BONN ECON DISCUSSION PAPERS

Discussion Paper 1/2006

A New Panel Data Treatment for Heterogeneity in Time Trends

by

Alois Kneip, Robin C. Sickles, Wonho Song

December 2005

Bonn Graduate School of Economics
Department of Economics
University of Bonn
Adenauerallee 24 - 42
D-53113 Bonn

The Bonn Graduate School of Economics is
sponsored by the

Deutsche Post World Net

MAIL EXPRESS LOGISTICS FINANCE

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 2005

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.

JEL Classification: C13, C14, C23, G21.

Key words and phrases: Time trends, panel models, principal component analysis, smoothing splines, banking efficiency.

---

*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 citicisms 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.

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 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

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

Although we consider non-constant individual effects, we will assume that $u_i(t)$ is varying "slowly" with t, and that $u_1, \ldots, u_n$ 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. It must be emphasized that identifiability of (1) requires that all variables X_{itj}, $j = 1, \ldots, p$ possess a considerable variation over t. Additional variables, like e.g. socioeconomic attributes, which characterize individuals but do not change over time may be analyzed in a second step by studying possible effects on the structure of the corresponding functions u_i.

Based on (1), the coefficients β as well as the functions u_i can be estimated by semi-parametric 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 $L \in \{0, 1, 2, \ldots\}$ there exist some basis functions (common factors) $g_1, \ldots, g_L$ such that

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

Together with (1) this leads to the model

$$Y_{it} = \sum_{j=1}^{p} \beta_j X_{itj} + w(t) + \sum_{r=1}^{L} \theta_{ir} g_r(t) + \epsilon_{it}, \quad i = 1, \ldots, n, t = 1, \ldots, T \tag{3}$$

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), (3) additionally incorporates a fixed effect term. This is similar to the approach by Ahn et al. (2005). Note that by (2) only the linear factor space spanned by $g_1, \ldots, g_L$ is identified but not the particular basis. We will thus additionally assume that

(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 (2) 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. Individual differences are captured by the coefficients θ_{ir} and standard methods can be applied in order to study the effects of additional explanatory variables, like socioeconomic characteristics, on the distribution of these coefficients.

Note that if individual effects are constant then (2) is satisfied with $L = 1$ and $g_1(t) \equiv 1$. If instead, as proposed by Cornwell, Schmidt, and Sickles (1990), the u_i can be modelled by quadratic polynomials then $L = 3$ and g_1, g_2, g_3 correspond to a polynomial basis. To give another example, assume that $u_i(t) = \vartheta_i r_t$, where r_t is a realization of a random walk. Then, $L = 1$, $w(t) = \bar{\vartheta}_i r_t$, $g_r(t) = \frac{r_t}{\sqrt{T}}$ and $\theta_{1i} = \sqrt{T}(\vartheta_i - \bar{\vartheta}_i)$. Indeed, the general model (2) does not impose any strong restriction on the structure of the functions v_i. It is only assumed that for *some* L relation (2) 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

$$\Sigma_{n,T} = \frac{1}{n}\sum_i v_i {v_i}' \tag{4}$$

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

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

$$\theta_{ir} = \frac{1}{T}\sum_t v_i(t) g_r(t) \quad \text{for all} \quad r = 1, 2, \ldots,\ i = 1, \ldots, n, \tag{6}$$

$$\lambda_r = \frac{T}{n}\sum_i \theta_{ir}^2 \quad \text{for all} \quad r = 1, 2, \ldots \tag{7}$$

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

$$\sum_{r=l+1}^{T} \lambda_r = \sum_{i,t}(v_i(t) - \sum_{r=1}^{l} \theta_{ir} g_r(t))^2 = \min_{\tilde{g}_1,\ldots,\tilde{g}_l} \sum_i \min_{\vartheta_{i1},\ldots,\vartheta_{il}} \sum_t (v_i(t) - \sum_{r=1}^{l} \vartheta_{ir} \tilde{g}_r(t))^2 \qquad (8)$$

One can infer from relation (8) 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 (2) 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, estimation procedures necessarily rely on sample characteristics and, as will be seen in Subsection 2.2, our theoretical setup also covers situations where $\mathbf{E}(\theta_{ir}^2) \to \infty$ as $T \to \infty$. Furthermore, the real object of interest in model (2) 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 (3) 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

$$\sum_i \frac{1}{T} \sum_t (Y_{it} - \bar{Y}_t - \sum_{j=1}^{p} \beta_j (X_{itj} - \bar{X}_{tj}) - v_i(t))^2 + \sum_i \kappa \frac{1}{T} \int_1^T (v_i^{(m)}(s))^2 ds \qquad (9)$$

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 (9) 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

$$\sum_i \left(\|Y_i - \bar{Y} - (X_i - \bar{X})\beta - Z\zeta_i\|^2 + \kappa \zeta_i' A \zeta_i \right), \tag{10}$$

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

$$\hat{\beta} = \left(\sum_i (X_i - \bar{X})'(I - \mathcal{Z}_\kappa)(X_i - \bar{X}) \right)^{-1} \sum_i (X_i - \bar{X})'(I - \mathcal{Z}_\kappa)(Y_i - \bar{Y}) \tag{11}$$

as well as

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

Therefore,

$$\hat{v}_i = Z\hat{\zeta}_i = \mathcal{Z}_\kappa (Y_i - \bar{Y} - (X_i - \bar{X})\hat{\beta}) \tag{12}$$

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

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

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

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

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

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

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

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

$$\sum_t (Y_{it} - \bar{Y}_t - (X_i - \bar{X})\hat{\beta} - \sum_{r=1}^{L} \vartheta_{ri}\hat{g}_r(t))^2 \tag{13}$$

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

Remark: In principle, it is possible to iterate this procedure. In addition to estimating θ_{ri}, (13) 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 to determine new approximations $\hat{g}_r^{(1)}$, etc. Such iterations may possess the potential to enhance efficiency of estimates. A precise analysis is, however, not in the scope of the present paper, and our theoretical results will only refer to Step 1 - 4 as described above.

