Time Series Modelling With Semiparametric Factor Dynamics

Byeong U. PARK, Enno MAMMEN, Wolfgang HÄRDLE, and Szymon BORAK

High-dimensional regression problems, which reveal dynamic behavior, are typically analyzed by time propagation of a few number of factors. The inference on the whole system is then based on the low-dimensional time series analysis. Such high-dimensional problems occur frequently in many different fields of science. In this article we address the problem of inference when the factors and factor loadings are estimated by semiparametric methods. This more flexible modeling approach poses an important question: Is it justified, from an inferential point of view, to base statistical inference on the estimated times series factors? We show that the difference of the inference based on the estimated time series and "true" unobserved time series is asymptotically negligible. Our results justify fitting vector autoregressive processes to the estimated factors, which allows one to study the dynamics of the whole high-dimensional system with a low-dimensional representation. We illustrate the theory with a simulation study. Also, we apply the method to a study of the dynamic behavior of implied volatilities and to the analysis of functional magnetic resonance imaging (fMRI) data.

KEY WORDS: Asymptotic inference; Factor models; Implied volatility surface; Semiparametric models; Vector autoregressive process.

1 INTRODUCTION

Modeling for high-dimensional data is a challenging task in statistics especially when the data comes in a dynamic context and is observed at changing locations with different sample sizes. Such modeling challenges appear in many different fields. Examples are Stock and Watson (2005) in empirical macroeconomics, Lee and Carter (1992) in mortality analysis, Nelson and Siegel (1987) and Diebold and Li (2006) in bond portfolio risk management or derivative pricing, Martinussen and Scheike (2000) in biomedical research. Other examples include the studies of radiation treatment of prostate cancer by Kauermann (2000) and evoked potentials in Electroencephalogram (EEG) analysis by Gasser, Möcks, and Verleger (1983). In financial engineering, it is common to analyze the dynamics of implied volatility surface for risk management. For functional magnetic resonance imaging data (fMRI), one may be interested in analyzing the brain's response over time as well as identifying its activation area, see Worsley et al. (2002).

A successful modeling approach utilizes factor type models, which allow low-dimensional representation of the data. In an orthogonal L-factor model an observable J-dimensional random vector $Y_t = (Y_{t,1}, \ldots, Y_{t,J})^{\mathrm{T}}$ can be represented as

$$Y_{t,j} = m_{0,j} + Z_{t,1}m_{1,j} + \cdots + Z_{t,L}m_{L,j} + \varepsilon_{t,j}, \qquad (1)$$

where $Z_{t,l}$ are common factors, $\epsilon_{t,j}$ are errors or specific factors, and the coefficients $m_{l,j}$ are factor loadings. In most applications, the index $t = 1, \ldots, T$ reflects the time evolution of the whole system, and Y_t can be considered as a multidimensional time series. For a method to identify common factors in this model we refer to Peña and Box (1987). The study of high-dimensional Y_t is then simplified to the modeling of $Z_t = (Z_{t,1}, \ldots, Z_{t,L})^{\mathrm{T}}$, which is a more feasible task when $L \ll J$. The model (1) reduces to a special case of the generalized dynamic factor model considered by Forni, Hallin, Lippi, and Reichlin (2000), Forni and Lippi (2001) and Hallin and Liska (2007), when $Z_{t,l} = a_{l,1}(B)U_{t,1} + \cdots + a_{l,q}(B)U_{t,q}$ where the q-dimensional vector process $U_t = (U_{t,1}, \ldots, U_{t,q})^{\mathrm{T}}$ is an orthonormal white noise and B stands for the lag operator. In this case, the model (1) is expressed as $Y_{t,j} = m_{0,j} + \sum_{k=1}^{q} b_{k,j}(B)U_{t,k} + \varepsilon_{t,j}$, where $b_{k,j}(B) = \sum_{l=1}^{L} a_{l,k}(B)m_{l,j}$.

In a variety of applications, one has explanatory variables $X_{t,j} \in \mathbb{R}^d$ at hand that may influence the factor loadings m_l. An important refinement of the model (1) is to incorporate the existence of observable covariates $X_{t,j}$. The factor loadings are now generalized to functions of $X_{t,j}$, so that the model (1) is generalized to

$$Y_{t,j} = m_0(X_{t,j}) + \sum_{l=1}^{L} Z_{t,l}m_l(X_{t,j}) + \varepsilon_{t,j}, \ 1 \le j \le J_t. \qquad (2)$$

In this model, $Z_{t,l}$ for each l: $1 \le l \le L$ enters into all $Y_{t,j}$ for j such that $m_l(X_{t,j}) \neq 0$. Note that the probability of the event that $m_l(X_{t,j}) = 0$ for some $1 \le j \le J$ equals zero if $m_1(x) = 0$ at countably many points of x and the density f_t of $X_{t,j}$ is supported on an interval with nonempty interior, as we assume at (A2) in Section 5.

The model (2) can be interpreted as a discrete version of the following functional extension of the model (1):

$$Y_t(x) = m_0(x) + \sum_{l=1}^{L} Z_{t,l}m_l(x) + \varepsilon_t(x), \qquad (3)$$

where $\varepsilon_t(\cdot)$ is a mean zero stochastic process, and also regarded as a regression model with embedded time evolution. It is different from varying-coefficient models, such as in Fan, Yao, and Cai (2003) and Yang, Park, Xue, and Härdle (2006), because Z_t is unobservable. Our model also has some similarities to the one considered in Connor and Linton (2007) and Connor, Hagmann, and Linton (2007), which generalized the study of Fama and French (1992) on the common movements of stock price returns. There, the covariates, denoted by $X_{l,j}$, are

Byeong U. Park is Professor, Department of Statistics, Seoul National University Seoul 151-747, Korea (E-mail: *bupark@stats.snu.ac.kr*). Enno Mammen is Professor, Department of Economics, University of Mannheim, 68131 Mannheim, Germany (E-mail: *emammen@rumms.uni-mannheim.de*). Wolfgang Härdle is Professor, Institute for Statistics and Econometrics, Humboldt Universität zu Berlin, D-10178 Berlin, Germany (E-mail: *haerdle@wiwi.hu-berlin.de*). Szymon Borak is Ph.D. Student, Institute for Statistics and Econometrics, Humboldt Universität zu Berlin, D-10178 Berlin, Germany (E-mail: *szymon.borak@gmail.de*). The authors gratefully acknowledge financial support Deutsche Forschungsgemeinschaft and the Sonderforschungsbereich 649 "Ökonomisches Risiko." Byeong U. Park's research was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2005-070-C00021). The authors thank the associate editor and referees for their helpful comments and suggestions.

Journal of the American Statistical Association
March 2009, Vol. 104, No. 485, Theory and Methods
DOI 10.1198/jasa.2009.0105

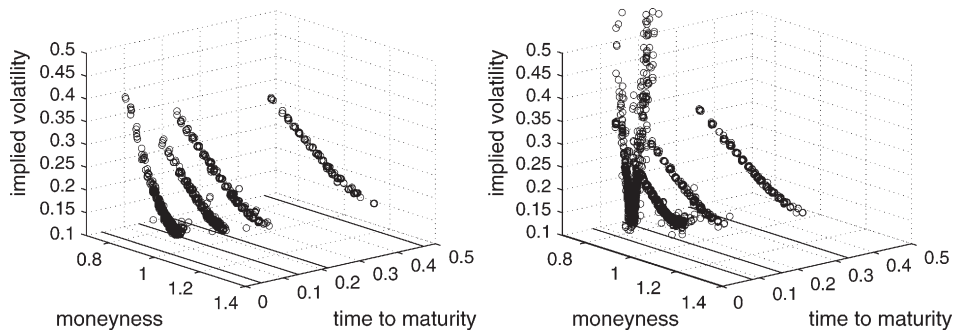

Figure 1. The typical IV data design on two different days. In the maturity direction observations appear in the discrete points for each particular day. Bottom solid lines indicate the observed maturities. Left panel: observations on 2004.07.08, $J_t = 5{,}606$. Right panel: observations on 2004.08.19, $J_t = 8{,}152$.

time-invariant and are different for different m_l, which allows a direct application of backfitting procedures and makes the problem quite different from our setting. Some linear models, which allow time-varying coefficients, as considered in Hansen, Nielsen, and Nielsen (2004) and Brumback and Rice (1998), may be recognized as a special case of (2).

In this article we consider the model (2) with unknown nonparametric functions m_l. We call this model a dynamic semiparametric factor model (DSFM). The evolution of complex high-dimensional objects may be described by (2), so that their analysis can be reduced to the study of a low-dimensional vector of factors Z_t. In the present article, we consider an efficient nonparametric method of fitting the model. We provide relevant theory for the method as well as illustrate its empirical aspects through a simulation and a real data application. Fengler, Härdle, and Mammen (2007) used a kernel smoothing approach for the same model, but it was focused on a particular data application without offering any discussion of numerical issues, statistical theory, and simulation analysis.

One of the main motivations for the model (2) comes from a special structure of the implied volatility (IV) data, as is observed in Figure 1. The IV is a volatility parameter that matches the observed plain vanilla option prices with the theoretical ones given by the formula of Black and Scholes (1973). Figure 1 shows the special "string" structure of the IV data obtained from the European option prices on the German stock index DAX (ODAX) for two different days. The volatility strings shift toward expiry, which is indicated by the bottom line in the figure. Moreover the shape of the IV strings is subject to stochastic deformation. Fengler et al. (2007) proposed to use the model (2) to describe the dynamics of the IV data, where $Y_{t,j}$ are the values of IV or those of its transformation on the day t, and $X_{t,j}$ are the two-dimensional vectors of the moneyness and time-to-maturity. For more details on the data design and econometric motivation, we refer to Fengler et al. (2007).

One may find another application of the model (2) in the analysis of functional magnetic resonance imaging (fMRI) data. The fMRI is a noninvasive technique of recording brain's signals on spatial area in every particular time period (usually 1–4 sec). One obtains a series of three-dimensional images of the blood-oxygen-level-dependent (BOLD) fMRI signals, whereas an exercised person is subject to certain stimuli. An example of the images in 15 different slices at one particular time point is presented in Figure 2. For the more detailed description on the fMRI methodology we refer to Logothetis and Wandell (2004). The main aims of the statistical methods in this field are identification of the brain's activation areas and analysis of its response over time. For this purpose the model (2) can be applied. DSFM may be applied to many other problems, such as modeling of yield curve evolution where the standard approach is to use the parametric factor model proposed by Nelson and Siegel (1987).

Our methods produce estimates of the true unobservable Z_t, say $\hat{Z}_t$, as well as estimates of the unknown functions m_l. In practice, one operates on these estimated values of Z_t for further statistical analysis of the data. In particular, for the IV application, one needs to fit an econometric model to the estimated factors $\hat{Z}_t$. For example, Hafner (2004) and Cont and da Fonseca (2002) fitted an AR(1) process to each factor, and Fengler et al. (2007) considered a multivariate VAR(2) model. The main question that arises from these applications is whether the inference based on $\hat{Z}_t$ is equivalent to the one based on Z_t. Attempting to give an answer to this question forms the core of this article.

It is worthwhile to note here that Z_t is not identifiable in the model (2). There are many versions of (Z_t, m), where $m = (m_0, \ldots, m_L)^{\mathrm{T}}$, that give the same distribution of Y_t. This means that estimates of Z_t and m_l are not uniquely defined. We show that for any version of $\{Z_t\}$ there exists a version of $\{\hat{Z}_t\}$ whose lagged covariances are asymptotically the same as those of $\{Z_t\}$. This justifies the inference based on $\{\hat{Z}_t\}$ when $\{Z_t\}$ is a VAR process, in particular. We confirm this theoretical result by a Monte Carlo simulation study. We also discuss fitting the model to the real ODAX IV and fMRI data.

The article is organized as follows. In the next section we propose a new method of fitting DSFM and an iterative algorithm that converges at a geometric rate. In Section 3 we present the results of a simulation study that illustrate the theoretical findings given in Section 5. In Section 4 we apply the model to the ODAX IV and fMRI data. Section 5 is devoted to the asymptotic analysis of the method. Technical details are provided in the Appendix.

2. METHODOLOGY

We observe $(X_{t,j}, Y_{t,j})$ for $j = 1, \ldots, J_t$ and $t = 1, \ldots, T$ such that

$$Y_{t,j} = \mathcal{Z}_t^{\top} m(X_{t,j}) + \varepsilon_{t,j}. \tag{4}$$

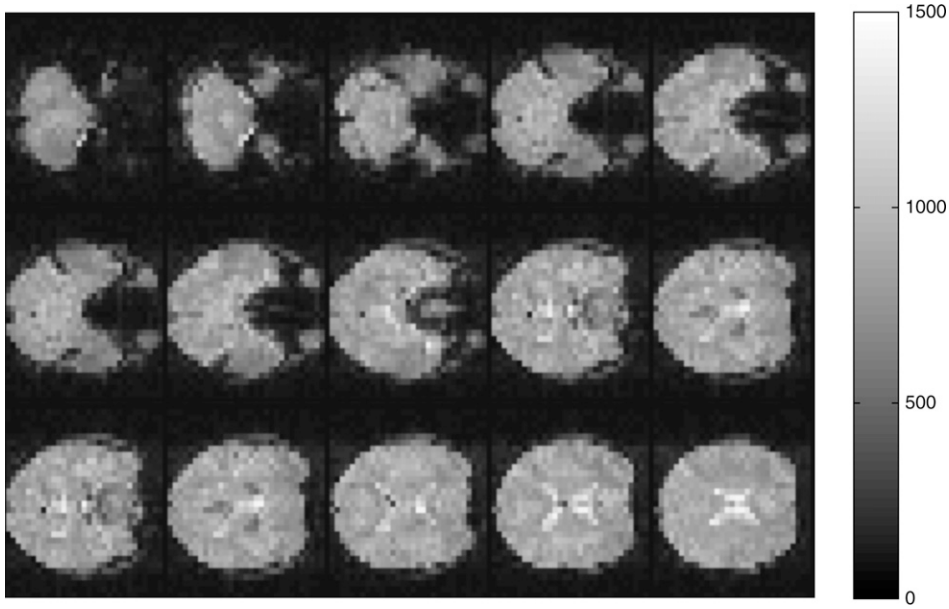

Figure 2. Typical fMRI data in one particular time point. The figure presents 15 parallel horizontal images. The brightness corresponds to the strength of the observed signals.

Here $\mathcal{Z}_t^\top = (1,\ Z_t^\top)$ and $Z_t = (Z_{t,1}, \ldots, Z_{t,L})^\top$ is an unobservable L-dimensional process. The function m is an $(L+1)$-tuple $(m_0, \ldots, m_L)$ of unknown real-valued functions m_l defined on a subset of $\mathbb{R}^d$. The variables $X_{1,1}, \ldots, X_{T,J_T}, \varepsilon_{1,1}, \ldots, \varepsilon_{T,J_T}$ are independent. The errors $\epsilon_{t,j}$ have zero means and finite second moments. For simplicity of notation, we will assume that the covariates $X_{t,j}$ have support $[0, 1]^d$, and also that $J_t \equiv J$ do not depend on t.

