

REPORTS
OF THE
INDIAN STATISTICAL INSTITUTE



INDIAN STATISTICAL INSTITUTE, CHENNAI CENTRE

SETS Building, MGR Knowledge City Road
CIT Campus, Taramani, Chennai - 600113

Time Series of Functional Data

Rituparna Sen and Claudia Klüppelberg

ISIC Preprint ASU-2015-1, May 2015

<http://isichennai.res.in/tr/asu/2015/1>

TIME SERIES OF FUNCTIONAL DATA

Rituparna Sen^{†,1} and Claudia Klüppelberg²

¹ Indian Statistical Institute, Chennai, India

² Technische Universität München, Germany

ABSTRACT. We develop time series analysis of functional data observed discretely, treating the whole curve as a random realization from a distribution on functions that evolve over time. The method consists of principal components analysis of functional data and subsequently modeling the principal component scores as vector ARMA process. We carry out the estimation of VARMA parameters using the equivalent state space representation. We derive asymptotic properties of the estimators and the fits. We apply the method to two different data sets. For term structures of interest rates, this provides a unified framework for studying the time and maturity components of interest rates under one set-up with few parametric assumptions. We compare our forecasts to the parametric Diebold and Li (2006) model. Secondly, we apply this approach to hourly spot prices of electricity and obtain fits and forecasts that are better than those existing in the electricity literature.

KEY WORDS: Diffusion model, Functional Principal Component, Functional Regression, Market Returns, Noise contamination, Prediction, Volatility Process, Trajectories of Volatility.

1. INTRODUCTION

Functional data analysis (see Ramsay and Silverman (2005) for a comprehensive introduction to FDA methods) is an extension of multivariate data analysis to functional data. In this framework, each individual is characterized by one or more real valued functions, rather than by a vector in \mathbb{R}^n . An important feature of FDA is its ability to take into account dependencies between numerical measurements that describe an individual, especially smoothness, ordering and neighborhood. In order to deal with irregular measurements and to allow numerical manipulation of functions, FDA replaces actual observations by a simple functional representation. Spline-based approximation is the most commonly used method, as it represents each individual by a smooth function. Kernel or wavelet-based approximations are also used. FDA has been successfully applied to real-life problems such as climatic variation forecasting, land usage prediction based on satellite images, forecasting electricity consumption, electrocardiograms etc.

An important tool of functional data analysis (FDA) is functional principal component analysis (FPCA, see Castro et al. (1986); Rice and Silverman (1991)). Functional processes can be characterized by their mean function and the eigenfunctions of the autocovariance operator. This is a consequence of the Karhunen-Loève representation of the functional process. We estimate the components of this representation. Individual trajectories are then represented by their functional principal component scores, which are available for subsequent statistical analysis. This often leads to substantial dimension reduction.

Most of the development in FDA has been with independent and identical replications of data. This permits the use of information from multiple data values to identify patterns. However, in certain situations, it is unrealistic to assume that the functions across time are independent. We do need some process structure. One idea to follow up here is to work with the replication principle implicit in stationary time series, where the values of the process are functions. Besse et al. (2000) develop an AR(1) model for FDA for forecasting climatic variations. Kargin and Onatski (2008) use an AR(1) model for forecasting Eurodollar futures. Hörmann and Kokoszka (2010) study weakly dependent functional processes, but they ignore the issue of smoothing. This is common to a lot of

[†]Corresponding author, E-mail: rsen@isichennai.res.in.

work following Bosq (2000) where the theory is developed assuming that the functions are observed continuously. In practice, however, we only observe the functions at a dense but discrete subset of the support and need to interpolate smoothly to infer about the whole function. This raises new questions about the behavior of the estimators. We develop the theory, where the functions follow a general ARMA(p, q) model and are observed discretely. We start with kernel smoothing, followed by dimension reduction using FPCA. Based on the time series of the first few significant principal components, we fit a VAR or VARMA model. We provide techniques for estimation of the model parameters and selection of the optimal model.

The paper is organized as follows: In section 2 we provide some background information on principal components analysis of functional data, term structures of interest rates and electricity spot markets. In section 3 we present our main set-up and methodology for time series analysis of functional data and some related results. In section 4 we propose the estimation techniques. In section 5 we describe two applications, namely forecasting and regression. We derive asymptotic properties of the estimators, fits and forecasts in section 6. We present the analysis of real data of interest rates and electricity spot prices in section 7 and finally we conclude in section 8.

2. BACKGROUND

Consider a sample of n smooth random trajectories $(f_i(t))_{t \in \mathcal{T}}$ for $i = 1, \dots, n$ generated from a process f . Throughout we assume that f is an element of the Hilbert space $\mathcal{H} := L^2(\mathcal{T})$ endowed with the inner product $\langle f, g \rangle_{\mathcal{H}} = \int_{\mathcal{T}} f(t)g(t)dt$ and the norm $\|f\| = \sqrt{\langle f, f \rangle_{\mathcal{H}}} < \infty$ a.s.. The observed measurements are available on a dense grid of support points t_{ij} on the domain $\mathcal{T} = [a_1, a_2]$ with additive white noise error W_{ij} which is independent of the underlying process. The measurements are for $i = 1, \dots, n$ and $j = 1, \dots, m$:

$$\tilde{f}_i(t_{ij}) = f_i(t_{ij}) + W_{ij} \quad \text{with} \quad E(W_{ij}) = 0, \text{Var}(W_{ij}) = \sigma^2. \quad (2.1)$$

2.1. Principal Components Analysis of Functional Data. We represent the smooth functional f in terms of its decomposition into functional principal components, a common approach in FDA. For the domain \mathcal{T} , setting

$$\mathbf{G}_f(s, t) = \text{Cov}(f(s), f(t)), \quad E(f(t)) = \mu_f(t), \quad s, t \in \mathcal{T}, \quad (2.2)$$

the functional principal components are the eigenfunctions of the auto-covariance operator $\mathbf{G}_f : \mathcal{H} \mapsto \mathbb{R}$, a linear operator on the space \mathcal{H} , that is given by

$$\mathbf{G}_f(g)(s) = \int_{\mathcal{T}} \mathbf{G}_f(s, t)g(t) dt.$$

We denote the orthonormal eigenfunctions by ϕ_k , with associated eigenvalues λ_k for $k = 1, 2, \dots$, such that $\lambda_1 \geq \lambda_2 \geq \dots$ and $\sum_k \lambda_k < \infty$. The Karhunen-Loève theorem (see Rice and Silverman (1991)) provides a representation of individual random trajectories of the functional f , given by

$$f(t) = \mu_f(t) + \sum_{k=1}^{\infty} \xi_k \phi_k(t), \quad t \in \mathcal{T}, \quad (2.3)$$

where the ξ_k are uncorrelated random variables that satisfy

$$\xi_k = \int (f(t) - \mu_f(t))\phi_k(t) dt, \quad E\xi_k = 0, \quad \text{Var}(\xi_k) = \lambda_k. \quad (2.4)$$

Under the data generating mechanism in (2.1), one has with indicator function $I(\cdot)$

$$E(\tilde{f}_i(t)) = \mu_f(t), \quad \text{Cov}(\tilde{f}_i(s), \tilde{f}_i(t)) = \mathbf{G}_f(s, t) + \sigma^2 I(s = t). \quad (2.5)$$

This implies that the smooth mean function μ_f and the smooth covariance surface \mathbf{G}_f can be consistently estimated from available data by pooling the sample of n trajectories and smoothing the resulting scatterplot. The exception for targeting points on \mathbf{G}_f with $s = t$ in (2.5) is necessitated by the presence of W . This does not pose a problem, since it follows from the smoothness of the

surface \mathbf{G}_f that the areas of $\mathbf{G}_f(s, t)$, for which $s = t$, can still be consistently estimated. Well-known procedures exist to infer eigenfunctions and eigenvalues (Rice and Silverman (1991); Müller et al. (2006)).

Processes f are then approximated by substituting estimates and using a judiciously chosen finite number K of terms in sum (2.3). This choice can be made using one-curve-leave-out cross-validation (Rice and Silverman (1991)), pseudo-AIC criteria (Yao et al. (2005)) or a scree plot, a tool from multivariate analysis, where one uses estimated eigenvalues to obtain a prespecified fraction of variance explained as a function of K or looks for a change-point.

The above procedure is also known in numerical analysis under the acronym *proper orthogonal decomposition* and as such it is used to price and hedge financial derivatives on forward curves; see Heppenger (2010) for examples from the energy market and further references.

2.2. Term Structure Modeling. Term structures of interest rates, also known as yield curve, represent the relationship between spot rates of zero-coupon securities and their term to maturity. This interest rate pattern is used to discount cash flows appropriately. The yield curve is also changing over time. Yield curves are used by fixed income analysts, who analyze bonds and related securities, to understand conditions in financial markets and to seek trading opportunities. Economists use the curves to understand economic conditions. Term structure modeling is a very interesting and active field. There are two popular approaches to term structure modeling. The no-arbitrage tradition focuses on perfectly fitting the term structure at a point in time to ensure that no arbitrage possibilities exist, which is important for pricing derivatives. The equilibrium tradition focuses on modeling the dynamics of the instantaneous rate, typically using affine models, after which yields at other maturities can be derived under various assumptions about the risk premium. Prominent contributions in the no-arbitrage vein include Hull and White (1990) and Heath et al. (1992), and prominent contributions in the affine equilibrium tradition include Vasicek (1977), Cox et al. (1985), and Duffie and Kan (1996). Diebold and Li (2006) use factor models imposing structure on the factor loadings to distill the entire yield curve, period-by-period, by regression onto a three-dimensional parameter that evolves dynamically. This is the closest one comes to simultaneous treatment of maturity and time evolution of the term structure. We propose an FDA analysis of the yield curve, treating the whole curve over different maturities as a random realization from a distribution on functions. Our proposed nonparametric approach requires no assumptions from the yield curve beyond smoothness and integrability in contrast to currently used approaches which include parametric components and assumptions. Our analysis provides a unified framework for studying the time and maturity components of interest rates under a set-up without too many parametric assumptions. This gives better modeling, data visualization and understanding of the interest rate process.

2.3. Electricity Hourly Spot Rates. In the beginning of the 1990s a liberalization of the electricity and gas markets started, resulting in the emergence of markets for spot prices and derivative products in numerous countries and regions spread over the world. Wolak (2000) gives a description of worldwide electricity market organization after deregulation. Johnson and Barz (1999) provides a comparative study of several electricity markets. These markets are in many ways distinct in nature and definition compared to what we find in the more classical commodity markets as oil, coal, metals and agriculture. Hence, new and challenging modeling problems appear. Electricity has very limited storage possibilities. Electrical power is only useful for practical purposes if it can be delivered during any period of time. This is why electricity has been called a flow commodity. Deregulated power markets have market mechanisms to balance supply and demand, where electricity is traded in an auction system for standardized contracts. All contracts guarantee the delivery of a given amount of power for a specified future time period.

Electricity spot prices have a strong periodicity. This can be explained from a microeconomic viewpoint by looking at the market price of electricity as an equilibrium price based on supply and demand curves. Since the demand is very inelastic, the marginal costs of the supply side determine the price to a large extent. If the total load is low, the power plants with the lowest variable production

costs are used, if the total load is high, gas or oil fired plants with high fuel costs are additionally running. The periodicity of the total load is responsible for the periodicity of the electricity prices. The total load has a random component, depending on short term weather conditions and other uncertain parameters, but it also has a clear predictable part, and so do the prices for electricity.

There is a lot of literature on the time series dynamics of daily average prices of electricity, see eg. Pilipovic (1998). However the models developed for average prices cannot be directly applied to describe the dynamics of the hourly prices. Agents in electricity markets are exposed to hourly variation. Prices need to be quoted on an hourly basis for the following day. Companies that use electricity in a certain profile during the day have demand for contracts that deliver only in a few hours of the day. Other applications include power risk management, contract structuring and derivative pricing, see Eydeland and Wolyniec (2003). Longstaff and Wang (2004) study the day-ahead hourly risk premium, calculated as the difference between the day-ahead price and the expected real time price. Szkuta et al. (1999) apply neural networks to model the dynamics of intraday prices. Huisman et al. (2007) propose a panel framework for modeling the characteristics of hourly electricity prices in day-ahead markets. Li and Flynn (2004) examine hourly rates of price change in fourteen deregulated markets. Knittel and Roberts (2005) fit a range of traditional financial models and less conventional electricity price models to hourly time series of real-time Californian electricity prices.

