Statistica Sinica **28** (2018), 2469-2496
doi:https://doi.org/10.5705/ss.202017.0199

FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS FOR DERIVATIVES OF MULTIVARIATE CURVES

Maria Grith[1], Heiko Wagner[2], Wolfgang K. Härdle[3,4] and Alois Kneip[2]

[1]*Erasmus University Rotterdam,* [2]*University of Bonn,*
[3]*Humboldt University of Berlin and* [4]*Singapore Management University*

Abstract: We propose two methods based on the functional principal component analysis (FPCA) to estimate smooth derivatives for a sample of observed curves with a multidimensional domain. We apply the eigendecomposition to a) the dual covariance matrix of the derivatives; b) the dual covariance matrix of the observed curves, and take derivatives of their eigenfunctions. To handle noisy and discrete observations, we rely on local polynomial regression. We show that if the curves are contained in a finite-dimensional function space, the second method performs better asymptotically. We apply our methodology in simulations and an empirical study of option implied state price density surfaces. Using call data for the DAX 30 stock index between 2002 and 2011, we identify three components that are interpreted as volatility, skewness and tail factors, and we find evidence of term structure variation.

Key words and phrases: Derivatives, dual method, functional principal component analysis, multivariate functions, option prices, state price densities.

1. Introduction

Over the last two decades functional data analysis has became a popular tool to handle data entities that are random functions. Usually, discrete and noisy versions of them are observed. Oftentimes, these entities are multivariate functions. Examples include brain activity recordings generated during fMRI or EEG experiments (van Bömmel et al. (2014), Majer et al. (2015)). In a variety of applications though, the object of interest is not directly observable but can be recovered from the observed data by means of derivatives. Typical examples of financial applications are functionals retrieved from the observed prices, such as the implied state price density (Grith, Härdle and Schienle (2012)), pricing kernel (Grith, Härdle and Park (2013)) or market price of risk (Härdle and Lopez-Cabrera (2012)). Motivated by such, we address the problem of estimating derivatives of multivariate functions from existing discrete and noisy data.

Functions that are objects on an infinite-dimensional vector space require specific methods that allow a reasonable approximation of their variability with a small number of components. FPCA is a convenient tool in this because it allows one to explain complicated data structures with only a few orthogonal principal components that fulfill the optimal basis property in terms of L^2 accuracy. These components are given by the Karhunen-Loève theorem, see for instance Bosq (2000). In addition, the corresponding principal loadings to this basis system can be used to study the variability of the observed phenomena. An important contribution in the treatment of the finite-dimensional PCA was by Dauxois, Pousse and Romain (1982), followed by subsequent studies that fostered the applicability of the method to samples of observed noisy curves. Besse and Ramsay (1986), among others, derived theoretical results for observations that are affected by additive errors. Some of the most important contributions for the extension of the PCA to functional data belong to Cardot, Ferraty and Sarda (1999), Cardot, Mas and Sarda (2007), Ferraty and Vieu (2006), Mas (2002) and Mas (2008). Simple one-dimensional spatial curves are well understood from both numerical and theoretical perspectives and FPCA is then easy to implement. Multivariate objects with more complicated spatial and temporal correlation structures, or not directly observable functions of these objects, such as derivatives, often lack a sound theoretical framework. The computational issues are considerable in higher-dimensional domain.

To our best knowledge, FPCA for derivatives has been tackled by Hall, Müller and Yao (2009) and Liu and Müller (2009). The first study handles one-dimensional directional derivatives and gradients. The second paper analyses a particular setup in one-dimensional domain where the observations are sparse. The method is applicable to non-sparse data but can be computationally inefficient when dealing with large amounts of observations per curve. For the study of observed curves, there are a series of empirical studies for the two-dimensional domain case, see Cont and da Fonseca (2002) for an application close to our empirical study. Further proposals to implement FPCA in more than two dimensions to analyze functions, rather than their derivatives, have been done particularly in the area of brain imaging, see for instance, Zipunnikov et al. (2011) who implement multilevel FPCA (Staicu, Crainiceanu and Carroll (2010), Di et al. (2009)) to analyze brain images of different groups of individuals. A thorough derivation of statistical properties of the estimators is missing in these works.

In this article, we aim to contribute to the literature on FPCA for the study of derivatives of multivariate functions. We present two approaches to estimating

the derivatives. They are not tailored to handle sparse data sets, compared to other methods that aim to estimate the mean or covariance function of a sample of curves, see for instance Cai and Yuan (2011) and Zhang and Wang (2016). Our approaches are feasible when the spatial dimension increases only under suitable smoothness assumptions of the underlying curves. Otherwise, the estimators that we propose suffer from the curse of dimensionality.

The paper is organized as follows: the theoretical framework, estimation procedure and statistical properties are derived in Section 2. Our empirical study in Section 3 is guided by the estimation and the dynamics analysis of the option-implied state price densities. It includes a simulation study and a data example.

2. Methodology

2.1. Two approaches to model derivatives using FPCA

Let X be a centered smooth random function in $L^2([0,1]^g)$, where g denotes the spatial dimension, with finite second moment $\int_{[0,1]^g} \mathsf{E}\left\{X(t)^2\right\} dt < \infty$ for $t = (t_1, \ldots, t_g)^\top$. The underlying dependence structure can be characterized by the covariance function $\sigma(t,v) \stackrel{\text{def}}{=} \mathsf{E}\{X(t)X(v)\}$ and the corresponding covariance operator Γ

$$(\Gamma\vartheta)(t) = \int_{[0,1]^g} \sigma(t,v)\vartheta(v)dv.$$

Mercer's lemma guarantees the existence of a set of eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots$ and a corresponding system of orthonormal eigenfunctions $\gamma_1, \gamma_2, \ldots$ called functional principal components such that

$$\sigma(t,v) = \sum_{r=1}^{\infty} \lambda_r \gamma_r(t)\gamma_r(v), \tag{2.1}$$

where the eigenvalues and eigenfunctions satisfy $(\Gamma\gamma_r)(t) = \lambda_r\gamma_r(t)$. Moreover, $\sum_{r=1}^{\infty} \lambda_r = \int_{[0,1]^g} \sigma(t,t)dt$. The Karhunen-Loève decomposition applied to the random function X gives

$$X(t) = \sum_{r=1}^{\infty} \delta_r \gamma_r(t), \tag{2.2}$$

where the loadings δ_r are random variables defined as $\delta_r = \int_{[0,1]^g} X(t)\gamma_r(t)dt$ that satisfy $\mathsf{E}\left(\delta_r^2\right) = \lambda_r$, as well as $\mathsf{E}\left(\delta_r\delta_s\right) = 0$ for $r \neq s$. Throughout the paper, we use this notation for the derivatives of a function X

$$X^{(d)}(t) \stackrel{\text{def}}{=} \frac{\partial^{|d|}}{\partial t^d} X(t) = \frac{\partial^{d_1}}{\partial t_1^{d_1}} \cdots \frac{\partial^{d_g}}{\partial t_g^{d_g}} X(t_1, \ldots, t_g), \tag{2.3}$$

for $d = (d_1, \ldots, d_g)^\top$ and $d_j \in \mathbb{N}$ the partial derivative in the spatial direction $j = 1, \ldots, g$. We take $|d| = \sum_{j=1}^{g} |d_j|$ and require that X is at least $|d| + 1$ times continuously differentiable.

Building on (2.1) and (2.2), we propose two approaches to model derivatives. The first one is stated in terms of the Karhunen-Loève decomposition applied to $X^{(d)}$ and uses the FPCA of the covariance function $\sigma^{(d)}(t, v) \stackrel{\text{def}}{=} \mathsf{E}\left\{X^{(d)}(t)X^{(d)}(v)\right\}$, assumed to be continuous for all $t, v \in [0, 1]^g$. With $\lambda_1^{(d)} \geq \lambda_2^{(d)} \geq \cdots$ denoting the eigenvalues of the corresponding covariance operator, functional principal components $\varphi_r^{(d)}$, $r = 1, 2, \ldots$ are solutions to the eigenequation

$$\int_{[0,1]^g} \sigma^{(d)}(t, v)\varphi_r^{(d)}(v)dv = \lambda_r^{(d)}\varphi_r^{(d)}(t). \tag{2.4}$$

Similarly to (2.2), the decomposition of $X^{(d)}$ in terms of principal components $\varphi_r^{(d)}(t)$ is given by

$$X^{(d)}(t) = \sum_{r=1}^{\infty} \delta_r^{(d)}\varphi_r^{(d)}(t), \tag{2.5}$$

for $\delta_r^{(d)} = \int_{[0,1]^g} X^{(d)}(t)\varphi_r^{(d)}(t)dt$.

By abuse of notation, $\varphi_r^{(d)}$ denotes the r-th eigenfunction of the covariance operator of $\sigma^{(d)}$ and not the d-th derivative of φ_r. For $|d| = 0$ we introduce the equivalent notations $\gamma_r(t) \equiv \varphi_r^{(0)}(t)$, $\sigma(t, v) \equiv \sigma^{(0)}(t, v)$, $\lambda_r \equiv \lambda_r^{(0)}$ and $\delta_r \equiv \delta_r^{(0)}$.

A different way to obtain a decomposition of $X^{(d)}$ is to differentiate the left and right hand sides of (2.2), which leads to

$$X^{(d)}(t) = \sum_{r=1}^{\infty} \delta_r\gamma_r^{(d)}(t), \tag{2.6}$$

where the d-th derivative of the r-th eigenfunction is the solution to

$$\int_{[0,1]^g} \frac{\partial^{|d|}}{\partial v^d} \left\{\sigma(t, v)\gamma_r(v)\right\} dv = \lambda_r\gamma_r^{(d)}(t). \tag{2.7}$$

In general, for $|d| > 0$ it holds that $\varphi_r^{(d)}(t) \neq \gamma_r^{(d)}(t)$, but both basis systems span the same function space. In particular, there always exists a projection a with $a_{rp} = \langle\gamma_r^{(d)}, \varphi_p^{(d)}\rangle = \int_{[0,1]^g} \gamma_r^{(d)}(t)\varphi_p^{(d)}(t)dt$ such that $\sum_{r=1}^{\infty} a_{rp}\varphi_r^{(d)}(t) = \gamma_p^{(d)}(t)$, for all pairs $r, p = 1, 2, \ldots$. However, if we consider a truncation of (2.2) after a finite number of components this is no longer true in general. An

advantage of using $\varphi_r^{(d)}$ instead of $\gamma_r^{(d)}$ is that the Karhunen-Loève decomposition uses orthonormal bases that fulfill the best basis property, such that for any fixed $L \in \mathbb{N}$ and every other orthonormal basis system φ'

$$\sum_{r=L+1}^{\infty} \lambda_r^{(d)} = \mathsf{E}\left\| X^{(d)} - \sum_{r=1}^{L} \left\langle X^{(d)}, \varphi_r^{(d)} \right\rangle \varphi_r^{(d)} \right\| \leq \mathsf{E}\left\| X^{(d)} - \sum_{r=1}^{L} \left\langle X^{(d)}, \varphi_r' \right\rangle \varphi_r' \right\|. \tag{2.8}$$

This guarantees that, by using $\varphi_r^{(d)}$, $r = 1, \ldots, L$, we always achieve the best L-dimensional subset selection in terms of the L^2 error function. We show that estimating the basis functions with such a property comes at the cost of slower rate of convergence. In addition, if the true underlying structure of X lies in a L-dimensional function space, which is equivalent to a factor model, the advantage of deriving the best L-orthogonal basis vanishes because $span\{\gamma_1^{(d)}, \ldots, \gamma_L^{(d)}\} = span\{\varphi_1^{(d)}, \ldots, \varphi_L^{(d)}\}$.

