# Time Series Modelling with Semiparametric Factor Dynamics *

Szymon Borak

CASE – Center for Applied Statistics and Economics
Humboldt-Universität zu Berlin,
Spandauer Straße 1, 10178 Berlin, Germany


Wolfgang Härdle

CASE – Center for Applied Statistics and Economics
Humboldt-Universität zu Berlin,
Spandauer Straße 1, 10178 Berlin, Germany


Enno Mammen

Department of Economics, University of Mannheim
L 7, 3–5, 68131 Mannheim, Germany


Byeong U. Park

Department of Statistics, Seoul National University
Seoul 151-747, Korea

April 25, 2007

---

# Time Series Modelling with Semiparametric Factor Dynamics

## Abstract

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 paper we address the problem of inference when the factors and factor loadings are estimated by semiparametric methods. This more flexible modelling approach poses an important question: Is it justified, from 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 discuss other possible applications in finance and economics.

# 1 Introduction

Modelling 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 size. Such modelling challenges appear in many different fields. In empirical macroeconomics one is interested in analyzing the dynamics of the plethora of economic indicators that reflect the state of the economy, see Stock and Watson (2005). In mortality analysis one builds a model for forecasting death rate or life expectancy, as is done in Lee and Carter (1992). The statistical analysis of financial term structure dynamics is crucial for bond portfolio risk management or derivative pricing, see Nelson and Siegel (1987) and Diebold and Li (2006). In bio-medical research the interest can be focussed on time improvement of the patients' treatment, as in the children growth history analysis by Martinussen and Scheike (2000). Other examples include the studies of radiation treatment of prostate cancer by Kauermann (2000) and evoked potentials in EEG analysis by Gasser et al. (1983). In financial engineering, it is common to analyze the dynamics of implied volatility surface for pricing exotic options.

A successful modelling 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 $Y_t = (Y_{t,1}, \ldots, Y_{t,J})^\top$ can be represented as

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

where $Z_{t,l}$ are common factors, $\varepsilon_{t,j}$ 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 multi-dimensional 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 modelling of $Z_t = (Z_{t,1}, \ldots, Z_{t,L})^\top$, which is a more feasible task when $L \ll J$.

In a variety of applications, one has explanatory variables $X_{t,j}$ 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
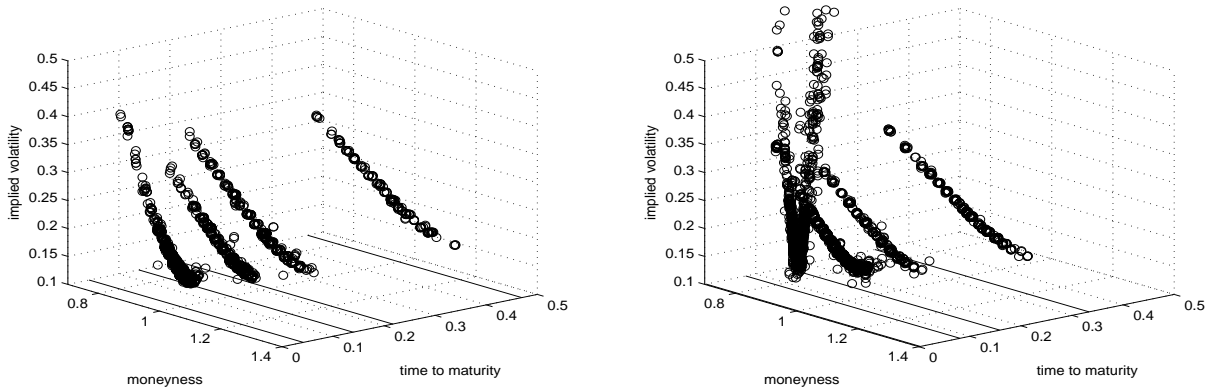
Figure 1: *The typical IV data design on two different days. In the maturity direction observations appear in the discrete points for each particular day. Bottom solid lines indicate the observed maturities, which move towards the expiry. Left panel: observations on 20040701, $J_t = 5606$. Right panel: observations on 20040819, $J_t = 8152$.*

of $X_{t,j}$, so that the model (1) is generalized to

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

The model (2) can be regarded as a regression model with embedded time evolution. In particular, lagged observations of the component series $(Y_{t-1,j}, \ldots, Y_{t-p,j})^\top$ may be the regressor $X_{t,j}$, but we allow here any external variable $X_t = (X_{t,1}, \ldots, X_{t,J})^\top$. Additionally, the regularity of the multi-dimensional time series may be omitted by allowing $J$ depending on $t$, say $J_t$. The model (2) is different from varying-coefficient models, such as in Fan et al. (2003) and Yang et al. (2006), since $Z_t$ is unobservable. Our model also has some similarities to the one considered in Connor and Linton (2007) and Connor et al. (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 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 et al. (2004) and Brumback and Rice (1998), may be recognized as a special case of (2).

In this paper we consider the model (2) with nonparametric functions $m_l$. We call this

4

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 paper, 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.

One of the main motivations for the model (2) comes from a special structure of the implied volatility (IV) data. 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). Enforcing relatively simple assumptions their model and pricing formula have achieved enormous popularity, and it is now a common market standard to communicate in terms of IV. Due to the trading regulation one may observe, on a particular day, only those options with certain times-to-maturity. This feature causes a typical 'string' structure in the IV data, which can be clearly observed in Figure 1. The data were obtained from the European option prices on the German stock index DAX (ODAX). The plots represent the values of IV calculated from option trades with maturities and moneyness–a monotone transformation of strike price. While the observations lie relatively dense in the moneyness direction, they appear only at a few discrete points in the maturity direction. Moreover, the volatility strings shift towards expiry. This is indicated by the bottom lines in the figure, where one may also observe that the maturities of the options on the two different days are different. The latter means that the observed design points changes each day.

These IV dynamics may be represented well through the model (2). In this particular example, $Y_{t,j}$ 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 the analysis of IV data, Hafner (2004) considered a stochastic implied volatility model which assumes a parametric form for the functions $m_l$. This method, however, may incur a significant model-bias when the actual model is far from the assumed parametric one. Cont and da Fonseca (2002) applied the Karhunen-Loéve decomposition in functional principal component analysis to the daily variations of IV surfaces that are estimated on the whole design space by a kernel method. This approach depends crucially on the initial estimation of IV surfaces, which may fail due to the design-sparseness that we discussed in the previous paragraph. DSFM is a flexible semiparametric model which neither assumes any parametric form for the functions

$m_l$ nor requires the initial smoothing as in Cont and da Fonseca (2002). The model has been used by Fengler et al. (2007) to study the dynamic behavior of implied volatility surfaces obtained from the ODAX data.

DSFM, fitted to the IV data, has numerous applications. First, the model can be seen as a low dimensional projection of the IV risk factors. Certain exotic options, whose prices strongly depend on the whole IV surface, have to be hedged additionally against undesirable surface movements. This, so called 'vega hedging', can be refined by adjusting the hedging portfolio according to the new hedging ratios, defined as the derivatives of the option price with respect to $Z_t$. Second, the model provides a useful tool for distributional forecasting of future IV, which is essential in risk management of large options' portfolios. It also serves for pricing certain exotic options as an extension of the Monte Carlo pricing methodology.

Our methods produce estimates of the true unobservable $Z_t$, say $\widehat{Z}_t$, 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 applications that we discuss in the above paragraph, one needs to fit an econometric model to the estimated factors $\widehat{Z}_t$. 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 $\widehat{Z}_t$ is equivalent to the one based on $Z_t$. Attempting to give an answer to this question forms the core of this paper.

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 $m = (m_0, \ldots, m_L)^\top$, 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 $\{\widehat{Z}_t\}$ whose lagged covariances are asymptotically the same as those of $\{Z_t\}$. This justifies the inference based on $\{\widehat{Z}_t\}$ 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 data.

Although DSFM has been motivated from analyzing IV surfaces, it may be applied to many other problems. A prominent example is modelling of yield curve evolution. Here, the standard approach is to use the parametric factor model proposed by Nelson and Siegel (1987), where the empirical form of the yield curve is fitted with some pre-specified functions

6

of the bonds' maturities. A possible refinement is the penalized spline smoothing, employed simultaneously in time and maturity dimensions, adopted by Krivobokova et al. (2006). Apart from this, other modelling strategies based on principal component analysis or factor models are commonly used, see Rebonato (1998), Bliss (1997) and Molgedey and Galic (2001) among others. Clearly, DSFM enhances flexibility of these approaches for modelling the term structure of interest rates. In a similar manner, DSFM may be employed to analyze the term structure of the variance swaps as in Detlefsen and Härdle (2006), and to study future prices of $CO_2$ emission allowances as in Trück et al. (2006). Other examples include mortality trend fitting, where the current standard is to use the model proposed by Lee and Carter (1992). In that model the age-specific death rates are regressed additively on a time-invariant age-specific component and another age-specific component multiplied by a time-varying factor. Here, one can let the age specific components be nonparametric functions of some particular covariates and extend the model to the reduced DSFM with $L = 1$. The mortality forecast is then obtained by a standard statistical method applied to the estimated factor time series.

For each $t$, the set of observations $\mathcal{Y}_t = \{Y_{t,j} : 1 \le j \le J_t\}$ may be viewed as representing discretized values of a smooth surface $\mathcal{S}_t$. Therefore interest may be similarly put on the functional objects $\mathcal{S}_t$, which has a direct linkage to the functional data analysis (Ramsay and Silverman, 1997). The standard practice is to obtain an estimate of $\mathcal{S}_t$ from $\mathcal{Y}_t$ and proceed to build up a factor model as in Cont and da Fonseca (2002). However this procedure requires an initial fit for $\mathcal{S}_t$, which may suffer from the design-sparseness problem as we mentioned earlier. DSFM avoids this preliminary estimation and shifts the discrete representation directly to the functions $m_l$.

The paper 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 data and discuss some further applications. 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

$$Y_{t,j} = Z_t^\top m(X_{t,j}) + \varepsilon_{t,j}. \tag{3}$$

Here $Z_t = (Z_{t,0}, \ldots, Z_{t,L})^\top$ is an unobservable $L + 1$-dimensional process with $Z_{t,0} \equiv 1$. The function $m$ is a tuple $(m_0, \ldots, m_L)$ of unknown real-valued functions $m_j$ 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 $\varepsilon_{t,j}$ 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 normed so that $\int_{[0,1]^d} \psi_k^2(x) \, dx = 1$. For example, one may take $\{\psi_k : 1 \leq k \leq K\}$ to be a tensor B-spline basis, see e.g. de Boor (2001). Then, a tuple of functions $m = (m_0, \ldots, m_L)^\top$ may be approximated by $A\psi$, where $A = (\alpha_{l,k})$ is a $(L + 1) \times K$ matrix and $\psi = (\psi_1, \ldots, \psi_K)^\top$. We define the least squares estimators $\widehat{Z}_t = (\widehat{Z}_{t,0}, \ldots, \widehat{Z}_{t,L})^\top$ and $\widehat{A} = (\widehat{\alpha}_{l,k})$:

$$S(A, Z) \equiv \sum_{t=1}^{T} \sum_{j=1}^{J} \left\{ Y_{t,j} - Z_t^\top A \psi(X_{t,j}) \right\}^2 = \min_{A,Z}! \tag{4}$$

where $Z = \left( Z_1^\top, \ldots, Z_T^\top \right)^\top$. The minimization runs over all values of $\widehat{Z}_t$ with

$$Z_{t,0} = 1. \tag{5}$$

With $\widehat{A}$ at hand, we estimate $m$ by $\widehat{m} = \widehat{A}\psi$.