Electricity spot prices cannot be treated as a traditional time series model. Time series models assume that the information set is updated by moving from one observation to the next in time. This assumption is not valid for hourly electricity spot prices. Hourly prices for the whole of the next day are determined at the same time, when agents submit their bids and offers for delivery of electricity on the previous day. Therefore, the information set is constant during the day and updated every day. Additionally, a time series approach would require too many parameters to model all the periodic components and one runs into the curse of dimensionality.

The FDA approach is very appropriate in this situation since one can pick up these periodic structures with a few data-driven basis functions. It is inappropriate to assume that the pattern for the individual days are independent. Even after taking into account the daily periodicity, there is a dependence between the days. Hence ARMA modeling of FDA is a suitable tool for this application.

It should be noted that electricity prices also exhibit extreme spikes and short periods of extreme activity. See Bernhardt et al. (2008) and Weron et al. (2004), respectively, for treatment of these issues, in the context of daily spot prices. García et al. (2011) addresses the extreme behavior in the context of intraday spot prices. Their analysis shows, in particular, that no finite variance exists, so that the Hilbert space approach used in this paper lacks foundation. It has, however, been shown, that standard L^2 -procedures can be applied also in this heavy-tailed setup and give correct answers; cf. Davis (1996) and Mikosch et al. (1995) and references therein.

3. TIME SERIES OF FUNCTIONAL DATA

Coming back to the situation as described in (2.1), we assume that the series of functions follows the ARMAH(p, q) model with mean $\mu \in \mathcal{H}$:

$$f_i(\cdot) - \mu = \theta_1(f_{i-1}(\cdot) - \mu) + \cdots + \theta_p(f_{i-p}(\cdot) - \mu) + \epsilon_i(\cdot). \quad (3.1)$$

where $\epsilon_i(\cdot) = \eta_i(\cdot) + \psi_1\eta_{i-1}(\cdot) + \cdots + \psi_q\eta_{i-q}(\cdot)$, and $\eta_i(\cdot)$ is \mathcal{H} white noise. $\theta_1, \dots, \theta_p$ are linear functions. Combining (3.1) and (2.3) we have,

$$\sum_{k=1}^{\infty} \xi_{ki}\phi_k(\cdot) + \mu_f - \mu = \theta_1\left(\sum_{k=1}^{\infty} \xi_{ki-1}\phi_k(\cdot) + \mu_f - \mu\right) + \cdots + \theta_p\left(\sum_{k=1}^{\infty} \xi_{ki-p}\phi_k(\cdot) + \mu_f - \mu\right) + \epsilon(\cdot). \quad (3.2)$$

Using linearity of $\theta_1, \dots, \theta_p$, this implies,

$$\sum_{k=1}^{\infty} \xi_{ki}\phi_k(\cdot) + \mu_f - \mu = \sum_{k=1}^{\infty} \xi_{ki-1}\theta_1(\phi_k(\cdot)) + \theta_1(\mu_f - \mu) + \cdots + \sum_{k=1}^{\infty} \xi_{ki-p}\theta_p(\phi_k(\cdot)) + \theta_p(\mu_f - \mu) + \epsilon(\cdot). \quad (3.3)$$

Combining all the terms involving μ and μ_f into $\tilde{\mu}$ and using vector notation, we have:

$$\Phi(\cdot)\Xi_i = \tilde{\mu} + \theta_1(\Phi(\cdot))\Xi_{i-1} + \cdots + \theta_p(\Phi(\cdot))\Xi_{i-p} + \epsilon(\cdot). \quad (3.4)$$

where $\Phi = (\phi_1, \phi_2, \dots)$ and $\Xi = (\xi_{1i}, \xi_{2i}, \dots)^T$. Since the columns of Φ are orthonormal, we can premultiply equation (3.4) by Φ^T to get:

$$\Xi_i = \Phi^T \tilde{\mu} + \Phi^T \theta_1(\Phi(\cdot))\Xi_{i-1} + \cdots + \Phi^T \theta_p(\Phi(\cdot))\Xi_{i-p} + \phi^T \epsilon(\cdot). \quad (3.5)$$

This implies a VARMA(p, q) structure on the vector of principal component scores Ξ_i .

4. ESTIMATION

4.1. Inferring the Functional Process. At the core of the estimation procedure is the principal analysis of random trajectories (PART), applied to the data \tilde{f}_{ij} from (2.1), which is an algorithm to obtain mean and eigenfunctions, as well as FPC scores, from densely sampled functional data, as described in Müller et al. (2006). The smoothing steps in this algorithm are implemented with weighted local linear smoothing as in Fan and Gijbels (1996), which works well in practice; alternative smoothing methods can also be used. In order to estimate the overall mean function μ_f , we pool all available data into one big scatterplot $\{(t_j, \tilde{f}_{ij}), i = 1, \dots, n, j = 1, \dots, m\}$, and then obtain the nonparametric regression of \tilde{f} versus t by local linear smoothing. Formally, one finds the minimizers $\hat{\beta}_0(\tau), \hat{\beta}_1(\tau)$ of

$$\sum_{i=1}^n \sum_{j=1}^m \kappa_1 \left(\frac{t_j - \tau}{b_f} \right) \{ \tilde{f}_{ij} - \beta_0(\tau) - \beta_1(\tau)(t_j - \tau) \}^2, \quad (4.1)$$

where b_f is the smoothing bandwidth, chosen in practice by (generalized) cross-validation, and κ_1 is a kernel function, which is required to be a square integrable and compactly supported symmetric density function, with finite variance and absolutely integrable Fourier transform. Then one sets $\hat{\mu}_f(\tau) = \hat{\beta}_0(\tau)$ for which one has an explicit representation that is linear in W_j (Fan and Gijbels (1996)).

Analogously, surface data are smoothed by fitting local planes by weighted least squares. Specifically, estimation of the covariance surface \mathbf{G}_f is based on the collection of all available pairwise ‘‘empirical covariances’’ $\mathbf{G}_i(t_{j_1}, t_{j_2}) = (\tilde{f}_{ij_1} - \hat{\mu}_f(t_{j_1}))(\tilde{f}_{ij_2} - \hat{\mu}_f(t_{j_2}))$, assembling these into a two-dimensional scatterplot $\{(t_{j_1}, t_{j_2}), \mathbf{G}_i(t_{j_1}, t_{j_2})\}, i = 1, \dots, n, j_1, j_2 = 1, \dots, m\}$, and fitting a two-dimensional smoother to obtain the nonparametric regression of $\mathbf{G}_i(t_{j_1}, t_{j_2})$ versus (t_{j_1}, t_{j_2}) . Formally, one minimizes

$$\sum_{i=1}^n \sum_{1 \leq j_1 \neq j_2 \leq m} \kappa_2 \left(\frac{t_{j_1} - \tau_1}{h_f}, \frac{t_{j_2} - \tau_2}{h_f} \right) \times \{ \mathbf{G}_i(t_{j_1}, t_{j_2}) - [\beta_0(\tau_1, \tau_2) + \beta_1(\tau_1, \tau_2)(\tau_1 - t_{j_1}) + \beta_2(\tau_1, \tau_2)(\tau_2 - t_{j_2})] \}^2 \quad (4.2)$$

with respect to $\hat{\beta}_0(\tau_1, \tau_2), \hat{\beta}_1(\tau_1, \tau_2), \hat{\beta}_2(\tau_1, \tau_2)$ and defines $\hat{\mathbf{G}}_f(\tau_1, \tau_2) = \hat{\beta}_0(\tau_1, \tau_2)$. In (4.2), κ_2 is a kernel function, which is required to be a square integrable and compactly supported radially symmetric bivariate density function, with finite variance and absolutely integrable Fourier transform. The smoothing bandwidth h_f can again be chosen by (generalized) cross-validation.

We note that the diagonal terms $(j_1, j_2), j_1 = j_2$, are missing in the summation over j_1, j_2 in (4.2). This omission is motivated by the dependence structure of the targets \tilde{f}_{ij} . Due to the assumed smoothness of the covariance surface \mathbf{G}_f , the diagonal, on the other hand, is not essential in the surface estimation step, and can be omitted from the data that are used to construct the surface, without incurring any asymptotic penalty.

Once mean and covariance functions of the functional process f have been determined, a next step is the estimation of the (eigenvalue/eigenfunction) pairs, which are defined as the solutions of the eigen-equations $\int \mathbf{G}_f(s, t)\phi_k(s)ds = \lambda_k\phi_k(t)$, substituting the estimated covariance surface $\hat{\mathbf{G}}_f$ for \mathbf{G}_f . Solutions $(\hat{\lambda}_k, \hat{\phi}_k)$ are obtained by numerical eigenanalysis, based on an initial discretization step,

under orthonormality constraints for the eigenfunctions. Positive definiteness of the corresponding covariance surface can be guaranteed by a projection of the initial estimate $\hat{\mathbf{G}}_f$ on a positive definite version $\tilde{\mathbf{G}}_f$, as described in Yao et al. (2003).

In a last step, the PART algorithm yields estimates of the individual FPC scores. Motivated by (2.4), these are implemented as

$$\hat{\xi}_{ik} = \sum_{j=2}^m (\tilde{f}_{ij} - \hat{\mu}_f(t_{ij}))(t_{ij} - t_{ij-1})\hat{\phi}_k(t_{ij}), \quad i = 1, \dots, n, k = 1, 2, \dots \quad (4.3)$$

Individual trajectories can then be represented by an empirical version of the Karhunen-Loève expansion (2.3), for appropriate K ,

$$\hat{f}_i^{(K)}(t) = \hat{\mu}_f(t) + \sum_{k=1}^K \hat{\xi}_{ik}\hat{\phi}_k(t). \quad (4.4)$$

4.2. VARMA modeling of the principal component scores. The estimated principal component score vectors $\hat{\xi}_i = (\hat{\xi}_{i1}, \dots, \hat{\xi}_{iK})$ form a vector time series of length n . The infinite dimension of the functional data has been reduced to a finite dimension K . We fit Vector Autoregressive Moving Average (VARMA) models of order p, q to the finite dimensional time series of estimated principal component scores $\hat{\xi}_i$.

A VARMA(p, q) process is defined in vector notation as:

$$\xi_i = \mu + \theta_1 \xi_{i-1} + \dots + \theta_p \xi_{i-p} + \epsilon_i + \psi_1 \epsilon_{i-1} + \dots + \psi_q \epsilon_{i-q}, \quad i = p+1, \dots, n,$$

which can be further simplified by adopting the representation of a lag polynomial

$$\Theta(L)\xi_i = \mu + \Psi(L)\epsilon_i, \quad i = p+1, \dots, n. \quad (4.5)$$

Here μ and ξ_i, ϵ_i , for $i = 1, \dots, n$ are vectors of dimension K and $\theta_1, \dots, \theta_p, \psi_1, \dots, \psi_q$ are $K \times K$ matrices.

Note that in the above model each ξ_{ik} depends not only on its own history but also on other series' history (cross dependencies). This gives us several additional tools for analyzing causality as well as feedback effects.

A basic assumption in the above model is that the residual vectors follow a multivariate white noise, i.e.

$$E(\epsilon_i) = 0 \quad (4.6)$$

$$E(\epsilon_i \epsilon_j') = \begin{cases} \Sigma_\epsilon & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad (4.7)$$

The coefficient matrices must satisfy certain constraints in order that the VARMA-model is stationary. It is required that roots of

$$\det(I - \theta_1 z - \dots - \theta_p z^p) = 0 \quad (4.8)$$

lie outside the unit circle. Here I is the identity matrix. For more details on VARMA models see Chapter 11 of Brockwell and Davis (2009).

Model selection and forecasts can be done conveniently by using the equivalent representation of VARMA using state space models proposed by Akaike (1976). For details consider Aoki and Havenner (1991). The main advantage of the state space approach is its capability to find the best model in terms of the Akaike information criterion.

$$AIC = \log |\det(\tilde{\Sigma}_{\epsilon,p,q})| + 2s/n \quad (4.9)$$

Here s is the number of estimated parameters, n is the sample size and $\tilde{\Sigma}_{\epsilon,p,q}$ is the estimated covariance matrix obtained as:

$$\tilde{\Sigma}_{\epsilon,p,q} = \frac{1}{n} \sum_{i=p+1}^n \hat{\epsilon}_{i,p,q} \hat{\epsilon}'_{i,p,q}$$

where $\hat{\epsilon}_{i,p,q}$ is the estimated error vector for the i -th data vector after fitting a VARMA(p, q) model.