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 $n, T \to \infty$. The smoothing parameter $\kappa \equiv \kappa(n, T)$ may either remain fixed or may increase with n, T. Model (2) is assumed to possess a fixed dimension L for all n, T. The following assumptions then provide the basis of our analysis.

We will write $\lambda_{min}(A)$ and $\lambda_{max}(A)$ to denote the minimal and maximal eigenvalues of a symmetric matrix A, 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 monotonically increasing function $c(T)$ of T such that as $n, T \to \infty$

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

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

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

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

   satisfy

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

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

4) $\frac{\mathbf{E}(\frac{1}{T}\sum_{t=1}^T \bar{X}_{it,j}^2)}{\mathbf{E}(\frac{1}{T}\sum_{t=1}^T w(t)^2)} = O(1)$, and there exists a monotonically increasing 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,

$$\lambda_{max}\left([\sum_i (X_i - \bar{X})'(I - \mathcal{Z}_\kappa)(X_i - \bar{X})]^{-1}\right) = O_p(\frac{1}{nT}) \tag{14}$$

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

$$a'(I - \mathcal{Z}_\kappa) \cdot \mathbf{E}\left((X_{ij} - \bar{X})(X_{ij} - \bar{X})'\right)(I - \mathcal{Z}_\kappa)a \leq D \cdot \|(I - \mathcal{Z}_\kappa)a\|^2. \tag{15}$$

   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.

First note that Assumptions 1) and 2) formalize our model introduced in the proceeding sections. Different from existing literature on factor models $\mathbf{E}(\frac{1}{T}\sum_{t=1}^T v_i(t)^2)$ is allowed to increase with T. The growth rate is given by $c(T)$.

Assumption 3) quantifies our requirement of "smooth functions" v_i. Spline theory provides a basis to understand the impact of this assumption (see, for example, de Boor 1978, or Eubank 1988). We will concentrate on 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 $z_1, \dots, z_T$ is the natural spline basis used to construct our estimator in Section 2.1, while $z_1^*, \dots, z_T^*$ is a basis for all natural splines defined on $[0, 1]$ with knots $1/T, 2/T, \dots, 1$. Obviously, $z_j'' = z_j^{*''}/T^2$. Defining the matrices Z^* and $A^* = \{\int_{1/T}^1 z_j^{*(m)}(s) z_k^{*(m)}(s) ds\}_{j,k=1,\dots,T}$ similar to Z, A in Section 2.1, some straightforward arguments show that $\mathcal{Z}_\kappa = (I - \kappa(Z')^{-1}AZ^{-1})^{-1} = (I - \frac{\kappa}{T^4}T \cdot (Z^{*\prime})^{-1}A^*(Z^*)^{-1})^{-1}$. 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

$$\frac{1}{T}\|(I - \mathcal{Z}_\kappa)v_i\|^2 \leq \kappa\frac{1}{T}\int_1^T |\tilde{v}_i''(t)|^2 dt, \; tr(\mathcal{Z}_\kappa^2) =\leq q \cdot \frac{T}{\kappa^{1/4}}, \;\; \frac{1}{T}\sum_t \mathbf{Var}_\epsilon(\hat{v}_i(t)) = O_P(\frac{\sigma^2 tr(\mathcal{Z}_\kappa^2)}{T}) \tag{16}$$

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

Note that it is only required that the above assumptions hold as "$n, T \to \infty$". Of course, $n \to \infty$ will correspond to drawing more and more individuals at random, but different asymptotic setups may be used to describe the situation as "$T \to \infty$". The point is that any asymptotic theory aims to provide first order approximations of a complex finite sample behavior. In practice, one has always to consider the question which asymptotic setup is best suited to approximate the respective finite sample situation. In this paper we will mainly concentrate on the following Situation 1, which formalizes smoothness in the most stringent way.

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

Formally this setup can be described as follows. 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. smooth functions $\nu_1, \dots, \nu$ on $L^2[0,1]$ such that $\mu_i(\frac{t}{T}) = w(t)$ and $\nu_i(\frac{t}{T}) = v_i(t)$ for $t = 1, \dots, T$. Smoothness then naturally translates into the assumption that μ as and $\nu_1, \dots, \nu$ are a.s. *twice continuously differentiable* with $\mathbf{E}(\int_0^1 \nu_i''(t)^2 dt) < \infty$.