For the estimation of m, we use a series estimator. For an integer $K \geq 1$, we choose functions $\psi_1, \ldots, \psi_K$: $[0, 1]^d \to \mathbb{R}$, which are normalized so that $\int_{[0,1]^d} \psi_k^2(x)dx = 1$. For example, one may take $\{\psi_k\colon 1 \leq k \leq K\}$ to be a tensor B-spline basis (e.g., see de Boor 2001). Then, an $(L+1)$-tuple of functions $m = (m_0, \ldots, m_L)^\top$ may be approximated by $\mathcal{A}\psi$, where $\mathcal{A} = (\alpha_{l,k})$ is an $(L+1) \times K$ matrix and $\psi = (\psi_1, \ldots, \psi_K)^\top$. We define the least squares estimators $\widehat{Z}_t = (\widehat{Z}_{t,1}, \ldots, \widehat{Z}_{t,L})^\top$ and $\widehat{\mathcal{A}} = \left(\hat{\alpha}_{l,k}\right)$:

$$S(\mathcal{A}, z) \equiv \sum_{t=1}^{T} \sum_{j=1}^{J} \left\{Y_{t,j} - (1, z_t^\top)\mathcal{A}\psi(X_{t,j})\right\}^2 = \min_{\mathcal{A},z}! \qquad (5)$$

where $z = \left(z_1^\top, \ldots, z_T^\top\right)^\top$ for L-dimensional vectors z_t. With $\widehat{\mathcal{A}}$ at hand, we estimate m by $\hat{m} = \widehat{\mathcal{A}}\psi$.

We note that, given z or $\mathcal{A}$, the function S in (5) is quadratic with respect to the other variables, and thus has an explicit unique minimizer. However, minimization of S with respect to $\mathcal{A}$ and z simultaneously is a fourth-order problem. The solution is neither unique nor explicit. It is unique only up to the values of $\widehat{\mathcal{Z}}_1^\top \widehat{\mathcal{A}}, \ldots, \widehat{\mathcal{Z}}_T^\top \widehat{\mathcal{A}}$, where $\widehat{\mathcal{Z}}_t^\top = (1, \widehat{Z}_t^\top)$. We will come back to this identifiability issue later in this section.

To find a solution $(\widehat{\mathcal{A}}, \widehat{Z})$ of the minimization problem (5), one might adopt the following iterative algorithm: (i) Given an initial choice $Z^{(0)}$, minimize $S(\mathcal{A}, Z^{(0)})$ with respect to $\mathcal{A}$, which is an ordinary least squares problem and thus has an explicit unique solution. Call it $\mathcal{A}^{(1)}$. (ii) Minimize $S(\mathcal{A}^{(1)}, z)$ with respect to z now, which is also an ordinary least squares problem. (iii) Iterate (i) and (ii) until convergence. This is the approach taken by Fengler et al. (2007). However, the procedure is not guaranteed to converge to a solution of the original problem.

We propose to use a Newton-Raphson algorithm. Let $\alpha \equiv \alpha(\mathcal{A})$ denote the stack form of $\mathcal{A} = (\alpha_{l,k})$ [i.e., $\alpha = (\alpha_{0,1}, \ldots, \alpha_{L,1}, \alpha_{0,2}, \ldots, \alpha_{L,2}, \ldots, \alpha_{0,K}, \ldots, \alpha_{L,K})^\top$]. In a slight abuse of notation we write $S(\alpha, z)$ for $S(\mathcal{A}, z)$. Define

$$F_{10}(\alpha, z) = \frac{\partial}{\partial \alpha} S(\alpha, z), \quad F_{01}(\alpha, z) = \frac{\partial}{\partial z} S(\alpha, z),$$
$$F_{20}(\alpha, z) = \frac{\partial^2}{\partial \alpha^2} S(\alpha, z), \quad F_{11}(\alpha, z) = \frac{\partial^2}{\partial \alpha \partial z} S(\alpha, z),$$
$$F_{02}(\alpha, z) = \frac{\partial^2}{\partial z^2} S(\alpha, z).$$

Let $\Psi_t = [\psi(X_{t,1}), \ldots, \psi(X_{t,J})]$ be a $K \times J$ matrix. Define A to be the $L \times K$ matrix obtained by deleting the first row of $\mathcal{A}$. Writing $\zeta_t^\top = (1, z_t^\top)$, it can be shown that

$$F_{10}(\alpha, z) = 2\sum_{t=1}^{T} \left[(\Psi_t \Psi_t^\top) \otimes (\zeta_t \zeta_t^\top)\right] \alpha - 2\sum_{t=1}^{T} (\Psi_t Y_t) \otimes \zeta_t,$$
$$F_{20}(\alpha, z) = 2\sum_{t=1}^{T} \left[(\Psi_t \Psi_t^\top) \otimes (\zeta_t \zeta_t^\top)\right],$$

$F_{01}(\alpha, z)^\top = 2\left(\zeta_1^\top \mathcal{A}\Psi_1\Psi_1^\top A^\top - Y_1^\top \Psi_1^\top A^\top, \ldots, \zeta_T^\top \mathcal{A}\Psi_T\Psi_T^\top A^\top - Y_T^\top \Psi_T^\top A^\top\right)$, and $F_{02}(\alpha, z)$ equals a $(TL) \times (TL)$ matrix that consists of T diagonal blocks $A\Psi_t\Psi_t^\top A^\top$ for $t = 1, \ldots, T$. Here and later, $\otimes$ denotes the Kronecker product operator. Also, by some algebraic manipulations it can be shown that

$$\left[(\Psi_t \Psi_t^\top) \otimes (\zeta_t \zeta_t^\top)\right]\alpha = (\Psi_t \Psi_t^\top \mathcal{A}^\top \zeta_t) \otimes \zeta_t. \qquad (6)$$

Let $\mathcal{I}$ be an $(L+1) \times L$ matrix such that $\mathcal{I}^{\mathrm{T}} = (0, I_L)$ and I_L denote the identity matrix of dimension L. Define $F_{11,t}(\alpha, z) = (\Psi_t \Psi_t^\top A^\top) \otimes \zeta_t + (\Psi_t \Psi_t^\top \mathcal{A}^\top \zeta_t) \otimes \mathcal{I} - (\Psi_t Y_t) \otimes$

$\mathcal{I}$. Then, we get $F_{11}(\alpha, z) = 2\,(F_{11,1}(\alpha, z), F_{11,2}(\alpha, z), \ldots, F_{11,T}(\alpha, z))$. Let

$$F(\alpha, z) = \begin{pmatrix} F_{10}(\alpha, z) \\ F_{01}(\alpha, z) \end{pmatrix}, \; F'(\alpha, z) = \begin{pmatrix} F_{20}(\alpha, z) & F_{11}(\alpha, z) \\ F_{11}(\alpha, z)^\top & F_{02}(\alpha, z) \end{pmatrix}.$$

We need to solve the equation $F(\alpha, z) = 0$ simultaneously for α and z. We note that the matrices $(\Psi_t \Psi_t^\top) \otimes (\zeta_t \zeta_t^\top) = (\Psi_t \otimes \zeta_t)(\Psi_t \otimes \zeta_t)^\top$ and $A \Psi_t \Psi_t^\top A^\top$ are nonnegative definite. Thus, by Miranda's existence theorem (for example, see Vrahatis 1989) the nonlinear system of equations $F(\alpha, z) = 0$ has a solution.

Given $(\alpha^{\mathrm{OLD}}, Z^{\mathrm{OLD}})$, the Newton-Raphson algorithm gives the updating equation for $(\alpha^{\mathrm{NEW}}, Z^{\mathrm{NEW}})$:

$$\begin{pmatrix} \alpha^{\mathrm{NEW}} \\ Z^{\mathrm{NEW}} \end{pmatrix} = \begin{pmatrix} \alpha^{\mathrm{OLD}} \\ Z^{\mathrm{OLD}} \end{pmatrix} - F_*'(\alpha^{\mathrm{OLD}}, Z^{\mathrm{OLD}})^{-1} F(\alpha^{\mathrm{OLD}}, Z^{\mathrm{OLD}}), \tag{7}$$

where $F_*'(\alpha, z)$ for each given (α, z) is the restriction to $\mathcal{F}_*$ of the linear map defined by the matrix $F'(\alpha, z)$ and $\mathcal{F}_*$ is the linear space of values of (α, z) with $\sum_{t=1}^T z_t = 0$ and $\sum_{t=1}^T Z_t^{(0)} (z_t - Z_t^{(0)})^\top = 0$. We denote the initial value of the algorithm by $(\alpha^{(0)}, Z^{(0)})$. We will argue later that under mild conditions, $(\hat{\alpha}, \widehat{Z})$ can be chosen as an element of $\mathcal{F}_*$.

The algorithm (7) is shown to converge to a solution of (5) at a geometric rate under some weak conditions on the initial choice $(\alpha^{(0)}, Z^{(0)})$, as is demonstrated by Theorem 1 later. We collect the conditions for the theorem.

(C1) It holds that $\sum_{t=1}^T Z_t^{(0)} = 0$. The matrix $\sum_{t=1}^T Z_t^{(0)} Z_t^{(0)\top}$ and the map $F_*'(\alpha^{(0)}, Z^{(0)})$ are invertible.
(C2) There exists a version $(\hat{\alpha}, \widehat{Z})$ with $\sum_{t=1}^T \widehat{Z}_t = 0$ such that $\sum_{t=1}^T \widehat{Z}_t Z_t^{(0)\top}$ is invertible. Also, $\hat{\alpha}_l = (\hat{\alpha}_{l1}, \ldots, \hat{\alpha}_{lK})^\top$ for $l = 0, \ldots, L$ are linearly independent.

Let $\alpha^{(k)}$ and $Z^{(k)}$ denote the kth updated vectors in the iteration with the algorithm (7). Also, we write $\mathcal{A}^{(k)}$ for the matrix that corresponds to $\alpha^{(k)}$, and $\mathcal{Z}_t^{(k)\top} = (1, Z_t^{(k)\top})$.

Theorem 1. Let T, J and K be held fixed. Suppose that the initial choice $(\alpha^{(0)}, Z^{(0)})$ satisfies (C1) and (C2). Then, for any constant $0 < \gamma < 1$ there exist $r > 0$ and $C > 0$, which are random variables depending on $\{(X_{t,j}, Y_{t,j})\}$, such that, if $\sum_{t=1}^T \| \mathcal{Z}_t^{(0)\top} \mathcal{A}^{(0)} - \widehat{\mathcal{Z}}_t^\top \widehat{\mathcal{A}} \|^2 \le r$, then

$$\sum_{t=1}^T \| \mathcal{Z}_t^{(k)\top} \mathcal{A}^{(k)} - \widehat{\mathcal{Z}}_t^\top \widehat{\mathcal{A}} \|^2 \le C 2^{-2(k-1)} \gamma^{2(2^k-1)}.$$

We now argue that under (C1) and (C2), $(\hat{\alpha}, \widehat{Z})$ can be chosen as an element of $\mathcal{F}_*$. Note first that one can always take $Z_t^{(0)}$ and $\widehat{Z}_t$ so that $\sum_{t=1}^T Z_t^{(0)} = 0$ and $\sum_{t=1}^T \widehat{Z}_t = 0$. This is because, for any version $(\hat{\alpha}, \widehat{Z})$, one has

$$\begin{aligned} \widehat{\mathcal{Z}}_t^\top \widehat{\mathcal{A}} = \hat{\alpha}_0^\top + \sum_{l=1}^L \widehat{Z}_{t,l} \hat{\alpha}_l^\top &= \left(\hat{\alpha}_0^\top + \sum_{l=1}^L \overline{\widehat{Z}}_l \hat{\alpha}_l^\top \right) \\ &+ \sum_{l=1}^L (\widehat{Z}_{t,l} - \overline{\widehat{Z}}_l) \hat{\alpha}_l^\top \overset{\text{let}}{=} \hat{\alpha}_0^{*\top} + \sum_{l=1}^L \widehat{Z}_{t,l}^* \hat{\alpha}_l^\top = \widehat{\mathcal{Z}}_t^{*\top} \widehat{\mathcal{A}}^*, \end{aligned}$$

where $\overline{\widehat{Z}}_l = T^{-1} \sum_{t=1}^T \widehat{Z}_{t,l}$, $\widehat{\mathcal{Z}}_t^{*\top} = (1, \widehat{Z}_t^{*\top})$ and $\widehat{\mathcal{A}}^*$ is the matrix obtained from $\widehat{\mathcal{A}}$ by replacing its first row by $\hat{\alpha}_0^{*\top}$. Furthermore, the minimization problem (5) has no unique solution. If $(\widehat{Z}_t, \widehat{\mathcal{A}})$ or $(\widehat{Z}_t, \hat{m} = \widehat{\mathcal{A}}\psi)$ is a minimizer, then also $(B^\top \widehat{Z}_t, \widetilde{B}^{-1} \hat{m})$ is a minimizer. Here

$$\widetilde{B} = \begin{pmatrix} 1 & 0 \\ 0 & B \end{pmatrix} \tag{8}$$

and B is an arbitrary invertible matrix. The special structure of $\widetilde{B}$ assures that the first component of $\widetilde{B}^\top \widehat{\mathcal{Z}}_t$ equals 1. In particular, with the choice $B = (\sum_{t=1}^T Z_t^{(0)} \widehat{Z}_t^\top)^{-1} \sum_{t=1}^T Z_t^{(0)} Z_t^{(0)\top}$ we get for $\widehat{Z}_t^* = B^\top \widehat{Z}_t$ that $\sum_{t=1}^T Z_t^{(0)} (\widehat{Z}_t^* - Z_t^{(0)})^\top = 0$.

In Section 5, we will show that, for any solution $\widehat{Z}_t$ and for any version of true Z_t, there exists a random matrix B such that $\widetilde{Z}_t = B^\top \widehat{Z}_t$ has asymptotically the same covariance structure as Z_t. This means that the difference of the inferences based on $\widetilde{Z}_t$ and Z_t is asymptotically negligible.

We also note that one can always choose $\hat{m} = \widehat{\mathcal{A}}\psi$ such that the components $\hat{m}_1, \ldots, \hat{m}_L$ are orthonormal in $L_2([0, 1]^d)$ or in other L_2 [e.g., in $L_2(T^{-1} \sum_{t=1}^T \widehat{f}_t)$ where $\widehat{f}_t$ is a kernel estimate of the density of $X_{t,j}$]. If one selects $\hat{m}$ in this way, then the matrix B should be an orthogonal matrix and the underlying time series Z_t is estimated up to such transformations.

In practice one needs to choose an initial estimate $(\alpha^{(0)}, Z^{(0)})$ to run the algorithm. One may generate normal random variates for $Z_{t,l}^{(0)}$, and then find the initial $\alpha^{(0)}$ by solving the equation $F_{10}(\alpha, Z^{(0)})$. This initial choice was found to work well in our numerical study presented in Sections 3 and 4.