4.3. VAR modeling of the principal component scores. If the MA part of the VARMA model has coefficients $\psi_1 = \dots = \psi_q = 0$, the VARMA model reduces to a VAR (vector autoregressive) model. A VAR(p) model is defined, in vector notation, as

$$\xi_i = \mu + \theta_1 \xi_{i-1} + \dots + \theta_p \xi_{i-p} + \epsilon_i,$$

which can be further simplified by adopting the representation of a lag polynomial leading to a VAR version of (4.5)

$$\Theta(L)\xi_i = \mu + \epsilon_i.$$

The same condition (4.8) for stationarity holds as in the VARMA model. Estimation can be carried out by single equation least squares. See Hamilton (1994) for further details on VAR processes.

A linearly transformed finite order VAR(p) process, in general, does not admit a finite order VAR representation but becomes a VARMA process, see Chapter 11.6 of Lütkepohl (2005). Because transformations of variables are quite common in practice, this result is a powerful argument in favor of the more general VARMA class. However, often, as in our data sets, the MA order q of the optimal VARMA model selected by AIC is zero. In such cases, it is sufficient to fit a vector autoregressive (VAR) models of order p to the finite dimensional time series of estimated principal component scores $\hat{\xi}_{ik}$. Estimation and model specification of the VAR class is in general less complicated than of VARMA models.

For the VAR(p) model, the number of estimated parameters s in equation (4.9) equals $K(1 + pK) + K(K + 1)/2$. $|\tilde{\Sigma}_{\epsilon,p}|$ is the estimated covariance matrix.

5. PRACTICAL APPLICATIONS

5.1. Forecasting. Our primary aim is forecasting the curve for a future date based on the information available upto a certain point of time. The final VARMA(p, q) model, chosen in section 4.2, is used to produce model forecasts $\tilde{\xi}_{ik}$ of future principal component scores. Plugging these into equation (4.4) we obtain the forecasts $\hat{f}_i(t)$ of the original process f .

$$\hat{f}_i(t) = \hat{\mu}_f(t) + \sum_{k=1}^K \tilde{\xi}_{ik} \hat{\phi}_k(t). \quad (5.1)$$

Diebold and Li (2006), henceforth referred to as DL, use parametric functions involving variations of Nelson-Siegel exponential components to model the yield curve and then use univariate AR(1) models componentwise to estimate and forecast the factors. This method performs very well for forecasting the yield curve since these parametric functions are specifically designed for this situation. However, the problem of forecasting curves can arise in a lot of other situations, as in our electricity spot rate example. In such cases the DL method fails completely. We need the set of basis functions to be able to adapt to the data to be of broad and general use. In particular, the basis functions we use are eigenfunctions of the covariance of the dataset. Hence they can be used in any general setup.

Kargin and Onatski (2008) use predictive factors, similar to simultaneous linear predictions and an alternative to canonical correlations, together with an AR(1) model to predict the term structure of Eurodollar futures. It is not clear if and how the canonical correlation idea can be extended beyond AR(1). Also, in the empirical application presented in their paper, this method performs worse than the DL method.

5.2. Correlation and regression involving long and short term interest rates. The correlation between long and short term interest rates is a matter of debate among economists, see eg Brown and Schaefer (1994). The most powerful and widely accepted theory regarding the relationship between short- and long-term interest rates is the expectations theory of the term structure. Under the expectations theory, long-term interest rates are described as functions of the weighted averages of expected future short rates plus term premia. According to the theory, therefore, it can be surmised that, a rise (fall) in current short rates will lead to an increase (decrease) in long-term rates. In fact, this is the situation that is usually observed in the real world. However, there are occasional exceptions: For example, in the USA, the Federal Reserve increased the Federal funds rate by 1 percentage point in May 1994, but the interest rates of long maturities fell after that. In the UK, the Bank of England decided to decrease the repo rate by 0.25 percentage points in February 2003, but relatively long-maturity interest rates rose compared with those of the previous day.

There is no clear consensus among researchers in regard to how long-term rates will react to changes in short-term rates. Romer and Romer (2000) argue that a contractionary monetary policy should be followed by a fall in long rates because the rate of inflation is expected to decline in the future. They explain that a positive correlation between short and long rates, which is typically observed in the real world, is due to the Federal Reserves information advantage over the public in forecasting inflation. On the other hand, Campbell (1995) asserts that such a usually observed phenomenon stems from bond-market participants increasing requirement for excess return on long-term bonds. Ellingsen and Söderström (2001) show that, if market participants consider an unexpected change in the Federal funds rate as the Federal Reserves reaction to economic shocks, then interest rates of all maturities will move in the same direction. In contrast, if a change in short-term rates is regarded as being caused by an unexpected shift in exogenous parameters, such as the relative weight on output variability, the Federal funds rate and the interest rates of sufficiently long maturities will move in opposite directions.

We describe a method to quantitatively compute the relation between short and long term interest rates by extending the functional regression techniques of Müller et al. (2011) to the time series setting. Let the short term interest rate, say, up to three months maturity, be denoted by $f_{X_i}(\tau)$ and the long term interest rate for maturities above three months, be denoted by $f_{Y_i}(\tau)$. We are interested in predicting $f_{Y_{i+1}}$ given the entire past $(f_{X_1}, f_{Y_1}, \dots, f_{X_i}, f_{Y_i})$ and $f_{X_{i+1}}$. As before, $i = 1, \dots, n$ denotes the time and τ denotes the maturity. We carry out a functional principal component analysis as described in section 2.1 of the two series separately. The Karhunen-Loève expansions of the two series in terms of the principal component scores and eigenfunctions as described in section 4.1 are given by:

$$\begin{aligned} f_{X_i}(t) &= \mu_X(t) + \sum_{k=1}^{K_X} \xi_{ik}^X \phi_k^X(t), \\ f_{Y_i}(t) &= \mu_Y(t) + \sum_{k=1}^{K_Y} \xi_{ik}^Y \phi_k^Y(t). \end{aligned} \quad (5.2)$$

For each $i = 1, \dots, n$ the vector $(\xi_i^X, \xi_i^Y)^T = (\xi_{i1}^X, \dots, \xi_{iK_X}^X, \xi_{i1}^Y, \dots, \xi_{iK_Y}^Y)^T$ is now modeled as a VARMA process

$$\theta(L) \begin{pmatrix} \xi_i^X \\ \xi_i^Y \end{pmatrix} = \Psi(L) \begin{pmatrix} \epsilon_i^X \\ \epsilon_i^Y \end{pmatrix} \quad (5.3)$$

where $(\epsilon_i^X, \epsilon_i^Y)^T$ are independent vectors with mean 0 and common covariance matrix Σ . The predictor of ξ_{i+1}^Y given $(\xi_1^X, \xi_1^Y, \dots, \xi_i^X, \xi_i^Y, \xi_{i+1}^X)$ is obtained by substituting in (5.3) the least squares regression prediction of ϵ_{i+1}^Y on ϵ_{i+1}^X , with regression coefficient β_e . Subsequently these are used in

(5.2) to get the functional regression:

$$\widehat{\mathbb{E}}(f_{Y_{i+1}}|f_{X_1}, f_{Y_1}, \dots, f_{X_i}, f_{Y_i}, f_{X_{i+1}}) \quad (5.4)$$

$$= \mu_Y(t) + g(f_{X_1}, f_{Y_1}, \dots, f_{X_i}, f_{Y_i}) \quad (5.5)$$

$$+ \int (f_{X_{i+1}}(s) - \mu_X(s)) \beta(s, t) ds, \quad (5.6)$$

where the regression surface

$$\beta(s, t) = \sum_{k,m=1}^{\infty} \beta_{\epsilon}(k, m) \phi_k^X(s) \phi_m^Y(t) \quad (5.7)$$

and g is a linear function of the past.

6. ASYMPTOTICS

We derive some consistency results for eigenfunctions, eigenvalues, FPC scores and fitted trajectories. All proofs and details regarding the assumptions (M1)-(M8) can be found in the Appendix. In the following, the observation interval $\mathcal{T} = [a_1, a_2] \subset (0, T]$.

Recollecting that we estimate the overall mean trajectory μ_f in (4.1) with bandwidth b_f , and the covariance surface \mathbf{G}_f (2.2) in (4.2) with bandwidth h_f , we obtain for the estimation of these key constituents the following result. All convergence results in the following are for $n \rightarrow \infty$ and $\Delta = \sup |t_j - t_{j-1}| \rightarrow 0$.

Theorem 1. Assuming (M1)-(M4), we have

$$\sup_{t \in \mathcal{T}} |\hat{\mu}_f(t) - \mu_f(t)| = O_P\left(\frac{1}{\sqrt{nb_f}}\right), \quad (6.1)$$

$$\sup_{s, t \in \mathcal{T}} |\widehat{\mathbf{G}}_f(s, t) - \mathbf{G}_f(s, t)| = O_P\left(\frac{1}{\sqrt{nh_f^2}}\right). \quad (6.2)$$

This result provides justification for the mean and covariance function estimates. Next, let \mathfrak{J}' denote the set of indices of the eigenfunctions ϕ_k corresponding to eigenvalues λ_k of multiplicity one. As a consequence of the following theorem, we obtain consistency for the estimation of eigenvalues $\hat{\lambda}_k$ and eigenfunctions $\hat{\phi}_k$ for $k \in \mathfrak{J}'$, justifying the use of these estimates in the subsequent analysis.

Theorem 2. Assume (M1)-(M4). Then

$$|\hat{\lambda}_k - \lambda_k| = O_P\left(\frac{1}{\sqrt{nh_f^2}}\right) \quad (6.3)$$

$$\|\hat{\phi}_k - \phi_k\|_H = O_P\left(\frac{1}{\sqrt{nh_f^2}}\right) \quad k \in \mathfrak{J}' \quad (6.4)$$

$$\sup_{t \in \mathcal{T}} |\hat{\phi}_k(t) - \phi_k(t)| = O_P\left(\frac{1}{\sqrt{nh_f^2}}\right), \quad k \in \mathfrak{J}'. \quad (6.5)$$

One is also interested in the consistency of estimated principal component scores ξ_i and estimates $\hat{f}_i^{(K)}(t)$ as in (4.4) of individual trajectories (2.3).

Theorem 3. Assuming (M1)-(M6),

$$\sup_{1 \leq k \leq K} |\hat{\xi}_{ik} - \xi_{ik}| \xrightarrow{P} 0 \quad (6.6)$$

$$\sup_{t \in \mathcal{T}} |\hat{f}_i^{(K)}(t) - f_i(t)| \xrightarrow{P} 0. \quad (6.7)$$

The following result is regarding the forecast $\widehat{f}_i(t)$ as in (5.1) using the fitted ARMA(p, q) model.

Theorem 4. Assuming (M1)-(M6), if the distribution of the innovations is Normal, then the forecasted ξ_i are asymptotically normal with variance Σ_ϵ . The forecasted functions $(\phi_1, \dots, \phi_K)^T(\widehat{f}_i - f_i)$ are asymptotically normal with variance Σ_ϵ .

7. EMPIRICAL EXAMPLES

7.1. Data. Euribor (Euro Interbank Offered Rate) is the rate at which Euro interbank term deposits are being offered by one prime bank to another within the European Monetary Union. Historical data is available at www.euribor.org The choice of banks quoting for Euribor is based on market criteria. These banks are of first class market standing and they have been selected to ensure that the diversity of the euro money market is adequately reflected, thereby making Euribor an efficient and representative benchmark. Thomson Reuters has been chosen as the screen service provider responsible for computing and also publishing Euribor. A representative panel of banks provide daily quotes of the rate, rounded to two decimal places, that each panel bank believes one prime bank is quoting to another prime bank for interbank term deposits within the euro zone. Panel banks contribute for fifteen maturities: one, two and three weeks, and then for every month from one to twelve. Thomson Reuters, for each maturity, eliminates the highest and lowest 15% of all the quotes collected. The remaining rates are averaged and rounded to three decimal places. After the calculation has been processed at 11:00 a.m. CET, Thomson Reuters instantaneously publishes the Euribor reference rate, which is made available to all its subscribers and to other data vendors. At the same time, the underlying panel bank rates are published on a series of fifteen composite pages which display all the rates by maturity. Since its launch, Euribor has become a reality on the derivatives markets and is the underlying rate of many derivatives transactions, both, over-the-counter and exchange-traded.

