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*

Supplementary Material

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.

S1. Assumptions summary

Assumption 1. $X_1, \ldots, X_N$ are i.i.d. centered random functions which are a.s. m times continuously differentiable. All corresponding partial derivatives possess finite fourth moments for all $t \in [0,1]^g$.

Assumption 2. Following model (2.11) functions are observed at a random grid $t_{i1}, \ldots, t_{iT_i}$, $t_{ik} \in [0,1]^g$ having a common bounded and continuously differentiable density f with support $\operatorname{supp}(f) = [0,1]^g$ and the integrand $u \in \operatorname{supp}(f)$ and $\inf_u f(u) > 0$. The random variables t_{ij} and X_i are independent.

Assumption 3. $\mathsf{E}(\varepsilon_{ik}) = 0$, $\operatorname{Var}(\varepsilon_{ik}) = \sigma^2_{i\varepsilon} > 0$, $\mathsf{E}\left[\varepsilon^4_{ik}\right] < D$ for some $D < \infty$ and all $i = 1, \ldots, N$, and ε_{ik} are independent of X_i and t_{ij}, $\forall i, k, j$.

Assumption 4. Let $K_B(u) = \frac{1}{b_1 \times \cdots \times b_g} K(u \circ b)$. K is a product kernel based on symmetric univariate kernels. B is a diagonal matrix with $b = (b_1, \ldots, b_g)^\top$ at the diagonal. The kernel K is bounded and has compact support on $[-1, 1]^g$ such that for $u \in \mathbb{R}^g$ $\int uu^T K(u) du = \mu(K) I$ where $\mu(K) \neq 0$ is a scalar and I is the $g \times g$ identity matrix. Conditions 2 and 3 from Masry (1996) are fulfilled.

Assumption 5. For ρ, p, $d_j \in \mathbb{N}$, $j = 1, \ldots, g$, $\rho - \sum_{l=1}^{g} d_l$ and $p - \sum_{l=1}^{g} d_l$ are odd.

Assumption 6. Estimators of the error variances satisfy $|\sigma_{i\varepsilon}^2 - \hat{\sigma}_{i\varepsilon}^2| = \mathcal{O}_P(T^{-1/2})$.

Assumption 7.

$$\sup_{r \in \mathbb{N}} \sup_{t \in [0,1]^g} |\varphi_r^{(d)}(t)| < \infty \ , \ \sup_{r \in \mathbb{N}} \sup_{t \in [0,1]^g} |\gamma_r^{(d)}(t)| < \infty$$

$$\sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \mathsf{E}\left[\left(\delta_{ri}^{(\nu)}\right)^2 \left(\delta_{si}^{(\nu)}\right)^2\right] < \infty \ , \ \sum_{q=1}^{\infty} \sum_{s=1}^{\infty} \mathsf{E}\left[\left(\delta_{ri}^{(\nu)}\right)^2 \delta_{si}^{(\nu)} \delta_{qi}^{(\nu)}\right] < \infty,$$

for $\nu = (0, \ldots, 0)^\top$ as well as $\nu = d$. Recall that $\mathsf{E}\left[(\delta_{ri}^{(\nu)})^2\right] = \lambda_r^{(\nu)}$, and hence $\delta_{ri}^{(\nu)} = 0$ a.s. iff $\lambda_r = 0$.

Assumption 8. Let $\nu = (0, \ldots, 0)^\top$ or $\nu = d$. For any $r \in \mathbb{N}^*$ with $\lambda_r^{(\nu)} > 0$ there exists some $0 < C_{1,r} < \infty$ such that

$$\min_{s \in \mathbb{N}^*; s \neq r} |\lambda_r^{(\nu)} - \lambda_s^{(\nu)}| \geq C_{1,r}.$$

S2. Proof of Proposition 1

The proposition is an immediate consequence of the following lemma.

Lemma 1. *Let the assumptions of Proposition 1 hold, and in addition* $T \to \infty$, $\max(b)^{\rho+1} b^{-\nu} \to 0$, $\frac{\log(T)}{T b_1 \times \cdots \times b_g} \to 0$, $T b_1 \times \cdots \times b_g b^{4\nu} \to \infty$. *Then*

$$\mathsf{E}\left(\hat{M}_{ij}^{(\nu)} \middle| \hat{X}_{i,b}^{(\nu)}, \hat{X}_{j,b}^{(\nu)}\right) - \int_{[0,1]^g} X_i^{(\nu)}(t) X_j^{(\nu)}(t) dt =$$
$$= \mathcal{O}_p\left(\max(b)^{\rho+1} b^{-\nu} + \frac{1}{T^{3/2}(b^{2\nu} b_1 \times \cdots \times b_g)}\right) \qquad \text{(S2.1)}$$
$$\mathrm{Var}\left(\hat{M}_{ij}^{(\nu)} \middle| \hat{X}_{i,b}^{(\nu)}, \hat{X}_{j,b}^{(\nu)}\right) = \mathcal{O}_p\left(\frac{1}{T^2 b_1 \times \cdots \times b_g b^{4\nu}} + \frac{1}{T}\right).$$

S3. Proof of Lemma 1

For the proof of the lemma we will concentrate on the diagonal entries $\hat{M}_{ii}^{(\nu)}$ which are slightly more difficult due to the necessity of a diagonal correction.

S3.1 Univariate case (g = 1)

Following Ruppert and Wand (1994), we show that equation (2.16) can be stated up to a vanishing constant using equivalent kernels. Equivalent kernels can be understood as an asymptotic version of $W_\nu^{T_i}$. Let e_l be a vector of length $\rho+1$ with 1 at the $l+1$ position and zero elsewhere. Then

$$W_\nu^{T_i}\left(\frac{t_{ij}-u}{b}\right) = \frac{e_\nu^\top S_{T_i}(u)^{-1}}{b^{\nu+1}T_i}\left(1, \left(\frac{t_{ij}-u}{b}\right)^1, \dots, \left(\frac{t_{ij}-u}{b}\right)^\rho\right)^\top K\left(\frac{t_{ij}-u}{b}\right),$$

where $S_{T_i}(u)$ is a $(\rho+1)\times(\rho+1)$ symmetric matrix

$$S_{T_i}(u) = \begin{pmatrix} S_{T_i,0}(u) & S_{T_i,1}(u) & \dots & S_{T_i,\rho}(u) \\ S_{T_i,1}(u) & S_{T_i,2}(u) & \dots & S_{T_i,\rho+1} \\ \vdots & \vdots & \ddots & \vdots \\ S_{T_i,\rho}(u) & S_{T_i,\rho+1}(u) & \dots & S_{T_i,2\rho}(u) \end{pmatrix}, \tag{S3.2}$$

with entries $S_{T_i,k}(u) = (T_i b)^{-1}\sum_{j=1}^{T_i} K\left(\frac{t_{ij}-u}{b}\right)\left(\frac{t_{ij}-u}{b}\right)^k$. Then it holds that

$$\begin{aligned}\mathsf{E}(S_{T_i,k}(t_{ij})) =& (T_i b)^{-1}\int_0^1 \sum_{l=1}^{T_i} K\left(\frac{x-u}{b}\right)\left(\frac{x-u}{b}\right)^k f(x)dx \\ =& b^{-1}\int_u^{1+u} K\left(\frac{x}{b}\right)\left(\frac{x}{b}\right)^k f(x)dx = \int_{ub^{-1}}^{(1+u)b^{-1}} K(t)\, t^k f(tb)dt.\end{aligned}$$

Since $K(t)$ has compact support and is bounded, for a point at the left boundary with $c \geq 0$ it holds that u is of the form $u = cb$ and at the right boundary $u = 1-cb$. We define $S_{k,-c} = \int_{-c}^{\infty} t^k K(t)dt$ and $S_{k,c} = \int_{-\infty}^{c} t^k K(t)dt$, respectively and for interior points $S_k = \int_{-\infty}^{\infty} t^k K(t)dt$. Further we construct the $(\rho+1)\times(\rho+1)$ matrix corresponding to (S3.2)

$$S(u) = \begin{cases} (S_{j+l,-c})_{0\leq j,l\leq\rho} & , u \text{ is a left boundary point} \\ (S_{j+l})_{0\leq j,l\leq\rho} & , u \text{ is an interior point} \\ (S_{j+l,c})_{0\leq j,l\leq\rho} & , u \text{ is a right boundary point.} \end{cases} \tag{S3.3}$$

The equivalent kernel is then defined as

$$K^*_{\nu,\rho}\left(\frac{t_{ij}-u}{b}\right) = e_\nu^\top S(u)^{-1}\left(1, \left(\frac{t_{ij}-u}{b}\right)^1, \ldots, \left(\frac{t_{ij}-u}{b}\right)^\rho\right)^\top K\left(\frac{t_{ij}-u}{b}\right).$$

Then the estimator in equation (2.16) can be rewritten as

$$\hat{X}^{(\nu)}_{i,b}(t) = \nu!\hat{\beta}_{i,\nu}(t) = \frac{\nu!}{T_i b^{\nu+1} f(t)} \sum_{l=1}^{T_i} K^*_{\nu,\rho}\left(\frac{t_{il}-t}{b}\right) Y_i(t_{il})\{1+\mathcal{O}_P(1)\}.$$

Following Masry (1996) we can further state that for a bandwidth b fulfilling $\frac{\log(T_i)}{T_i b} \to 0$ we have uniformly in $u \in [0,1]$ that $S_{T_i}(u)^{-1} \to \frac{S(u)^{-1}}{f(u)}$ almost surely as $T_i \to \infty$.