As an alternative way of fitting the model (2), one may extend the idea of the principal component method that is used to fit the orthogonal factor model (1). In this way, the data $\{Y_{t,j}: 1 \le j \le J\}$ are viewed as the values of a functional datum $Y_t(\cdot)$ observed at $x = X_{t,j}$, $1 \le j \le J$, and the functional factor model given at (3) may be fitted with smooth approximations of Y_t obtained from the original dataset. If one assumes $EZ_t = 0$, $\mathrm{var}(Z_t) = I_L$, as is typically the case with the orthogonal factor model (1), then one can estimate m_l and Z_t by performing functional principal component analysis with the sample covariance function

$$\widehat{K}(x, x') = T^{-1} \sum_{t=1}^T \{Y_t(x) - \overline{Y}(x)\}\{Y_t(x') - \overline{Y}(x')\},$$

where $\overline{Y}(x) = T^{-1} \sum_{t=1}^T Y_t(x)$. There are some limitations for this approach. First, it requires initial fits to get smooth approximations of $Y_t(\cdot)$, which may be difficult when the design points $X_{t,j}$ are sparse as is the case with the IV application. Our method avoids the preliminary estimation and shifts the discrete representation directly to the functions m_l. Second, for the method to work one needs at least stationarity of Z_t and ε_t, whereas our theory does not rely on these assumptions.

3. SIMULATION STUDY

In Theorem 3 we will argue that the inference based on the covariances of the unobserved factors Z_t is asymptotically equivalent to the one based on $B^\top \widehat{Z}_t$ for some invertible B. In this section we illustrate the equivalence by a simulation study. We compare the covariances of Z_t and $\widetilde{Z}_t \equiv B^\top \widehat{Z}_t$, where

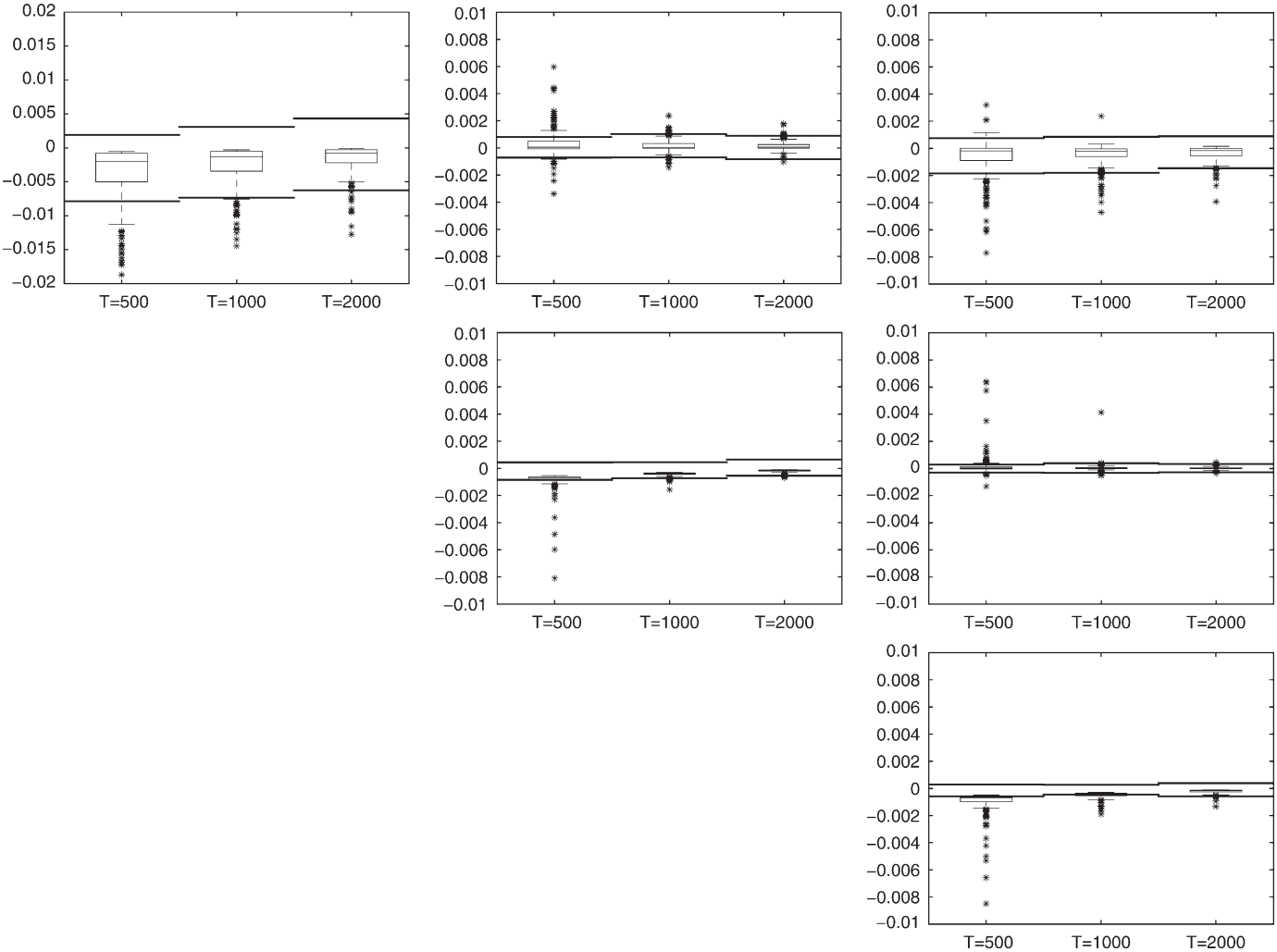

Figure 3. The boxplots based on 250 values of the entries of the scaled difference of the covariance matrices given at (10). The lengths of the series Z_t and $\tilde{Z}_t$ were 500, 1,000, 2,000. The thick lines represent the upper and lower quartiles of (11).

$$B = \left(T^{-1}\sum_{t=1}^{T} Z_{c,t}\widehat{Z}_{c,t}^{\top}\right)^{-1} T^{-1}\sum_{t=1}^{T} Z_{c,t}Z_{c,t}^{\top}, \tag{9}$$

$Z_{c,t} = Z_t - T^{-1}\sum_{s=1}^{T} Z_s$ and $\widehat{Z}_{c,t} = \widehat{Z}_t - T^{-1}\sum_{s=1}^{T}\widehat{Z}_s$. Note that B at (9) minimizes $\sum_{t=1}^{T} \| \widehat{Z}_{c,t} - (B^{\top})^{-1}Z_{c,t} \|^2$. In the Appendix we will prove that Theorem 3 holds with the choice at (9).

We took $T = 500, 1{,}000, 2{,}000$, $J = 100, 250, 1{,}000$ and $K = 36, 49, 64$. We considered $d = 2$, $L = 3$ and the following tuple of 2-dimensional functions:

$$\begin{aligned} &m_0(x_1, x_2) = 1, \quad m_1(x_1, x_2) = 3.46(x_1 - .5), \\ &m_2(x_1, x_2) = 9.45\left\{(x_1 - .5)^2 + (x_2 - .5)^2\right\} - 1.6, \\ &m_3(x_1, x_2) = 1.41\sin(2\pi x_2). \end{aligned}$$

The coefficients in these functions were chosen so that m_1, m_2, m_3 are close to orthogonal. We generated Z_t from a centered VAR(1) process $Z_t = \mathcal{R}Z_{t-1} + U_t$, where U_t is $N_3(0, \Sigma_U)$ random vector, the rows of $\mathcal{R}$ from the top equal $(0.95, -0.2, 0)$, $(0, 0.8, 0.1)$, $(0.1, 0, 0.6)$, and $\Sigma_U = 10^{-4}I_3$. The design points $X_{t,j}$ were independently generated from a uniform distribution on the unit square, $\varepsilon_{t,j}$ were iid $N(0, \sigma^2)$ with $\sigma = 0.05$, and $Y_{t,j}$ were obtained according to the model (4). The simulation experiment was repeated 250 times for each combination of (T, J, K). For the estimation we employed, for ψ_j, the tensor products of linear B-splines. The one-dimensional linear B-splines $\widetilde{\psi}_k$ are defined on a consecutive equidistant knots x^k, x^{k+1}, x^{k+2} by $\widetilde{\psi}_k(x) = (x - x^k)/(x^{k+1} - x^k)$ for $x \in (x^k, x^{k+1}]$, $\widetilde{\psi}_k(x) = (x^{k+2} - x)/(x^{k+2} - x^{k+1})$ for $x \in (x^{k+1}, x^{k+2}]$, and $\widetilde{\psi}_k(x) = 0$ otherwise. We chose $K = 8 \times 8 = 64$.

We plotted in Figure 3 the entries of the scaled difference of the covariance matrices

$$\widetilde{D} = \frac{1}{\sqrt{T}}\left\{\sum_{t=1}^{T}\left(\widetilde{Z}_t - \overline{\widetilde{Z}}\right)\left(\widetilde{Z}_t - \overline{\widetilde{Z}}\right)^{\top} - \sum_{t=1}^{T}\left(Z_t - \overline{Z}\right)\left(Z_t - \overline{Z}\right)^{\top}\right\}. \tag{10}$$

Each panel of Figure 3 corresponds to one entry of the matrix $\widetilde{D}$, and the three boxplots in each panel represent the distributions of the 250 values of the corresponding entry for $T = 500, 1{,}000, 2{,}000$. In the figure we also depicted, by thick lines, the upper and lower quartiles of

$$D = \frac{1}{\sqrt{T}}\left\{\sum_{t=1}^{T}\left(Z_t - \overline{Z}\right)\left(Z_t - \overline{Z}\right)^{\top} - T\Gamma\right\}, \tag{11}$$

where Γ is the true covariance matrix of the simulated VAR process. We refer to Lütkepohl (1993) for a representation of Γ.

Our theory in Section 5 tells that the size of $\widetilde{D}$ is of smaller order than the normalized error D of the covariance estimator based on Z_t. It is known that the latter converges to a nondegenerate law as $T \to \infty$. This is well supported by the plots in Figure 3 showing that the distance between the two thick lines in each panel is almost invariant as T increases. The fact that the additional error incurred by using $\widetilde{Z}_t$ instead of Z_t is negligible for large T is also confirmed. In particular, the long stretches at tails of the distributions of $\widetilde{D}$ get shorter as T increases. Also, the upper and lower quartiles of each entry of $\widetilde{D}$, represented by the boxes, lie within those of the corresponding entry of D, represented by the thick lines, when $T =$ 1,000 and 2,000.

4. APPLICATIONS

This section presents an application of DSFM. We fit the model to the intraday IV based on ODAX prices and to fMRI data.

For our analysis we chose the data observed from July 1, 2004 to June 29, 2005. The one year period corresponds to the financial regulatory requirements. The data were taken from Financial and Economic Data Center of Humboldt-Universität zu Berlin. The IV data were regressed on the two-dimensional space of future moneyness and time-to-maturity, denoted by $(\kappa_t, \tau_t)^\top$. The future moneyness κ_t is a monotone function of the strike price K: $\kappa_t = K/(S_t e^{-r_t \tau_t})$, where S_t is the spot price at time t and r_t is the interest rate. We chose r_t as a daily Euro Interbank Offered Rate (EURIBOR) taken from the Ecowin Reuters database. The time-to-maturity of the options were measured in years. We took all trades with $10/365 < \tau < 0.5$. We limit also the moneyness range to $\kappa \in [0.7, 1.2]$.

The structure of the IV data, described already in Section 1, requires a careful treatment. Apart from the dynamic degeneration, one may also observe nonuniform frequency of the trades with significantly greater market activities for the options closer to expiry or at-the-money. Here, "at-the-money" means a condition in which the strike price of an option equals the spot price of the underlying security (i.e., $K = S_t$). To avoid the computational problems with the highly skewed empirical distribution of $X_t = (\kappa_t, \tau_t)$, we transformed the initial space $[0.7, 1.2] \times [0.03, 0.5]$ to $[0, 1]^2$ by using the marginal empirical distribution functions. We applied the estimation algorithm to the transformed space, and then transformed back the results to the original space.

Because the model is not nested, the number of the dynamic functions needs to be determined in advance. For this, we used

$$RV(L) = \frac{\sum_t^T \sum_j^{J_t} \left\{ Y_{t,j} - \hat{m}_0(X_{t,j}) - \sum_{l=1}^L \widehat{Z}_{t,l} \hat{m}_l(X_{t,j}) \right\}^2}{\sum_t^T \sum_j^{J_t} \left(Y_{t,j} - \overline{Y} \right)^2}, \tag{12}$$

although one may construct an Akaike information (AIC) or Bayesian information (BIC) type of criterion, where one penalizes the number of the dynamic functions in the model, or performs some type of cross-validation. The quantity $1 - RV(L)$ can be interpreted as a proportion of the variation explained by the model among the total variation. The computed values of $RV(L)$ are given in Table 1 for various L. Because the third, fourth, and fifth factor made only a small improvement in the fit, we chose $L = 2$.

Table 1. Proportion of the explained variation by the models with L = 1, . . ., 5 dynamic factors

No. factors	$L = 1$	$L = 2$	$L = 3$	$L = 4$	$L = 5$
$1 - RV(L)$	0.848	0.969	0.976	0.978	0.980

For the series estimators of $\hat{m}_l$ we used tensor B-splines that are cubic in the moneyness and quadratic in the maturity direction. In the transformed space we placed 10×5 knots, 10 in the moneyness and 5 in the maturity direction. We found that the results were not sensitive to the choice of the number of knots and the orders of splines. For several choices of knots in the range 5×5–15×10 and for the spline orders (2, 1), (2, 2), (3, 2), the values of $1 - RV(2)$ were between 0.949 and 0.974. Because the model is identifiable only up to the transformation (8), one has a freedom for the choice of factors. Here, we chose the approach taken by Fengler et al. (2007) with $L_2[0,1]^2$ norm. Specifically, we orthonormalized $\hat{m}_l$ and transformed $\widehat{Z}_t$ according to their Equation (19) with $\Gamma = \int \hat{m}(x)\hat{m}(x)^\top \, dx$, where $\hat{m} = (\hat{m}_1, \ldots, \hat{m}_L)^\top$. Call them $\hat{m}_l^*$ and $\widehat{Z}_t^*$, respectively. Then, we transformed them further by $\hat{m}_l^{**} = p_l^\top \hat{m}^*$ and $\widehat{Z}_{t,l}^{**} = p_l^\top \widehat{Z}_t^*$, where p_l were the orthonormal eigenvectors of the matrix $\sum_{t=1}^T \widehat{Z}_t^* \widehat{Z}_t^{*\top}$ that correspond to the eigenvalues $\lambda_1 > \lambda_2$. Note that $\widehat{Z}_t^{*\top} \hat{m}^* = \widehat{Z}_t^{**\top} \hat{m}^{**}$. In this way, $\{\widehat{Z}_{t,1}^{**} \hat{m}_1^{**}\}$ makes a larger contribution than $\{\widehat{Z}_{t,2}^{**} \hat{m}_2^{**}\}$ to the total variation $\sum_{t=1}^T \int (\widehat{Z}_t^{**\top} \hat{m}^{**})^2$ because $\sum_{t=1}^T \int (\widehat{Z}_{t,1}^{**} \hat{m}_1^{**})^2 = \lambda_1$ and $\sum_{t=1}^T \int (\widehat{Z}_t^{**\top} \hat{m}^{**})^2 = \lambda_1 + \lambda_2$. Later, we continue to write $\widehat{Z}_t$ and $\hat{m}$ for such $\widehat{Z}_t^{**}$ and $\hat{m}^{**}$, respectively.

