Functional Data Analysis for Volatility

July 2011

Hans-Georg Müller[1] *University of California, Davis*

Rituparna Sen[2] *University of California, Davis*

Ulrich Stadtmüller[3] *University of Ulm*

Abstract

We introduce a functional volatility process for modeling volatility trajectories for high frequency observations in financial markets and describe functional representations and data-based recovery of the process from repeated observations. A study of its asymptotic properties, as the frequency of observed trades increases, is complemented by simulations and an application to the analysis of intra-day volatility patterns of the S&P 500 index. The proposed volatility model is found to be useful to identify recurring patterns of volatility and for successful prediction of future volatility, through the application of functional regression and prediction techniques.

Keywords: Diffusion model; Functional Principal Component; Functional Regression; High Frequency Trading; Market Returns; Prediction; Volatility Process; Trajectories of Volatility.

Journal of Economics Literature Classification Codes: C14, C51, C52, G12, G17.

[1]Department of Statistics, UC Davis, One Shields Ave, Davis, CA 95616, mueller@wald.ucdavis.edu

[2]Corresponding author. Department of Statistics, UC Davis, One Shields Ave, Davis, CA 95616, rsen@wald.ucdavis.edu

[3]Institut f. Mathematik, Universität Ulm, Helmholtzstr. 18, 89069 Ulm, Germany, ulrich.stadtmueller@uni-ulm.de

1. Introduction

Recently, there has been increasing interest in the modeling of patterns of volatility for high-frequency financial data. An example is provided by intra-day trading patterns in exchange and equity markets which have been analyzed by Andersen and Bollerslev (1997) and Speight, McMillan and Gwilym (2000). In this and similar applications, one is dealing with repeated (daily) observations of patterns of returns that are assumed to be generated by an underlying but unknown stochastic process. The observed daily returns are then associated with realizations of this process. A central problem for the nonparametric analysis of patterns of volatility for financial returns arises from the fact that volatility is not directly observable. Theoretically, volatility could be determined from the quadratic variation of the log price process, if this process were to be observed continuously. However, in practice the log price process is observed only at discrete times when a trade takes place, and volatility must therefore be inferred from discretely observed data on returns. As an example, the observed patterns of returns for six days for the S&P 500 index are displayed in Figure 1.

Volatility is a measure of the extent of variation around a "mean" trajectory. One way of estimating instantaneous volatility is based on the assumption that the volatility process is a function of the observable state variable itself and nonparametric techniques can be applied (Florens-Zmirou, 1993; Bandi and Phillips, 2003; Renò, 2008). Fully nonparametric methods, where volatility is not constrained to be a function of the state variable, include the idea of rolling sample volatility estimators, as in Foster and Nelson (1996); the kernel method in Fan and Wang (2008) and Kristensen (2010); Fourier analysis in Malliavin and Mancino (2009) and Mancino and Sanfelici (2008); and wavelet analysis in Genon-Catalot et al. (1992). Bandi and Renò (2008) estimate spot volatility by differentiation of integrated volatility, while Ogawa and Sanfelici (2008) propose a two-step regularization scheme designed to filter microstructure noise. All of these methods are developed within the traditional framework of volatility modeling for one observed series, corresponding to one realization of the volatility process, and do not target repeated realizations of this process. In contrast, the approach proposed here aims at modeling data and extract

information about volatility for situations where repeated realizations of the volatility process are available, estimating recurring patterns of a suitably defined functional volatility process. A key feature is that the tools of functional data analysis methodology are harnessed for the analysis of financial returns.

We develop a concept of smooth nonparametric volatility trajectories and theory within the framework of a diffusion model for the returns, aiming to extract trajectories of volatility from sequences of returns, such as those shown in Figure 1. We target smooth volatility functions (cf. Fan and Yao, 1998; Kogura, 1996; Stanton, 1997), which differentiates our approach from alternative volatility models with non-smooth paths, such as diffusion models for volatility, e.g., Ornstein-Uhlenbeck processes or the Heston model (cf. Barndorff-Nielsen and Shephard, 2002). Other assumptions include integrability and the existence of second moments. Neither distributional nor parametric model assumptions are required, in contrast to commonly used time series models for volatility. Our approach is nonparametric in spirit; we refer to Fan and Yao (2003) and Fan (2005) for excellent overviews of nonparametric approaches to volatility modeling. We demonstrate the usefulness of the functional volatility process in the setting of intra-day trading application examples, in terms of both modeling and predicting volatility. In financial data analysis, the issue of modeling intra-day high-frequency data, especially for the case where the spacing of the observations tends to 0 in the limit, has been much discussed in recent years (Aït-Sahalia, 1996; Aït-Sahalia and Mykland, 2003; Aït-Sahalia, Mykland and Zhang, 2005; Aït-Sahalia, Mykland and Zhang, 2009; Fan, Jiang, Zhang and Zhou, 2003; Fan, 2005; Zhang, Mykland and Aït-Sahalia, 2005).

The proposed characterization of the functional volatility process is adapted to situations in which series of returns are repeatedly observed, as illustrated in Figure 1, and where the observed trajectories may be viewed as repeated realizations of an underlying unknown functional volatility process. For modeling functional volatility processes, we develop an extension of the concept of a functional variance error model developed in Müller, Stadtmüller and Yao (2006). The extension of this earlier simpler approach to a volatility model requires additional substantial theoretical developments that are provided in this paper. A precursor of these models are variance and noise estimation approaches that have been

developed within the framework of nonparametric regression models over the past decades (e.g., Eubank and Thomas, 1993; Yao and Tong, 2000; Wang et al., 2008). The present approach is novel in the sense that the problem of analyzing volatility from repeated realizations of the functional volatility process has not yet been explicitly studied, and therefore models that specifically address the problem of extracting information from a sample of volatility processes have not yet been considered. It intuitively makes sense to extract information about patterns of volatility from samples of repeatedly observed volatility paths. As we demonstrate in this paper, this idea can be formalized by developing a connection to the methodology of functional data analysis (FDA).

An important tool for the analysis of trajectories of volatility within the framework of FDA is functional principal component analysis (Castro, Lawton and Sylvestre, 1986; Rice and Silverman, 1991). The proposed functional volatility processes can be characterized by their mean function and the eigenfunctions of the autocovariance operator. This is a consequence of the Karhunen-Loève representation of the functional volatility process. For data analysis based on returns, we develop suitable estimates for the components of this representation. Individual trajectories of volatility may be represented by their functional principal component scores, which can serve as input for subsequent statistical analysis. An application that we explore in more detail below is functional regression (Ramsay and Dalzell, 1991) for the prediction of future volatility. The fact that a sizable fraction of future volatility can be predicted with the proposed method supports this approach. Background on FDA can be found in Ramsay and Silverman (2005). For other applications of FDA to economic time series, we refer to Malfait and Ramsay (2003) and Ramsay and Ramsey (2001).

The paper is organized as follows: Background on diffusion processes for financial returns and the relationship to volatility and repeated realizations of the volatility process is discussed in Section 2. Auxiliary asymptotic approximations in the Appendix, derived in the framework of a diffusion process for the returns, guarantee that the leading terms of suitably standardized differences of log closing prices in high frequency settings can be decomposed into a product of a volatility term and a noise term. This

justifies the definition of the functional volatility process which is obtained by transforming these terms and thus decoupling the noise from the volatility signal. The availability of repeated realizations of the functional volatility process makes it possible to obtain consistent estimation procedures for the main components of this process (mean and eigenfunctions). Functional principal components to represent individual realizations of the functional volatility process and functional regression for volatility are introduced in Section 3, where also asymptotic results on the convergence of these estimates are derived, including a result on consistency of functional regression for volatility processes. In Section 4 we describe simulation results for various volatility models. The implementation of the methodology to recover functional volatility processes from financial data and an application to high-frequency intra-day trading data of the S&P 500 index is described in Section 5, while Section 6 contains a discussion of some salient features of the proposed method and concluding remarks.

2. The Functional Volatility Process

2.1 Background on diffusion models for volatility

For equity prices $X(t)$, as well as related market prices, the by now classical continuous time model for returns is the diffusion equation (Black and Scholes, 1973)

$$\frac{dX(t)}{X(t)} = \mu\, dt + \sigma\, dW(t)\,,\ t \geq 0\,, \tag{1}$$

where $W(\cdot)$ denotes a standard Wiener process, $\sigma > 0$ is the volatility and μ a drift term. Both volatility and drift term are time-independent in this model. This model is a simplification, which does not reflect the patterns observed in real data. Furthermore, these processes are not observed in the continuum, but rather on a discrete grid of time points at which trades take place or are recorded, which may be randomly or regularly spaced. In accordance with daily trading data, which motivate our approaches, we assume that data are observed on a regular grid of times

$$t_j = j\Delta, \quad j = 1, \ldots, [T/\Delta], \tag{2}$$

where the overall time interval on which processes are observed is $[0, T]$, and $[r]$ stands for the largest integer smaller or equal to $r \geq 0$. The motivating daily trading data are recorded on a regular grid with $\Delta = 5$min. Generally, data with small Δ are referred to as high-frequency financial data.

A class of diffusion models that has been developed specifically for high-frequency trading data such as intra-day trading are the stochastic volatility models of Barndorff-Nielsen and Shephard (2002). These models consist of two equations, one governing the log returns according to

$$d \log X(t, \omega) = \mu \, dt \; + \; \beta \sigma^2(t, \omega) \, dt \; + \; \sigma(t, \omega) \, dW(t, \omega), \tag{3}$$

where we include the arguments ω in the probability space Ω to emphasize the stochastic parts, and a second diffusion equation for the volatility σ. The model also includes a non-random drift term μ and the so-called risk premium β. Typically, σ is assumed to be a stationary predictable process, the so-called spot volatility, which for example has been modeled as an Ornstein-Uhlenbeck process.

Previous studies for the volatility of high frequency trading data focused either on a diffusion model for volatility with its associated non-smooth volatility processes, a GARCH type or other time series model, or alternatively, the nonparametric estimation of smooth but essentially non-random trajectories of volatility. In addition, wavelets allowing for jump discontinuities have also been considered (e.g., Fan and Wang, 2007). Our approach retains both the random nature of volatility trajectories and the nonparametric flavor by modeling volatility as a smooth random process that otherwise is not governed by a known or assumed equation of any kind. The assumptions for this process ((M1)-(M5) in the Appendix) include smoothness, but otherwise are minimal. That we still can study the volatility process under weak assumptions is due to the functional paradigm and the availability of a sample of i.i.d. realizations of the process, which is a crucial ingredient for the proposed method.

2.2 An asymptotic volatility model for repeated realizations

We consider a variant of model (3), where within the framework of a general diffusion model with random drift function, observations are available for a sample of n realizations of the underlying processes,

$$d \log X_i(t, \omega) = \widetilde{\mu}_i(t, \omega) \, dt + \widetilde{\sigma}_i(t, \omega) \, dW_i(t, \omega), \quad 0 \leq t \leq T, \quad i = 1, \ldots, n. \tag{4}$$

Here $\widetilde{\mu}_i(t,\omega)$, $\widetilde{\sigma}_i(t,\omega)$ are i.i.d. copies of the stochastic processes $\widetilde{\mu}$ and $\widetilde{\sigma}$, which are assumed to be smooth but not stationary, and W_i, $i = 1,\ldots,n$, are n independent standard Wiener processes. The availability of multiple copies is crucial for the application of functional data analysis methodology and sets our approach apart from other volatility models. In the following, we will not indicate dependency on $\omega \in \Omega$ and also omit indices i, focusing on a generic realization of the underlying processes, while keeping in mind that one observes n copies of these processes.

As our focus is on the high-frequency case, the relatively dense time grid that underlies such high-frequency data motivates the asymptotic assumption $\Delta \to 0$, reflecting increasingly frequent trading. Importantly, one does not actually observe continuous data and it is therefore necessary to consider discretized versions, defining scaled log-returns and associated diffusion terms

$$\begin{aligned} Z_\Delta(t) &= \frac{1}{\sqrt{\Delta}} \log\left(\frac{X(t+\Delta)}{X(t)}\right), \\ W_\Delta(t) &= \frac{1}{\sqrt{\Delta}} (W(t+\Delta) - W(t)). \end{aligned}$$

We rewrite model (4) for the actual high-frequency observations $Z_\Delta(t)$ as follows,

$$\begin{aligned} Z_\Delta(t) &= \frac{1}{\sqrt{\Delta}} \int_t^{t+\Delta} \widetilde{\mu}(v)\, dv + \frac{1}{\sqrt{\Delta}} \int_t^{t+\Delta} \widetilde{\sigma}(v)\, dW(v) \qquad (5) \\ &= \widetilde{\mu}(t)\sqrt{\Delta} + \widetilde{\sigma}(t) W_\Delta(t) \,+\, R_1(t,\Delta) + R_2(t,\Delta). \end{aligned}$$

The remainder terms R_1, R_2 reflect the discretization step,

$$\begin{aligned} R_1(t,\Delta) &= \frac{1}{\sqrt{\Delta}} \int_t^{t+\Delta} \widetilde{\mu}(v)\, dv - \widetilde{\mu}(t)\sqrt{\Delta} \qquad (6) \\ R_2(t,\Delta) &= \frac{1}{\sqrt{\Delta}} \int_t^{t+\Delta} \widetilde{\sigma}(v)\, dW(v) - \widetilde{\sigma}(t) W_\Delta(t). \end{aligned}$$

According to Lemma 1 in the Appendix, under suitable regularity assumptions, these remainder terms are uniformly small and therefore may be neglected asymptotically. For small Δ one arrives at the approximate model

$$Z_\Delta(t) \approx \widetilde{\mu}(t)\sqrt{\Delta} \,+\, \widetilde{\sigma}(t)\, W_\Delta(t). \qquad (7)$$

Noting that $\sup_{t\in[0,T]} |\widetilde{\mu}(t)| = O_P(1)$, providing uniform boundedness in t, we find that the first term is uniformly $O_P(\sqrt{\Delta})$, and therefore is also negligible. This leads to the approximation

$$Z_\Delta(t) \approx \widetilde{\sigma}(t)\, W_\Delta(t). \tag{8}$$

For related results and background we refer to Jacod and Shiryaev (2003) and Barndorff-Nielsen et al. (2006).

