$ is a natural spline function with knots at $1, \dots, T$ and $\tilde{v}_i(t) = v_i(t)$ for $t = 1, \dots, T$. By definition, the vector $(I - \mathcal{Z}_\kappa)v_i$ is obtained by evaluating the function v minimizing $\frac{1}{T}\sum_t (v_i(t) - v(t))^2 + \kappa \frac{1}{T}\int_1^T |v''(t)|^2 dt$ at $t = 1, \dots, T$. Consequently, $\frac{1}{T}\|(I - \mathcal{Z}_\kappa)v_i\|^2 \leq \kappa \frac{1}{T}\int_1^T |\tilde{v}_i''(t)|^2 dt$.

When analyzing properties of $\mathcal{Z}_\kappa$ it turns out that all eigenvalues are between 0 and 1, and for any fixed κ, $tr(\mathcal{Z}_\kappa^2) \leq T$ and $tr(I - \mathcal{Z}_\kappa) = O(T)$ as $T \to \infty$. Our setup is slightly different from usual spline theory which considers smoothing over a fixed (non-increasing) interval. But we have $z_j(t) = z_j^*(t/T)$, where $z_1, \ldots, z_T$ is the natural spline basis used to construct our estimator in Section 2.1, while $z_1^*, \ldots, z_T^*$ is a basis for all natural splines defined on $[0, 1]$ with knots $1/T, 2/T, \ldots, 1$. Obviously, $z_j'' = z_j^{*''}/T^2$. Defining the matrices Z^* and $A^* = \{\int_{1/T}^1 z_j^{*(m)}(s) z_k^{*(m)}(s) ds\}_{j,k=1,\ldots,T}$ similar to Z, A in Section 2.1, some straightforward arguments show that $\mathcal{Z}_\kappa = (I + \kappa(Z')^{-1} A Z^{-1})^{-1} = (I + \frac{\kappa}{T^4} T \cdot (Z^{*\prime})^{-1} A^* (Z^*)^{-1})^{-1}$. The structure of the eigenvalues of $T \cdot (Z^{*\prime})^{-1} A^* (Z^*)^{-1}$ is well-known (see, for example, Utreras, 1983) and can be used to show the existence of a constant $0 < q < \infty$ such that $tr(\mathcal{Z}_\kappa^2) \leq q \cdot \frac{T}{\kappa^{1/4}}$. 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

$$\frac{1}{T}\|(I - \mathcal{Z}_\kappa) v_i\|^2 \leq \kappa \frac{1}{T} \int_1^T |\tilde{v}_i''(t)|^2 dt, \; tr(\mathcal{Z}_\kappa^2) =\leq q \cdot \frac{T}{\kappa^{1/4}}, \quad \frac{1}{T} \sum_t \mathbf{Var}_\epsilon(\hat{v}_i(t)) = O_P(\frac{\sigma^2 tr(\mathcal{Z}_\kappa^2)}{T}) \tag{15}$$

where $\mathbf{Var}_\epsilon$ denotes conditional variance given v_i, X_{it}. Similar relations can, of course, be obtained with respect to w.

A central theme of Assumptions 1) - 5) below is a quantification of our requirement of “smooth” components v and w, while X_{it} is assumed to be considerably less smooth. This is translated into the assumption that v_i is well approximated by a cubic smoothing spline, or more precisely that the approximation bias denoted by $b_v(\kappa)$ can be made sufficiently small by a suitable choice of the smoothing parameter κ. The smoothest possible function is a constant as assumed in the standard panel model (2). Then $b_v(\kappa) = 0$ for all possible choices of κ. The bias will be small if $v_i(t)$ is “slowly” varying over t. No reasonable approximations are possible if the values $v_i(t)$ and $v_i(t+1)$ are essentially unrelated, as for example for independent white noise processes. In this case Assumption 2) below will not be fulfilled (note however that by (2) any white noise component of Y_{it} will be captured by the error term ϵ_{it} and will thus not appear in v_i, w).

In existing literature the mixed effects and the time series approach to panel data analysis appear to be largely incompatible, and very different methodologies are applied. Although “$n \to \infty$” will of course correspond to drawing more and more individuals at random, completely different asymptotic setups are used to describe the situation as "$T \to \infty$". Our admittedly unusual way of measuring smoothness via goodness-of-fit of spline approximations is motivated by an attempt to “unify” these approaches and to provide theoretical results which are able to cope with arbitrary "smooth" temporal pattern.

Nonparametric versions of mixed effects models, see for example Brumback and Rice (1998), suppose that for each individual the values $v_i(t)$, $t = 1, \ldots, T$, correspond to discretized measurements of an underlying smooth, at least twice differentiable random function. In this context, similar to nonparametric regression, a straightforward asymptotic setup consists in assuming that the distance between adjacent observational points tends to zero as $T \to \infty$. In other words, 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. With increasing T the discrete values $v_i(t)$, $t = 1, \ldots, T$, then provide more and more information about the true underlying functions $\nu_i(\cdot)$. Note that in this situation $\frac{1}{T}\sum_t v_i(t)^2$ will remain stochastically bounded and will not increase with T.

The mixed effects point of view is commonly adopted in applications, where t does not represent chronological time, but for example measurements at different ages of individuals. Furthermore, generalizations of (2) in stochastic frontier analysis are usually based on assuming smoothly varying functions representing individual inefficiencies.

The time series approach usually relies on very different methodological reasoning. Our assumptions then translate into the requirement that $v_i(t)$ represent "smooth" trends, where the degree of smoothness is measured by spline approximations. Recall that our procedure is not based on trend elimination by differencing, but tries to estimate the structure of individual trends. By (4) the $v_i(t)$ may contain important information about effects of explanatory variables. Time series asymptotics is based on adding more and more equidistant time periods as $T \to \infty$. Different from above $\frac{1}{T}\sum_t v_i(t)^2$ will then generally increase as $T \to \infty$. In order to cope with this setup we will use functions $c(T)$ and $d(T)$ to quantify resulting growth rates of $\frac{1}{T}\sum_t v_i(t)^2$ and $\frac{1}{T}\sum_t X_{it,j}^2$.

The following assumptions now provide the basis of our theoretical analysis. We will write $\lambda_{min}(A)$ and $\lambda_{max}(A)$ to denote the minimal and maximal eigenvalues of a symmetric matrix A, and g_r will be used to represent the vector $(g_r(1), \ldots, g_r(T))'$.

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 nondecreasing function $c(T)$ of T such that

   - $\mathbf{E}(\frac{1}{T}\sum_{t=1}^T v_i(t)^2) = O(c(T)), \quad \mathbf{E}(\frac{1}{T}\sum_{t=1}^T w(t)^2) = O(c(T))$,
   - $\frac{1}{n}\sum_i \theta_{ir}^2 = O_P(c(T)), \quad \frac{1}{n}\sum_i \theta_{ir}^4 = O_P(c(T)^2)$,
   - $c(T) = O_P(\frac{1}{n}\sum_i \theta_{ir}^2), \quad c(T) = O_P(|\frac{1}{n}\sum_i \theta_{ir}^2 - \frac{1}{n}\sum_i \theta_{is}^2|)$

   hold for all $r, s = 1, \ldots, L$, $r \neq s$, $j = 1, \ldots, p$, as $n, T \to \infty$.

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

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

satisfy

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

as $n, T \to \infty$. Furthermore, $tr(\mathcal{Z}_\kappa^2) \to \infty$ as $n, T \to \infty$.

4) $\frac{\mathbf{E}(\frac{1}{T}\sum_{t=1}^T \bar{X}_{tj}^2)}{\mathbf{E}(\frac{1}{T}\sum_{t=1}^T w(t)^2)} = O(1)$, and there exists a nondecreasing function $d(T) \leq c(T)$ of T with $d(T) = o(T)$ such that as $n, T \to \infty$ $\mathbf{E}(\frac{1}{T}\sum_{t=1}^T X_{it,j}^2) = O(d(T))$ holds for all $j = 1, \ldots, p$ as $n, T \to \infty$. Furthermore,

$$\lambda_{max}\left([\sum_i (X_i - \bar{X})'(I - \mathcal{Z}_\kappa)(X_i - \bar{X})]^{-1}\right) = O_p(\frac{1}{nT}) \tag{16}$$

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

$$a'(I - \mathcal{Z}_\kappa) \cdot \mathbf{E}\left((X_{ij} - \bar{X}_j)(X_{ij} - \bar{X}_j)'\right)(I - \mathcal{Z}_\kappa)a \leq D \cdot \|(I - \mathcal{Z}_\kappa)a\|^2. \tag{17}$$

holds for all sufficiently large n, T.

5) 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) -5). It is therefore important to understand these assumptions correctly in view of the different asymptotic setups discussed above. First note that by requiring that $\frac{1}{n}\sum_i \theta_{ir}^2 = O_P(c(T))$ as well as $c(T) = O_P(\frac{1}{n}\sum_i \theta_{ir}^2)$ we assume that $\frac{1}{n}\sum_i \theta_{ir}^2$ increases *exactly* with rate $c(T)$.