2.2. Sample inference

In practice, the true eigenfunctions are unknown. Let $X_1, \ldots, X_N \in L^2([0,1]^g)$ be a sample of i.i.d. realizations of the smooth random function X. For some m assume that X_i is a.s. m-times continuously differentiable in each direction $j = 1, \ldots, g$. Let $\nu = (\nu_1, \ldots, \nu_g)^\top$, $\nu_j \in \mathbb{N}$, $|\nu| < m$. The two cases of interest are $\nu = (0, \ldots, 0)^\top$ and $\nu = d$.

The empirical approximation of the covariance function based on a sample of N curves is given by

$$\tilde{\sigma}^{(\nu)}(t,v) = \frac{1}{N} \sum_{i=1}^{N} X_i^{(\nu)}(t) X_i^{(\nu)}(v) \tag{2.9}$$

and of the covariance operator by

$$(\tilde{\Gamma}^{(\nu)} \tilde{\varphi}_r^{(\nu)})(t) = \int_{[0,1]^g} \tilde{\sigma}^{(\nu)}(t,v) \tilde{\varphi}_r^{(\nu)}(v) dv = \tilde{\lambda}_r^{(\nu)} \tilde{\varphi}_r^{(\nu)}(t), \tag{2.10}$$

where eigenfunction $\tilde{\varphi}_r^{(\nu)}$ corresponds to the r-th eigenvalue $\tilde{\lambda}_r^{(\nu)}$ of $\tilde{\Gamma}^{(\nu)}$. Then we get $X_i^{(\nu)}(t) = \sum_{r=1}^{N} \tilde{\delta}_{ri}^{(\nu)} \tilde{\varphi}_r^{(\nu)}(t)$, where $\tilde{\delta}_{ri}^{(\nu)} = \int_{[0,1]^g} X_i^{(\nu)}(t) \tilde{\varphi}_r^{(\nu)}(t) dt$. Note that for $\nu = (0, \ldots, 0)^\top$ we have $\tilde{\gamma}_r \equiv \tilde{\varphi}_r^{(0)}(t)$, and $\tilde{\lambda}_r \equiv \tilde{\lambda}_r^{(0)}$, $\tilde{\delta}_{ri} \equiv \tilde{\delta}_{ri}^{(0)}$. Following (2.6), $X_i^{(\nu)}(t) = \sum_{r=1}^{N} \tilde{\delta}_{ri} \tilde{\gamma}_r^{(\nu)}(t)$.

Theoretical properties of $\tilde{\varphi}_r^{(\nu)}$ are well studied. Under some regularity conditions we obtain $||\varphi_r^{(\nu)} - \tilde{\varphi}_r^{(\nu)}|| = \mathcal{O}_p(N^{-1/2})$ and $|\lambda_r^{(\nu)} - \tilde{\lambda}_r^{(\nu)}| = \mathcal{O}_p(N^{-1/2})$, see for instance Dauxois, Pousse and Romain (1982) or Hall and Hosseini-Nasab (2006).

2.3. The model

In most applications, the curves are only observed at discrete points and data is noisy. To model these aspects, we assume that each curve in the sample is observed at a random grid $t_i = (t_{i1}, \ldots, t_{iT_i})^\top,\ t_{ik} \in [0,1]^g,\ k = 1, \ldots, T_i, i = 1, \ldots, N$ having a common bounded and continuously differentiable density f with support $\text{supp}(f) = [0,1]^g$ and the integrand $u \in \text{supp}(f)$ with $\inf_u f(u) > 0$. Then

$$Y_i(t_{ik}) = X_i(t_{ik}) + \varepsilon_{ik} = \sum_{r=1}^{\infty} \delta_{ri}\gamma_r(t_{ik}) + \varepsilon_{ik}, \tag{2.11}$$

where ε_{ik} are i.i.d. random variables with $\mathsf{E}[\varepsilon_{ik}] = 0$, $\text{Var}(\varepsilon_{ik}) = \sigma^2_{i\varepsilon}$ and let ε_{ik} be independent of X_i. We take $Y_i = (Y_i(t_{i1}), \ldots, Y_i(t_{iT_i}))^\top$ to be the vector of observations of X_i.

2.4. Estimation procedure

2.4.1. Dual method

Under (2.11), the empirical principal components have to be recovered from the discrete, noisy data. An efficient estimation procedure when the number of observations per individual curves T_i is larger than the sample size N relies on the duality relation between the row and column space. The method was first used in a functional context by Kneip and Utikal (2001) to estimate density functions and later adapted by Benko, Härdle and Kneip (2009) to general functions. The underlying idea is that integrals of smooth functions can be estimated more accurately than specific functional values.

Let $M^{(\nu)}$ be the dual matrix of $\tilde{\sigma}^{(\nu)}(t,v)$ from (2.9) consisting of entries

$$M^{(\nu)}_{ij} = \int_{[0,1]^g} X^{(\nu)}_i(t) X^{(\nu)}_j(t) dt. \tag{2.12}$$

Let $l^{(\nu)}_r$, $r = 1, \ldots, N$ be the eigenvalues of matrix $M^{(\nu)}$ and $p^{(\nu)}_r = (p^{(\nu)}_{1r}, \ldots, p^{(\nu)}_{Nr})^\top$ the corresponding eigenvectors. For $\nu = d$, the eigendecomposition of $M^{(d)}$ is relevant for the empirical version of equation (2.5), leading to

$$\tilde{\varphi}^{(d)}_r(t) = \frac{1}{\sqrt{l^{(d)}_r}} \sum_{i=1}^{N} p^{(d)}_{ir} X^{(d)}_i(t)\ ,\ \tilde{\lambda}^{(d)}_r = \frac{l^{(d)}_r}{N} \text{ and } \tilde{\delta}^{(d)}_{ri} = \sqrt{l^{(d)}_r} p^{(d)}_{ir}. \tag{2.13}$$

Important for an empirical version of equation (2.6) are the eigenvalues and eigenvectors of $M^{(0)}$ denoted by $l_r \equiv l^{(0)}_r, p_r \equiv p^{(0)}_r$. Then

$$\tilde{\gamma}_r^{(d)}(t) = \frac{1}{\sqrt{l_r}} \sum_{i=1}^{N} p_{ir} X_i^{(d)}(t) \ , \ \tilde{\lambda}_r = \frac{l_r}{N} \text{ and } \tilde{\delta}_{ri} = \sqrt{l_r} p_{ir}. \tag{2.14}$$

2.4.2. Estimation of $M^{(0)}$ and $M^{(d)}$

There are challenges when estimating $M^{(0)}$ and $M^{(d)}$: we observe discrete noisy curves in (2.11) and each curve is observed at irregular points. To handle such difficulties smoothing methods are commonly used. We implement local polynomial regressions as better suited to estimate integrals like (2.12) than other kernel methods, e.g., Nadaraya-Watson or Gasser-Müller estimators, because the bias and variance are of the same order of magnitude near the boundary as well as in the interior, see Fan and Gijbels (1992).

For any vectors a, $b \in \mathbb{R}^g$ and $c \in \mathbb{N}^g$, we take $|a| \stackrel{\text{def}}{=} \sum_{j=1}^{g} |a_j|$, $a^{-1} \stackrel{\text{def}}{=} (a_1^{-1}, \ldots, a_g^{-1})^\top$, $a^b \stackrel{\text{def}}{=} a_1^{b_1} \times \cdots \times a_g^{b_g}$, $a \circ b \stackrel{\text{def}}{=} (a_1 b_1, \ldots, a_g b_g)^\top$ and $c! \stackrel{\text{def}}{=} c_1! \times \cdots \times c_g!$.

For arbitrary $i = 1, \ldots, N$, consider the multivariate local polynomial estimator $\hat{\beta}_i(t) \in \mathbb{R}^\rho$ defined by a lexicographical arrangement, see Masry (1996), that solves

$$\min_{\beta_i(t)} \sum_{l=1}^{T_i} \left\{ Y_i(t_{il}) - \sum_{0 \le |k| \le \rho} \beta_{i,k}(t)(t_{il} - t)^k \right\}^2 K_B(t_{il} - t). \tag{2.15}$$

K_B is a non-negative, symmetric and bounded multivariate kernel function, B a $g \times g$ bandwidth matrix. For simplicity, we assume that B has main diagonal entries $b = (b_1, \ldots, b_g)^\top$ and zeros elsewhere. ρ satisfying $|\nu| < \rho \le m$ is the order of the local polynomial expansion used in (2.15). In our application, for the two dimensional case, if $\nu = (0, 0)^\top$ then $\rho = 1$ and if $\nu = (2, 0)^\top$ then $\rho = 3$.

As noted by Fan et al. (1997) the solution for the minimization problem in (2.15) can be represented using a weight function $W_\nu^{T_i}$, see the online Supplementary Material S3, such that the local polynomial estimator of $X_i^{(\nu)}$ is given by

$$\hat{X}_{i,b}^{(\nu)}(t) = \nu! \hat{\beta}_{i,\nu}(t) = \nu! \sum_{l=1}^{T_i} W_\nu^{T_i} \left((t_{il} - t) \circ b^{-1} \right) Y_i(t_{il}). \tag{2.16}$$

We estimate off-diagonal terms $M_{ij}^{(\nu)}$, $j \neq i$, by $\int_{[0,1]^g} \hat{X}_{i,b}^{(\nu)}(t) \hat{X}_{j,b}^{(\nu)}(t) dt$ given (2.16) and, due to the presence of squared error terms, a diagonal correction is additionally applied for $i = j$. This leads to

$$\hat{M}_{ij}^{(\nu)} = \begin{cases} \nu!^2 \sum_{k=1}^{T_i} \sum_{l=1}^{T_j} w_\nu(t_{ik}, t_{jl}, b) Y_j(t_{jl}) Y_i(t_{ik}) & \text{if } i \neq j, \\ \nu!^2 \left\{ \sum_{k=1}^{T_i} \sum_{l=1}^{T_i} w_\nu(t_{ik}, t_{il}, b) Y_i(t_{il}) Y_i(t_{ik}) \right. & \\ \qquad \left. - \hat{\sigma}_{i\varepsilon}^2 \sum_{k=1}^{T_i} w_\nu(t_{ik}, t_{ik}, b) \right\} & \text{if } i = j, \end{cases} \tag{2.17}$$

where $w_\nu(t_{ik}, t_{jl}, b) := \int_{[0,1]^g} W_\nu^{T_i}\left((t_{ik} - s) \circ b^{-1}\right) W_\nu^{T_j}\left((t_{jl} - s) \circ b^{-1}\right) ds$. The estimator for $M^{(0)}$ is given by setting $\nu = (0, \ldots, 0)^\top$ and the estimator for $M^{(d)}$ by $\nu = d$.