2.3 Trajectories of volatility

The empirical observations $Z_\Delta(t_j)$ target the processes $\widetilde{\sigma}(t_j)\, W_\Delta(t_j)$, where trades are assumed to be recorded on a dense discrete time grid $t_j = j\Delta, \quad j = 1, 2, ..., [T/\Delta]$, as in (2). Our target for inference is the smooth process V defined by

$$V(t) = \log(\{\widetilde{\sigma}(t)\}^2). \tag{9}$$

We refer to V as *functional volatility process*. It is related to the observations $Z_\Delta(t_j)$ by

$$\log(\{Z_\Delta(t_j)\}^2) - q_0 \approx Y_\Delta(t_j) = V(t_j) + U_\Delta(t_j), \tag{10}$$

where q_0 is a numerical constant and $Y_\Delta(t), U_\Delta(t)$ are stochastic processes, defined as follows,

$$\begin{aligned} q_0 &= E(\log W_\Delta^2(t)) = \frac{4}{\sqrt{2\pi}} \int_0^\infty \log(x) e^{-x^2/2}\, dx \approx -1.27, \\ Y_\Delta(t) &= \log(\{\widetilde{\sigma}(t) W_\Delta(t)\}^2) - q_0, \\ U_\Delta(t) &= \log(\{W_\Delta(t)\}^2) - q_0. \end{aligned} \tag{11}$$

We note that the adjustment by the constant q_0 has the consequence that $EU_\Delta(t) = 0$ for all t, while $\mathrm{cov}(U_\Delta(s), U_\Delta(t)) = 0$ for $|t - s| > \Delta$ (independent increments property). We also remark that the functional volatility process V does not depend on Δ.

Since our focus is on spot volatility, we do not encounter problems with accumulation of noise. Such problems are known to be critical when estimating integrated volatility with measures such as realized variance. In that case, one would be summing numerous contaminated squared return data. As $\Delta \to 0$,

the number of summands grows to infinity, and as each summand is contaminated by an error with constant variance, the sum would diverge.

We represent the smooth functional volatility process V in terms of its decomposition into functional principal components, a common approach in FDA. For a domain $\mathcal{T}$, setting

$$G_V(s,t) = \mathrm{cov}(V(s), V(t)), \quad E(V(t)) = \mu_V(t), \quad s,t \in \mathcal{T}, \tag{12}$$

the functional principal components are the eigenfunctions of the auto-covariance operator of V, a linear operator on the space L^2 that is given by

$$\boldsymbol{G}_V(f)(s) = \int_{\mathcal{T}} G_V(s,t) f(t)\, dt.$$

We assume the orthonormal eigenfunctions of this operator are ϕ_k, with associated eigenvalues λ_k, $k = 1, 2, \ldots$, such that $\lambda_1 \geq \lambda_2 \geq \cdots$ and $\sum_k \lambda_k < \infty$. The Karhunen-Loève theorem then provides a representation of individual random trajectories of the functional volatility process V, given by

$$V(t) = \mu_V(t) + \sum_{k=1}^{\infty} \xi_k \phi_k(t), \tag{13}$$

where the ξ_k are uncorrelated random variables that satisfy

$$\xi_k = \int (V(t) - \mu_V(t))\phi_k(t)\, dt, \quad E\xi_k = 0, \quad \mathrm{var}(\xi_k) = \lambda_k. \tag{14}$$

It is a consequence of Lemma 2 in the Appendix that under suitable regularity conditions regarding the dependence between $\widetilde{\sigma}$ and W, one has

$$E(Y_\Delta(t)) = \mu_V(t), \quad \mathrm{cov}(Y_\Delta(s), Y_\Delta(t)) = O(\sqrt{\Delta}) + G_V(s,t). \tag{15}$$

Equation (10) suggests that the smooth mean function μ_V and the smooth covariance surface G_V can be consistently estimated from available data, and we demonstrate in the following section that this is indeed correct. Once estimates for functions μ_V and G_V have been obtained, well-known procedures exist to infer eigenfunctions and eigenvalues (Rice and Silverman, 1991; Müller et al., 2006). Processes V are then approximated by substituting estimates and using a judiciously chosen finite number of terms in the sum for representation (13).

3. Inferring the Functional Volatility Process

3.1 Estimation procedures

A central goal is to identify the stochastic structure of the underlying functional volatility process $V(t) = \log(\{\widetilde{\sigma}(t)\}^2)$. Key components include the mean volatility trajectory μ_V, which reflects overall trends of volatility, and the eigenfunctions ϕ_k, $k \geq 1$, of the process V, representing *modes of volatility* and indicating the patterns of variation among individual volatility trajectories (cf. Castro et al., 1986). A second goal is to estimate the functional principal component (FPC) scores for the individual trajectories, which then allows to represent such trajectories in the eigenfunction expansion (13) and therefore to obtain estimates of predicted individual trajectories. Furthermore, and perhaps more importantly, the estimated FPC scores can be used for subsequent statistical analysis.

For simplicity, we assume that the grid of observation times t_j , see (2), for trades $X_i(t_j)$ is the same for all price trajectories $X_1, \ldots, X_n$; it is possible to relax this assumption. The price trajectories, from which the observations $X_i(t_j)$ are derived, are assumed to be an i.i.d. sample from the price process X. The high frequency observations on differences of log-transformed closing prices that form the basis of volatility analysis are then

$$Z_{ij\Delta} = \frac{1}{\sqrt{\Delta}} \log\left(\frac{X_i(t_j + \Delta)}{X_i(t_j)}\right), \ i = 1, \ldots, n, \ j = 1, \ldots, [\frac{T}{\Delta}].$$

From these, we form transformed and adjusted data,

$$Y_{ij\Delta} = \log(Z^2_{ij\Delta}) - q_0, \tag{16}$$

with q_0 as defined in (11).

At the core of the estimation procedure is the principal analysis of random trajectories (PART), applied to the data $Y_{ij\Delta}$, which is an algorithm to obtain mean and eigenfunctions, as well as FPC scores, from densely sampled functional data, as described in Müller et al. (2006). The smoothing steps in this algorithm are implemented with weighted local linear smoothing (Fan and Gijbels, 1996), which works well in practice; alternative smoothing methods can also be used. In order to estimate the overall mean

function μ_V, we pool all available data into one big scatterplot $\{(t_j, Y_{ij\Delta}),\ i = 1, \ldots, n,\ j = 1, \ldots, [T/\Delta]\}$, and then obtain the nonparametric regression of Y versus t by local linear smoothing. Formally, one finds the minimizers $\hat{\beta}_0$, $\hat{\beta}_1$ of

$$\sum_{i=1}^{n} \sum_{j=1}^{[T/\Delta]} \kappa_1(\frac{t_j - t}{b_V})\{Y_{ij\Delta} - \beta_0 - \beta_1(t - t_j)\}^2, \tag{17}$$

where b_V is the smoothing bandwidth, chosen in practice by (generalized) cross-validation, and κ_1 is a kernel function, which is required to be a square integrable and compactly supported symmetric density function, with absolutely integrable Fourier transform. Then one sets $\hat{\mu}_V(t) = \hat{\beta}_0(t)$.

Estimation of the covariance surface G_V is based on the collection of all available pairwise "empirical covariances" $G_i(t_{j_1}, t_{j_2}) = (Y_{ij_1\Delta} - \hat{\mu}_V(t_{j_1}))(Y_{ij_2\Delta} - \hat{\mu}_V(t_{j_2}))$, assembling these into a two-dimensional scatterplot $\{[(t_{j_1}, t_{j_2}), G_i(t_{j_1}, t_{j_2})],\ i = 1, \ldots, n,\ j_1, j_2 = 1, \ldots, [T/\Delta]\}$, and fitting a two-dimensional smoother to obtain the nonparametric regression of $G_i(t_{j_1}, t_{j_2})$ versus (t_{j_1}, t_{j_2}). Formally, one minimizes

$$\sum_{i=1}^{n} \sum_{1 \le j_1 \ne j_2 \le [T/\Delta]} \kappa_2(\frac{t_{j_1} - s}{h_V}, \frac{t_{j_2} - t}{h_V})\{G_i(t_{j_1}, t_{j_2}) - [\beta_0 + \beta_1(s - t_{j_1}) + \beta_2(t - t_{j_2})]\}^2 \tag{18}$$

w.r. to $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ and defines $\widehat{G}_V(s,t) = \hat{\beta}_0(s,t)$. In (18), the kernel κ_2 is a compactly supported and square integrable bivariate kernel function, which is a density with mean zero and finite variance that possesses an absolutely integrable Fourier transform. The smoothing bandwidths h_V can again be chosen by (generalized) cross-validation.

We note that the diagonal terms (j_1, j_2), $j_1 = j_2$, are missing in the summation over j_1, j_2 in (18). This omission is motivated by the dependence structure of the targets $Y_\Delta(t)$ of the transformed volatility observations $Y_{ij\Delta}$, as given in (16). We note that due to the fact that $|t_{j_1} - t_{j_2}| \ge \Delta$ whenever $j_1 \ne j_2$, excluding the diagonal terms suffices to keep the "empirical covariances" on target. Due to the assumed smoothness of the covariance surface G, the diagonal on the other hand is not essential in the surface estimation step, and can be omitted from the data that are used to construct the surface, without

incurring any asymptotic penalty. We note that along the diagonal,

$$\begin{aligned} \mathrm{var}(Y_{ij\Delta}) &= G_V(t_j,t_j) + \mathrm{var}(\log(W_\Delta(t))^2) + \mathrm{cov}(\log\widetilde{\sigma}^2(t),\ \log W_\Delta^2(t)) \\ &= G_V(t_j,t_j) + \left\{ \frac{8}{\sqrt{2\pi}} \int_0^\infty (\log(x))^2 \exp\{-\frac{x^2}{2}\}\, dx \ - \ q_0^2 \right\}, \end{aligned} \tag{19}$$

where the second equality holds under the assumptions of Lemma 2.

Once mean and covariance functions of functional volatility processes V have been determined, a next step is the estimation of the (eigenvalue/eigenfunction) pairs, which are defined as the solutions of the eigen-equations $\int G_V(s,t)\phi_k(s)ds = \lambda_k\phi_k(t)$, substituting the estimated covariance surface $\hat{G}_V$ for G_V. Solutions $(\hat{\lambda}_k,\ \hat{\phi}_k)$ are obtained by numerical eigen-analysis, based on an initial discretization step, under orthonormality constraints for the eigenfunctions. Positive definiteness of the corresponding covariance surface can be guaranteed by a projection of the initial estimate $\hat{G}$ on a positive definite version $\tilde{G}$, as described in Yao et al. (2003). In a last step, the PART algorithm yields estimates of the individual FPC scores. Motivated by (14), these are implemented as

$$\hat{\xi}_{ik} \ = \ \Delta \sum_{j=2}^{[T/\Delta]} (Y_{ij\Delta} - \hat{\mu}_V(t_j))\hat{\phi}_k(t_j), \quad i = 1,\ldots,n,\ k = 1,2,\ldots. \tag{20}$$

Individual trajectories of volatility can then be represented by an empirical version of the Karhunen-Loève expansion (13),

$$\hat{V}_i(t) = \hat{\mu}_V(t) \ + \sum_{k=1}^{K} \hat{\xi}_{ik}\hat{\phi}_k(t). \tag{21}$$

One choice that has to be made is the selection of the number K of included components. Options include one-curve-leave-out cross-validation (Rice and Silverman 1991), pseudo-AIC criteria (Yao et al. 2005a) or a scree plot, a tool from multivariate analysis, where one uses estimated eigenvalues to obtain a pre-specified fraction of variance explained as a function of K or looks for a change-point.

3.2 Functional regression for trajectories of volatility

It is of practical interest to predict future volatility patterns from observed trajectories. In intra-day trading, one may have observed the first half day of trading and then, based on these data, may wish to

predict the trajectory of volatility for the remainder of the day. This prediction naturally will take into account the general relationship of volatility trajectories between first and second half day, and especially the observed patterns in the morning of the given day, for which one wishes to predict the afternoon volatility. We explore this example further in the next section. Functional regression models, where both predictors and responses are random functions, have been introduced by Ramsay and Dalzell (1991). While there exists a large body of literature on functional regression, considerably less work has been done for functional responses (e.g., Faraway 1997), the case we contemplate here.

In a function to function regression setting, one observes a sample of i.i.d. pairs of random trajectories (Q_i, R_i), $i = 1, \ldots, n$, sampled from square integrable stochastic processes (Q, R), with mean functions μ_Q, μ_R and eigenvalue, eigenfunction pairs $\{(\lambda_k, \phi_k),\ k = 1, 2, \ldots\}$ for random predictor functions Q and $\{(\tau_m, \psi_m),\ m = 1, 2, \ldots\}$, say, for random response functions R. The functional linear regression model

$$E(R(t)|Q) = \mu_R(t) + \int_{\mathcal{T}_Q} (Q(s) - \mu_Q(s))\beta(s,t)\, ds, \quad t \in \mathcal{T}_R, \tag{22}$$

features the regression parameter surface β, a nonparametric smooth function. We note that the domains $\mathcal{T}_Q$ of Q and $\mathcal{T}_R$ of R can be arbitrary, but predictor and response function are always drawn from the same joint realization of processes Q_i and R_i, which have a joint distribution. For example, in the intra-day volatility example that we study in Section 4, both predictor and response processes are jointly observed on the same day.