The estimated functions $\hat{m}_1$ and $\hat{m}_2$ are plotted in Figure 4 in the transformed estimation space. The intercept function $\hat{m}_0$ was almost flat around zero, thus is not given. By construction, $\hat{m}_0 + \widehat{Z}_{t,1} \hat{m}_1$ explain the principal movements of the surface. It was observed by Cont and da Fonseca (2002) and Fengler et al. (2007) that most dominant innovations of the entire surface are parallel level shifts. Note that VDAX is an estimated at-the-money IV for an option with 45 days to maturity, and thus indicates up-and-down shifts. The left panel of Figure 5 shows the values of VDAX together with $\hat{m}_0(X_{t,0}) + \widehat{Z}_{t,1} \hat{m}_1(X_{t,0})$, where $X_{t,0}$ is the moneyness and maturity corresponding to an option at-the-money with 45 days to maturity. The right panel of Figure 5 depicts the factor $\widehat{Z}_t$, where one can find that $\widehat{Z}_t$ shows almost the same dynamic behavior as the index VDAX. This similarity supports that DSFM catches leading dynamic effects successfully. Obviously the model in its full setting explains other effects, such as skew or term structure changes, which are not explicitly stated here.

Statistical analysis on the evolution of a high-dimensional system ruling the option prices can be simplified to a low-dimensional analysis of the $\widehat{Z}_t$. In particular, as our theory in Section 5 and the simulation results in Section 3 assert, the inference based on the $\widehat{Z}_t$ is well justified in the VAR context. To select a VAR model we computed the Schwarz (SC), the Hannan-Quinn (HQ), and the Akaike criterion, as given in

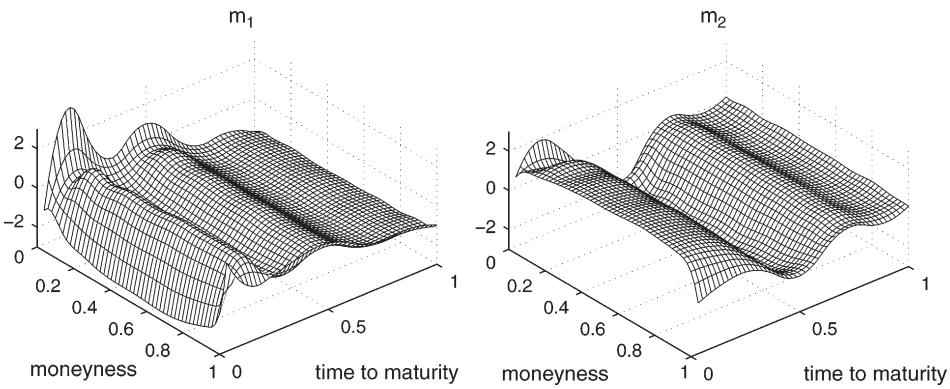

Figure 4. The estimated factor functions for the ODAX IV data in the period 20040701–20050629.

Table 2. One can find that SC and HQ suggest a VAR(1) process, whereas AIC selects VAR(2). The parameter estimates for each selected model are given in Table 3. The roots of the characteristic polynomial lie inside the unit circle, so the specified models satisfy the stationarity condition. For each of VAR(1) and VAR(2) models, we conducted a portmanteau test for the hypothesis that the autocorrelations of the error term at lags up to 12 are all zero, and also a series of LM tests, each of which tests whether the autocorrelation at a particular lag up to 5 equals zero. Some details on selection of lags for these tests can be found in Hosking (1980, 1981) and Brüggemann, Lütkepohl, and Saikkonen (2006). We found that in any test the null hypothesis was not rejected at 5% level. A closer inspection on the autocorrelations of the residuals, however, revealed that the autocorrelation of $\widehat{Z}_{t,2}$ residuals at lag one is slightly significant in the VAR(1) model, see Figure 6. But, this effect disappears in the VAR(2) case, see Figure 7. Similar analyses of characteristic polynomials, portmanteau and Lagrange multiplier (LM) tests supported VAR(2) as a successful model for $\widehat{Z}_t$.

As a second application of the model, we considered fitting an fMRI dataset. The data were obtained at Max-Planck Institut für Kognitions-und-Neurowissenschaften Leipzig by scanning a subject's brain using a standard head coil. The scanning was done every two seconds on the resolution of $3 \times 3 \times 2$ mm^3 with 1 mm gap between the slices. During the experiment, the subject was exposed to three types of objects (bench, phone and motorbike) and rotated around randomly changing axes for four seconds, followed by relaxation phase of six to ten seconds. Each stimulus was shown 16 times in pseudo-randomized order. As a result, a series of 290 images with $64 \times 64 \times 30$ voxels was obtained.

To apply the model (2) to the fMRI data, we took the voxel's index (i_1, i_2, i_3) as covariate $X_{t,j}$, and the BOLD signal as $Y_{t,j}$. For numerical tractability we reduced the original data to a series of $32 \times 32 \times 15$ voxels by taking every second slice in each direction. Thus, $J_t \equiv 32 \times 32 \times 15$ and $T = 290$. The voxels' indices (i_1, i_2, i_3) for $1 \le i_1, i_2 \le 32$;$1 \le i_3 \le 15$ are associated with $32 \times 32 \times 15$ equidistant points in $\mathbb{R}^3$. The function m_0 represents the "average" signal as a function of the three-dimensional location, and m_l for each $l \ge 1$ determines the effect of the lth common factor $Z_{t,l}$ on the brain's signal. In Figure 8, each estimated function $\hat{m}_l$ is represented by its sections on the 15 slices in the direction of i_3 [i.e., by those $\hat{m}_l(\cdot, \cdot, x_3)$ for which x_3 are fixed at the equidistant points corresponding to $i_3 = 1, \ldots, 15$]. We used quadratic tensor B-splines on equidistant knots. The number of knots in each direction was 8, 8, 4, respectively, so that $K = 9 \times 9 \times 5 = 405$. For the model identification we used the same method as in the IV application, but normalized $\widehat{Z}$ to have mean zero.

In contrast to the IV application, there was no significant difference between the values of $1 - RV(L)$ for different $L \ge 1$.

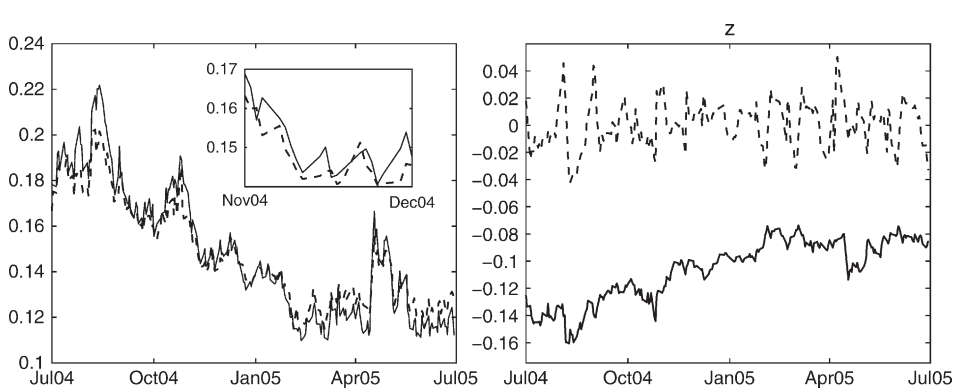

Figure 5. Left panel: VDAX in the period 20040701–20050629 (solid) and the dynamics of the corresponding IV given by the submodel $\hat{m}_0 + \widehat{Z}_{t,1}\hat{m}_1$ (dashed). Right panel: The obtained time series $\widehat{Z}_t$ on the ODAX IV data in the period 20040701–20050629. The solid line represents $\widehat{Z}_{t,1}$, the dashed line $\widehat{Z}_{t,2}$.

Table 2. The VAR model selection criteria. The smallest value for each criterion is marked by an asterisk

Order	AIC	SC	HQ
1	−14.06	−13.98*	−14.03*
2	−14.07*	−13.93	−14.02
3	−14.06	−13.86	−13.98
4	−14.06	−13.81	−13.96
5	−14.07	−13.76	−13.95

All the values for $L \geq 1$ were around 0.871. The fMRI signals $Y_{t,j}$ were explained mostly by $\hat{m}_0(X_{t,j}) + Z_{t,1}\hat{m}_1(X_{t,j})$, and the effects of the common factors $Z_{t,l}$ for $l \geq 2$ were relatively small. The slow increase in the value of $1 - RV(L)$ as $L \geq 1$ grows in the fMRI application, contrary to the case of the IV application, can be explained partly by the high complexity of human brain. Because the values of $1 - RV(L)$ were similar for $L \geq 1$, one might choose $L = 1$. However, we chose $L = 4$, which we think still allows relatively low complexity, to demonstrate some further analysis that might be possible with similar datasets. The estimated functions $\hat{m}_l$ for $0 \leq l \leq 4$ and the time series $\widehat{Z}_{t,l}$ for $1 \leq l \leq 4$ are plotted in Figures 8 and 9, respectively. The function $\hat{m}_0$ can be recognized as a smoothed version of the original signal. By construction the first factor and loadings incorporate the largest variation. One may see the strong positive trend in $\widehat{Z}_{t,1}$ and relatively flat patterns of $\widehat{Z}_{t,2}, \widehat{Z}_{t,3}, \widehat{Z}_{t,4}$. These effects could be typically explained by the mixture of several components, such as physiological pulsation, subtle head movement, machine noise, and so on. For a description of different artifacts, which significantly influence the fMRI signals, we refer to Biswal, Yetkin, Haughton, and Hyde (1995). The function estimates $\hat{m}_l$ for $1 \leq l \leq 4$ appear to have a clear peak, and $\widehat{Z}_{t,l}$ for $2 \leq l \leq 4$ show rather mild mean reverting behavior.

To see how the recovered signals interact with the given stimuli, we plotted $\widehat{Z}_{t+s,l} - \widehat{Z}_{s,l}$ against t in Figure 10, where s is the time when a stimulus appears. The mean changes of $\widehat{Z}_{t,1}$ and $\widehat{Z}_{t,3}$ show mild similarity, up to sign change, to the hemodynamic response (see Worsley et al. 2002). The case of $\widehat{Z}_{t,4}$ has a similar pattern as those of $\widehat{Z}_{t,1}$ and $\widehat{Z}_{t,3}$ but with larger amplitude, whereas the changes in $\widehat{Z}_{t,2}$ seem to be independent of the stimuli. In fitting the fMRI data, we did not use any external information on the signal. From the biological perspective it could be hardly expected that a pure statistical procedure gives full insight into understanding of the complex dynamics of MR images. For the latter one needs to incorporate into the procedure the shape of hemodynamic response, for example, or consider physiologically motivated identification of the factors. It goes however beyond the scope of this illustrative example.

5. ASYMPTOTIC ANALYSIS

In the simulation study and the real data application in Sections 3 and 4, we considered the case where Z_t is a VAR-process. Here, we only make some weak assumptions on the average behavior of the process. In our first theorem we allow that it is a deterministic sequence. In our second result we assume that it is a mixing sequence. For the asymptotic analysis, we let $K, J, T \to \infty$. This is a very natural assumption often also made in cross-sectional or panel data analysis. It is appropriate for data with many observations per data point that are available for many dates. It allows us to study how J and T have to grow with respect to each other for a good performance of a procedure. The distance between m and its best approximation $\mathcal{A}\psi$ does not tend to zero unless $K \to \infty$, see Assumption (A5) later. One needs to let J $\to \infty$ to get consistency of $\widehat{\mathcal{Z}}_t^\top \widehat{\mathcal{A}}$ and $\hat{m} = \widehat{\mathcal{A}}\psi$ as estimates of $\mathcal{Z}_t^\top \mathcal{A}^*$ and m, respectively, where $\mathcal{A}^*$ is defined at (A5). One should let $T \to \infty$ to describe the asymptotic equivalence between the lagged covariances of Z_t and those of $\widetilde{Z}_t$, see Theorem 3 below. In our analysis the dimension L is fixed. Clearly, one could also study our model with L growing to infinity. We treat the case where X_{it} are random. However, a theory for deterministic designs can be developed along the lines of our theory.

Our first result relies on the following assumptions.

(A1) The variables $X_{1,1}, \ldots, X_{T,J}, \varepsilon_{1,1}, \ldots, \varepsilon_{T,J}$, and $Z_1, \ldots, Z_T$ are independent. The process Z_t is allowed to be nonrandom.
(A2) For $t = 1, \ldots, T$ the variables $X_{t,1}, \ldots, X_{t,J}$ are identically distributed, have support $[0, 1]^d$ and a density f_t that is bounded from below and above on $[0, 1]^d$, uniformly over $t = 1, \ldots, T$.
(A3) We assume that $E\varepsilon_{t,j} = 0$ for $1 \leq t \leq T$, $1 \leq j \leq J$, and for $c > 0$ small enough $\sup_{1\leq t\leq T,\ 1\leq j\leq J} E\exp(c\varepsilon_{t,j}^2) < \infty$.
(A4) The functions ψ_k may depend on the increasing indices T and J, but are normed so that $\int_{[0,1]^d} \psi_k^2(x)dx = 1$ for $k = 1, \ldots, K$. Furthermore, it holds that $\sup_{x\in[0,1]} \| \psi(x) \| = \mathcal{O}(K^{1/2})$.
(A5) The vector of functions $m = (m_0, \ldots, m_L)^\top$ can be approximated by ψ_k, i.e.,

$$\delta_K \equiv \sup_{x\in[0,1]^d} \inf_{\mathcal{A}\in\mathbb{R}^{(L+1)\times K}} \|m(x) - \mathcal{A}\psi(x)\| \to 0$$

as $K \to \infty$. We denote $\mathcal{A}$ that fulfills $\sup_{x\in[0,1]^d}\|m(x) - \mathcal{A}\psi(x)\| \leq 2\delta_K$ by $\mathcal{A}^*$.
(A6) There exist constants $0 < C_L < C_U < \infty$ such that all eigenvalues of the matrix $T^{-1}\sum_{t=1}^T \mathcal{Z}_t\mathcal{Z}_t^\top$ lie in the interval $[C_L, C_U]$ with probability tending to one.