Rates of convergence for these estimators are given by Proposition 1 that relies on Assumptions 1 - 6 as given in the Supplementary Material S1. Apart from regularity conditions, essential requirements are that, for some integer m, sample functions are m times continuously differentiable. Furthermore, it is assumed that consistent estimators of the error variances are used that satisfy $|\sigma_{i\varepsilon}^2 - \hat{\sigma}_{i\varepsilon}^2| = \mathcal{O}_P(T^{-1/2})$.

Proposition 1. *Suppose Assumptions* 1 - 6 *hold and that* $m \geq \max(|\nu|+2, 2|\nu|)$, *and that the local polynomial regression is of order* ρ *with* $|\nu| < \rho \leq m$. *If* $T := \min_i(T_i) \to \infty$ *with* $\max(b)^{\rho+1} b^{-\nu} \to 0$, $\log(T)/(Tb_1 \times \cdots \times b_g) \to 0$ *and* $Tb_1 \times \cdots \times b_g b^{4\nu} \to \infty$, *then, for all* $i, j \in \{1, \ldots, N\}$,

$$|M_{ij}^{(\nu)} - \hat{M}_{ij}^{(\nu)}| = \mathcal{O}_P\left(\max(b)^{\rho+1} b^{-\nu} + \left(\frac{1}{T^2 b_1 \times \cdots \times b_g b^{4\nu}} + \frac{1}{T}\right)^{1/2}\right).$$

Notation $b^\nu = b_1^{\nu_1} \times \cdots \times b_1^{\nu_g}$ was introduced earlier. A proof of the proposition is given in the Supplementary Material S2. By Proposition 1, estimating $M^{(d)}$ gives an asymptotic higher bias and also a higher variance than estimating $M^{(0)}$. This effect is more pronounced for larger g, but one can get parametric rates within each method if using local polynomial regression with large polynomial order ρ is feasible.

Remark 1. We can infer from Proposition 1 that if

$$m > \rho \geq \frac{g}{2} - 1 + 3\sum_{l=1}^{g} \nu_l, \quad b_j = C_j T^{-\alpha} \text{ for } 0 < C_j < \infty, \tag{2.18}$$

for $j = 1, \ldots, g$, $1/\{2(\rho + 1 - \sum_{l=1}^g \nu_l)\} \leq \alpha \leq 1/(g + 4\sum_{l=1}^g \nu_l)$, then $|M_{ij}^{(\nu)} - \hat{M}_{ij}^{(\nu)}| = \mathcal{O}_P(1/\sqrt{T})$.

Here the orders of polynomial expansion and the bandwidths for estimating $M^{(\nu)}$ differ for $\nu = (0, \ldots, 0)^\top$ and $\nu = d$. In particular, the estimator of $M^{(d)}$ requires higher smoothness assumptions via $m > \rho$, and a higher bandwidth to achieve the same parametric convergence rate as the estimator for $M^{(0)}$.

Remark 2. If the dimension g is larger than 1, then it is nontrivial to calculate integrals of the form $\int_{[0,1]^g} \hat{X}_{i,b}^{(\nu)}(t)\hat{X}_{j,b}^{(\nu)}(t)dt$ numerically. In our Matlab implementation, such integrals are determined via Monte Carlo integration. We draw random samples u_l, $l = 1, \ldots, T^*$, $T^* \geq T = \min_i(T_i)$, uniform on $[0,1]^g$, and use $\int_{[0,1]^g} \hat{X}_{i,b}^{(\nu)}(t)\hat{X}_{j,b}^{(\nu)}(t)dt \approx (1/T^*)\sum_{l=1}^{T^*} \hat{X}_{i,b}^{(\nu)}(u_l)\hat{X}_{j,b}^{(\nu)}(u_l)$. Since u_l is independent of all observations $(Y_i(t_{ik}), t_{ik})$, and since all $\hat{X}_{i,b}^{(\nu)}(t)$ are continuous functions, we have

$$\mathsf{E}\left\{\frac{1}{T^*}\sum_{l=1}^{T^*}\hat{X}_{i,b}^{(\nu)}(u_l)\hat{X}_{j,b}^{(\nu)}(u_l)\middle|\ \hat{X}_{i,b}^{(\nu)}, \hat{X}_{j,b}^{(\nu)}\right\} = \int_{[0,1]^g}\hat{X}_{i,b}^{(\nu)}(t)\hat{X}_{j,b}^{(\nu)}(t)dt$$

$$\operatorname{Var}\left(\frac{1}{T^*}\sum_{l=1}^{T^*}\hat{X}_{i,b}^{(\nu)}(u_l)\hat{X}_{j,b}^{(\nu)}(u_l)\middle|\ \hat{X}_{i,b}^{(\nu)}, \hat{X}_{j,b}^{(\nu)}\right) = \frac{1}{T^*}\left[\int_{[0,1]^g}\hat{X}_{i,b}^{(\nu)}(t)^2\hat{X}_{j,b}^{(\nu)}(t)^2dt - \left\{\int_{[0,1]^g}\hat{X}_{i,b}^{(\nu)}(t)\hat{X}_{j,b}^{(\nu)}(t)dt\right\}^2\right].$$

When computing $\hat{M}_{ij}^{(\nu)}$ according to (2.17) this of course means replacing $w_\nu(t_{ik}, t_{jl}, b)$ by $(1/T^*)\sum_{l=1}^{T^*} W_\nu^{T_i}\left((t_{ik} - u_l)\circ b^{-1}\right) W_\nu^{T_j}\left((t_{jl} - u_l)\circ b^{-1}\right)$. Since $T^* \geq T = \min_i(T_i)$, this implies that this approximation has asymptotically no effect on the rate of convergence derived in Proposition 1, since the additional error is of order $T^{-1/2}$ regardless of dimension g.

In Proposition 1 it is required that $|\sigma_{i\varepsilon}^2 - \hat{\sigma}_{i\varepsilon}^2| = \mathcal{O}_p(T^{-1/2})$, which ensures parametric rates of convergence for $\hat{M}^{(\nu)}$ under the conditions of Remark 1. By Assumption 3, in the univariate case, a simple class of estimators for $\sigma_{i\varepsilon}^2$ that achieve the desired convergence rate are given by successive differences, see von Neumann et al. (1941) and Rice (1984). However, as pointed out in Munk et al. (2005), difference estimators are no longer consistent for $g \geq 4$. A possible solution is to generalize the kernel-based variance estimator proposed by Hall and Marron (1990) to the multidimensional case

$$\hat{\sigma}_{i\varepsilon}^2 = \frac{1}{v_i}\sum_{l=1}^{T_i}\left\{Y_i(t_{il}) - \sum_{k=1}^{T_i} w_{ilk}Y(t_{ik})\right\}^2, \tag{2.19}$$

where $w_{ilk} = K_{s,B'}(t_{il} - t_{ik})/\sum_{k=1}^{T_i} K_{s,B'}(t_{il} - t_{ik})$, $v_i = T_i - 2\sum_l w_{ilk} + \sum_{l,k} w_{ilk}^2$

and $K_{s,B'}$ is a g-dimensional product kernel of order s with bandwidth matrix B'. Munk et al. (2005) show that if $4s > g$ and if the elements of the diagonal matrix B' are of order $\mathcal{O}(T^{-2/(4s+g)})$ then the estimator $\hat{\sigma}^2_{i\varepsilon}$ in (2.19) achieves parametric rates of convergence.

In the special case that the curves are observed at a common uniform random grid with $T = T_i = T_j$, $i, j = 1, \ldots, N$, a simple estimator for $M^{(0)}$ is constructed by approximating (2.12) directly through Monte-Carlo integral (see the Supplementary Material S4). This estimator is given by

$$\tilde{M}^{(0)}_{ij} = \begin{cases} \dfrac{1}{T}\displaystyle\sum_{l=1}^{T} Y_i(t_l)Y_j(t_l) & \text{if } i \neq j, \\ \dfrac{1}{T}\displaystyle\sum_{k=1}^{T} Y_i(t_l)^2 - \hat{\sigma}^2_{i\varepsilon} & \text{if } i = j. \end{cases} \tag{2.20}$$

When working with more than one spatial dimension, data is often recorded using an equidistant grid with T points in each direction. For our approach, this strategy does not improve the convergence rate of $\tilde{M}^{(0)}$. If it is possible to influence how data is recorded, we recommend using a common uniform random grid that keeps computing time and the storage space of data at a minimum and still gives parametric convergence rates for the estimator of $M^{(0)}_{ij}$. This constitutes a very special situation and, in particular, for estimating $M^{(d)}$ smoothing is always necessary.

The estimation of the eigensystem through the dual method involves the estimation of the $N \times N$ matrix $M^{(\nu)}$. The consistency of $\hat{M}^{(\nu)}_{ij}$ is shown in Proposition 1 to depend only on $T \leq \min_i(T_i)$ and not on the sample size N. Furthermore, in Remark 1 we derive the bandwidth rule under which $\hat{M}^{(\nu)}_{ij}$ achieves $1/\sqrt{T}$ rate. We use this convergence rate for $\hat{M}^{(\nu)}_{ij}$ to establish asymptotic results for the empirical eigenvalues and loadings, as well as pointwise convergence for the empirical eigenfunctions, the derivatives of eigenfunctions and our proposed estimators for the derivatives of the individual curves.

2.4.3. Estimation of the basis functions

In order to estimate $\tilde{\varphi}^{(d)}_r$ under (2.11) we first determine eigenvalues $\hat{l}^{(d)}_r$ and eigenvectors $\hat{p}^{(d)}_r$ of $\hat{M}^{(d)}$. Following (2.13) a corresponding estimator is given by

$$\hat{\varphi}^{(d)}_{r,T}(t) = \frac{1}{\sqrt{\hat{l}^{(d)}_r}} \sum_{i=1}^{N} \hat{p}^{(d)}_{ir} \hat{X}^{(d)}_{i,h}(t), \tag{2.21}$$

where, similar to (2.16), $\hat{X}_{i,h}^{(d)}$ denotes the local polynomial kernel estimator of $X_i^{(d)}$ with polynomial order p and bandwidth vector $h = (h_1, \ldots, h_g)^\top$. Analogously, based on eigenvalues $\hat{l}_r \equiv \hat{l}_r^{(0)}$ and eigenvectors $\hat{p}_r \equiv \hat{p}_r^{(0)}$ of $\hat{M}^{(0)}$, estimators $\hat{\gamma}_{r,T}^{(d)}$ of $\tilde{\gamma}_r^{(d)}$ are obtained using (2.14).