We note that, given $Z$ or $A$, the function $S$ in (4) is quadratic with respect to the other variables, and thus has an explicit unique minimizer. However, minimization of $L$ with respect to $A$ and $Z$ simultaneously is a fourth-order problem. The solution is neither unique nor explicit. It is unique only up to the values of $\widehat{Z}_1^\top \widehat{A}, \ldots, \widehat{Z}_T^\top \widehat{A}$. We will come back to this identifiability issue later in this section.

To find a solution $(\widehat{A}, \widehat{Z})$ of the minimization problem (4), one might adopt the following iterative algorithm: (i) Given an initial choice $Z^{(0)}$, minimize $S(A, Z^{(0)})$ with respect to $A$,

which is an ordinary least squares problem and thus has an explicit unique solution. Call it $A^{(1)}$. (ii) Minimize $S(A^{(1)}, Z)$ 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 $\alpha \equiv \alpha(A)$ denote the stack form of $A = (\alpha_{l,k})$, 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 $S(A, Z)$. Define

$$F_{10}(\alpha, Z) = \frac{\partial}{\partial \alpha} S(\alpha, Z), \quad F_{01}(\alpha, Z) = \frac{\partial}{\partial Z} S(\alpha, Z),$$

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

Let $\Psi_t = \left[ \psi(X_{t,1}), \ldots, \psi(X_{t,J}) \right]$ be a $K \times J$ matrix. Then, it can be shown that

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

$$F_{20}(\alpha, Z) = 2 \sum_{t=1}^{T} \left[ (\Psi_t \Psi_t^\top) \otimes (Z_t Z_t^\top) \right],$$

$$F_{01}(\alpha, Z) = 2 \begin{pmatrix} A\Psi_1 \Psi_1^\top A^\top Z_1 - A\Psi_1 Y_1 \\ A\Psi_2 \Psi_2^\top A^\top Z_2 - A\Psi_2 Y_2 \\ \vdots \\ A\Psi_T \Psi_T^\top A^\top Z_T - A\Psi_T Y_T \end{pmatrix},$$

$$F_{02}(\alpha, Z) = 2 \begin{pmatrix} A\Psi_1 \Psi_1^\top A^\top & 0 & \cdots & 0 \\ 0 & A\Psi_2 \Psi_2^\top A^\top & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & A\Psi_T \Psi_T^\top A^\top \end{pmatrix}.$$

Here and below, $\otimes$ denotes the Kronecker product operator. Also, by some algebraic manipulations it can be shown that

$$\left[ (\Psi_t \Psi_t^\top) \otimes (Z_t Z_t^\top) \right] \alpha = (\Psi_t \Psi_t^\top A^\top Z_t) \otimes Z_t.$$

9

Thus, we get

$$F_{11}(\alpha, Z) = 2\left(F_{11,1}(\alpha, Z),\ F_{11,2}(\alpha, Z),\ \ldots,\ F_{11,T}(\alpha, Z)\right),$$

where

$$F_{11,t}(\alpha, Z) = (\Psi_t \Psi_t^\top A^\top) \otimes Z_t + (\Psi_t \Psi_t^\top A^\top Z_t) \otimes I_{L+1} - (\Psi_t Y_t) \otimes I_{L+1}$$

and $I_q$ denotes the identity matrix of dimension $q$. Let

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

We need to solve the equation $F(\alpha, Z) = 0$ simultaneously for $\alpha$ and $Z$. Given $(\alpha^{\text{OLD}}, Z^{\text{OLD}})$, the Newton-Raphson algorithm gives the updating equation for $(\alpha^{\text{NEW}}, Z^{\text{NEW}})$:

$$\begin{pmatrix} \alpha^{\text{NEW}} \\ Z^{\text{NEW}} \end{pmatrix} = \begin{pmatrix} \alpha^{\text{OLD}} \\ Z^{\text{OLD}} \end{pmatrix} - F'(\alpha^{\text{OLD}}, Z^{\text{OLD}})^{-1} F(\alpha^{\text{OLD}}, Z^{\text{OLD}}). \tag{6}$$

The algorithm (6) is shown to converge to a solution of (4) at a geometric rate under some week conditions on the initial choice $(\alpha^{(0)}, Z^{(0)})$, as is demonstrated by Theorem 2.1 below. We collect the conditions for the theorem.

(C1) $F'(\alpha^{(0)}, Z^{(0)})$ is invertible.

(C2) There exists a solution $(\widehat{\alpha}, \widehat{Z})$ of the equation $F(\alpha, Z) = 0$ such that $\sum_{t=1} \widehat{Z}_t \widehat{Z}_t^\top$ and $\sum_{t=1} \widehat{Z}_t Z_t^{(0)\top}$ are invertible, and $\widehat{\alpha}_l = (\widehat{\alpha}_{l1}, \ldots, \widehat{\alpha}_{lK})^\top$ for $l = 0, \ldots, L$ are linearly independent, i.e., the matrix $\widehat{A}$ that corresponds to $\widehat{\alpha}$ has full rank.

Let $\alpha^{(k)}$ and $Z^{(k)}$ denote the $k$th updated vectors in the iteration with the algorithm (6). Also, we write $A^{(k)}$ for the matrix that corresponds to $\alpha^{(k)}$.

**Theorem 2.1.** *Suppose that the initial choice $(\alpha^{(0)}, Z^{(0)})$ satisfies (C1) and (C2). Then, for any $0 < \gamma < 1$ there exists $r > 0$ and $C > 0$ such that, if $\sum_{t=1}^T \|Z_t^{(0)\top} A^{(0)} - \widehat{Z}_t^\top \widehat{A}\|^2 \le r$, then*

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

The minimization problem (4) has no unique solution. If $(\widehat{Z}_t, \widehat{A})$ or $(\widehat{Z}_t, \widehat{m} = \widehat{A}\psi)$ is a minimizer, then also $(\widetilde{B}^\top \widehat{Z}_t, \widetilde{B}^{-1}\widehat{m})$ is a minimizer. Here $\widetilde{B}$ is an arbitrary matrix of the form

$$\widetilde{B} = \begin{pmatrix} 1 & 0 \\ 0 & B \end{pmatrix} \tag{7}$$

for an invertible matrix $B$. The special structure of $\widetilde{B}$ assures that the first component of $\widetilde{B}^\top \widehat{Z}_t$ equals 1. 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 $\widetilde{B}$ such that $\widetilde{Z}_t = \widetilde{B}^\top \widehat{Z}_t$ 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 $\widehat{m} = \widehat{A}\psi$ such that the components $\widehat{m}_1, \ldots, \widehat{m}_L$ 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 $\widehat{m}$ 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.

The following proposition will be used to prove Theorem 2.1, but it has an important implication in its own right. Let $(\alpha^{\mathrm{OLD}}, Z^{\mathrm{OLD}})$ be the 'old' value to be updated by (6) and $(\alpha^{\mathrm{NEW}}, Z^{\mathrm{NEW}})$ be the updated value as defined there. For an $(L+1) \times (L+1)$ nonsingular square matrix $\widetilde{B}$, define

$$\widetilde{A}^{\mathrm{OLD}} = \widetilde{B}^{-1} A^{\mathrm{OLD}}, \quad \widetilde{Z}_t^{\mathrm{OLD}} = \widetilde{B}^\top Z_t^{\mathrm{OLD}}.$$

The stack form of $\widetilde{A}^{\mathrm{OLD}}$ can be written as

$$\widetilde{\alpha}^{\mathrm{OLD}} = (I_K \otimes \widetilde{B}^{-1})\alpha^{\mathrm{OLD}}. \tag{8}$$

Also, we have

$$\widetilde{Z}^{\mathrm{OLD}} = (\widetilde{Z}_1^{\mathrm{OLD}\top}, \ldots, \widetilde{Z}_T^{\mathrm{OLD}\top})^\top = (I_T \otimes \widetilde{B}^\top)Z^{\mathrm{OLD}}. \tag{9}$$

The proposition demonstrates that updating $\widetilde{\alpha}^{\mathrm{OLD}}$ and $\widetilde{Z}^{\mathrm{OLD}}$ directly by the formula (6) is equivalent to first updating $\alpha^{\mathrm{OLD}}$ and $Z^{\mathrm{OLD}}$ by (6) and then transforming the updated vectors $\alpha^{\mathrm{NEW}}$ and $Z^{\mathrm{NEW}}$ according to (8) and (9), respectively.

**Proposition 2.2.** *Let $\widetilde{\alpha}^{\mathrm{NEW}}$ and $\widetilde{Z}^{\mathrm{NEW}}$ be obtained from $\widetilde{\alpha}^{\mathrm{OLD}}$ and $\widetilde{Z}^{\mathrm{OLD}}$ by the updating equation (6). Then $\widetilde{\alpha}^{\mathrm{NEW}} = (I_K \otimes \widetilde{B}^{-1})\alpha^{\mathrm{NEW}}$ and $\widetilde{Z}^{\mathrm{NEW}} = (I_T \otimes \widetilde{B}^\top)Z^{\mathrm{NEW}}.$*

PROOF. Given $(\alpha, Z)$, let $\widetilde{\alpha} = (I_K \otimes \widetilde{B}^{-1})\alpha$ and $\widetilde{Z} = (I_T \otimes \widetilde{B}^\top)Z$. We prove

$$F'(\widetilde{\alpha}, \widetilde{Z})^{-1} F(\widetilde{\alpha}, \widetilde{Z}) = \begin{pmatrix} I_K \otimes \widetilde{B}^{-1} & O \\ O & I_T \otimes \widetilde{B}^\top \end{pmatrix} F'(\alpha, Z)^{-1} F(\alpha, Z). \tag{10}$$