The estimation of β is an inverse problem and requires regularization. We use the projection of the trajectories on FPC scores, regularizing by controlling the number of included components. With Karhunen-Loève expansions for Q and R,

$$Q(t) = \mu_Q(t) + \sum_{k=1}^{\infty} \xi_k \phi_k(t),\ t \in \mathcal{T}_Q\,, \quad R(t) = \mu_R(t) + \sum_{m=1}^{\infty} \zeta_m \psi_m(t),\ t \in \mathcal{T}_R\,, \tag{23}$$

where ξ_k are FPC scores for Q and ζ_m for R, it is well known that the functional regression relation (22) can be alternatively written as

$$R^*(t|Q) = E(R(t)|Q) = \mu_R(t) + \sum_{k,m=1}^{\infty} \xi_k \frac{\gamma_{km}}{\lambda_k} \psi_m(t)\,,\ t \in \mathcal{T}_R\,. \tag{24}$$

Here $\gamma_{km} = E(\xi_k \zeta_m)$, which can be estimated in various ways (Müller, Chiou and Leng 2008). The functional regression can be decomposed into a series of simple linear regressions through the origin of response scores ζ_m versus predictor scores ξ_k.

The estimated prediction of a new trajectory R^*, given any trajectory Q^* with FPC scores ξ_k^*, is

$$\widehat{R}^*(t|Q) = \hat{\mu}_R(t) \; + \; \sum_{k,m=1}^{\infty} \xi_k^* \frac{\hat{\gamma}_{km}}{\hat{\lambda}_k} \, \hat{\psi}_m(t), \; t \in \mathcal{T}_R \, . \tag{25}$$

To obtain estimates $\hat{\gamma}_{km}$ of γ_{km}, we adopt a functional regression method described in Yao et al. (2005), which is based on the fact that

$$\gamma_{km} = \int_{\mathcal{T}_\mathcal{Q}} \int_{\mathcal{T}_\mathcal{R}} \phi_k(s) \psi_m(t) G_{QR}(s,t) \, ds \, dt, \tag{26}$$

where $G_{QR}(s,t) = \mathrm{cov}(Q(s), R(t))$. The cross-covariance surface can be estimated by similar procedures as used for the auto-covariance function G in (18), where we designate the bandwidth used in the two-dimensional smoothing step of the raw cross-covariances by $\tilde{h}$. Estimates $\hat{\gamma}_{km}$ are then obtained by plugging the resulting estimates $\hat{G}_{QR}(s,t)$ and the previously determined estimates $\hat{\phi}_k$ and $\hat{\psi}_m$ of the eigenfunctions of processes Q and R into (26).

The regression parameter surface $\beta(s,t)$ appearing in (22) is customarily represented as

$$\beta(s,t) = \sum_{k,m=1}^{\infty} \frac{\gamma_{km}}{\lambda_k} \phi_k(s) \psi_m(t) \, , \; s \in \mathcal{T}_Q, \; t \in \mathcal{T}_R, \tag{27}$$

and an estimate $\hat{\beta}(s,t)$ is obtained simply by plugging in estimates for the unknown quantities. When one is interested to quantify the strength of a functional relationship, the functional coefficient of determination is useful. With eigenvalues τ_m for processes R, it is defined as

$$R^2_{\mathrm{det}} = \frac{\int \mathrm{var}(E[R(t)|Q])dt}{\int \mathrm{var}(R(t))dt} = \frac{\sum_{k,m=1}^{\infty} \gamma_{km}^2 / \lambda_k}{\sum_{m=1}^{\infty} \tau_m}, \tag{28}$$

and estimation is straightforward by replacing the unknown quantities by their respective estimates. We illustrate these estimates in Section 4.

3.3 Asymptotics

Consistency results for eigenfunctions, eigenvalues, FPC scores and fitted trajectories for estimation of volatility processes and for regressing volatility trajectories on each other are the topic of this section. The starting point for the asymptotic analysis is model (4) for the returns. As this model differs in various ways from usual functional models, the situation does not conform with the usual functional approaches and additional asymptotic considerations are required. A first step is a bound that demonstrates the uniform approximation of the observations $Z_{ij\Delta}$ in (5) and (6). Proofs and details regarding the assumptions can be found in the Appendix. In the following, $\mathcal{T} \subset (0, T]$ is an arbitrary compact interval.

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

$$E[\sup_{t\in\mathcal{T}} |Z_\Delta - \widetilde{\sigma}(t)W_\Delta(t)|] = O(\sqrt{\Delta}). \tag{29}$$

Due to the $1/\sqrt{\Delta}$ normalization used for Z_Δ, the mean price trajectory has the same order of convergence as the remainder term and therefore does not contribute to the leading term. Recollecting that we estimate the overall mean trajectory μ_V of volatility processes V in (17) with bandwidths b_V, and the covariance surface G_V (12) in (18) with bandwidths h_V, we obtain for the estimation of these key constituents the following result. All convergence results in the following are for $n \to \infty$ and $\Delta \to 0$.

Theorem 2. Assuming (M1)-(M8),

$$\sup_{t\in\mathcal{T}} |\hat{\mu}_V(t) - \mu_V(t)| = O_P(\sqrt{\Delta} + (n^{1/2}b_V)^{-1}), \tag{30}$$

$$\sup_{s,t\in\mathcal{T}} |\widehat{G}_V(s,t) - G_V(s,t)| = O_P(\sqrt{\Delta} + (n^{1/2}h_V^2)^{-1}). \tag{31}$$

This result provides justification for the mean and covariance function estimates. As a consequence of Theorem 2, we also obtain consistency for the estimation of eigenvalues λ_k and eigenfunctions ϕ_k of volatility processes, justifying the use of these estimates in volatility analysis.

Theorem 3. Assume (M1)-(M9) for a sequence $K = K(n) \to \infty$. Then for all $1 \le k \le K$ and

eigenvalues λ_k of multiplicity one, eigenfunctions $\hat{\phi}_k$ can be chosen such that, uniformly over $1 \leq k \leq K$,

$$\sup_{t \in \mathcal{T}} |\hat{\phi}_k(t) - \phi_k(t)| \xrightarrow{P} 0, \qquad \hat{\lambda}_k \xrightarrow{P} \lambda_k. \tag{32}$$

We are also interested in the consistency of estimates $\hat{V}_i(t) = \hat{\mu}_V(t) + \sum_{k=1}^{K} \hat{\xi}_{ik}\hat{\phi}_k(t)$ (21) of individual volatility trajectories $V_i(t) = \mu_V(t) + \sum_{k=1}^{\infty} \xi_k \phi_k(t)$, as estimating volatility from available trading data is a major objective.

Theorem 4. Assuming (M1)-(M11), as $K = K(n) \longrightarrow \infty$,

$$\sup_{t \in \mathcal{T}} |\widehat{V}_i(t) - V_i(t)| \xrightarrow{P} 0. \tag{33}$$

An immediate consequence of this result is that also the function $\widetilde{\sigma}^2(t)$ that appears in (9) can be consistently estimated by $\widetilde{\sigma}_i^2(t) = \exp(\hat{V}_i(t))$. Finally, we analyze the prediction of new volatility trajectories from available trajectories via functional regression for volatility processes. Assuming that K components are included for predictor processes and M for response processes, the estimated prediction via functional regression $\widehat{E}(R(t)|Q^*)$ (25) is consistent for its target $E(R(t)|Q^*)$ (24), as the following result shows.

Theorem 5. Assume (M1)-(M11) for predictor processes Q with $K = K(n) \to \infty$, and also for response processes R with M substituting for K, $M = M(n) \to \infty$; furthermore that (M12) holds for eigenfunctions ϕ_k of Q and for eigenfunctions ψ_k of R. In addition, the bandwidths $\tilde{h}$ that are used for the auxiliary cross-covariance surface estimate $\hat{G}_{QR}$ (as introduced in Section 3.2 at (26)) are assumed to satisfy (M13). Then the estimate $\hat{\beta}$ of the regression parameter surface β (27) satisfies

$$\sup_{(s,t) \in \mathcal{T}_Q \times \mathcal{T}_R} |\hat{\beta}(s,t) - \beta(s,t)| \xrightarrow{P} 0, \tag{34}$$

and the estimate $\widehat{R}^*(t|Q)$ (25) of the predicted response trajectory $R^*(t|Q) = E(R(t)|Q)$ (24), evaluated at predictor trajectory $Q = Q^*$, satisfies

$$\sup_{t \in \mathcal{T}_R} |\widehat{R}^*(t|Q^*) - R^*(t|Q^*)| \xrightarrow{P} 0. \tag{35}$$

This result provides asymptotic justification for predicting volatility trajectories for high-frequency data from observations of related trajectories.

4. Simulations

4.1 Heston and smooth volatility models

We compared the behavior of the proposed functional approach with standard methods for spot volatility estimation for simulated price curves. This comparison was carried out for both the traditional Heston model with non-smooth volatility function and a smooth volatility model that is in accordance with assumption (M5) in the Appendix.

Heston Model. Here the diffusion for the price process is given by

$$d\log X(t) = \mu X(t)dt + \sqrt{V(t)}X(t)dW_1(t) \tag{36}$$

for a Wiener process W_1. The stochastic volatility function V follows the Cox-Ingersoll-Ross square root diffusion model

$$dV(t) = a(b - V(t))dt + c\sqrt{V(t)}dW_2(t), \tag{37}$$

with constants a, b, c, where W_2 is a second Wiener process and W_1 and W_2 have correlation ρ. The non-differentiability of the resulting volatility process means that assumption (M5) is not satisfied for the Heston model. For the simulation study, we generated data from a Heston model with $X_0 = 20, V_0 = 1, \mu = 0.1, a = 0.5, b = 1, c = 1$, and two levels for the correlation, $\rho = 0.2$ and $\rho = 0.8$. The parameters were chosen so that both the price process and volatility process are positive with a high probability.

Smooth Volatility Process Model. For the simulation study, we adopted the following specific example for smooth volatility processes $V(t) = \log(\{\widetilde{\sigma}(t)\}^2)$ that satisfy assumption (M5) in the Appendix: For $\alpha \geq 0$, $\delta > 0$, let

$$V(t) = \log\left[\int_{t-\delta}^{t} \kappa(t-v)\, dW(v) \,+\, b\right]^2 + \alpha \log\left[S(t)\right]^2, \quad t \in [0,1], \tag{38}$$

where we chose the kernel $\kappa(x) = x^3(1-x)^3$ and $S(t) = \xi_1\phi_1(t) + \xi_2\phi_2(t) + b$. Here, ξ_1, ξ_2 are independent Gaussian random variables with mean zero and variance 1 and 0.5 respectively, which are also independent of W, with orthonormal functions $\phi_1(t) = \frac{1}{\sqrt{2}}\sin(k\pi t)$, $\phi_2(t) = \frac{1}{\sqrt{2}}\cos(k\pi t)$. It is easy to see that these

processes satisfy (M5). By changing α, or increasing the number of components included in the smooth processes S, this construction reflects various degrees of correlatedness between V and W, similarly to the situation in the Heston model.

As in the Heston model, we chose $X_0 = 20, V_0 = 1, \mu = 0.1, b = 1$ in model (38), as well as $\delta = 1/78$, corresponding to the fraction of 5 minutes relative to the trading day. This parameter controls the degree of smoothing of the Wiener process that is reflected in (38), while α controls the relative contributions of the two terms on the right hand side of (38), or more geneally of $\tilde{W}(t)$ and $H(t)$ in (M5), to the total volatility, while k determines the period of the smooth component. We carried out simulations for two cases, *Scenario 1* with $\alpha = 0.05$ and $k = 1$, and *Scenario 2* with $\alpha = 0.02$ and $k = 2$.

We adopted the sample size and the number of trades generated in the simulation to the number of days ($n = 92$) and to the 78 daily trades, corresponding to 5 minute intervals, that were actually observed in the application to financial data, which is described in more detail in Section 5. Accordingly, we generated for all simulations 92 independent functional realizations from both Heston and smooth volatility models, and for each random trajectory 78 trades were generated, for a total of 100 Monte Carlo runs. Wiener processes were approximated with 300 increments for each observation, so that the step size corresponds to 1 second.

4.2 Simulation results

For comparing the behavior of estimates for the functional volatility process, in addition to the proposed functional methods, we implemented the kernel approach of Fan and Wang (2008). In this approach the spot volatility is estimated using a kernel-smoothed version of squared log returns of the high frequency data. We used a two-sided exponential kernel and chose optimal bandwidths by minimizing the L_2 distance between estimates and the actual volatility curves. The simulation results are reported in terms of $E[\int(\hat{V}(t) - V(t))^2]^{1/2}$, the mean root integrated squared error (MRISE) of curve estimates $\hat{V}(t)$, which we approximated by taking the average of $\frac{1}{92}\sum_{i=1}^{92}[\frac{1}{78}\sum_{j=1}^{78}(\hat{V}_i(t_j) - V_i(t_j))^2]^{1/2}$ over 100 simulation runs.

When simulating data from the Heston model, for the case with correlation $\rho = 0.2$, MRISE was

found to be 0.35 (standard deviation 0.02) for the proposed functional method and 0.31 (0.02) for the kernel method. For a Heston model in the case where $\rho = 0.8$, MRISE was 0.36 (0.02) for the functional and 0.31 (0.02) for the kernel method, respectively. For data generated from the smooth volatility process model, in Scenario 1, MRISE was 0.16 (0.02) for the proposed functional and 0.23 (0.01) for the kernel method. In Scenario 2, the corresponding values were 0.14 (0.02) and 0.23 (0.01), respectively. The functional approach is found to perform substantially better than an established kernel method under smooth volatility models, while its performance is slightly worse under the Heston model.