In (2.21), h is not equal to b, the bandwidth used to smooth the entries of $\hat{M}^{(0)}$ or $\hat{M}^{(d)}$. We show below that the optimal order for the bandwidth vector h differs asymptotically from that of b derived in the previous section. An advantage of using local polynomial estimators, compared for example to spline or wavelet estimators, is that the bias and variance can be derived analytically. For the univariate case these results can be found in Fan and Gijbels (1996), and for the multivariate case in Masry (1996) and Gu, Li and Yang (2015). We summarize them as

$$\begin{aligned} \operatorname{Bias}\left(\hat{X}_{i,h}^{(d)}(t)|Y_i,\, t_i\right) &= \mathcal{O}_p(\max(h)^{p+1}h^{-d}), \\ \operatorname{Var}\left(\hat{X}_{i,h}^{(d)}(t)\big|Y_i,\, t_i\right) &= \mathcal{O}_p\left(\frac{1}{Th_1 \times \cdots \times h_g h^{2d}}\right). \end{aligned} \tag{2.22}$$

These results provide a basis for inference of eigenfunctions. We consider fixed components with nonzero eigenvalues $\lambda_r^{(\nu)} > 0$. Under Assumptions 1 - 8, the results of Hall and Hosseini-Nasab (2006) imply that $\lambda_r^{(\nu)} - \tilde{\lambda}_r^{(\nu)} = \mathcal{O}_p(N^{-1/2})$, and hence $l_r^{(\nu)} = N\lambda_r \cdot \{1 + O_P(N^{-1/2})\}$. Using (2.22), it follows that for $\max(h)^{p+1}h^{-d} \to 0$, $\left\{\max(h)^{p+1}Th^{-d}\right\}^{-1} \to 0$ as $T \to \infty$, and for p chosen such that $p - |d|$ is odd, one has that

$$\begin{aligned} &\mathsf{E}\left[\frac{1}{\sqrt{l_r^{(\nu)}}}\sum_{i=1}^{N} p_{ir}^{(\nu)}\left\{X_i^{(d)}(t) - \hat{X}_{i,h}^{(d)}(t)\right\}\Big|Y_i,\, t_i\right] \\ &= \frac{1}{\sqrt{l_r^{(\nu)}}}\sum_{j=1}^{N} p_{jr}^{(\nu)} \operatorname{Bias}\left(\hat{X}_{j,h}^{(d)}(t)\big|Y_i,\, t_i\right) + o_p\left(\max(h)^{p+1}h^{-d}\right) \\ &= \mathcal{O}_p(\max(h)^{p+1}h^{-d}), \\ &\operatorname{Var}\left(\frac{1}{\sqrt{l_r^{(\nu)}}}\sum_{i=1}^{N} p_{ir}^{(\nu)}\hat{X}_{i,h}^{(d)}(t)\Big|Y_i,\, t_i\right) \\ &= \frac{1}{l_r^{(\nu)}}\sum_{j=1}^{N}\left(p_{jr}^{(\nu)}\right)^2 \operatorname{Var}\left(\hat{X}_{j,h}^{(d)}(t)\big|Y_i,\, t_i\right) + o_p\left(\frac{1}{NTh_1 \times \cdots \times h_g h^{2d}}\right) \\ &= \mathcal{O}_p\left(\frac{1}{NTh_1 \times \cdots \times h_g h^{2d}}\right). \end{aligned}$$

We show that, under certain assumptions, the asymptotic mean squared error of $\hat{\varphi}_{r,T}^{(d)}$ and $\hat{\gamma}_{r,T}^{(d)}$ is dominated by these terms, while the effects of replacing $p_r^{(d)}$ by $\hat{p}_r^{(d)}$ and p_r by $\hat{p}_r$ are asymptotically of smaller order of magnitude.

Since eigenfunctions are only unique up to sign, all results concerning a comparison of eigenfunctions implicitly assume that signs have been chosen appropriately.

Proposition 2. *Suppose that the requirements of Proposition* 1, (2.18) *and Assumptions* 7 - 8 *are satisfied. If* $\max(h)^{p+1}h^{-d} \to 0$, $\left\{\max(h)^{p+1}Th^{-d}\right\}^{-1} \to 0$, *and* $NTh_1 \dots h_g h^{2d} \to \infty$ *as* $T, N \to \infty$, *then*

a) $|\tilde{\varphi}_r^{(d)}(t) - \hat{\varphi}_{r,T}^{(d)}(t)| = \mathcal{O}_p\left(\max(h)^{p+1}h^{-d} + (NTh_1 \times \cdots \times h_g h^{2d})^{-1/2}\right)$,

b) $|\tilde{\gamma}_r^{(d)}(t) - \hat{\gamma}_{r,T}^{(d)}(t)| = \mathcal{O}_p\left(\max(h)^{p+1}h^{-d} + (NTh_1 \times \cdots \times h_g h^{2d})^{-1/2}\right)$.

A proof of Proposition 2 is provided in the Supplementary Material S5. Under our assumptions the results of Dauxois, Pousse and Romain (1982) or Hall and Hosseini-Nasab (2006) imply that the difference between $\varphi_r^{(\nu)}$ and $\tilde{\varphi}_r^{(\nu)}$ is of order $\mathcal{O}_p(N^{-1/2})$. The proposition implies that the additional error, generated by estimating eigenfunctions from discrete noisy data, depends on N and T. If T is sufficiently large compared to N, then resulting global optimal bandwidths are given by $h_{r,opt} = \mathcal{O}_p\left((NT)^{-1/(g+2p+2)}\right)$. Even if the optimal bandwidth for both approaches and each basis function is of the same order of magnitude, the values of the actual bandwidths may differ. A simple rule of thumb for the choice of bandwidths in practice is given in Section S7.2.

2.4.4. Estimation of the eigenvalues and loadings

We keep notations $\nu = d$ to refer to the decomposition used in equation (2.5) and $\nu = (0, \dots, 0)^\top$ to (2.6). Empirical estimators of the eigenvalues and loadings are given by $\hat{\lambda}_{r,T}^{(\nu)} = \hat{l}_r^{(\nu)}/N$ and $\hat{\delta}_{ir,T}^{(\nu)} = \sqrt{\hat{l}_r^{(\nu)}}\hat{p}_{ir}^{(\nu)}$. Since $\lambda_r^{(\nu)} - \tilde{\lambda}_r^{(\nu)} = \mathcal{O}_p(N^{-1/2})$, equation (S5) (in the Supplementary Material S5) implies $\lambda_r^{(\nu)} - \hat{\lambda}_{r,T}^{(\nu)} = \mathcal{O}_p(N^{-1/2} + T^{-1/2}N^{-1/2} + T^{-1}) = \mathcal{O}_p(N^{-1/2})$, when $N/T \to 0$. In equation (S6.18) we show that $\tilde{\delta}_{ir}^{(\nu)} - \hat{\delta}_{ir,T}^{(\nu)} = \mathcal{O}_p(T^{-1/2} + N^{1/2}T^{-1})$.

2.5. Consistency results for the derivatives of individual curves

In this section, the focus lies on approximating $X_i^{(d)}$ by a fixed number of components. By (2.8) the mean squared error difference between $X_i^{(d)}$ and $X_{i,L,\varphi}^{(d)}(t) := \sum_{r=1}^{L} \delta_{ir}^{(d)}\varphi_r^{(d)}(t)$ is equal to $\sum_{r=L+1}^{\infty} \lambda_r^{(d)}$. In most applications, the eigenvalues decrease rapidly, and the approximation error becomes small if L is

sufficiently large. Recall that under model (2.11), the eigenvalues and eigenfunctions have to be estimated from discrete, noisy observations. Then there will always exist a dimension L such that the influence of additional functional components is small and cannot be distinguished from the pure random components generated by noise.

Following Kneip and Utikal (2001), the analysis is simplified when relying on the additional semiparametric assumption that there exists a finite dimension L, such that a **factor model** with L components holds $X_i(t) = \sum_{r=1}^{L} \delta_{ir}\gamma_r(t)$ and $0 = \lambda_{L+1} = \lambda_{L+2} = \cdots$. From a theoretical point of view this is a restrictive requirement, but in practice there will exist some L such that the estimated eigenvalue $\hat{\lambda}_{L+1,T}$ does not differ significantly from 0. Model selection criteria based on a factor model may thus provide an appropriate dimension to determine well-fitting approximations whose error is mainly due to random noise. Such criteria will be proposed in Section 2.5.1.

When considering derivatives, a factor model with L components leads to

$$X^{(d)}(t) = \sum_{r=1}^{L_d} \delta_r^{(d)}\varphi_r^{(d)}(t) = \sum_{r=1}^{L} \delta_r\gamma_r^{(d)}(t), \text{ where } L_d \le L. \tag{2.23}$$

In general it cannot be excluded that derivatives $\gamma_r^{(d)}(t)$, $r = 1, \ldots, L$ are collinear, and thus (2.8) leads to $L_d \le L$. As observed by Kneip and Utikal (2001), a factor model with $L \le N$ components implies that, with probability 1, the empirical eigenfunctions constitute a different basis of the same L-dimensional linear space. In our context, we have a.s. $span\{\tilde{\varphi}_1^{(d)}, \ldots, \tilde{\varphi}_{L_d}^{(d)}\} = span\{\varphi_1^{(d)}, \ldots, \varphi_{L_d}^{(d)}\}$, and hence we get $X^{(d)}(t) = \sum_{r=1}^{L_d} \delta_r^{(d)}\varphi_r^{(d)}(t) = \sum_{r=1}^{L_d} \tilde{\delta}_r^{(d)}\tilde{\varphi}_r^{(d)}(t)$.

Based on our methodology, we use the estimators

$$\hat{X}_{i,L_d,\varphi}^{(d)}(t) \stackrel{\text{def}}{=} \sum_{r=1}^{L_d} \hat{\delta}_{ir,T}^{(d)}\hat{\varphi}_{r,T}^{(d)}(t), \quad \hat{X}_{i,L,\gamma}^{(d)}(t) \stackrel{\text{def}}{=} \sum_{r=1}^{L} \hat{\delta}_{ir,T}\hat{\gamma}_{r,T}^{(d)}(t) \tag{2.24}$$

for approximations $X_{i,L_d,\varphi}^{(d)}(t) := \sum_{r=1}^{L_d} \delta_{ir}^{(d)}\varphi_r^{(d)}(t)$, $X_{i,L,\gamma}^{(d)}(t) := \sum_{r=1}^{L} \delta_{ir}\gamma_r^{(d)}(t)$ of the individual derivatives. In a factor model, L_d and L have a clear interpretation. In the general case, we employ the same notation to emphasize that different criteria are used to select the number of components following the two approaches to represent derivatives.