First, we collect several useful algebraic identities. Let $M$ be an arbitrary $u \times v$ matrix. For matrices $P$, $Q$, $R$ where $PQR$ is defined,

$$M \otimes (PQR) = (I_u \otimes P)(M \otimes Q)(I_v \otimes R). \tag{11}$$

In particular, for a vector $q$ where $Pq$ is defined, we have from (11) by taking $R = 1$ that

$$M \otimes (Pq) = (I_u \otimes P)(M \otimes q). \tag{12}$$

On the other hand, for a vector $q$ and a matrix $P$ where $MP$ is defined, we obtain

$$(MP) \otimes q = (M \otimes q)P. \tag{13}$$

Finally, for an $r \times r$ invertible square matrix $Q$

$$M \otimes I_r = (I_u \otimes Q)(M \otimes I_r)(I_v \otimes Q^{-1}). \tag{14}$$

For simplicity, write $F_{ij} = F_{ij}(\alpha, Z)$ and $\widetilde{F}_{ij} = F(\widetilde{\alpha}, \widetilde{Z})$ for $(i, j) = (1, 0), (0, 1), (2, 0), (1, 1)$ and $(0, 2)$. Using the facts (10)–(14), it can be shown that

$$\begin{aligned} \widetilde{F}_{20} &= (I_K \otimes \widetilde{B}^\top) F_{20} (I_K \otimes \widetilde{B}), \\ \widetilde{F}_{02} &= (I_T \otimes \widetilde{B}^{-1}) F_{02} (I_T \otimes (\widetilde{B}^\top)^{-1}), \\ \widetilde{F}_{11} &= (I_K \otimes \widetilde{B}^\top) F_{11} (I_T \otimes (\widetilde{B}^\top)^{-1}). \end{aligned} \tag{15}$$

From (15) it can be also seen that

$$\widetilde{F}_{02} - \widetilde{F}_{11}^\top \widetilde{F}_{20}^{-1} \widetilde{F}_{11} = (I_T \otimes \widetilde{B}^{-1}) \left( F_{02} - F_{11}^\top F_{20}^{-1} F_{11} \right) (I_T \otimes (\widetilde{B}^\top)^{-1}). \tag{16}$$

For each of $(i, j) = (2, 0), (1, 1), (0, 2)$, let $G_{ij}$ denote the block matrix of $F'(\alpha, Z)^{-1}$ whose the dimension equals that of $F_{ij}$. Likewise, define $\widetilde{G}_{ij}$ with $F'(\widetilde{\alpha}, \widetilde{Z})^{-1}$. Then, from (15), (16) and the inversion formula for a partitioned matrix, we obtain

$$\begin{aligned} \widetilde{G}_{20} &= (I_K \otimes \widetilde{B}^{-1}) G_{20} (I_K \otimes (\widetilde{B}^\top)^{-1}), \\ \widetilde{G}_{02} &= (I_T \otimes \widetilde{B}^\top) G_{02} (I_T \otimes \widetilde{B}), \\ \widetilde{G}_{11} &= (I_K \otimes \widetilde{B}^{-1}) G_{11} (I_T \otimes \widetilde{B}). \end{aligned} \tag{17}$$

The proposition follows immediately from (17) and the facts $\widetilde{F}_{10} = (I_K \otimes \widetilde{B}^\top) F_{10}$ and $\widetilde{F}_{01} = (I_T \otimes \widetilde{B}^{-1}) F_{01}$. $\qquad\square$

# 3 Simulation Study

In Section 5 we will argue that the inference based on the covariances of the unobserved factors $Z_t$ is asymptotically equivalent to the one based on $\widetilde{B}^\top \widehat{Z}_t$ for some invertible $\widetilde{B}$, see Eq. (34). In this section we illustrates the equivalence by a simulation study. For this, we compare the covariances of $Z_t$ and $\widetilde{Z}_t \equiv \widetilde{B}^\top \widehat{Z}_t$.

We took $T = 500, 1000, 2000$, $J = 100, 250, 1000$ and $K = 36, 49, 64$. We considered $d = 2, L = 3$ and the following tuple of 2-dimensional functions:

$$\begin{pmatrix} m_0 \\ m_1 \\ m_2 \\ m_3 \end{pmatrix} (x_1, x_2) = \begin{pmatrix} 1 \\ 3.46(x_1 - \frac{1}{2}) \\ 9.45\left\{(x_1 - \frac{1}{2})^2 + (x_2 - \frac{1}{2})^2\right\} - 1.6 \\ 1.41\sin(2\pi x_2) \end{pmatrix}. \tag{18}$$
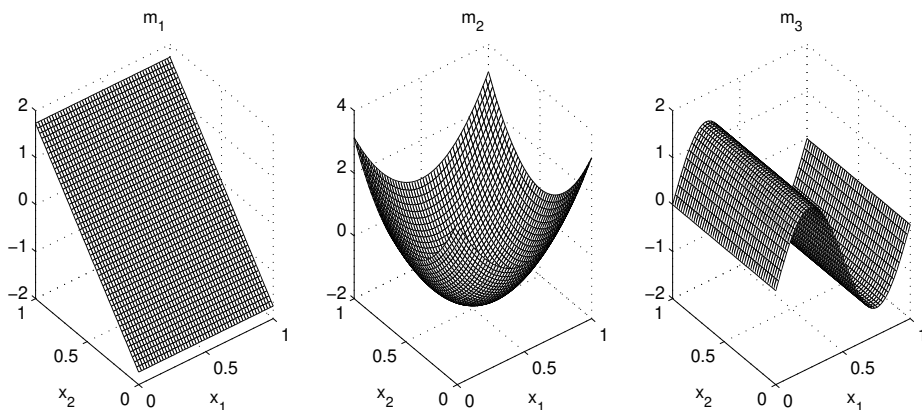


Figure 2: *True functions $m_1, m_2, m_3$ from which the data were generated.*

The coefficients in (18) were chosen so that $m_1, m_2, m_3$ are close to orthogonal. The factor loading functions are displayed in Figure 2. We generated $Z_t$ from a centered VAR(1) process $Z_t = \mathcal{A}Z_{t-1} + U_t$, where $U_t$ is $N_3(0, \Sigma_U)$ random vector and

$$\mathcal{A} = \begin{pmatrix} 0.95 & -0.2 & 0 \\ 0 & 0.8 & 0.1 \\ 0.1 & 0 & 0.6 \end{pmatrix}, \quad \Sigma_U = \begin{pmatrix} 10^{-4} & 0 & 0 \\ 0 & 10^{-4} & 0 \\ 0 & 0 & 10^{-4} \end{pmatrix}.$$

13

The design points $X_{t,j}$ were independently generated from a uniform distribution on the unit square, $\varepsilon_{t,j}$ were i.i.d. $N(0, \sigma^2)$ with $\sigma = 0.05$, and $Y_{t,j}$ were obtained according to the model (3). The simulation experiment was repeated 250 times for each combination of $(T, J, K)$.
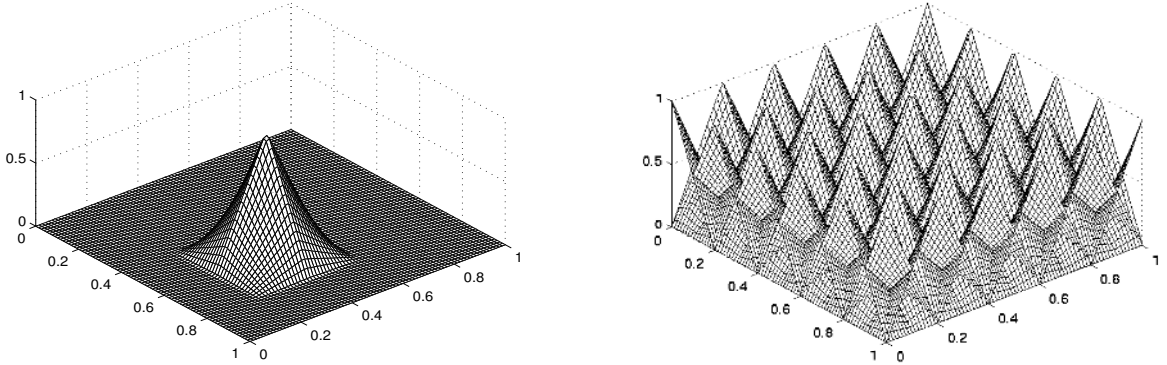


Figure 3: *Tensor linear B-spline basis used in the estimation. Left panel: one particular basis function $\psi_k$. Right panel: the whole set of basis functions for $K = 36$.*

For the estimation we employed, for $\psi_j$, 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. The tensor spline basis functions in the case $K = 36$ are plotted in Figure 3. In the simulation we increased the number of the basis functions $K$ to 49 and 64, which correspond to more dense layouts in the right panel of Figure 3.

We plotted in Figure 4 the entries of the scaled difference of the covariance matrices

$$\widetilde{D} = \frac{1}{\sqrt{T}} \left\{ \sum_{t=1}^{T} \left( \widetilde{Z}_t - \overline{\widetilde{Z}} \right) \left( \widetilde{Z}_t - \overline{\widetilde{Z}} \right)^\top - \sum_{t=1}^{T} \left( Z_t - \overline{Z} \right) \left( Z_t - \overline{Z} \right)^\top \right\}. \tag{19}$$

Each panel of Figure 4 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 $T = 500, 1000, 2000$. In the figure we also depicted, by thick lines, the 95% and 5% quantiles of

$$D = \frac{1}{\sqrt{T}} \left\{ \sum_{t=1}^{T} \left( Z_t - \overline{Z} \right) \left( Z_t - \overline{Z} \right)^\top - \Gamma \right\}, \tag{20}$$
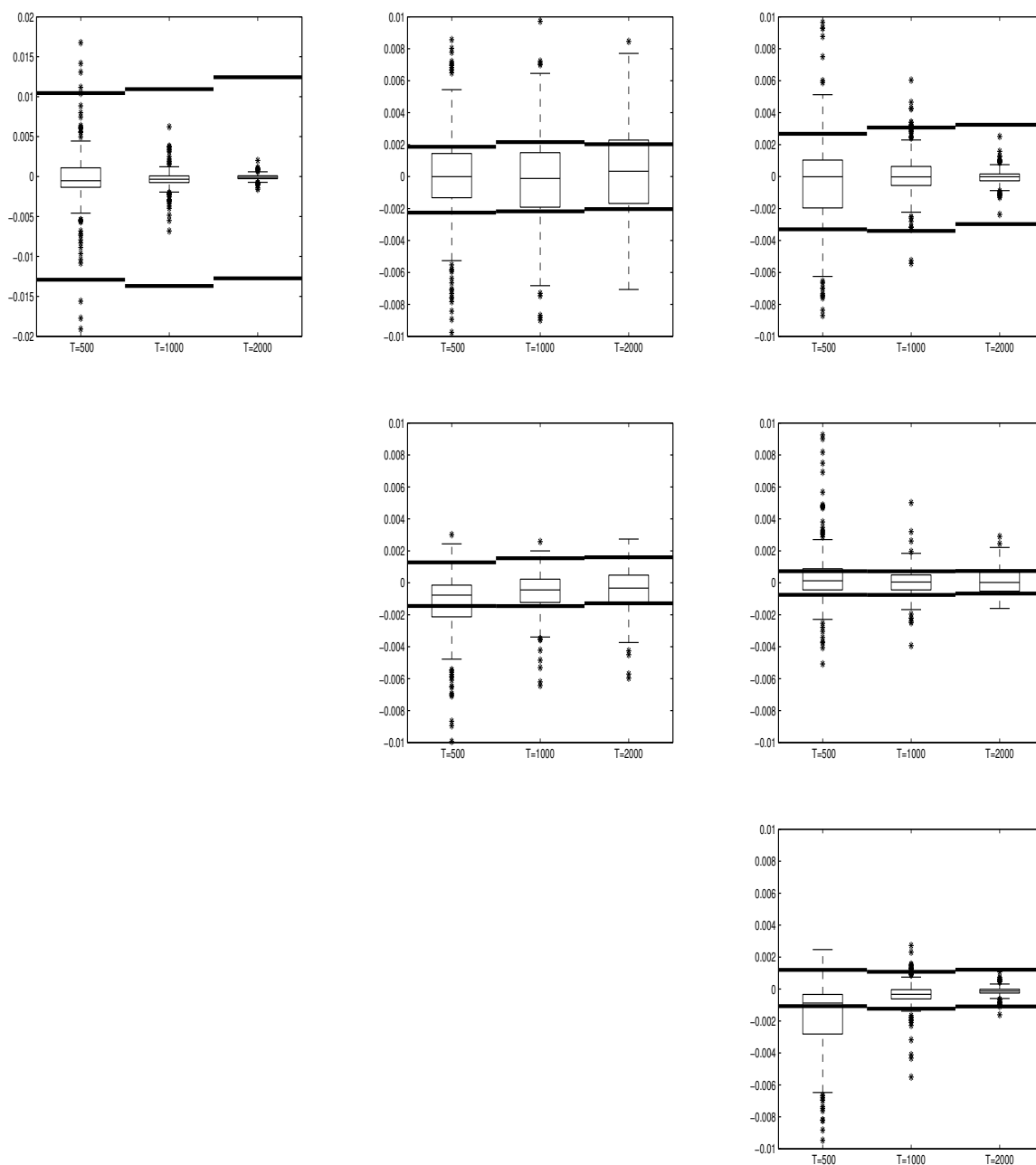
14