Typical behaviors of predicted volatility trajectories, using functional predictors as in (25), are illustrated in Figure 2. Here the prediction is for the volatility trajectory V of the second half day, based on the data observed first half day for the daily volatilities generated as described above. More details about predicting second half day volatility from observed data for the first half day can be found in the following section. This forecast is shown for two randomly selected curves and the two Heston models as well as the two smooth volatility models described above. A simple forecast based on the value of Z_Δ at the end of the first-half day is also shown as horizontal line, as are the actual observed volatilities Z_Δ for the second day, which are not used for constructing the predictions. We find that the predicted volatility process approximately tracks the observed patterns of volatility, with improved tracking for smooth volatility processes, while the simple extrapolation forecast produces very poor predictions.

5. Functional Data Analysis Of Intra-Day Trading Patterns

5.1 Description of data and initial processing

The data consist of returns in high-frequency trading of the Standard and Poor 500 index, based on closing prices that are recorded every 5 minutes for the period December 2003 to March 2006, comprising 548 trading days. Here $\Delta = 5$ minutes, which is the unit of time in the graphs. The length of the daily trading period is 390 minutes. Our use of five-minute sampling frequency parallels many other studies in the literature, and use of the closing prices is in accordance with the tick-method as discussed in

Dacorogna et al. (2001). Figure 1 displays daily trajectories of log returns for these data, observed on randomly selected six days.

While volatility patterns for consecutive days may not be independently distributed, it is expected that this dependence vanishes as days are spaced sufficiently apart from each other. To enhance independence between the trading days entering our analysis, we randomly selected a subsample of days, consisting of pairs of days satisfying the following constraints: (a) the time distance in between days that belong to different pairs is at least a month; (b) the interval between the days in each pair is six days. Constraint (b) resulted from a computation of the functional coefficient of determination R^2_{det} (28), when regressing the volatility pattern of the second day in each pair on the first day, varying the interval between the two days in the pair. The results for R^2_{det} and also associated p-values obtained from 1000 bootstrap samples generated under the null hypothesis of no functional relationship (Müller, Chiou and Leng 2008) are shown in Table 1, suggesting that pairs of days that are six days apart (where the first day of each pair is randomly chosen) can be treated as independent for practical purposes. The sample then consists of data for $n = 92$ trading days that satisfy constraints (a) and (b).

We applied the following variant of the PART algorithm (see Section 3.1). As inference about the functional volatility process is based on approximating $Z_\Delta(t)$ by $\widetilde{\sigma}(t)W_\Delta(t)$, (7) suggests that the approximation might be improved in practice if in a first step one estimates individual drifts $\tilde{\mu}_i(t)$, by smoothing scatterplots $\{(t_j, Z_{ij\Delta}), j = 1, \ldots, [T/\Delta]\}$, for each fixed $1 \leq i \leq n$. Denoting the smoothed trajectories obtained from this smoothing step by $\hat{z}_i(t)$, which are substituted for $\tilde{\mu}_i(t)\sqrt{\Delta}$, one then forms $Z'_{ij\Delta} = Z_{ij\Delta} - \hat{z}_i(t_j)$ and substitutes the transformed observations, the *raw volatilities* $Y'_{ij\Delta} = \log(\{Z'_{ij\Delta}\}^2) - q_0$, for $Y_{ij\Delta}$ in the PART algorithm. We adopted this variant, which led to slightly improved results in our application. Overall, the differences as compared to the analysis obtained from the unmodified PART algorithm were minor. For the initial smoothing step to obtain the initial estimates $\hat{z}_i(t)$ we used a cross-validation bandwidth choice.

5.2 Estimation of components of the functional volatility process

When applying the PART algorithm to the transformed residuals or raw volatilities $Y'_{ij\Delta}$, one needs to choose the included number of components K, see (21). Using an AIC criterion with pseudo-likelihood (marginal likelihood) led to the choice $K = 3$. Also from the scree plot, $K = 3$ seems to be a good choice, explaining 97.66% of total variance. The first three estimated eigenvalues are 23.42, 2.36 and 1.95. In Figure 3 we present the estimated mean function $\hat{\mu}_V$ and covariance surface $\hat{G}_V(s,t)$ of the functional volatility process V, which reflect the smoothing steps (17) and (18), respectively. The estimated covariance surface is obtained by plugging the fitted eigenfunctions and eigenvalues into $\sum_{k=1}^{K} \lambda_k \phi_k(s)\phi_k(t)$. The mean function of the volatility process shows a dip during midday. A noon dip in volatility has been reported before (e.g., Andersen and Bollerslev, 1998) and may be attributed to a drop in trading activity during lunch break.

Estimated eigenfunctions $\hat{\phi}_k$ are illustrated in Figure 4. The first eigenfunction, corresponding to the primary mode of variation of daily volatility, parallels the mean function (note that the sign of the eigenfunctions is arbitrary) and primarily reflects the overall level of volatility on a given day. The second eigenfunction differentiates between morning and afternoon trading volatility, and also separates out the volatility of the very last trades on a given day. The third eigenfunction reflects shape changes between more moderate or more pronounced volatility dips around noon. Raw volatilities $Y'_{ij\Delta}$ and the estimated trajectories $\hat{V}_i(t)$ of the functional volatility process, as defined in (21), are shown for six randomly chosen days in Figure 5. The individual time courses of volatility reflect the trends of the raw volatilities in a smooth fashion and exhibit substantial variability from day to day.

5.3 Regressing volatility trajectories of second half day on first half day of trading

One aim of volatility analysis and a yardstick for the usefulness of the proposed methodology is the prediction of future volatility from past observations. For this purpose, we adopt functional regression analysis as described in Section 3.2 to predict afternoon volatility trajectories from those observed in the morning. Predictor processes are thus chosen as $Q(t) = V(t)$, with t restricted to the first half of the

trading day, which then is the domain $\mathcal{T}_Q$ of Q, and response processes are $R(t) = V(t)$, with t restricted to the second half of the trading day, which is the domain $\mathcal{T}_R$ of R. This method applies the accumulated knowledge about the relationship of volatility patterns as learned in a training set to the prediction task. Specifically, we are aiming at predicting the volatility process for the second half day (response) from that of the first half day (predictor).

Of special interest is the quantitative gain in predicting the afternoon volatility trajectories when applying the prediction based on functional regression over a baseline estimate. As baseline we use the estimated mean afternoon volatility function $\hat{\mu}_R^{(-i)}$, which is not influenced by morning trading and is obtained as in (17), while also omitting the data from day i. In order to quantify the performance of these two predictors, we first obtain target afternoon volatility trajectories by polynomial smoothing of the raw volatilities $Y_{ij\Delta}$, obtaining for each afternoon a target volatility trajectory R_i. The predicted afternoon volatility trajectory for the i-th day is $\widehat{R}^{(-i)*}(t|Q_i)$ (25), obtained by substituting ξ_k^* by $\hat{\xi}_{ki}$ (20) for the predictor process Q_i, while omitting the data from day i. Thus the situation for day i corresponds to prediction of the response at a new level of the predictor. Data from the afternoon for which volatility is to be predicted are not used in any way.

The ratios

$$r_i = \frac{\int_{\mathcal{T}'}(R_i(t) - \widehat{R}^{(-i)*}(t|Q_i))^2\,dt}{\int_{\mathcal{T}'}(R_i(t) - \hat{\mu}_R^{(-i)}(t))^2\,dt}, \tag{39}$$

where $\mathcal{T}' \subset \mathcal{T}_R$ is chosen to eliminate the influence of boundary effects that are inherent in R_i, provide a measure of the relative performance of these predictors, in terms of L^2 distance of the predicted to the observed target trajectory. We compute the median of this ratio over the 92 days in the sample and minimize this median over a grid of (K, M), chosen from $1 \le K, M \le 5$. The resulting optimal value is $K = M = 1$.

To visualize the functional regression of afternoon volatility on morning volatility, we display the eigenfunctions and a scatterplot of the FPC scores of afternoon volatility trajectories R_i vs. morning volatility trajectories Q_i in Figure 6, choosing $K = M = 1$ for the representations (23) of predictor and

response processes Q and R. We find that the functional regression of these volatility processes on each other both explains a sizable fraction of the variance of the afternoon processes and also is significant. This can be seen from the size of the functional coefficient of determination $R^2_{\text{det}} = 0.45$ and the bootstrap p-value $p < .001$.

The median of the ratio r_i in 39 over the 92 days in the sample is 0.703, implying that the squared prediction error of the predictor based on functional regression for volatility processes is approximately 30% smaller than the simple prediction using the overall mean trajectory. We present average squared errors between the target function $R_i(t)$ and the two predictors, substituting for $E(R_i(t) - \hat{\mu}_R^{(-i)}(t))^2$ and $E(R_i(t) - \widehat{R}^{(-i)*}(t|Q_i))^2$, as functions of t, in Figure 7. This demonstrates the superiority of the functional prediction, as its error is uniformly lower than that of the mean, especially for early afternoon trading; both errors increase somewhat towards the end of the trading day. Our analysis provides evidence that substantial gains can be achieved in volatility prediction by combining the volatility process with functional regression.

6. Discussion and Concluding Remarks

A central feature of the proposed approach that distinguishes it from established methods for spot volatility modeling and estimation is that it is geared towards the analysis of observations drawn from repeated realizations of the volatility process, rather than observations from a single realization. Existing nonparametric approaches for volatility modeling implement smoothing of the volatility function separately for each realization of the volatility process and do not aim to determine recurrent patterns of this process. If one has repeated realizations of the process available, as in day-to-day volatility modeling, it is clearly of interest to deploy methods that combine information and extract patterns across these realizations. The methodology proposed in this paper can be viewed as a first step in this direction.

To facilitate the analysis of a repeatedly observed volatility process, we develop a version of functional principal component analysis (FPCA) that is suitable for volatility modeling. The substantial dimension reduction afforded by FPCA enables financial data analysts to effectively model and study samples of

repeatedly observed volatility trajectories. For effective dimension reduction through FPCA one needs to choose the number of components to be included, which corresponds to the reduced dimension of the originally infinite-dimensional trajectories of the functional volatility process. Depending on the context, this can be done by targeting specific outcomes such as fraction of total variance explained or one-leave-out prediction error, if one aims at predicting a response. A comprehensive theory for the choice of this dimension parameter remains an open problem.

An important assumption that we need to make is smoothness of the underlying process $\tilde{\sigma}$ in (5) and of the functional volatility process defined as $V(t) = \log(\{\widetilde{\sigma}(t)\}^2)$ in (9), reflected in assumptions (M1)-(M5) in the Appendix. We note that the observed data that are generated by these processes are not assumed to be smooth, as they result from stochastic integrals of these smooth random processes with respect to the Wiener process. While there is no evidence for non-smoothness in the daily volatility data, there certainly are situations for which the smoothness assumptions on the volatility process are too restrictive. Further discussion of the smooth volatility assumption (M5) can be found in Appendix A.1, and explicit examples of smooth volatility processes are given in Section 3.4.

While the proposed methods may still give some insights in cases of mild violations of the assumed smoothness of volatility, the theoretical results will not apply for non-smooth processes. Another necessary assumption for the theoretical results to hold is that a sample of i.i.d. realizations of the processes is available. The general consistency results are expected to be robust in the presence of weak dependence and also with respect to mild violations of the required smoothness of the volatility process. When fitting the model, one can use subsamples consisting of practically independent trajectories, as we have illustrated in the study of intra-day trading. Nevertheless, this assumption is strong and restricts the applicability of the proposed methods. To develop theory for the analysis of repeated realizations for more general types of volatility processes under weaker assumptions will be a task for future work.

In many situations, the first few principal components and the associated dominant characteristics of the volatility function lend themselves to interpretation and general descriptions of recurrent patterns in

financial transactions. The principal components and associated eigenfunctions can be used to summarize the salient features of the volatility process. Further statistical analyses are then feasible without running into the curse of dimensionality, which otherwise would require to invoke strong parametric assumptions. For example, one can carry out a functional regression analysis by using the functional principal components of the functional volatility process as predictors, coupled with scalar or functional responses. An example is the prediction of afternoon volatility paths from morning volatility trajectories that we study in section 4. Other applications of interest include cluster analyses based on the estimated FPCA scores. Such procedures are either impossible or inefficient without a prior dimension reduction step. For these and other applications, the development of asymptotic distribution theory will be desirable; in practical applications, inference via bootstrap procedures is an option.

We have established that under certain assumptions the functional paradigm can be applied to advantage within the framework of a diffusion model for financial returns such as (3). The proposed functional volatility process approach can be used for applications in modeling, regression and prediction. It provides tools for studying the dependency of volatility trajectories on other predictors and modeling the relationship of volatility with other financial processes, such as the trade volume process. We envision benefits for portfolio allocation and risk management from using functional methods for financial data. As we have seen in Section 4.3, the difficult problem of predicting volatility can be successfully approached with the functional volatility process. We conclude that this methodology is a promising tool for the analysis of financial data.

Appendix

A.1 Assumptions

We begin by listing some assumptions for processes $\widetilde{\mu}$, $\widetilde{\sigma}$, as needed for some basic bounds. Throughout we consider $\Delta \to 0$. "Smooth" in the following refers to twice differentiable, the domain$\mathcal{T}$ is taken to be the interval $[0, T]$ and $c, c_0, c_1, \ldots$ denote generic positive constants.