By construction, the equivalent kernel fulfills

$$\int u^k K^*_{\nu,\rho}(u)\, du = \delta_{\nu,k} \quad 0 \leq \nu, k \leq \rho, \tag{S3.4}$$

for $\delta_{\nu,k}$ the Kronecker delta. As mentioned by Fan et al. (1997), the design of the kernel automatically adapts to the boundary which gives the same order of convergence for the interior and boundary points, see also Ruppert and Wand (1994). For $u \in [0,1]$

$$\int \nu!^2 \sum_{j=1}^{T_i}\sum_{l=1}^{T_i} W_\nu^{T_i}\left(\frac{t_{ij}-u}{b}\right) W_\nu^{T_i}\left(\frac{t_{il}-u}{b}\right) Y_i(t_{il})Y_i(t_{ij})du = \int \frac{\nu!^2}{T_i^2 b^{2\nu+2}}$$
$$\frac{1}{f(u)^2}\sum_{l=1}^{T_i}\sum_{j=1}^{T_i} K^*_{\nu,\rho}\left(\frac{t_{ij}-u}{b}\right) K^*_{\nu,\rho}\left(\frac{t_{il}-u}{b}\right) Y_i(t_{il})Y_i(t_{ij})\{1+\mathcal{O}_P(1)\}du.$$

For the expectation we get

$$\begin{aligned}
&\mathsf{E}\left(\hat{M}_{ii}^{(\nu)} \middle| \hat{X}_{i,b}^{(\nu)}\right)\\
&= \int \nu!^2 \sum_{j=1}^{T_i}\sum_{l=1}^{T_i} W_\nu^{T_i}\left(\frac{t_{ij}-u}{b}\right) W_\nu^{T_i}\left(\frac{t_{il}-u}{b}\right) X_i(t_{il})X_i(t_{ij})du\\
&\quad + \nu!^2\left(\sigma_{i\varepsilon}^2 - \hat{\sigma}_{i\varepsilon}^2\right)\int \sum_{j=1}^{T_i} W_\nu^{T_i}\left(\frac{t_{ij}-u}{b}\right)^2 du\\
&= \Bigg\{\nu!^2 \int\int\int \frac{f(x)f(y)}{b^{2(\nu+1)}f(z)^2} K_{\nu,\rho}^*\left(\frac{x-z}{b}\right) K_{\nu,\rho}^*\left(\frac{y-z}{b}\right) X_i(x)X_i(y)dxdydz\\
&\quad + \mathcal{O}_P\left(\frac{1}{T_i^{3/2}b^{2\nu+1}}\right)\Bigg\}\{1+o_P(1)\}\\
&= \Bigg\{\int X_i^{(\nu)}(z)X_i^{(\nu)}(z)dz\\
&\quad + 2\frac{\nu!}{(\rho+1)!}\int \frac{b^{\rho+1}}{b^\nu}\left(\int u^{\rho+1}K_{\nu,\rho}^*(u)\,du\right) X_i^{(\rho+1)}(z)X_i^{(\nu)}(z)dz\\
&\quad + \frac{\nu!^2}{(\rho+1)!^2}\int \frac{b^{2\rho+2}}{b^{2\nu}}\left(\int u^{\rho+1}K_{\nu,\rho}^*(u)\,du\right)^2 X_i^{(\rho+1)}(z)X_i^{(\rho+1)}(z)dz\\
&\quad + \mathcal{O}_P\left(\frac{1}{T_i^{3/2}b^{2\nu+1}}\right)\Bigg\}\{1+o_P(1)\}
\end{aligned}$$

These results were obtained by substitution with $x = z + ub, y = z + vb$ and using a $\rho+1$ order Taylor expansion of $X_i(z+ub)$ and $X_i(z+vb)$ together with (S3.4). We get

$$\int X_i^{(\nu)}(u)^2du - \mathsf{E}\left(\hat{M}_{ii}^{(\nu)} \middle| \hat{X}_{i,b}^{(\nu)}\right) = \mathcal{O}_p\left(b^{\rho+1-\nu} + \left(T_i^{3/2}b^{2\nu+1}\right)^{-1}\right).$$

First note that by the second mean value integration theorem there exits some $c \in (0,1)$ such that

$$\int \frac{1}{f(z)^2} K_{\nu,\rho}^*\left(\frac{y-z}{b}\right) K_{\nu,\rho}^*\left(\frac{x-z}{b}\right) dz = \frac{1}{f(c)^2}\int K_{\nu,\rho}^*\left(\frac{y-z}{b}\right) K_{\nu,\rho}^*\left(\frac{x-z}{b}\right) dz.$$

We introduce a kernel convolution with

$$K_{\nu,\rho}^C(y-x) \overset{\text{def}}{=} \int K_{\nu,\rho}^*(y-z)\,K_{\nu,\rho}^*(x-z)\,dz$$

and thus using $z = \frac{u}{b}$

$$K^C_{\nu,\rho}\left(\frac{y-x}{b}\right) = \int K^*_{\nu,\rho}\left(\frac{y}{b}-z\right)K^*_{\nu,\rho}\left(\frac{x}{b}-z\right)dz$$
$$= \int b^{-1}K^*_{\nu,\rho}\left(\frac{y-u}{b}\right)K^*_{\nu,\rho}\left(\frac{x-u}{b}\right)du.$$

Note that the integral over $K^C_{\nu,\rho}$ is computed over an parallelogram D bounded by the lines $x+y=2, x+y=0, x-y=1, x-y=-1$. Using the substitution $x = \frac{v+u}{2}b,\ y = \frac{u-v}{2}b$

$$\int\int_D K^C_{\nu,\rho}\left(\frac{y-x}{b}\right)dydx = \frac{b}{2}\int_0^2\int_{-1}^1 K^C_{\nu,\rho}(v)\,dvdu = b\int K^C_{\nu,\rho}(v)\,dv.$$

Note that the variance can be decomposed

$$\begin{aligned}
&\operatorname{Var}\left(\hat{M}^{(\nu)}_{ii}\middle|\ \hat{X}^{(\nu)}_{i,b}\right)\\
=&\frac{\nu!^4}{T_i^4(b^{4\nu+2})f(c)^4}\left\{\sum_{l=1}^{T_i}K^C_{\nu,\rho}(0)^2\operatorname{Var}(Y_i(t_{il})^2)\right. && \text{(S3.5)}\\
&+2\sum_{l=1}^{T_i}\sum_{k\neq l}^{T_i}\operatorname{Var}(Z_{ilk}) && \text{(S3.6)}\\
&+4\sum_{l=1}^{T_i}\sum_{k\neq l}^{T_i}\sum_{k'\neq k}^{T_i}\operatorname{Cov}(Z_{ilk},Z_{ilk'}) && \text{(S3.7)}\\
&\left.+24\sum_{l=1}^{T_i}\sum_{k\neq l}^{T_i}\sum_{k'\neq k}^{T_i}\sum_{l'\neq k'}^{T_i}\operatorname{Cov}(Z_{ilk},Z_{ilk'})\right\} && \text{(S3.8)}\\
&+\mathcal{O}_P\left(\frac{1}{T_i}\right),
\end{aligned}$$

where $Z_{ilk} = K^C_{\nu,\rho}\left(\frac{t_{il}-t_{ik}}{b}\right)Y_i(t_{il})Y_i(t_{ik})$. Expression (S3.8) vanishes, while (S3.5) is of order $\mathcal{O}_P\left(\frac{1}{T_i^3 b^{4\nu+2}}\right)$ and is thus dominated by (S3.6), since

$$\begin{aligned}
&\frac{2\nu!^4}{T_i^4(b^{4\nu+2})f(c)^4}\sum_{l=1}^{T_i}\sum_{k\neq l}^{T_i}K^C_{\nu,\rho}\left(\frac{t_{il}-t_{ik}}{b}\right)^2\operatorname{Var}(Y_i(t_{il})Y_i(t_{ik}))\\
=&\frac{2\nu!^4}{T_i^4(b^{4\nu+2})f(c)^4}\sum_{l=1}^{T_i}\sum_{k\neq l}^{T_i}K^C_{\nu,\rho}\left(\frac{t_{il}-t_{ik}}{b}\right)^2\left\{\mathsf{E}(Y_i(t_{il})^2Y_i(t_{ik})^2)-\mathsf{E}(Y_i(t_{il})Y_i(t_{ik}))^2\right\}\\
=&\frac{2\nu!^4\int(\sigma^4_{i\epsilon}+2\sigma^2_{i\epsilon}X(x)^2)f(x)^2dx}{T_i^2b^{4\nu+1}f(c)^4}\int\left(K^C_{\nu,\rho}(u)\right)^2du+o_P\left(\frac{1}{T_i^2b^{4\nu+1}}\right).
\end{aligned}$$

Before looking at expression (S3.7), note that with $m \geq 2\nu$

$$\begin{aligned}
&\int\int \frac{\nu!^2}{b^{2\nu+1}} K^C_{\nu,\rho}\left(\frac{x-y}{b}\right) X_i(x)dxdy \\
=&\frac{\nu!^2}{b^{2\nu}} \int\int\int K^*_{\nu,\rho}\left(m\right) K^*_{\nu,\rho}\left(z\right) X_i\left\{y+(m-z)b\right\} dzdmdy \\
=&(-1)^d \int X_i^{(2\nu)}(y)dy + \mathcal{O}_P(1)
\end{aligned} \tag{S3.9}$$

by performing two Taylor expansions with mb first and then $-zb$.