The electricity data was obtained on day-ahead spot prices from European Energy Exchange (EEX). EEX is an exchange under public law with location in Leipzig. EEX operates the spot trading for power with physical delivery on the day to follow and the futures market. With high trading volumes, the EEX is one of the largest and most important power exchanges in Europe. Clearing prices resulting at the EEX markets are also used as reference prices for other electricity contracts in Germany and elsewhere in Europe. The EEX operates a day-ahead market for hourly and block electricity contracts. Hourly power contracts are traded daily for physical delivery in the next days 24-hour period (midnight to midnight). Each morning, the players submit their bids for purchasing or selling a certain volume of electricity for the different hours of the following day. Once the spot market is closed for bids, at noon each day, the day-ahead price is derived for each hour next day.

7.2. Programs. The initial fitting of functional data to obtain mean, covariance and principal components is done by employing the PACE package for functional data analysis written in Matlab. We use the Gaussian kernel. The package is available at

<http://anson.ucdavis.edu/~wyang/PACE/>

VAR model fitting and diagnostics is done using the econometrics toolbox in Matlab.

VARMA and related state space model computations are done using the Dynamic Systems Estimation (dse) package in R available at

<http://cran.r-project.org/web/packages/dse/index.html>.

It should be noted that in all the actual data applications, the models chosen by AIC criterion had the MA degree zero.

7.3. Analysis of Euribor Data. We separate the data into years because for long time horizons the stationarity assumption of the time series may not be valid. We present the results for the years 1999 and 2007. Together they are representative of the other years. In Figure 1, we present the raw data for the two years. For each weekday of the year we have data of dimension 13 (for 1999) and 15 (for 2007) and we think of it as a time series of functions. The maturities, smoothing bandwidth

choices, number of significant principal components and degree of VARMA models are presented in table 1. The summary statistics for the different maturities are given in tables 2 and 3. The mean functions, covariance surfaces and significant eigenfunctions for both years are presented in Figures 2-4.

We present the distance between the observed and estimated functions in Figure 5. Distance is defined as:

$$\int (f_i(t) - \hat{f}_i^{(K)}(t))^2 dt$$

It is seen that the estimation based on FDA performs comparably with the DL method which was specifically designed for the specific purpose of forecasting the term structure of interest rates. In Figure 6 we present the distance between the observed functions and the predictions based on VARMA models for the principal component scores. In this case, at least for the year 1999, the FDA method outperforms the DL method. In general, the errors from the FDA method have smaller variability than those from the DL method.

For regression, the smoothing bandwidth choices and number of significant principal components for different maturity horizons, that is, short and long, for the two years are presented in table 4. In that table we also present the degree of VARMA models and regression R^2 . The regression coefficient of ϵ^Y on ϵ^X is (0.9025 1.6098) for 1999 and (-1.0037 3.5752) for 2007. In figure 7 we present the regression surface. The surface is far from flat which shows that the relation between short and long term interest rates cannot be expressed as one number, but depends in an involved way on the short term rates of different maturities. The surface is similar for the two years and doesn't vary too much over the y-axis. This implies that the relationship is stable in time and does not vary much with the different maturities of the long term rates. In figure 8 we present the distance between the observed and predicted functions of long term interest rates. Prediction based on time series and regression on the short term interest rates far outperforms the prediction based on only time series of long term rates.

7.4. Analysis of Energy 2008 Data. Data are hourly spot prices for weekdays from Jan 1 to Sept 30 2008. The number of days is $n = 196$. As noted in section 2, electricity prices over long time horizons exhibit nonstationarity. We choose a period that looks like a stationary period from time series plots of the raw data and the principal component scores. The price is observed in one hour intervals over the whole day, so $\tau = 1, 2, \dots, 24$. The data are of dimension 24 and we think of it as a time series of functions representing spot prices over each day. In figure 9 we present the raw data. The summary statistics for the different hours are given in table 3. The smoothing bandwidth chosen for the mean is 2.75 hours and for the covariance surface is 1.0698 hours. The number of significant principal components is chosen to be 4. The estimated smooth mean, covariance surface and eigenfunctions are presented in Figures 9-10. In figure 11 we present the distance between the observed and estimated functions and the distance between the observed functions and the predictions based on VARMA(1,0) models for the principal component scores. In figure 12 we present the observed, estimated and predicted curves for 12 randomly chosen days.

The measure we use for evaluating the model is the percentage of variance explained compared to the simple invariant estimate given by the overall mean. This is like an R^2 criterion and is defined as:

$$1 - \frac{(y_{ij} - \hat{y}_{ij})^2}{(y_{ij} - \bar{y})^2}$$

This value equals 0.9503 for the fitted model and equals 0.7713 for the forecast using VARMA(1,0).

We compare the forecasts with the AR and iterated Hsieh-Manski (IHMAR) models as described in Weron and Misiorek (2008). The basic AR model considered there looks upon the data as a single time series and involves as predictors the prices for the same hours on the previous two days, the previous week, the minimum of the previous day's 24 hourly prices and a dummy variable for Monday. The IHM estimator is an iterated version of an adaptive maximum likelihood estimator for ordinary regression. Cao et al. (2003) suggest that this estimator should perform well when the distribution

of the model errors is far from normal. The forecasts are compared using the weekly-weighted mean absolute error (WMAE) defined as:

$$\text{WMAE} = \frac{\sum_{i=1}^5 \sum_{j=1}^{24} |y_{ij} - \hat{y}_{ij}|}{\sum_{i=1}^5 \sum_{j=1}^{24} y_{ij}}. \quad (7.1)$$

The WMAE for the 38 weeks are displayed in table 6 for the three models: pure AR, FDA and IHMAR. The summary statistics are presented in the bottom rows. These include the mean WMAE over all weeks, the number of times a given model was best and the mean deviation from the best model (m.d.f.b.) in each week. The latter measure indicates which approach is closest to the optimal model on the average and is defined as:

$$\text{m.d.f.b.} = \frac{1}{T} \sum_{t=1}^T (E_{i,t} - E_{\text{bestmodel},t}), \quad (7.2)$$

where i ranges over all models, T is the number of weeks and E is the WMAE.

It is seen that IHMAR does not give best forecast for any week. This model is supposed to perform well if the data is far from normality. Since we have chosen a stationary period for our analysis, this is almost always dominated by the simple AR model. The FDA method does not dominate the AR model, but performs better on the average according to all three summary statistics.

8. FUTURE DIRECTIONS

We have developed the method for studying the time series of functional data. It should be straight forward to extend this method to incorporate seasonality. Further work needs to be done for extensions to nonstationary time series eg. change point, regime switching, heavy-tailed innovations etc. One can use the methods of Aneiros-Pérez and Vieu (2008) and Damon and Guillas (2002) to incorporate the effects of covariates.

APPENDIX

We begin by listing some assumptions, on moments of \tilde{f}_{ij} (2.1) and smoothing bandwidths b_f and h_f as used in (4.1) and (4.2). Throughout we consider $\Delta \rightarrow 0$ and $n \rightarrow \infty$.

$$(M1) \sup_j E[\tilde{f}_{ij}]^4 < \infty$$

$$(M2) \begin{aligned} b_f \rightarrow 0, \quad nb_f^4 \rightarrow \infty, \quad \limsup_n b_f^6 < \infty, \\ h_f \rightarrow 0, \quad nh_f^6 \rightarrow \infty, \quad \limsup_n h_f^8 < \infty, \\ \limsup_n^{1/2} b_f \Delta < \infty, \quad \limsup_n^{1/2} h_f^2 \Delta < \infty. \end{aligned}$$

(M3) The kernel functions κ_1 and κ_2 are compactly supported kernel functions of order (0,2) and (0,0,2) respectively as defined in equation (25) of Yao et al. (2005). The Fourier transforms of $\kappa_1(u)$ and $\kappa_2(u, v)$, namely, $\zeta_1(t) = \int e^{-iut} \kappa_1(u) du$ and $\zeta_2(t, s) = \int e^{-iut+ivs} \kappa_2(u, v) dudv$ are absolutely integrable. That is, $\int |\zeta_1(t)| dt < \infty$ and $\int \int |\zeta_2(t, s)| dt ds < \infty$.

Let \mathfrak{F}_i^k be the σ -algebra of events generated by the random functions $\{f_j(t), t \in \mathcal{T}, i \leq j \leq k\}$ and let $L_2(\mathfrak{F}_i^k)$ denote the collection of all second order random variables which are \mathfrak{F}_i^k measurable. The stationary process $f_i(\cdot)$ is called ρ -mixing (Kolmogorov and Rozanov(1960)) if

$$\sup_{U \in L_2(\mathfrak{F}_{-\infty}^0), V \in L_2(\mathfrak{F}_k^\infty)} \frac{\text{Cov}(U, V)}{\text{Var}^{1/2}(U) \text{Var}^{1/2}(V)} = \rho(k) \rightarrow 0 \quad \text{as } \rho \rightarrow \infty.$$

(M4) The process f_j is ρ -mixing with $\sum \rho(l) < \infty$.

Proof of Theorem 1: The proof borrows arguments from the proofs of Lemma 1,2 and Theorem 1 of Yao et al. (2005). Assumptions (M1)-(M3) ensure that proper versions of the above mentioned Theorem apply here. The difference of the present paper from the setup in Yao et al. (2005) is that the functions are no longer independent, but come from an ARMA process. Assumption (M4) allows us to suitably modify the last step of Lemma 1 to bound the variance term under this dependence

structure. Compare to Masry and Fan (1997) on local polynomial estimation of regression functions for mixing processes. \square

Proof of Theorem 2: The proof is analogous to that of Theorem 2 of Yao et al. (2005). \square

Further assumptions are needed for the remaining results. For each $j \geq 0$, define $\delta_j^f = \frac{1}{2} \min\{|\lambda_l - \lambda_j| : l \neq j\}$, and $\mathbf{A}_{\delta_j^f} = \{z \in \mathcal{C} : |z - \lambda_j| = \delta_j^f\}$, where \mathcal{C} are the complex numbers. Furthermore, define $A_{\delta_j^f} = \sup\{\|\mathbf{R}_f(z)\|_F : z \in \mathbf{A}_{\delta_j^f}\}$, where $\mathbf{R}_f(z) = (\mathbf{G}_f - zI)^{-1}$ is the resolvent of operator \mathbf{G}_f and $\|\cdot\|_F$ is an operator norm, defined on the separable Hilbert space F generated by the Hilbert-Schmidt operators on H , endowed with the inner product $\langle T_1, T_2 \rangle_F = \sum_j \langle T_1 u_j, T_2 u_j \rangle_H$ and the norm $\|T\|_F^2 = \langle T, T \rangle_F$, where $T_1, T_2, T \in F$, and $\{u_j : j \geq 1\}$ is any complete orthonormal system in H . Then we assume

$$(M5) \quad K = K(n) \rightarrow \infty. \text{ and } \sum_{j=1}^K (\delta_j^f A_{\delta_j^f} \sup_{t \in [0, T]} |\phi_j(t)|) / (\sqrt{nh_f^2} - A_{\delta_j^f}) \rightarrow 0.$$

$$(M6) \quad \sum_{j=1}^K \sup_{t \in [0, T]} |\phi_j(t)| = o(\min\{\sqrt{nb_f}, \sqrt{\Delta^{-1}}\}), \text{ and} \\ \sum_{j=1}^K \sup_{t \in [0, T]} |\phi_j(t)| \sup_{t \in [0, T]} |\phi_j'(t)| = o(\Delta^{-1}).$$

The assumptions (M5) and (M6) describe how the number of included eigenfunctions K increases when n tends to infinity. The quantities δ_j reflect the decay of the eigenvalues of the covariance operators, while $A_{\delta_j^f}$ depend on the local properties of the covariance operator G around the eigenvalues λ_j . In practice, the eigenvalues usually decrease rapidly to zero, the number of included eigenfunctions K is much less than n ; i.e., $n \ll K$, which suggests the assumptions (M5) and (M6) can be easily fulfilled for such processes.