Let us now analyze the situation where $v_i(t)$, $t = 1, \ldots, T$, are assumed to be discretized values of at least twice differentiable random functions, and where the local asymptotics of nonparametric mixed effects models is considered. Then $c(T) = d(T) = 1$, and Assumptions 2) - 3) can be made more explicit by posing the following condition on the structure of underlying functions:

Situation 1. For each individual there are data from T equidistant observations in a *fixed* time interval. There exists a smooth function μ as well as i.i.d. nonzero random functions $\nu_1, \ldots, \nu_n$ on $L^2[0,1]$ such that $\mu(\frac{t}{T}) = w(t)$ and $\nu_i(\frac{t}{T}) = v_i(t)$

for $t = 1, \dots, T$. The functions μ as well as $\nu_1, \dots, \nu_n$ are a.s. *twice continuously differentiable* with $\mathbf{E}(\int_0^1 \nu_i''(t)^2 dt) < \infty$ and $0 < \mathbf{E}(\int_0^1 \nu_i(t)^2 dt) < \infty$.

Then Assumption 2) is fulfilled with $c(T) = 1$. Moreover, $v_i''(t) = \frac{1}{T^2}\nu_i''(t)$, and $\kappa \frac{1}{T} \int_1^T |\tilde{v}_i''(t)|^2 dt \leq \kappa \frac{1}{T} \int_1^T |v_i''(t)|^2 dt = \kappa \frac{1}{T^4} \int_0^1 |\nu_i''(t)|^2 dt$.

>From (15) we can then infer that

$$b_v(\kappa)^2 = O(\kappa \frac{1}{T^4}), \quad tr(\mathcal{Z}_\kappa^2) = O(\frac{T}{\kappa^{1/4}}) \tag{18}$$

The bias thus depends on the rate of decrease of $\kappa \equiv \kappa_{n,T}$ as $n, T \to \infty$. An optimal smoothing parameter for estimating v_i then satisfies $\kappa \sim T^{-4/5} \cdot T^4$, which means that $b_v(\kappa)^2 = O(T^{-4/5})$ as $T \to \infty$ in Assumption 3). Similar results are to be obtained with respect to w and $b_w(\kappa^*)$. It will be seen from the results of Theorem 1, that undersmoothing, i.e. choosing a smaller smoothing parameter than the individually optimal one, leads to still better rates of convergence for our estimates of g_r. Also note that in order to satisfy Assumption 4) we implicitly assume that X_{itj} are generated by *non-smooth* stochastic processes. This is a natural condition, since due to the error terms ϵ_{it} also the time path of our dependent variable Y_{it} is *non-smooth* and cannot be well approximated by splines.

In the mixed effect approach formalized in Situation 1 our Assumption 1) corresponds to assuming that all v_i lie in a fixed L-dimensional, non-random space $\mathcal{L}_T$. From a time series point of view stochastic trends are, however, often described by discrete time stochastic processes. An approach relying on the existence of underlying smooth functions $\nu_i(\cdot)$ is not feasible in this context. The components $g_r(t)$ as well as the structure of $\mathcal{L}_T$ may then depend on particular realizations of such processes and may thus be random. Our approach of measuring smoothness by spline approximations can also be applied in such a situation: Construction of spline smoothers implies that the value of the integral $\frac{1}{T} \int_1^T |\tilde{v}_i''(t)|^2 dt$ in (15) 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, if $\frac{1}{T} \sum_t (v_i(t+1) - 2v_i(t) + v_i(t-1))^2$ is reasonably *small*, then a fairly large smoothing parameter κ will still result in a small bias.

An example is provided by the process $u_i(t) = w(t) + v_i(t) = \vartheta_i r_t$, where $r(t) = \sqrt{|r_0 + \delta_1 + \delta_2 + \cdots + \delta_t|}$ for some fixed r_0 and i.i.d. random variables $\delta_1, \delta_2, \dots$ with $\mathbf{E}(\delta_t) = \mu$, $var(\delta_t) = \sigma_\delta^2$. Since for large T we have $r_t - r_{t-1} \approx \frac{\mu + \delta_t}{2r_{t-1}}$, such a process may possibly be assumed if the innovations in certain period depend on the level of the process reached in the previous period. The stochastic trend induced by this process cannot be eliminated by differencing, since for any $q = 1, 2, \dots$ the q-th order differences of r_t are *not* stationary. On the other hand, the resulting $v_i(t)$ are reasonably smooth. It is easily checked that then Assumptions 1) and 2) hold with $L = 1$ and $c(T) = T^{1/2}$. Furthermore $\frac{1}{T} \int_1^T |\tilde{v}_i''(t)|^2 dt = O_P(T^{-1/2})$, and hence $b_v(\kappa) = O_P(\kappa T^{-1/2})$.

$w(t)$, $v_i(t)$ may instead be generated by more traditional $I(1)$ processes. Reasonable convergence results can then 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 $\frac{1}{T}\sum_t v_i(t)^2$. To give some explicit results in this context we will concentrate on the simple case of a random walk:

Situation 2.: Assume that for some fixed $r_1 \in \mathbb{R}$

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

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

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

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 $\tilde{r}(t)$ of r_t we obtain $\mathbf{E}(\frac{1}{T}\int_1^T |\tilde{r}''(t)|^2 dt) = O(1)$ 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

$$b_v(\kappa) = \mathbf{E}(\frac{1}{T}\|(I - \mathcal{Z}_\kappa)v_i\|) = O(1), \quad tr(\mathcal{Z}_\kappa^2)/T = O(1). \tag{19}$$

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:

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

Although, as shown above, our approach is able to cope with trends which do not fit into the usual $I(q)$ framework, some of our assumptions are restrictive from a time series point of view. Apart from assuming i.i.d. errors in 5), Assumption 4) contains regularity conditions which impose restrictions on the design matrix. It is essentially required 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 paths. Note, however, that some interesting cases, as for example cointegration between Y and X, are excluded. We believe that more general results can be obtained, but part of the methodology may have to be adapted to the specific situation.

When considering the simplest case $p = 1$, Assumption 4) 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 (16) as well as (17) follow from the well-known structure of such autocovariance matrices.

Assumption 4) 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 construction of $\mathcal{Z}_\kappa$, $\mathcal{Z}_\kappa \mathbf{1} = \mathbf{1}$ for $\mathbf{1} = (1,1,\ldots,1)'$ we arrive at

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

The maximal eigenvalue of Γ remains bounded as $T\to\infty$, and relation (17) is an immediate consequence of the structure of $\mathcal{Z}_\kappa$. Moreover, it is easily checked that $\lambda_{min}(\mathbf{E}[(X_i-\bar{X})'(I-\mathcal{Z}_\kappa)(X_i-\bar{X})])$ is proportional to T. Since $\frac{1}{n}\sum_i (X_i-\bar{X})'(I-\mathcal{Z}_\kappa)(X_i-\bar{X}) \to_P \mathbf{E}[(X_i-\bar{X})'(I-\mathcal{Z}_\kappa)(X_i-\bar{X})]$, relation (16) follows from the continuity of $\lambda_{min}(A)$ as a function of A.

Before stating our main theorem we have to introduce some additional notation. We will say that v_i and X_i are "uncorrelated up to linear components" (ulc-uncorrelated) if there exist *linear* functions $z_{v,i,T}(t)$ and $z_{x,i,j,T}(t)$, possibly depending on $i=1,\ldots,n$, $j=1,\ldots,p$, or T, so that $\mathbf{E}(v_i^* v_l^*|X^*) = \mathbf{E}(v_i^* v_l^*)$ holds for all $i,l\in\{1,\ldots,n\}$, where $v_i^*(t) = v_i(t) - z_{v,i,T}(t)$, $X_{itj}^* = X_{itj} - z_{x,i,j,T}(t)$, and $X^* = (X_{itj}^*)_{i,t,j}$. In the standard model (2) v_i and X_i are necessarily ulc-uncorrelated, since $v_i^*(t) = v_i(t) - z_{v,i,T}(t) = 0$ for the constant function $z_{v,i,T}(t) \equiv \sum_{j=p+1}^{p+q}\beta_j(X_{ij}-\bar{X}_j)+\tau_i-\bar{\tau}$, and hence $v_i^*\equiv 0$ does not depend at all on X.

We will use "$\mathbf{E}_\epsilon$" to denote conditional expectation given v_i and X_i, $i=1,\ldots n$. Moreover, $\tilde{X}_i = X_i - \bar{X}$. Additionally note that eigenvectors are only unique up to sign changes. In the following we will always assume that the right "versions" are used. This will go without saying.

Theorem 1. Under Assumptions 1) - 5) we obtain as $n,T\to\infty$

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

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

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

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

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

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

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

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

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

Furthermore, if $\frac{b_v(\kappa)^2}{c(T)^{1/2}} + d(T)^{1/2}(b_\beta(\kappa) + \frac{1}{\sqrt{nT}}) + \frac{1}{T^2 c(T)^{3/2}} = o(T^{-1/2})$, then

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

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

$$\frac{n\sum_{r=L+1}^{T}\hat{\lambda}_r \;\; - (n-1)\sigma^2 \cdot tr(\mathcal{Z}_\kappa \hat{\mathcal{P}}_L \mathcal{Z}_\kappa)}{\sigma^2\sqrt{2n \cdot tr((\mathcal{Z}_\kappa \hat{\mathcal{P}}_L \mathcal{Z}_\kappa)^2)}} \rightarrow_d \mathbf{N}(0,1), \qquad (20)$$

$$\frac{n \cdot tr(\mathcal{P}_L \hat{\Sigma}_{n,T}) - (n-1)\sigma^2 \cdot tr(\mathcal{Z}_\kappa \mathcal{P}_L \mathcal{Z}_\kappa)}{\sigma^2\sqrt{2n \cdot tr((\mathcal{Z}_\kappa \mathcal{P}_L \mathcal{Z}_\kappa)^2)}} \rightarrow_d \mathbf{N}(0,1), \qquad (21)$$

where $\hat{\mathcal{P}}_L = I - \sum_{r=1}^{L}\hat{\gamma}_r\hat{\gamma}_r'$, and $\mathcal{P}_L$ is the projection matrix projecting into the $n-L$ dimensional linear space orthogonal to $span\{\mathcal{Z}_\kappa g_1, \dots, \mathcal{Z}_\kappa g_L\}$.