Proposition 3. *Under the requirements of Proposition* 2 *let* $N, T \to \infty$, $NT^{-1} \to 0$.

a) *If additionally a factor model* $X(t) = \sum_{r=1}^{L} \delta_r\varphi_r(t) = \sum_{r=1}^{L} \delta_r\gamma_r(t)$ *holds for a fixed dimension* L, *such that* $X^{(d)}(t) = \sum_{r=1}^{L_d} \delta_r^{(d)}\varphi_r^{(d)}(t) = \sum_{r=1}^{L} \delta_r\gamma_r^{(d)}(t)$,

$L_d \leq L$, then

$$|X_i^{(d)}(t) - \hat{X}_{i,L_d,\varphi}^{(d)}(t)| = \mathcal{O}_p(T^{-1/2} + \max(h)^{p+1}h^{-d} + (NTh_1 \times \cdots \times h_g h^{2d})^{-1/2})$$
$$|X_i^{(d)}(t) - \hat{X}_{i,L,\gamma}^{(d)}(t)| = \mathcal{O}_p(T^{-1/2} + \max(h)^{p+1}h^{-d} + (NTh_1 \times \cdots \times h_g h^{2d})^{-1/2}).$$

b) *In the general case, for* $X_{i,L_d,\varphi}^{(d)}(t) = \sum_{r=1}^{L_d} \delta_{ir}^{(d)} \varphi_r^{(d)}(t)$ *and* $X_{i,L,\gamma}^{(d)}(t) = \sum_{r=1}^{L} \delta_{ir}$ $\gamma_r^{(d)}(t)$

$$|X_{i,L_d,\varphi}^{(d)}(t) - \hat{X}_{i,L_d,\varphi}^{(d)}(t)| = \mathcal{O}_p(N^{-1/2} + \max(h)^{p+1}h^{-d} + (NTh_1 \times \cdots \times h_g h^{2d})^{-1/2})$$
$$|X_{i,L,\gamma}^{(d)}(t) - \hat{X}_{i,L,\gamma}^{(d)}(t)| = \mathcal{O}_p(N^{-1/2} + \max(h)^{p+1}h^{-d} + (NTh_1 \times \cdots \times h_g h^{2d})^{-1/2})$$

Furthermore, $|X_i^{(d)}(t) - X_{i,L_d,\varphi}^{(d)}(t)| \to_P 0$ *and* $|X_i^{(d)}(t) - X_{i,L,\gamma}^{(d)}(t)| \to_P 0$ *as* $L_d, L \to \infty$.

A proof of Proposition 3 is given in the Supplementary Material S6.1. Compared with the convergence rates of the local polynomial estimators for the individual curves, see (2.22), for a factor model, the error of the proposed FPCA-based estimators reduces not only in T but also in N. Equations (2.13) and (2.14) can be interpreted as weighted averages over N curves for a finite number of components. The intuition behind this is that only those components are truncated that are related to the error term and thus a more accurate representation is possible. If N increases at a certain rate, it is possible to get close to parametric rates. Such rates are not possible when smoothing the curves individually.

For the estimation of $\hat{X}_{i,L_d,\varphi}^{(d)}$ stronger assumptions on the smoothness of the curves are necessary to guarantee that the elements of $\hat{M}^{(d)}$ and $\hat{M}^{(0)}$ achieve the same rates of convergence, as illustrated in Remark 1. For raising g and $|d|$ it is required that the true curves are very smooth, which might be unrealistic in many applications. In contrast, the estimation of $M^{(0)}$ still gives parametric rates if less smooth curves are assumed. In addition, if the sample size is small, using a high degree polynomial needed to estimate $M^{(d)}$ might lead to unreliable results. To learn more about these issues, we check the performance of both approaches in a simulation study in Section 3.1 using different sample sizes.

2.5.1. Estimation of the number of components

Model selection criteria can be applied to select a suitable dimension for the FPCA approximations. For orthogonal basis expansion there exists a wide range of criteria that can be adapted to our case. The easiest way to determine the number of components is by choosing the model accuracy by an amount of variance explained by the eigenvalues. In (S5) we show that under the assumptions of

Proposition 2 $\tilde{\lambda}_r^{(\nu)} - \hat{\lambda}_{r,T}^{(\nu)} = \mathcal{O}_p(N^{-1/2}T^{-1/2} + T^{-1})$ and $\lambda_r^{(\nu)} - \tilde{\lambda}_r^{(d)} = \mathcal{O}_p(N^{-1/2})$. The assumptions in Corollary 1 from Bai and Ng (2002) can be adapted to our case and give several criteria for finding L or L_d by generalizing Mallows (1973) C_p criteria for panel data settings. These criteria imply minimizing the sum of squared residuals when k factors are estimated and penalizing the overfitting. One such formulation suggests choosing the number of factors using the criteria

$$PC^{(\nu)}(k^*) = \min_{k\in\mathbb{N}, k\le L_{\max}} \left(\left(\sum_{r=k+1}^{N} \hat{\lambda}_{r,T}^{(\nu)} \right) + k \left(\sum_{r=L^{\max}}^{N} \hat{\lambda}_{r,T}^{(\nu)} \right) \left\{ \frac{\log(C_{NT}^2)}{C_{NT}^2} \right\} \right), \tag{2.25}$$

for the constant $C_{NT} = \min(\sqrt{N}, \sqrt{T})$ and a prespecified $L^{\max} < \min(N, T)$. Bai and Ng (2002) propose information criteria that do not depend on the choice of $L_{\max}$. We consider the modified version

$$IC^{(\nu)}(k^*) = \min_{k\in\mathbb{N}, k\le L} \left(\log \left(\frac{1}{N} \sum_{r=k+1}^{N} \hat{\lambda}_{r,T}^{(\nu)} \right) + k \left\{ \frac{\log(C_{NT}^2)}{C_{NT}^2} \right\} \right). \tag{2.26}$$

For $\nu = (0, \ldots, 0)^\top$, k^* approximates L which is an upper bound for L_d, while for $\nu = d$, k^* estimates L_d.

Another possibility for the choice of number of components is to compute the variance explained by each nonorthogonal basis by

$$\operatorname{Var}\left(\hat{\delta}_{r,T}^{(d)} \hat{\gamma}_{r,T}^{(d)} \right) = \left\langle \hat{\gamma}_{r,T}^{(d)}, \hat{\gamma}_{r,T}^{(d)} \right\rangle \hat{\lambda}_r, \tag{2.27}$$

and sort the variances in decreasing order. Then one could use either (2.25) or (2.26) to select the number of components. A thorough treatment of this criterion is left for future research.

3. Application to SPDs Implied From Option Prices

In this section, we analyze the state price densities (SPDs) implied by the stock index option prices. As state-dependent contingent claims, options contain information about the risk factors driving the underlying asset price process and provide information about expectations and risk patterns on the market. Mathematically, SPDs are densities of some equivalent martingale measures for the discounted asset price and their existence is guaranteed in the absence of arbitrage. In the mathematical-finance terminology they are known as risk neutral densities (RNDs). A restrictive model, with log-normal marginals for the asset price, is the Black-Scholes model. This model results in log-normal SPDs of the underlying asset, which are equivalent to a constant implied volatility surface across

strikes and maturities. This feature is inconsistent with the empirically documented volatility smile or skew and the term structure, see Rubinstein (1985). Therefore, richer specifications for the option surface dynamics have to be used. Many earlier works adopt a static viewpoint; they estimate curves separately at different moments in time, see the reviews by Bahra (1997), Jackwerth (1999) and Bliss and Panigirtzoglou (2002). In order to exploit the information content from all available data, it is reasonable to consider them as collection of curves.

The relation between the SPDs and the European call prices has been demonstrated by Breeden and Litzenberger (1987) and Banz (1978) for a continuum of strike prices spanning the possible range of future realizations of the underlying asset. For a fixed maturity, the SPD is proportional to the second derivative of the European call options with respect to the strike price. In this case, SPDs are one-dimensional functions. A two-dimensional point of view can be adopted if maturities are taken as an additional argument and the SPDs are viewed as a family of curves.

Let $C : \mathbb{R}^2_{\geq 0} \to \mathbb{R}$ denote the price function of a European call option with strike price k and maturity τ such that

$$C(k,\tau) = \exp(-r_\tau \tau) \int_0^\infty (s_\tau - k)^+ q(s_\tau, \tau)\, ds_\tau, \tag{3.1}$$

where r_τ is the annualized risk free interest rate for maturity τ, s_τ the future price of the underlying asset at maturity τ, k the strike price and q the state price density of the stochastic variable s_τ. One can show that

$$q(s_\tau, \tau) = \exp(r_\tau \tau) \left. \frac{\partial^2 C(k,\tau)}{\partial k^2} \right|_{k=s_\tau}. \tag{3.2}$$

Let s_0 be the asset price at the moment of pricing and assume it to be fixed. Then by the no-arbitrage condition, the forward price for maturity τ is

$$F_\tau = \int_0^\infty s_\tau q(s_\tau, \tau) ds_\tau = s_0 \exp(r_\tau \tau). \tag{3.3}$$

Suppose that the call price is homogeneous of degree one in the strike price. Then

$$C(k,\tau) = F_\tau C\left(\frac{k}{F_\tau}, \tau\right). \tag{3.4}$$

If we denote $m = k/F_\tau$ to be the moneyness, then

$$\frac{\partial^2 C(k,\tau)}{\partial k^2} = \frac{1}{F_\tau} \frac{\partial^2 C(m,\tau)}{\partial m^2}. \tag{3.5}$$

One can show that for $d = (2,0)^\top$, $C^{(d)}(m,\tau)|_{m=s_\tau/F_\tau} = q(s_\tau/s_0, \tau) = s_0 q(s_\tau, \tau)$. In practice, it is preferable to work with densities of returns instead of prices when analyzing them jointly because prices are not stationary. Also, notice that

call price curves are not centered. This leads to an additional additive term in (2.4) and (2.6), which refers to the population mean. We illustrate in the Supplementary Material S7 how to handle this in practice.

In our application, X refers to the rescaled and centered random call price surface. Thus we have $g = 2$ with $t = (m, \tau)$ and observation points $t_{ij} = (m_{ij}, \tau_{ij})$. Henceforth, we also assume that the indices $i = 1, \ldots, N$ refer to ordered time-points.

The code used to generate the results reported in this section is published online at www.github.com/QuantLet/FPCA and www.quantlet.de. The data used in the empirical study is available from the authors upon request. Further details regarding the implementation are presented in the Supplementary Material S7.