Proof of Theorem 3: The proof is immediate from Theorem 1 of Yao and Lee (2006). \square

Proof of Theorem 4: Replacing f_i by $\hat{f}_i^{(K)}$ in (6) and following through the argument of (7)-(10), we have,

$$\hat{\xi}_i = \mu_\xi + \theta_{1\xi} \hat{\xi}_{i-1} + \dots + \theta_{p\xi} \hat{\xi}_{i-p} + \epsilon_\xi + (\phi_1 \dots \phi_K)^T R_i \tag{8.1}$$

$$R_i(t) = \hat{f}_i^{(K)}(t) - \mu - \theta_1(\hat{f}_{i-1}^{(K)}(t) - \mu) - \dots - \theta_p(\hat{f}_{i-p}^{(K)}(t) - \mu) - \epsilon_i(t). \\ = \hat{f}_i^{(K)}(t) - f_i(t) - \theta_1(\hat{f}_{i-1}^{(K)}(t) - f_{i-1}(t)) - \dots - \theta_p(\hat{f}_{i-p}^{(K)}(t) - f_{i-p}(t)).$$

Combining this with (6.6)

$$\sup_{t \in [0, T]} |R_i(t)| \leq p(1 + \|\theta_1\|^2 + \dots + \|\theta_p\|^2) \max_{i-p \leq j \leq i} \sup_{t \in [0, T]} |\hat{f}_j^{(K)}(t) - f_j(t)| \\ \xrightarrow{P} 0$$

where $\|\theta_i\| = \sup_f (\theta_i(f)(t)/f(t))$ is the maximum eigenvalue of the operator θ_i . For a vector ARMA(p, q) model with normally distributed innovations, the MSE of the forecast error is $(1 + K(p + q)/n)\Sigma_\epsilon$ as shown in given in Hung and Alt(1994). Following the same argument and combining with (8.1) using Slutsky's theorem gives the result. \square

REFERENCES

- [1] Akaike, H. (1976) "An information criterion (AIC)." *Math. Sci.*, 14(153), 59.
- [2] Aneiros-Pérez, G. and P. Vieu (2008) "Nonparametric time series prediction: A semi-functional partial linear modeling." *Journal of Multivariate Analysis*, 99, 834-857.
- [3] Aoki, M. and A. Havenner (1991) "State space modeling of multiple time series." *Econometric Reviews*, 10(1), 159.
- [4] Besse, P.C., H. Cardot and D. B. Stephenson (2000) "Autoregressive forecasting of some functional climatic variations." *Scandinavian Journal of Statistics*, 27(4), 673-687.

- [5] Bernhardt, C., C. Klüppelberg and T. Meyer-Brandis (2008) “Estimating high quantiles for electricity prices by stable linear models.” *Journal of Energy Markets*, 1(1), 3-19.
- [6] Brown, R. H. and S. M. Schaefer (1994) “The term structure of real interest rates and the Cox, Ingersoll, and Ross model.” *Journal of Financial Economics*, 35, 3-42.
- [7] Brockwell, J. and R. Davis (2009) “Time Series: Theory and Methods.” Springer, New York, NY.
- [8] Campbell, J. Y. (1995) “Some lessons from the yield curve.” *Journal of Economic Perspectives*, 9, 129-152.
- [9] Cao, R., J.D. Hart and A. Saavedra (2003) “Nonparametric maximum likelihood estimators for AR and MA time series.” *Journal of Statistical Computation and Simulation*, 73(5), 347-360.
- [10] Castro, P. E., W. H. Lawton and E. A. Sylvestre (1986) “Principal modes of variation for processes with continuous sample curves.” *Technometrics*, 28, 329-337.
- [11] Cox, J.C., J.E. Ingersoll and S.A. Ross (1985) “A Theory of the Term Structure of Interest Rates,” *Econometrica*, 53, 385-407.
- [12] Damon J. and S. Guillas (2002) “The inclusion of exogenous variables in functional autoregressive ozone forecasting.” *Environmetrics*, 13, 759-774.
- [13] Davis, R. A. (1996) “Gauss-Newton and M-estimation for ARMA processes with infinite variance.” *Stochastic Processes and Their Applications*, 63, 75-95.
- [14] Diebold, F.X. and C. Li (2006) “Forecasting the term structure of government bond yields.” *Journal of Econometrics*, 130(2), 337-364.
- [15] Duffie, D. and R. Kan (1996) “A yield-factor model of interest rates.” *Mathematical Finance*, 6, 379-406.
- [16] Ellingsen, T. and U. Söderström (2001) “Monetary policy and market interest rates.” *American Economic Review*, 91, 1594-607.
- [17] Eydeland, A. and K. Wolyniec (2003) “Energy and Power Risk Management: New Developments in Modeling, Pricing and Hedging.” John Wiley and Sons, Hoboken, NJ.
- [18] Fan, J. and I. Gijbels (1996) “Local polynomial modelling and its applications.” Chapman and Hall, London.
- [19] García, I., C. Klüppelberg and G. Müller (2011) “Estimation of stable CARMA models with an application to electricity spot prices.” *Statistical Modelling*, 11(5), 447-470.
- [20] Hamilton, J. D. (1994) “Time Series Analysis.” Princeton University Press, Princeton, NJ.
- [21] Heath, D., R. Jarrow and A. Morton (1992) “Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation.” *Econometrica*, 60, 77-105.
- [22] Hepperger, P. (2010) “Option pricing in Hilbert space valued jump-diffusion models using partial integro-differential equations.” *SIAM Journal on Financial Mathematics* 1(1), 454-489.
- [23] Hörmann, S. and P. Kokoszka (2010) “Weakly dependent functional data.” *The Annals of Statistics*, 38, 1845-1884.
- [24] Huisman, R., C. Hurman and R. Mahieu (2007) “Hourly electricity prices in day-ahead markets.” *Energy Economics*, 29, 2402-48.
- [25] Hull, J. and A. White (1990) “Pricing interest-rate-derivative securities.” *Review of Financial Studies*, 3, 573-592.
- [26] Hung, K. and F. B. Alt (1994) “The approximation of the one-step ahead forecast error covariance for vector ARMA models.” *International Journal of Forecasting*, 10, 59-64.
- [27] Johnson, B. and G. Barz (1999) “Selecting stochastic processes for modelling electricity prices.” In Jamson, R. (ed.) “Energy Modelling and the Management of Uncertainty,” pp. 322, Risk Books, London.
- [28] Kargin, V. and A. Onatski (2008) “Curve forecasting by functional autoregression.” *Journal of Multivariate Analysis*, 99, 2508-2526.
- [29] Knittel, C. and M. Roberts (2005) “An empirical examination of deregulated electricity prices.” *Energy Economics*, 27(5), 791-817.
- [30] Kolmogorov, A. N. and Y.A. Rozanov (1960) “On strong mixing conditions for stationary Gaussian processes.” *Theory of Probability and its Applications*, 5, 204-208.
- [31] Li, Y. and P. Flynn (2004) “Deregulated power prices: comparison of volatility.” *Energy Policy*, 32, 1591-1601.
- [32] Longstaff, F. and A. Wang (2004) “Electricity forward prices: A high-frequency empirical analysis.” *Journal of Finance*, 59, 1877-1900.
- [33] Lütkepohl, H. (2005) “New Introduction to Multiple Time Series Analysis.” Springer-Verlag, Berlin.
- [34] Masry, E., and J. Fan (1997) “Local polynomial estimation of regression functions for mixing processes.” *Scandinavian Journal of Statistics*, 24, 165-179.
- [35] Mikosch, T., Gadrich, T. Klüppelberg, C. and Adler, R. (1995) “Parameter estimates for ARMA models with infinite variance innovations.” *Annals of Statistics*, 23, 305-326.
- [36] Müller, H. G., U. Stadtmüller and F. Yao (2006) “Functional variance processes.” *Journal of the American Statistical Association*, 101, 1007-1018.
- [37] Müller, H. G., R. Sen and U. Stadtmüller (2011) “Functional data analysis for volatility process.” *Journal of Econometrics*, 165: 233 - 245.
- [38] Pilipovic, D. (1998) “Energy risk: valuing and managing energy derivatives.” McGraw-Hill, New York.

- [39] Ramsay, J.O. and B.W. Silverman (2005) "Functional Data Analysis." Springer Series in Statistics. New York, NY.
- [40] Rice, J. A. and B. W. Silverman (1991) "Estimating the mean and covariance structure nonparametrically when the data are curves." *Journal of Royal Statistical Society Series B*, 53, 233-243.
- [41] Romer, C. D. and D. H. Romer (2000) "Federal reserve information and the behavior of interest rates." *American Economic Review*, 90, 429-57.
- [42] Szkuta, B., L. Sanabria and T. Dillon (1999) "Electricity price short-term forecasting using artificial networks." *IEEE Transactions on Power Systems*, 4, 851-857.
- [43] Vasicek, O. (1977) "An equilibrium characterization of the term structure." *Journal of Financial Economics*, 5, 177-188.
- [44] Weron, R., M. Bierbrauer and S. Trück (2004) "Modeling electricity prices: jump diffusion and regime switching." *Physica A*, 336, 39-48.
- [45] Weron R. and A. Misiorek (2008) "Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models" *International Journal of Forecasting*, 24, 744-763.
- [46] Wolak, F. A. (2000) "Market design and price behavior in restructured electricity markets: an international comparison." In: "Pricing in Competitive Electricity Markets" ed. Faruqui, A. and K. Eakin. Kluwer Academic Publishers. Norwell, MA.
- [47] Yao, F., H. G. Müller, A. J. Clifford, S. R. Dueker, J. Follett, Y. Lin, B. Buchholz and J. S. Vogel (2003) "Shrinkage estimation for functional principal component scores, with application to the population kinetics of plasma folate." *Biometrics*, 59, 676-685.
- [48] Yao, F., H. G. Müller and J. L. Wang (2005) "Functional data analysis for sparse longitudinal data." *Journal of the American Statistical Association*, 100, 577-590.
- [49] Yao, F. and T. Lee (2006) "Penalized spline models for functional principal component analysis." *Journal of the Royal Statistical Society: Series B*, 68(1), 325.

Year	Trading Days (n)	Maturities (τ)	Bandwidth for mean	Bandwidth for covariance	PC (K)	VARMA (p, q)
1999	260	0.25, 1, 2, \dots , 12	2.75	1.0673	3	$p = 3, q = 0$
2007	255	0.25, 0.5, 0.75, 1, 2, \dots , 12	2.75	0.6853	3	$p = 1, q = 0$

TABLE 1. Forecasting the term structure as in Section 5.1

Maturity	Mean	SD	Minimum	Maximum
0.25	2.8190	0.2981	2.431	4.099
1	2.8637	0.3046	2.557	3.576
2	2.9148	0.3201	2.565	3.517
3	2.9625	0.3309	2.570	3.531
4	3.0024	0.3180	2.576	3.544
5	3.0336	0.3086	2.581	3.555
6	3.0547	0.3032	2.586	3.569
7	3.0739	0.3056	2.596	3.604
8	3.0940	0.3119	2.629	3.660
9	3.1148	0.3222	2.652	3.719
10	3.1371	0.3354	2.659	3.780
11	3.1604	0.3506	2.665	3.837
12	3.1834	0.3679	2.668	3.895

TABLE 2. Summary Statistics for Euribor 1999 data grouped by maturity in months

Maturity	Mean	SD	Minimum	Maximum
0.25	3.9618	0.2226	3.567	4.548
0.50	3.9938	0.2471	3.582	4.946
0.75	4.0310	0.2855	3.594	4.969
1	4.0804	0.3197	3.606	4.947
2	4.1952	0.3596	3.648	4.952
3	4.2776	0.3799	3.725	4.953
4	4.3045	0.3593	3.770	4.937
5	4.3276	0.3392	3.808	4.927
6	4.3520	0.3200	3.857	4.917
7	4.3711	0.3039	3.896	4.908
8	4.3891	0.2900	3.924	4.898
9	4.4067	0.2776	3.956	4.891
10	4.4224	0.2677	3.974	4.889
11	4.4362	0.2602	3.984	4.885
12	4.4500	0.2538	3.995	4.885

TABLE 3. Summary Statistics for Euribor 2007 data grouped by maturity in months

Year	Maturity	Bandwidth for mean	Bandwidth for covariance	PC	VARMA	R^2
1999	Short	1.9663	0.9154	2	$p = 1, q = 0$	0.4223
	Long	4.4	2.4603	1		
2007	Short	1.1812	0.4927	2	$p = 3, q = 0$	0.4019
	Long	4.4	2.7106	1		

TABLE 4. Regression of the long term on the short term interest rates as in Section 5.2