Figure 4: *The boxplots based on 250 values of the entries of the scaled difference of the covariance matrices given at (19). The lengths of the series $Z_t$ and $\widetilde{Z}_t$ were $500, 1000, 2000$. The thick lines represent the 95% and 5% quantiles of (20).*

15

where $\Gamma$ is the true covariance matrix of the simulated VAR process. It is known that $\Gamma$ can be represented as $\gamma = (I_{L^2} - \mathcal{A} \otimes \mathcal{A})^{-1}\sigma_U$, where $\gamma$ and $\sigma_U$ are the stack forms of $\Gamma$ and $\Sigma_U$ respectively. We refer to Lütkepohl (1993) for more details.

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 non-degenerate law as $T \to \infty$. This is well supported by the plots in Figure 4 showing that the distance between the two thick lines in each panel is 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 $\widetilde{D}$ as well as their interquartile ranges get shorter as $T$ increases.

# 4  Applications

This section presents an application of DSFM. We fit the model to the intra-day IV based on ODAX prices and discuss related issues. We also give a brief overview for other possible applications.

For our analysis we chose the the data observed from July 1st 2004 to June 29th 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 $\kappa_t$ 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 EURIBOR rate taken from Ecowin Reuters database. The time-to-maturity of the option 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 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.

Since the model is not nested, the number of the dynamic functions needs to be determined in advance. Based on computed values

$$RV(L) = \frac{\sum_t^T \sum_j^{J_t} \{Y_{t,j} - \sum_{l=0}^L \widehat{Z}_{t,l} \widehat{m}_l(X_{t,j})\}^2}{\sum_t^T \sum_j^{J_t} (Y_{t,j} - \bar{Y})^2}, \tag{21}$$

which are given in Table 1 for various $L$, we chose $L = 2$. The quantity $1 - RV(L)$ can be interpreted as a proportion of the variation explained by the model among the total variation. Table 1 indicates that the third, fourth and fifth factor make only a small improvement in the fit.

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

Table 1: *Proportion of the explained variation by the models with $L = 1, \ldots, 5$ dynamic factors.*

For the series estimators of $\widehat{m}_l$ we used tensor B-splines that are cubic in the moneyness and quadratic in the maturity direction. In the transformed space we placed $10 \times 5$ 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, see Table 2. Since the model is identifiable only up to the transformation (7), one has a freedom for the choice of factors. Here, we chose $\widehat{m}_1$ and $\widehat{m}_2$ that are orthogonal to each other in $L_2[0, 1]^2$ in such a way that $\sum_{t=1}^T \widehat{Z}_{t,1}^2$ is maximized, as is described in Fengler et al. (2007).

| knots | | order | | $1 - RV(2)$ |
| moneyness | maturity | moneyness | maturity | |
|---|---|---|---|---|
| 15 | 10 | 3 | 2 | 0.974 |
| 10 | 10 | 3 | 2 | 0.972 |
| 10 | 5 | 3 | 2 | 0.969 |
| 5 | 5 | 3 | 2 | 0.961 |
| 15 | 10 | 2 | 2 | 0.972 |
| 10 | 10 | 2 | 2 | 0.969 |
| 10 | 5 | 2 | 2 | 0.965 |
| 5 | 5 | 2 | 2 | 0.951 |
| 15 | 10 | 2 | 1 | 0.971 |
| 10 | 10 | 2 | 1 | 0.968 |
| 10 | 5 | 2 | 1 | 0.967 |
| 5 | 5 | 2 | 1 | 0.949 |

Table 2: *Proportion of the explained variation by the models when $L = 2$ for different numbers of knots and different orders of splines.*

The estimated functions $\widehat{m}_1$ and $\widehat{m}_2$ are plotted in Figure 5 in the transformed estimation space. The intercept function $\widehat{m}_0$ is almost flat around zero, thus is not given. By construction, $\widehat{m}_0 + \widehat{Z}_{t,1}\widehat{m}_1$ explain the principal movements of the surface. It was observed by Cont and da Fonseca (2002) 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 6 shows the values of VDAX together with $\widehat{m}_0(X_{t,0}) + \widehat{Z}_{t,1}\widehat{m}_1(X_{t,0})$, 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 6 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
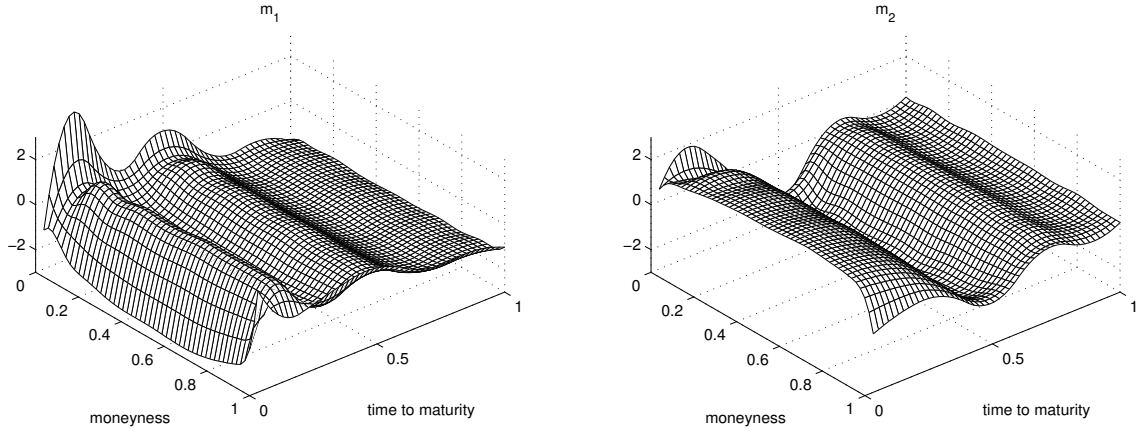
Figure 5: *The estimated factor functions for the ODAX IV data from 20040701 to 20050629.*
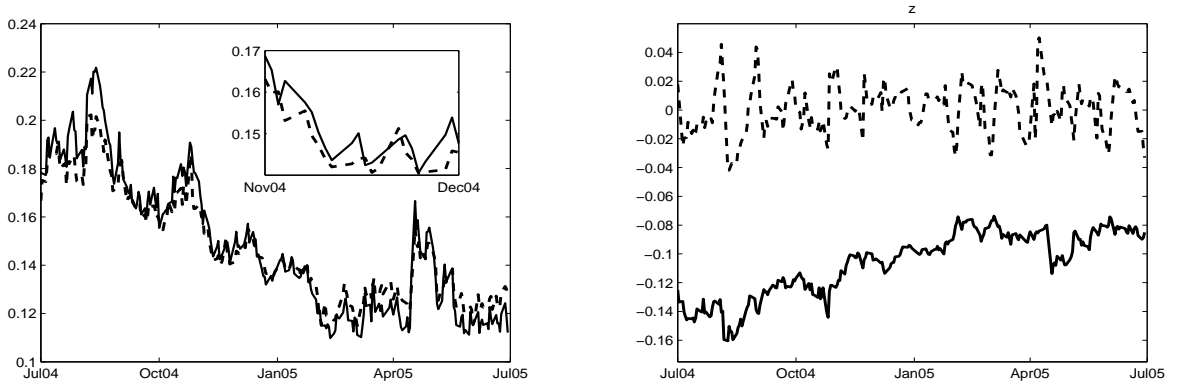


Figure 6: *Left panel: VDAX from 20040701 to 20050629 (solid) and the dynamics of the corresponding IV given by the sub-model $\widehat{m}_0 + \widehat{Z}_{t,1}\widehat{m}_1$ (dashed). Right panel: The obtained time series $\widehat{Z}_t$ on the ODAX IV data from 20040701 to 20050629. The solid line represents $\widehat{Z}_{t,1}$, the dashed line $\widehat{Z}_{t,2}$.*

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 Table 3. One can find that SC and HQ suggest a VAR(1) process, while AIC selects VAR(2). The parameter estimates for each selected model are given in Table 4. The roots of the characteristic polynomial lie

| 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 |

Table 3: *The VAR model selection criteria. The smallest value for each criterion is marked by (\*).*

| | 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 |

Table 4: *The estimated parameters for VAR(1) and VAR(2) models.*

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 et al. (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 7. But, this effect disappears in the VAR(2) case, see Figure 8. Similar analyses of characteristic polynomials, portmanteau and LM tests supported VAR(2) as a successful model for $\widehat{Z}_t$.