(M1) Processes $(\widetilde{\mu}(t))_{t\in[0,T]}$ and $(\widetilde{\sigma}(t))_{t\in[0,T]}$ (4) have sample paths that are uniformly Lipschitz contin-

uous of order 1, i.e., there exists a constant $L > 0$ so that a.s.

$$|\widetilde{\mu}(t) - \widetilde{\mu}(s)| < L|t-s|, \quad |\widetilde{\sigma}(t) - \widetilde{\sigma}(s)| < L|t-s|.$$

(M2) Processes $\widetilde{\mu}$ satisfy $E[\sup_{t\in[0,T]} |\widetilde{\mu}(t)|] < \infty$

(M3) Processes $\widetilde{\sigma}$ satisfy $\sup_{t\in[0,T]} E[\widetilde{\sigma}(t)^2] < \infty$

(M4) Sample paths of processes $\widetilde{\sigma}$ are smooth and derivatives $\frac{d}{dt}\widetilde{\sigma}(t)$ satisfy

$$E[\sup_{t\in[0,T]} |\frac{d}{dt}\widetilde{\sigma}(t)|^2] = O(1).$$

(M5) The log-variance process can be represented as

$$\log(\widetilde{\sigma}^2(t)) = \widetilde{W}(t) + H(t), \text{ with } \widetilde{W}(t) = f(\int_{t-\delta}^{t} \kappa(t-v)dW(v)) \text{ for a } \delta > 0,$$

where $E(|\widetilde{W}(t)W(t)^p|) < \infty$ for $p = 1, 2$. Here H is a continuously differentiable process, which is independent of W, f is a continuously differentiable function and κ a twice continuously differentiable kernel function with support supp $\subseteq [0, \delta]$, which satisfies

$$\kappa(0) = \kappa'(0) = \kappa''(0) = 0, \quad \kappa(\delta) = \kappa'(\delta) = \kappa''(\delta) = 0, \quad \int_0^{\tilde{\delta}} \kappa^2(u)du > 0 \quad \text{for all} \quad \tilde{\delta} > 0.$$

Assumption (M5) merits some discussion. This assumption implies continuously differentiable volatility functions $\log(\widetilde{\sigma}^2(t))$. This differentiates our approach from the Heston model in (36), (37), for which both price and volatility processes are non-smooth, while in our approach, according to (M5), the price process is non-smooth but the volatility process is differentiable. Applying the properties of the kernel in (M5) and integration by parts, and defining $\widetilde{W}_1(t) = f'(\int_{t-\delta}^{t} \kappa(t-v)dW(v))$, we find

$$\frac{d}{dt}\widetilde{W}(t) = \widetilde{W}_1(t)\frac{d}{dt}\int_{t-\delta}^{t} \kappa'(t-v)W(v)dv = \widetilde{W}_1(t)\int_{t-\delta}^{t} \kappa''(t-v)W(v)dv. \tag{40}$$

We also note that the correlations between process W and the log variance process $\log(\widetilde{\sigma}^2)$ vary in dependence on the function f and processes H. Specific examples of smooth volatility processes that satisfy (M5) are given in Section 3.4.

Additional assumptions are as follows. To apply the logarithmic transformation, we assume that the data are truncated near zero, i.e., $|\widetilde{\sigma}(t)W_\Delta(t)| \geq \varepsilon$ for a small $\varepsilon > 0$, and when conducting the transformation we replace the original $Z_{ij\Delta}$ by $\tilde{Z}_{ij\Delta} = \max(Z_{ij\Delta}, \varepsilon)$. Then (29) immediately implies that

(M6) $$E[\sup_{1\leq j\leq T/\Delta} |\log(\{Z_\Delta(t_j)\}^2) - q_0 - Y_\Delta(t_j)|] = O(\sqrt{\Delta}).$$

This assumption will be utilized for some of the results. We note that implementing such a lower threshold might also be occasionally necessary in practice to execute the transformation smoothly; this was not an issue in our data illustrations in Section 4.

For the moments of $Y_{ij\Delta}$ (16) and the smoothing bandwidths b_V and h_V as used in (17) and (18), some of the results require the following assumptions.

(M7) $\sup_j E[Y_{ij\Delta}]^4 < \infty$

(M8) As $n \to \infty$,

$$b_V \to 0, \quad nb_V^4 \to \infty, \quad \text{limsup}_n\, nb_V^6 < \infty,$$

$$h_V \to 0, \quad nh_V^6 \to \infty, \quad \text{limsup}_n\, nh_V^8 < \infty,$$

$$\text{limsup}_n\, n^{1/2} b_V \Delta < \infty, \quad \text{limsup}_n\, n^{1/2} h_V^2 \Delta < \infty.$$

For the proofs of Theorems 4 and 5 we require further assumptions on the nature of the volatility process V. Further details on these assumptions and notation can be found in Müller et al. (2006). For each $j \geq 0$, define $\delta_j^V = \frac{1}{2}\min\{|\rho_l - \rho_j| : l \neq j\}$, and $\mathbf{\Lambda}_{\delta_j^V} = \{z \in \mathcal{C} : |z - \rho_j| = \delta_j^V\}$, where $\mathcal{C}$ are the complex numbers. Furthermore, define $A_{\delta_j^V} = \sup\{\|\boldsymbol{R}_V(z)\|_F : z \in \mathbf{\Lambda}_{\delta_j^V}\}$, where $\boldsymbol{R}_V(z) = (\boldsymbol{G}_V - zI)^{-1}$ is the resolvent of operator $\boldsymbol{G}_V$ and $\|\cdot\|_F$ is an operator norm, defined on the separable Hilbert space F generated by the Hilbert-Schmidt operators on H, endowed with the inner product $\langle T_1, T_2\rangle_F = \sum_j \langle T_1 u_j, T_2 u_j\rangle_H$ and the norm $\|T\|_F^2 = \langle T, T\rangle_F$, where $T_1, T_2, T \in F$, and $\{u_j : j \geq 1\}$ is any complete orthonormal system in H. Then we assume

(M9) $\sum_{j=1}^K (\delta_j^V A_{\delta_j^V} \sup_{t\in[0,T]} |\phi_j(t)|)/(\sqrt{n}h_V^2 - A_{\delta_j^V}) \to 0$, as $K = K(n) \to \infty$.

Further assumptions are needed for Theorem 5. For the sequence $K = K(n) \to \infty$,

(M10) $\sum_{j=1}^{K} \sup_{t\in[0,T]} |\phi_j(t)| = o(\min\{\sqrt{n}b_V, \sqrt{\Delta^{-1}}\})$, and
$\sum_{j=1}^{K} \sup_{t\in[0,T]} |\phi_j(t)| \sup_{t\in[0,T]} |\phi_j'(t)| = o(\Delta^{-1})$

(M11) $E\{[\sup_{t\in[0,T]} |V(t) - V^{(K)}(t)|]^2\} = o(1)$, where $V^{(K)}(t) = \mu_V(t) + \sum_{k=1}^{K} \xi_k \phi_k(t)$

(M12) $\sum_{j=1}^{K} \sup_{t\in[0,T]} |\phi_j(t)|^2 = o_p(\Delta^{-1/2})$ and
$\sum_{j=1}^{K} (\delta_j^V A_{\delta_j^V} \sup_{t\in[0,T]} |\phi_j(t)|/(\Delta^{-1/2} - A_{\delta_j^V}) \to 0$, as $n \to \infty$

(M13) $\tilde{h} \to 0, \quad n\tilde{h}^6 \to \infty, \quad \text{limsup}_n\, n\tilde{h}^8 < \infty, \quad \text{limsup}_n\, n^{1/2}\tilde{h}^2\Delta < \infty.$

Here, $\tilde{h}$ is the bandwidth used for smoothing the cross-covariance of predictor and response processes, as described after eq. (26).

A.2 Auxiliary results and proofs

Next we state two auxiliary lemmas with proofs.

Lemma 1. Under (M1)-(M4), it holds that

$$E[\sup_{t\in[0,T]} |R_1(t,\Delta)|] = O(\Delta^{3/2}), \quad E[\sup_{t\in[0,T]} |R_2(t,\Delta)|] = O(\Delta^{1/2}).$$

Proof: The result for $R_1(t,\Delta)$ follows from

$$\begin{aligned} \left| \frac{1}{\sqrt{\Delta}} \int_t^{t+\Delta} \widetilde{\mu}(v)\, dv - \widetilde{\mu}(t)\sqrt{\Delta} \right| &\leq \frac{1}{\sqrt{\Delta}} \int_t^{t+\Delta} |\widetilde{\mu}(v) - \widetilde{\mu}(t)|\, dv \\ &= O(\frac{1}{\sqrt{\Delta}}) \int_t^{t+\Delta} |v-t|\, dv \;=\; O(\Delta^{3/2}), \end{aligned}$$

where the O-terms are a.s. and uniform in t. For $R_2(t,\Delta)$, by partial integration,

$$\begin{aligned} R_2(t,\Delta) &= \frac{1}{\sqrt{\Delta}} \int_t^{t+\Delta} (\widetilde{\sigma}(v) - \widetilde{\sigma}(t))\, dW(v) \\ &= \frac{1}{\sqrt{\Delta}} \left(W(t+\Delta)(\widetilde{\sigma}(t+\Delta) - \widetilde{\sigma}(t)) - \int_t^{t+\Delta} W(v) d(\widetilde{\sigma}(v) - \widetilde{\sigma}(t)) \right). \end{aligned}$$

Observing $E\sup_{t\in[0,T]}|W(t)|^2<\infty$ for the Wiener process (e.g., Adler, 1990, Thm. 3.2) and $d[\widetilde{\sigma}(v)-\widetilde{\sigma}(t)]=\widetilde{\sigma}'(v)\,dv$, as well as (M4), it follows that

$$\begin{aligned} E[\sup_{t\in[0,T]}|R_2(t,\Delta)|] &\leq \frac{1}{\sqrt{\Delta}}\{E[\sup_{t\in[0,T]}|W(t+\Delta)|^2]\,E[\sup_{t\in[0,T]}|\widetilde{\sigma}(t+\Delta)-\widetilde{\sigma}(t)|^2]\}^{1/2} \\ &\quad +\frac{1}{\sqrt{\Delta}}\int_t^{t+\Delta}\{E[\sup_{t\in[0,T]}|W(v)|^2]\,E[\sup_{t\in[0,T]}|\widetilde{\sigma}'(v)|^2]\}^{1/2}\,dv \\ &= O(\Delta^{1/2}). \end{aligned}$$

The analysis of $Y_\Delta(t)=V(t)+U_\Delta(t)$ with $V(t)=\log(\widetilde{\sigma}^2(t))$ and $U_\Delta(t)=\log(W_\Delta^2(t))-q_0$ will be facilitated by the following result.

Lemma 2. Assume that the variance process satisfies (M5). Then for any $\delta>0$ there exists a constant $c>0$ such that uniformly for all s,t with $0<\delta\leq s,t\leq T$, for sufficiently small Δ,

$$\big|\operatorname{cov}(\log(\widetilde{\sigma}^2(t)),\log(W_\Delta^2(s)))\big|\leq c\sqrt{\Delta}\,. \tag{41}$$

Proof: We may assume w.l.o.g. that $\Delta\leq 1-\delta<1$. With $\Psi(t)=\int_{t-1}^t\kappa(t-v)dW(v)$, writing $W_\Delta(s)=\frac{1}{\sqrt{\Delta}}\int_s^{s+\Delta}dW(v)$, one finds

$$\operatorname{var}(\Psi(t))=\int_{t-1}^t\kappa^2(t-v)\,dv=\int_0^{1\wedge t}\kappa^2(v)\,dv,\quad \operatorname{cov}(W_\Delta(s),\Psi(t))=\frac{1}{\sqrt{\Delta}}\int_s^{s+\Delta}\kappa(t-v)\,dv,$$

which is zero for $t\leq s$ or $t>s+\Delta+1$, so that it remains to establish (41) for the case $s<t\leq s+\Delta+1$. Then the conditional distribution of $W_\Delta(s)\mid\Psi(t)$ is normal $N(m(s,t,\Delta),\,\varsigma^2(s,t,\Delta))$ with

$$\begin{aligned} m(s,t,\Delta) &= \frac{\frac{1}{\sqrt{\Delta}}\int_s^{s+\Delta}\kappa(t-v)\,dv}{\int_{t-1}^t\kappa^2(t-v)\,dv}\,\Psi(t)=\sqrt{\Delta}\,\theta_1(s,t)\Psi(t), \\ \varsigma^2(s,t,\Delta) &= 1-\frac{\Delta[\frac{1}{\Delta}\int_s^{s+\Delta}\kappa(t-v)\,dv]^2}{\int_0^{1\wedge t}\kappa^2(v)\,dv}=1-\Delta\theta_2(s,t), \end{aligned}$$

with suitable functions θ_1,θ_2. Given $\delta>0$ as in the assumption of Lemma 2, according to (M5) the functions θ_1,θ_2 are uniformly bounded,

$$\sup_{\delta<s,t\leq T}|\theta_1(s,t)|<\infty,\quad \sup_{\delta<s,t\leq T}|\theta_2(s,t)|\leq\frac{\int_s^{s+\Delta}\kappa^2(t-v)\,dv}{\int_{t-1}^t\kappa^2(t-v)\,dv}\leq 1\,,$$

where we used the Cauchy-Schwarz-inequality for θ_2. Next we will prove

$$|D_\Delta| < \sqrt{\Delta}\left(c_0 + c_1|\Psi(t)| + c_2\Psi^2(t)\right) \quad \text{for} \quad D_\Delta = E\left(\log(W_\Delta^2(s)) - E(\log(W_\Delta^2(s))) \mid \Psi(t)\right). \tag{42}$$

Observe that

$$\begin{aligned} D_\Delta &= \int_{-\infty}^{\infty} \log y^2 \exp\left(-\frac{1}{2}\frac{(y-\sqrt{\Delta}\,\theta_1(s,t)\Psi(t))^2}{1-\Delta\,\theta_2(s,t)}\right) \frac{dy}{\{2\pi(1-\Delta\,\theta_2(s,t))\}^{1/2}} \left(1-\{1-\Delta\,\theta_2(s,t)\}^{1/2}\right) \\ &\quad + \int_{-\infty}^{\infty} \log y^2 \left(\exp\left(-\frac{1}{2}\frac{(y-\sqrt{\Delta}\,\theta_1(s,t)\Psi(t))^2}{1-\Delta\,\theta_2(s,t)}\right) - \exp\left(-\frac{y^2}{2}\right)\right)\frac{dy}{\sqrt{2\pi}} \\ &= A_\Delta + B_\Delta, \qquad \text{say.} \end{aligned}$$