3.1. Simulation study

We investigate the finite sample behavior of our estimators in a simulation study guided by the data application in Section 3.2. Simulated SPDs are modeled as mixtures of G components, $q(m, \tau) = \sum_{l=1}^{G} w_l q^l(m, \tau)$, where q^l are fixed components and w_l are random weights. For fixed τ we consider log-normal density functions $q^l(\cdot, \tau)$, with mean $\left\{\mu_l - (1/2)\sigma_l^2\right\}\tau$ and variance $\sigma_l^2\tau$, and simulate weights w_{il} satisfying $\sum_{l=1}^{G} w_{il} = 1$, where $i = 1, \ldots, N$ is the index for the day. Then

$$q_i(m, \tau) = \sum_{l=1}^{G} w_{il} \frac{1}{m\sqrt{2\pi\sigma_l^2\tau}} \exp\left(-\frac{1}{2}\left[\frac{\log(m) - \left\{\mu_l - (1/2)\sigma_l^2\right\}\tau}{\sigma_l\sqrt{\tau}}\right]^2\right). \quad (3.6)$$

Following Brigo and Mercurio (2002) the prices of call options for these SPDs are

$$C_i(m, \tau) = \exp(-r_{i\tau}\tau) \sum_{l=1}^{G} w_{il} \left\{\exp(\mu_l\tau)\Phi(y_1) - m\Phi(y_2)\right\}, \quad (3.7)$$

where $y_1 = [\log(m^{-1}) + \{\mu_l + (1/2)\sigma_l^2\}\tau]/(\sigma_l\sqrt{\tau})$, $y_2 = [\log(m^{-1}) + \{\mu_l - (1/2)\sigma_l^2\}\tau]/(\sigma_l\sqrt{\tau})$ and Φ is the standard normal cdf. This representation corresponds to a factor model in which the mixture components can be interpreted as densities associated with a particular state of nature and the weights as probabilities of these states.

We illustrate the finite sample behavior for $G = 3$ with $\mu_1 = 0.4$, $\mu_2 = 0.7$, $\mu_3 = 0.1$, and $\sigma_1 = 0.5$, $\sigma_2 = 0.3$, $\sigma_3 = 0.3$. The weights are simulated from the positive half-standard normal distribution, then standardized to sum up to one. As a result, the covariance operator of the SPD curves has $L = G - 1$

nonzero eingenvalues. In this example, using a mixture of three factors means that only two principal components are necessary to explain the variance in the true curves. In this application $L = L_d = 2$.

Without loss of generality, we set $r_{i\tau} = 0$, for each day $i = i, \ldots, N$. We construct a random grid for each observed curve X_i by simulating points $t_{ik} = (m_{ik}, \tau_{ik})$, $k = 1, \ldots, T$, from a uniform distribution with continuous support $[0.5, 1.8] \times [0.2, 0.7]$. Finally, we record noisy discrete observations of the call functions with additive error term i.i.d. $\varepsilon_{ik} \sim \mathrm{N}(0, 0.1^2)$.

The true SPDs given by (3.6) are used to verify the performance of $\hat{X}^{(d)}_{i,L,\varphi}$, $\hat{X}^{(d)}_{i,L,\gamma}$ and of the individually estimated curves $\hat{X}^{(d)}_{i,LP}$ by local polynomial regression. To derive the optimal bandwidth in each case we use the rule-of-thumb approach presented in Section S7. To illustrate the empirical performance for the estimators of curve derivatives we compute the relative mean integrated square error

$$RMISE\left(X^{(d)}_i, \hat{X}^{(d)}_{i,L,\varphi}\right) = \frac{N^{-1}\sum_{i=1}^{N}\int_{[0,1]^g}\left\{X^{(d)}_i(t) - \hat{X}^{(d)}_{i,L,\varphi}(t)\right\}^2 dt}{N^{-1}\sum_{i=1}^{N}\int_{[0,1]^g}\left\{X^{(d)}_i(t)\right\}^2 dt}.$$

Similarly, we take $RMISE(X^{(d)}, \hat{X}^{(d)}_{i,L,\gamma})$ and $RMISE(X^{(d)}, \hat{X}^{(d)}_{i,LP.})$.

The bandwidth for the individually smoothed curve i is derived by replacing $\hat{p}^{(\nu)}_{ir}$ in (S7.23) by one and zero otherwise. The performance is recorded for sample sizes N of 10 and 25 with T observations per day of size 50 and 250. This procedure is repeated for 500 samples to get reliable results. The mean, variance, median and the interquartile range based on the $RMISE$ of all replications are reported in Table 1.

Both FPCA-based approaches give better estimates for the derivatives of call functions than the simple local polynomial regression applied to the individual curves, as shown by the mean and the median of their corresponding $RMISE$. However, the estimator $\hat{X}^{(d)}_{L,i,\gamma}$ performs decisively better for small T than the other two, in terms of the mean and variance of $RMISE$. With small T, the variability of $RMISE$ for $\hat{X}^{(d)}_{L,i,\varphi}$ and individually smoothed curves is much larger than for $\hat{X}^{(d)}_{L,i,\gamma}$, while the median of $RMISE$ for $\hat{X}^{(d)}_{L,i,\gamma}$ and $\hat{X}^{(d)}_{L,i,\varphi}$ are comparable. This means that individual local polynomial smoothers and $\hat{X}^{(d)}_{L,i,\varphi}$ might behave worse than $\hat{X}^{(d)}_{L,i,\gamma}$ in some instances while $\hat{\gamma}^{(d)}_{r,T}$-based expansion provides more to stable estimates. To get the same effect using $\hat{X}^{(d)}_{L,i,\varphi}$ a higher T is needed. A possible explanation for this behavior is given by Proposition 1. The rates of

Table 1. Simulation results for $g = 2$. Based on the mean and the median of $RMISE$, $\hat{X}^{(d)}_{i,L,\varphi}$ and $\hat{X}^{(d)}_{i,L,\gamma}$ yield superior results compared to $\hat{X}^{(d)}_{i,LP}$; $\hat{X}^{(d)}_{i,L,\gamma}$ outperforms $\hat{X}^{(d)}_{i,L,\varphi}$ in all cases. Results for $\hat{X}^{(d)}_{i,L,\varphi}$ and $\hat{X}^{(d)}_{i,L,\gamma}$ improve with raising N and T. These results support our asymptotic results given by Proposition 1 and 3. All results are multiplied by 10^2.

$RMISE$	T	50				250			
N	$\hat{X}^{(d)}_{\bullet}$	Mean	Var	Med	IQR	Mean	Var	Med	IQR
10	$\hat{X}^{(d)}_{i,L,\varphi}$	32.28	18.83	18.95	23.33	8.24	0.35	6.85	5.36
	$\hat{X}^{(d)}_{i,L,\gamma}$	29.85	22.33	16.80	22.22	7.48	0.35	5.80	4.76
	$\hat{X}^{(d)}_{i,LP}$	74.71	75.45	49.59	63.29	14.47	1.06	11.50	10.35
25	$\hat{X}^{(d)}_{i,L,\varphi}$	17.35	3.89	12.18	11.48	6.10	0.08	5.28	4.19
	$\hat{X}^{(d)}_{i,L,\gamma}$	15.32	3.63	10.12	9.78	4.67	0.14	4.05	2.66
	$\hat{X}^{(d)}_{i,LP}$	80.83	82.15	54.18	67.78	14.57	1.05	11.77	10.39

Table 2. Simulation results for $g = 2$. The mean and the median of $RISE$ for $\hat{\gamma}^{(d)}_{r,T}$ and $\hat{\varphi}^{(d)}_{r,T}$ improve with raising N and T.

$RISE$		Mean				Median			
T	N	$\hat{\gamma}^{(d)}_{1,T}$	$\hat{\gamma}^{(d)}_{2,T}$	$\hat{\varphi}^{(d)}_{1,T}$	$\hat{\varphi}^{(d)}_{1,T}$	$\hat{\gamma}^{(d)}_{1,T}$	$\hat{\gamma}^{(d)}_{2,T}$	$\hat{\gamma}^{(d)}_{1,T}$	$\hat{\varphi}^{(d)}_{2,T}$
50	10	19.97	6.99	0.91	1.02	2.69	0.05	0.96	1.02
250	10	51.7	1.35	0.64	1.03	2.54	0.19	0.6	1.06
50	25	9.02	0.24	0.84	1.01	0.96	0.05	0.87	1.03
250	25	3.22	1.54	0.54	1.06	0.49	0.19	0.45	1.09

convergence for the estimators of the dual matrix entries rely on T. Thus in finite samples, when T is small, the estimated loadings might be biased.

We evaluate the empirical performance of $\hat{\gamma}^{(d)}_{r,T}$ and $\hat{\varphi}^{(d)}_{r,T}$ based on the relative integrated square error

$$RISE\left(\tilde{\gamma}^{(d)}_{r}, \hat{\gamma}^{(d)}_{r,T}\right) = \frac{\int_{[0,1]^g} \left\{\tilde{\gamma}^{(d)}_{r}(t) - \hat{\gamma}^{(d)}_{r,T}(t)\right\}^2 dt}{\int_{[0,1]^g} \left\{\hat{\gamma}^{(d)}_{r,T}(t)\right\}^2 dt}.$$

The results are summarized in Table 2. Here the mean and median of $RISE$ do not have the same order of magnitude and the results for the corresponding r are not comparable because they refer to different functions $\hat{\varphi}^{(d)}_{r,T}$ and $\hat{\gamma}^{(d)}_{r,T}$.

To compare the performance of our methodology with an existing FPCA-based method, we include in the Supplementary Material S8 a simulation study for the unidimensional case, in which we report the results from our second

method in (2.6) and (2.7) and the results from Liu and Müller (2009). Both these approaches aim to estimate $\tilde{\gamma}_r^{(d)}$.

3.2. Data example

3.2.1. Data description

We use settlement European call option prices written on the underlying DAX 30 stock index. These prices are computed by EUREX at the end of each trading day based on the intraday transaction prices. The data range runs between January 2, 2002 and December 3, 2011, and includes 2,557 days. The expiration dates of the options are set on the third Friday of the month. Therefore, on a particular day, option prices with only a few maturities are available, as illustrated in the upper panels of Figure 1. Methods that analyze curves jointly are generally better tailored to this type of data, because they provide better estimates at grid points with only a few observations of the individual curves. We include call options with maturity between one day and one year. The sample contains prices of options with an average of six maturities and sixty-five strikes per day.

By assuming 'sticky' coordinates for the daily observations, in accordance with (3.4), we divide the strike and the call prices within one day by the stock index forward price to ensure that the observation points are in the same range. Afterward, we apply the estimation methodology of Section 2 to the rescaled call prices, which are functions of moneyness and maturity. The proxy for the risk-free interest rates are the EURIBOR rates, which are listed daily for several maturities. We apply a linear interpolation to calculate the rate values for desired maturities.