Although DSFM is motivated from modelling IV data, it can be applied to many other problems. A number of possible applications have been already discussed in Section 1. We close this section by adding two other applications of DSFM. One is to analyze some
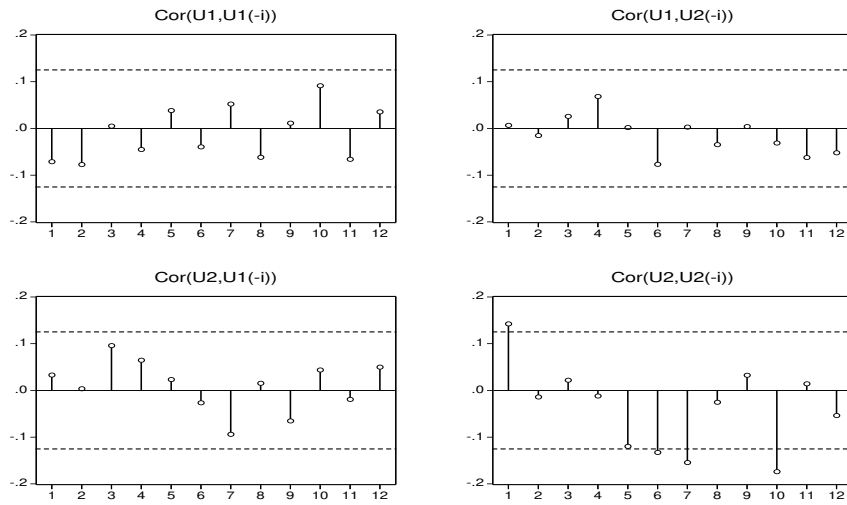
Figure 7: *Cross-autocorrelogram for the VAR(1) residuals. The dashed line-bounds indicate $\pm 2 \times$ (standard deviations), which correspond to an approximate 95% confidence bound.*
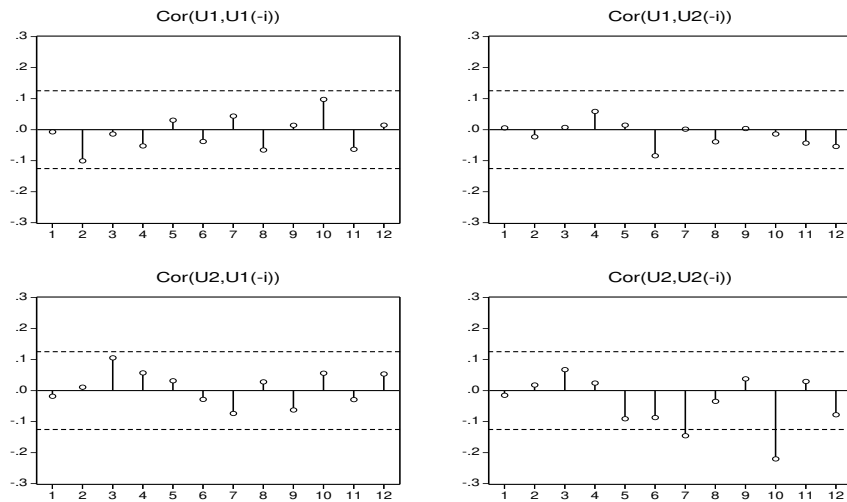


Figure 8: *Cross-autocorrelogram for the VAR(2) residuals. The dashed line-bounds indicate $\pm 2 \times$ (standard deviations), which correspond to an approximate 95% confidence bound.*

important characteristics of state price density (SPD). SPD describes the distribution of the risk-neutral price of an underlying asset at some future date. The knowledge of SPD

can be beneficial for an objective interpretation of market expectations. The SPD dynamics may also be modelled via (3), where the functions $m_l$ are one-dimensional nonparametric functions of the future underlying states. They can be described in terms of a symmetric or asymmetric contribution of the probability to reflect the changes in skewness and kurtosis over time. The other application is the analysis of book orders. It allows to recover at any time the supply and demand structure, usually given as *bid* and *ask* curves. These curves may serve to construct a measure of liquidity. A tradable asset is called 'liquid' if it is possible to exchange a big volume 'rapidly' without too much damage to the price of the asset. In practical term this means that the difference between the bid and ask prices (as a function of volume) is not too large. One can obtain for every day a surface of bid-ask differences as a function of time and volume. DSFM may help here in quantifying the dynamics of this liquidity measure.

# 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. For the asymptotic analysis, we let $K, J, T \to \infty$. Our first result relies on the following assumptions.

(A1) The variables $X_{1,1}, \ldots, X_{T,J}, \varepsilon_{1,1}, \ldots, \varepsilon_{T,J}$ are independent.

(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

$$\begin{aligned} \mathrm{E}[\varepsilon_{t,j}] &= 0 \text{ for } t = 1, \ldots, T, j = 1, \ldots, J, \\ \sup_{t=1,\ldots,T,j=1,\ldots,J} \mathrm{E}[\varepsilon_{t,j}^2] &< \infty. \end{aligned}$$

(A4) The functions $\psi_k$ 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 $k = 1, \ldots, K$.

(A5) The components $m_0, \ldots, m_L$ can be approximated by $\psi_1, \ldots, \psi_K$, i.e.

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

as $K \to \infty$. We denote a matrix that fulfills $\sup_{x \in [0,1]^d} |m(x) - A\psi(x)| \le 2\delta_K$ by $A^*$. We assume that $\delta_K = \mathcal{O}(K^{1/2} J^{-1/2})$ for $K, J \to \infty$.

(A6) There exist constants $0 < C_L < C_U < \infty$ such that all eigenvalues of the random matrix $T^{-1} \sum_{t=1}^T Z_t Z_t^\top$ lie in the interval $[C_L, C_U]$ with probability tending to one.

(A7) It holds that $(K \log K)/J \to 0$ and $\log T/J \to 0$.

(A8) The minimization (4) runs over all values of $(A, z)$ with $z_{t,0} = 1$ and with

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

where the constant $M_T$ fulfils $\max_{1 \le t \le T} \|Z_t\| \le M_T/C_m$ for a constant $C_m$ such that $\sup_{x \in [0,1]} \|m(x)\| < C_m$, and $M_T^2(K \log K/J) \to 0$, $M_T^2(\log T/J) \to 0$.

Condition (A8) and the additional bound $M_T$ in the minimization is introduced for purely technical reasons.

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

**Theorem 5.1.** *Suppose that model (3) holds and that $(\widehat{Z}_t, \widehat{A})$ is defined by the minimization problem (4) under the constraint (5). Make the assumptions (A1)–(A8). Then it holds that*

$$\frac{1}{T} \sum_{1 \le t \le T} \left\| \widehat{Z}_t^\top \widehat{A} - Z_t^\top A^* \right\|^2 = \mathcal{O}_P(KJ^{-1}). \tag{22}$$

The proof of Theorem 5.1 is given in Section A.2.

At this point we have made no assumptions on the sequence $Z_t : 1 \le t \le T$, besides the bound in (A8). 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) The bound $\max_{1 \leq t \leq T} \|Z_t\| \leq M_T$ holds with probability tending to one, and it holds that $M_T^2 \{(K \log K)/J\} \to 0$ and $M_T^2 (\log T/J) \to 0$.

(A10) $Z_t$ is strictly stationary with $\mathrm{E}(Z_t) = 0$ and $\mathrm{E}\|Z_t\|^\gamma < \infty$ for some $\gamma > 2$. It is strongly mixing with $\sum_{i=1}^\infty \alpha(i)^{(\gamma-2)/\gamma} < \infty$. The matrix $\mathrm{E}Z_t Z_t^\top$ has full rank. The process $Z_t$ is independent of $X_{11}, \ldots, X_{TJ}, \varepsilon_{11}, \ldots, \varepsilon_{TJ}$.

(A11) The functions $m_0, \ldots, m_L$ are linearly independent. In particular, no function is equal to 0. Furthermore, it holds that $\sup_{x \in [0,1]} \|\psi(x)\| = \mathcal{O}(K^{1/2})$.

(A12) It holds that $K/J + \delta_K = o(T^{-1/2})$, $\log T = o(K)$, $K^5 J^{-4}(\log K)^2 = o(T^{-1})$, and $K^7 J^{-5}(\log K)^2 = o(T^{-1})$.

Condition (A10) implies that $T^{-1} \sum_{t=1}^T Z_t$ has a bounded second moment, see e.g. Corollary 1.1 in Bosq (1998).

**Theorem 5.2.** *Suppose that model (3) holds and that $(\widehat{Z}_t, \widehat{A})$ is defined by the minimization problem (4) under the constraint (5). Make the assumptions (A1)–(A12). Then there exists a random matrix $\widetilde{B}$, which is of the form (7), such that for $h \geq 0$*

$$\frac{1}{T} \sum_{t=h+1}^T \left( \widetilde{Z}_t - \overline{\overline{Z}} \right) \left( \widetilde{Z}_{t-h} - \overline{\overline{Z}} \right)^\top - \frac{1}{T} \sum_{t=h+1}^T \left( Z_t - \overline{Z} \right) \left( Z_{t-h} - \overline{Z} \right)^\top = \mathcal{O}_P(T^{-1/2}),$$

$$\overline{\overline{Z}} - \overline{Z} = \mathcal{O}_P(T^{-1/2}),$$

*where $\widetilde{Z}_t = \widetilde{B}^\top \widehat{Z}_t$, $\overline{\overline{Z}} = T^{-1} \sum_{t=1}^T \widetilde{Z}_t$ and $\overline{Z} = T^{-1} \sum_{t=1}^T Z_t$.*

To illustrate an implication of Theorem 5.2, suppose that the factor loading process $Z_t$ in (3) is a stationary VAR(p) process in a mean adjusted form:

$$Z_t - \mu = \Theta_1(Z_{t-1} - \mu) + \cdots + \Theta_p(Z_{t-p} - \mu) + U_t, \tag{23}$$

where $\mu = E(Z_t)$, $\Theta_j$ is a $L \times L$ matrix of coefficients and $U_t$ is a white noise with some nonsingular covariance matrix $\Sigma_U$. Here, we take the $L$ components of $Z_t$, omitting the first $Z_{t,0} \equiv 1$. In a slight abuse of notation, we continue to refer to the resulting vector as $Z_t$.

Let $\Gamma_h$ 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 (23) can be rewritten as $Y = \Theta W + U$ and the least squares estimator of $\Theta$ 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 $\mu$ is replaced by $\overline{Z}$. Likewise, fitting a VAR(p) model with the estimated factor loading 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 5.2, we have $\sqrt{T}(\widetilde{\Theta} - \widehat{\Theta}) = o_P(1)$.

# A    Appendix

## A.1    Proof of Theorem 2.1.

We use the following lemma to prove the theorem.

**Lemma A.1.** (Newton-Kantorovich) *Let $X$ and $Y$ be Banach spaces and $F : D \subset X \to Y$. Suppose that on an open convex set $D_0 \subset D$, $F$ is Fréchet differentiable and*

$$\|F'(x) - F'(y)\| \leq \gamma \|x - y\|, \quad x, y \in D_0.$$

*Let $x_0 \in D_0$ be the initial point. Assume that $F'(x_0)^{-1}$ is defined on all of $Y$ and*

$$\|F'(x_0)^{-1}\| \leq \alpha, \quad \|F'(x_0)^{-1} F(x_0)\| \leq \beta$$

*for some constants $\alpha$ and $\beta$ such that $q \equiv 2\alpha\beta\gamma \leq 1$. Define*

$$t^* = \frac{1}{\alpha\gamma}(1 - \sqrt{1-q}), \quad t^{**} = \frac{1}{\alpha\gamma}(1 + \sqrt{1-q}),$$

*and assume that* $S \equiv \{x \big| \|x - x_0\| \leq t^*\} \subset D_0$. *Then, the Newton iterates*

$$x_{k+1} = x_k - F'(x_k)^{-1} F(x_k), \quad k = 0, 1, \ldots$$

*are well defined, lie in* $S$ *and converge to a solution* $x^*$ *of* $F(x) = 0$ *at a geometric rate:*

$$\|x_k - x^*\| \leq \beta \, 2^{-(k-1)} q^{2^k - 1}.$$

*When* $q < 1$, *the solution* $x^*$ *is unique in* $D_0 \cap \{x \big| \|x - x_0\| < t^{**}\}$. *When* $q = 1$ *(in this case* $t^* = t^{**}$), *the solution* $x^*$ *is unique in* $D_0 \cap \{x \big| \|x - x_0\| \leq t^{**}\}$.