We can thus derive for expression (S3.7) that

$$\begin{aligned}
&H(T)\sum_{l=1}^{T_i}\sum_{k\neq l}^{T_i}\sum_{k'\neq k}^{T_i} \mathrm{Cov}\left(Z_{ilk}, Z_{ilk'}\right) \\
=&H(T)\sum_{l=1}^{T_i}\sum_{k\neq l}^{T_i}\sum_{k'\neq k}^{T_i} K^C_{\nu,\rho}\left(\frac{t_{ik}-t_{il}}{b}\right) K^C_{\nu,\rho}\left(\frac{t_{il}-t_{ik'}}{b}\right)\left\{\mathsf{E}\left(Y_i(t_{ik})Y_i(t_{il})^2Y_i(t_{ik'})\right)\right. \\
&\left.-\mathsf{E}\left(Y_i(t_{ik})Y_i(t_{il})\right)\mathsf{E}\left(Y_i(t_{il})Y_i(t_{ik'})\right)\right\} \\
=&H(T)\sum_{l=1}^{T_i}\sum_{k=1}^{T_i}\sum_{k'=1}^{T_i} K^C_{\nu,\rho}\left(\frac{t_{ik}-t_{il}}{b}\right) K^C_{\nu,\rho}\left(\frac{t_{il}-t_{ik'}}{b}\right) X_i(t_{ik})\sigma^2_{i\epsilon}X_i(t_{ik'}) \\
&-\frac{2\nu!^4}{T_i^4(b^{4\nu+2})f(c)^4}\sum_{k=1}^{T_i}\sum_{k'=1}^{T_i} K^C_{\nu,\rho}\left(\frac{t_{il}-t_{ik'}}{b}\right)^2 X_i(t_{ik})\sigma^2_{i\epsilon}X_i(t_{ik'}) \\
=&\frac{4\sigma^2_{i\epsilon}}{T_i f(c)}\int X_i^{(2\nu)}(y)X_i^{(2\nu)}(y)dy - \mathcal{O}_P\left(\frac{1}{T_i^2(b^{4\nu+1})}\right),
\end{aligned}$$

where $H(T) \overset{\text{def}}{=} \frac{4\nu!^4}{T_i^4(b^{4\nu+2})f(c)^4}$. Thus $\mathrm{Var}\left(\hat{M}^{(\nu)}_{ii}\middle|\, \hat{X}^{(\nu)}_{i,b}\right) = \mathcal{O}_P\left(\frac{1}{T_i^2(b^{4\nu+1})}+\frac{1}{T}\right)$.

S3.2 Multivariate case ($g > 1$)

The same strategy also works in the multivariate case by using multivariate Taylor series. Using a vector of partial derivatives $a = (a_1, ..., a_g)^\top$, $a_l \in \mathbb{N}$, a multivariate Taylor expansion of degree $k < \rho$ is given by

$$X_i(x - u\circ b) = \sum_{0\leq|a|\leq k}\frac{X_i^{(a)}(x)}{a!}(u\circ b)^a + \mathcal{O}_P\left(u^{k+1}\max(b)^{k+1}\right). \tag{S3.10}$$

Using the equivalent kernel by Ruppert and Wand (1994) extended to the case and using Masry (1996) we can further state that with a bandwidth fulfilling $\frac{log(T)}{Tb_1 \times \cdots \times b_g} \to 0$ we have uniformly in $u \in [0,1]^g$ that $S_{T_i}(u)^{-1} \to \frac{S(u)^{-1}}{f(u)}$ almost surely as $T \to \infty$. Furthermore, the multivariate equivalent kernel has the properties that with $v = (v_1, \ldots, v_g)^\top,\ v_l \in \mathbb{N}$

$$\int u^v K^*_{\nu,\rho}(u)\, du = \delta_{\nu,v},\ |v| \leq \rho,\ 0 \leq v_i\ \forall i = 1, \ldots g. \tag{S3.11}$$

Let c be the position of $\max(b)$ in b and $\tilde{\rho}$ be a vector of length g which is $\rho + 1$ at the $c - th$ position and 0 else. Then for $z \in [0,1]^g$

$$\begin{aligned}
&\mathsf{E}\left(\hat{M}_{ii}^{(\nu)} \middle| \hat{X}_{i,b}^{(\nu)}\right) \\
&= \Bigg\{ \int X_i^{(\nu)}(z) X_i^{(\nu)}(z) dz \\
&\quad + 2\frac{\nu!}{(\rho+1)!} \int \frac{\max(b)^{\rho+1}}{b^\nu} \left(\int u^{\tilde{\rho}} K^*_{\nu,\rho}(u)\, du \right) X_i^{(\tilde{\rho})}(z) X_i^{(\nu)}(z) dz \\
&\quad + \mathcal{O}_P\left(\frac{\max(b)^{\rho+1}}{b^\nu} + \frac{1}{T_i^{3/2}(b^{2\nu} b_1 \times \cdots \times b_g)} \right) \Bigg\} \{1 + o_P(1)\}.
\end{aligned} \tag{S3.12}$$

Further note that for the convoluted kernel we get

$$\begin{aligned}
&K^C_{\nu,\rho}\left((y-x) \circ b^{-1}\right) \\
&= \int (b_1 \times \cdots \times b_g)^{-1} K^*_{\nu,\rho}\left\{(y-u) \circ b^{-1}\right\} K^*_{\nu,\rho}\left\{(x-u) \circ b^{-1}\right\} du.
\end{aligned}$$

Accordingly, we get for the multivariate equivalent of expression (S3.6) that

$$\begin{aligned}
&\frac{2\nu!^4}{T^4 f(c)^4 (b_1^2 \times \cdots \times b_g^2 b^{4\nu})} \sum_{l=1}^{T_i} \sum_{k \neq l}^{T_i} K^C_{\nu,\rho}\left((t_{il} - t_{ik}) \circ b^{-1}\right)^2 \operatorname{Var}(Y_i(t_{il}) Y_i(t_{ik})) \\
=& \frac{2\nu!^4 \int (\sigma_{i\epsilon}^4 + 2\sigma_{i\epsilon}^2 X(x)^2) f(x)^2 dx}{T_i^2 f(c)^4 b_1 \times \cdots \times b_g b^{4\nu}} \int \left(K^C_{\nu,\rho}(u)\right)^2 du \{1 + o_P(1)\}.
\end{aligned}$$

Because we assume that $m \geq 2|\nu|$ we obtain the multivariate equivalent of expression

(S3.7) for $Z_{ilk} = K_{\nu,\rho}^{C}\left((t_{il} - t_{ik}) \circ b^{-1}\right) Y(t_{il}) Y_i(t_{ik})$

$$
\begin{aligned}
&A(T) \sum_{l=1}^{T_i} \sum_{k \neq l}^{T_i} \sum_{k' \neq k}^{T_i} \operatorname{Cov}\left(Z_{ilk}, Z_{ilk'}\right) \\
&= A(T) \sum_{l=1}^{T_i} \sum_{k \neq l}^{T_i} \sum_{k' \neq k}^{T_i} K_{\nu,\rho}^{C}\left((t_{ik} - t_{il}) \circ b^{-1}\right) K_{\nu,\rho}^{C}\left((t_{il} - t_{ik'}) \circ b^{-1}\right) X_i(t_{ik}) \sigma_{i\epsilon}^2 X_i(t_{ik'}) \\
&= \frac{4\sigma_{i\epsilon}^2}{T_i f(c)} \int X_i^{(2\nu)}(y) X_i^{(2\nu)}(y) dy + \mathcal{O}_P\left(\frac{1}{T_i^2 (b^{4\nu} b_1 \times \cdots \times b_g)}\right)
\end{aligned}
$$

where $A(T) \stackrel{\text{def}}{=} \frac{4\nu!^4}{T_i^4 (b^{4\nu} b_1^2 \times \cdots \times b_g^2) f(c)^4}$. These arguments imply (S2.1) for $i = j$.