3.2.2. Estimation results

We report the results for the empirical loadings based on the spectral decomposition of empirical dual covariance matrix $\hat{M}^{(0)}$ of the rescaled option price functions, and for the empirical second partial derivative of the eigenfunctions $\hat{\gamma}_{r,T}^{(d)}$, $d = (2, 0)^\top$. Recall that this method does not estimate orthogonal basis. In the Supplementary Material S9.1 we explain the selection of three interpretable components that we now analyze. They correspond to $\gamma_{1,T}^{(d)}$, $\gamma_{3,T}^{(d)}$ and $\gamma_{7,T}^{(d)}$. Their estimates together with the empirical loadings are displayed in Figure 2. They describe three types of variation present in the dynamics of the SPDs. There is a long left tail, specific to the negatively skewed densities, and a peak located at a value of moneyness slightly above one. For positive loadings, this component

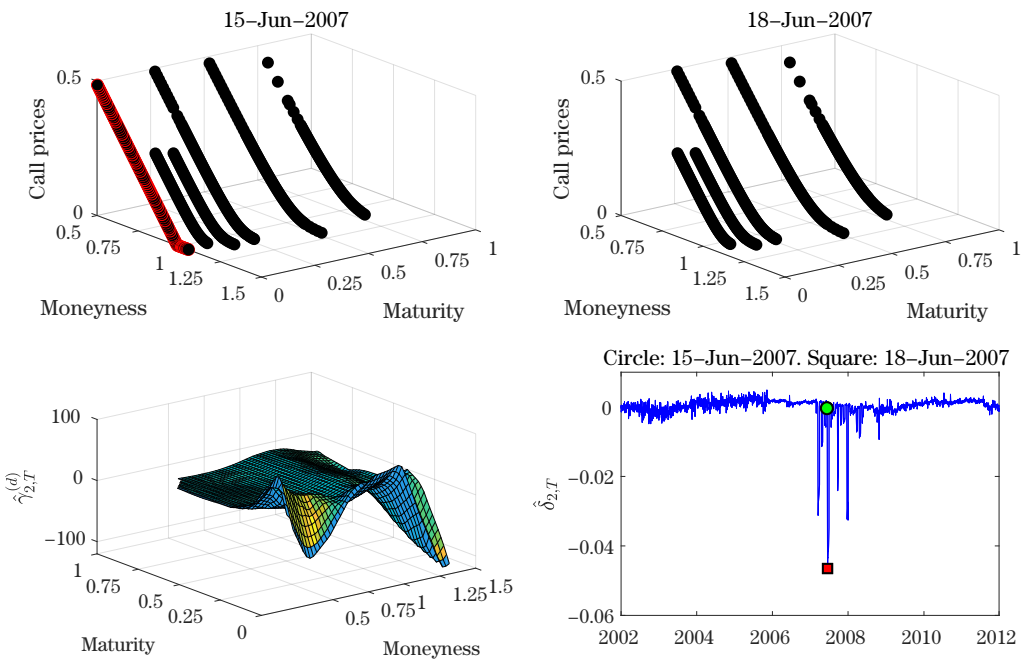

Figure 1. Option prices around expiration date (upper panels). Estimated component $\hat{\gamma}^{(d)}_{2,T}$ (lower panel left). The effect of expiration date on the level of empirical loadings $\hat{\delta}_{2,T}$ (lower panel right).

increases the mass of SPD around the mode of the empirical mean and decreases the mass in the tails. We find that this component is related to the time-varying volatility of the index returns. The next component, $\hat{\gamma}^{(d)}_{3,T}$ has a 'valley-hill' pattern, which shifts mass around the central region of the density. A positive shock in the direction of this component increases the negative skewness, and a large negative shock can reverse the sign of the SPD skewness. This component is interpreted as a skewness factor. Further justifications for the interpretation of these two components are provided in the Supplementary Material S9.2. The last component, $\hat{\gamma}^{(d)}_{7,T}$ takes negative values in the left tail and displays a prominent positive-valued peak at the right of the mode of the empirical SPD mean. This component can be interpreted as a tail factor, and we show in the next section that its loadings are related to the volatility of volatility index.

We conjecture that the other components selected by $PC^{(0)}$ and $IC^{(0)}$ criteria are related to reactions of option prices along the maturity direction. In addition, their loadings contain regular spikes around the expiration date of options between mid-February 2007 and mid-September 2008. We illustrate this in

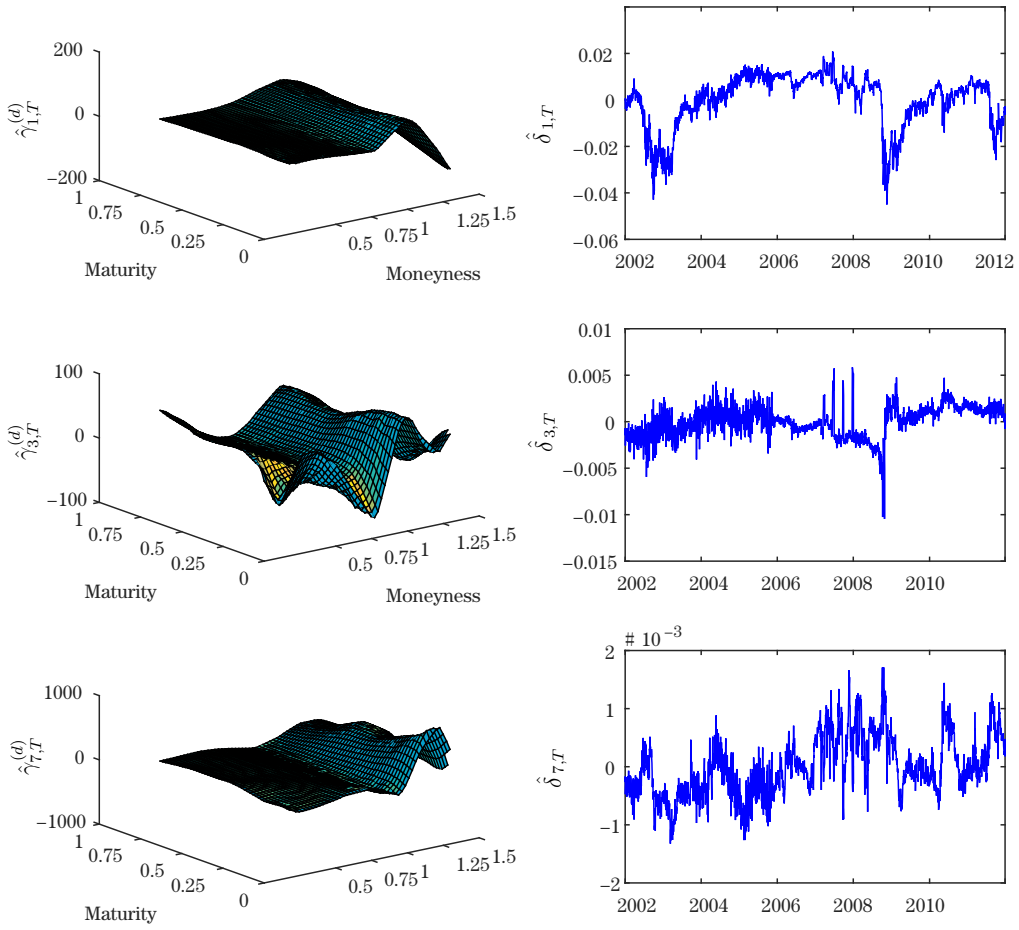

Figure 2. Estimated components $\hat{\gamma}^{(d)}_{1,T}$, $\hat{\gamma}^{(d)}_{3,T}$ and $\hat{\gamma}^{(d)}_{7,T}$ and their loadings.

the lower panels of Figure 1 for the second component.

3.2.3. Dynamic analysis of the loadings

In this section, we investigate the dynamics of the loadings in the approximating model. In the Supplementary Material S9.3 we discuss the preliminary analysis of these loadings in a time series context. This suggests that we consider the following time-varying autoregressive model for the loadings

$$\hat{\delta}_{ir,T} = b_r \hat{\delta}_{i-1r,T} + e_{ir}, \quad \text{Var}(e_{ir}) = \sigma^2_{er}, \quad r = 1, 2, 3, \tag{3.8}$$

where b_r is the autoregressive coefficient. We reestimate (3.8) daily based on a rolling window of 250 past observation using OLS. This adaptive estimation procedure helps detect the possible sources of non-stationarity in the estimated

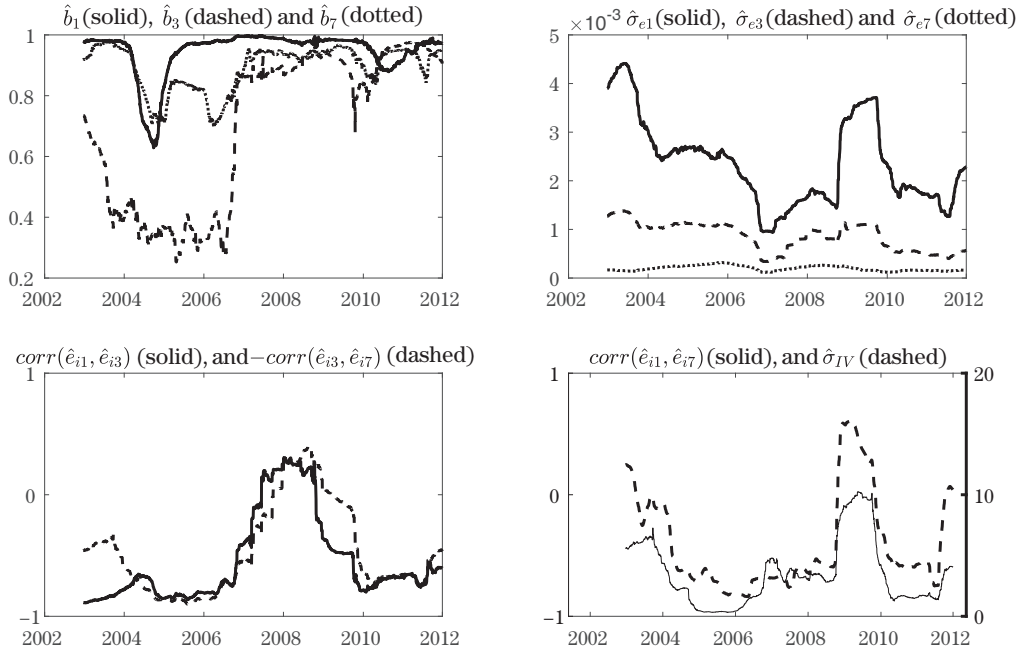

Figure 3. Time-varying autoregressive coefficients $\hat{b}_r$, standard deviation $\hat{\sigma}_{er}$ and pairwise correlations $corr(\hat{e}_{ir}, \hat{e}_{ir'})$ of residuals in univariate AR(1) regressions of loadings, and standard deviation of the VDAX volatility index $\hat{\sigma}_{IV}$ estimated daily with a rolling window of 250 observations.

loadings, by allowing the autoregressive coefficient and the error variance to vary over time.

The upper left panel of Figure 3 displays the estimated autoregressive coefficients. $\hat{\delta}_{1,T}$ is very persistent ($\hat{b}_1$ is close to one), except for 2004. Interestingly, $\hat{b}_3$ is relatively small between 2003–2006 and increases significantly thereafter, suggesting a possible regime shift. $\hat{b}_7$ is relatively high and its variation seems sensitive to the changes in the other two parameters.

We also compute the time-varying cross-equation correlations between the error terms. The two lower panels of Figure 3 illustrate the results. The error correlation of the skewness with the volatility and with the tail factor move closely together, suggesting a strong relationship between the volatility and the tail factors. We focus on $corr(\hat{e}_{i1}, \hat{e}_{i3})$, which describes the dynamic relationship between the changes in SDP volatility and skewness. Most of the time, the plotted correlation is negative, meaning that positive changes in the SPD variance are associated on average with increases in the negative skewness. The negative correlation between an asset return and its changes of volatility is generally known

as the leverage effect. The correlation reverses sign and becomes positive between 2007-2009. This implies that when volatility increases, there is a change in the concentration mass in the left side of the density, in the area of medium-ranged negative returns. We identify this behavior with the implied volatility skew puzzle, as documented by Constantinides and Lian (2015). The authors rationalize this behavior through the reduction in put option supply from credit-constrained market makers together with an increase in the demand for OTM puts required for hedging purposed, see also net buying pressure in Bollen and Whaley (2004), Gârleanu, Pedersen and Poteshman (2009).

Typically, the error correlation $corr(\hat{e}_{i1}, \hat{e}_{i7})$ is negative. Its magnitude decreases and reaches values close to zero in 2009. In the lower right panel of Figure 3, we also plot the 250-observation standard deviation $\hat{\sigma}_{IV}$ of the VDAX implied volatility index. The two time-series are strongly correlated (the correlation coefficient is 90.78%). This suggests that the tail component can be interpreted as the volatility of volatility risk factor. Similar interpretations were proposed in Du and Kapadia (2012), Huang and Shaliastovich (2014) and Park (2015), who use different measures of the volatility-of-volatility implied by VIX (the implied volatility index of the S&P 500) as a tail risk indicator. The tail factor takes highest positive values during the financial crisis, consistent with fat tail and high risk hypothesis.

To verify the stability of the results reported, we repeat the regression analysis by including a constant in (3.8). The root mean square error does not improve significantly. We also estimate the model by including the lagged values of the other two loadings as additional explanatory variables. Some of the estimated autoregressive coefficients take values above one. Independently of these modeling choices, the estimated cross-error term interactions are very similar to those shown in Figure 3. These suggest that changes in the correlation sign for the levels of the loadings are driven mainly by the error term correlation structure and not by the changes in the other lagged variables.

Several stylized facts emerge from the dynamic analysis of the loadings that summarize the variation of SPDs. When volatility is small, the innovations to the volatility, skewness and volatility of volatility loading equations are very strongly correlated. When volatility increases, the correlation structure changes. In particular, the leverage parameter changes sign during the financial crises. By including volatility of volatility as an additional factor, see also Huang and Shaliastovich (2014), our study distinguishes between the volatility induced skewness through the leverage effect and by the volatility of volatility induced skew-

ness, see also Feunou and Tédongap (2012). These findings may have important consequences for the formulation of stochastic volatility models for option pricing.

Supplementary Materials

The online Supplementary Material includes a summary of technical assumptions, proofs of Proposition 1, 2 and 3, the convergence of Monte-Carlo integrals that approximate the elements of the dual matrix, practical aspects for the implementation of proposed methods, comparison to an existing FPCA-based method for estimating derivatives, supporting results for the analysis of DAX 30 SPDs and additional references.

Acknowledgment

Financial support from the German Research Foundation for the joint project "Functional Principal Components for Derivatives and Higher Dimensions", between Humboldt University of Berlin and University of Bonn, is gratefully acknowledged. We would like to thank as well the Collaborative Research Center 649 "Economic Risk" for providing the data and the International Research Training Group (IRTG) 1792 "High-Dimensional Non-Stationary Time Series Analysis" at Humboldt University of Berlin for additional funding.

References

Bahra, B. (1997). Implied risk-neutral probability density functions from option prices: theory and application. Technical report, Bank of England Working Paper No 66.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70**, 191–221.

Banz, R. W. (1978). Prices for state-contingent claims: Some estimates and applications. *The Journal of Business* **51**, 653–672.

Benko, M., Härdle, W. and Kneip, A. (2009). Common functional principal components. *The Annals of Statistics* **37**, 1–34.

Besse, P. and Ramsay, J. (1986). Principal components analysis of sampled functions. *Psychometrika* **51**, 285–311.

Bliss, R. R. and Panigirtzoglou, N. (2002). Testing the stability of implied probability density functions. *Journal of Banking & Finance* **26**, 381–422.

Bollen, N. P. B. and Whaley, R. E. (2004). Does net buying pressure affect the shape of implied volatility functions? *Journal of Finance* **59**, 711–753.

Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer.

Breeden, D. T. and Litzenberger, R. H. (1987). Prices of state-contingent claims implicit in option prices. *Journal of Business* **51**, 621–651.

Brigo, D. and Mercurio, F. (2002). Lognormal-mixture dynamics and calibration to market volatility smiles. *International Journal of Theoretical and Applied Finance* **5**, 427–446.

Cai, T. T. and Yuan, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *The Annals of Statistics* **39**, 2330–2355.

Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters* **45**, 11–22.

Cardot, H., Mas, A. and Sarda, P. (2007). CLT in functional linear regression models. *Probability Theory and Related Fields* **138**, 325–361.

Constantinides, G. M. and Lian, L. (2015). The supply and demand of S&P 500 put options. *SSRN Electronic Journal*.

Cont, R. and da Fonseca, J. (2002). Dynamics of implied volatility surfaces. *Quantitative Finance* **2**, 45—60.

Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis* **12**, 136–154.

Di, C., Crainiceanu, C., Caffo, B. and Punjabi, N. (2009). Multilevel functional principal component analysis. *The Annals of Applied Statistics* **3**, 458–488.

Du, J. and Kapadia, N. (2012). Tail and volatility indices from option prices. Working paper.

Fan, J., Gasser, T., Gijbels, I., Brockmann, M. and Engel, J. (1997). Local polynomial regression: Optimal kernels and asymptotic minimax efficiency. *Annals of the Institute of Statistical Mathematics* **49**, 79–99.

Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *The Annals of Statistics* **20**, 2008–2036.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis, Theory and Practice*. Springer.

Feunou, B. and Tédongap, R. (2012). A stochastic volatility model with conditional skewness. *Journal of Business & Economic Statistics* **30**, 576–591.

Gârleanu, N., Pedersen, L. H. and Poteshman, A. M. (2009). Demand-based option pricing. *Review of Financial Studies* **22**, 4259–4299.

Grith, M., Härdle, W. K. and Park, J. (2013). Shape invariant modeling of pricing kernels and risk aversion. *Journal of Financial Econometrics* **11**, 370–399.

Grith, M., Härdle, W. K. and Schienle, M. (2012). Nonparametric estimation of risk-neutral densities, *In Handbook of Computational Finance*, 277–305. Springer Verlag.

Gu, J., Li, Q. and Yang, J.-C. (2015). Multivariate local polynomial kernel estimators: Leading bias and asymptotic distribution. *Econometric Reviews* **34**, 978–1009.

Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of The Royal Statistical Society Series B Statistical Methodology* **68**, 109–126.

Hall, P. and Marron, J. S. (1990). On variance estimation in nonparametric regression. *Biometrika* **77**, 415–419.

Hall, P., Müller, H.-G. and Yao, F. (2009). Estimation of functional derivatives. *The Annals of Statistics* **37**, 3307–3329.

Härdle, K. W. and Lopez-Cabrera, B. (2012). The implied market price of weather risk. *Applied*

Mathematical Finance **19**, 59–95.

Huang, D. and Shaliastovich, I. (2014). Volatility-of-volatility risk. University of Pennsylvania Preprint.

Jackwerth, J. C. (1999). Option-implied risk-neutral distributions and implied binomial trees: a literature review. *Journal of Derivatives* **2**, 66–82.

Kneip, A. and Utikal, K. J. (2001). Inference for density families using functional principal component analysis. *Journal of the American Statistical Association* **96**, 519–542.

Liu, B. and Müller, H.-G. (2009). Estimating derivatives for samples of sparsely observed functions, with application to online auction dynamics. *Journal of the American Statistical Association* **104**, 704–717.

Majer, P., Mohr, P., Heekeren, H. and Härdle, K. W. (2015). Portfolio decisions and brain reactions via the CEAD method. *Psychometrika* **81**, 881–903.

Mallows, C. (1973). Some comments on c_p. *Technometrics* **15**, 661–675.

Mas, A. (2002). Weak convergence for the covariance operators of a hilbertian linear process. *Stochastic Processes and Their Applications* **99**, 117–135.

Mas, A. (2008). Local functional principal component analysis. *Complex Analysis and Operator Theory* **2**, 135–167.

Masry, E. (1996). Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis* **17**, 571–599.

Munk, A., Bissantz, N., Wagner, T. and Freitag, G. (2005). On difference-based variance estimation in nonparametric regression when the covariate is high dimensional. *Journal of the Royal Statistical Society Series B Statistical Methodology* **67**, 19–41.

Park, Y. H. (2015). Volatility-of-volatility and tail risk hedging returns. *Journal of Financial Markets* **26**, 38–63.

Rice, J. (1984). Bandwidth choice for nonparametric regression. *The Annals of Statistics* **12**, 1215–1230.

Rubinstein, M. (1985). Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active CBOE option classes from August 23, 1976 through August 31, 1978. *The Journal of Finance* **40**, 455–480.

Staicu, A. M., Crainiceanu, C. M. and Carroll, R. J. (2010). Fast methods for spatially correlated multilevel functional data. *Biostatistics* **11**, 177–194.

van Bömmel, A., Song, S., Majer, P., Mohr, P. N. C., Heekeren, H. R. and Härdle, W. K. (2014). Risk patterns and correlated brain activities. Multidimensional statistical analysis of fMRI data in economic decision making study. *Psychometrika* **79**, 489–514.

von Neumann, J., Kent, R. H., Bellinson, H. R. and Hart, B. I. (1941). The mean square successive difference. *The Annals of Mathematical Statistics* **12**, 153–162.

Zhang, X. and Wang, J.-L. (2016). From sparse to dense functional data and beyond. *The Annals of Statistics* **44**, 2281–2321.

Zipunnikov, V., Caffo, B. C., Yousem, D. M., Davatzikos, C., Schwartz, B. S. and Crainiceanu, C. M. (2011). Functional principal component model for high-dimensional brain imaging. *NeuroImage* **58**, 772–784.

Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, Burg Oudlaan 50, 3062 PA Rotterdam, The Netherlands.

E-mail: grith@ese.eur.nl

Institute for Financial Economics and Statistics, Department of Economics, University of Bonn, Adenauerallee 24-26, 53113 Bonn, Germany.

E-mail: heikowagner@uni-bonn.de

Ladislaus von Bortkiewicz Chair of Statistics and C.A.S.E. - Center for Applied Statistics and Economics, School of Business and Economics, Humboldt University of Berlin, Spandauer Stra 1, 10178 Berlin, Germany. Sim Kee Boon Institute for Financial Economics, Singapore Management University, 81 Victoria Street, Singapore 188065.

E-mail: haerdle@hu-berlin.de

Institute for Financial Economics and Statistics, Department of Economics, University of Bonn, Adenauerallee 24-26, 53113 Bonn, Germany.

E-mail: akneip@uni-bonn.de

(Received September 2016; accepted March 2018)