The proof of the lemma can be found in Kantorovich and Akilov (1982), for example.

We now prove the theorem. Suppose that $\sum_{t=1}^{T} \|Z_t^{(0)\top} A^{(0)} - \widehat{Z}_t^\top \widehat{A}\|^2 \leq r$ for some $r > 0$ which will be chosen later. Define

$$B_0 = \left( \sum_{t=1}^{T} \widehat{Z}_t Z_t^{(0)\top} \right)^{-1} \sum_{t=1}^{T} \widehat{Z}_t \widehat{Z}_t^\top.$$

Also, define $\widetilde{A}^{(0)} = B_0^{-1} A^{(0)}$ and $\widetilde{Z}_t^{(0)} = B_0^\top Z_t^{(0)}$. Let $\widetilde{\alpha}^{(0)}$ be the stack form of $\widetilde{A}^{(0)}$, i.e., $\widetilde{\alpha}^{(0)} = (I_K \otimes B_0^{-1}) \alpha^{(0)}$. Note that $\sum_{t=1}^{T} \widehat{Z}_t \widehat{Z}_t^\top = \sum_{t=1}^{T} \widehat{Z}_t \widetilde{Z}_t^{(0)\top}$. With the Frobenius norm $\|M\|$ for a matrix $M$, we get

$$
\begin{aligned}
\|\widetilde{\alpha}^{(0)} - \widehat{\alpha}\|^2 &= \|\widetilde{A}^{(0)} - \widehat{A}\|^2 \\
&\leq \left\| \left( \sum_{t=1}^{T} \widehat{Z}_t \widehat{Z}_t^\top \right)^{-1} \right\|^2 \cdot \left\| \sum_{t=1}^{T} \widehat{Z}_t \widehat{Z}_t^\top (\widetilde{A}^{(0)} - \widehat{A}) \right\|^2 \\
&= \left\| \left( \sum_{t=1}^{T} \widehat{Z}_t \widehat{Z}_t^\top \right)^{-1} \right\|^2 \cdot \left\| \sum_{t=1}^{T} \widehat{Z}_t \widetilde{Z}_t^{(0)\top} \widetilde{A}^{(0)} - \sum_{t=1}^{T} \widehat{Z}_t \widehat{Z}_t^\top \widehat{A} \right\|^2 \quad (24) \\
&\leq \left\| \left( \sum_{t=1}^{T} \widehat{Z}_t \widehat{Z}_t^\top \right)^{-1} \right\|^2 \left( \sum_{t=1}^{T} \left\| \widehat{Z}_t Z_t^{(0)\top} A^{(0)} - \widehat{Z}_t \widehat{Z}_t^\top \widehat{A} \right\| \right)^2 \\
&\leq r \left\| \left( \sum_{t=1}^{T} \widehat{Z}_t \widehat{Z}_t^\top \right)^{-1} \right\|^2 \left( \sum_{t=1}^{T} \|\widehat{Z}_t\|^2 \right) \\
&\equiv r \, c_1.
\end{aligned}
$$

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

get

$$
\begin{aligned}
\|\widehat{A}^\top (\widetilde{Z}_t^{(0)} - \widehat{Z}_t)\| &\geq \|\widehat{A}\|_2^{-1} \cdot \|\widehat{A}\widehat{A}^\top (\widetilde{Z}_t^{(0)} - \widehat{Z}_t)\| \\
&\geq \|\widehat{A}\|_2^{-1} \cdot \|(\widehat{A}\widehat{A}^\top)^{-1}\|^{-1} \cdot \|\widetilde{Z}_t^{(0)} - \widehat{Z}_t\|.
\end{aligned}
\tag{25}
$$

On the other hand,

$$
\begin{aligned}
\|(\widetilde{Z}_t^{(0)} - \widehat{Z}_t)^\top \widehat{A}\| &\leq \|\widetilde{Z}_t^{(0)\top}(\widehat{A} - \widetilde{A}^{(0)})\| + \|\widetilde{Z}_t^{(0)\top}\widetilde{A}^{(0)} - \widehat{Z}_t^\top \widehat{A}\| \\
&\leq \|\widetilde{Z}_t^{(0)}\| \cdot \|\widehat{A} - \widetilde{A}^{(0)}\| + \|Z_t^{(0)\top}A^{(0)} - \widehat{Z}_t^\top \widehat{A}\|.
\end{aligned}
\tag{26}
$$

The two inequalities (25) and (26) together with (24) give

$$
\begin{aligned}
\|\widetilde{Z}^{(0)} - \widehat{Z}\|^2 = \sum_{t=1}^T \|\widetilde{Z}_t^{(0)} - \widehat{Z}_t\|^2 &\leq 2\,r\,\|\widehat{A}\|_2^2 \cdot \|(\widehat{A}\widehat{A}^\top)^{-1}\|^2 \\
&\quad \times \left[ 1 + \Big(\sum_{t=1}^T \|\widetilde{Z}_t^{(0)}\|^2\Big)\Big(\sum_{t=1}^T \|\widehat{Z}_t\|^2\Big)\Big\|\Big(\sum_{t=1}^T \widehat{Z}_t\widehat{Z}_t^\top\Big)^{-1}\Big\|^2 \right] \\
&\equiv r\,c_2.
\end{aligned}
\tag{27}
$$

Since $F'(\alpha, Z)$ is quadratic in $(\alpha, Z)$, there exists $0 < c_3 < \infty$ for any compact set $D$ in $\mathbb{R}^{(K+T)(L+1)}$ 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 $c_4 = \|F'(\widetilde{\alpha}^{(0)}, \widetilde{Z}^{(0)})^{-1}\|_2 < \infty$. Since $F$ is continuous and $F(\widehat{\alpha}, \widehat{Z}) = 0$, there exists $r' > 0$ such that, if $\|\widetilde{\alpha}^{(0)} - \widehat{\alpha}\|^2 + \|\widetilde{Z}^{(0)} - \widehat{Z}\|^2 \leq r'$, then

$$
\|F'(\widetilde{\alpha}^{(0)}, \widetilde{Z}^{(0)})^{-1} F(\widetilde{\alpha}^{(0)}, \widetilde{Z}^{(0)})\| \leq \frac{\gamma}{2c_3 c_4}.
$$

By Lemma A.1, the Newton iterates $(\widetilde{\alpha}^{(k)}, \widetilde{Z}^{(k)})$ based on the algorithm (6) starting from $(\widetilde{\alpha}^{(0)}, \widetilde{Z}^{(0)})$ converges to $(\widehat{\alpha}, \widehat{Z})$ at a geometric rate:

$$
\|\widetilde{\alpha}^{(k)} - \widehat{\alpha}\|^2 + \|\widetilde{Z}^{(k)} - \widehat{Z}\|^2 \leq C_1 2^{-(k-1)} \gamma^{2^k - 1}
\tag{28}
$$

for some $C_1 > 0$. By Proposition 2.2,

$$
\widetilde{\alpha}^{(k)} = (I_K \otimes B_0^{-1})\alpha^{(k)}, \quad \widetilde{Z}^{(k)} = (I_T \otimes B_0^\top)Z^{(k)},
$$

or equivalently

$$
\widetilde{A}^{(k)} = B_0^{-1}A^{(k)}, \quad \widetilde{Z}_t^{(k)} = B_0^\top Z_t^{(k)}
$$

27

for $k \geq 1$. This and (28) give that, if $\|\widetilde{\alpha}^{(0)} - \widehat{\alpha}\|^2 + \|\widetilde{Z}^{(0)} - \widehat{Z}\|^2 \leq r'$, then

$$
\begin{aligned}
\sum_{t=1}^{T} \|Z_t^{(k)\top} A^{(k)} - \widehat{Z}_t^\top \widehat{A}\|^2 & = \sum_{t=1}^{T} \|\widetilde{Z}_t^{(k)\top} \widetilde{A}^{(k)} - \widehat{Z}_t^\top \widehat{A}\|^2 \\
& \leq C_2 \left( \|\widetilde{\alpha}^{(k)} - \widehat{\alpha}\|^2 + \|\widetilde{Z}^{(k)} - \widehat{Z}\|^2 \right) \\
& \leq C2^{-(k-1)} \gamma^{2^k - 1}
\end{aligned}
$$

for some $C, C_2 > 0$. We take $r = (c_1 + c_2)^{-1} r'$. Then, by (24) and (27), $\|\widetilde{\alpha}^{(0)} - \widehat{\alpha}\|^2 + \|\widetilde{Z}^{(0)} - \widehat{Z}\|^2 \leq r'$ if $\sum_{t=1}^{T} \|Z_t^{(0)\top} A^{(0)} - \widehat{Z}_t^\top \widehat{A}\|^2 \leq r$. This completes the proof of the theorem. $\qquad \square$

## A.2 Proof of Theorem 5.1.

We write $d_t^\top = \widehat{Z}_t^\top \widehat{A} - Z_t^\top A^*$. We have to show that

$$
\frac{1}{T} \sum_{1 \leq t \leq T} \|d_t\|^2 = \mathcal{O}_P(KJ^{-1} + \delta_K^2). \tag{29}
$$

For the proof of this claim we first note that by definition of our estimator it holds that

$$
\sum_{t=1}^{T} \sum_{j=1}^{J} \left\{ Y_{t,j} - \widehat{Z}_t^\top \widehat{A} \psi(X_{t,j}) \right\}^2 \leq \sum_{t=1}^{T} \sum_{j=1}^{J} \left\{ Y_{t,j} - Z_t^\top A^* \psi(X_{t,j}) \right\}^2. \tag{30}
$$

This implies that

$$
\sum_{t=1}^{T} \sum_{j=1}^{J} \left\{ d_t^\top \psi(X_{t,j}) \right\}^2 \leq 2 \sum_{t=1}^{T} \sum_{j=1}^{J} d_t^\top \psi(X_{t,j}) \varepsilon_{t,j} \tag{31}
$$

$$
+ 2 \sum_{t=1}^{T} \sum_{j=1}^{J} d_t^\top \psi(X_{t,j}) \left\{ Z_t^\top m(X_{t,j}) - Z_t^\top A^* \psi(X_{t,j}) \right\}. \tag{32}
$$