A proof of the theorem can be found in the appendix. Let us interpret the results on estimating β and g_r in terms of the specific additional assumptions made in the two situations discussed above.

Situation 1. Recall that $c(T) = d(T) = 1$ and that by (19) we obtain $b_v(\kappa)^2 = O(\kappa\frac{1}{T^4})$. As noted above the optimal smoothing parameter to obtain best possible estimates of the *individual* functions $v_i(t)$ is of order $\frac{\kappa}{T^4} = \kappa_T \sim T^{-4/5}$. However, different from individual estimates of v_i variance of the estimated functional components $\hat{g}_r$ decrease as n increases. Best rates of convergence with respect to g_r can thus be obtained by undersmoothing. If $n = o(T^4)$ and $T = o(n^4)$, then $n^{-4/5}\kappa_T$ may be used instead of κ_T. This yields $b_v(\kappa) = O_P((nT)^{-2/5})$, $tr(\mathcal{Z}_\kappa^2) = O(nT)^{1/5}))$, as well as

$$T^{-1/2}\|g_r - \hat{g}_r\| = O_P((nT)^{-2/5}).$$

Also note that in this situation $(nT)^{-2/5} = o(T^{-1/2})$, $(nT)^{-2/5} = o(n^{-1/2})$, and the additional requirements ensuring the distributional results in Theorem 1c) - 1e) are necessarily fulfilled. Moreover, $b_\beta(\kappa) = o(1/\sqrt{nT})$, and Theorem 1a) simplifies to

$$V_{n,T}^{-1/2}(\hat{\beta} - \beta) \sim \mathbf{N}(0, I).$$

One might compare these results with the general theory of existing econometric factor models as derived by Bai (2003). If T is not too small compared to n, Bai's results imply that in his context the rate of convergence of estimated factors is $n^{-1/2}$ instead of $(nT)^{-2/5}$ as obtain for our method. One must, however, be careful when interpreting this difference. Our results crucially depend on the data-dependent normalization of $g_1, g_2, \ldots$ given by (a) - (c) above, while in standard factor models normalization usually refers to population characteristics. If for example, the sample means in (a) - (c) were replaced by their population analogues, then even in our context only a rate of convergence $n^{-1/2}$ of $\hat{g}_r$ to this "re-normalized" factors could be achieved, since at best $\frac{1}{n}\sum_i \theta_{ir}^2$ is only a $\sqrt{n}$-consistent estimator of $\mathbf{E}(\theta_{ir}^2)$ (in Situation 1 this will usually be the case). But recall that factor spaces are identical, and in order to characterize this space as precisely as possible, one should definitely look for the "best estimable" orthogonal basis. Therefore, a crucial point is that standard factor approaches (not applying smoothing techniques) *will always lead to* $T^{-1/2}\|g_r - \hat{g}_r\| = O_P(n^{-1/2})$, *even if* g_r *is defined according to our particular normalization (a) - (c).* Smoothing here dramatically improves upon the rate of convergence.

Situation 2. Consider the case of a random walk as discussed above. Note that this situation does not fit into the framework of traditional econometric factor models. Additionally assume that as for $ARMA(p,q)$-processes X_{it} satisfies Assumption 4 with $d(T) = 1$. Then, $c(T) = T$ and a constant, non-increasing smoothing parameter κ provides best possible estimates of individual functions. Then $\frac{b_v(\kappa)}{c(T)^{1/2}} = O(T^{-1/2})$, and consequently

$$T^{-1/2}\|g_r - \hat{g}_r\| = O_P(T^{-1/2}).$$

The additional requirements ensuring the distributional results in Theorem 1c) - 1e) hold if v_i and X_i are ulc-uncorrelated. The bias in estimating β is of order $b_\beta(\kappa) = O(1/\sqrt{nT})$ if if X_i and v_i are ulc-uncorrelated, and $b_\beta(\kappa) = O(1/\sqrt{T})$ else. It will thus not be negligible compared to the standard error.

In order to avoid further complications in the presentation of results, the effect of undersmoothing is not covered by the theorem. Formally, in the case of a random walk undersmoothing will mean to use a sequence of smoothing parameters with $\kappa \to 0$ as $n, T \to \infty$, which is not compatible with Assumption 2. For example, let $\kappa \sim n^{-\tau}$ for some $\tau > 0$ with $T^{1/2}n^{-\tau} \to \infty$. Then $b_v(\kappa) = O(n^{-\tau})$. It follows from the results of Utreras (1983) that we still have $tr(\mathcal{Z}_\kappa^2) = O(T)$, but $tr((I - \mathcal{Z}_\kappa)) = O(\kappa T)$. For simplicity assume that β is estimated with respect to a constant, non-increasing smoothing parameter κ, and that undersmoothing is only applied in Step 3 of our estimation procedure by analyzing re-estimated functions $\hat{v}_i = \mathcal{Z}_\kappa(Y_i - \bar{Y} - (X_i - \bar{X})\hat{\beta})$ with $\kappa \sim n^{-\tau}$. The arguments used to derived Theorem 1(c) then readily generalize, and it may then be shown that $T^{-1/2}\|g_r - \hat{g}_r\| = O_P(\frac{n^{-\tau}}{T^{1/2}} + \frac{d(T)^{1/2}}{T\sqrt{n}} + \sqrt{\frac{1}{nT}})$ if v_i and X_i are ulc-uncorrelated, and $T^{-1/2}\|g_r - \hat{g}_r\| = O_P(\frac{n^{-\tau}}{T^{1/2}} + \frac{d(T)^{1/2}}{T} + \sqrt{\frac{1}{nT}})$, else. In both cases the rate of convergence is $o(T^{-1/2})$, which shows that undersmoothing may be beneficial even in this situation.

Remarks:

a) We have seen that, depending on the situation, the bias of our estimator $\hat{\beta}$ may not be negligible. Fortunately there seem to exist some ways to reduce bias. As can be seen from the proof of the theorem one obtains $\mathbf{E}_\epsilon(\hat{\beta}) - \beta = (\sum_i \tilde{X}_i'(I - \mathcal{Z}_\kappa)\tilde{X}_i)^{-1} \sum_i \tilde{X}_i'(I - \mathcal{Z}_\kappa)v_i$. By the results of the theorem this bias term may be estimated when replacing v_i by $\hat{v}_i$. Another approach to bias reduction may consist in iterating our estimation procedure. In addition to estimating θ_{ri}, (14) might also be used to obtain updated least squares estimates $\hat{\beta}^{(1)}$. These new estimates of β might in turn be plugged into Step 2 and 4 of our algorithm to determine new approximations $\hat{g}_r^{(1)}$, etc. A precise analysis is, however, not in the scope of the present paper.

b) The question arises whether it is possible to determine the best smoothing parameter for estimating $g_1, g_2, \dots$ directly from the data. A straightforward approach consists in a "leave-one-individual-out" cross-validation. For a fixed L and $i = 1, \dots, n$ let $\hat{\beta}_{-i}$ and $\hat{g}_{r,-i}$ denote the respective estimates of β and g_r obtained from the data (Y_{kj}, X_{kj}), $k = 1, \dots, i-1, i+1, \dots, n$, $j = 1, \dots, T$, and let $\hat{\theta}_{r,-i}$ denote the corresponding estimates of θ_{ri} to be obtained when using $\hat{\beta}_{-i}$, $\hat{g}_{r,-i}$ instead of $\hat{\beta}$, $\hat{g}_r$ in Step 4 of our estimation procedure. All these estimates depend on κ, and one may approximate an optimal smoothing parameter by minimizing

$$CV(\kappa) := \frac{1}{nT} \sum_i \sum_t (Y_{it} - \bar{Y}_t - (X_{it} - \bar{X}_t)\hat{\beta}_{-i} - \sum_{r=1}^{L} \hat{\theta}_{r,-i}\hat{g}_{r,-i}(t))^2$$

over κ. Note that by (4) and by the independence of $\hat{\beta}_{-i}, \hat{g}_{r,-i}$ from ϵ_{it}

$$\mathbf{E}_\epsilon[CV(\kappa)] = \frac{1}{nT} \sum_i \sum_t ((X_{it} - \bar{X}_t)\hat{\beta} + \sum_{r=1}^{L} \theta_{ri} g_r(t) - (X_{it} - \bar{X}_t)\hat{\beta}_{-i} - \sum_{r=1}^{L} \hat{\theta}_{r,-i}\hat{g}_{r,-i}(t))^2 + \frac{(T-L)}{T}\sigma^2$$

holds for all κ. It therefore seems to be reasonable to expect that this approach "in tendency" selects a κ providing a small mean squared error between true an estimated model. A precise theoretical analysis is not in the scope of the present paper.