In this case we, of course, obtain $\frac{1}{T}\sum_t v_i(t) = O(1), \frac{1}{T}\sum_t w(t) = O(1)$ as $T \to \infty$ and, hence, Assumptions 2) and 3) refer to a constant function $c(T) = 1$. 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 (16) we can infer that an optimal smoothing parameter for estimating v_i then satisfies $\frac{\kappa}{T^4} = \kappa_T \sim T^{-4/5}$, which means that the smoothing parameter κ in (9) has to increase rapidly as $T \to \infty$. Similar results are to be obtained with respect to w. Assumption 3) then holds with

$$b_v(n,T)^2 = \mathbf{E}(\frac{1}{T}\|(I - \mathcal{Z}_\kappa)v_i\|^2) = O(T^{-4/5}), \quad tr(\mathcal{Z}_\kappa^2)/T = O(T^{-4/5}). \tag{17}$$

Similar rates of convergence then can also be derived for $b_w(n,T)$. 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*.

From a practical point of view it is important to interpret this asymptotic setup in an appropriate manner. Construction of spline smoothers implies that the value of the integral $\frac{1}{T}\int_1^T |\tilde{v}_i''(t)|^2 dt$ in (16) is of the same order of magnitude as the average squared second differences $\frac{1}{T}\sum_t (v_i(t+1) - 2v_i(t) + v_i(t-1))^2$ Therefore, for a given finite sample theoretical results based on the above setup will provide a reasonable first order approximation if it can be assumed that the functions v_i are smooth enough such that $\frac{1}{T}\sum_t (v_i(t+1) - 2v_i(t) + v_i(t-1))^2$ is *smaller* than the error variance σ^2. In this case a fairly large smoothing parameter κ will still result in a small bias while at the same time the average variance of the estimator will be smaller than σ^2 (due to $tr(\mathcal{Z}_\kappa^2) \ll T$).

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

Let us consider the example of a random walk. Assume that 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 $L = 1$, $w(t) = \bar{\vartheta}_i r_t$, $g_r(t) = \frac{r_t}{\sqrt{T}}$ and $\theta_{1i} = \sqrt{T}(\vartheta_i - \bar{\vartheta}_i)$. Since $\frac{1}{T}\sum_{t=1}^T \mathbf{E}(\vartheta_i^2 r_t^2) = O(T)$, Assumptions 2) and 3) are then satisfied with $c(T) = T$.

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

$$b_v(n,T) = \mathbf{E}(\frac{1}{T}\|(I - \mathcal{Z}_\kappa)v_i\|) = O(1), \quad tr(\mathcal{Z}_\kappa^2)/T = O(1). \tag{18}$$

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:

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

Assumption 4) contains regularity conditions which imposes a restriction on the design matrix. It essentially requires that the time paths $\{X_{itj} - \bar{X}_{ij}\}_t$ are "less smooth" than those of $\{v_i(t)\}_t$. In particular, stationary processes generate non-smooth time parts.

When considering the simplest case $p = 1$, Assumption 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 (14) as well as (15) 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 assumption $\mathcal{Z}_\kappa \mathbf{1} = \mathbf{1}$ for $\mathbf{1} = (1, 1, \ldots, 1)'$ we arrive at

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

The maximal eigenvalue of Γ remains bounded as $T \to \infty$, and relations (14) as well as (15) are an immediate consequence of the structure of $\mathcal{Z}_\kappa$.

We are now ready to state our main theorem. A proof can be found in the appendix. We will use the notation "$\mathbf{E}_\epsilon$" to denote conditional expectation given v_i and X_i, $i = 1, \ldots n$. Moreover, $\tilde{X}_i = X_i - \bar{X}$, and we will say that v_i and X_i are $\mathcal{Z}_\kappa$-uncorrelated, if $\mathbf{E}(v_i|(I - \mathcal{Z}_\kappa)X_i) = 0$ as well as $\mathbf{E}(v_i(s)v_i(t)|(I - \mathcal{Z}_\kappa)X_i) = \mathbf{E}(v_i(s)v_i(t))$ for all, s, t. $\mathcal{Z}_\kappa$-uncorrelatedness is weaker than assuming that X_i and v_i are uncorrelated. In the above $MA(q)$-example X_i and v_i are correlated if δ_i and v_i are correlated. At the same time v_i and X_i are necessarily $\mathcal{Z}_\kappa$-uncorrelated. 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) $\|\beta - \mathbf{E}_\epsilon(\hat{\beta})\| = O_P(b_\beta(n, T))$, where

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

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

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

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

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

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

Furthermore, if $\frac{b_v(n,T)^2}{Tc(T)^2} + \frac{1}{T^3c(T)^3} + Td(t)b_\beta(n,T)^2 = o(\sqrt{tr(\mathcal{Z}_\kappa^2)}/n)$ then

$$\frac{T^{-1}\|g_r - \hat{g}_r + S_r(\Sigma_{n,T} - \mathcal{Z}_\kappa\Sigma_{n,T}\mathcal{Z}_\kappa)g_r\|^2 - \sigma^2\frac{\lambda_r}{n}tr(\mathcal{Z}_\kappa S_r^2\mathcal{Z}_\kappa)}{\sigma^2\frac{\lambda_r}{n}\sqrt{2tr((\mathcal{Z}_\kappa S_r^2\mathcal{Z}_\kappa)^2)}} \to_d \mathbf{N}(0,1) \qquad (19)$$

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

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

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

Furthermore, if $\frac{b_v(n,T)^2}{c(T)^{1/2}} + d(T)^{1/2}(b_\beta(n,T) + \frac{1}{\sqrt{nT}}) + \frac{1}{T^2 c(T)^{3/2}} = o(T^{-1/2})$, 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 $T d(T) b_\beta(n,T)^2 + \frac{d(T)}{n} + \frac{1}{Tc(T)} = o\left(\sqrt{tr(\mathcal{Z}_\kappa^2)/n}\right)$, then

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

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

Let us interpret the results on estimating g_r in terms of the two situations analyzed above. Recall that $c(T) = 1$ (and $d(T) = 1$) in Situation 1. The optimal smoothing parameter to obtain best possible estimates of the *individual* functions $v_i(t)$ is of order $\frac{\kappa}{T^4} = \kappa_T \sim T^{-4/5}$. Then $b_v(n,T) = O_P(T^{-2/5})$, and Theorem 1c) shows that $T^{-1/2}\|g_r - \hat{g}_r\|$ possesses the same rate of convergence. However, different from individual estimates of v_i variance of the estimated functional components $\hat{g}_r$ decrease as n increases. An improvement 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 $b_v(n,T) = O_P((nT)^{-2/5})$, $tr(\mathcal{Z}_\kappa^2) = O(nT)^{1/5}))$, 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, $\|\beta - \hat{\beta}\| = O_P(1/\sqrt{nT})$.

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 $g_1, g_2, \dots$ 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.

In Situation 2 consider the random walk example 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 $\frac{b_v(n,T)}{c(T)^{1/2}} = O(T^{-1/2})$, 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 $\mathcal{Z}_\kappa$-uncorrelated. 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 $b_v(n,T) = O(n^{-\tau})$. It follows from the results of Utreras (1983) that we still have $tr(\mathcal{Z}_\kappa^2) = O(T)$, but $tr((I - \mathcal{Z}_\kappa)) = O(\kappa T)$. On the right hand side of condition (14) in Assumption 4 $O(1/(nT)$ has to be replaced by $O((1/(\kappa nT)$. Theoretical analysis of this setup may follow the lines of the proof of Theorem 1, but some of the arguments have to be adapted to the modified structure of $\mathcal{Z}_\kappa$. It may then be shown that $T^{-1/2}\|g_r - \hat{g}_r\| = O_P(\frac{n^{-\tau}}{T^{1/2}} + \frac{1}{T\sqrt{n}} + \sqrt{\frac{1}{nT}})$ if v_i and X_i are $\mathcal{Z}_\kappa$-uncorrelated, and $T^{-1/2}\|g_r - \hat{g}_r\| = O_P(\frac{n^{-\tau}}{T^{1/2}} + \frac{1}{T} + \sqrt{\frac{1}{nT}})$, else. In both cases the rate of convergence is $n^{-\tau}T^{-1/2} = o(T^{-1/2})$, which shows that undersmoothing may be beneficial even in this situation.

Remark: The question arises whether it is possible to determine the best smoothing parameter for estimating $g_1, g_2, \ldots$ directly from the data. A straightforward approach consists in a "leave-one-individual-out" cross-validation. For a fixed L and $i = 1, \ldots, n$ 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}), $k = 1, \ldots, i-1, i+1, \ldots, n$, $j = 1, \ldots, T$. These estimates $\hat{\beta}_{-i}$ as well as $\hat{g}_{r,-i}$ depend on κ, and one may approximate an optimal smoothing parameter by minimizing

$$\sum_i \sum_t (Y_{it} - \bar{Y}_t - (X_i - \bar{X})\hat{\beta}_{-i} - \sum_{r=1}^{L} \hat{\theta}_{ri}\hat{g}_{r,-i}(t))^2$$

over κ. It seems to be reasonable to expect that this approach works under fairly general conditions, although a precise theoretical analysis is not in the scope of the present paper.

2.3 Dimensionality and model tests

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

$$\hat{\sigma}^2 := \frac{1}{(n-1) \cdot tr(I - \mathcal{Z}_\kappa)^2} \sum_i \|(I - \mathcal{Z}_\kappa)(Y_i - \bar{Y} - (X_i - \bar{X})\hat{\beta})\|^2. \tag{20}$$

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

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

$$\Delta(l) := \frac{n \sum_{r=l+1}^{T} \hat{\lambda}_r \ - (n-1)\hat{\sigma}^2 \cdot tr(\mathcal{Z}_\kappa \hat{\mathcal{P}}_l \mathcal{Z}_\kappa)}{\hat{\sigma}^2 \sqrt{2n \cdot tr((\mathcal{Z}_\kappa \hat{\mathcal{P}}_l \mathcal{Z}_\kappa)^2)}}. \tag{21}$$

Choose $\hat{L}$ as the smallest $l = 1, 2, \ldots$ 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 $Td(T)b_\beta(n,T)^2 + \frac{d(T)}{n} + \frac{1}{Tc(T)} = o\left(\sqrt{tr(\mathcal{Z}_\kappa^2)/n}\right)$. Then,

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

Based on the above theoretical results, our methodology allows to test 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. In this case, if θ_{1i} possesses finite second and fourth moments, then assumptions 1 -2) are fulfilled with $L = 1$, $c(T) = 1$, $g_1(t) \equiv 1$, $v_i(t) \equiv \theta_{1i} \cdot 1$, and $w(t) \equiv \beta_0$. Spline smoothing of constant functions does not produce any bias, and for any reasonable choice of smoothing parameters Assumption 3) holds with $b_v(n,T) = b_w(n,T) = 0$ If X_{it} and ϵ_{it} satisfy Assumption 4 and 5, one may invoke Theorem 1. We then obtain $b_\beta(n,T) = 0$. The additional requirements ensuring the distributional results in Theorem 1c) - 1e) and Theorem 2 are automatically satisfied in this situation provided that $n = o(T^2)$.

Therefore, a test of the null-hypothesis $H_0 : v_i(t) \equiv \theta_{1i} \cdot 1$ (constant individual effects) may proceed as follows:

1) In order to quantify possibly time-varying effects $v_i(t)$ under the alternative, choose some reasonable smoothing parameters and determine nonparametric estimates $\hat{v}_i$, $\hat{w}$ as proposed by the general estimation procedure of Section 3. Then compute a dimension estimate $\hat{L}$ using the method explained above. By Theorem 2, the hypothesis of constant individual effects has to be rejected if $\hat{L} > 1$.

2) Even if $\hat{L} = 1$, the null-model additionally requires that $g_1(t) \equiv 1$. This structural assumption implies that $g_1 = \mathbf{1}$, $\mathbf{1} = (1, \ldots, 1)'$, and $\Sigma_{n,T} = \mathcal{Z}_\kappa \Sigma_{n,T} \mathcal{Z}_\kappa$. A test may be based on relation (19). Note that H_0 leads to $S_1 = \frac{1}{-\lambda_1}(I - \frac{1}{T}\mathbf{1}\mathbf{1}')$. Hence, (19) simplifies to

$$\frac{T^{-1}\|\mathbf{1} - \hat{g}_1\|^2 - \sigma^2 \frac{1}{\lambda_1 n} tr(\mathcal{Z}_\kappa (I - \frac{1}{T}\mathbf{1}\mathbf{1}')\mathcal{Z}_\kappa)}{\sigma^2 \frac{1}{\lambda_1 n}\sqrt{2tr((\mathcal{Z}_\kappa (I - \frac{1}{T}\mathbf{1}\mathbf{1}')\mathcal{Z}_\kappa)^2)}} \rightarrow_d \mathbf{N}(0,1)$$

Under the null-model estimates $\hat{\theta}_{1i}$, $\hat{\beta}_j$ and $\hat{\sigma}^2$ can simply be obtained by least squares. If H_0 is true, $\hat{\lambda}_1 = \frac{T}{n}\sum_{j=1}^n \hat{\theta}_{1i}^2$ then yields a consistent estimate of λ_1 with $\hat{\lambda}_1 = \lambda_1(1 + O_P(\frac{1}{\sqrt{nT}}))$. A sensible test statistics is then given by $Z = \frac{T^{-1}\|\mathbf{1} - \hat{g}_1\|^2 - \hat{\sigma}^2 \frac{1}{\hat{\lambda}_1 n} tr(\mathcal{Z}_\kappa (I - \frac{1}{T}\mathbf{1}\mathbf{1}')\mathcal{Z}_\kappa)}{\hat{\sigma}^2 \frac{1}{\hat{\lambda}_1 n}\sqrt{2tr((\mathcal{Z}_\kappa (I - \frac{1}{T}\mathbf{1}\mathbf{1}')\mathcal{Z}_\kappa)^2)}}$, which is asymptotically standard normal under H_0. The null-hypothesis is rejected if Z is too large.

A similar approach may be used to test more complex parametric models of the form $v_i(t) = \sum_{j=1}^L \vartheta_{ri}\psi_r(t)$ for some pre-specified basis functions ψ_r. A first step then consists in using our methodology to conduct a dimension test. If the estimated dimension turns out to be appropriate, then the assumed structure of the basis functions may be tested by plugging in estimates of the unknown quantities $\lambda_r, \Sigma_{n,T}$ and S_r in (19). Consistent estimates under the null-model may be determined from least squares estimates of the model coefficients.

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):

$$X_{it} = RX_{i,t-1} + \eta_{it}, \text{ where } \eta_{it} \sim \mathbf{N}(0, I_2), \tag{22}$$

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 (22) 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 (2) 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:

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

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 thought 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

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

where $M_Q = I - Q(Q'Q)^{-1}Q'$, $Q = \text{diag}(W_i)$, $i = 1, \ldots, n$, and $W_{it} = [1, t, t^2]$. 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 (13), 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.

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

[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.

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

$$Y_{it} = Y_{it}^{*\prime}\gamma + X_{it}'\beta + u_i(t) + \epsilon_{it}. \tag{23}$$

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 $TE = \exp\{u_i(t) - \max_{j=1,\ldots,N}(u_i(t))\}$ 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 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

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

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

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 7[5].

Table 6 presents parameter estimates from Within, BC, CSSW, and KSS[6]. Table 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. Results for the respective estimators do not indicate any significant scale economies. Ray returns to scale are estimated to be 1.085, 1.045, 0.939, and 1.079 by Within, BC, CSSW, and KSS. Average technical efficiencies for Within, BC, CSSW, and KSS are 0.4383, 0.5921, 0.6189, 0.6056. 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

[5] When we assume $L = 1$ and test the null hypothesis that the individual effect is constant, the test statisitc 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 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.

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 ouyr 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.

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, $v_i = \sum_{r=1}^L \vartheta_{ir} b_r$. Let X_{ij} denote the T-vectors with elements X_{itj}, $t=1,\dots,T$. 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

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

But by Assumptions 2) - 4) we obtain

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

Condition (14) of Assumption 4) then leads to $\|\mathbf{E}_\epsilon(\hat{\beta})-\beta\| = O_P((\frac{b_v(n,T)}{T^{1/2}})$. On the other hand, if v_i and X_i are $\mathcal{Z}_\kappa$-uncorrelated, then

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

and $\|\mathbf{E}_\epsilon(\hat{\beta})-\beta\| = O_P((nT)^{-1/2}\cdot b_v(n,T))$. By Assumptions 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,

$$\hat{\Sigma}_{n,T} = \Sigma_{n,T} + B, \quad B = \frac{1}{n}\sum_i(v_i r_i' + r_i v_i' + r_i r_i'). \tag{24}$$

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

$$\gamma_r - \hat{\gamma}_r = S_r B\gamma_r + R, \quad \text{with} \ \ \|R\| \leq \frac{6\sup_{\|a\|=1} a'B'Ba}{\min_s |\lambda_r - \lambda_s|^2} \tag{25}$$

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

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

$$\sup_{\|a\|=1} \|\frac{1}{n}\sum_i v_i v_i'(I - \mathcal{Z}_\kappa)a\| \leq \sup_{\|a\|=1} \frac{1}{n}\sum_i \sqrt{v_i'v_i} \ |v_i'(I - \mathcal{Z}_\kappa)a| = O_P(Tc(T)^{1/2} b_v(n,T)), \tag{27}$$

$$\begin{aligned}\sup_{\|a\|=1} \|\frac{1}{n}\sum_i (\mathcal{Z}_\kappa\tilde{X}_i(\beta - \hat{\beta}))v_i'a\| &\leq \frac{1}{n}\sum_i |v_i'a|\sqrt{(\beta - \hat{\beta})'\tilde{X}_i'\mathcal{Z}_\kappa^2\tilde{X}_i(\beta - \hat{\beta})} \\ &= O_P\left(Tc(T)^{1/2} d(T)^{1/2}(b_\beta(n,T) + \frac{1}{\sqrt{nT}})\right).\end{aligned} \tag{28}$$

By similar arguments

$$\sup_{\|a\|=1} \|\frac{1}{n}\sum_i v_i(\mathcal{Z}_\kappa\tilde{X}_i(\beta - \hat{\beta}))'a\| = O_P(Tc(T)^{1/2} d(T)^{1/2} b_\beta(n,T)) \tag{29}$$

Obviously, $\mathbf{E}_\epsilon(tr((\frac{1}{n}\sum_i v_i\epsilon_i'\mathcal{Z}_\kappa) \cdot (\frac{1}{n}\sum_i \mathcal{Z}_\kappa\epsilon_i v_i'))) = O(\frac{Tc(T)\cdot tr(\mathcal{Z}_\kappa^2)}{n})$, and $\frac{1}{n}\sum_i v_i\bar{\epsilon}'\mathcal{Z}_\kappa = 0$. Therefore

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

Similarly,

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

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

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

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

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

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

Assumptions 1) and 2) additionally imply that $\frac{1}{\min_s|\lambda_r-\lambda_s|} = O_P(\frac{1}{T\cdot c(T)})$. When combining (25) with (26) - (34) we thus obtain

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

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

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

Let us now consider the remainder term R in (25). Note that all eigenvalues of $\mathcal{Z}_\kappa$ are less or equal to 1, and thus $\sup_{\|a\|=1}a'\mathcal{Z}_\kappa^4a \leq 1$. Relations (26) - (34) then imply

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

By (25), (35), (36) and (37) the asserted rate of convergence follows from

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

Furthermore, it is easily seen that

$$\|S_r\frac{1}{n}\sum_i v_i(\epsilon_i-\bar{\epsilon})'\mathcal{Z}_\kappa\gamma_r\|^2 = \|S_r\frac{1}{n}\sum_i v_i\epsilon_i'\mathcal{Z}_\kappa\gamma_r\|^2 = O_P\left(\frac{1}{Tc(T)n}\right) \quad (39)$$

Therefore, if $\frac{b_v(n,T)^2}{Tc(T)^2} + \frac{1}{T^3c(T)^3} + Td(t)b_\beta(n,T)^2 = o(\sqrt{tr(\mathcal{Z}_\kappa^2)}/n)$, the above relations imply

$$T^{-1}\|g_r - \hat{g}_r + S_r(\Sigma_{n,T} - \mathcal{Z}_\kappa \Sigma_{n,T}\mathcal{Z}_\kappa)g_r\|^2 = \|S_r \frac{1}{n}\sum_i \mathcal{Z}_\kappa(\epsilon_i - \bar{\epsilon})v_i'\gamma_r\|^2 + o_p(\frac{\sqrt{tr(\mathcal{Z}_\kappa^2)}}{nTc(T)})$$

Since $v_i'\gamma_r = \sqrt{T}\theta_{ir}$, it follows that $S_r\frac{1}{n}\sum_i \mathcal{Z}_\kappa(\epsilon_i - \bar{\epsilon})v_i'\gamma_r = S_r\frac{\sqrt{T}}{n}\sum_i \theta_{ir}\mathcal{Z}_\kappa\epsilon_i$. Using $T\frac{1}{n}\sum_i \theta_{ir}^2 = \lambda_r$, some straightforward computations then lead to
$\mathbf{E}_\epsilon(\|S_r\frac{\sqrt{T}}{n}\sum_i \theta_{ir}\mathcal{Z}_\kappa\epsilon_i\|^2) = \sigma^2\frac{\lambda_r}{n}tr(\mathcal{Z}_\kappa S_r^2 \mathcal{Z}_\kappa) = O(\frac{tr(\mathcal{Z}_\kappa^2)}{nTc(T)})$ and
$\mathbf{Var}_\epsilon(\|S_r\frac{\sqrt{T}}{n}\sum_i \theta_{ir}\mathcal{Z}_\kappa\epsilon_i\|^2) = 2\sigma^4\frac{\lambda_r^2}{n^2}tr((\mathcal{Z}_\kappa S_r^2\mathcal{Z}_\kappa)^2)(1+o(1)) = O(\frac{tr(\mathcal{Z}_\kappa^2)}{n^2T^2c(T)^2})$.
The asserted asymptotic distribution then follows from standard arguments.

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

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

Moreover, one can infer from relations (25) - (39) that

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

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,

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

as well as

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

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

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

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 $\frac{b_v(n,T)^2}{c(T)^{1/2}} + d(T)^{1/2}b_\beta(n,T) + \frac{tr(\mathcal{Z}_\kappa^2)}{nT} = o(T^{-1/2})$, the assertion on the multivariate distribution of $\sqrt{T}(\hat{\theta}_{1i} - \theta_{1i}, \ldots, \hat{\theta}_{Li} - \theta_{Li})'$ follows from standard arguments.

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

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

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

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

$\tilde{\Sigma}_n$ possesses only L nonzero eigenvalues $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_L$ with corresponding eigenvectors $\tilde{\gamma}_1, \ldots, \tilde{\gamma}_L$. Our assumptions and arguments similar to (25) - (39) then show that $\tilde{\lambda}_r = O(Tc(T))$, $\frac{1}{\min_s |\tilde{\lambda}_r - \tilde{\lambda}_s|} = O_P(\frac{1}{T \cdot c(T)})$, $\|\gamma_r - \tilde{\gamma}_r\| = O_P(\frac{b_v(n,T)}{c(T)^{1/2}})$, and $\|\hat{\gamma}_r - \tilde{\gamma}_r\| = O_P\left(\frac{d(T)^{1/2} b_\beta(n,T)}{c(T)^{1/2}} + \frac{1}{T^2 c(T)^2} + \sqrt{\frac{tr(\mathcal{Z}_\kappa^2)}{nTc(T)}}\right)$ for all $r, s = 1, \ldots, L$, $r \neq s$.

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

$$\sum_{r=L+1}^{T} \hat{\lambda}_r = tr(\mathcal{P}_L \tilde{B}) + R^*, \quad \text{with } R^* \leq \frac{6L \sup_{\|a\|=1} a' \tilde{B}' \tilde{B} a}{\min_s |\tilde{\lambda}_r - \tilde{\lambda}_s|} \tag{40}$$

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

$$\frac{6L \sup_{\|a\|=1} a' \tilde{B}' \tilde{B} a}{\min_s |\tilde{\lambda}_r - \tilde{\lambda}_s|} = O_P\left(Td(T) b_\beta(n,T)^2 + \frac{1}{Tc(T)} + \frac{tr(\mathcal{Z}_\kappa^2)}{n}\right). \tag{41}$$

On the other hand,

$$tr(\mathcal{P}_L \tilde{B}) = tr\left(\frac{1}{n}\sum_i \mathcal{P}_L \mathcal{Z}_\kappa \tilde{X}_i (\beta - \hat{\beta})(\beta - \hat{\beta})' \tilde{X}_i' \mathcal{Z}_\kappa\right) + tr\left(\mathcal{P}_L \mathcal{Z}_\kappa (\frac{1}{n}\sum_i (\epsilon_i - \bar{\epsilon})(\epsilon_i - \bar{\epsilon})')\mathcal{Z}_\kappa\right) \tag{42}$$

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

$$\frac{\sum_{r=L+1}^{h} \hat{\lambda}_r - \sigma^2 \left(1 - \frac{1}{n}\right) tr(\mathcal{Z}_\kappa \mathcal{P}_L \mathcal{Z}_\kappa)}{\sqrt{\frac{2\sigma^4}{n} \cdot tr((\mathcal{Z}_\kappa \mathcal{P}_L \mathcal{Z}_\kappa)^2)}} \to_d \mathbf{N}(0,1). \tag{43}$$

By (38), Relation (43) remains valid when $\mathcal{P}_L$ is replaced by $\hat{P}_L$. This proves assertion (e). $\square$

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

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

Clearly,

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

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

$$0\le R_{n,T}=\frac{1}{(n-1)\cdot tr((I-\mathcal{Z}_\kappa)^2)}\sum_i v_i'(I-\mathcal{Z}_\kappa)^2 v_i = O_p(b_v(n,T)^2) \tag{44}$$

we obtain

$$\hat{\sigma}^2=\sigma^2+R_{n,T}+o_p(1). \tag{45}$$

Let us now consider the behavior of $\Delta(l)$ for $l<L$. We can immediately infer from (45) that

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

$$\left.+\frac{n\sum_{r=L+1}^{T}\hat{\lambda}_r\;-(n-1)\sigma^2\cdot tr(\mathcal{Z}_\kappa\hat{\mathcal{P}}_L\mathcal{Z}_\kappa)}{\hat{\sigma}^2\sqrt{2n\cdot tr((\mathcal{Z}_\kappa\hat{\mathcal{P}}_l\mathcal{Z}_\kappa)^2)}}\right](1+o_P(1)). \tag{47}$$

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 term on the right hand side of

(46) increases as $n, T \to \infty$, while the first term in (47) is still bounded in probability. We can thus infer that for $l < L$

$$\mathbf{P}(\Delta(l) > z_{1-\alpha}) \to 1 \quad \text{and therefore } \mathbf{P}(\hat{L} \neq l) \to 1 \tag{48}$$

as $n, T \to \infty$.

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

$$\limsup_{n,T \to \infty} \mathbf{P}(\Delta(L) \geq z_{1-\alpha}) \leq \alpha. \tag{49}$$

The assertion of the theorem now is an immediate consequence of (48) and (49). $\square$

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 Reseve 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, Cambrige.*

Mark, N. C. and D. Sul (2003) "Cointergration 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*, Hiedelberg: 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," *Numererical 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 for Small Banks

	Within	BC	CSSW	KSS
CD	-0.0351 (0.0047)	-0.0320 (0.0042)	-0.0099 (0.0032)	-0.0008 (0.0019)
DD	-0.0904 (0.0160)	-0.0351 (0.0135)	-0.0813 (0.0138)	-0.0410 (0.0109)
OD	-0.1525 (0.0097)	-0.1474 (0.0091)	-0.1245 (0.0071)	-0.0440 (0.0200)
lab	-0.1786 (0.0171)	-0.1557 (0.0142)	-0.1508 (0.0146)	-0.1254 (0.0093)
cap	-0.0427 (0.0054)	-0.0502 (0.0048)	-0.0458 (0.0054)	-0.0289 (0.0053)
purf	-0.5855 (0.0215)	-0.6243 (0.0169)	-0.5263 (0.0195)	-0.7598 (0.0268)
ciln	0.1603 (0.0045)	0.1601 (0.0042)	0.1470 (0.0037)	0.1202 (0.0031)
inln	0.3712 (0.0061)	0.3622 (0.0055)	0.3516 (0.0056)	0.3237 (0.0050)
time	0.0145 (0.0009)	0.0016 (0.0013)	-	-
Avg TE	0.4389	0.6011	0.6230	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

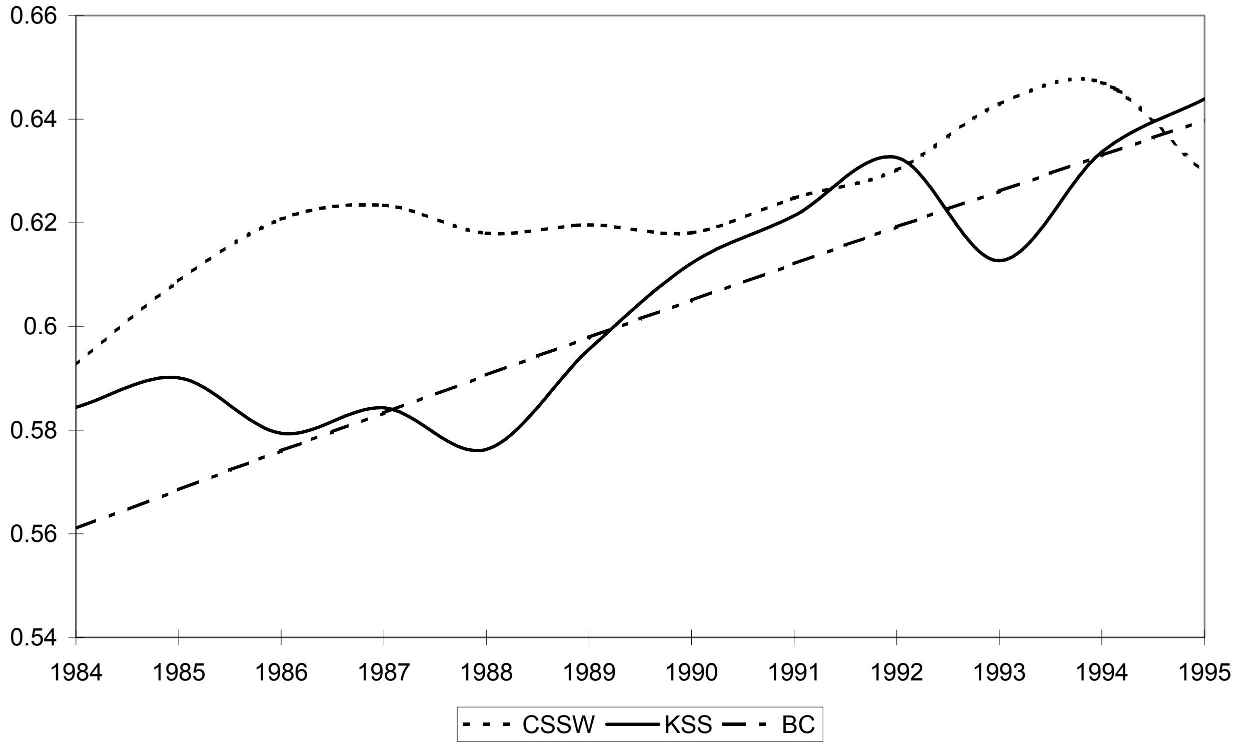

Figure 1:

Figure 1. Temporal Patterns of Efficiencies