With $\|d\|_{J,T}^2 = T^{-1} J^{-1} \sum_{1 \leq j \leq J, 1 \leq t \leq T} \left\{ d_t^\top \psi(X_{t,j}) \right\}^2$ and $\|d\|^2 = T^{-1} \sum_{1 \leq t \leq T} \|d_t\|^2$ this gives with (A5)

$$
\|d\|_{J,T}^2 \leq 2\|d\| \left\{ \frac{1}{T} \sum_{t=1}^{T} \left\| \frac{1}{J} \sum_{j=1}^{J} \psi(X_{t,j}) \varepsilon_{t,j} \right\|^2 \right\}^{1/2} + 4\|d\|_{J,T} \delta_K. \tag{33}
$$

28

We now argue that with $c, C > 0$ small enough it holds for all vectors $e \in \mathcal{E}(M_T)$ with $\mathcal{E}(C) = \{e \in \mathbb{R}^K : \|e\| = 1 \text{ and } \sup_{x \in [0,1]} |e^\top \psi(x)| \le C\}$ and for $1 \le t \le T$

$$P\left[J^{-1} \sum_{1 \le j \le J} \left\{e^\top \psi(X_{t,j})\right\}^2 \le c\right] \le \exp(-CJ/M_T^2).$$

This follows by application of Bernstein's inequality, see also (A4). Now we use that for all $C' > 0$ the unit sphere in $\mathbb{R}^K$ can be covered by $o\{(C')^{-K} K^K\}$ balls with radius $C' K^{-1}$. Furthermore, it holds almost surely with a constant $C''$ (not depending on $C'$):

$$\sup_{\|e\|=\|f\|=1, \|e-f\| \le C'K^{-1}} \left| J^{-1} \sum_{1 \le j \le J} \left\{e^\top \psi(X_{t,j})\right\}^2 - J^{-1} \sum_{1 \le j \le J} \left\{f^\top \psi(X_{t,j})\right\}^2 \right| \le C''C'.$$

This implies that with a constant $C'''$

$$P\left[\min_{1 \le t \le T} \inf_{e \in \mathcal{E}(M_T)} J^{-1} \sum_{1 \le j \le J} \left\{e^\top \psi(X_{t,j})\right\}^2 \le c/2\right] \le C'''(C')^{-K} K^K T \exp(-CJ/M_T^2).$$

The right hand side converges to 0 because of (A8). This shows that

$$\|d\|_{J,T}^2 \ge \frac{c}{2}\|d\|^2$$

with probability tending to one. Therefore we get from (33) that with probability tending to one

$$\sqrt{\frac{c}{2}}\|d\| \le \|d\|_{J,T} \le c\left\{\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{J}\sum_{j=1}^{J}\psi(X_{t,j})\varepsilon_{t,j}\right\|^2\right\}^{1/2} + 4\delta_K.$$

The right hand side is of order $\mathcal{O}_P(K^{1/2} J^{-1/2})$. This concludes the proof of Theorem 5.1.

## A.3 Proof of Theorem 5.2.

We choose $\widetilde{B}$ as

$$\widetilde{B} = \left(T^{-1}\sum_{t=1}^{T} Z_t \widehat{Z}_t^\top\right)^{-1} T^{-1}\sum_{t=1}^{T} Z_t Z_t^\top. \tag{34}$$

For this definition we have to check that the matrix $T^{-1}\sum_{t=1}^{T} Z_t \widehat{Z}_t^\top$ is invertible. We suppose that this is not the case and we define a random vector $e$ (depending on $T$) with $\|e\| = 1$

29

and $e^\top \sum_{t=1}^T Z_t \widehat{Z}_t^\top = 0$. We now use that

$$\left\| T^{-1} \sum_{t=1}^T Z_t \widehat{Z}_t^\top \widehat{A} - T^{-1} \sum_{t=1}^T Z_t Z_t^\top A^* \right\| \tag{35}$$

$$\leq T^{-1} \sum_{t=1}^T \left\| Z_t \widehat{Z}_t^\top \widehat{A} - Z_t Z_t^\top A^* \right\|$$

$$= T^{-1} \sum_{t=1}^T \left\| Z_t \left( \widehat{Z}_t^\top \widehat{A} - Z_t^\top A^* \right) \right\|$$

$$\leq T^{-1} \sum_{t=1}^T \| Z_t \| \left\| \widehat{Z}_t^\top \widehat{A} - Z_t^\top A^* \right\|$$

$$\leq \left( T^{-1} \sum_{t=1}^T \| Z_t \|^2 \right)^{1/2} \left( T^{-1} \sum_{t=1}^T \left\| \widehat{Z}_t^\top \widehat{A} - Z_t^\top A^* \right\|^2 \right)^{1/2}$$

$$= \mathcal{O}_P(K^{1/2}/J^{1/2}),$$

because of (A6) and Theorem 5.1. This gives with $f = T^{-1} \sum_{t=1}^T Z_t Z_t^\top e$,

$$\| f^\top m \| = \| f^\top (A^* \psi) \| + \mathcal{O}(\delta_K) = \mathcal{O}_P(K^{1/2}/J^{1/2})$$

and this would imply that $m_0, \ldots, m_d$ are linearly dependent, in contrast to assumption (A11). Note also that $\| T^{-1} \sum_{t=1}^T Z_t Z_t^\top - E Z_t Z_t^\top \| = \mathcal{O}_P(T^{-1/2}) = o_P(1)$, because of (A6).

We now define $\widetilde{Z}_t = \widetilde{B}^\top \widehat{Z}_t$ and $\widetilde{A} = \widetilde{B}^{-1} \widehat{A}$. Then $\widetilde{Z}_t^\top \widetilde{A} = \widehat{Z}_t^\top \widehat{A}$ and $T^{-1} \sum_{t=1}^T Z_t \widetilde{Z}_t^\top = T^{-1} \sum_{t=1}^T Z_t Z_t^\top$. This gives with (35)

$$\left\| \widetilde{A} - A^* \right\| = \left\| T^{-1} \sum_{t=1}^T Z_t Z_t^\top (\widetilde{A} - A^*) \right\| \mathcal{O}_P(1) \tag{36}$$

$$= \left\| T^{-1} \sum_{t=1}^T Z_t \widetilde{Z}_t^\top \widetilde{A} - T^{-1} \sum_{t=1}^T Z_t Z_t^\top A^* \right\| \mathcal{O}_P(1)$$

$$= \mathcal{O}_P(K^{1/2}/J^{1/2}).$$

Furthermore, from (A11), (A5),(36) and Theorem 5.1 one gets that

$$T^{-1} \sum_{t=1}^{T} \left\| \widetilde{Z}_t - Z_t \right\|^2 \tag{37}$$

$$= T^{-1} \sum_{t=1}^{T} \left\| \widetilde{Z}_t^\top m - Z_t^\top m \right\|^2 \mathcal{O}_P(1)$$

$$= T^{-1} \sum_{t=1}^{T} \left\| \widetilde{Z}_t^\top A^* - Z_t^\top A^* \right\|^2 \mathcal{O}_P(1) + \mathcal{O}_P(\delta_K^2)$$

$$= T^{-1} \sum_{t=1}^{T} \left\| \widetilde{Z}_t^\top \widetilde{A} - \widetilde{Z}_t^\top A^* \right\|^2 \mathcal{O}_P(1) + T^{-1} \sum_{t=1}^{T} \left\| \widetilde{Z}_t^\top \widetilde{A} - Z_t^\top A^* \right\|^2 \mathcal{O}_P(1) + \mathcal{O}_P(\delta_K^2)$$

$$= T^{-1} \sum_{t=1}^{T} \left\| \widetilde{Z}_t \right\|^2 \left\| \widetilde{A} - A^* \right\|^2 \mathcal{O}_P(1) + \mathcal{O}_P(K/J)$$

$$= T^{-1} \sum_{t=1}^{T} \left\| \widetilde{Z}_t - Z_t \right\|^2 \left\| \widetilde{A} - A^* \right\|^2 \mathcal{O}_P(1) + \left\| \widetilde{A} - A^* \right\|^2 \mathcal{O}_P(1) + \mathcal{O}_P(K/J)$$

$$= \mathcal{O}_P(K/J).$$

We will show that for $h \geq 0$

$$T^{-1} \sum_{t=h+1}^{T} \left( \widetilde{Z}_t - Z_t \right) Z_{t-h}^\top = o_P(T^{-1/2}). \tag{38}$$

Because of $Z_{t,0} \equiv 1$, the equation (38) with the choice $h = 0$ implies that

$$T^{-1} \sum_{t=1}^{T} \left( \widehat{Z}_t - Z_t \right) = o_P(T^{-1/2}).$$

These two equations imply the statement of Theorem 5.2, because of

$$T^{-1} \sum_{t=h}^{T} \left( \widetilde{Z}_t - Z_t \right) \left( \widetilde{Z}_{t-h}^\top - Z_{t-h}^\top \right) = \mathcal{O}_P(K/J) = o_P(T^{-1/2}),$$

see (A12).

For the proof of (38) we use the representation $Z_t^\top A^* \psi(X_{t,j}) = \{\psi(X_{t,j}) \otimes Z_t\}^\top \alpha^*$, where $\alpha^*$ is the stack form of $A^*$. Similarly, $\widetilde{Z}_t^\top \widetilde{A} \psi(X_{t,j}) = \{\psi(X_{t,j}) \otimes \widetilde{Z}_t\}^\top \widetilde{\alpha}$, where $\widetilde{\alpha}$ is the stack

form of $\widetilde{A}$. Now, by definition

$$\widetilde{Z}_t \;=\; \widetilde{S}_{t,Z}^{-1} J^{-1} \sum_{j=1}^{J} Y_{t,j} \widetilde{A} \psi(X_{t,j}), \tag{39}$$

$$\widetilde{\alpha} \;=\; \widetilde{S}_{\alpha}^{-1} T^{-1} J^{-1} \sum_{t=1}^{T} \sum_{j=1}^{J} \{\psi(X_{t,j}) \otimes \widetilde{Z}_t\} Y_{t,j}, \tag{40}$$

where

$$\widetilde{S}_{t,Z} \;=\; J^{-1} \sum_{j=1}^{J} \widetilde{A} \psi(X_{t,j}) \psi(X_{t,j})^\top \widetilde{A}^\top,$$

$$\widetilde{S}_{\alpha} \;=\; T^{-1} J^{-1} \sum_{t=1}^{T} \sum_{j=1}^{J} \{\psi(X_{t,j}) \otimes \widetilde{Z}_t\} \{\psi(X_{t,j}) \otimes \widetilde{Z}_t\}^\top.$$