Table 3. The estimated parameters for VAR(1) and VAR(2) models. Those that are not significant at 5% level are marked by asterisk

	VAR(1)			VAR(2)				
	$\widehat{Z}_{t-1,1}$	$\widehat{Z}_{t-1,2}$	Const.	$\widehat{Z}_{t-1,1}$	$\widehat{Z}_{t-1,2}$	$\widehat{Z}_{t-2,1}$	$\widehat{Z}_{t-2,2}$	Const.
$\widehat{Z}_{t,1}$	0.984	−0.029*	−0.001	0.913	−0.025	0.071	−0.004	−0.001
$\widehat{Z}_{t,2}$	0.055	0.739	0.005	0.124	0.880	−0.065	−0.187*	0.006

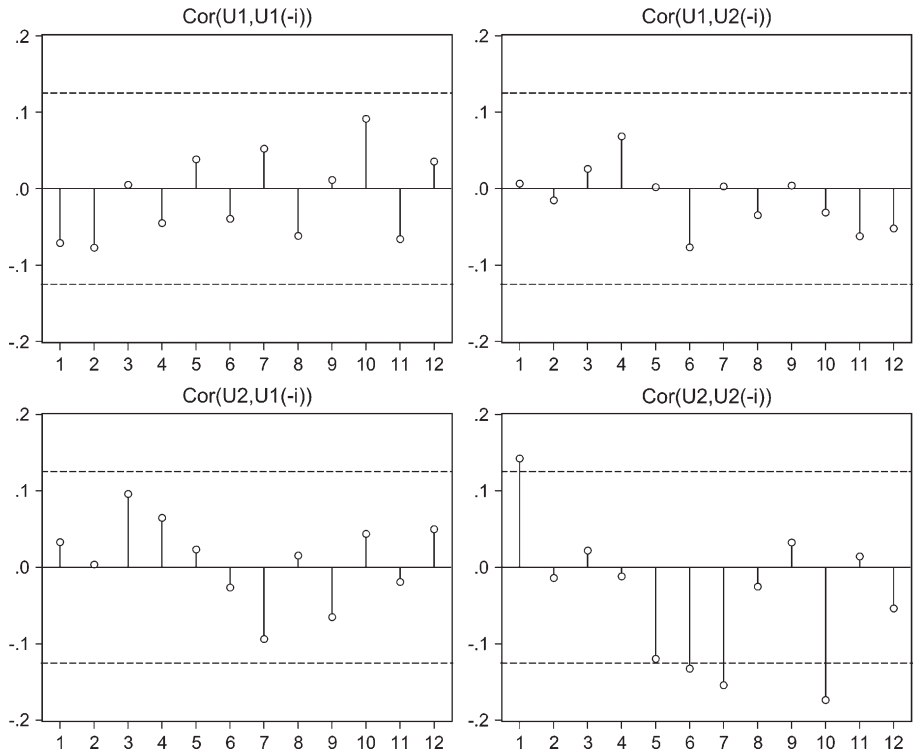

Figure 6. Cross-autocorrelogram for the VAR(1) residuals. The dashed line-bounds indicate $\pm 2 \times$ (standard deviations), which correspond to an approximate 95% confidence bound.

(A7) The minimization (5) runs over all values of $(\mathcal{A}, z)$ with

$$\sup_{x\in[0,1]} \max_{1\le t\le T} \| (1, z_t^\top)\mathcal{A}\psi(x) \| \le M_T,$$

where the constant M_T fulfils $\max_{1\le t\le T} \|Z_t\| \le M_T/C_m$ (with probability tending to one) for a constant C_m such that $\sup_{x\in[0,1]}\|m(x)\| < C_m$.

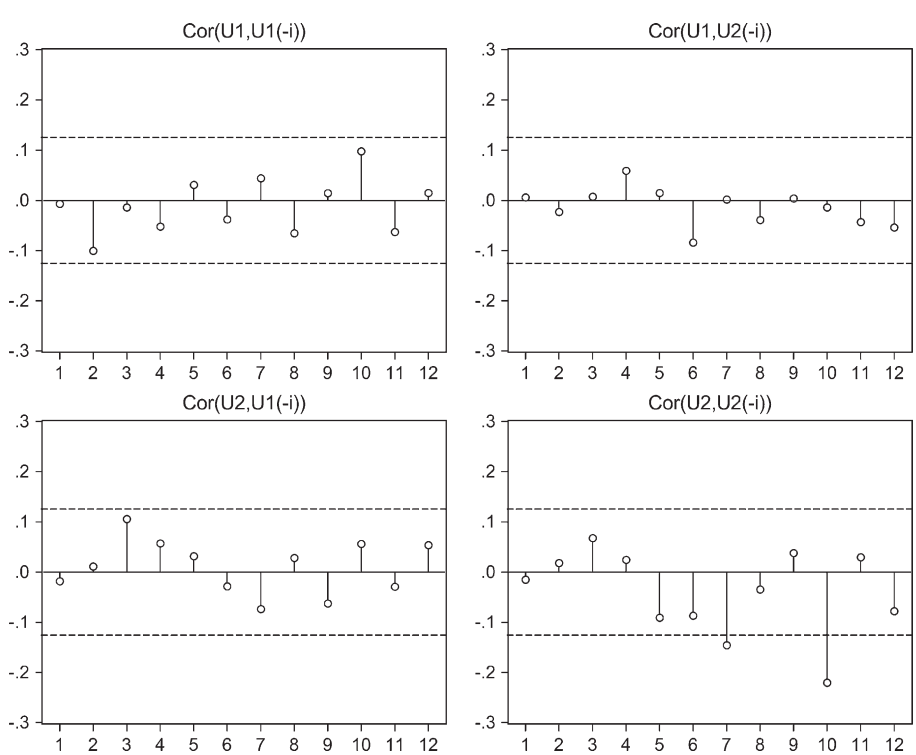

Figure 7. Cross-autocorrelogram for the VAR(2) residuals. The dashed line-bounds indicate $\pm 2 \times$ (standard deviations), which correspond to an approximate 95% confidence bound.

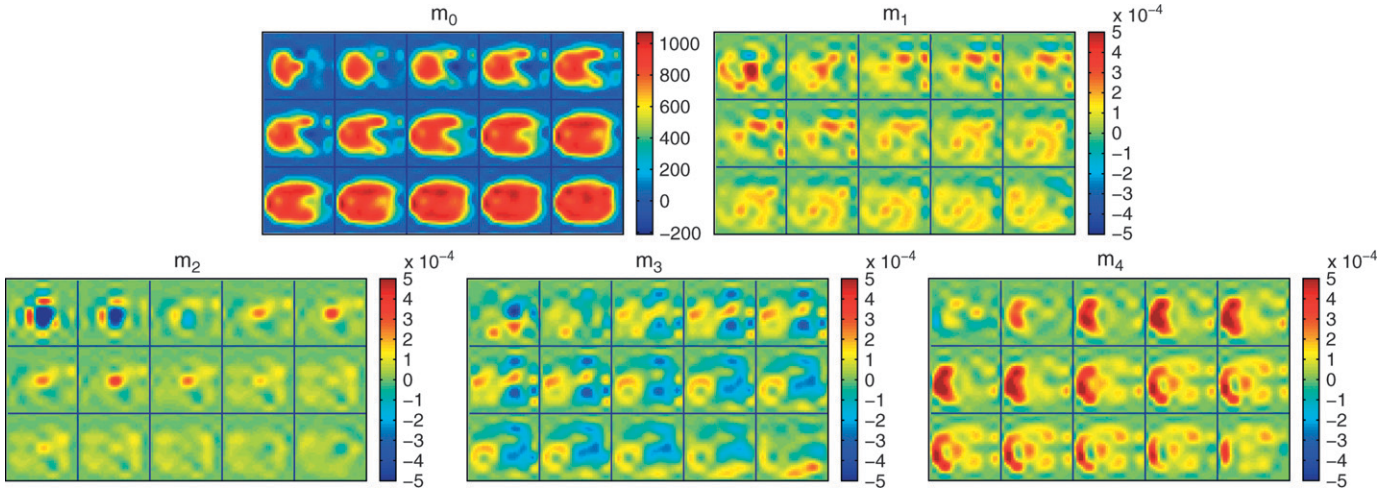

Figure 8. The estimated functions $\hat{m}_l$ for the fMRI signals.

(A8) It holds that $\rho^2 = (K+T)M_T^2 \log(JTM_T)/(JT) \to 0$. The dimension L is fixed.

Assumption (A7) and the additional bound M_T in the minimization is introduced for purely technical reasons. We conjecture that to some extent the asymptoic theory of this article could be developed under weaker conditions. The independence assumptions in (A1) and Assumption (A3) could be relaxed to assuming that the errors $\epsilon_{t,j}$ have a conditional mean zero and have a conditional distribution with subgaussian tails, given the past values $X_{s,i}$, Z_s $(1 \le i \le J,\ 1 \le s \le t)$. Such a theory would require an empirical process theory that is more explicitly designed for our model and it would also require a lot of more technical assumptions. We also expect that one could proceed with the assumption of subexponential instead of subgaussian tails, again at the cost of some additional conditions. Recall that the number of parameters to be estimated equals $TL + K(L+1)$. Because L is fixed, Assumption (A8) requires basically that, neglecting the factor $M_T{}^2 \log(JTM_T)$, the number of parameters grows slower than the number of observations, JT.

Our first result gives rates of convergence for the least squares estimators $\widehat{Z}_t$ and $\widehat{\mathcal{A}}$.

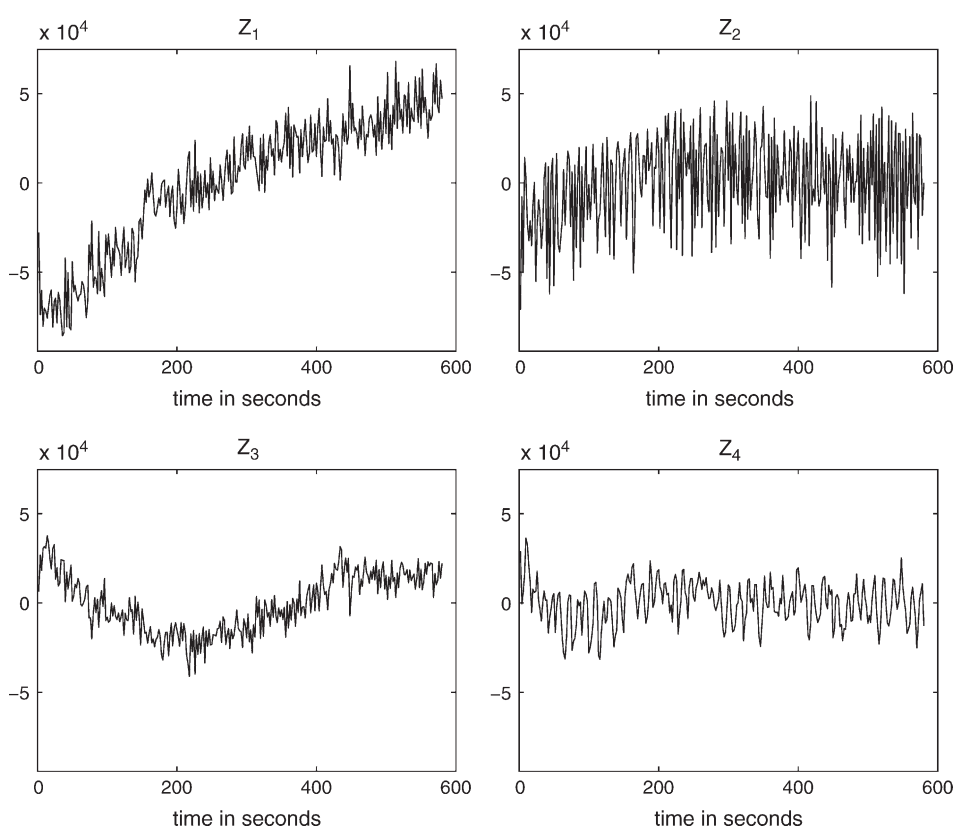

Figure 9. The estimated time series $\widehat{Z}_{t,l}$ for the fMRI signals.

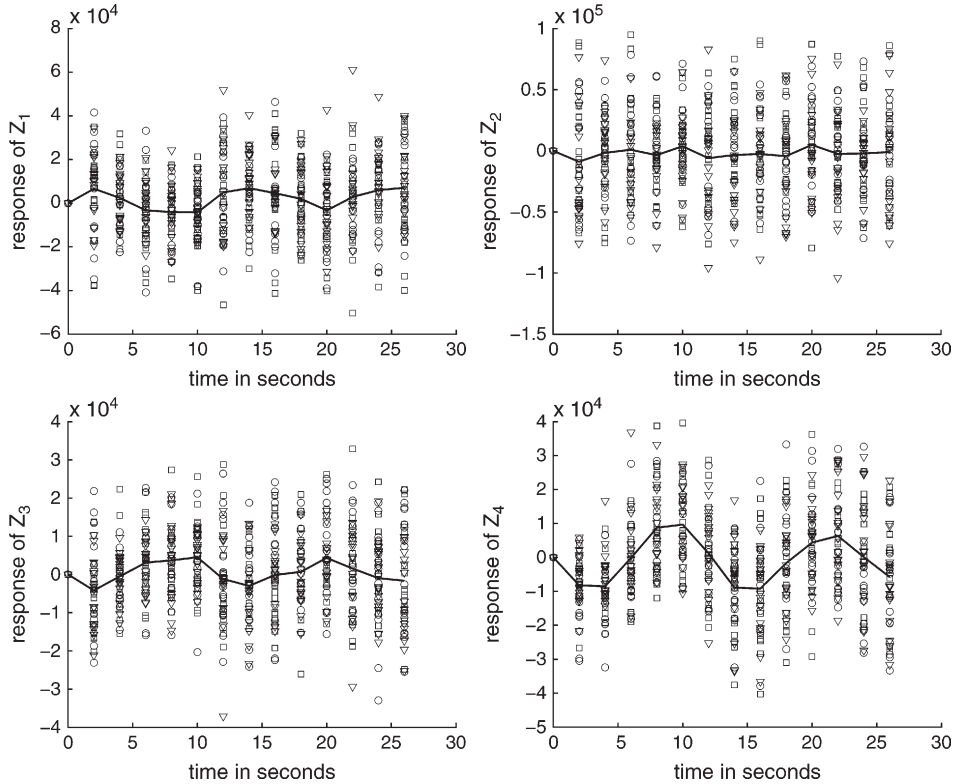

Figure 10. The responses of $\widehat{Z}_{t,l}$ to the stimuli.