2.3 Dimensionality and model tests

Result (20) of 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

$$\hat{\sigma}^2 := \frac{1}{(n-1) \cdot tr(I - \mathcal{Z}_\kappa)^2} \sum_i \|(I - \mathcal{Z}_\kappa)(Y_i - \bar{Y} - (X_i - \bar{X})\hat{\beta})\|^2. \qquad (22)$$

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

$$\Delta(l) := \frac{n \sum_{r=l+1}^{T} \hat{\lambda}_r \; - (n-1)\hat{\sigma}^2 \cdot tr(\mathcal{Z}_\kappa \hat{\mathcal{P}}_l \mathcal{Z}_\kappa)}{\hat{\sigma}^2 \sqrt{2n \cdot tr((\mathcal{Z}_\kappa \hat{\mathcal{P}}_l \mathcal{Z}_\kappa)^2)}}. \tag{23}$$

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

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

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

The following theorem provides a theoretical justification of this procedure. A proof is given in the appendix.

Theorem 2. In addition to the assumptions of Theorem 1 assume that $tr(\mathcal{Z}_\kappa^2)/n \to 0$ as well as $Td(T)b_\beta(\kappa)^2 + \frac{d(T)}{n} + \frac{1}{Tc(T)} = o\left(\sqrt{tr(\mathcal{Z}_\kappa^2)/n}\right)$. Then,

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

Relation (21) may serve to test the validity of a pre-specified parametric model of the form $v_i(t) = \sum_{j=1}^{L} \vartheta_{ri}\psi_r(t)$ for some known basis functions ψ_r. If $\mathcal{P}_{\psi,L}$ denotes the projection matrix projecting into the $n-L$ dimensional linear space orthogonal to $span\{\mathcal{Z}_\kappa\psi_1, \dots, \mathcal{Z}_\kappa\psi_L\}$, then the null hypothesis: $H_0 : v_i(t) = \sum_{j=1}^{L} \vartheta_{ri}\psi_r(t)$ is rejected if

$$\frac{n \cdot tr(\mathcal{P}_{\psi,L}\hat{\Sigma}_{n,T}) \; - (n-1)\hat{\sigma}^2 \cdot tr(\mathcal{Z}_\kappa \mathcal{P}_{\psi,L} \mathcal{Z}_\kappa)}{\hat{\sigma}^2 \sqrt{2n \cdot tr((\mathcal{Z}_\kappa \mathcal{P}_{\psi,L} \mathcal{Z}_\kappa)^2)}} > z_{1-\alpha}$$

Obviously, under H_0 we have $\mathcal{P}_{\psi,L} = \mathcal{P}_L$, and by (21) the test possesses an asymptotically correct size. But the derivation of (21) is based on the fact that $tr(\mathcal{P}_L\Sigma_{n,T}) = 0$ and hence $tr(\mathcal{P}_L\hat{\Sigma}_{n,T}) = tr(\mathcal{P}_L(\hat{\Sigma}_{n,T} - \Sigma_{n,T}))$. If H_0 is false, then generally $tr(\mathcal{P}_{\psi,L}\Sigma_{n,T}) = O_P(Tc(T))$, and therefore $tr(\mathcal{P}_{\psi,L}\hat{\Sigma}_{n,T}) = tr(\mathcal{P}_{\psi,L}\Sigma_{n,T}) + tr(\mathcal{P}_{\psi,L}(\hat{\Sigma}_{n,T} - \Sigma_{n,T}))$ will in tendency be too large.

This test can of course be particularly applied to verify the validity of a standard panel model $Y_{it} = \beta_0 + \sum_{j=1}^{p} \beta_j X_{itj} + \theta_{1i} + \epsilon_{it}$ with constant individual effects. Then $L = 1$ and $\mathcal{P}_{\psi,L} = I - \frac{1}{T}\mathbf{1}\mathbf{1}'$ with $\mathbf{1} = (1, \dots, 1)'$. Also note that in this case $c(T) = 1$ as well as $b_v(\kappa) = b_w(\kappa^*) = 0$ for all possible choices of κ, κ^*.

3 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 experiments. A competing time-varying individual effects estimator is based on the

Cornwell, Schmidt, and Sickles fixed effects estimator (CSSW, 1990). The CSSW estimator allows for an arbitrary polynomial in time (usually truncated at powers larger than two) with different parameters for each firm. We also consider the classical time-invariant fixed and the random effects estimators (Baltagi, 2005). These estimators have been used extensively in the productivity literature which interprets time varying firm effects (time trends) as technical efficiencies.

We consider the panel data model (1):

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

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 (2003, 2005):

$$X_{it} = RX_{i,t-1} + \eta_{it}, \text{ where } \eta_{it} \sim \mathbf{N}(0, I_2), \tag{24}$$

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 (24) for $t \geq 2$. Then, the obtained values of X_{it} are shifted around three different means to obtain three 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. In all of our data generating processes (DGP's) we set the mean function $w(t) = 0$. Thus in equation (3) above $u_i(t) = v_i(t)$ and the model considered in our experiments becomes:

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

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

$$\begin{aligned} \text{DGP1} &: \; v_i(t) = \theta_{i0} + \theta_{i1}\frac{t}{T} + \theta_{i2}\left(\frac{t}{T}\right)^2 \\ \text{DGP2} &: \; v_i(t) = \phi_i r_t \\ \text{DGP3} &: \; v_i(t) = \upsilon_{i1} g_{1t} + \upsilon_{i2} g_{2t} \\ \text{DGP4} &: \; v_i(t) = \xi_i \end{aligned}$$

where θ_{ij} $(j = 0, 1, 2) \sim i.i.d.\mathbf{N}(0, 0.5^2)$, $r_{t+1} = r_t + \delta_t$, $\phi_i, \delta_t, \upsilon_{ij}(j = 1, 2) \sim i.i.d.\mathbf{N}(0, 1)$, $g_{1t} = \sin(\pi t/4)$ and $g_{2t} = \cos(\pi t/4)$. Even though there is no correlation between the effects and regressors in DGP1, the fixed effects treatment (CSSW) is used in the experiments. DGP2 is the random walk process. DGP3 is considered to model effects with large temporal variations. DGP4 is the usual constant effects model with symmetric effects. Thus, we may consider DGP3 and DGP4 as two extreme cases among the possible functional forms of time varying individual effects.

The CSSW (within) fixed effects estimator is

$$\beta_{CSSW} = (X'M_QX)^{-1}X'M_Qy$$

where $M_Q = I - Q(Q'Q)^{-1}Q'$, $Q = \text{diag}(W_i)$, $i = 1, \ldots, n$, and $W_{it} = [1, t, t^2]$. A second-order time polynomial is used to approximate $v_i(t)$ based on the CSSW (within) residuals.

For the KSS estimator, cubic smoothing splines were used to approximate $v_i(t)$ in step 1, and the smoothing parameter κ was selected by using 'leave-one-individual-out' cross-validation.[1] The coefficient parameter β is updated using $\hat{g}_r(t)$ obtained in step 4 of (14), which is found to generate substantial efficiency gains. However, the updated estimates $\hat{\beta}^{(1)}$ are not plugged into step 2 again because there is no efficiency gain observed for $\hat{g}_r(t)$. Most simulation experiments were repeated 1,000 times except the cases for $n = 300$ for which 500 replications were carried out. To measure the performances of the various estimators of the effects, we used normalized mean squared error (MSE):

$$R(\widehat{v}, v) = \frac{\sum_{i,t} \left(\widehat{v}_i(t) - v_i(t)\right)^2}{\sum_{i,t} v_i^2(t)}.$$

We now present the simulation results. Tables 1-4 present mean squared errors (MSE) of coefficients[2] and effects for each DGP. Also, average optimal dimensions, L, chosen by $\Delta(l)$ criterion are reported in the last column of second panel in each table. We note that the optimal dimension, L, is correctly chosen for the KSS estimator in all DGPs. Thus, we can verify the validity of the dimension test $\Delta(l)$ discussed in Section 2.

For DGP1, the performances of the KSS estimator are better than those of 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 inefficient CSSW estimator (effects and regressors are not correlated) is no better than the other estimators. The performances of Within and GLS estimators generally get worse as T increases.

[1] We let $\kappa = (1 - p)/p$ and choose p among a selected grid of 9 equally spaced values between 0.1 and 0.9.

[2] The MSE of coefficients are scaled by 10^3.

DGP2 is considered to see the performance of the estimators for arbitrary functional form of individual effects. Hence, estimators based on relatively simple function of time such as used in the CSS estimator is 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 the results confirm such belief. KSS estimator dominantly outperforms the other estimators by any standards in the order of three to ten times. It is particularly conspicuous in terms of MSE of effects and efficiencies. CSSW performs reasonably well for effects, but it is no better than the others for other criteria.

DGP3 generates effects with large temporal variations. As T increases, the variations become large. The other estimators assume pre-specified and simple functional forms, thus they are expected to perform less satisfactorily for this DGP. On the contrary, the KSS estimator allows arbitrary functional forms as well as multiple individual effects. Hence, it is expected to perform well even under this DGP. Indeed, the results show that the KSS estimator performs very well, especially for large T, with correct number of L chosen. On the other hand, the other estimators suffer from severe distortions in the estimates of the effects, although coefficient estimates look reasonably good.

DGP4 represents the reverse situation so that there is no temporal variation in the effects. Thus, 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 performance is fairly good 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 sum, simulations show that the KSS estimator is safely applicable regardless of the assumption on the temporal patterns of effects, and may therefore be preferred to other existing estimators in these types of empirical settings, among potentially many others. On the other hand, either if constant effects are assumed when true effects are time-variant, or if the temporal patterns of effects are misspecified, parameter estimates as well as effect estimates become severely biased. In these cases, large T increases the bias, and large n does not help solve the problem.