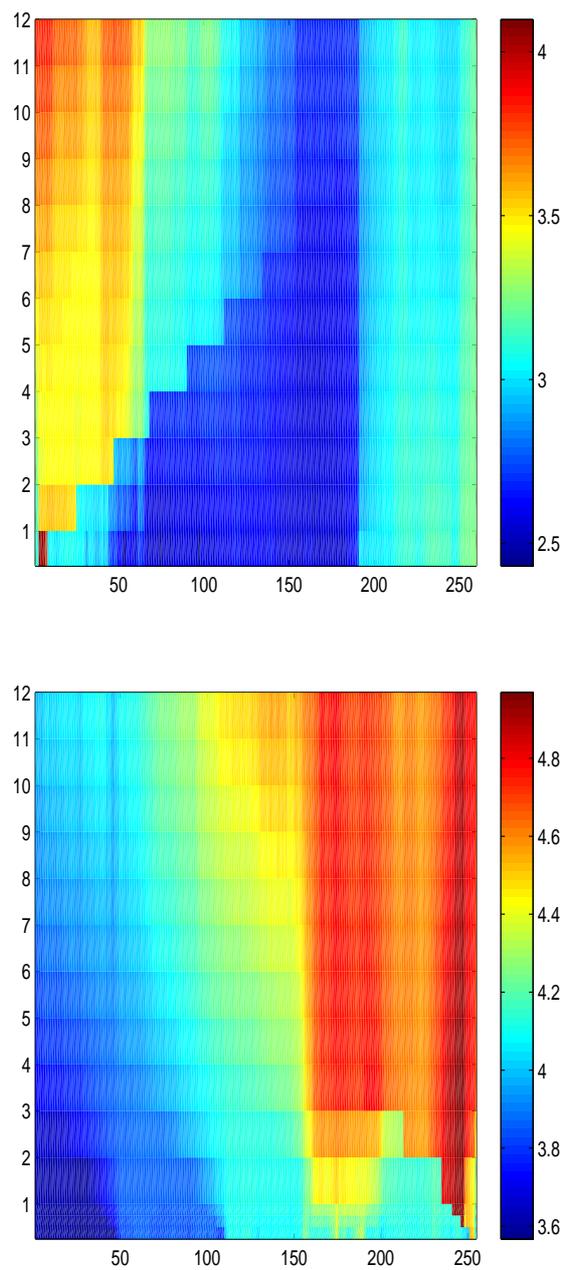


FIGURE 1. Raw data of interest rates. Top: 1999, bottom: 2007; x-axis: day, y-axis: maturity in months, z-axis (colors): interest rate.

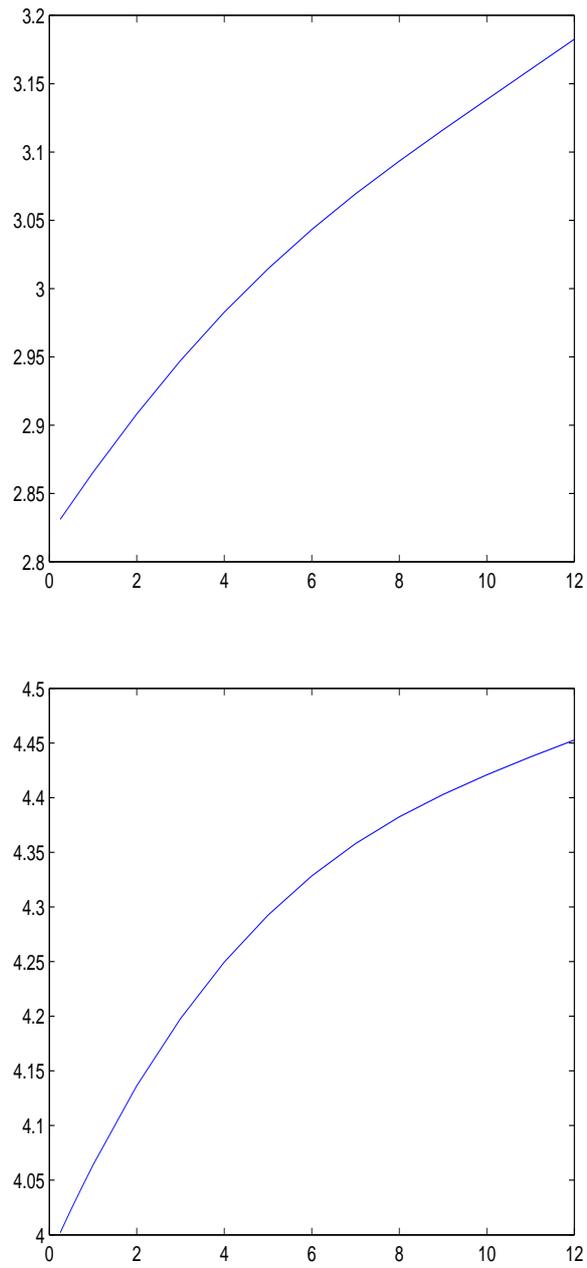


FIGURE 2. Estimated smooth mean function of Euribor data. Top: 1999, bottom: 2007; x-axis: maturity in months, y-axis: mean interest rate.

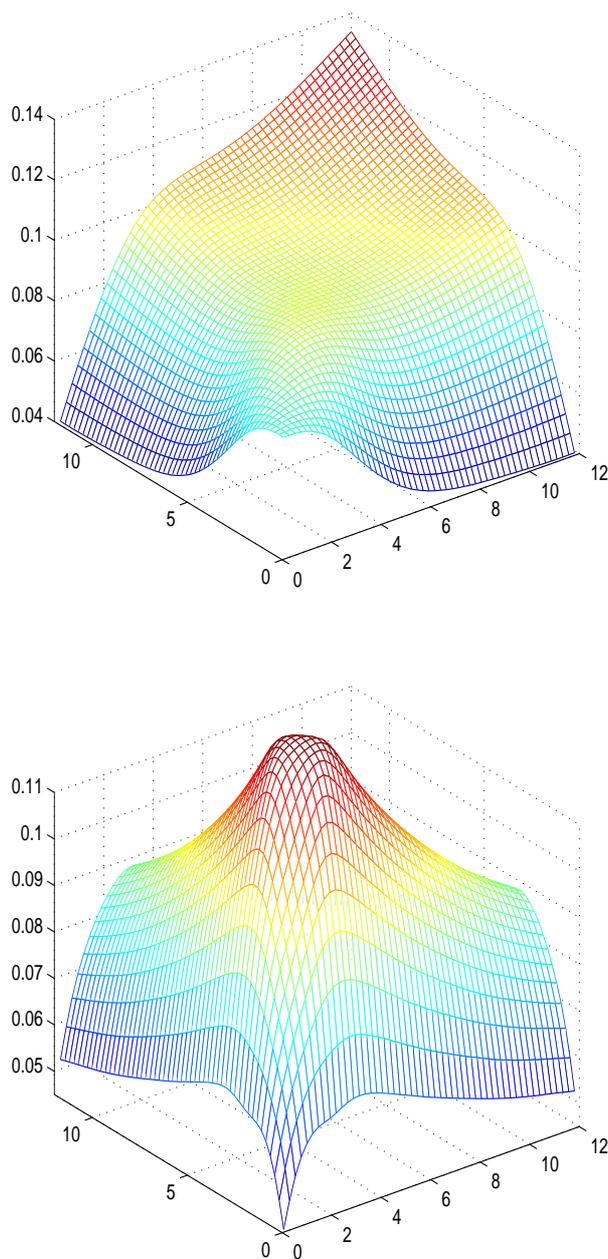


FIGURE 3. Estimated smooth covariance function of Euribor data. Top: 1999, bottom: 2007; x-axis and y-axis: maturity in months, z-axis: covariance between interest rates.

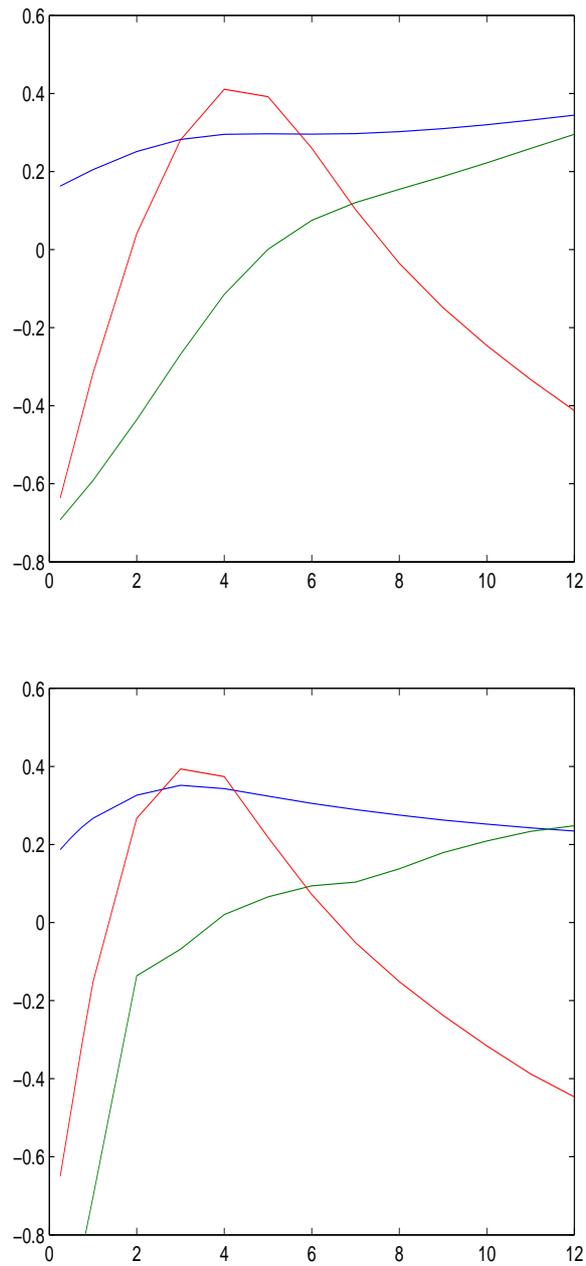


FIGURE 4. Estimated first three significant eigenfunctions of the Euribor data. x-axis: maturity in months. y-axis: estimated eigenfunctions. Top 1999 blue(79.1), green(8), red(0.5). Bottom 2007: blue(81.5), green(1.2), red(0.2). Numbers in brackets denote percentage of variance explained.

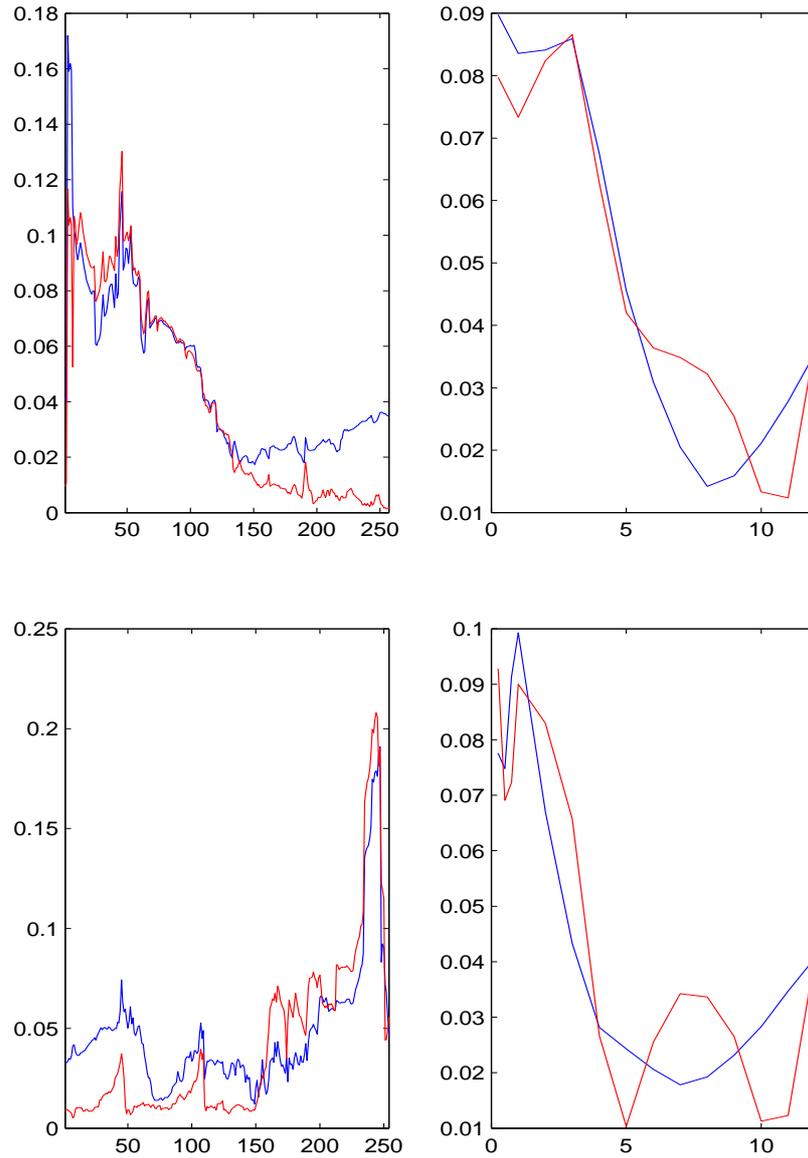


FIGURE 5. Distance between observed functions and estimated functions for the Eu-ribor data. The blue lines are estimates using the first three principal components. The red lines are for the fitted functions using the method of Diebold and Li (2006). Left panel: time series of RMSE plotted over days. Right panel: RMSE plotted over maturity. Top: 1999, bottom: 2007.