Theorem 2. Suppose that model (4) holds and that $(\widehat{\mathcal{Z}}_t, \widehat{\mathcal{A}})$ is defined by the minimization problem (5). Make the Assumptions (A1)–(A8). Then it holds that

$$\frac{1}{T}\sum_{1\le t\le T}\left\|\widehat{\mathcal{Z}}_t^\top\widehat{\mathcal{A}}-\mathcal{Z}_t^\top\mathcal{A}^*\right\|^2=\mathcal{O}_P(\rho^2+\delta_K^2). \qquad (13)$$

At this point we have made no assumptions on the sequence Z_t: $1 \le t \le T$, besides the bound in (A7). Up to now it is allowed to be a deterministic or a random sequence. We now assume that it is a random process. We discuss how a statistical analysis differs if inference on Z_t is based on $\widehat{Z}_t$ instead of using (the unobserved) process Z_t. We will show that the differences are asymptotically negligible (except an orthogonal transformation). This is the content of the following theorem, where we consider estimators of autocovariances and show that these estimators differ only by second order terms. This asymptotic equivalence carries over to classical estimation and testing procedures in the framework of fitting a vector autoregresssive model. For the statement of the theorem we need the following assumptions:

(A9) Z_t is a strictly stationary sequence with $E(Z_t) = 0$, $E(\|Z_t\|^\gamma) < \infty$ for some $\gamma > 2$. It is strongly mixing with $\sum_{i=1}^\infty \alpha(i)^{(\gamma-2)/\gamma} < \infty$. The matrix $EZ_tZ_t^{\mathrm{T}}$ has full rank. The process Z_t is independent of $X_{11}, \ldots, X_{TJ}, \epsilon_{11}, \ldots, \epsilon_{TJ}$.
(A10) The functions $m_0, \ldots, m_L$ are linearly independent. In particular, no function is equal to 0.
(A11) It holds that $[\log(KT)^2\{(KM_T/J)^{1/2} + T^{1/2}M_T^4J^{-2} + K^{3/2}J^{-1} + K^{4/3}J^{-2/3}T^{-1/6}\} + 1]T^{1/2}(\rho^2 + \delta_K^2) = o(1)$.

Assumption (A11) poses very weak conditions on the growth of J, K, and T. Suppose, for example, that M_T is of logarithmic order and that K is of order $(TJ)^{1/5}$ so that the variance and the bias are balanced for twice differentiable functions. In this setting, (A11) only requires that T/J^2 times a logarithmic factor converges to zero. Define $\widetilde{Z}_t = B^\top\widehat{Z}_t$,

$$\widetilde{Z}_{c,t} = \widetilde{Z}_t - T^{-1}\sum\nolimits_{s=1}^T \widetilde{Z}_s,$$
$$Z_{c,t} = Z_t - T^{-1}\sum\nolimits_{s=1}^T Z_s,$$
$$\widetilde{Z}_{n,t} = (T^{-1}\sum\nolimits_{s=1}^T \widetilde{Z}_{c,s}\widetilde{Z}_{c,s}^\top)^{-1/2}\widetilde{Z}_{c,t},$$

and $Z_{n,t} = (T^{-1}\sum_{s=1}^T Z_{c,s}Z_{c,s}^\top)^{-1/2}Z_{c,t}$.

Theorem 3. Suppose that model (4) holds and that $(\widehat{Z}_t, \widehat{\mathcal{A}})$ is defined by the minimization problem (5). Make the Assumptions (A1)–(A11). Then there exists a random matrix B such that for $h \neq 0$

$$\frac{1}{T}\sum_{t=\max[1,-h+1]}^{\min[T,T-h]}\widetilde{Z}_{c,t}\left(\widetilde{Z}_{c,t+h}-\widetilde{Z}_{c,t}\right)^\top-\frac{1}{T}\sum_{t=\max[1,-h+1]}^{\min[T,T-h]}Z_{c,t}\left(Z_{c,t+h}-Z_{c,t}\right)^\top=o_P(T^{-1/2}),$$

$$\frac{1}{T}\sum_{t=\max[1,-h+1]}^{\min[T,T-h]}\widetilde{Z}_{n,t}\widetilde{Z}_{n,t+h}^\top-\frac{1}{T}\sum_{t=\max[1,-h+1]}^{\min[T,T-h]}Z_{n,t}Z_{n,t+h}^\top=o_P(T^{-1/2}).$$

To illustrate an implication of Theorem 3, suppose that the factor process Z_t in (4) is a stationary VAR(p) process in a mean adjusted form:

$$Z_t - \mu = \Theta_1(Z_{t-1} - \mu) + \ldots + \Theta_p(Z_{t-p} - \mu) + U_t, \quad (14)$$

where $\mu = E(Z_t)$, Θ_j is a $L \times L$ matrix of coefficients and U_t is a white noise with a nonsingular covariance matrix. Let Γ_h be the autocovariance matrix of the process Z_t with the lag $h \geq 0$, which is estimated by $\widehat{\Gamma}_h = T^{-1}\sum_{t=h+1}^{T}(Z_t - \overline{Z})(Z_{t-h} - \overline{Z})^\top$. Let $Y = (Z_{p+1} - \mu, \ldots, Z_T - \mu)$, $\Theta = (\Theta_1, \ldots, \Theta_p)$, and $U = (U_{p+1}, \ldots, U_T)$. Define $W_t = \left((Z_t - \mu)^\top, \ldots, (Z_{t-p+1} - \mu)^\top\right)^\top$ and $W = (W_p, \ldots, W_{T-1})$. Then, the model (14) can be rewritten as $Y = \Theta W + U$ and the least squares estimator of Θ is given by $\widehat{\Theta} = \widehat{Y}\widehat{W}^\top(\widehat{W}\widehat{W}^\top)^{-1}$, where $\widehat{Y}$ and $\widehat{W}$ are the same as Y and W, respectively, except that μ is replaced by $\overline{Z}$. Likewise, fitting a VAR(p) model with the estimated factor process $\widetilde{Z}_t$ yields $\widetilde{\Theta} = \widetilde{Y}\widetilde{W}^\top(\widetilde{W}\widetilde{W}^\top)^{-1}$, where $\widetilde{Y}$ and $\widetilde{W}$ are defined as $\widehat{Y}$ and $\widehat{W}$ with Z_t being replaced by $\widetilde{Z}_t$. Both $\widehat{Y}$ and $\widehat{W}$ are matrices composed of $\widehat{\Gamma}_h$ for various h. The matrices $\widetilde{Y}$ and $\widetilde{W}$ have the same forms as $\widehat{Y}$ and $\widehat{W}$, respectively, but with $\widehat{\Gamma}_h$ being replaced by $\widetilde{\Gamma}_h = T^{-1}\sum_{t=h+1}^{T}(\widetilde{Z}_t - \overline{\widetilde{Z}})(\widetilde{Z}_{t-h} - \overline{\widetilde{Z}})^\top$. It is well known that $\sqrt{T}(\widehat{\Theta} - \Theta) = \mathcal{O}_P(1)$, see Lütkepohl (1993). By Theorem 3, we have $\sqrt{T}(\widetilde{\Theta} - \widehat{\Theta}) = o_P(1)$.

APPENDIX: PROOFS OF THEOREMS

A.1 Proof of Theorem 1

We use the Newton-Kantorovich theorem to prove the theorem. The statement of the theorem may be found in Kantorovich and Akilov (1982), for example.

Suppose that $\sum_{t=1}^{T} \| \mathcal{Z}_t^{(0)\top}\mathcal{A}^{(0)} - \widehat{\mathcal{Z}}_t^\top\widehat{\mathcal{A}} \|^2 \leq r$ for some $r > 0$, which will be chosen later. With the Frobenius norm $\|M\|$ for a matrix M, we get

$$\begin{aligned}
\| \mathcal{A}^{(0)} - \widehat{\mathcal{A}} \|^2 \leq & \left\| \left(\sum_{t=1}^{T} \mathcal{Z}_t^{(0)} \mathcal{Z}_t^{(0)\top} \right)^{-1} \right\|^2 \cdot \\
& \left\| \sum_{t=1}^{T} \mathcal{Z}_t^{(0)} \mathcal{Z}_t^{(0)\top} (\mathcal{A}^{(0)} - \widehat{\mathcal{A}}) \right\|^2 \\
= & \left\| \left(\sum_{t=1}^{T} \mathcal{Z}_t^{(0)} \mathcal{Z}_t^{(0)\top} \right)^{-1} \right\|^2 \cdot \left\| \sum_{t=1}^{T} \mathcal{Z}_t^{(0)} \mathcal{Z}_t^{(0)\top} \mathcal{A}^{(0)} \right. \\
& \left. - \sum_{t=1}^{T} \mathcal{Z}_t^{(0)} \widehat{\mathcal{Z}}_t^\top \widehat{\mathcal{A}} \right\|^2 \leq \left\| \left(\sum_{t=1}^{T} \mathcal{Z}_t^{(0)} \mathcal{Z}_t^{(0)\top} \right)^{-1} \right\|^2 \\
& \times \left(\sum_{t=1}^{T} \left\| \mathcal{Z}_t^{(0)} \mathcal{Z}_t^{(0)\top} \mathcal{A}^{(0)} - \mathcal{Z}_t^{(0)} \widehat{\mathcal{Z}}_t^\top \widehat{\mathcal{A}} \right\| \right)^2 \leq \\
& r \left\| \left(\sum_{t=1}^{T} \mathcal{Z}_t^{(0)} \mathcal{Z}_t^{(0)\top} \right)^{-1} \right\|^2 \left(\sum_{t=1}^{T} \| \mathcal{Z}_t^{(0)} \|^2 \right) \\
\equiv & \; r c_1^2.
\end{aligned} \quad (A.1)$$

For a matrix M, define $\|M\|_2 = \sup_{\|x\|=1} \|Mx\|$. It is known that $\|M\|_2 \leq \|M\|$. We get

$$\begin{aligned}
\| \widehat{\mathcal{A}}^\top(\mathcal{Z}_t^{(0)} - \widehat{\mathcal{Z}}_t) \| \geq & \| \widehat{\mathcal{A}} \|_2^{-1} \\
& \cdot \| (\widehat{\mathcal{A}}\widehat{\mathcal{A}}^\top)^{-1} \|^{-1} \cdot \| \mathcal{Z}_t^{(0)} - \widehat{\mathcal{Z}}_t \|,
\end{aligned} \quad (A.2)$$

$$\begin{aligned}
\| (\mathcal{Z}_t^{(0)} - \widehat{\mathcal{Z}}_t)^\top \widehat{\mathcal{A}} \| \leq & \| \mathcal{Z}_t^{(0)\top}(\widehat{\mathcal{A}} - \mathcal{A}^{(0)}) \| + \| \mathcal{Z}_t^{(0)\top}\mathcal{A}^{(0)} - \\
& \widehat{\mathcal{Z}}_t^\top \widehat{\mathcal{A}} \| \leq \| \mathcal{Z}_t^{(0)} \| \cdot \| \widehat{\mathcal{A}} - \mathcal{A}^{(0)} \| + \| \mathcal{Z}_t^{(0)\top}\mathcal{A}^{(0)} - \widehat{\mathcal{Z}}_t^\top \widehat{\mathcal{A}} \| .
\end{aligned} \quad (A.3)$$

The two inequalities (A.2) and (A.3) together with (A.1) give

$$\begin{aligned}
\| \mathcal{Z}^{(0)} - \widehat{\mathcal{Z}} \|^2 \leq & \; 2r \; \| \widehat{\mathcal{A}} \|_2^2 \cdot \| (\widehat{\mathcal{A}}\widehat{\mathcal{A}}^\top)^{-1} \|^2 \\
& \times \left(1 + c_1 \sum_{t=1}^{T} \| \mathcal{Z}_t^{(0)} \|^2 \right) \equiv r c_2^2.
\end{aligned} \quad (A.4)$$

Because $F'(\alpha, z)$ is quadratic in (α, z), there exists $0 < c_3 < \infty$ for any compact set D in $\mathbb{R}^{K(L+1)+TL}$ such that $\|F'(\alpha', z') - F'(\alpha, z)\|_2 \leq c_3\|(\alpha'^\top, z'^\top)^\top - (\alpha^\top, z^\top)^\top\|$ for all $(\alpha^\top, z^\top)^\top$, $(\alpha'^\top, z'^\top)^\top \in D$. Let $c_4 = \|F'_*(\alpha^{(0)}, Z^{(0)})^{-1}\|_2 < \infty$. Because F is continuous and $F(\hat{\alpha}, \widehat{Z}) = 0$, there exists $r' > 0$ such that, if $\| \alpha^{(0)} - \widehat{\alpha} \| + \| Z^{(0)} - \widehat{Z} \| \leq r'$, then

$$\| F'_*(\alpha^{(0)}, Z^{(0)})^{-1} F(\alpha^{(0)}, Z^{(0)}) \| \leq \frac{\gamma}{2c_3c_4}.$$

By the Newton-Kantorovich theorem,

$$\| \alpha^{(k)} - \widehat{\alpha} \| + \| Z^{(k)} - \widehat{Z} \| \leq C_1 2^{-(k-1)} \gamma^{2^k - 1} \quad (A.5)$$

for some $C_1 > 0$. This gives that if $\| \alpha^{(0)} - \widehat{\alpha} \| + \| Z^{(0)} - \widehat{Z} \| \leq r'$, then

$$\begin{aligned}
\sum_{t=1}^{T} \| \mathcal{Z}_t^{(k)\top}\mathcal{A}^{(k)} - \widehat{\mathcal{Z}}_t^\top \widehat{\mathcal{A}} \|^2 \leq & \; C_2(\| \alpha^{(k)} - \widehat{\alpha} \|^2 + \\
& \| Z^{(k)} - \widehat{Z} \|^2) \leq C 2^{-2(k-1)} \gamma^{2(2^k - 1)}
\end{aligned}$$

for some C, $C_2 > 0$. We take $r = (c_1 + c_2)^{-2} r'^2$. Then, by (A.1) and (A.4), $\| \alpha^{(0)} - \widehat{\alpha} \| + \| Z^{(0)} - \widehat{Z} \| \leq r'$ if $\sum_{t=1}^{T} \| \mathcal{Z}_t^{(0)\top}\mathcal{A}^{(0)} - \widehat{\mathcal{Z}}_t^\top \widehat{\mathcal{A}} \|^2 \leq r$. This completes the proof of the theorem.

A.2 Proof of Theorem 2

For functions $g(t, x)$ we define the norms $\| g \|_1^2 = (1/TJ)\sum_{t=1}^{T}\sum_{j=1}^{J} g(t, X_{t,j})^2$, $\|g\|_2^2 = (1/T)\sum_{t=1}^{T}\int g(t,x)^2 f_t(x)\, dx$, and $\| g \|_3^2 = (1/T)\sum_{t=1}^{T}\int g(t,x)^2\, dx$. Note that because of Assumption (A2) the last two norms are equivalent. Thus, for the statement of the theorem we have to show for $\Delta(t,x) = (\widehat{\mathcal{Z}}_t^\top \widehat{\mathcal{A}} - \mathcal{Z}_t^\top \mathcal{A}^*)\psi(x)$ that

$$\| \Delta \|_2^2 = \mathcal{O}_P(\rho^2 + \delta_K^2). \quad (A.6)$$

We start by showing that

$$\| \Delta \|_1^2 = \mathcal{O}_P([(K+T)\log(JTM_T)]/(JT) + \delta_K^2). \quad (A.7)$$

For this aim we apply Theorem 10.11 in Van de Geer (2000) that treats rates of convergence for least squares estimators on sieves. In our case we have the following sieve: $\mathcal{G}_T^* = \{g\colon \{1, \ldots, T\} \times [0,1]^d \to \mathbb{R}, g(t,x) = (1, z_t^\top)\mathcal{A}\psi(x)$ for an $(L+1) \times K$ matrix $\mathcal{A}$ and $z_t \in \mathbb{R}^L$ with the following properties: $|(1, z_t^\top)\mathcal{A}\psi(x)| \leq M_T$ for $1 \leq t \leq T$ and $x \in [0,1]^d\}$. With a constant C the δ-entropy $H_T(\delta, \mathcal{G}_T^*)$ of $\mathcal{G}_T^*$ with respect to the empirical norm $\|g\|_1$ is bounded by

$$H_T(\delta, \mathcal{G}_T^*) \leq CT\log(M_T/\delta) + CK \log(KM_T/\delta). \quad (A.8)$$

For the proof of (A.8) note first that each element $g(t,x) = (1, z_t^\top)\mathcal{A}\psi(x)$ of $\mathcal{G}_T^*$ can be chosen such that $T^{-1}\sum_{t=1}^T z_t z_t^\top$ is equal to the $L \times L$ identity matrix I_L. Then the bound $|(1, z_t^\top)\mathcal{A}\psi(x)| \le M_T$ implies that $\| \mathcal{A}\psi(x) \| \le M_T$. For the proof of (A.8) we use that the (δ/M_T)-entropy of a unit ball in $\mathbb{R}^T$ is of order $\mathcal{O}(T \log(M_T/\delta))$ and that the δ-entropy with respect to the sup-norm for functions $\mathcal{A}\psi(x)$ with $\| \mathcal{A}\psi(x) \| \le M_T$ is of order $\mathcal{O}(K \log(KM_T/\delta))$. In the last entropy bound we used that for each x it holds that $||\psi(x)|| \le K^{1/2}$. These two entropy bounds imply (A.8). Application of Theorem 10.11 in Van de Geer (2000) gives (A.7).