4 Efficiency Analysis of Banking Industry

4.1 Empirical Model

We next compare the various estimators in an empirical illustration of efficiency changes in the US banking industry after a series of deregulatory initiatives in the early 1980's. We model the multiple output/multiple input banking technology using the output distance function (Adams, Berger, and Sickles, 1996). 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 aggregated 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 $u_i^*(t)$ for the firm specific level of radial technical inefficiency, with variations in time allowed. The Cobb-Douglas stochastic distance frontier is thus

$$0 = \sum_j \gamma_j \ln y_{j,it} - \sum \beta_k \ln x_{k,it} + u_i^*(t) + \epsilon_{it}.$$

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

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

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 $u_i(t) = -u_i^*(t)$, in which case we can write the stochastic distance frontier as

$$Y_{it} = Y_{it}^{*\prime}\gamma + X_{it}'\beta + u_i(t) + \epsilon_{it}. \qquad (25)$$

This model can be viewed as a generic panel data model we introduced in equation (1) above 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)[3]. Once the individual effects $u_i(t)$ are estimated, technical efficiency for a particular firm at time t is calculated as $TE = \exp\{u_i(t) - \max_{j=1,\ldots,N}(u_j(t))\}$ for the CSSW and the KSS estimators (Cornwell, Schmidt, and Sickles, 1990). Technical efficiency is calculated similarly for the standard time-invariant fixed effects and random effects estimators following Schmidt and Sickles (1984). We also consider the Battese and Coelli (BC, 1992) estimator which is a likelihood-based random effects estimator wherein the likelihood function

[3] In keeping with the stochastic frontier paradigm we allow the technical efficiency to be correlated with the potentially distorted relative output allocations Y_{it}^*.

is derived from a mixture of normal noise and an independent one-sided efficiency error, usually specified as a half-normal. In the BC estimator, effect levels are allowed to differ across cross-sectional units but their temporal pattern is fixed across cross-sectional units and are specified as technical efficiencies $u_i(t) = -\exp(-\eta(t-T))\xi_i$ where ξ_i are independent half normal random effects and η parameterizes the temporal pattern in the firms' efficiencies.

4.2 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 667 banks or 8,004 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).

4.3 Empirical Results

The Hausman-Wu test, which tests the correlation assumptions for regressors and individual effects, was performed. The test statistic is 203.58, and the null hypothesis of no correlation is rejected at the 1% significance level. Thus there is strong evidence against the exogeneity assumption underlying the random effects GLS estimator. Consequently, in the following analysis we do not report the results from the random effects GLS estimator. The assumption is also fatal to the consistency of the random effects BC estimator. However, we will provide estimation results for the BC estimator as well to compare them with those from the other estimators (Within, CSSW, and KSS) which are robust to the existence of correlation between regressors and effects.

We test the dimensionality using $\Delta(l)$ test. The dimension L is chosen according to the rule described in Section 2 with the maximum dimension set to 8. Using the 1% significance level, the critical value is 2.33. With $L = 7$ the test statistic is 1.36 which is below the critical value. The optimal choice of dimensionality is thus seven[5].

Table 6 presents parameter estimates from Within, BC, CSSW, and KSS[6]. Ta-

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

[5] When we assume $L = 1$ and test the null hypothesis that the individual effect is constant, the test statistic Z is 165.02. Thus the null hypothesis of linear individual effect is strongly rejected.

[6] To calculate efficiency scores from the effects estimators, the effects estimates are trimmed at the top and bottom 5% level (see Berger, 1993). 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

ble 7 provides Spearman rank correlations among the estimators and shows relatively close correspondences (ranging from 0.7667 to 0.9854) among the rankings of efficiencies based on the different treatments of time-varying firm specific effects[7]. Average technical efficiencies for Within, BC, CSSW, and KSS are 0.4553, 0.6111, 0.6220, 0.6056 respectively. One may expect that during the period of deregulation firms tend to become more efficient due to increased competitive pressures in the industry. Figure 1 displays the temporal pattern of efficiency changes for time-variant efficiency estimators. We also construct an estimate of efficiency change over the sample period based on a pooled estimator that combines estimates from each of the time-varying measures. These results indicate a consensus growth of about 0.8% per year in efficiency during the sample period. Were these rates of cost diminution applied to the US banking industry the implied savings based on 1995 revenues and costs (Klee and Natalucci, 2005) would be on the order of $30 billion-our estimated measure of the benefits from deregulation of this key service industry.

5 Conclusion

In this paper we have introduced a new approach to estimating temporal heterogeneity in panel data models. 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. As we have pointed out, these methods can be transparently ported to the time series literature to address the issues of proper detrending filters in time series models.

Simulation experiments show that previous estimators, which do not allow for general temporal variations in effects terms or which misspecify the temporal pattern of variations, may suffer from serious distortions. 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 the limited branching regulatory environment for relatively small banks for the period of 1984-1995, and discovered that the relatively small banks in our sample have became more efficient over the years. The implied savings to the banking industry by 1995, were all banks to have enjoyed a similar efficiency gain as did our sample banks, is on the order of $30b.

---

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 with those from the CSS and KSS estimators.

[7] We report results with ray returns to scale set to one. No significant ray scale economies appear to exist using these treatments and in other analysis conducted by the authors with these data. Moreover, the equivalence of input and output oriented technical efficiency is preserved when scale economies are unity, thus avoiding difficutlies in interpretation that have been pointed out often in the productivity literature.

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 $b_1,\dots,b_L$ of $\mathcal{L}_T$ with $\frac{1}{T}\|b_r\|^2=1$, $r=1,\dots,L$, which can be chosen independent of X_{it}. Therefore, $v_i=\sum_{r=1}^L \vartheta_{ir}b_r$ with $\sum_{i=1}^n \vartheta_{ir}=0$. Let X_{ij} denote the T-vectors with elements X_{itj}, $t=1,\dots,T$, and recall that by Chebychev-type inequalities we have $\mathbf{P}(|Z_{n,T}|\geq\delta)\leq \mathbf{E}(|Z_{n,T}|^r)/\delta^r$ for all possible sequences of random variables $|Z_{n,T}|$ with $\mathbf{E}(|Z_{n,T}|^r)<\infty$ and all $\delta>0$. We thus necessarily have $Z_{n,T}=O_P(\mathbf{E}(|Z_{n,T}|^r)^{1/r})$.

In the general case, the $j=1,\dots,p$ 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 2) - 4) we obtain

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

Condition (16) of Assumption 4) then leads to $\|\mathbf{E}_\epsilon(\hat{\beta})-\beta\|=O_P((\frac{b_v(\kappa)}{T^{1/2}})$.

Note that $\mathcal{Z}_\kappa z=z$ and $(I-\mathcal{Z}_\kappa)z=(I-\mathcal{Z}_\kappa)^{1/2}z=0$ for all κ, if $z=(z(1),\dots,z(T))'$ is a linear function. If v_i and X_i are ulc-uncorrelated, then in the notation used in the definition of ulc-uncorrelatedness $\tilde{X}_{ij}'(I-\mathcal{Z}_\kappa)^{1/2}=\tilde{X}_{ij}^{*\prime}(I-\mathcal{Z}_\kappa)^{1/2}$, $(I-\mathcal{Z}_\kappa)^{1/2}v_i=(I-\mathcal{Z}_\kappa)^{1/2}v_i^*$, and therefore

$$\begin{aligned}\mathbf{E}(\tilde{X}_{ij}'(I-\mathcal{Z}_\kappa)v_i)^2 &= tr\left(\mathbf{E}((I-\mathcal{Z}_\kappa)^{1/2}\tilde{X}_{ij}\tilde{X}_{ij}'(I-\mathcal{Z}_\kappa)^{1/2})\cdot\mathbf{E}((I-\mathcal{Z}_\kappa)^{1/2}v_iv_i'(I-\mathcal{Z}_\kappa)^{1/2})\right) \\ &= \mathbf{E}\left(\mathbf{E}(\vartheta_{ir}^2)|b_r'(I-\mathcal{Z}_\kappa)\mathbf{E}(\tilde{X}_{ij}\tilde{X}_{ij}'))(I-\mathcal{Z}_\kappa)b_r\right)=O(T\cdot b_v(\kappa)^2)\end{aligned}$$

Since due to our normalization $\mathbf{E}(v_i(t)v_l(t))=O(\mathbf{E}(v_i(t)^2/n)$, it can be shown by similar arguments that $\mathbf{E}(\tilde{X}_{ij}'(I-\mathcal{Z}_\kappa)v_i)(\tilde{X}_{lj}'(I-\mathcal{Z}_\kappa)v_l)=O(T\cdot b_v(\kappa)^2/n)$ for $i\neq l$.

Therefore,
$\mathbf{E}(\sum_i \tilde{X}_{ij}'(I - \mathcal{Z}_\kappa)v_i)^2 = O(nT \cdot b_v(\kappa)^2)$, and $|\sum_i \tilde{X}_{ij}'(I - \mathcal{Z}_\kappa)v_i| = O_P(\sqrt{nT \cdot b_v(\kappa)^2})$, which leads to $\|\mathbf{E}_\epsilon(\hat{\beta}) - \beta\| = O_P((nT)^{-1/2} \cdot b_v(\kappa))$. By Assumptions 4) and 5) the assertion on $\hat{\beta} - \mathbf{E}_\epsilon(\hat{\beta}) = (\sum_i \tilde{X}_i'(I - \mathcal{Z}_\kappa)\tilde{X}_i)^{-1} \sum_i \tilde{X}_i'(I - \mathcal{Z}_\kappa)(\epsilon_i - \bar{\epsilon}) = (\sum_i \tilde{X}_i'(I - \mathcal{Z}_\kappa)\tilde{X}_i)^{-1} \sum_i \tilde{X}_i'(I - \mathcal{Z}_\kappa)\epsilon_i$ 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 $T^{-1/2}\|\mathcal{Z}_{\kappa^*}\bar{\epsilon}\| = O_P(\sqrt{tr(\mathcal{Z}_{\kappa^*}^2)/(nT)})$. The assertion then follows from Assumptions 2) and 4) as well as from the above results on the convergence of $\|\beta - \hat{\beta}\|$.