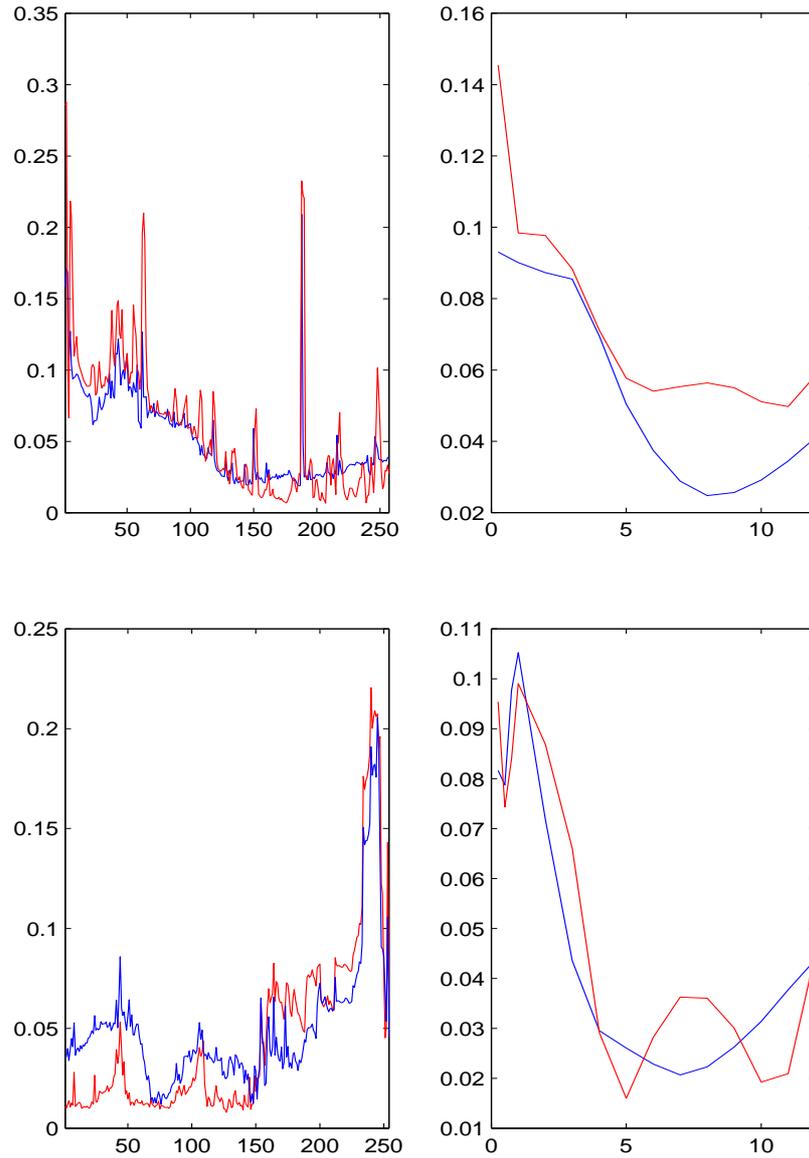


FIGURE 6. Distance between observed functions and predicted functions using VARMA modeling for the Euribor data. The blue lines are predictions using the first three principal components. The red lines are predictions using the method of Diebold and Li (2006). Left panel: time series of RMSE plotted over days. Right panel: RMSE plotted over maturity. Top: 1999 the order of ARMA is chosen to be (3,0), bottom: 2007 the order of ARMA is chosen to be (1,0).

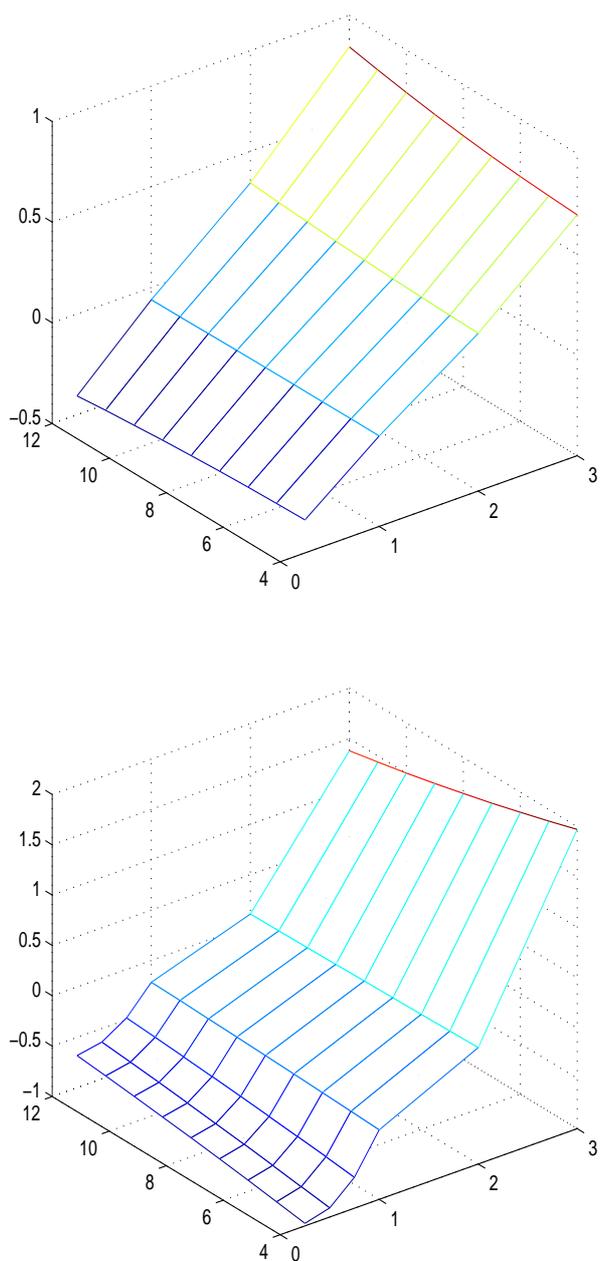


FIGURE 7. Estimated regression surface of the Euribor data. Top: 1999, bottom: 2007; x-axis and y-axis: maturity in months, z-axis: regression surface of interest rates.

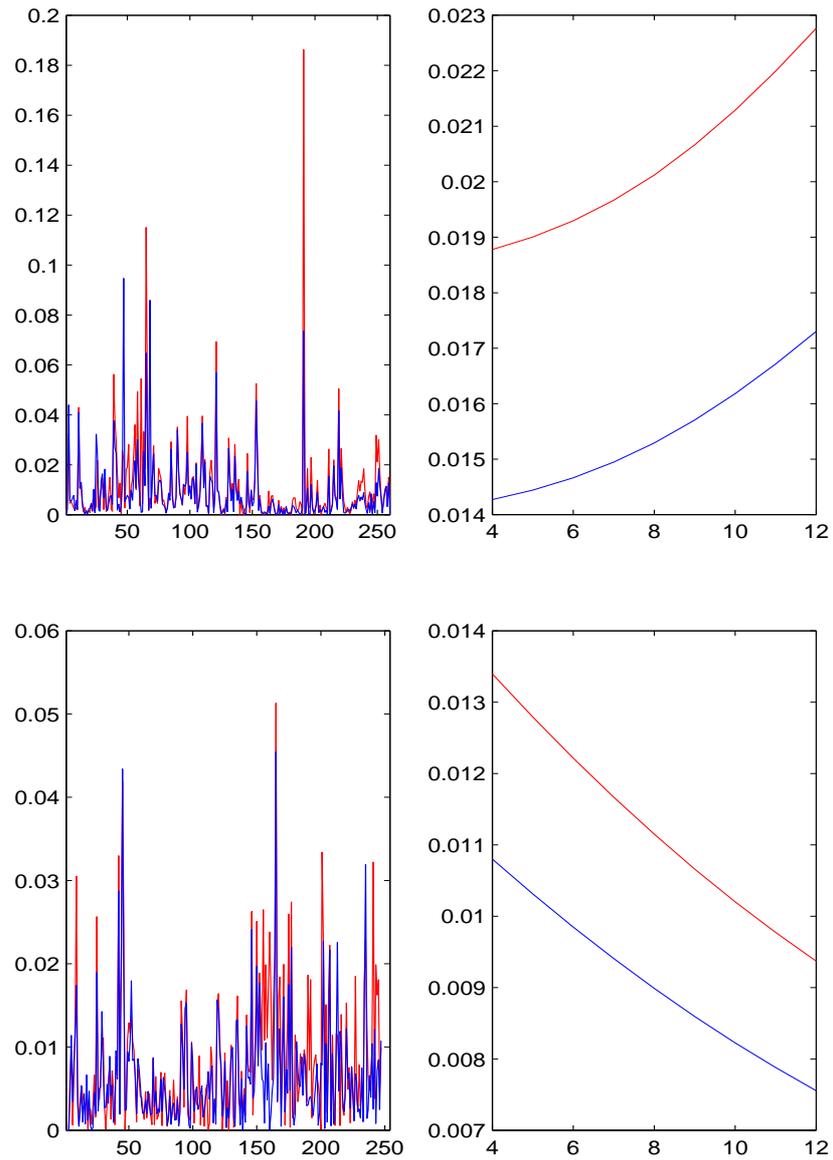


FIGURE 8. Distance between observed functions and predicted functions of long term interest rates for Euribor data. The red lines are predictions using only the time series of long term maturity rates. The blue lines are predictions using time series and regression on short term maturity rates. Left panel: time series of RMSE plotted over days. Right panel: RMSE plotted over maturity. Top: 1999, bottom: 2007.

Hour	Mean	SD	Minimum	Maximum
1	51.1599	11.7999	6.98	74.01
2	44.4660	12.6925	1.10	69.63
3	39.3625	12.5969	1.77	64.09
4	35.7377	12.5233	1.00	60.20
5	37.5708	11.7297	2.05	61.18
6	47.3861	11.6892	0.00	70.28
7	60.0419	13.8962	0.00	94.51
8	80.8758	17.7730	0.00	135.66
9	85.3308	20.9992	0.00	157.91
10	90.9880	24.4998	0.00	200.78
11	95.8930	28.0496	0.00	199.92
12	105.1220	35.8101	17.00	274.95
13	92.9396	22.8618	18.44	166.06
14	88.3578	21.4724	19.06	159.00
15	84.9993	22.5315	17.05	163.90
16	79.3881	19.7112	19.14	133.62
17	75.1704	16.9914	21.06	119.93
18	75.5914	14.6505	26.27	113.48
19	78.9728	14.5556	30.05	149.92
20	77.4729	15.5560	32.05	138.47
21	76.8082	15.7122	33.54	125.02
22	69.3300	13.2287	40.03	105.93
23	65.8497	11.3978	38.25	94.82
24	55.0160	11.3874	26.56	80.98

TABLE 5. Summary statistics for the electricity spot data 2008 grouped by the hour of the day

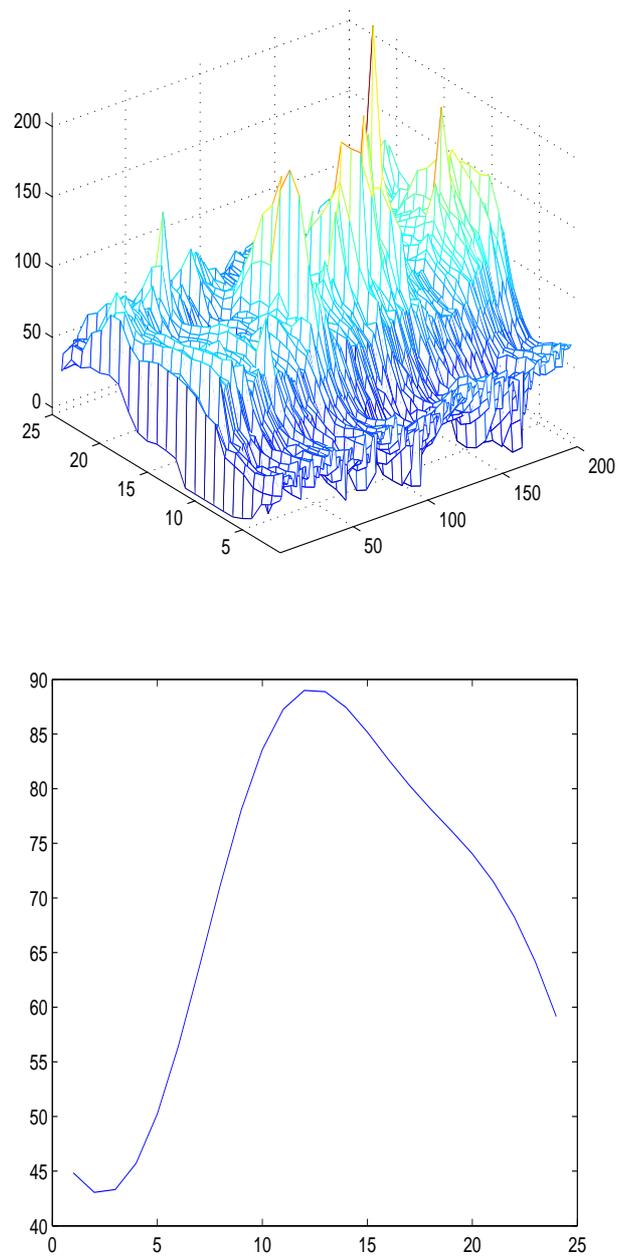


FIGURE 9. Top: raw data for electricity spot prices; x-axis: Day, y-axis: hour of the day, z-axis: spot price. Bottom: estimated smooth mean function. x-axis: hour of the day, y-axis: estimated mean.