For the first term,

$$\begin{aligned} |A_\Delta| &\le c_3\,\Delta \int_{|y|\le 1} \log y^2 \frac{dy}{\sqrt{2\pi}} \\ &\quad + c_4\Delta \int_{|y|\ge 1} (y^2-1)\exp\left(-\frac{1}{2}\frac{(y-\sqrt{\Delta}\,\theta_1(s,t)\Psi(t))^2}{1-\Delta\,\theta_2(s,t)}\right)\frac{dy}{\{2\pi(1-\Delta\,\theta_2(s,t))\}^{1/2}} \le \Delta(c_5 + c_6\Delta\Psi(t)^2). \end{aligned}$$

For the second term, observing $|e^x - e^y| \le |x-y|e^x$,

$$\begin{aligned} |B_\Delta| &\le c_7\sqrt{\Delta}\int_{-\infty}^{\infty} \log y^2 \left|2y\theta_1(s,t)\Psi(t) + \sqrt{\Delta}\{\theta_1(s,t)\Psi(t)\}^2 + y^2\sqrt{\Delta}\theta_2(s,t)\right| \exp\left(-\frac{y^2}{2}\right)\frac{dy}{\sqrt{2\pi}} \\ &\le c_8\sqrt{\Delta}\left(|\Psi(t)| + \sqrt{\Delta}\Psi^2(t) + c_9\sqrt{\Delta}\right), \end{aligned}$$

which implies (42), from which we may infer with (M5) that

$$\begin{aligned} \left|\operatorname{cov}(\log(\tilde\sigma^2(t)), \log W_\Delta^2(s))\right| &= \left|E\left(\log(\tilde\sigma^2(t)) - E(\log(\tilde\sigma^2(t)))\right)\left(\log W_\Delta^2(s) - q_0\right)\right| \\ &= \left|E\left(\left(\widetilde{W}(t) + H(t) - E(\tilde\sigma^2(t))\right)\left(\log W_\Delta^2(s) - q_0\right) \mid \Psi(t)\right)\right| \\ &= \left|E\left(\left(\widetilde{W}(t) - E(\log\tilde\sigma^2(t))\right)E\left(\log W_\Delta^2(s) - q_0 \mid \Psi(t)\right)\right)\right| \\ &\le c\sqrt{\Delta}, \end{aligned}$$

whence the result follows.

Corollary. For any given $\delta > 0$, it holds for all s, t with $\delta < s, t \in \mathcal{T}$ that

$$\operatorname{cov}(V(t) + U_\Delta(t), V(s) + U_\Delta(s)) = G_V(t,s) + O(\sqrt{\Delta}), \tag{43}$$

where the O-term is uniform in s and t.

We are now in a position to give proofs for the main results.

Proof of Theorem 1. The result follows directly from (5) and Lemma 1, noting that $E \sup_{t\in[0,T]} |\widetilde{\mu}(t)| < \infty$.

Proof of Theorem 2. The proof borrows arguments from the proofs of Lemma C1 and Theorem 2 of Müller et al. (2006), with some variations due to the fact that in the present context of volatility estimation there is neither an issue of estimating the smooth price trajectories $\widetilde{\mu}$ nor of estimating the variance of the errors $W_\Delta(t_j)$. Assumptions (M1)-(M8) and those made about the kernels in the estimation section 3.1 ensure that proper versions of Lemma C1 and the above-mentioned Theorem 2 apply here. A first step is to replace in the basic mean function and covariance function estimates (17) and (18) the transformed data $Y_{ij\Delta}$ by their (unknown) target values $Y_\Delta(t_j)$. For the resulting mean function estimates $\tilde{\mu}_V$ and covariance surface estimates $\tilde{G}_V$, Lemma C1 implies that $\sup_{t\in\mathcal{T}} |\tilde{\mu}_V(t) - \mu_V(t)| = O_P((n^{1/2}b_V)^{-1})$, and $\sup_{s,t\in\mathcal{T}} |\widetilde{G}_V(s,t) - G_V(s,t)| = O_P((n^{1/2}h_V^2)^{-1})$. In a second step, combining with (M6) and (43), results (30), (31) follow.

Proof of Theorem 3. The proof is analogous to that of (C3) in Lemma C1 in Müller et al. (2006).

Proof of Theorem 4. The proof is a slightly modified version of the proof of Theorem 3 in Müller et al. (2006).

Proof of Theorem 5. We first observe that

$$\sup_{s,t\in\mathcal{T}_{QR}} |\widehat{G}_{QR}(s,t) - G_{QR}(s,t)| = O_P(\sqrt{\Delta} + (n^{1/2}\tilde{h}^2)^{-1}), \tag{44}$$

which is proved similarly to (31). Next, the assumptions imply, analogously to arguments used in the proof of Theorem 3 in Müller et al. (2006), that

$$\max_{1\le k\le K} \sup_{t\in\mathcal{T}_Q} |\widehat{\phi}_k(t) - \phi_k(t)| \xrightarrow{P} 0, \quad \max_{1\le m\le M} \sup_{t\in\mathcal{T}_R} |\widehat{\psi}_m(t) - \psi_m(t)| \xrightarrow{P} 0, \tag{45}$$

and

$$\max_{1\le k\le K} |\hat{\lambda}_k - \lambda_k| \xrightarrow{P} 0. \tag{46}$$

Combining (44), (45), (46), we find with (26) that

$$\max_{1\leq k\leq K, 1\leq m\leq M} |\hat{\gamma}_{km} - \gamma_{km}| \xrightarrow{P} 0. \tag{47}$$

The result (34) is an immediate consequence of (45), (46) and (47). Result (35) follows by observing in addition that the ξ_k^* are given and that (30) holds for response processes $R = V$.

Acknowledgments

The research of H.G.M. was supported in part by NSF grants DMS-0806199 and DMS-1104426. We are grateful to the referees and an associate editor for their detailed reading and very helpful comments, which led to major improvements in the paper.

References

Adler, R. J., 1990, An Introduction to Continuity, Extrema and Related Topics for General Gaussian Processes, Hayward, CA. Institute of Mathematical Statistics, Lecture Notes – Monograph Series, Vol. 12.

Aït-Sahalia, Y., 1996, Nonparametric pricing of interest rate derivative securities. Econometrica 64, 527-560.

Aït-Sahalia, Y. and Mykland, P. A., 2003, The effects of random and discrete sampling when estimating continuous-time diffusions. Econometrica 71, 483-549.

Aït-Sahalia, Y., Mykland, P. A. and Zhang, L., 2005, How often to sample a continuous-time process in the presence of market microstructure noise. Review of Financial Studies 18, 351-416.

Aït-Sahalia, Y., Mykland, P. A. and Zhang, L., 2010, Ultra high frequency volatility estimation with dependent microstructure noise. Forthcoming in the Journal of Econometrics.

Andersen, T. G. and Bollerslev, T., 1997, Intraday periodicity and volatility persistence in financial markets. J. Empirical Finance 4, 115-158.

Andersen, T. G and Bollerslev T., 1998, Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review 39, 885-905.

Bandi, F. and Phillips P., 2003, Fully nonparametric estimation of scalar diffusion models. Econometrica 71, 241283.

Bandi, F. and Renò R., 2008, Nonparametric stochastic volatility. Working Paper.

Barndorff-Nielsen, O.E., Graversen, S.E., Jacod, J. and Shephard, N., 2006, Limit theorems for realised bipower variation in econometrics. Econometric Theory, 22, 677-719.

Barndorff-Nielsen, O. E. and Shephard, N., 2002, Econometric analysis of realized volatility and its use in estimating stochastic volatility models. J. Royal Statistical Society Series B 64, 253-280.

Black, F. and Scholes, M., 1973, The pricing of options and corporate liabilities. J. Political Economy 81, 637-654.

Castro, P. E., Lawton, W. H. and Sylvestre, E. A., 1986, Principal modes of variation for processes with continuous sample curves. Technometrics 28, 329-337.

Dacorogna, M., Gencay, R., Muller, U., Olsen, R. B and Pictet, O. V., 2001, An Introduction to High-Frequency Finance. San Diego, CA: Academic Press.

Eubank, R. and Thomas, W., 1993, Detecting heteroscedasticity in nonparametric regression. J. Royal Statistical Society Series B 55, 145-155.

Fan, J., 2005, A selective overview of nonparametric methods in financial econometrics (with discussion). Statistical Science 20, 317-357.

Fan, J., Jiang, J., Zhang, C. and Zhou, Z., 2003, Time-dependent diffusion models for term structure dynamics. Statistica Sinica 13, 965-992.

Fan, J. and Gijbels, I., 1996, Local Polynomial Modelling and its Applications. London, UK: Chapman and Hall.

Fan, J. and Wang, Y., 2007, Multiscale jump and volatility analysis for high-frequency financial data. Journal of American Statistical Association 102, 1349-1362.

Fan, J. and Wang Y., 2008, Spot volatility estimation for high-frequency data. Statistics and Its Interface 1, 279-288.

Fan, J. and Yao, Q.W., 1998, Efficient estimation of conditional variance functions in stochastic regression. Biometrika 85, 645–660.

Fan, J. and Yao, Q.W., 2003, Nonlinear Time Series: Nonparametric and Parametric Methods. Springer.

Faraway, J. J., 1997, Regression analysis for a functional response. Technometrics 39, 254–262.

Florens-Zmirou, D., 1993, On estimating the diffusion coefficient from discrete observations. Journal of Applied Probability 30, 790-804.

Foster, D. and Nelson D., 1996, Continuous record asymptotics for rolling sample variance estimators. Econometrica 64, 139–174.

Genon-Catalot, V., Laredo C. and Picard D., 1992, Non-parametric estimation of the diffusion coefficient by wavelet methods. Scandinavian Journal of Statistics 19, 317–335.

Gerety, M. S. and Mulherin, J. H., 1994, Price formation on stock exchanges: the evolution of trading within the day. Review of Financial Studies 7, 609–629.

Jacod, J. and Shiryaev, A., 2003, Limit Theorems for Stochastic Processes. Springer, New York.

Kogure, A., 1996, Nonparametric prediction for the time-dependent volatility of the security price. Financial Engineering Jap. Market 3, 1–22.

Kristensen, D., 2010, Nonparametric filtering of the realized spot volatility: A kernel-based approach. Econometric Theory 26, 60-93.

Malliavin, P. and Mancino M., 2009, A Fourier transform method for non- parametric estimation of volatility. Annals of Statistics 37, 1983-2010.

Mancino, M. and Sanfelici S., 2008, Robustness of Fourier estimator of integrated volatility in the presence of microstructure noise. Computational Statistics and Data Analysis 52, 2966–2989.

Malfait, N., and Ramsay, J. O., 2003, The historical functional linear model. Canadian J. of Statistics 31, 115-128.

Müller, H.G., Chiou, J.M., Leng, X., 2008, Inferring gene expression dynamics via functional regression analysis. BMC Bioinformatics 9:60.

Müller, H. G., Stadtmüller, U. and Yao, F., 2006, Functional variance processes. J. American Statistical Association 101, 1007-1018.

Ogawa, S. and Sanfelici S., 2008, An improved two-step regularization scheme for spot volatility estimation. Working paper.

Ramsay, J. O. and Dalzell, C. J., 1991, Some tools for functional data analysis. J. Royal Statistical Society Series B 53, 539-572.

Ramsay, J. O. and Ramsey, J. B., 2001, Functional data analysis of the dynamics of the monthly index of non durable goods production. J. Econometrics 107, 327-344.

Ramsay, J. O. and Silverman, B. W., 2005, Functional Data Analysis. New York, NY: Springer.

Renò, R., 2008, Nonparametric estimation of the diffusion coefficient of stochastic volatility models. Econometric Theory 24, 1174-1206.

Rice, J. A. and Silverman, B. W. ,1991, Estimating the mean and covariance structure nonparametrically when the data are curves. J. Royal Statistical Society Series B 53, 233–243.

Speight A. E. H., McMillan, D. C. and Gwilym, O. A. P., 2000, Intra-day volatility components in FTSE-100 stock index futures. J. Futures Markets 20, 425–444.

Stanton, R., 1997, A nonparametric model of term structure dynamics and the market price of interest rate risk. J. Finance 52, 1973–2002.

Wang, L., Brown, L. D., Cai, T. T.and Levine, M. (2008), Effect of mean on variance function estimation in nonparametric regression. Annals of Statistics 36, 646–664.

Yao, F., Müller, H. G., Clifford, A. J., Dueker, S. R., Follett, J., Lin, Y., Buchholz, B. and Vogel, J. S., 2003, Shrinkage estimation for functional principal component scores, with application to the population kinetics of plasma folate. Biometrics 59, 676–685.

Yao, F., Müller, H. G. and Wang, J. L., 2005a, Functional data analysis for sparse longitudinal data. J.

American Statistical Association 100, 577–590.

Yao, F., Müller, H. G. and Wang, J. L., 2005b, Functional linear regression analysis for longitudinal data. Annals of Statistics 33, 2873–2903.

Yao, Q. W. and Tong, H., 2000, Nonparametric estimation of ratios of noise to signal in stochastic regression. Statistica Sinica 10, 751-770.

Zhang, L., Mykland, P. and Aït-Sahalia, Y., 2005, A tale of two time scales: determining integrated volatility with noisy high-frequency data. J. American Statistical Association 100, 1394-1411.