We now show that (A.7) implies (A.6). For this aim note first that by Bernstein's inequality for a, $d > 0$, $g \in \mathcal{G}_T^*$ with $||g||_2^2 \le d$

$$P(|\| g \|_1^2 - \| g \|_2^2 | \ge a) \le 2 \exp\left(-\frac{a^2 JT}{2(a+d)M_T^2}\right). \tag{A.9}$$

Furthermore, for $g, h \in \mathcal{G}_T^*$ it holds with constants C, C' that

$$\begin{aligned} &|\| g \|_1^2 - \| h \|_1^2 | \le CK\left(T^{-1}\sum_{t=1}^T \| e_t - f_t \|^2\right)^{1/2} \\ &\left(T^{-1}\sum_{t=1}^T \| e_t + f_t \|^2\right)^{1/2} \le C'K \| g - h \|_2 (\| g \|_2 + \| h \|_2), \end{aligned} \tag{A.10}$$

where e_t and f_t are chosen such that $g(x,t) = e_t^\top \psi(x)$ and $h(x,t) = f_t^\top \psi(x)$. From (A.9) and (A.10) we get with a constant $C > 0$ that for $d = 1, 2, \ldots$

$$\begin{aligned} &P(\sup_{g \in \mathcal{G}_T^*, d\rho^2 \le \|g\|_2^2 \le (d+1)\rho^2} |\| g \|_1^2 - \| g \|_2^2 | \ge d\rho^2/2) \\ &\le C \exp((C + K + T) \log(dKM_T) - d\rho^2 JT/[20M_T^2]). \end{aligned}$$

By summing these inequalities over $d \ge 1$ we get $||\Delta||_2^2 \le \rho^2$ or

$$||\Delta||_2^2 \le |||\Delta||_1^2 - ||\Delta||_2^2| + ||\Delta||_1^2 \le ||\Delta||_2^2/2 + ||\Delta||_1^2$$

with probability tending to one. This shows Equation (A.6) and concludes the proof of Theorem 2.

A.3 Proof of Theorem 3

We will prove the first equation of the theorem for $h \ne 0$. The second equation follows from the first equation. We first prove that the matrix $T^{-1}\sum_{t=1}^T \mathcal{Z}_{c,t}\widehat{\mathcal{Z}}_{c,t}^\top$ is invertible, where $\mathcal{Z}_{c,t}^\top = (1, Z_{c,t}^\top)$, $\widehat{\mathcal{Z}}_{c,t}^\top = (1, \widehat{Z}_{c,t}^\top)$, and $\widehat{Z}_{c,t} = \widehat{Z}_t - T^{-1}\sum_{s=1}^T \widehat{Z}_s$. This implies that $T^{-1}\sum_{t=1}^T Z_{c,t}\widehat{Z}_{c,t}^\top$ is invertible. Suppose that the assertion is not true. We can choose a random vector e such that $\|e\| = 1$ and $e^\top \sum_{t=1}^T \mathcal{Z}_{c,t}\widehat{\mathcal{Z}}_{c,t}^\top = 0$. Let $\widehat{A}$ and A^* be the $L \times K$ matrices that are obtained by deleting the first rows of $\widehat{\mathcal{A}}$ and $\mathcal{A}^*$, respectively. Let $\widehat{\mathcal{A}}_c$ and $\mathcal{A}_c^*$ be the matrices obtained from $\widehat{\mathcal{A}}$ and $\mathcal{A}^*$ by replacing their first rows by $\widehat{\alpha}_0^\top + (T^{-1}\sum_{t=1}^T \widehat{Z}_t)^\top \widehat{A}$ and $\alpha_0^{*\top} + (T^{-1}\sum_{t=1}^T Z_t)^\top A^*$, respectively. By definition, it follows that

$$\widehat{\mathcal{Z}}_t^\top \widehat{\mathcal{A}} = \widehat{\mathcal{Z}}_{c,t}^\top \widehat{\mathcal{A}}_c, \quad \mathcal{Z}_t^\top \mathcal{A}^* = \mathcal{Z}_{c,t}^\top \mathcal{A}_c^*. \tag{A.11}$$

Note that

$$\begin{aligned} &\left\| T^{-1}\sum_{t=1}^T \mathcal{Z}_{c,t}\widehat{\mathcal{Z}}_{c,t}^\top \widehat{\mathcal{A}}_c - T^{-1}\sum_{t=1}^T \mathcal{Z}_{c,t}\mathcal{Z}_{c,t}^\top \mathcal{A}_c^* \right\| \\ &\le T^{-1}\sum_{t=1}^T \left\| \mathcal{Z}_{c,t}\left(\widehat{\mathcal{Z}}_{c,t}^\top \widehat{\mathcal{A}}_c - \mathcal{Z}_{c,t}^\top \mathcal{A}_c^*\right) \right\| \\ &\le \left(T^{-1}\sum_{t=1}^T \left\| \mathcal{Z}_{c,t} \right\|^2\right)^{1/2} \left(T^{-1}\sum_{t=1}^T \left\| \widehat{\mathcal{Z}}_t^\top \widehat{\mathcal{A}} - \mathcal{Z}_t^\top \mathcal{A}^* \right\|^2\right)^{1/2} \\ &= \mathcal{O}_P(\rho + \delta_K), \end{aligned} \tag{A.12}$$

because of Assumption (A6) and Theorem 2. Thus with $f = T^{-1}\sum_{t=1}^T \mathcal{Z}_{c,t}\mathcal{Z}_{c,t}^\top e$, we obtain

$$\begin{aligned} \| f^\top m \| &= \| f^\top (\mathcal{A}_c^* \psi) \| + \mathcal{O}_P(T^{-1/2} + \delta_K) \\ &= \left\| e^\top T^{-1}\sum_{t=1}^T \mathcal{Z}_{c,t}\widehat{\mathcal{Z}}_{c,t}^\top \widehat{\mathcal{A}}_c \psi \right\| + \mathcal{O}_P(T^{-1/2} + \rho + \delta_K) \\ &= \mathcal{O}_P(T^{-1/2} + \rho + \delta_K). \end{aligned}$$

This implies that $m_0, \ldots, m_d$ are linearly dependent, contradicting to Assumption (A10).

Let $\widetilde{B}$ be the matrix given at (8) with B defined as in (9). Define $\widetilde{\mathcal{Z}}_{c,t} = \widetilde{B}^\top \widehat{\mathcal{Z}}_{c,t}$ and $\widetilde{\mathcal{A}}_c = \widetilde{B}^{-1}\widehat{\mathcal{A}}_c$. Then $\widetilde{\mathcal{Z}}_{c,t}^\top \widetilde{\mathcal{A}}_c = \widehat{\mathcal{Z}}_{c,t}^\top \widehat{\mathcal{A}}_c$ and $T^{-1}\sum_{t=1}^T \mathcal{Z}_{c,t}\widetilde{\mathcal{Z}}_{c,t}^\top = T^{-1}\sum_{t=1}^T \mathcal{Z}_{c,t}\mathcal{Z}_{c,t}^\top$. This gives with (A.12)

$$\begin{aligned} \left\| \widetilde{\mathcal{A}}_c - \mathcal{A}_c^* \right\| &= \left\| T^{-1}\sum_{t=1}^T \mathcal{Z}_{c,t}\mathcal{Z}_{c,t}^\top (\widetilde{\mathcal{A}}_c - \mathcal{A}_c^*) \right\| \mathcal{O}_P(1) \\ &= \left\| T^{-1}\sum_{t=1}^T \mathcal{Z}_{c,t}\widetilde{\mathcal{Z}}_{c,t}^\top \widetilde{\mathcal{A}}_c - T^{-1}\sum_{t=1}^T \mathcal{Z}_{c,t}\mathcal{Z}_{c,t}^\top \mathcal{A}_c^* \right\| \mathcal{O}_P(1) \\ &= \mathcal{O}_P(\rho + \delta_K). \end{aligned} \tag{A.13}$$

Because of Theorem 2 this implies

$$\left\| \widetilde{\mathcal{A}} - \mathcal{A}^* \right\| = \mathcal{O}_P(\rho + \delta_K). \tag{A.14}$$

Define $\widetilde{Z}_{c,t}$ by $\widetilde{\mathcal{Z}}_{c,t}^\top = (1, \widetilde{Z}_{c,t}^\top)$. Note that $\widetilde{Z}_{c,t} = B^\top \widehat{Z}_{c,t}$. Also, define $\widetilde{A} = B^{-1}\widehat{A}$, which equals $\widetilde{\mathcal{A}}_c$ without the first row. From (A10), (A5), (A.14), and Theorem 2, we get

$$\begin{aligned} &T^{-1}\sum_{t=1}^T \left\| \widetilde{Z}_t - Z_t \right\|^2 = T^{-1}\sum_{t=1}^T \left\| \widetilde{\mathcal{Z}}_t - \mathcal{Z}_t \right\|^2 \\ &= T^{-1}\sum_{t=1}^T \left\| \widetilde{\mathcal{Z}}_t^\top (m_0, \ldots, m_L)^\top - \mathcal{Z}_t^\top (m_0, \ldots, m_L)^\top \right\|^2 \mathcal{O}_P(1) \\ &= T^{-1}\sum_{t=1}^T \left\| \widetilde{Z}_t^\top A^* - \widetilde{Z}_t^\top \widetilde{A} \right\|^2 \mathcal{O}_P(1) + T^{-1}\sum_{t=1}^T \left\| \widetilde{Z}_t^\top \widetilde{A} - Z_t^\top A^* \right\|^2 \\ &\quad \times \mathcal{O}_P(1) + \mathcal{O}_P(\delta_K^2) \\ &\le T^{-1}\sum_{t=1}^T \| \widetilde{Z}_t - Z_t \|^2 \| \widetilde{A} - A^* \|^2 \mathcal{O}_P(1) + T^{-1}\sum_{t=1}^T \| Z_t \|^2 \\ &\quad \times \| \widetilde{A} - A^* \|^2 \mathcal{O}_P(1) \\ &\quad + T^{-1}\sum_{t=1}^T \| \widetilde{Z}_t^\top \widetilde{A} - Z_t^\top A^* \|^2 \mathcal{O}_P(1) + \mathcal{O}_P(\rho^2 + \delta_K^2) \\ &= \mathcal{O}_P(\rho^2 + \delta_K^2). \end{aligned} \tag{A.15}$$

From Equation (A.15) one gets

$$T^{-1}\sum_{t=1}^{T}\left\|\widetilde{Z}_{c,t}-Z_{c,t}\right\|^2=\mathcal{O}_P(\rho^2+\delta_K^2). \tag{A.16}$$

We will show that for $h\neq 0$

$$T^{-1}\sum_{t=h+1}^{T}\{(\widetilde{Z}_{c,t+h}-Z_{c,t+h})-(\widetilde{Z}_{c,t}-Z_{c,t})\}Z_{c,t}^{\top}=o_P(T^{-1/2}). \tag{A.17}$$

This implies the first statement of Theorem 3, because by (A.16)

$$T^{-1}\sum_{t=-h+1}^{T}(\widetilde{Z}_{c,t}-Z_{c,t})(\widetilde{Z}_{c,t+h}^{\top}-Z_{c,t+h}^{\top})=\mathcal{O}_P(\rho^2+\delta_K^2)$$
$$=\mathcal{O}_P(T^{-1/2}).$$

For the proof of (A.17), let $\widetilde{\alpha}_c$ be the stack form of $\widetilde{\mathcal{A}}_c$ and $\widetilde{\alpha}_{c,0}^{\top}$ be its first row. Using the representation (6) and the first identity of (A.11), it can be verified that

$$\widetilde{Z}_{c,t}=\widetilde{S}_{t,Z}^{-1}J^{-1}\sum_{j=1}^{J}\{Y_{t,j}\widetilde{A}\psi(X_{t,j})-\widetilde{A}\psi(X_{t,j})\psi(X_{t,j})^{\top}\widetilde{\alpha}_{c,0}\}, \tag{A.18}$$

$$\widetilde{\alpha}_c=\widetilde{\mathcal{S}}_{\alpha}^{-1}T^{-1}J^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J}\{\psi(X_{t,j})\otimes\widetilde{\mathcal{Z}}_{c,t}\}Y_{t,j}, \tag{A.19}$$

where $\widetilde{S}_{t,Z}=J^{-1}\sum_{j=1}^{J}\widetilde{A}\psi(X_{t,j})\psi(X_{t,j})^{\top}\widetilde{A}^{\top}$ and $\widetilde{\mathcal{S}}_{\alpha}=T^{-1}J^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J}\{\psi(X_{t,j})\otimes\widetilde{\mathcal{Z}}_{c,t}\}\{\psi(X_{t,j})\otimes\widetilde{\mathcal{Z}}_{c,t}\}^{\top}$. Define $\widetilde{\mathcal{S}}_{t,Z}$ as $\widetilde{S}_{t,Z}$ with $\widetilde{\mathcal{A}}_c$ replacing $\widetilde{A}$. Also, define $\mathcal{S}_{t,Z}=\mathcal{A}_c^{*}E\{\psi(X_{t,j})\psi(X_{t,j})^{\top}\}\mathcal{A}_c^{*\top}$, $S_{t,Z}=A^{*}E\left\{\psi(X_{t,j})\psi(X_{t,j})^{\top}\right\}A^{*\top}$ and

$$\mathcal{S}_{\alpha}=T^{-1}\sum_{t=1}^{T}E[\{\psi(X_{t,j})\otimes\mathcal{Z}_{c,t}\}\{\psi(X_{t,j})\otimes\mathcal{Z}_{c,t}\}^{\top}|Z_t].$$