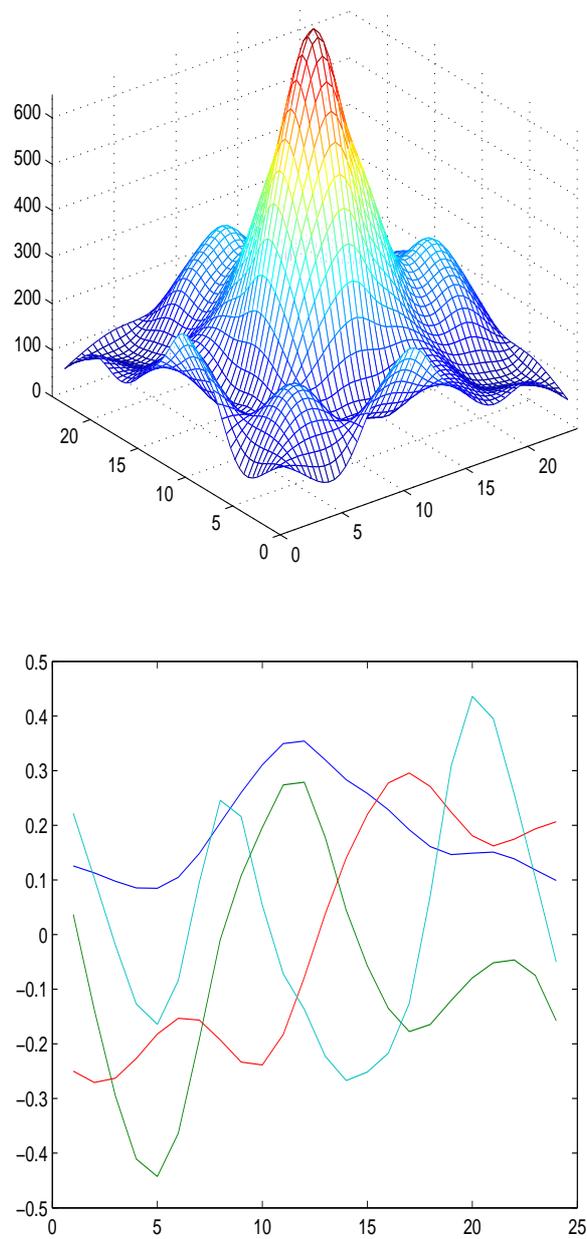


FIGURE 10. Top: smooth covariance surface of electricity data; x-axis and y-axis: hour of the day, z-axis: covariance. Bottom: estimated first four significant eigenfunctions. x-axis: hour of the day, y-axis: estimated eigenfunction. Blue(74.0), green(10.0), red(4.0), cyan(2.8). Numbers in brackets denote percentage of variance explained.

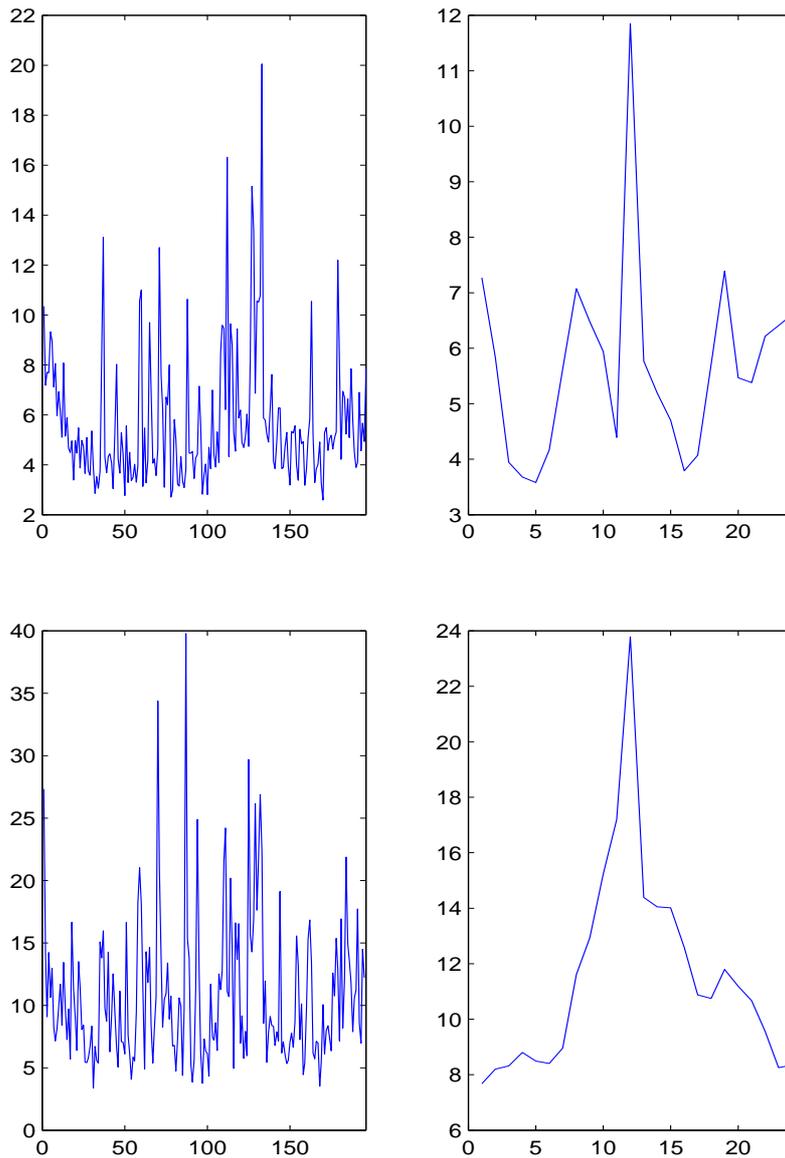


FIGURE 11. Electricity data. Top: distance between observed functions and fitted functions using the first four principal components. Bottom: distance between observed functions and predicted functions using VARMA(1,0) modeling of principal components. Left panel: time series of RMSE plotted over days. Right panel: RMSE plotted over maturity.

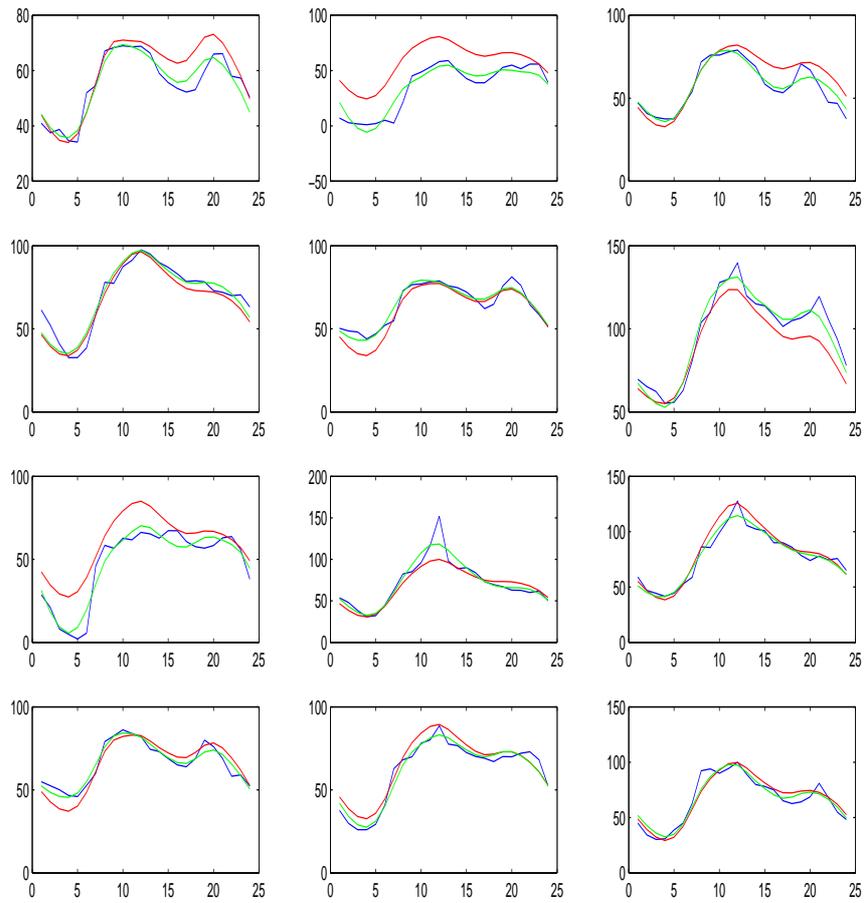


FIGURE 12. Electricity data. For 12 randomly selected days: daily observed functions(blue), fitted functions(green) and predicted functions using an VARMA(1,0) model(red) .

Week	AR	FDA	IHMAR
1	0.0874	0.1214	0.1160
2	0.1431	0.1374	0.1413
3	0.1301	0.1456	0.1576
4	0.1747	0.1352	0.1746
5	0.1009	0.0787	0.1484
6	0.0584	0.0721	0.1095
7	0.1691	0.1422	0.1672
8	0.1486	0.1469	0.1576
9	0.1105	0.1017	0.1478
10	0.1215	0.1297	0.1361
11	0.1988	0.1775	0.2164
12	0.2203	0.1530	0.2527
13	0.1229	0.1166	0.1577
14	0.1507	0.1432	0.1697
15	0.0942	0.1040	0.1191
16	0.1248	0.1016	0.1377
17	0.3689	0.2644	0.3529
18	0.1679	0.1556	0.1749
19	0.1612	0.0945	0.1948
20	0.0971	0.0941	0.1198
21	0.1055	0.1030	0.1357
22	0.1707	0.1368	0.1863
23	0.1299	0.1307	0.1438
24	0.0867	0.0736	0.1066
25	0.1478	0.1554	0.1827
26	0.1665	0.1442	0.1841
27	0.0822	0.0891	0.1238
28	0.1019	0.1158	0.1394
29	0.0839	0.0717	0.1274
30	0.0953	0.1123	0.1318
31	0.1199	0.1079	0.1270
32	0.2217	0.1785	0.2253
33	0.0833	0.0694	0.1411
34	0.0786	0.0792	0.1456
35	0.0902	0.0879	0.1397
36	0.1139	0.1107	0.1411
37	0.1111	0.0987	0.1409
38	0.0945	0.0914	0.1341
WMAE	0.1325	0.1203	0.1581
# best	11	27	0
m.d.f.b	0.5909	0.1280	1.5643

TABLE 6. WMAE errors for Electricity data. Best results in each row are emphasized in bold. Measures of fit are summarized in the bottom rows. They include the mean WMAE over all weeks, the number of times a given model was best and the mean deviation from the best model in each week (m.d.f.b.).