In order to prove Assertion (c) first note that

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

Therefore,

$$\hat{\Sigma}_{n,T} = \Sigma_{n,T} + B, \quad B = \frac{1}{n}\sum_i (v_i r_i' + r_i v_i' + r_i r_i'). \tag{26}$$

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

$$\gamma_r - \hat{\gamma}_r = S_r B \gamma_r + R, \quad \text{with } \|R\| \leq \frac{6 \sup_{\|a\|=1} a'B'Ba}{\min_s |\lambda_r - \lambda_s|^2} \tag{27}$$

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 the eigenvalue λ_s of $\Sigma_{n,T}$.

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) - 4) some straightforward arguments now lead to

$$\sup_{\|a\|=1} \|\frac{1}{n}\sum_i (I - \mathcal{Z}_\kappa)v_i v_i' a\| \leq \frac{1}{n}\sum_i \sup_{\|a\|=1} |v_i'a| \sqrt{v_i'(I - \mathcal{Z}_\kappa)(I - \mathcal{Z}_\kappa)v_i} = O_P(Tc(T)^{1/2} b_v(\kappa)), \tag{28}$$

$$\sup_{\|a\|=1} \|\frac{1}{n}\sum_i v_i v_i'(I - \mathcal{Z}_\kappa)a\| \leq \sup_{\|a\|=1} \frac{1}{n}\sum_i \sqrt{v_i'v_i}\, |v_i'(I - \mathcal{Z}_\kappa)a| = O_P(Tc(T)^{1/2} b_v(\kappa)), \tag{29}$$

$$\begin{aligned} \sup_{\|a\|=1} \|\frac{1}{n}\sum_i (\mathcal{Z}_\kappa \tilde{X}_i(\beta - \hat{\beta}))v_i'a\| &\leq \frac{1}{n}\sum_i |v_i'a| \sqrt{(\beta - \hat{\beta})'\tilde{X}_i'\mathcal{Z}_\kappa^2 \tilde{X}_i(\beta - \hat{\beta})} \\ &= O_P\left(Tc(T)^{1/2} d(T)^{1/2} (b_\beta(\kappa) + \frac{1}{\sqrt{nT}})\right). \end{aligned} \tag{30}$$

By similar arguments

$$\sup_{\|a\|=1} \|\frac{1}{n}\sum_i v_i(\mathcal{Z}_\kappa \tilde{X}_i(\beta - \hat{\beta}))'a\| = O_P(Tc(T)^{1/2} d(T)^{1/2} b_\beta(\kappa)) \tag{31}$$

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 $\frac{1}{n}\sum_i v_i\bar{\epsilon}'\mathcal{Z}_\kappa = 0$. Therefore

$$\sup_{\|a\|=1}\|\frac{1}{n}\sum_i \mathcal{Z}_\kappa(\epsilon_i-\bar{\epsilon})v_i'a\| \leq [tr((\frac{1}{n}\sum_i v_i\epsilon_i'\mathcal{Z}_\kappa)\cdot(\frac{1}{n}\sum_i \mathcal{Z}_\kappa\epsilon_i v_i'))]^{\frac{1}{2}} = O_P\left(\sqrt{\frac{Tc(T)\cdot tr(\mathcal{Z}_\kappa^2)}{n}}\right), \tag{32}$$

Similarly,

$$\sup_{\|a\|=1}\|\frac{1}{n}\sum_i v_i(\epsilon_i-\bar{\epsilon})'\mathcal{Z}_\kappa a\| = O_P\left(\sqrt{\frac{Tc(T)\cdot tr(\mathcal{Z}_\kappa^2)}{n}}\right). \tag{33}$$

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

$$\sup_{\|a\|=1}\|\frac{1}{n}\sum_i (I-\mathcal{Z}_\kappa)v_i v_i'(I-\mathcal{Z}_\kappa)a\| = O_p(T\cdot b_v(\kappa)^2), \tag{34}$$

$$\sup_{\|a\|=1}\|\frac{1}{n}\sum_i (\mathcal{Z}_\kappa\tilde{X}_i(\beta-\hat{\beta}))(\mathcal{Z}_\kappa\tilde{X}_i(\beta-\hat{\beta}))'a\| = O_P\left(Td(T)\cdot(b_\beta(\kappa)^2+\frac{1}{nT})\right). \tag{35}$$

Obviously,
$\mathbf{E}_\epsilon(tr[(\frac{1}{n}\sum_i \mathcal{Z}_\kappa\epsilon_i\epsilon_i'\mathcal{Z}_\kappa - \sigma^2\mathcal{Z}_\kappa^2)\cdot(\frac{1}{n}\sum_i \mathcal{Z}_\kappa\epsilon_i\epsilon_i'\mathcal{Z}_\kappa - \sigma^2\mathcal{Z}_\kappa^2)]) = \frac{1}{n}\mathbf{E}(tr[\mathcal{Z}_\kappa\epsilon_i\epsilon_i'\mathcal{Z}_\kappa\mathcal{Z}_\kappa\epsilon_i\epsilon_i'\mathcal{Z}_\kappa - \sigma^4\mathcal{Z}_\kappa^4]) = O_P(\frac{tr(\mathcal{Z}_\kappa^4)}{n})$, and construction of $\mathcal{Z}_\kappa$ implies that $tr(\mathcal{Z}_\kappa)$ is of the same order of magnitude as $tr(\mathcal{Z}_\kappa^s)$ for all $s = 1, 2, 4$. Therefore

$$\sup_{\|a\|=1}\|\frac{1}{n}\sum_i (\mathcal{Z}_\kappa(\epsilon_i-\bar{\epsilon})(\epsilon_i-\bar{\epsilon})'\mathcal{Z}_\kappa - \sigma^2\mathcal{Z}_\kappa^2)a\| = O_P\left(\sqrt{\frac{tr(\mathcal{Z}_\kappa^2)}{n}}\right) \tag{36}$$

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 (27) with (28) - (36) we thus obtain

$$\begin{aligned}\|S_r B\gamma_r\| &\leq \|\sigma^2 S_r\mathcal{Z}_\kappa^2\gamma_r\| + \frac{1}{\min_s|\lambda_r-\lambda_s|}\|(B-\sigma^2\mathcal{Z}_\kappa^2)\gamma_r\| \\ &= \|\sigma^2 S_r\mathcal{Z}_\kappa^2\gamma_r\| + O_P\left(\frac{b_v(\kappa)}{c(T)^{1/2}} + \sqrt{\frac{tr(\mathcal{Z}_\kappa^2)}{nTc(T)}}\right)\end{aligned} \tag{37}$$

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

$$\|\sigma^2 S_r\mathcal{Z}_\kappa^2\gamma_r\| \leq \|\sigma^2 S_r(I-\mathcal{Z}_\kappa)\gamma_r\| + \|\sigma^2 S_r\mathcal{Z}_\kappa(I-\mathcal{Z}_\kappa)\gamma_r\| = O_P(\frac{b_v(\kappa)}{Tc(T)^{3/2}}), \tag{38}$$

Let us now consider the remainder term R in (27). 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 (28) - (36) then imply

$$\begin{aligned}\frac{\sup_{\|a\|=1} a'B'Ba}{\min_s|\lambda_r-\lambda_s|^2} &\leq 2\frac{\sup_{\|a\|=1} a'(B-\sigma^2\mathcal{Z}_\kappa^2)'(B-\sigma^2\mathcal{Z}_\kappa^2)a}{\min_s|\lambda_r-\lambda_s|^2} + 2\frac{\sup_{\|a\|=1} a'\mathcal{Z}_\kappa^4 a}{\min_s|\lambda_r-\lambda_s|^2} \\ &= O_P\left(\frac{b_v(\kappa)^2}{c(T)} + \frac{1}{T^2c(T)^2} + \frac{tr(\mathcal{Z}_\kappa^2)}{nTc(T)})\right)\end{aligned} \tag{39}$$

By (27), (37), (38) and (39) the asserted rate of convergence follows from

$$T^{-1/2}\|g_r - \hat{g}_r\| = \|\gamma_r - \hat{\gamma}_r\| = O_P\left(\frac{b_v(\kappa)}{c(T)^{1/2}} + \frac{1}{T^2 c(T)^2} + \sqrt{\frac{tr(\mathcal{Z}_\kappa^2)}{nTc(T)}}\right). \tag{40}$$