When $i \neq j$, the proof is analogous except that the diagonal correction

$$
\nu!^2 \hat{\sigma}_{i\varepsilon}^2 \int_{[0,1]^g} \sum_{k=1}^{T_i} W_\nu^{T_i}\left((t_{ik} - t) \circ b^{-1}\right)^2 dt
$$

is no longer necessary.

S4. Convergence of $\tilde{M}_{ij}^{(0)}$

We have

$$
\begin{aligned}
M_{ij}^{(0)} - \tilde{M}_{ij}^{(0)} &= \int_{[0,1]^g} X_i(t) X_j(t) dt - \frac{1}{T} \sum_{l=1}^{T} Y_i(t_{il}) Y_j(t_{jl}) + I(i = j) \hat{\sigma}_{i\varepsilon}^2 \\
&= \int_{[0,1]^g} X_i(t) X_j(t) dt - \frac{1}{T} \sum_{l=1}^{T} \left(X_i(t_l) + \varepsilon_{il}\right)\left(X_j(t_l) + \varepsilon_{jl}\right) + I(i = j) \hat{\sigma}_{i\varepsilon}^2 \\
&= \int_{[0,1]^g} X_i(t) X_j(t) dt - \frac{1}{T} \sum_{l=1}^{T} X_i(t_l) X_j(t_l) \\
&\quad - \frac{1}{T} \sum_{l=1}^{T} X_i(t_l) \varepsilon_{jl} - \frac{1}{T} \sum_{l=1}^{T} X_j(t_l) \varepsilon_{il} - \frac{1}{T} \sum_{l=1}^{T_i} \varepsilon_{il} \varepsilon_{jl} + I(i = j) \hat{\sigma}_{i\varepsilon}^2.
\end{aligned}
$$

By assumption, the random variables $X_i(t_l)$ and ε_{il} are independent, and $\mathsf{E}\left[\varepsilon_{il}\varepsilon_{jl}\right] = 0,\ i \neq j$, $\mathsf{E}\left[\varepsilon_{il}{}^2\right] = \sigma_{i\varepsilon}^2$. Additionally, $\hat{\sigma}_{i\varepsilon}$ is $T^{-1/2}$ consistent. It follows from standard arguments that $M_{ij}^{(0)} - \tilde{M}_{ij}^{(0)} = \int_{[0,1]^g} X_i(t) X_j(t) dt - \frac{1}{T} \sum_{l=1}^{T} X_i(t_l) X_j(t_l) + \mathcal{O}_P(T^{-1/2})$.

But recall that t_l is independent of X_i. Hence $\mathsf{E}\left(\frac{1}{T}\sum_{l=1}^{T} X_i(t_l)X_j(t_l)\middle| X_i, X_j\right) = \int_{[0,1]^g} X_i(t)X_j(t)dt$ and $\text{Var}\left(X_i(t_l)X_j(t_l)\middle| X_i, X_j\right) = \mathcal{O}(T^{-1})$, similarly to Remark 2.

S5. Proof of Proposition 2

Under the assumptions of Proposition 2 together with the requirements of Proposition 1 and the setup of Remark 1 for $\nu \in \{0,\, d\}$

$$||\hat{M}^{(\nu)} - M^{(\nu)}|| \leq \text{tr}\left\{\left(\hat{M}^{(\nu)} - M^{(\nu)}\right)^\top \left(\hat{M}^{(\nu)} - M^{(\nu)}\right)\right\}^{1/2} = \mathcal{O}_p\left(NT^{-1/2}\right).$$

Given that $\sum_{l=1}^{T} p_{lr}^{(\nu)} = 0,\ \sum_{l=1}^{T}\left(p_{lr}^{(\nu)}\right)^2 = 1\ \forall r$ and applying Cauchy-Schwarz inequality gives $\sum_{l=1}^{N} |p_{lr}^{(\nu)}| = \mathcal{O}_p\left(N^{1/2}\right)$. This together with Lemma A from Kneip and Utikal (2001) leads to

$$\mathsf{E}\left[\left(p_r^{(\nu)}\right)^\top (\hat{M}^{(\nu)} - M^{(\nu)})p_r^{(\nu)}\middle| \hat{X}_{i,b}^{(\nu)};\, i = 1, \ldots, N\right]^2 = \mathcal{O}_p\left(\frac{N}{T}\right).$$

We can now make a statement about the basis that span the factor space

$$\begin{aligned}&\left|\frac{1}{\sqrt{l_r^{(\nu)}}}\sum_{i=1}^{N} p_{ir}^{(\nu)} X_i^{(d)}(t) - \frac{1}{\sqrt{\hat{l}_r^{(\nu)}}}\sum_{i=1}^{N} \hat{p}_{ir}^{(\nu)}\hat{X}_{i,h}^{(d)}(t)\right| \leq \\ &\left|\frac{1}{\sqrt{l_r^{(\nu)}}}\sum_{i=1}^{N} p_{ir}^{(\nu)}\left[X_i^{(d)}(t) - \hat{X}_{i,h}^{(d)}(t)\right]\right| + \left|\sum_{i=1}^{N}\left(\frac{1}{\sqrt{l_r^{(\nu)}}}p_{ir}^{(\nu)} - \frac{1}{\sqrt{\hat{l}_r^{(\nu)}}}\hat{p}_{ir}^{(\nu)}\right)\hat{X}_{i,h}^{(d)}(t)\right|.\end{aligned} \tag{S5.13}$$

The first term is discussed in equation (2.4.3). Therefore we take a look at the second term here. Recall that $l_r^{(\nu)} = N\lambda_r \cdot (1 + O_P(N^{-1/2}))$. As a consequence of Assumption (8), Lemma A (a) from Kneip and Utikal (2001) together with equation (S5) give

$$l_r^{(\nu)} - \hat{l}_r^{(\nu)} = (p_r^{(\nu)})^\top(\hat{M}^{(\nu)} - M^{(\nu)})p_r^{(\nu)}) + \mathcal{O}_p(NT^{-1}) = \mathcal{O}_p(N^{1/2}T^{-1/2} + NT^{-1}),$$

where

$$\frac{1}{\sqrt{\hat{l}_r^{(\nu)}}} - \frac{1}{\sqrt{l_r^{(\nu)}}} = \frac{l_r^{(\nu)} - \hat{l}_r^{(\nu)}}{\sqrt{\hat{l}_r^{(\nu)}}\sqrt{l_r^{(\nu)}}(\sqrt{\hat{l}_r^{(\nu)}} + \sqrt{l_r^{(\nu)}})} = \mathcal{O}_p\left(T^{-1/2}N^{-1} + T^{-1}N^{-1/2}\right). \tag{S5.14}$$

Using Lemma A (b) from Kneip and Utikal (2001) we further get

$$|\hat{p}_{ir}^{(\nu)} - p_{ir}^{(\nu)}| = \mathcal{O}_p\left((NT)^{-1/2}\right) \text{ and } ||\hat{p}_r^{(\nu)} - p_r^{(\nu)}|| = \mathcal{O}_p\left(T^{-1/2}\right). \tag{S5.15}$$

Putting all results together for the second term gives

$$\begin{aligned}
&\left|\sum_{i=1}^{N}\left(\frac{1}{\sqrt{l_r^{(\nu)}}}p_{ir}^{(\nu)} - \frac{1}{\sqrt{\hat{l}_r^{(\nu)}}}\hat{p}_{ir}^{(\nu)}\right)\hat{X}_{i,h}^{(d)}(t)\right| = \\
=&\left|\sum_{i=1}^{N}\left(\frac{1}{\sqrt{l_r^{(\nu)}}} - \frac{1}{\sqrt{\hat{l}_r^{(\nu)}}}\right)\hat{p}_{ir}^{(\nu)}\hat{X}_{i,h}^{(d)}(t) + \frac{1}{\sqrt{l_r^{(\nu)}}}\sum_{i=1}^{N}\left(\hat{p}_{ir}^{(\nu)} - p_{ir}^{(\nu)}\right)\hat{X}_{i,h}^{(d)}(t)\right| \\
\leq&\left|\left(\frac{1}{\sqrt{l_r^{(\nu)}}} - \frac{1}{\sqrt{\hat{l}_r^{(\nu)}}}\right)\right|\sum_{i=1}^{N}|p_{ir}^{(\nu)}|\left|\left(\hat{X}_{i,h}^{(d)}(t)\right)\right| \\
&+\left|\left(\frac{1}{\sqrt{l_r^{(\nu)}}} - \frac{1}{\sqrt{\hat{l}_r^{(\nu)}}}\right)\right| ||\hat{p}_r^{(\nu)} - p_r^{(\nu)}||\left|\hat{X}_{i,h}^{(d)}(t)\right| + \frac{1}{\sqrt{l_r^{(\nu)}}}||\hat{p}_r^{(\nu)} - p_r^{(\nu)}||\left|\hat{X}_{i,h}^{(d)}(t)\right| \\
=&\mathcal{O}_p\left((NT)^{-1/2}\right)\left|\hat{X}_{i,h}^{(d)}(t) - X_{i,h}^{(d)}(t) + X_{i,h}^{(d)}(t)\right| \\
\leq&\mathcal{O}_p\left((NT)^{-1/2}\right)(\text{Bias}\left(\hat{X}_{i,h}^{(d)}(t)\middle|Y_i,\, t_i\right) + \sqrt{\text{Var}\left(\hat{X}_{i,h}^{(d)}(t)\middle|Y_i,\, t_i,\right)} + \left|X_{i,h}^{(d)}(t)\right|).
\end{aligned}$$