We now argue that

$$\|\widetilde{S}_{t,Z} - S_{t,Z}\| \;=\; o_P(J^{1/2} K^{-1/2} T^{-1/2}), \tag{41}$$

$$\|\widetilde{S}_{\alpha} - S_{\alpha}\| \;=\; o_P(J^{1/2} K^{-1/2} T^{-1/2}), \tag{42}$$

where

$$S_{t,Z} \;=\; A^* E\left\{\psi(X_{t,j}) \psi(X_{t,j})^\top\right\} A^{*\top},$$

$$S_{\alpha} \;=\; T^{-1} \sum_{t=1}^{T} E\left[\{\psi(X_{t,j}) \otimes Z_t\} \{\psi(X_{t,j}) \otimes Z_t\}^\top \big| Z_t\right].$$

We show (41). Claim (42) can be shown similarly. For the proof of (41) it suffices to show

that

$$J^{-1} \sum_{j=1}^{J} A^* \left[ \psi(X_{t,j}) \psi(X_{t,j})^\top - E\left\{ \psi(X_{t,j}) \psi(X_{t,j})^\top \right\} \right] \left( \widetilde{A} - A^* \right)^\top \tag{43}$$

$$= o_P(J^{1/2} K^{-1/2} T^{-1/2}),$$

$$J^{-1} \sum_{j=1}^{J} \left( \widetilde{A} - A^* \right) \left[ \psi(X_{t,j}) \psi(X_{t,j})^\top - E\left\{ \psi(X_{t,j}) \psi(X_{t,j})^\top \right\} \right] \left( \widetilde{A} - A^* \right)^\top \tag{44}$$

$$= o_P(J^{1/2} K^{-1/2} T^{-1/2}),$$

$$J^{-1} \sum_{j=1}^{J} A^* \left[ \psi(X_{t,j}) \psi(X_{t,j})^\top - E\left\{ \psi(X_{t,j}) \psi(X_{t,j})^\top \right\} \right] A^{*\top} \tag{45}$$

$$= o_P(J^{1/2} K^{-1/2} T^{-1/2}),$$

$$J^{-1} \sum_{j=1}^{J} A^* E\left\{ \psi(X_{t,j}) \psi(X_{t,j})^\top \right\} \left( \widetilde{A} - A^* \right)^\top \tag{46}$$

$$= o_P(J^{1/2} K^{-1/2} T^{-1/2}),$$

$$J^{-1} \sum_{j=1}^{J} \left( \widetilde{A} - A^* \right) E\left\{ \psi(X_{t,j}) \psi(X_{t,j})^T \right\} \left( \widetilde{A} - A^* \right)^\top \tag{47}$$

$$= o_P(J^{1/2} K^{-1/2} T^{-1/2}).$$

The proof of (45)–(47) follows by simple arguments, note also that because of (A12) it holds that $K^{1/2} J^{-1/2} = o(J^{1/2} K^{-1/2} T^{-1/2})$ by assumption. We now show (43). Claim (44) can be shown similarly. For the proof of (43) we use Bernstein's inequality for the following sum:

$$P\left( |\sum_{j=1}^{J} W_j| > x \right) \le 2 \exp\left( \frac{1}{2} \frac{x^2}{V + Mx/3} \right). \tag{48}$$

Here for a value of $t$ with $1 \le t \le T$ and for a vector $e \in \mathbb{R}^K$ with $\|e\| = 1$ the random variable $W_j$ is a row of the vector

$$J^{-1} A^* \left[ \psi(X_{t,j}) \psi(X_{t,j})^\top - E\left\{ \psi(X_{tj}) \psi(X_{tj})^\top \right\} \right] e.$$

In (48), $V$ 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 $t$, $e$ and the row number we get $V \leq C_1 J^{-1}$ and $M \leq C_2 K^{1/2} J^{-1}$. One can check that for $x$ with $x/[JK^{-1}T^{-1/2}] \to 0$ slowly enough, it holds that

$$K^K T \exp\left(\frac{1}{2}\frac{x^2}{V + Mx/3}\right) \to 0.$$

Using similar arguments as in the proof of Theorem 5.1 this implies claim (43).

From (36), (37), (39), (40), (41) and (42) we get the following expansions (uniformly for $1 \leq t \leq T$):

$$
\begin{aligned}
\widetilde{Z}_t - Z_t &= S_{t,Z}^{-1}\Bigg\{ J^{-1}\sum_{j=1}^{J}\varepsilon_{t,j}A^*\psi(X_{tj}) + J^{-1}\sum_{j=1}^{J}\varepsilon_{t,j}(\widetilde{A}-A^*)\psi(X_{t,j}) \\
&\qquad + J^{-1}\sum_{j=1}^{J}Z_t^{\top}(\widetilde{A}-A^*)\psi(X_{t,j})(\widetilde{A}-A^*)\psi(X_{t,j}) \\
&\qquad + J^{-1}\sum_{j=1}^{J}\Delta(X_{t,j})\widetilde{A}\psi(X_{t,j}) + \sum_{j=1}^{J}Z_t^{\top}(\widetilde{A}-A^*)\psi(X_{t,j})A^*\psi(X_{t,j})\Bigg\} + \mathcal{O}_P(T^{-1/2}) \\
&= S_{t,Z}^{-1}J^{-1}\sum_{j=1}^{J}\varepsilon_{t,j}A^*\psi(X_{t,j}) + S_{t,Z}^{-1}J^{-1}\sum_{j=1}^{J}\varepsilon_{t,j}(\widetilde{A}-A^*)\psi(X_{t,j}) \\
&\qquad + S_{t,Z}^{-1}J^{-1}\sum_{j=1}^{J}Z_t^{\top}(\widetilde{A}-A^*)\psi(X_{t,j})A^*\psi(X_{t,j}) + \mathcal{O}_P(T^{-1/2}) \\
&= \Delta_{t,1,Z} + \Delta_{t,2,Z} + \Delta_{t,3,Z} + \mathcal{O}_P(T^{-1/2}),
\end{aligned}
\tag{49}
$$

$$\widetilde{\alpha} - \alpha^* \;=\; S_\alpha^{-1} T^{-1} J^{-1} \sum_{t=1}^{T} \sum_{j=1}^{J} \{\psi(X_{t,j}) \otimes Z_t\} \varepsilon_{t,j} \tag{50}$$

$$+ S_\alpha^{-1} T^{-1} J^{-1} \sum_{t=1}^{T} \sum_{j=1}^{J} \{\psi(X_{t,j}) \otimes (\widetilde{Z}_t - Z_t)\} \varepsilon_{t,j}$$

$$+ S_\alpha^{-1} T^{-1} J^{-1} \sum_{t=1}^{T} \sum_{j=1}^{J} \{\psi(X_{t,j}) \otimes Z_t\} \{\psi(X_{t,j}) \otimes (\widetilde{Z}_t - Z_t)\}^\top \alpha$$

$$+ S_\alpha^{-1} T^{-1} J^{-1} \sum_{t=1}^{T} \sum_{j=1}^{J} \{\psi(X_{t,j}) \otimes (\widetilde{Z}_t - Z_t)\} \{\psi(X_{t,j}) \otimes (\widetilde{Z}_t - Z_t)\}^\top \alpha$$

$$+ S_\alpha^{-1} T^{-1} J^{-1} \sum_{t=1}^{T} \sum_{j=1}^{J} \{\psi(X_{t,j}) \otimes Z_t\} \Delta(X_{t,j})$$

$$+ S_\alpha^{-1} T^{-1} J^{-1} \sum_{t=1}^{T} \sum_{j=1}^{J} \{\psi(X_{t,j}) \otimes (\widetilde{Z}_t - Z_t)\} \Delta(X_{t,j}) + R,$$

where $\Delta(x) = m(x) - A^* \psi(x)$ and $R$ is a vector with $\|R\| = o_P(T^{-1/2})$.

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

$$T^{-1} \sum_{t=h+1}^{T} \Delta_{t,j,Z} Z_{t-h} = o_P(T^{-1/2}). \tag{51}$$

This can be easily checked for $j = 1$. For $j = 2$ it follows from

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

and $\|\widetilde{A} - A^*\|^2 = \mathcal{O}(K J^{-1})$. For the proof of (51) for $j = 3$ we note first that for $0 \leq l \leq L$

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

$$= Q_T(\widetilde{\alpha} - \alpha^*) + \|\widetilde{\alpha} - \alpha^*\| \mathcal{O}_P(K^{1/2} J^{-1/2} T^{-1/2})$$

$$= Q_T(\widetilde{\alpha} - \alpha^*) + o_P(T^{-1/2}),$$

where

$$Q_T = E\left[ T^{-1} \sum_{t=h+1}^{T} S_{t,Z}^{-1} A^* \psi(X_{t,j}) Z_{t-h,l} \{\psi(X_{t,j}) \otimes Z_t\}^\top \right].$$

Thus claim (51) for $j = 3$ follows from

$$Q_T(\widetilde{\alpha} - \alpha^*) = o_P(T^{-1/2}).$$

This claim can be checked by noting that $\|Q_T\| = \mathcal{O}(1)$ and by applying the expansion (50). Note that under our assumptions it does not hold in general that $\|\widetilde{\alpha} - \alpha^*\| = o_P(T^{-1/2})$. For this reason a careful treatment that we have just described above is needed. This concludes the proof of Theorem 5.2.

# References

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654.

Bliss, R. (1997). Movements in the term structure of interest rates. Economic Review Q IV, Federal Reserve Bank of Atlanta.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes*. Springer, New York.

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*, forthcoming.

Cont, R. and da Fonseca, J. (2002). The dynamics of implied volatility surfaces. *Quantitative Finance*, 2(1):45–60.

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

Detlefsen, K. and Härdle, W. (2006). Forecasting the term structure for variance swaps. Discussion Paper 2006-052, SfB 649, Humboldt-Universität zu Berlin.

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. *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 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(2):189–218.

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*. Springer, Berlin.

Hansen, L. H., Nielsen, B., and Nielsen, J. P. (2004). Two sided analysis of the variance with a latent time series. Nuffield Colledge 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 B*, 43(2):219–230.

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

Kauermann, G. (2000). Modeling longitudinal data with ordinal response by varying coefficients. *Biometrics*, 56(3):1692–698.

Krivobokova, T., Kauermann, G., and Archontakis, T. (2006). Estimating the term structure of interest rates using penalized splines. *Statistical Papers*, 47(3):443–459.

Lee, R. D. and Carter, L. (1992). Modeling and forecasting the time series of u.s. mortality. *Journal of the American Statistical Association*, 87(419):659–671.

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

Martinussen, T. and Scheike, T. (2000). A nonparametric dynamic additive regression model for longitudinal data. *Annals of Statistics*, 28(4):1000–1025.

Molgedey, L. and Galic, E. (2001). Extracting factors for interest rate scenarios. *European Physical Journal B*, 20(4):517–522.

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.

Rebonato, R. (1998). *Interest-Rate Option Models: Understanding, Analyzing and Using Models for Exotic Interest-Rate Options.* Wiley Series in Financial Engineering. John Wiley & Son Ltd., 2nd edition.

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.

Trück, S., Borak, S., Härdle, W., and Weron, R. (2006). Convenience yields for $CO_2$ emission allowance futures contracts. Discussion Paper 2006-076, SfB 649, Humboldt-Universität zu Berlin.

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.