Lag	1	2	3	4	5	6
R^2_{det}	0.176	0.147	0.240	0.202	0.092	0.073
p-value	0.460	0.384	0.244	0.366	0.894	0.965

Table 1: Functional R^2_{det}, as defined in (28), and bootstrap p-values for regressing second day on first day for randomly selected pairs of trading days, in dependency on the lag (in days) between the two days.

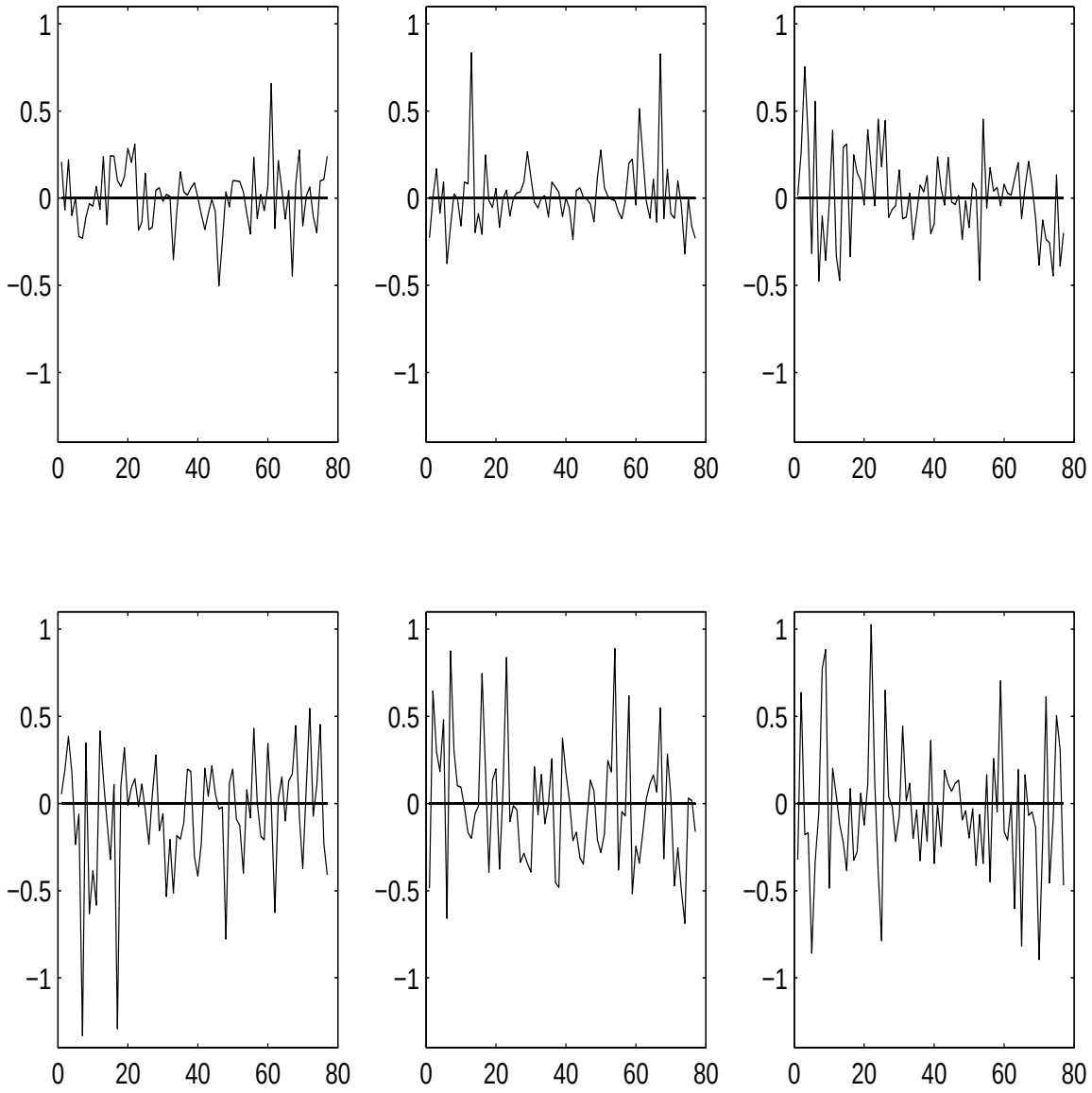

Figure 1: Log returns for six randomly chosen days of S&P500 trading data. The unit of the time axis is $\Delta = 5$min for all figures.

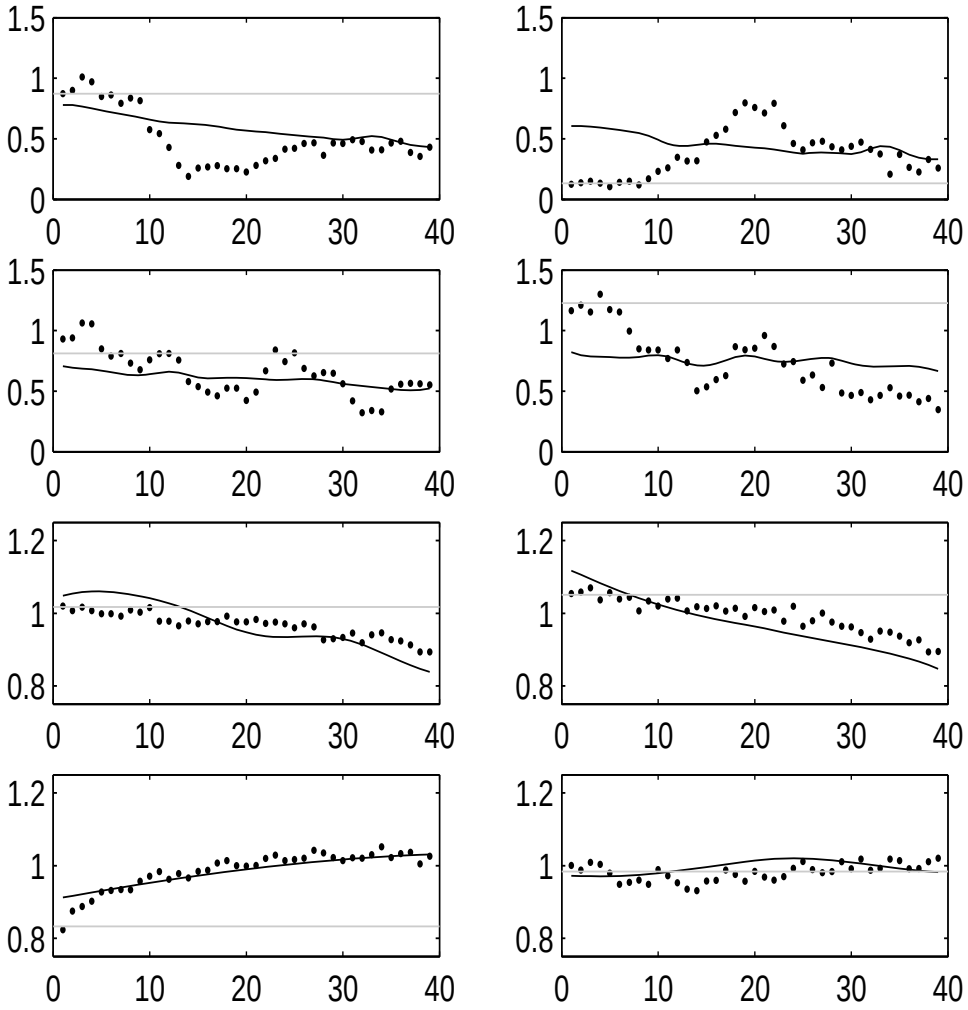

Figure 2: Predicted volatility trajectories (solid) and observed volatilities (dots) for second half days, predicted from first half day with functional linear regression (25), for two randomly selected days (in left and right column). The prediction is obtained for four simulation models: Heston model with correlation $\rho = 0.2$ (first row), Heston model with correlation $\rho = 0.8$ (second row), smooth volatility model from simulation scenario 1 (third row), smooth volatility model from simulation scenario 2 (fourth row). The functional volatility prediction is compared with a simple forecast, which is indicated by the gray horizontal line, and is based on the last observed volatility of the first half day.

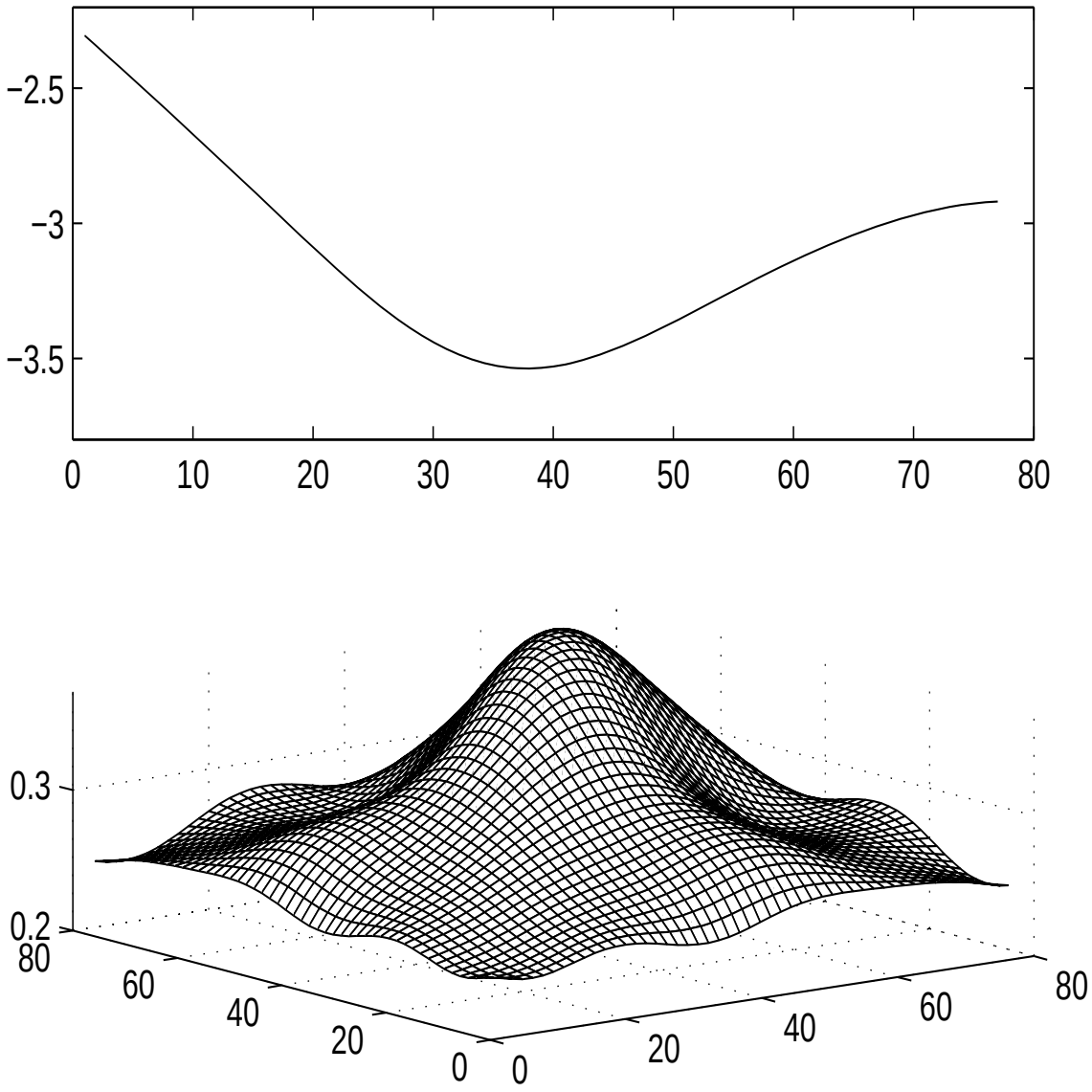

Figure 3: Estimated mean function $\hat{\mu}_V$ (top) and estimated covariance surface $\hat{G}_V$ (bottom) for the functional volatility process V for S&P500 data.