Using Cauchy-Schwarz and equation (S5.14) we see that the first term is of order $(NT)^{-1/2}$. For the second term, remember that $l_r^{(\nu)}$ is of order N; together with (S5.15) this also leads to order $(NT)^{-1/2}$. Then equation (S5.13) becomes

$$\begin{aligned}
&\mathcal{O}_p\left(\max(h)^{p+1}h^{-d}\right) + \mathcal{O}_p\left((NTh_1\dots h_g h^{2d})^{-1/2}\right) \\
&+ \mathcal{O}_p\left((NT)^{-1/2}\right)\mathcal{O}_p\left(\max(h)^{p+1}h^{-d}\right) \\
&+ \mathcal{O}_p\left((NT)^{-1/2}\right)\mathcal{O}_p\left((Th_1\dots h_g h^{2d})^{-1/2}\right) + \mathcal{O}_p\left((NT)^{-1/2}\right) \\
&= \mathcal{O}_p\left(\max(h)^{p+1}h^{-d}\right) + \mathcal{O}_p\left((NTh_1\dots h_g h^{2d})^{-1/2}\right).
\end{aligned}$$

S6. Proof of Proposition 3

We use the following notations:

$$X_i^{(d)}(t) = \sum_{r=1}^{\infty} \delta_{ir}^{(d)} \varphi_r^{(d)}(t) = \sum_{r=1}^{\infty} \delta_{ir} \sum_{j=1}^{\infty} a_{jr} \varphi_j^{(d)}(t) = \sum_{r=1}^{\infty} \delta_{ir} \gamma_r^{(d)}(t)$$

$$X_{i,L_d,\varphi}^{(d)}(t) = \sum_{r=1}^{L_d} \delta_{ir}^{(d)} \varphi_r^{(d)}(t) \quad X_{i,L,\gamma}^{(d)}(t) = \sum_{r=1}^{L} \delta_{ir} \gamma_r^{(d)}(t)$$

$$\tilde{X}_{i,L_d,\varphi}^{(d)}(t) = \sum_{r=1}^{L_d} \tilde{\delta}_{ir}^{(d)} \tilde{\varphi}_r^{(d)}(t) \quad \tilde{X}_{i,L,\gamma}^{(d)}(t) = \sum_{r=1}^{L} \tilde{\delta}_{ir} \tilde{\gamma}_r^{(d)}(t)$$

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

For brevity, we illustrate the results for $\hat{X}_{i,L,\varphi}^{(d)}(t)$. First notice that

$$\begin{aligned} |X_i^{(d)}(t) - \hat{X}_{i,L,\varphi}^{(d)}(t)| =& |(X_i^{(d)}(t) - X_{i,L,\varphi}^{(d)}(t)) + (X_{i,L,\varphi}^{(d)}(t) - \\ & \tilde{X}_{i,L,\varphi}^{(d)}(t)) + (\tilde{X}_{i,L,\varphi}^{(d)}(t) - \hat{X}_{i,L,\varphi}^{(d)}(t))|. \end{aligned} \tag{S6.16}$$

Furthermore

$$\mathsf{E}(X_i^{(d)}(t) - X_{i,L_d,\varphi}^{(d)}(t))^2 = \sum_{r=L+1}^{\infty} \lambda_r^{(d)} \varphi_t^{(d)}(t)^2 \to 0 \text{ as } L \to \infty.$$

S6.1 Proof of Proposition 3 a)

In the finite case with $L \leq N$, use Hall and Hosseini-Nasab (2006) to show that a.s. $|\lambda_r^{(d)} - \hat{\lambda}_r^{(d)}| = 0$ for $r > L$. This implies that $\sum_{r=1}^{L_d} \delta_{ir}^{(d)} \varphi_r^{(d)} = \sum_{r=1}^{L_d} \tilde{\delta}_{ir}^{(d)} \tilde{\varphi}_r^{(d)}$ even though $\varphi_r^{(d)} \neq \tilde{\varphi}_r^{(d)}$, $\delta_{ir}^{(d)} \neq \tilde{\delta}_{ir}^{(d)}$. Then $|X_i^{(d)}(t) - X_{i,L_d,\varphi}^{(d)}(t)| = |X_{i,L_d,\varphi}^{(d)}(t) - \tilde{X}_{i,L_d,\varphi}^{(d)}(t)| = 0$. Further, note that

$$\sqrt{l_r^{(d)}} - \sqrt{\hat{l}_r^{(d)}} = (l_r^{(d)} - \hat{l}_r^{(d)})(\sqrt{l_r^{(d)}} + \sqrt{\hat{l}_r^{(d)}})^{-1} = \mathcal{O}_p(N^{1/2}T^{-1/2} + NT^{-1}) \tag{S6.17}$$

and from equation (S5.15)

$$\begin{aligned} &\tilde{\delta}_{ir}^{(d)} - \hat{\delta}_{ir,T}^{(d)} = \sqrt{l_r^{(d)}} p_{ir}^{(d)} - \sqrt{\hat{l}_r^{(d)}} \hat{p}_{ir}^{(d)} \\ &= \left(\sqrt{l_r^{(d)}} - \sqrt{\hat{l}_r^{(d)}}\right) p_{ir}^{(d)} - \sqrt{\hat{l}_r^{(d)}} \left(\hat{p}_{ir}^{(d)} - p_{ir}^{(d)}\right) = \mathcal{O}_p(T^{-1/2} + N^{1/2}T^{-1}). \end{aligned} \tag{S6.18}$$

Using Proposition 2 and equation (S6.16) it follows that

$$
\begin{aligned}
&|X_i^{(d)}(t) - \hat{X}_{i,L_d,\varphi}^{(d)}(t)| = |\tilde{X}_{i,L_d,\varphi}^{(d)}(t) - \hat{X}_{i,L_d,\varphi}^{(d)}(t)| \\
=&|\sum_{r=1}^{L_d} \tilde{\delta}_r^{(d)} \tilde{\varphi}_r^{(d)}(t)(t) - \sum_{r=1}^{L_d} \hat{\delta}_{ir,T}^{(d)} \hat{\varphi}_{r,T}^{(d)}(t)| \\
=&|\sum_{r=1}^{L_d} (\tilde{\delta}_{ir}^{(d)} - \hat{\delta}_{ir,T}^{(d)}) \varphi_r^{(d)}(t) + \hat{\delta}_{ir,T}^{(d)} (\varphi_r^{(d)}(t) - \hat{\varphi}_{r,T}^{(d)}(t))| \\
=&\mathcal{O}_p \left(T^{-1/2} + N^{1/2}T^{-1} + \max(h)^{p+1}h^{-d} + (NTh_1 \times \cdots \times h_g h^{2d})^{-1/2} \right).
\end{aligned}
\tag{S6.19}
$$

For $N/T \to 0$, we get the result in Proposition 3 a). The proof for $\hat{X}_{i,L,\gamma}^{(d)}(t)$ follows similar considerations.

S6.2 Proof of Proposition 3 b)

Given equation (S6.16)

$$|X_i^{(d)}(t) - \hat{X}_{i,L_d,\varphi}^{(d)}| \leq |X_i^{(d)}(t) - X_{i,L_d,\varphi}^{(d)}(t)| + |X_{i,L_d,\varphi}^{(d)}(t) - \tilde{X}_{i,L_d,\varphi}^{(d)}(t)| + |\tilde{X}_{i,L_d,\varphi}^{(d)}(t) - \hat{X}_{i,L_d,\varphi}^{(d)}(t)|.$$

Note that

$$\mathsf{E}(X_i^{(d)}(t) - X_{i,L_d,\varphi}^{(d)}(t))^2 = \sum_{r=L+1}^{\infty} \lambda_r^{(d)} \varphi_t^{(d)}(t)^2. \tag{S6.20}$$