Let $\gamma=T^{-1/2}(\rho+\delta_K)^{-1}$. We argue that

$$\sup_{1\le t\le T}\|\widetilde{\mathcal{S}}_{t,Z}-\mathcal{S}_{t,Z}\|=o_P(\gamma),\quad \|\widetilde{\mathcal{S}}_{\alpha}-\mathcal{S}_{\alpha}\|=o_P(\gamma). \tag{A.20}$$

We show the first part of (A.20). The second part can be shown similarly. To prove the first part it suffices to show that, uniformly for $1\le t\le T$,

$$J^{-1}\sum_{j=1}^{J}\mathcal{A}_c^{*}[\psi(X_{t,j})\psi(X_{t,j})^{\top}-E\{\psi(X_{t,j})\psi(X_{t,j})^{\top}\}](\widetilde{\mathcal{A}}_c-\mathcal{A}_c^{*})^{\top}$$
$$=o_P(\gamma), \tag{A.21}$$

$$J^{-1}\sum_{j=1}^{J}(\widetilde{\mathcal{A}}_c-\mathcal{A}_c^{*})[\psi(X_{t,j})\psi(X_{t,j})^{\top}-E\{\psi(X_{t,j})\psi(X_{t,j})^{\top}\}]$$
$$(\widetilde{\mathcal{A}}_c-\mathcal{A}_c^{*})^{\top}=o_P(\gamma), \tag{A.22}$$

$$J^{-1}\sum_{j=1}^{J}\mathcal{A}_c^{*}[\psi(X_{t,j})\psi(X_{t,j})^{\top}-E\{\psi(X_{t,j})\psi(X_{t,j})^{\top}\}]\mathcal{A}_c^{*\top}$$
$$=o_P(\gamma), \tag{A.23}$$

$$J^{-1}\sum_{j=1}^{J}\mathcal{A}_c^{*}E\{\psi(X_{t,j})\psi(X_{t,j})^{\top}\}(\widetilde{\mathcal{A}}_c-\mathcal{A}_c^{*})^{\top}=o_P(\gamma), \tag{A.24}$$

$$J^{-1}\sum_{j=1}^{J}(\widetilde{\mathcal{A}}_c-\mathcal{A}_c^{*})E\{\psi(X_{t,j})\psi(X_{t,j})^{T}\}(\widetilde{\mathcal{A}}_c-\mathcal{A}_c^{*})^{\top}=o_P(\gamma). \tag{A.25}$$

The proof of (A.23)–(A.25) follows by simple arguments. We now show (A.21). Claim (A.22) can be shown similarly. For the proof of (A.21) we use Bernstein's inequality for the following sum:

$$P\left(|\sum_{j=1}^{J}W_j|>x\right)\le 2\exp\left(-\frac{1}{2}\frac{x^2}{V+Mx/3}\right). \tag{A.26}$$

Here for a value of t with $1\le t\le T$, the random variable W_j is an element of the $(L+1)\times 1$-matrix $S=J^{-1}\mathcal{A}_c^{*}[\psi(X_{t,j})\psi(X_{t,j})^{\top}e-E\{\psi(X_{tj})\psi(X_{tj})^{\top}e\}]$ where $e\in\mathbb{R}^K$ with $\|e\|=1$. In (A.26), V is an upper bound for the variance of $\sum_{j=1}^{J}W_j$ and M is a bound for the absolute values of W_j (i.e. $|W_j|\le M$ for $1\le j\le J$, a.s.). With some constants C_1 and C_2 that do not depend on t and the row number we get $V\le C_1J^{-1}$ and $M\le C_2K^{1/2}J^{-1}$. Application of Bernstein's inequality gives that, uniformly for $1\le t\le T$ and $e\in\mathbb{R}^K$ with $\|e\|=1$, all $(L+1)$ elements of S are of order $o_P(\gamma)$. This shows claim (A.21).

From (A.13), (A.15), (A.18), (A.19), and (A.20) it follows that uniformly for $1\le t\le T$,

$$\widetilde{Z}_{c,t}-Z_{c,t}=S_{t,Z}^{-1}J^{-1}\sum_{j=1}^{J}\varepsilon_{t,j}A^{*}\psi(X_{t,j})+S_{t,Z}^{-1}J^{-1}\sum_{j=1}^{J}\varepsilon_{t,j}$$
$$\times(\widetilde{A}-A^{*})\psi(X_{t,j}) \tag{A.27}$$
$$+S_{t,Z}^{-1}J^{-1}\sum_{j=1}^{J}(\widetilde{A}-A^{*})\psi(X_{t,j})\psi(X_{t,j})^{\top}\mathcal{A}_c^{*\top}\mathcal{Z}_{c,t}+o_P(T^{-1/2})$$
$$\equiv\Delta_{t,1,Z}+\Delta_{t,2,Z}+\Delta_{t,3,Z}+o_P(T^{-1/2}).$$

For the proof of the theorem it remains to show that for $1\le j\le 3$

$$T^{-1}\sum_{t=-h+1}^{T}(\Delta_{t+h,j,Z}-\Delta_{t,j,Z})Z_{c,t}^{\top}=o_P(T^{-1/2}). \tag{A.28}$$

This can be easily checked for $j=1$. For $j=2$ it follows from $\|\widetilde{A}-A^{*}\|=\mathcal{O}(\rho+\delta_k)$ and

$$E\left\{\|T^{-1}J^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J}\varepsilon_{t,j}S_{t,Z}^{-1}\mathcal{M}\psi(X_{t,j})\|^2\right\}=\mathcal{O}(KJ^{-1}T^{-1})$$

for any $L\times K$ matrix $\mathcal{M}$ with $\|\mathcal{M}\|=1$. For the proof of (A.28) for $j=3$, it suffices to show that

$$T^{-1}\sum_{t=1}^{T+h}\Delta_{t,j,Z}(Z_{c,t-h}-Z_{c,t})^{\top}=o_P(T^{-1/2}). \tag{A.29}$$

We note first that for $1\le l\le L$

$$T^{-1}\sum_{t=1}^{T+h}\Delta_{t,3,Z}(Z_{c,t-h,l}-Z_{c,t,l})$$
$$=T^{-1}J^{-1}\sum_{t=1}^{T+h}\sum_{j=1}^{J}\left\{\left(V_{h,t}^{\top}\mathcal{A}_c^{*}\psi(X_{t,j})\psi(X_{t,j})^{\top}\right)\otimes S_{t,Z}^{-1}\right\}(\widetilde{\alpha}-\alpha^{*}),$$

where $V_{h,t} = (\underset{\sim}{Z}_{c,t-h,l} - Z_{c,t,l})\mathcal{Z}_{c,t}$, and $\widetilde{\alpha}$ and α^* denote the stack forms of $\widetilde{A}$ and A^*, respectively. For the proof of (A.29) it suffices to show

$$T^{-1}J^{-1}\sum_{t=1}^{T+h}\sum_{j=1}^{J}\{(E[V_{h,t}]^\top \mathcal{A}_c^*\psi(X_{t,j})\psi(X_{t,j})^\top)\otimes S_{t,Z}^{-1}\} \times(\widetilde{\alpha}-\alpha^*) = o_P(T^{-1/2}), \tag{A.30}$$

$$\left\| T^{-1}J^{-1}\sum_{t=1}^{T+h}\sum_{j=1}^{J}\{(\{V_{h,t}-E[V_{h,t}]\}^\top \mathcal{A}_c^*\psi(X_{t,j})\psi(X_{t,j})^\top) \otimes S_{t,Z}^{-1}\}\right\|^2 = \mathcal{O}_P(KJ^{-1}T^{-1}). \tag{A.31}$$

Claim (A.31) can be easily shown by calculating the expectation of the left hand side of (A.31) and by using the mixing condition at Assumption (A9). For a proof of (A.30) we remark first that by construction

$$0 = T^{-1}\sum_{t=1}^{T}(\widetilde{Z}_{c,t} - Z_{c,t})Z_{c,t}^T.$$

Using (A.27) and similar arguments as in the proof of (A.28) for $j = 1, 2$ we get that

$$T^{-1}\sum_{t=1}^{T}\Delta_{t,3,Z}Z_{c,t}^T = T^{-1}J^{-1} \sum_{t=1}^{T}\sum_{j=1}^{J}\Big\{(Z_{c,t}\mathcal{Z}_{c,t}^\top\mathcal{A}_c^*\psi(X_{t,j})\psi(X_{t,j})^\top) \otimes S_{t,Z}^{-1}\Big\}(\widetilde{\alpha}-\alpha^*) = o_P(T^{-1/2}).$$

As in the proof of (A.31) one can show that

$$\left\| T^{-1}J^{-1}\sum_{t=1}^{T+h}\sum_{j=1}^{J}\{(\{Z_{c,t}\mathcal{Z}_{c,t}^\top - E[Z_{c,t}\mathcal{Z}_{c,t}^\top]\}\mathcal{A}_c^*\psi(X_{t,j})\psi(X_{t,j})^\top \otimes S_{t,Z}^{-1}\}\right\|^2 = \mathcal{O}_P(KJ^{-1}T^{-1}).$$

The last two equalities imply that

$$T^{-1}J^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J}\{(E[Z_{c,t}\mathcal{Z}_{c,t}^\top]\mathcal{A}_c^*\psi(X_{t,j})\psi(X_{t,j})^\top)\otimes S_{t,Z}^{-1}\} \times(\widetilde{\alpha}-\alpha^*) = o_P(T^{-1/2}).$$

Because of Assumption (A9) this implies claim (A.29) and concludes the proof of Theorem 3.

[Received June 2007. Revised August 2008.]

REFERENCES

Biswal, B., Yetkin, F., Haughton, V., and Hyde, J. (1995), "Functional Connectivity in the Motor Cortex of Resting Human Brain Using Echo-Planar MRI," *Magnetic Resonance in Medicine,* 34, 537–541.

Black, F., and Scholes, M. (1973), "The Pricing of Options and Corporate Liabilities," *The Journal of Political Economy,* 81, 637–654.

Brüggemann, R., Lütkepohl, H., and Saikkonen, P. (2006), "Residual Autocorrelation Testing for Vector Error Correction Models," *Journal of Econometrics,* 134, 579–604.

Brumback, B., and Rice, J. A. (1998), "Smooting Spline Models for the Analysis of Nested and Crossed Samples of Curves," *Journal of the American Statistical Association,* 93, 961–994.

Connor, G., Hagmann, M., and Linton, O. (2007). *Efficient Semiparametric Estimation of the Fama-French Model and Extensions*, Preprint.

Connor, G., and Linton, O. (2007), "Semiparametric Estimation of a Characteristic-based Factor Model of Stock Returns," *Journal of Empirical Finance,* 14, 694–717.

Cont, R., and da Fonseca, J. (2002), "The Dynamics of Implied Volatility Surfaces," *Quantitative Finance,* 2, 45–60.

de Boor, C. (2001). *A Practical Guide to Splines*, Berlin, Heidelberg: Springer-Verlag.

Diebold, F. X., and Li, C. (2006), "Forecasting the Term Structure of Government Bond Yields," *Journal of Econometrics,* 130, 337–364.

Fama, E. F., and French, K. R. (1992), "The Cross-Section of Expected Stock Returns," *The Journal of Finance,* 47, 427–465.

Fan, J., Yao, Q., and Cai, Z. (2003), "Adaptive Varying-Coefficient Linear Models," *Journal of the Royal Statistical Society:* Series B, 65, 57–80.

Fengler, M. R., Härdle, W., and Mammen, E. (2007), "A Semiparametric Factor Model for Implied Volatility Surface Dynamics," *Journal of Financial Econometrics,* 5, 189–218.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000), "The Generalized Dynamic Factor Model: Identification and Estimation," *The Review of Economics and Statistics,* 82, 540–554.

Forni, M., and Lippi, M. (2001), "The Generalized Factor Model: Representation Theory," *Econometric Theory,* 17, 1113–1141.

Gasser, T., Möcks, R., and Verleger, R. (1983), "Selavco: A Method to Deal With Trial-to-Trial Variability of Evoked Potential," *Electroencephalography and Clinical Neurophysiology,* 55, 717–723.

Hafner, R. (2004). *Stochastic Implied Volatility*, Berlin: Springer.

Hallin, M., and Liska, R. (2007), "Determining the Number of Factors in the Generalized Dynamic Factor Model," *Journal of the American Statistical Association,* 102, 603–617.

Hansen, L.H., Nielsen, B., and Nielsen, J.P. (2004). "Two Sided Analysis of the Variance With a Latent Time Series," Nuffield College Economic Working Paper 2004-W25, University of Oxford.

Hosking, J. R. M. (1980), "The Multivariate Portmanteau Statistic," *Journal of the American Statistical Association,* 75, 602–608.

Hosking, J. R. M. (1981), "Lagrange-Multiplier Tests of Multivariate Time-Series Models," *Journal of the Royal Statistical Society,* Series B, 43, 219–230.

Kantorovich, L. V., and Akilov, G. P. (1982). *Functional Analysis* (2nd ed.), Oxford, U.K: Pergamon Press.

Kauermann, G. (2000), "Modeling Longitudinal Data With Ordinal Response by Varying Coefficients," *Biometrics,* 56, 1692–1698.

Lee, R. D., and Carter, L. (1992), "Modeling and Forecasting the Time Series of U.S. Mortality," *Journal of the American Statistical Association,* 87, 659–671.

Logothetis, N., and Wandell, B. (2004), "Interpreting the Bold Signal," *Annual Review of Physiology,* 66, 735–769.

Lütkepohl, H. (1993). *Intorduction to Multiple Time Series Analysis*, Berlin, Heidelberg: Springer-Verlag.

Martinussen, T., and Scheike, T. (2000), "A Nonparametric Dynamic Additive Regression Model for Longitudinal Data," *Annals of Statistics,* 28, 1000–1025.

Nelson, C. R., and Siegel, A. F. (1987), "Parsimonious Modeling of Yield Curves," *Journal of Business,* 60, 473–489.

Peña, D., and Box, E. P. (1987), "Identifying a Simplifying Structure in Time Series," *Journal of the American Statistical Association,* 82, 836–843.

Stock, J. H., and Watson, M.W. (2005). "Implications of Dynamic Factor Models for VAR Analysis," NBER Working Papers 11467, National Bureau of Economic Research, Inc., available at http://ideas.repec.org/p/nbr/nberwo/11467.html.

Van de Geer, S. (2000). *Empirical Processes in M-Estimation*, Cambridge, U.K.: Cambridge University Press.

Vrahatis, M. N. (1989), "A Short Proof and a Generalization of Miranda's Existence Theorem," *Proceedings of the American Mathematical Society,* 107, 701–703.

Worsley, K., Liao, C., Aston, J., Petre, V., Duncan, G., Morales, F., and Evans, A. (2002), "A General Statistical Analysis for fMRI Data," *NeuroImange,* 15, 1–15.

Yang, L., Park, B. U., Xue, L., and Härdle, W. (2006), "Estimation and Testing for Varying Coefficients in Additive Models With Marginal Integration," *Journal of the American Statistical Association,* 101, 1212–1227.