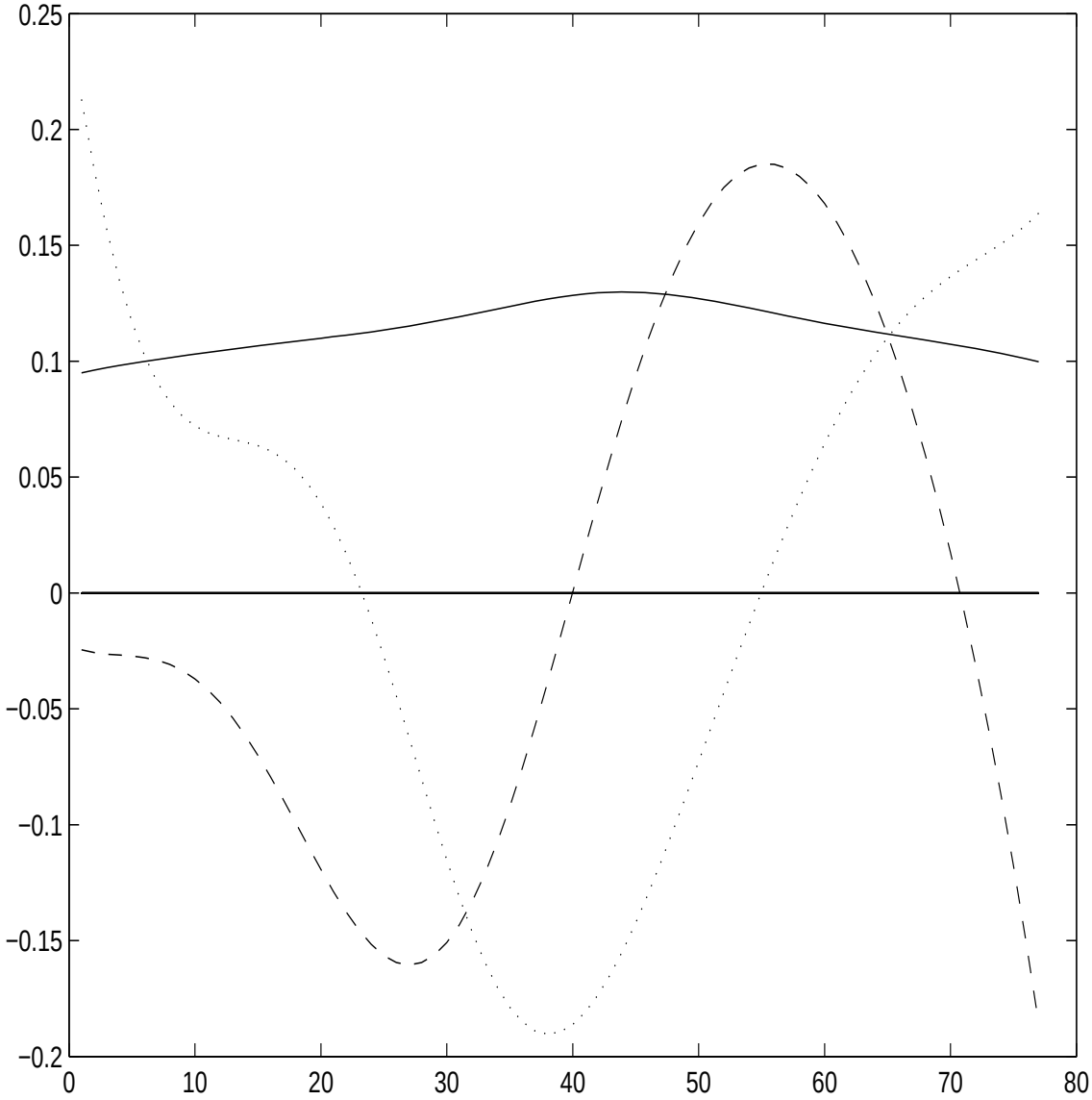

Figure 4: First three estimated eigenfunctions $\hat{\phi}_1$ (solid line), $\hat{\phi}_2$ (dashed line), and $\hat{\phi}_3$ (dotted line) of the functional volatility process V, explaining $82.48\%, 8.31\%$ and 6.87% of total variance, respectively, for S&P500 data.

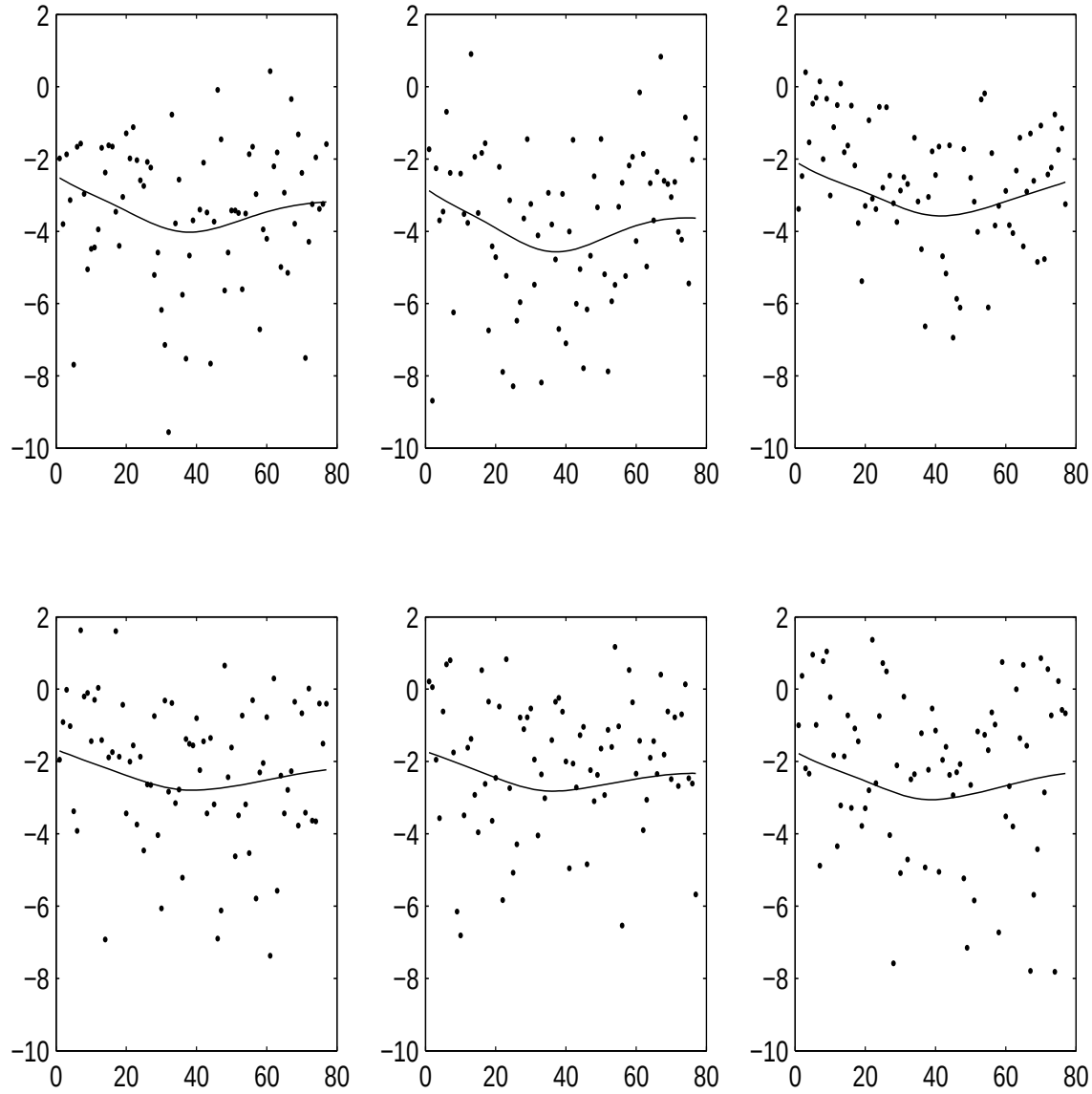

Figure 5: Raw volatilities $Y_{ij\Delta}$ (dots) and estimated volatility trajectories $\hat{V}_i$, for six days of S&P500 trading data.

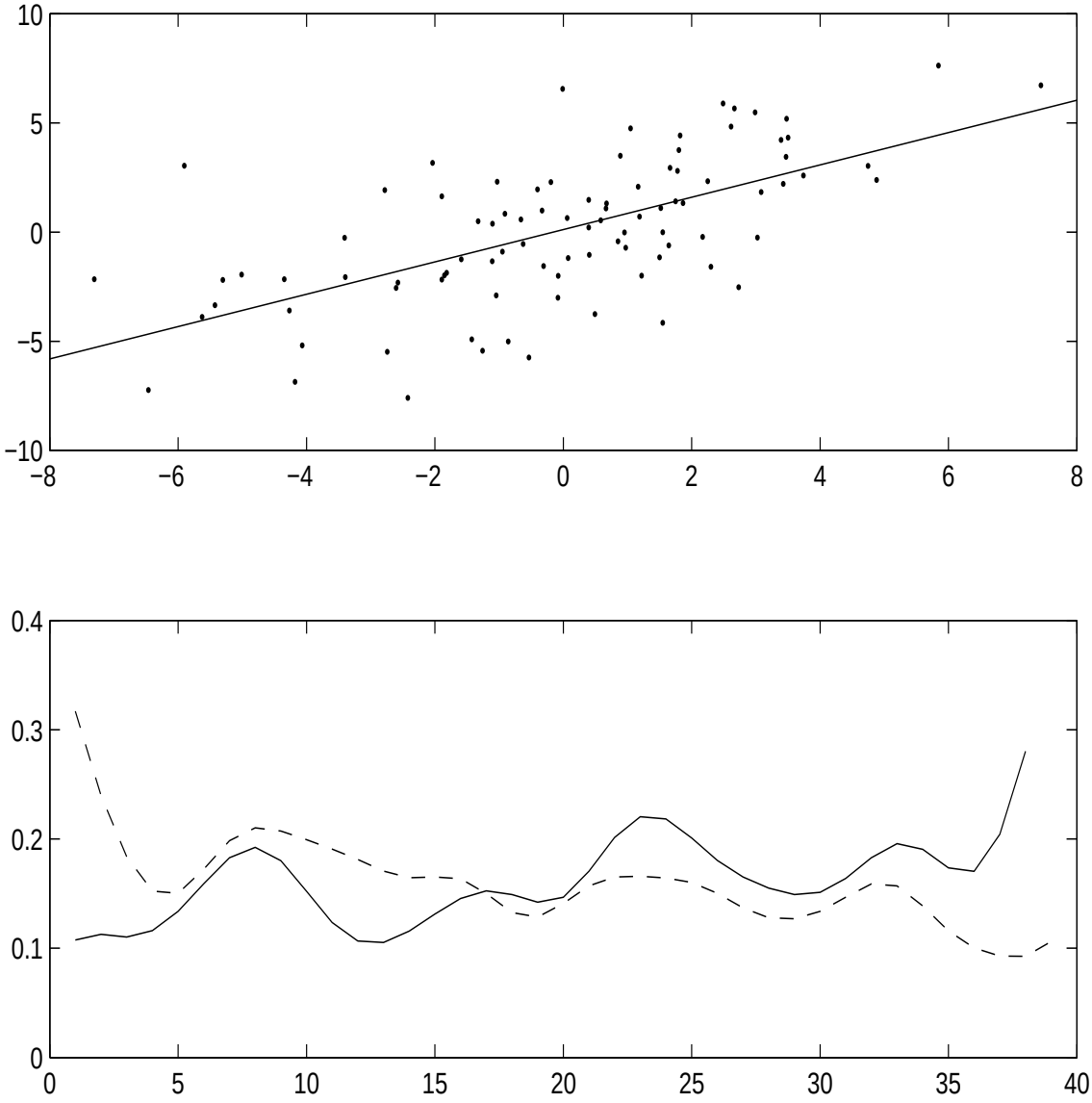

Figure 6: Scatterplot of estimated FPC scores of the afternoon volatility process (on y-axis) versus those of the morning volatility process (on x-axis), for the S&P500 data (top) and first eigenfunctions of morning (solid) and afternoon (dashed) processes.

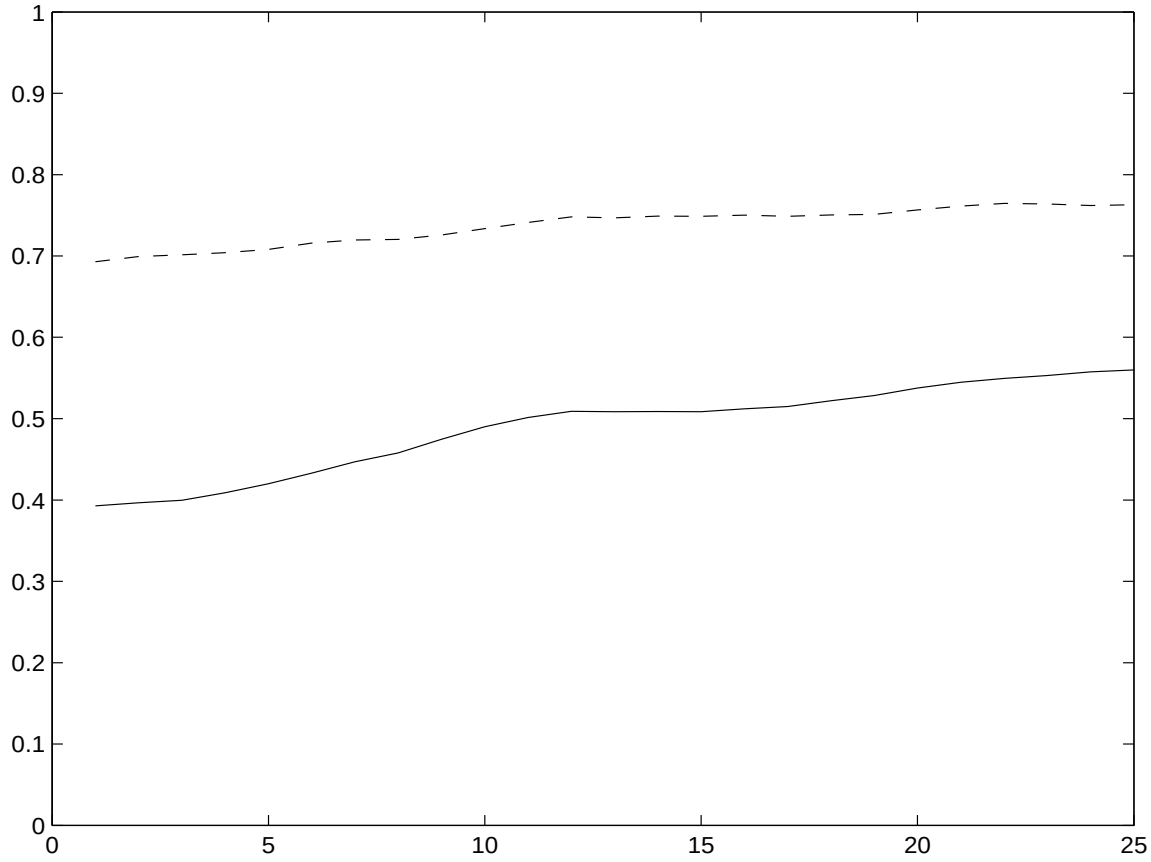

Figure 7: Average squared errors in dependency on time t for predictors based on functional regression (solid) and the overall mean (dashed).