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

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

Moreover, one can infer from relations (27) - (40) that

$$\frac{1}{T}(\hat{g}_r - g_r)'v_i = \frac{1}{n\sqrt{T}}\sum_j \gamma_r' v_j v_j'(I - \mathcal{Z}_\kappa)S_r v_i + \frac{1}{n\sqrt{T}}\sum_j \gamma_r'(I - \mathcal{Z}_\kappa)v_j v_j' S_r v_i$$
$$\frac{1}{n\sqrt{T}}\sum_j \gamma_r' v_j \epsilon_j' \mathcal{Z}_\kappa S_r v_i + O_P\left(\frac{b_v(\kappa)^2}{c(T)^{1/2}} + d(T)^{1/2}(b_\beta(\kappa) + \frac{1}{\sqrt{nT}}) + \frac{1}{T^2 c(T)^{3/2}}\right)$$

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

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

as well as

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

Furthermore, $\frac{1}{n\sqrt{T}}\sum_j \gamma_r' v_j \epsilon_j' \mathcal{Z}_\kappa S_r v_i = O_P(\frac{1}{\sqrt{nT}})$. This implies

$$(\hat{\theta}_{ri} - \theta_{ri}) = \frac{1}{T}g_r'\epsilon_i + O_P\left(\frac{b_v(\kappa)^2}{c(T)^{1/2}} + d(T)^{1/2}(b_\beta(\kappa) + \frac{1}{\sqrt{nT}}) + \frac{1}{T^2 c(T)^{3/2}}\right) + o_P\left(T^{-1/2}\right).$$

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 $\frac{b_v(\kappa)^2}{c(T)^{1/2}} + d(T)^{1/2}b_\beta(\kappa) + \frac{tr(\mathcal{Z}_\kappa^2)}{nT} = o(T^{-1/2})$, 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 (27) - (40) then show that $\tilde{\lambda}_r = O(Tc(T))$, $\frac{1}{\min_s |\tilde{\lambda}_r - \tilde{\lambda}_s|} = O_P(\frac{1}{T \cdot c(T)})$, $\|\gamma_r - \tilde{\gamma}_r\| = O_P(\frac{b_v(\kappa)}{c(T)^{1/2}})$, and

$$\|\hat{\gamma}_r - \tilde{\gamma}_r\| = O_P\left(\frac{d(T)^{1/2} b_\beta(\kappa)}{c(T)^{1/2}} + \frac{1}{T^2 c(T)^2} + \sqrt{\frac{tr(\mathcal{Z}_\kappa^2)}{nTc(T)}}\right) \tag{41}$$

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

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

$$\sum_{r=L+1}^{T} \hat{\lambda}_r = tr(\mathcal{P}_L \tilde{B}) + R^*, \quad \text{with } R^* \leq \frac{6L \sup_{\|a\|=1} a' \tilde{B}' \tilde{B} a}{\min_s |\tilde{\lambda}_r - \tilde{\lambda}_s|} \tag{42}$$

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

$$\frac{6L \sup_{\|a\|=1} a' \tilde{B}' \tilde{B} a}{\min_s |\tilde{\lambda}_r - \tilde{\lambda}_s|} = O_P\left(Td(T) b_\beta(\kappa)^2 + \frac{1}{Tc(T)} + \frac{tr(\mathcal{Z}_\kappa^2)}{n}\right). \tag{43}$$

On the other hand,

$$tr(\mathcal{P}_L \tilde{B}) = tr\left(\frac{1}{n}\sum_i \mathcal{P}_L \mathcal{Z}_\kappa \tilde{X}_i (\beta - \hat{\beta})(\beta - \hat{\beta})' \tilde{X}_i' \mathcal{Z}_\kappa\right) + tr\left(\mathcal{P}_L \mathcal{Z}_\kappa (\frac{1}{n}\sum_i (\epsilon_i - \bar{\epsilon})(\epsilon_i - \bar{\epsilon})')\mathcal{Z}_\kappa\right) \tag{44}$$

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

$$\frac{tr(\mathcal{P}_L\tilde{B}) \ - \sigma^2\left(1-\frac{1}{n}\right)tr(\mathcal{Z}_\kappa\mathcal{P}_L\mathcal{Z}_\kappa)}{\sqrt{\frac{2\sigma^4}{n}\cdot tr((\mathcal{Z}_\kappa\mathcal{P}_L\mathcal{Z}_\kappa)^2)}} \to_d \mathbf{N}(0,1). \tag{45}$$

Since $tr(\mathcal{P}_L\tilde{B}) = tr(\mathcal{P}_L\hat{\Sigma}_n)$, (21) is an immediate consequence. By (41)- (43) , Relation (45) remains valid when $tr(\mathcal{P}_L\tilde{B})$ is replaced by $\sum_{r=L+1}^T \hat{\lambda}_r$ as well as $\mathcal{P}_L$ by $\hat{P}_L$. This proves (20) and hence completes the proof of the theorem. □

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

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

Clearly,

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

By Assumption 2) the well-known properties of $\mathcal{Z}_\kappa$ imply $1/tr(I-\mathcal{Z}_\kappa) = O_P(T^{-1})$, and therefore

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

Consequently, with

$$0 \leq R_{n,T} = \frac{1}{(n-1)\cdot tr((I-\mathcal{Z}_\kappa)^2)}\sum_i v_i'(I-\mathcal{Z}_\kappa)^2 v_i = O_p(b_v(\kappa)^2) \tag{46}$$

we obtain

$$\hat{\sigma}^2 = \sigma^2 + R_{n,T} + o_p(1). \tag{47}$$

Let us now consider the behavior of $\Delta(l)$ for $l < L$. We can immediately infer from (47) that

$$\begin{aligned}\Delta(l) = \Bigg[&\frac{n\sum_{r=l+1}^L\hat{\lambda}_r \ - (n-1)(\sigma^2+R_{n,T})\cdot tr(\mathcal{Z}_\kappa(\hat{\mathcal{P}}_l-\hat{\mathcal{P}}_L)\mathcal{Z}_\kappa) - (n-1)R_{n,T}\cdot tr(\mathcal{Z}_\kappa\hat{\mathcal{P}}_l\mathcal{Z}_\kappa)}{\hat{\sigma}^2\sqrt{2n\cdot tr((\mathcal{Z}_\kappa\hat{\mathcal{P}}_l\mathcal{Z}_\kappa)^2)}} \\ &+\frac{n\sum_{r=L+1}^T\hat{\lambda}_r \ - (n-1)\sigma^2\cdot tr(\mathcal{Z}_\kappa\hat{\mathcal{P}}_L\mathcal{Z}_\kappa)}{\hat{\sigma}^2\sqrt{2n\cdot tr((\mathcal{Z}_\kappa\hat{\mathcal{P}}_l\mathcal{Z}_\kappa)^2)}}\Bigg](1+o_P(1)).\end{aligned} \tag{48}$$

By Assumption 2) and Theorem 1d) $n\sum_{r=l+1}^{L}\hat{\lambda}_r = \sum_{r=l+1}^{L} T\sum_i \hat{\theta}_{ir}^2$ 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 first term on the right hand side of (48) increases as $n, T \to \infty$, while the second term is still bounded in probability. We can thus infer that for $l < L$

$$\mathbf{P}(\Delta(l) > z_{1-\alpha}) \to 1 \quad \text{and therefore } \mathbf{P}(\hat{L} \neq l) \to 1 \tag{49}$$

as $n, T \to \infty$.

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

$$\limsup_{n,T\to\infty} \mathbf{P}(\Delta(L) \geq z_{1-\alpha}) \leq \alpha. \tag{50}$$

The assertion of the theorem now is an immediate consequence of (49) and (50). □

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.

Ahn, S.C., Y. Lee, and P.J. Schmidt (2001), "GMM estimation of linear panel data models with time-varying individual effects," *Journal of Econometrics,* 101, 219-255.

Ahn, S. C., Y. Lee, and P. J. Schmidt (2005) "Panel data models with multiple time-varying individual effects: application to a stochastic frontier production model," mimeo, Michigan State University.

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

Bai, J. (2003) "Inferential theory for factor models of large dimensions" *Econometrica,* vol. 71, issue 1, pages 135-171.

Bai, J. (2005) "Panel data models with interactive fixed effects," April 2005, mimeo, Department of Economics, New York University.

Bai, J. and S. Ng (2002) "Determining the number of factors in approximate factor models," *Econometrica,* vol. 70, issue 1, pages 191-221

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

Baltagi, B., P. Egger, and M. Pfaffermayr (2003), "A generalized design for bilateral trade flow models," *Economics Letters,* vol. 80, issue 3, pages 391-397.

Baltagi, B. (2005) *Econometric Analysis of Panel Data, 3rd ed.*, NJ: John Wiley & Sons.

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.

Chang, Y. (2004), "Bootstrap unit root tests in panels with cross-sectional dependency," *Journal of Econometrics,* 120, 263-293.

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. (1988), *Nonparametric Regression and Spline Smoothing*, Marcel Dekker, New York.

Ferré, L. (1995), "Improvement of some multivariate estimates by reduction of dimensionality," *Journal of Multivariate Analysis* 54, 147-162.

Forni, M. and M. Lippi (1997) "Aggregation and the microfoundations of dynamic macroeconomics," *Oxford, Oxford University Press.*

Forni, M. and L. Reichlin (1998) "Let's get real: a factor analytic approach to disaggregated business cycle dynamics," *Review of Economic Studies*, 65 653-473.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000), "The generalized dynamic factor model: identification and estimation, " *Review of Economics and Statistics*, 82, 540-554.

Han, C. , L. Orea, and P. Schmidt (2005) "Estimation of a panel data model with parametric temporal variation in individual effects" *Journal of Econometrics,* 126, 241-267

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.

Kao, C. and M. H. Chiang (2000) "On the estimation and inference of a cointegrated regression in panel data," *Advances in Econometrics*, 15,179-222.