Equation (S6.20) implies $|X_i^{(d)}(t) - X_{i,L_d,\varphi}^{(d)}(t)| \xrightarrow{p} 0$ as $L \to \infty$. Further, it can be verified by equations (S6.17) - (S6.18) and the results from equations (2.8) and (2.9) in Hall and Hosseini-Nasab (2006) - $|\delta_{ir}^{(d)} - \hat{\delta}_{ir}^{(d)}| = \mathcal{O}_p(N^{-1/2})$ and $|\varphi_r^{(d)}(t) - \hat{\varphi}_r^{(d)}(t)| = \mathcal{O}_p(N^{-1/2})$ - that

$$
\begin{aligned}
&|X_{i,L_d,\varphi}^{(d)}(t) - \tilde{X}_{i,L_d,\varphi}^{(d)}(t)| = |\sum_{r=1}^{L_d} (\delta_{ir}^{(d)} - \tilde{\delta}_{ir}^{(d)}) \varphi_r^{(d)}(t) + \tilde{\delta}_{ir} (\varphi_r^{(d)}(t) - \tilde{\varphi}_r^{(d)}(t))| = \mathcal{O}(N^{-1/2}) \\
&|\tilde{X}_{i,L_d,\varphi}^{(d)}(t) - \hat{X}_{i,L_d,\varphi}^{(d)}(t)| = |\sum_{r=1}^{L_d} (\tilde{\delta}_{ir} - \hat{\delta}_{ir,T}) \tilde{\varphi}_r^{(d)} + \hat{\delta}_{ir,T} (\tilde{\varphi}_r^{(d)} - \hat{\varphi}_{r,T}^{(d)})| \\
&= \mathcal{O}_p \left(N^{-1/2} + T^{-1/2} + N^{1/2}T^{-1} + \max(h)^{p+1}h^{-d} + (NTh_1 \times \cdots \times h_g h^{2d})^{-1/2} \right) \\
&= \mathcal{O}_p \left(N^{-1/2} + \max(h)^{p+1}h^{-d} + (NTh_1 \times \cdots \times h_g h^{2d})^{-1/2} \right)
\end{aligned}
$$

when $N/T \to 0$. Notice that to obtain consistency of the estimators we need only that $N^{1/2}T^{-1}$. For L fixed, both $|X^{(d)}_{i,L_d,\varphi}(t) - \tilde{X}^{(d)}_{i,L_d,\varphi}(t)| \xrightarrow{p} 0$ and $|\tilde{X}^{(d)}_{i,L_d,\varphi}(t) - \hat{X}^{(d)}_{i,L_d,\varphi}(t)| \xrightarrow{p} 0$ as $N \to \infty$, given assumption $\max(h)^{p+1}h^{-d} \to 0$ in Proposition 2. The proof for $\hat{X}^{(d)}_{i,L,\gamma}(t)$ is analogous.

S7. Implementation

S7.1 Centering the observed curves

Throughout the theoretical part of the paper it has been assumed that the curves are centered. To satisfy this assumption, we subtract the empirical mean $\bar{X}^{(\nu)}(t) = \frac{1}{N}\sum_{i=1}^{N} \hat{X}^{(\nu)}_{i,b}(t)$ from the the observed call prices to obtain centered curves. A centered version $\overline{M}^{(\nu)}$, $\nu \in \{0, d\}$ is given by

$$\overline{M}^{(\nu)}_{ij} = \hat{M}^{(\nu)}_{ij} - \int_{[0,1]^g} \left(\bar{X}^{(\nu)}(t)\hat{X}^{(\nu)}_{i,b}(t) + \bar{X}^{(\nu)}(t)\hat{X}^{(\nu)}_{j,b}(t) - \bar{X}^{(\nu)}(t)^2 \right). \tag{S7.21}$$

The integral in (S7.21) is evaluated by Monte Carlo integration. Improving the centering of the curves is possible. First, one can use a different bandwidth than b to compute the mean $\bar{X}^{(\nu)}(t)$ because averaging will lower its variance. Second, by the arguments of Section 2.4.2, the term $\int_{[0,1]^g} \bar{X}^{(\nu)}(t)^2 dt$ can be improved accordingly to Lemma 1 by subtracting $\hat{\sigma}^2_{\varepsilon} = N^{-1}\sum_{i=1}^{N} \hat{\sigma}^2_{i\varepsilon}$ weighted by suitable parameters. We decided to omit these fine tunings in our application because they involve a significant additional computational effort for only minor improvements.

S7.2 Bandwidth selection

To get parametric rates of convergence for $\hat{M}^{(d)}$, as shown in Remark 1, we choose $\rho = 7$ and b between $\mathcal{O}(T^{-1/10})$ and $\mathcal{O}(T^{-1/12})$. The choice of b to estimate $\hat{M}^{(0)}$ is similar, with the difference that it is only required that $\rho > 0$, therefore we use $\rho = 1$, while b has to lie between $\mathcal{O}(T^{-1/3})$ and $\mathcal{O}(T^{-1/5})$. We use a simple criteria to choose the bandwidth because by Proposition 2 the dominating error depends mainly on the choice of h. Let $t_{ik} = (t_{ik1}, \dots t_{ikg})$, then the bandwidth for direction j is determined by $b_j = ((\max_k(t_{ikj}) - \min_k(t_{ikj}))T_i)^{\alpha}$. When estimating state price densities $t_{ik} = (\tau_{ik}, m_{ik})$ and T_i is replaced by the cardinality of $\tau_i = \{\tau_{i1}, \dots \tau_{iT_i}\}$ and m_i respectively. In the estimation of $\hat{M}^{(d)}$ we set $\alpha = -1/10$ and $\alpha = -1/3$ for $\hat{M}^{(0)}$ when $g = 2$ and $\alpha = -3/10$ when $g = 1$.

The choice of bandwidths h is a crucial parameter for the quality of the estimators. To derive an estimator for the bandwidths first note that in the univariate case ($g = 1$) the theoretical optimal univariate asymptotic bandwidth for the r-th basis is given by

$$h^{d,\nu}_{r,opt} = C_{d,p}(K) \left[T^{-1} \frac{\int_0^1 \sum_{i=1}^N (p^{(\nu)}_{ir})^2 \sigma^2_{\varepsilon i}(t) f_i(t)^{-1} dt}{\int_0^1 \left\{ \sum_{i=1}^N p^{(\nu)}_{ir} X_i^{(p+1)}(t) \right\}^2 dt} \right]^{1/(2p+3)} \tag{S7.22}$$

$$C_{d,p}(K) = \left[\frac{(p+1)!^2 (2d+1) \int K^{*2}_{p,d_j}(t) dt}{2(p+1-d)\{\int u^{p+1} K^*_{d,p}(t) dt\}^2} \right]^{1/(2p+3)} .$$

Like in the conventional local polynomial smoothing case $C_{d,p}(K)$ does not depend on the curves and is an easily computable constant. It only depends on the chosen kernel, the order of the derivative and the order of the polynomial, see for instance Fan and Gijbels (1996).

In the multivariate case, for our bandwidth estimator we treat every dimension separately, similar to choosing an optimal an optimal bandwidth for derivatives in the univariate case, and correct for the asymptotic order, see Section 2.4.4. In practice, we can not use equation (S7.22) to determine the optimal bandwidth because some variables are unknown and only discrete points are observed. As a rule-of-thumb, we replace these unknown variables with empirical quantities: estimates of $p^{(0)}_{ir}$ from $\hat{M}^{(0)}$ and of $p^{(d)}_{ir}$ from $\hat{M}^{(d)}$. With these approximations, a feasible rule for computing the optimal bandwidth in direction j for the r-th basis function gives

$$h^{d,\nu}_{jr,rot} = \left(T^{-1} \frac{C^{2p+3}_{d,p} \hat{\sigma}^2_{\varepsilon}}{f_j \int_0^1 \left\{ \sum_{i=1}^N \hat{p}^{(\nu)}_{ir} \tilde{X}_i^{(p+1)}(t_j) \right\}^2 dt_j} \right)^{1/(g+2p+2)} . \tag{S7.23}$$

In our application as well as our main simulation we have $g = 2$ and $d = (2, 0)$. If $\nu = (0, 0)^\top$ then $p = 1$ and if $\nu = (2, 0)^\top$ then $p = 3$. The integrals in (S7.23) are approximated by Riemann sums.

- The distribution of observed points is assumed to be uniform, hence the quantities f_j, $j = 1, 2$ in (S7.23) are approximated by $f_1 = \{\max_{i,j}(\tau_{ij}) - \min_{i,j}(\tau_{ij})\}^{-1}$, $f_2 = \{\max_{i,j}(m_{ij}) - \min_{i,j}(m_{ij})\}^{-1}$.

- To get a rough estimator for $X_i^{(p+1)}$ based on X_i, we use a polynomial regression. For our application, we take $p = 3$ and are thus interested in estimates for $X_i^{(4)}(m)$ and $X_i^{(4)}(\tau)$. We expect the curves to be more complex in the moneyness direction than in the maturity direction and we adjust the degree of the polynomials to reflect this issue. The estimates are then given by

$$\begin{aligned} a_i^* =& \arg\min_{a_i} \left(X_i(m,\tau) - a_{i0} + \sum_{l=1}^{5} a_{il} m^l + \sum_{l=6}^{9} a_{il} \tau^{(l-5)} \right) \\ \tilde{X}_i^{(4)}(m) =& 24 a_{i4}^* + 120 a_{i5}^* m \\ \tilde{X}_i^{(4)}(\tau) =& 24 a_{i9}^*. \end{aligned} \tag{S7.24}$$

- To estimate the variance for each curve we use the kernel approach given in (2.19) using a Epanechnikov kernel with a bandwidth of $T^{-2/(4+g)}$ for each spatial direction. In addition, these estimates are used to correct for the diagonal bias when $\hat{M}^{(0)}$ and $\hat{M}^{(d)}$ are estimated. In (S7.23) the average over all $\hat{\sigma}_{i\epsilon}$ is used.

We use the product Gaussian kernel to construct local polynomial estimators. For both $\hat{\gamma}_{r,T}^{(d)}$ and $\hat{\varphi}_{r,T}^{(d)}$, we employ the mean bandwidth $h_{i,rot}^{d,\nu} = L^{-1} \sum_{r=1}^{L} h_{ir,rot}^{d,\nu}$ to reduce the computation time. Since we demean the sample in (S7.21), we need to add $N^{-1} \sum_{i=1}^{N} \hat{X}_{i,h_{i,rot}^{d,\nu}}^{(d)}$ to the truncated decomposition to obtain the final estimators for the derivatives.

S7.3 Numerical integration

For simplicity, when calculating the integrals in $\hat{M}_{ij}^{(\nu)}$ according to equation (2.17), we use an equidistant grid to compute Riemann sums in the one-dimensional case, while for the two-dimensional case the common grid is randomly drawn from a uniform distribution with support $[\min_{i,j}(m_{ij}), \max_{i,j}(m_{ij})] \times [\min_{i,j}(\tau_{ij}), \max_{i,j}(\tau_{ij})]$ to evaluate the Monte Carlo integral.

S8. Comparison to FPCA-based method

Our second approach to represent derivatives and the approach in Liu and Müller (2009) are conceptually similar. In fact, equation (2.7) corresponds to equation (2) in their article. However, the estimation methods are different: they focus on estimating the covariance function and its derivatives, we estimate the dual matrix and smooth derivatives of the individual curves. The two approaches are motivated by the specific assumptions about the observations in a sparse or non-sparse setup.

Liu and Müller (2009) estimate the covariance function and its partial derivatives by applying a local linear smoother to the pooled "raw" covariances for sparse observations, extending an idea proposed by Yao et al. (2005). They give asymptotic results for the partial derivative of the covariance function in Theorem 1, in the case when T is fixed, and only N grows asymptotically.

Our setup is different; we do not consider sparse observations of functional data, but assume that T is sufficiently large, such that reasonable nonparametric estimation is possible. The dual matrix $M^{(0)}$ is not smooth and we estimate each entry $M_{ij}^{(0)}$ individually through Monte Carlo integration and local polynomial regression. When only the number of curves N increases asymptotically, while T is fixed, $\hat{M}_{ij}^{(0)}$ is not consistent, see Proposition 1. If T increase asymptotically, we show in Remark 1 that $\hat{M}_{ij}^{(0)}$ can achieve $1/\sqrt{T}$ rate if the underlying curves are smooth enough, under the given bandwidth rule. Thus, a comparison in terms of asymptotic behavior is possible only if we let both increase N and T grow asymptotically. We compare the finite sample performance of the two methods in a simulation study in Section 3.1.

Regarding the estimation of the principal loadings (scores), with only a few observations of the individual curves, the usual method to derive the loadings as an integral, see definition following equation (2.5), will not work. To better estimate the loadings when data is sparse, Liu and Müller (2009) use the conditional expectation under the assumption that the distribution of the data is Gaussian. If the distribution is not Gaussian, this affects the estimation of loadings in an obvious way, but the estimators for the individual curve derivatives can be interpreted as best linear predictors. Our estimators for the loadings $\hat{\delta}_{ir,T}^{(\nu)} = \sqrt{\hat{l}_r^{(\nu)}}\hat{p}_{ir}^{(\nu)}$ rely on $\hat{l}_r^{(\nu)}$ and $\hat{p}_{ir}^{(\nu)}$. When N grows too fast, the loadings for the individual curves do not converge. For this reason, we restrict N to grow such that $\sqrt{N}/T \to 0$.

		$RMISE\left(X^{(d)}, \hat{X}^{(d)}_{i,L,\gamma}\right)$				$RMISE\left(X^{(d)}, \hat{X}^{(d)}_{i,LM}\right)$			
T	N	Mean	Var	Med	IQR	Mean	Var	Med	IQR
25	25	59.92	9.54	28.31	38.68	142.23	11.37	114.91	99.86
50	25	27.65	0.64	20.42	19.20	129.88	2.25	117.79	56.50
100	25	14.01	0.10	11.42	9.60	71.50	4.49	56.42	74.57
25	50	24.93	0.28	20.51	18.10	85.31	1.25	85.60	50.41
50	50	16.03	0.10	13.33	11.10	93.58	1.20	87.07	44.04
100	50	12.55	0.10	9.98	9.00	58.37	1.86	55.68	66.80
25	100	40.62	10.09	24.03	20.19	80.46	1.73	72.05	46.30
50	100	12.67	0.08	10.43	8.74	68.29	0.82	61.81	31.70
100	100	7.95	0.04	6.19	5.43	55.17	0.98	54.31	37.51

Table 1: Simulation results for $g = 1$. Based on the mean and the median of $RMISE$, $\hat{X}^{(d)}_{i,L,\gamma}$ performs better than $\hat{X}^{(d)}_{i,LM}$ in all cases. Results for $\hat{X}^{(d)}_{i,L,\gamma}$ and $\hat{X}^{(d)}_{i,LM}$ improve with raising N and T. All results are multiplied by 10^3.

The approach in Liu and Müller (2009) can be computationally intensive because the double sum in their equation (7). Therefore, we use their Matlab code from the online repository '`PACE_matlab-master`' based on binning. We experimented with different binning points and chose to report the best performance results obtained with 15 bins. We use an Epanechnikov kernel and keep the default values from their code. We simulate call prices for the fixed time to maturity $\tau = 0.5$ years, for different N and T, perform 500 replications and report the mean and median of $RISE$ and $RMISE$, respectively. The standard deviation of the error is $\sigma_\varepsilon = 0.005$. The relative performance of the two methods is showed in Table 1 and Table 2.

$RISE$		Mean				Median			
T	N	$\hat{\gamma}_{1,T}^{(d)}$	$\hat{\gamma}_{2,T}^{(d)}$	$\hat{\gamma}_{1,LM}^{(d)}$	$\hat{\gamma}_{2,LM}^{(d)}$	$\hat{\gamma}_{1,T}^{(d)}$	$\hat{\gamma}_{2,T}^{(d)}$	$\hat{\gamma}_{1,LM}^{(d)}$	$\hat{\gamma}_{2,LM}^{(d)}$
25	25	2.25	340.19	5.84	32.16	1.78	60.80	4.98	27.75
50	25	1.60	103.78	3.02	25.45	1.16	47.05	3.00	23.41
100	25	1.23	47.35	8.02	187.15	0.85	27.91	2.85	177.37
25	50	1.47	34.74	3.34	26.39	1.08	24.43	2.95	26.01
50	50	0.98	38.17	2.94	22.67	0.94	32.71	2.84	22.74
100	50	0.75	24.37	2.75	100.12	0.78	20.15	2.68	25.23
25	100	1.37	73.15	3.36	22.24	1.25	31.34	2.92	21.89
50	100	0.63	34.66	2.89	23.12	0.54	18.90	2.88	23.12
100	100	0.42	36.75	2.34	28.66	0.39	19.86	2.73	23.91

Table 2: Simulation results for $g = 1$. Based on the mean and the median of $RISE$, $\hat{\gamma}_{1,T}^{(d)}$ performs better than $\hat{\gamma}_{1,LM}^{(d)}$ in all cases, while $\hat{\gamma}_{2,T}^{(d)}$ outperforms $\hat{\gamma}_{2,LM}^{(d)}$ for $T > N$. In general, results for all estimators improve with raising N and T.

S9. Supporting results for the analysis of DAX 30 SPDs

S9.1 Selection of components

The first eigenvalue of the empirical dual covariance matrix $\hat{M}^{(0)}$ has a dominant explanatory power. To detect the relative contribution of consecutive components, we construct the ratio of two adjacent estimated eigenvalues in descending order, see Ahn and Horenstein (2013). The first two terms are dominating the sequence and there are spikes at the fourth and seventh eigenvalue ratio. $PC^{(0)}$ criterion suggests at least seven components, see values of k^* for $L_{max} \geq 7$ in Table 3. $IC^{(0)}$ criterion, which does not depend on the truncation parameter L_{max}, suggests seven components. In the following, we investigate these components.

r, $L_{\max}$	1	2	3	4	5	6	7	8	9
$\hat{\lambda}_{r,T} \times 10^6$	133.29	18.90	2.69	1.62	0.49	0.34	0.26	0.09	0.08
$\hat{\lambda}_{r,T}/\hat{\lambda}_{r+1,T}$	7.05	7.01	1.66	3.28	1.44	1.31	2.83	1.18	1.70
$k^*(PC^{(0)})$	-	-	-	-	-	-	7	8	9
$k^*(IC^{(0)})$	-	-	-	-	-	-	7	-	-

Table 3: Selection of number of components

A closer look at the dynamics of the loadings $\hat{\delta}_{2,T}$ in the lower right panel of Figure 1 shows a highly volatile behavior from mid-February 2007 to mid-June 2008. This interval spans the financial crisis and extends until the end of the recession in the Euro Area, according to the Center for Economic and Policy Research (CEPR) recession indicator. In addition, during this interval, the loadings display a certain regularity of spikes. We identify the timing of these spikes with the Mondays following an expiration date (recall that expiration dates have a monthly frequency). Figure 1 highlights the dynamics of $\hat{\delta}_{2,T}$ on and following an expiration day. After roughly two weeks, the loadings revert to a 'normal' level. During this period, for small maturities, there are only few observations available for call prices with strikes larger than the current stock index. This potentially introduces bias in the presmoothed call surfaces for grid values outside the observation range, which translates in bias to the loadings. The shape of the second estimated component $\hat{\gamma}^{(d)}_{2,T}$, displayed in Figure 1, suggests that it is related to variations of the short end of the SPD term structure. A similar behavior is observed for the loadings of other components: $\hat{\delta}_{4,T}$, $\hat{\delta}_{5,T}$ and $\hat{\delta}_{6,T}$. The variance of these loadings remain important even if we exclude the financial crisis and recession observations from the sample. The corresponding components have similar shape features to the components $\hat{\delta}_{1,T}$, $\hat{\delta}_{3,T}$ and $\hat{\delta}_{7,T}$. We conjecture that they are related to reactions of option prices along the maturity direction.

The eigenfunctions of the covariance operator for the sample of approximating curves $\sum_{r\in\{1,3,7\}} \hat{\hat{\delta}}_{ir,T}\hat{\gamma}^{(d)}_{r,T}$, $i = 1, \ldots, N$ resemble closely the three components displayed in Figure 2. We find that when including additional components to the approximation of derivatives and perform spectral decomposition of their covariance operator, the shape of the resulting eigenfunctions changes to some degree and their loadings become 'contaminated' with spikes. Moreover, all the loadings estimated by decomposing $\hat{M}^{(d)}$, for $d = (2, 0)^\top$ feature the volatile behavior outlined above, between mid-February 2007 and mid-September 2008. For these reasons, we conjecture that $\hat{M}^{(0)}$ decomposition allows a better interpretation of the components, by separating the regular and irregular sources of variation in the SPDs.

S9.2 Interpretation of selected components

In this section we show that the first estimated component $\hat{\gamma}^{(d)}_{1,T}$ is related to the expected variance of the asset returns under the risk neutral measure, which admits the density q. Recall that under this measure, discounted prices are martingales. Then, equations (2.6) and (3.30) yield

$$\frac{\int_0^\infty mq(m,\tau)dm}{\exp(r_{i\tau}\tau)} = \int_0^\infty m\tilde{q}(m,\tau)dm + \sum_{r=1}^{\infty}\delta_{ir}\int_0^\infty m\gamma_r^{(d)}(m,\tau)dm = 1, \qquad \text{(S9.25)}$$

where $\tilde{q}$ is the population mean. The computation of the second moment gives

$$\frac{\int_0^\infty m^2q(m,\tau)dm}{\exp(r_{i\tau}\tau)^2} = \int_0^\infty m^2\tilde{q}(m,\tau)dm + \sum_{r=1}^{\infty}\delta_{ir}\int_0^\infty m^2\gamma_r^{(d)}(m,\tau)dm - 1. \qquad \text{(S9.26)}$$

We consider the empirical version of Equation (S9.26), for $\tau = 1$ month. Instead of computing the integrals, based on our estimates of $\tilde{q}$ and $\gamma_r^{(d)}$, we assume them to be fixed coefficients in a linear regression, in which the empirical loadings are used as explanatory variables of the real-data proxy for the standardized variance. In the numerator, we use the squared VDAX index multiplied by τ. This index is computed by Deutsche Börse AG from the prices of call and put options and reflects market expectation under the risk neutral measure of the 30 day ahead square root implied variance for the DAX 30 log-returns, which is then annualized. Duan and Yeh (2010) show that squared volatility index is a good approximation of the expected risk-neutral volatility when the jumps are small. While the volatility index refers to the standard deviation of the log-returns under the risk neutral measure, it can still be used in the regression because the transformation $q(\log m, \tau) = mq(m, \tau)$ maintains the linear-relationship between the dependent and explanatory variables. We find that the most important component in the regression is $\hat{\delta}_{1,T}$ (adjusted R-squared in the univariate regression is 93.97%). When including $\hat{\delta}_{3,T}$ as an additional regressor, it increases the adjusted R-squared to 94.06%, while $\hat{\delta}_{7,T}$ has a negative marginal contribution to the goodness of fit of multivariate regression.

No skewness index is readily available, and we take a simple measure instead, Pearson's skewness coefficient. In terms of equations (S9.25) and (S9.26), for a fixed maturity τ, this coefficient is equal to

$$\frac{1 - \arg\max_{m} \{q(m, \tau)\}}{\sqrt{\mathrm{Var}_i^Q\,(s_{i+\tau}/s_i)}/\exp(r_{i\tau}\tau)}. \tag{S9.27}$$

Since the first component $\hat{\gamma}^{(d)}_{1,T}$ is unimodal (as it is also $\hat{\gamma}^{(d)}_{2,T}$), the SPD mode is mostly affected by the loadings of the third component $\hat{\gamma}^{(d)}_{3,T}$ (and to some extend by those of the seventh component $\hat{\gamma}^{(d)}_{7,T}$).

S9.3 Preliminary analysis of the loadings

In this section we describe the preliminary analysis of the estimated loadings for first, third and seventh component. The partial autocorrelation function of all three time series display a salient spike at the first lag. This suggests that an autoregressive or perhaps an integrated model of order one might be appropriate to represent their dynamics. Their serial autocorrelations decay slowly, similarly to the integrated processes that feature a stochastic trend. Unit root and stationarity test results (not included in this draft) are ambiguous. When the null hypothesis assumes the existence of a unit root (augmented Dickey-Fuller unit-root test, Phillips-Perron test, variance-ratio test for random walk) the tests reject the null, while stationary tests that have the unit root hypothesis as an alternative (KPSS test, Leybourne-McCabe stationarity test) favor the alternative. Based on these results, we further investigate if the loadings are fractionally integrated of order $\alpha \in (0,\, 1)$, which is typical to long-memory processes. We employ Lo (1991)'s modified $R/\tilde{S}$ (range over standard deviation) rescaled statistic $\tilde{V}_N$, for a time series sample of N observations. The denominator of the statistic is computed as the square root of Newey and West (1987) estimator of the long run variance of the time series. For a maximum lag $q = [N^{1/4}] = 9$, we obtain $\tilde{V}_{N,1}(9) = 5.1582$, $\tilde{V}_{N,3}(9) = 4.5248$ and $\tilde{V}_{N,7}(9) = 4.9893$, with 95% confidence interval $(0, 809,\, 1, 862)$. The tests reject the hypothesis that loadings have short-memory. We also apply Geweke and Porter-Hudak (1983) log-periodogram regression model to estimate the Hurst exponent. The estimates are $H_1^{GPH} = 1.3736$, $H_3^{GPH} = 1.1761$ and $H_7^{GPH} = 1.1433$ for the cutoff $[N^{1/2}] = 50$. The 95% confidence interval (0.2981, 0.7019) for the the GPH estimator is calculated using a bootstrapping procedure proposed by Weron (2002). These estimates imply an order of integration $\hat{\alpha}_r^{GPH} = H_r^{GPH} - 0.5$, $r = 1, 3, 7$. It is known that in the presence of large autoregressive or moving average terms, $\hat{\alpha}_r^{GPH}$ is biased upwards. In general, these models are nontrivial to estimate by other methods. Furthermore, fractionally integrated processes lack a clear economic interpretation. Therefore, instead of including a large number of autoregressive terms we use a parsimonious AR(1) model with time varying coefficients to approximate the long memory process. This is appropriate also for $\alpha \in (1/2,\, 1)$, when the loadings are not stationary, see Comte and Renault (1998).

References

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*, 1203–1227.

Comte, F. and E. Renault (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance 8*, 291–323.

Duan, J. C. and C. Y. Yeh (2010). Jump and volatility risk premiums implied by VIX. *Journal of Economic Dynamics and Control 34*, 2232–2244.

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

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

Geweke, J. and S. Porter-Hudak (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis 4*, 221–238.

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

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

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

Lo, A. W. (1991). Long-term memory in stock market prices. *Econometrica 59*, 1279–1313.

Masry, E. (1996). Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *J. Time Ser. Anal 17*, 571–599.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Ruppert, D. and M. P. Wand (1994). Multivariate locally weighted least squares regression. *Annals of Statistics 22*, 1346–1370.

Weron, R. (2002). Estimating long-range dependence: finite sample properties and confidence intervals. *Physica A: Statistical Mechanics and its Applications 312*, 285–299.

Yao, F., H.-G. Müller, and J.-L. Wang (2005). Data analysis for sparse longitudinal data. *Journal of the American Statistical Association 100*, 577–590.