Klee, E. C., and F. M. Natalucci (2005) "Profits and balance sheet developments at U. S. commercial banks in 2004," *Federal Reserve Bulletin*, Spring, 2005.

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.

Maddala, G. S. and I. M. Kim (1998) "Unit Roots, cointegration and structural change," *Cambridge University Press, Cambridge.*

Mark, N. C. and D. Sul (2003) "Cointegration vector estimation by panel dlos and long-run money demand." *Oxford Bulletin of Economics and Statistics, Blackwell Publishing,* vol. 65, issue 5, pages 655-680.

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

Nelson, C. R. and C. I. Plosser (1982) "Trends and random walks in macroeconomics time series: some evidence and implications," *Journal of Monetary Economics*, 10, 139-162.

Park, B.U., R.C. Sickles. and L., Simar. (1998) "Stochastic frontiers: a semiparametric approach,"*Journal of Econometrics,* 84, 273-301.

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

Park, B.U., and R.C. Sickles and L. Simar (2005) "Semiparametric efficient estimation of dynamic panel data models," forthcoming in *Journal of Econometrics.*

Ramsay, J. and B. Silverman (1997) *Functional Data Analysis*, Heidelberg: Springer-Verlag.

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

Sickles, R. C. (2004), "Panel Estimators and the identification of firm-specific efficiency levels in parametric, semiparametric and nonparametric settings,"*Journal of Econometrics,* 50, 126, 305-334.

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

Stock, J. H., and M. W. Watson (2002), "Forecasting using principal components from a large number of predictors," *Journal of the American Statistical Association*, 97, 1167-1179.

Utreras, F. (1983). "Natural spline functions, their associated eigenvalue problem," *Numerical Mathematics*, 42, 107-117

Table 1. Monte Carlo Simulation Results for DGP1

MSE of Coefficients

N	T	Within	GLS	CSSW	KSS
30	12	0.07258	0.06381	0.00867	0.00874
	30	0.02832	0.02355	0.00240	0.00258
100	12	0.01862	0.01643	0.00266	0.00273
	30	0.00678	0.00649	0.00073	0.00075
300	12	0.00610	0.00609	0.00086	0.00087
	30	0.00210	0.00208	0.00023	0.00023

MSE of Effects

N	T	Within	GLS	CSSW	KSS	L
30	12	0.1770	0.1746	0.0091	0.0091	2.4070
	30	0.1666	0.1663	0.0036	0.0043	2.8050
100	12	0.1285	0.1280	0.0072	0.0073	2.9688
	30	0.1240	0.1240	0.0029	0.0030	3.0100
300	12	0.1025	0.1025	0.0059	0.0060	3.0040
	30	0.1001	0.1001	0.0024	0.0025	3.0060

Table 2. Monte Carlo Simulation Results for DGP2

MSE of Coefficients

N	T	Within	GLS	CSSW	KSS
30	12	0.02414	0.02085	0.01370	0.00477
	30	0.00699	0.00675	0.00662	0.00188
100	12	0.00974	0.00842	0.00488	0.00139
	30	0.00201	0.00195	0.00193	0.00052
300	12	0.00341	0.00430	0.00169	0.00047
	30	0.00071	0.00073	0.00063	0.00028

MSE of Effects

N	T	Within	GLS	CSSW	KSS	L
30	12	0.1655	0.1630	0.0601	0.0170	1.0050
	30	0.0976	0.0975	0.0692	0.0100	1.0000
100	12	0.1544	0.1547	0.0491	0.0117	1.0000
	30	0.0890	0.0890	0.0624	0.0072	1.0000
300	12	0.1480	0.1484	0.4500	0.0104	1.0000
	30	0.0860	0.0861	0.0597	0.0065	1.0000

Table 3. Monte Carlo Simulation Results for DGP3

MSE of Coefficients

N	T	Within	GLS	CSSW	KSS
30	12	0.01346	0.00589	0.02166	0.00662
	30	0.00464	0.00227	0.00598	0.00203
100	12	0.00465	0.00188	0.00708	0.00168
	30	0.00153	0.00074	0.00193	0.00041
300	12	0.00148	0.00066	0.00241	0.00038
	30	0.00049	0.00023	0.00062	0.00012

MSE of Effects

N	T	Within	GLS	CSSW	KSS	L
30	12	1.1064	1.0411	1.1410	0.3586	2.0184
	30	1.0541	1.0318	1.1158	0.2213	1.9382
100	12	1.0517	1.0311	1.0276	0.2086	2.1727
	30	1.0350	1.0285	1.0810	0.0879	2.0776
300	12	1.0398	1.0337	1.0144	0.1787	2.0859
	30	1.0308	1.0287	1.0728	0.0727	2.0432

Table 4. Monte Carlo Simulation Results for DGP4

MSE of Coefficients

N	T	Within	GLS	CSSW	KSS
30	12	0.00544	0.00484	0.00841	0.00615
	30	0.00188	0.00181	0.00221	0.00200
100	12	0.00176	0.00122	0.00262	0.00183
	30	0.00061	0.00051	0.00073	0.00062
300	12	0.00056	0.00080	0.00086	0.00058
	30	0.00020	0.00026	0.00024	0.00020

MSE of Effects

N	T	Within	GLS	CSSW	KSS	L
30	12	0.1213	0.1126	0.3387	0.1519	1.0320
	30	0.0472	0.0462	0.1288	0.0638	1.0100
100	12	0.0929	0.0876	0.2706	0.1032	1.0430
	30	0.0363	0.0354	0.1062	0.0414	1.0230
300	12	0.0795	0.0811	0.2366	0.0838	1.0280
	30	0.0319	0.0323	0.0947	0.0339	1.0200

Table 5.Summary Statistics for Small Banks

Variable	Definition	Mean
reln	Log of real estate loans	8.559
ciln	Log of commercial and industrial loans	7.338
inln	Log of installment loans	7.632
CD	Log of certificate of deposits	7.400
DD	Log of demand deposits	7.875
OD	Log of retail time and savings deposits	9.977
lab	Log of labor	4.499
cap	Log of capital	5.613
purf	Log of purchased funds	10.079
	Number of observations	8004

Table 6. Estimation Results

	Within	BC	CSSW	KSS
CD	-0.0357 (0.0047)	-0.0332 (0.0043)	-0.0095 (0.0032)	-0.0008 (0.0019)
DD	-0.0678 (0.0155)	-0.0244 (0.0124)	-0.0908 (0.0134)	-0.0410 (0.0109)
OD	-0.1451 (0.0097)	-0.1433 (0.0091)	-0.1295 (0.0069)	-0.0440 (0.0200)
lab	-0.1517 (0.0165)	-0.1403 (0.0130)	-0.1639 (0.0139)	-0.1254 (0.0093)
cap	-0.0456 (0.0054)	-0.0523 (0.0048)	-0.0461 (0.0054)	-0.0289 (0.0053)
purf	-0.5522 (0.0208)	-0.6065 (0.0151)	-0.5601 (0.0162)	-0.7598 (0.0268)
ciln	0.1583 (0.0045)	0.1596 (0.0042)	0.1468 (0.0037)	0.1202 (0.0031)
inln	0.3745 (0.0061)	0.3639 (0.0054)	0.3512 (0.0056)	0.3237 (0.0050)
time	0.0154 (0.0009)	0.0023 (0.0013)	-	-
Avg TE	0.4553	0.6111	0.6220	0.6056

Table 7. Spearman Rank Correlations of Efficiencies

	Within	BC	CSSW	KSS
Within	1.0000	.	.	.
BC	0.9854	1.0000	.	.
CSSW	0.8743	0.8785	1.0000	.
KSS	0.7667	0.7937	0.8974	1.0000

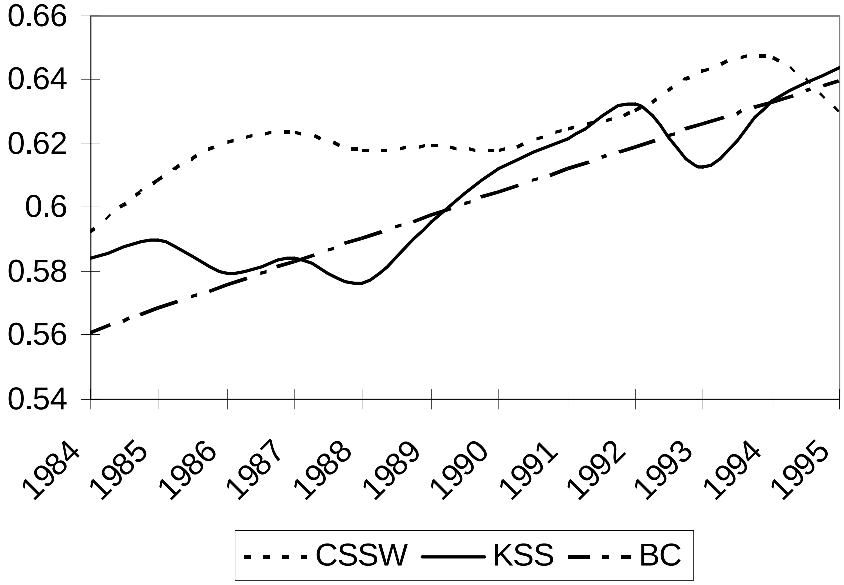
0.66
0.64
0.62
0.6
0.58
0.56
0.54
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
CSSW
KSS
BC

Figure 1: