International Journal of Forecasting 34 (2018) 75–88

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Dynamics of financial returns densities: A functional approach applied to the Bovespa intraday index

Eduardo Horta *, Flavio Ziegelmann

Universidade Federal do Rio Grande do Sul, Department of Statistics, 9500 Bento Gonçalves Av., 43–111, Porto Alegre, RS, 91509-900, Brazil

ARTICLE INFO

Keywords:
Stochastic processes
Autocovariance
Dimension reduction
High frequency data
Volatility forecasting

ABSTRACT

We model the stochastic evolution of the probability density functions (PDFs) of Ibovespa intraday returns over business days, in a functional time series framework. We find evidence that the dynamic structure of the PDFs reduces to a vector process lying in a two-dimensional space. Our main contributions are as follows. First, we provide further insights into the finite-dimensional decomposition of the curve process: it is shown that its evolution can be interpreted as a dynamic dispersion-symmetry shift. Second, we provide an application to realized volatility forecasting, with a forecasting ability that is comparable to those of HAR realized volatility models in the model confidence set framework.

1. Introduction

The adequate specification of the distribution of financial assets' prices or returns is a particularly relevant topic in the statistical modeling of financial data. A wide range of models have been introduced in the literature over recent decades, with the aim of accommodating the stylized features of financial returns distributions. These models include the ARCH and GARCH models introduced by Bollerslev (1986) and Engle (1982), respectively. We refer the reader to Mikosch, Kreiß, Davis, and Andersen (2009) for a very comprehensive review. However, a crucial difficulty that arises in this context is the fact that most of the aforementioned models are designed to describe the time-path dynamics of the returns (often driven by a latent volatility process), which usually requires certain restrictions to be imposed on the underlying distributions — for instance, that they belong to a parametric family — in order to make inference possible.

The present paper introduces a modeling approach that seeks to capture a different type of information that may potentially exist in the data, by relaxing the specification of the time-path dynamics while allowing for a greater flexibility of the underlying conditional distributions. Thus, rather than falling under the same framework as the aforementioned models, our approach is more likely to complement them, and may enrich traders' analysis toolkits. The approach that we take here is distinct because our analysis focuses on the dynamics of the returns' underlying conditional distributions, seeing the PDFs of intraday returns as a sequence of random variables that take values on a function space; that is, as a latent functional time series. Following Bathia, Yao, and Ziegelmann (2010), we adopt an essentially model-free environment in which the underlying density process evolves autonomously, giving rise to the observable returns process according to a specific conjugation property — see Eq. (1). Our aim in so doing is to establish an alternative setting for the modeling, estimation and forecasting of asset returns' probability density functions.

From a methodological point of view, our approach lies at the intersection of functional data analysis (FDA) with functional time series and nonparametric statistics. In recent years, the theory of estimation and inference when the observed data pertain to function spaces has received an increasing amount of attention from researchers from a wide spectrum of academic disciplines; see for

* Corresponding author.
E-mail address: eduardo.horta@ufrgs.br (E. Horta).

http://dx.doi.org/10.1016/j.ijforecast.2017.08.001
0169-2070/

instance the collection edited by Dabo-Niang and Ferraty (2008) for a discussion of recent developments and many applications. Unfortunately this blossom is not yet as widespread in the fields of economics and finance — not to say merely incipient. For example, Benko, Härdle, and Kneip (2009) provide an application to implied volatility estimation, and the cornerstone monograph by Ramsay and Silverman (1998) presents a thorough treatment of the topic. From a theoretical point of view, functional data can be viewed as realizations of function-valued random variables. Formal treatments of random elements that take values in the Hilbert and Banach spaces are provided by Bosq (2000), Ledoux and Talagrand (1991) and Vakhania, Tarieladze, and Chobanyan (1987). A more general theory considers random elements in metric spaces; see the classic texts by Billingsley (1999), Parthasarathy (2005) and van der Vaart and Wellner (1996). An approach that blends theory and applications is provided by Ferraty and Vieu (2006). Interestingly, most of the literature on FDA has dealt with the case in which the functional data are supposed to be independent realizations of a random function, with the case where the random functions display a dynamic dependence — that is, the case of a sequence of function-valued, nonindependent random variables — not attracting consideration until very recently. Bosq (2000) provides a good presentation on the theory of linear processes of such objects; see for example Damon and Guillas (2005) for further developments.

One technique that is central to FDA is that of principal components analysis. Such methodologies, the foundations of which lie in the Karhunen-Loève Theorem, seek a decomposition of the observed functions as a sum of orthogonal projections onto a suitable orthonormal basis that corresponds to the eigenfunctions of a covariance operator. However, as was pointed out by Hall and Vial (2006), if the observed functional data are imprecise (due to rounding, experimental measurement errors, non-observability, etc.) then so is the estimator of the covariance operator, which poses a major methodological problem for the application of principal components analysis to these observed data. One possible way to overcome such difficulties is to impose a condition in which the measurement errors vanish as the overall sample size goes to infinity. For example, Petersen and Müller (2016) show that, for a sample of density estimates, the covariance function of the latent density process can be estimated consistently as long as the individual sample sizes go to infinity together with the overall sample size. However, when dealing with a sample of density estimates in a time series framework — for instance, kernel density estimates of the conditional density of an asset's intra-day 5 min returns — the individual sample sizes are essentially fixed, removing the possibility of using the density estimates to estimate the true covariance function.

The methodology developed by Bathia et al. (2010), which we will be following throughout this paper, relies instead on the dynamic structure of the curve process as a way of filtering the noise from the observed functions, and of finding an appropriate orthogonal basis — related to the *lagged* covariance functions — which spans the linear space to which the curves pertain. This is an entirely original approach, in that it does not need to rely on the stronger assumption that the measurement errors would vanish in the face of a large sample. In terms of its implementation, the method reduces to the eigenanalysis of a finite-dimensional matrix, and the modeling of both the dynamic structure of the PDFs and prediction procedures can be carried out through traditional, and computationally less expensive, multivariate time series methods.

The following section presents this methodology in detail, after which Section 3 applies this methodology to intraday Ibovespa data. In particular, we find evidence that the dynamic structure of Ibovespa returns' PDFs lies in a two-dimensional subspace, and thus reduces to a $\mathbb{R}^2$ vector process whose evolution is shown to affect the dispersion and symmetry of returns distributions sequentially. We generate one-step-ahead density forecasts, and also provide an application to realized volatility forecasting, demonstrating a forecasting ability that is comparable to those of HAR realized volatility models in the model confidence set framework. Section 4 concludes.

2. Methodology

Let $f_1, f_2, \ldots$ be a sequence of *random densities*, and let $r_{1t}, \ldots, r_{n_t,t}$ denote the n_t observations of a financial asset's 5 min return process within day t. We consider the model

$$r_{it}|\mathscr{F} \quad \sim \quad f_t, \tag{1}$$

where $\mathscr{F}$ denotes the σ-algebra generated by $(f_t : t = 1, 2, \ldots)$. Eq. (1) says that, conditional on $\mathscr{F}$, the financial returns $r_{1t}, \ldots, r_{n_t,t}$ share the same marginal density within each day t, allowing these densities to evolve stochastically from day to day. It is convenient to consider the densities f_t as *random elements* in the Hilbert space $L^2 := L^2(I)$ of square integrable functions defined on a compact interval $I \subset \mathbb{R}$, equipped with inner product $\langle f, g\rangle := \int_I f(x)\, g(x)\, dx$ for all $f, g \in L^2$. Randomness of the f_t may also be interpreted in a Bayesian sense, in which case expressions such as $\mathbb{P}[f_t \in B]$ are understood as *prior* probabilities. In this context, the terminology can sometimes become confusing; for instance, when one says that r_{it} follows the distribution f_t (this is true for $r_{it}|\mathscr{F}$ only). The reader should be careful to distinguish between *conditional* and *unconditional* statements.

Now, the true densities f_t are not observable in applications, and the statistician only has access to a sample of estimates $g_1, \ldots, g_n$, obtained through some nonparametric method applied to the data $\{r_{it}\}$, for example. The densities g_t are taken to satisfy

$$g_t(x) = f_t(x) + \varepsilon_t(x), \qquad x \in I, \tag{2}$$

where ε_t is assumed to be noise, in the sense that (i) $\mathbb{E}(\varepsilon_t(x)) = 0$ for all t and all $x \in I$; (ii) $\mathrm{Cov}(\varepsilon_t(x), \varepsilon_{t+k}(y)) = 0$ for all $x, y \in I$ provided that $k \neq 0$; and (iii) $\mathrm{Cov}(f_t(x), \varepsilon_s(y)) = 0$ for all $x, y \in I$ and all t, s. These conditions can be interpreted as saying that the error in estimating f_t is intrinsic to day t and exogenous with respect to f_t. Note that the requirement $\mathbb{E}(\varepsilon_t(x)) = 0$ is a strong one, since many density estimators are biased, such as when g_t is a kernel density estimator of f_t, for example.

In any case, this issue can be overcome by substituting $\mathbb{E}(g_t \mid \mathscr{F})$ for f_t in Eq. (2) and considering the former as the time series of interest. In the above, since both g_t and f_t are probability density functions, we also have $\int \varepsilon_t(x)\,dx = 0$ for all t. Note also that, in principle, it is generally not necessary to assume that $r_{it}|\mathscr{F}$, $i = 1, \ldots, n_t$, is an iid sample: all that is required is that estimation of f_t be feasible, and hence, a condition such as ergodicity may suffice. However, it should be noted that the asymptotic properties of the estimators introduced below are linked to the dependence structure of $r_{1t}, \ldots, r_{n_t t}$, as this structure influences the quality of g_t as an estimator of f_t.

Now assume that $f_1, f_2, \ldots$ is stationary, so that the mean-density and the kth lag covariance function of f_t, defined respectively by

$$\mu(x) := \mathbb{E}(f_t(x)), \tag{3}$$

and

$$C_k(x, y) := \mathrm{Cov}(f_t(x), f_{t+k}(y)), \tag{4}$$

do not depend on t. As was shown by Bathia et al. (2010), the following spectral representation holds, under standard assumptions, for all t:

$$f_t(x) = \mu(x) + \sum_{j=1}^{d} \eta_{tj}\psi_j(x), \tag{5}$$

where $\psi_1, \ldots, \psi_d$ are eigenfunctions of the positive kernel

$$M(x, y) := \sum_{k=1}^{p} \int C_k(x, z)\, C_k(y, z)\, dz, \tag{6}$$

and where $\eta_{tj} := \langle f_t - \mu, \psi_j \rangle$ are centered, scalar random variables. Here, d is a positive integer called the *dimension* of the curve process (f_t), and we assume without loss of generality that $\{\psi_1, \ldots, \psi_d\}$ is an orthonormal set in L^2.

The motivation for introducing the kernel M is technical and relates to a fundamental difficulty that arises when one considers the estimation of the zero-lag covariance structure of random curves which are observed imperfectly; see Bathia et al. (2010) for a discussion. The importance of the induced representation in Eq. (5) is that it implies that the dynamic features of the curve sequence $f_1, f_2, \ldots$ are captured entirely by the d-dimensional time series $\boldsymbol{\eta}_t := (\eta_{t1}, \ldots, \eta_{td})$. The primary goal of the methodology is to identify d and to estimate the dynamic space $\mathcal{M}$ that is spanned by the eigenfunctions $\psi_1, \ldots, \psi_d$. In the present context, it is also relevant to acquire insights about the time series $\boldsymbol{\eta}_t$, to provide a meaningful interpretation of the eigenfunctions ψ_j, and to understand how they interact dynamically to give rise to the latent densities f_t via Eq. (5).

2.1. Estimation

We now describe the estimation of the quantities of interest. For a fixed positive integer p, let

$$\widehat{M}(x, y) := \sum_{k=1}^{p} \int \widehat{C}_k(x, z)\, \widehat{C}_k(y, z)\, dz, \tag{7}$$

where

$$\widehat{C}_k(x, y) := \frac{1}{n-p} \sum_{t=1}^{n-p} (g_t(x) - \widehat{\mu}(x))\,(g_{t+k}(y) - \widehat{\mu}(y)), \tag{8}$$

and $\widehat{\mu}(x) := n^{-1}\sum_{t=1}^{n} g_t(x)$. From Proposition 2 of Bathia et al. (2010), the D nonzero eigenvalues of $\widehat{M}$, say $\widehat{\theta}_1, \ldots, \widehat{\theta}_D$, coincide with the nonzero eigenvalues of the $(n-p) \times (n-p)$ matrix

$$\widehat{\boldsymbol{M}} := \frac{1}{(n-p)^2} \sum_{k=1}^{p} \boldsymbol{G}_k \boldsymbol{G}_0, \tag{9}$$

where, for $k = 0, \ldots, p$, we define $\boldsymbol{G}_k$ to be the $(n-p) \times (n-p)$ matrix whose (t, s)th entry is equal to $\langle g_{t+k} - \widehat{\mu}, g_{s+k} - \widehat{\mu} \rangle$. Moreover, letting $\widehat{\boldsymbol{\psi}}_j := \left(\widehat{\boldsymbol{\psi}}_{j1}, \ldots, \widehat{\boldsymbol{\psi}}_{j,n-p}\right)$ denote an eigenvector of $\widehat{\boldsymbol{M}}$ that is associated with its jth largest nonzero eigenvalue $\widehat{\theta}_j$, we have that, for $x \in I$, the functions $\widetilde{\psi}_j$ that are defined by

$$\widetilde{\psi}_j(x) := \sum_{t=1}^{n-p} \widehat{\boldsymbol{\psi}}_{jt} \times (g_t(x) - \widehat{\mu}(x)) \tag{10}$$

satisfy the identity $\int \widehat{M}(x, y)\, \widetilde{\psi}_j(y)\, dy = \widehat{\theta}_j \widetilde{\psi}_j(x)$. In other words, $\widetilde{\psi}_j$ is an eigenfunction of $\widehat{M}$ that is associated with its jth largest nonzero eigenvalue. This leads to a filtered estimator for f_t, defined by

$$\widehat{f}_t(x) := \widehat{\mu}(x) + \sum_{j=1}^{\widehat{d}} \widehat{\eta}_{tj} \widehat{\psi}_j(x), \tag{11}$$

where

$$\widehat{\eta}_{tj} := \langle g_t - \widehat{\mu}, \widehat{\psi}_j \rangle, \tag{12}$$

and where $\left\{\widehat{\psi}_1, \ldots, \widehat{\psi}_D\right\}$ denotes the orthonormal set of eigenfunctions of $\widehat{M}$ that are obtained by applying a Gram–Schmidt algorithm to the functions $\left\{\widetilde{\psi}_1, \ldots, \widetilde{\psi}_D\right\}$. In the above, $\widehat{d} \le D$ is an estimator of d that is determined through some statistical test, as discussed below. The consistency of the proposed estimators is established in Theorems 1, 2 and 3 of Bathia et al. (2010).

2.2. Bootstrap procedure

In spite of M having exactly d nonzero eigenvalues, the number of nonzero eigenvalues of $\widehat{M}$ may at times be much larger. Thus, D may not be a sensible estimator of d. Bathia et al. (2010) propose a bootstrap test for determining $\widehat{d}$, which we will present briefly. Let (θ_j) be the eigenvalues of M, repeated according to multiplicity. Say that the true dimension of $\mathcal{M}$ is d_0; that is, $\theta_1 \ge \cdots \ge \theta_{d_0} > 0$ and $\theta_{d_0+j} = 0$ for all $j \ge 1$. Then, we wish to reject the null hypothesis that $\theta_{d_0} = 0$, and to fail to reject the hypothesis that $\theta_{d_0+1} = 0$. If the null is stated as

$$H_0 : \theta_{d_0+1} = 0, \tag{13}$$

where d_0 is a known integer that is obtained for example by a visual inspection of the graph of the estimated eigenvalues $\widehat{\theta}_1 \ge \widehat{\theta}_2 \ge \cdots \ge 0$ of $\widehat{M}$, then we reject H_0 if $\widehat{\theta}_{d_0+1} > \ell_\alpha$,

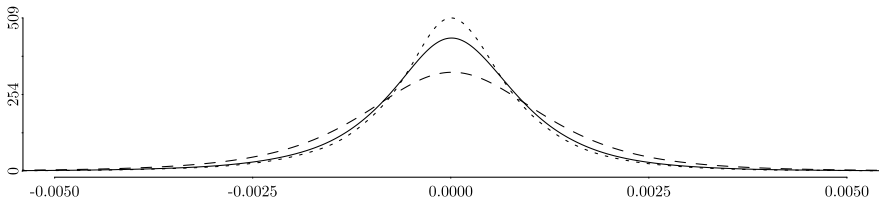


Fig. 1. Estimated mean-density $\widehat{\mu}$ using bandwidths $\widehat{h}_t$ (solid), $0.5\widehat{h}_t$ (dotted) and $2\widehat{h}_t$ (dashed).

where ℓ_α is the critical value at the $\alpha \in (0, 1)$ significance level. The following bootstrap procedure allows this critical value to be evaluated.

B1. Define $\widehat{\varepsilon}_t(x) = g_t(x) - \widehat{f}_t(x)$, with $\widehat{d}$ in Eq. (11) being set equal to d_0.

B2. Generate bootstrap observations $g_t^{[b]}$ that are defined by

$$g_t^{[b]}(x) := \widehat{f}_t(x) + \varepsilon_t^{[b]}(x), \qquad t = 1, \ldots, n$$

where b varies in $\{1, \ldots, B\}$ for some large integer B, and where the $\varepsilon_t^{[b]}$s are random, independent draws (with replacement) from $\{\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_n\}$.

B3. For each b, form a matrix $\widehat{\boldsymbol{M}}_{[b]}$ in the same way as $\widehat{\boldsymbol{M}}$ but with $g_t^{[b]}$ in place of g_t, and let $\widehat{\theta}_{d_0+1}^{[b]}$ be the (d_0+1)th largest eigenvalue of $\widehat{\boldsymbol{M}}_{[b]}$.

Now the conditional distribution of $\widehat{\theta}_{d_0+1}^{[b]}$ given observations $(g_1, \ldots, g_n)$ is taken as the distribution of $\widehat{\theta}_{d_0+1}$ under H_0, such that

$$\frac{1}{B}\sum_{b=1}^{B} \mathbb{I}_{\left[\widehat{\theta}_{d_0+1}, +\infty\right)}\left(\widehat{\theta}_{d_0+1}^{[b]}\right)$$

is taken to be the probability of obtaining an estimate that is equal to or greater than $\widehat{\theta}_{d_0+1}$ when θ_{d_0+1} is equal to zero. Thus, we reject the null whenever this magnitude is equal to or less than α.

The following section applies the methodology presented here to the 5 min returns series of the Bovespa index.

3. Data analysis

Our application is inspired by the similar approach of Bathia et al. (2010). Our data set consists of intraday Bovespa *logarithmic* stock indexes over $n = 305$ business days from September 1st, 2009, to November 6th, 2010. All data were kindly made available by Capse Investimentos. The tick-by-tick series is sampled at 5 min intervals, such that we have a sample $(p_{it},\ i = 1, \ldots, n_t + 1)$ for each day $t = 1, \ldots, n$, where $n_t + 1$ is the number of log-price observations within day t. We define the ith 5 min return on day t as

$$r_{it} := p_{i+1,t} - p_{it}, \qquad i = 1, \ldots, n_t. \tag{14}$$

Note that the overnight return is not computed in our definition. We assume throughout that Eq. (1) holds.

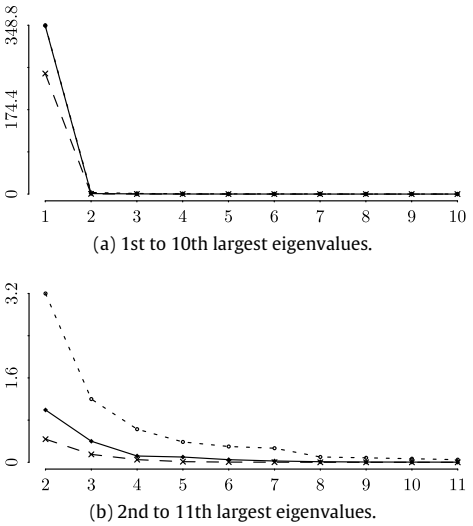
Now denote the kernel density estimate of the 5 min return PDF on day t by g_t. That is, for $x \in I$, define g_t as

$$g_t(x) := \frac{1}{n_t h_t}\sum_{i=1}^{n_t} K\left(\frac{x - r_{it}}{h_t}\right), \qquad t = 1, \ldots, n, \tag{15}$$

where $K(x) = \left(\sqrt{2\pi}\right)^{-1}\exp\left(-x^2/2\right)$ is a Gaussian kernel and h_t is a bandwidth. For the sake of rigor, we must mention that the method would require us to truncate the support of the densities g_t on some compact interval I and then normalize them so that $\int g_t = 1$. However, such corrections are of no consequence computationally. We use three scalings of Silverman's rule of thumb bandwidth $\widehat{h}_t = 1.06\widehat{\sigma}_t n_t^{-1/5}$, where $\widehat{\sigma}_t$ denotes the sample standard deviation of r_{it} $(i = 1, \ldots, n_t)$; that is, we set h_t in Eq. (15) equal to $0.5\widehat{h}_t$, $\widehat{h}_t$ and $2\widehat{h}_t$. Fig. 1 plots the estimated mean-density $\widehat{\mu}$ using these three levels of smoothing.

We now apply the methodology discussed in the previous section to the PDFs obtained. All of the computational work in this paper was carried out using the R statistical package. We set $p = 6$ in Eq. (8). The eigenvalues of $\widehat{\boldsymbol{M}}$ are displayed in Fig. 2. It can be seen that $\widehat{\theta}_1$ is much larger than $\widehat{\theta}_2$, and that this disparity does not depend on the bandwidth selected. Although this graphical inspection would in itself recommend setting $\widehat{d} = 1$, we evaluate the true dimension of $\mathcal{M}$ further by executing the bootstrap test presented in Section 2, with $B = 10,000$. We sequentially test the null hypotheses that $\theta_2 = 0$ and that $\theta_3 = 0$. Table 1 reports the p-values obtained. These results suggest that the particular choice of bandwidth does not have a major impact on the decisions implied. For instance, we can reject the null that $\theta_2 = 0$ at the $\alpha = 0.05$ significance level, while we cannot reject the null that $\theta_3 = 0$ at the $\alpha = 0.15$ level, whatever bandwidth is being considered. We also apply the Ljung–Box portmanteau test to each of the time series $\widehat{\eta}_{t1}$, $\widehat{\eta}_{t2}$, $\widehat{\eta}_{t3}$ and $\widehat{\eta}_{t4}$, obtaining p-values that are (to four decimal places) equal to 0.0000, 0.0003, 0.0015 and 0.2148, respectively. These results suggest that there is a lot of dynamic structure in the two-dimensional subspace corresponding to the eigenvalues θ_1 and θ_2, and perhaps also θ_3, but possibly none in the remaining directions. We therefore set $\widehat{d} = 2$.

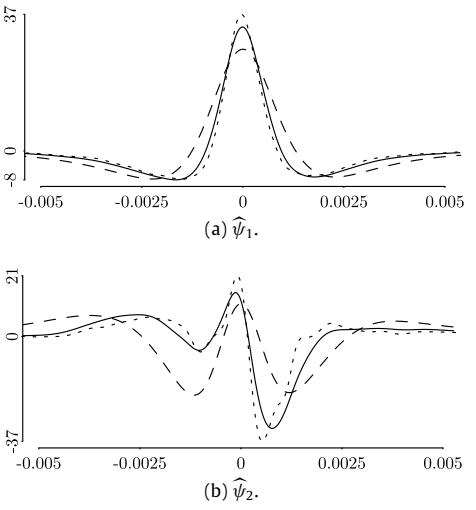
Fig. 3 displays the estimated eigenfunctions $\widehat{\psi}_1$ and $\widehat{\psi}_2$, and Fig. 4 plots the time series paths of the estimated loadings $\widehat{\eta}_{t1}$ and $\widehat{\eta}_{t2}$ for $t = 1, \ldots, 60$. These figures suggest that the choice of bandwidth has only a mild influence on the overall shapes of the eigenfunctions and the sample paths of the discrete scalar time series. With this in mind, we restrict the remainder of our analysis to the 5 min return PDFs built with the bandwidth set equal to $\widehat{h}_t$.



(a) 1st to 10th largest eigenvalues.

(b) 2nd to 11th largest eigenvalues.

Fig. 2. Estimated eigenvalues using bandwidths $\widehat{h}_t$ (solid), $0.5\widehat{h}_t$ (dotted) and $2\widehat{h}_t$ (dashed).



(a) $\widehat{\psi}_1$.

(b) $\widehat{\psi}_2$.

Fig. 3. Estimated eigenfunctions $\widehat{\psi}_1$ and $\widehat{\psi}_2$ using bandwidths $\widehat{h}_t$ (solid), $0.5\widehat{h}_t$ (dotted) and $2\widehat{h}_t$ (dashed).

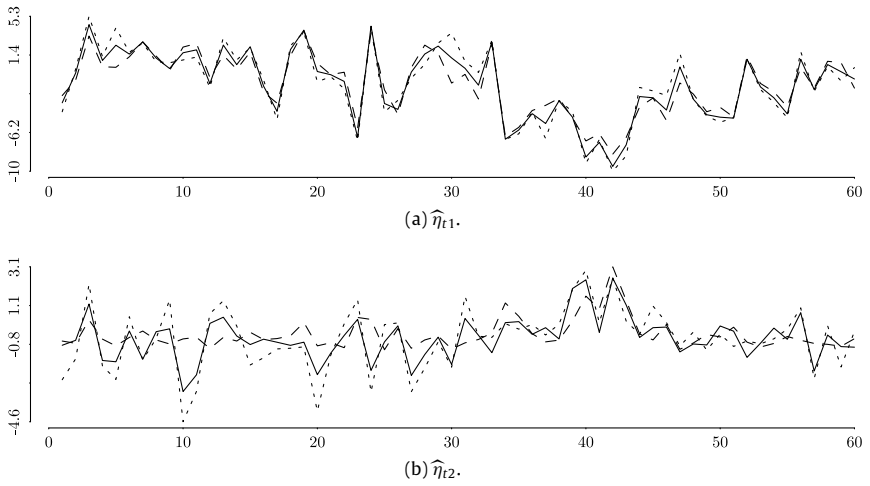
Table 1
Bootstrap test p-values.

	$h_t = \widehat{h}_t$	$h_t = 0.5\widehat{h}_t$	$h_t = 2\widehat{h}_t$
$H_0: \theta_2 = 0$	0.0362	0.0130	0.0269
$H_0: \theta_3 = 0$	0.1674	0.3435	0.1619

Notes: The reported values correspond to the proportions of $(d_0 + 1)$th eigenvalues generated under the null hypothesis through bootstrap loops that are greater than the estimated eigenvalue $\widehat{\theta}_{d_0+1}$.

Fig. 5 displays the nonparametric density estimates g_t and the filtered estimates $\widehat{f}_t$ that correspond to the first four business days in the sample. The estimates are corrected such that the functions obtained satisfy the conditions required by a PDF, namely that $\widehat{f}_t \geq 0$ and $\int \widehat{f}_t = 1$. This is done by setting all of the negative points of $\widehat{f}_t$ equal to zero and normalizing the resulting function. While these corrections may seem rather arbitrary, they are in fact minimal, since the negative points that appear are close to zero, and the estimated functions integrate nearly to one. This draws attention to the fact that the nonparametric and filtered estimates are similar in shape, with the filtered ones being slightly smoother.

Fig. 6 plots the mean-density $\widehat{\mu}$ and the eigenfunctions $\widehat{\psi}_1$ and $\widehat{\psi}_2$. The figure has a straightforward yet meaningful interpretation: for each t we have that, by definition, $\widehat{f}_t$ is formed by adding a weighted sum of the two eigenfunctions to the mean-density, with the weights being the random scalars $\widehat{\eta}_{t1}$ and $\widehat{\eta}_{t2}$. Thus, the filtered PDF $\widehat{f}_t$ can be seen as a random, dynamic deformation of $\widehat{\mu}$, with the shape-deforming structure being given by the eigenfunctions and the degree of deformation by $\widehat{\boldsymbol{\eta}}_t$. In particular, it appears from Fig. 6 that for positive values of $\widehat{\eta}_{t1}$, the eigenfunction



(a) $\widehat{\eta}_{t1}$.

(b) $\widehat{\eta}_{t2}$.

Fig. 4. Estimated loadings $\widehat{\eta}_{t1}$ and $\widehat{\eta}_{t2}$ for $t = 1, \ldots, 60$, using bandwidths $\widehat{h}_t$ (solid), $0.5\widehat{h}_t$ (dotted) and $2\widehat{h}_t$ (dashed).

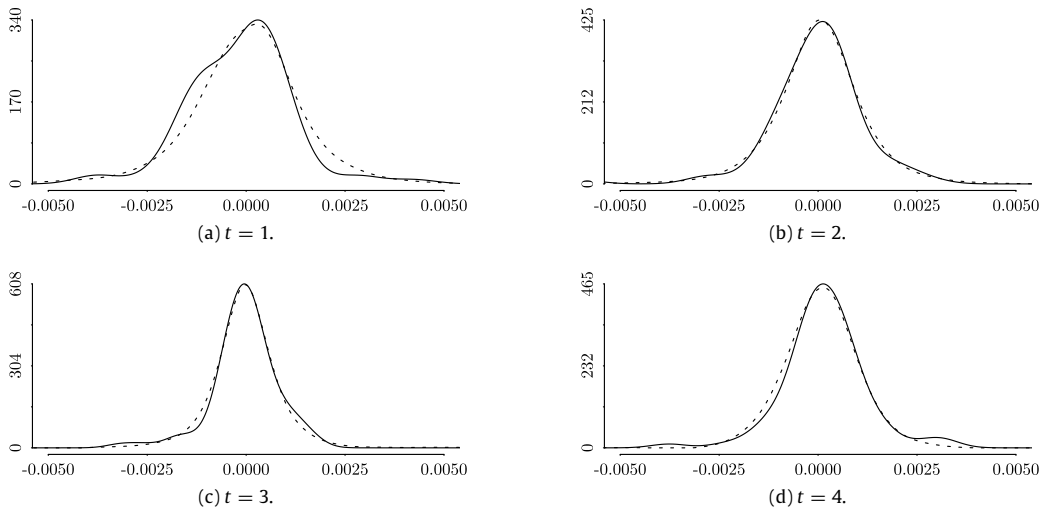


Fig. 5. Nonparametric density estimates g_t (solid) and filtered estimates $\widehat{f}_t$ (dotted), $t = 1, \ldots, 4$.

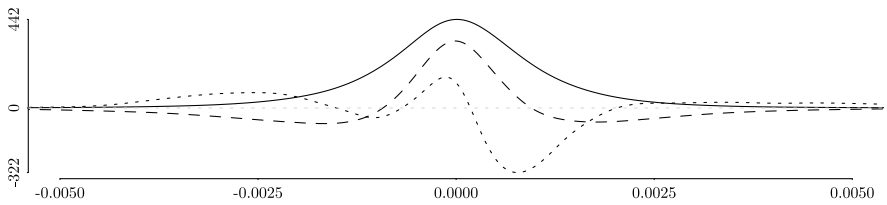


Fig. 6. The estimated mean-density $\widehat{\mu}$ (solid), and the eigenfunctions $\widehat{\psi}_1$ (dashed) and $\widehat{\psi}_2$ (dotted). The eigenfunctions are rescaled by a factor of 10 for better visualization.

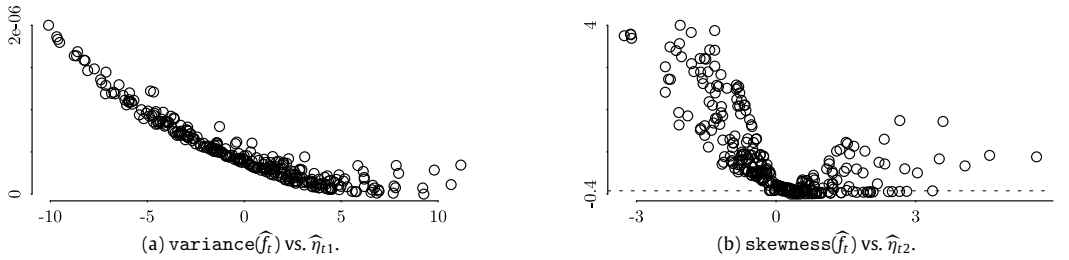
$\widehat{\psi}_1$ symmetrically adds probability mass for values close to the mode of $\widehat{\mu}$ and takes probability mass from the tails, with the converse being true for negative values of $\widehat{\eta}_{t1}$. Likewise, for positive values of $\widehat{\eta}_{t2}$, the eigenfunction $\widehat{\psi}_2$ seems to add little probability mass for values that are very close to the mode of $\widehat{\mu}$, to take probability mass asymmetrically from values to the left and right of this mode (with this asymmetry being heavier towards the right), and to add probability mass asymmetrically to the tails of the distribution. The reasoning for negative values of $\widehat{\eta}_{t2}$ is entirely analogous. Thus, $\widehat{\psi}_1$ represents a dispersion-shift, whilst $\widehat{\psi}_2$ amounts to a symmetry-shift over $\widehat{\mu}$.

Following this reasoning, it should be expected that positive (negative) values of $\widehat{\eta}_{t1}$ would be associated with distributions $\widehat{f}_t$ that show lower (higher) dispersions; and in the same sense, negative (positive) values of $\widehat{\eta}_{t2}$ should be associated with distributions $\widehat{f}_t$ that show positive (negative) skewness. Let $\texttt{mean}\big(\widehat{f}_t\big) := \int x\widehat{f}_t(x)\,dx$, and define $\texttt{variance}\big(\widehat{f}_t\big)$ and $\texttt{skewness}\big(\widehat{f}_t\big)$ similarly. This reasoning is supported further by the associations shown in the scatterplots of $\big(\widehat{\eta}_{t1},\ \texttt{variance}\big(\widehat{f}_t\big)\big)$ and $\big(\widehat{\eta}_{t2},\ \texttt{skewness}\big(\widehat{f}_t\big)\big)$, for $t = 1, \ldots, n$, which are displayed in Fig. 7.

We now consider the modeling of the time series $\widehat{\boldsymbol{\eta}}_t$. As can be seen in our application to the realized volatility below, the data generating process (DGP) governing the time evolution of $\widehat{\boldsymbol{\eta}}_t$ is tied to the DGP governing the conditional (daily) variance process in many instances, and therefore more sophisticated models may be necessary for capturing its dynamics fully. First, let us consider fitting a standard *VAR*(q) model to the data.

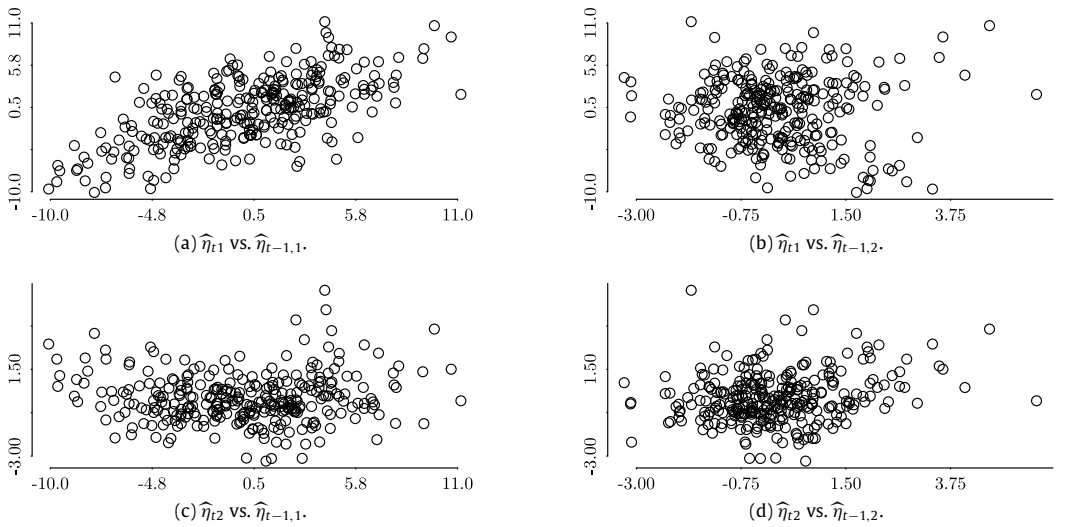
Fig. 8 displays the scatterplots of $\widehat{\eta}_{t1}$ and $\widehat{\eta}_{t2}$ against their lagged values. Panel (a) appears to show a linear relationship between $\widehat{\eta}_{t1}$ and $\widehat{\eta}_{t-1,1}$. Regarding the conditional expectation of $\widehat{\eta}_{t2}$ given $\widehat{\eta}_{t-1,1}$, panel (c) indicates the existence of a nonlinear, perhaps quadratic, relationship, albeit not a very strong one. The scatterplots of $\widehat{\eta}_{t1}$ against $\widehat{\eta}_{t-1,2}$, and of $\widehat{\eta}_{t2}$ against $\widehat{\eta}_{t-1,2}$ (panels (b) and (d), respectively), are inconclusive. We perform an augmented Dickey-Fuller test on the time series $\widehat{\eta}_{t1}$ and $\widehat{\eta}_{t2}$ in order to test the null hypothesis of the presence of a unit root, and obtain p-values that are virtually zero for both series regardless of the specification used, whether with a drift component, a drift and a trend component, or neither. We also perform a Priestley-Subba Rao test on these series in order to test the null of a time invariant Fourier spectrum, and obtain p-values of 0.8792934 and 0.3009572 respectively. Thus, we take $\widehat{\boldsymbol{\eta}}_t$ to be stationary.

Fig. 9 displays the ACF and PACF plots for $\widehat{\boldsymbol{\eta}}_t$. We fit a VAR($q$) model to the data $\widehat{\boldsymbol{\eta}}_1, \ldots, \widehat{\boldsymbol{\eta}}_n$, setting $q = 3$ based on the mean squared one-step-ahead forecast error cross-validation criterion (over 100 forecasts). The p-values of the F test for the overall significance of the two regression equations are nearly zero and 0.03304, respectively. Individually, at the 0.05 significance level, only three regression coefficients have significant t-statistics, namely (i) the coefficient of $\widehat{\eta}_{t-1,1}$ on $\widehat{\eta}_{t,1}$, with a value of 0.546; (ii) the



(a) variance($\widehat{f}_t$) vs. $\widehat{\eta}_{t1}$. (b) skewness($\widehat{f}_t$) vs. $\widehat{\eta}_{t2}$.

Fig. 7. Scatterplots of (a) variance($\widehat{f}_t$) vs. $\widehat{\eta}_{t1}$, and (b) skewness($\widehat{f}_t$) vs. $\widehat{\eta}_{t2}$. The dotted line in panel (b) indicates zero on the vertical axis.



(a) $\widehat{\eta}_{t1}$ vs. $\widehat{\eta}_{t-1,1}$. (b) $\widehat{\eta}_{t1}$ vs. $\widehat{\eta}_{t-1,2}$.

(c) $\widehat{\eta}_{t2}$ vs. $\widehat{\eta}_{t-1,1}$. (d) $\widehat{\eta}_{t2}$ vs. $\widehat{\eta}_{t-1,2}$.

Fig. 8. Scatterplots of $\widehat{\boldsymbol{\eta}}_t$.

coefficient of $\widehat{\eta}_{t-3,1}$ on $\widehat{\eta}_{t,1}$, with a value of 0.146; and (iii) the coefficient of $\widehat{\eta}_{t-1,2}$ on $\widehat{\eta}_{t,2}$, which is equal to 0.202.

Our first illustration concerns the problem of forecasting the curves f_t. In this setting it will be convenient, despite abusing the notation a little, to denote the full sample size by $N = 305$, and to write $\widehat{\mu}_{|n}$, $\widehat{\psi}_{j|n}$, $\widehat{\eta}_{tj|n}$ and $\widehat{f}_{t|n}$ (where $j = 1, 2$) to denote the estimates obtained from a subsample $g_1, \ldots, g_n$, with $n \leq N$. Thus, for instance, $\widehat{\mu}_{|N} = \widehat{\mu}$ in the previous notation. For the sake of generality, consider the following model. Assuming that $\boldsymbol{\eta}_t$ is stationary, we can write

$$\boldsymbol{\eta}_t = \alpha\left(\boldsymbol{\eta}_{t-1}, \ldots, \boldsymbol{\eta}_{t-q}\right) + \boldsymbol{u}_t, \tag{16}$$

where $\alpha = (\alpha_1, \alpha_2) : \mathbb{R}^{2q} \to \mathbb{R}^2$ is the regression function of $\boldsymbol{\eta}_t$ on its q lagged values, and $\mathbb{E}\left(\boldsymbol{u}_t \mid \boldsymbol{\eta}_{t-1}, \ldots, \boldsymbol{\eta}_{t-q}\right) = \mathbf{0}$ (or perhaps median $\left(u_{tj} \mid \boldsymbol{\eta}_{t-1}, \ldots, \boldsymbol{\eta}_{t-q}\right) = 0$, for $j = 1, 2$). Note that if α is differentiable (and setting $q = 1$ for simplicity), then, for $1 \leq t \leq n$, a straightforward calculation yields

$$\widehat{\boldsymbol{\eta}}_{t|n} = \alpha\big(\widehat{\boldsymbol{\eta}}_{t-1|n}\big) + \boldsymbol{u}_t + O_{\mathbb{P}}\left(\rho_{t|n}\right),$$

where $\rho_{t|n} = \max\{\|\widehat{\boldsymbol{\eta}}_{t|n} - \boldsymbol{\eta}_t\|,\ \|\widehat{\boldsymbol{\eta}}_{t-1|n} - \boldsymbol{\eta}_{t-1}\|\}$, and where it is understood that the asymptotics in the $O_{\mathbb{P}}$ symbol are in n. Hence, $\widehat{\boldsymbol{\eta}}_t$ inherits the dynamic behavior of $\boldsymbol{\eta}_t$ up to a negligible remainder. We then let $\widehat{\alpha}_{|n} = \big(\widehat{\alpha}_{1|n}, \widehat{\alpha}_{2|n}\big)$ denote an estimator of α obtained by fitting the data $\widehat{\boldsymbol{\eta}}_{1|n}, \ldots, \widehat{\boldsymbol{\eta}}_{n|n}$, with $n \leq N$, and write

$$\widehat{f}_{n+1|n}(x) = \widehat{\mu}_{|n}(x) + \sum_{j=1}^{2} \widehat{\alpha}_{j|n}\big(\widehat{\boldsymbol{\eta}}_{n|n}, \ldots, \widehat{\boldsymbol{\eta}}_{n-q+1|n}\big)\widehat{\psi}_{j|n}(x), \quad x \in I. \tag{17}$$

For the sake of simplicity, we stick to the VAR(3) model. We obtain the OLS estimate of the linear transformation α by fitting the data $\big(\widehat{\boldsymbol{\eta}}_{1|n}, \ldots, \widehat{\boldsymbol{\eta}}_{n|n}\big)$, then generate one-step-ahead predictions $\widehat{f}_{n+1|n}$ using Eq. (17), letting n range from $n_0 := N - 100$ to $N - 1$. Fig. 10 summarizes the performances of our forecasts. Since the *ex post* densities f_{n+1} are not observable, we consider two distinct proxies, namely the kernel density estimates g_{n+1} and the filtered estimates $\widehat{f}_{n+1|n+1}$. Here, we describe only the contents of panels (a), (c) and (e), which correspond to the case in which g_{n+1} are used as proxies. Panels (b), (d) and (f) are interpreted analogously, only with g_{n+1} replaced by $\widehat{f}_{n+1|n+1}$. Panel (a) displays the best forecasts according to both the L^2 and the supremum norms, whereas panel (c) plots the worst forecasts according to the same criteria. The forecasts are shown as dotted lines, and the *ex post* densities as solid lines. Panel (e) then displays the whole

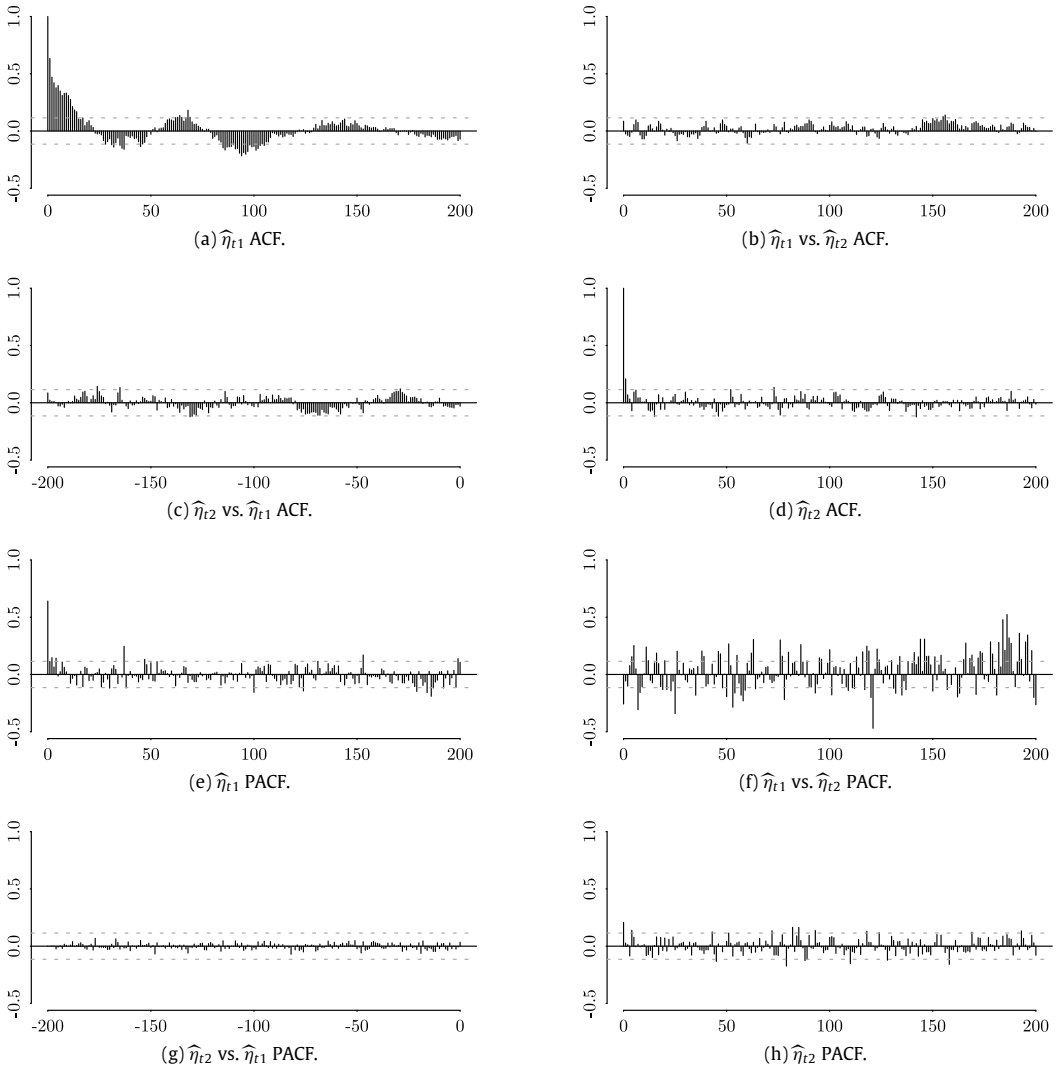


Fig. 9. The ACF and PACF of $\widehat{\boldsymbol{\eta}}_t$.

set of forecasting errors (shaded in gray), together with the empirical forecasting bias (solid black line). Although there is some variation in the forecasts, it can be seen that the proposed forecasting strategy is nearly unbiased overall, but with a mild positive bias for points in the domain near zero.

Note that, according to the model in Eq. (1), the forecasted densities are an estimate of the distribution of tomorrow's 'typical' 5 min return. If one is interested in obtaining an estimate of tomorrow's daily return distribution, then one need only scale up the forecasted density via a calculation similar to Eq. (18) below.

3.1. Realized volatility forecasting

Over recent years, realized volatility has become one of the most prominent areas of research in the statistical modeling of high-frequency financial data. The theoretical grounds for this were established by Andersen, Bollerslev, Diebold, and Labys (2001, 2003), and concurrently by Barndorff-Nielsen and Shephard (2002). Much of the recent work in the field has focused on to the study of microstructure noise — the fact that observed prices seldom correspond to efficient prices — and the way in which this noise affects the estimation of return variances through the realized volatility (see for instance Aït-Sahalia, Mykland, and Zhang, 2011; Bandi and Russell, 2011; Ghysels and Sinko, 2011; and Hansen and Lunde, 2006). For a comprehensive review of the subject, see McAleer and Medeiros (2008); and see also Barunik, Krehlik, and Vacha (2016).

We shall proceed as follows: let r_t denote the day t Ibovespa return, that is, $r_t := p_{n_t+1,t} - p_{1,t} = \sum_{i=1}^{n_t} r_{it}$ in the previous notation.[1] We are interested in forecasting the conditional variance $\varsigma_t^2 := \mathrm{Var}\,(r_t|\mathscr{F})$. Recall that this quantity is not observable, not even *ex post*; hence, we will use the realized volatility $\widehat{\varsigma}_t^2 := \sum_{i=1}^{n_t} r_{it}^2$ as a proxy.

[1] Notice that we are considering the daily return as being intrinsic to each day t: the overnight return $p_{1,t} - p_{n_{t-1},t-1}$ is not added to r_t.

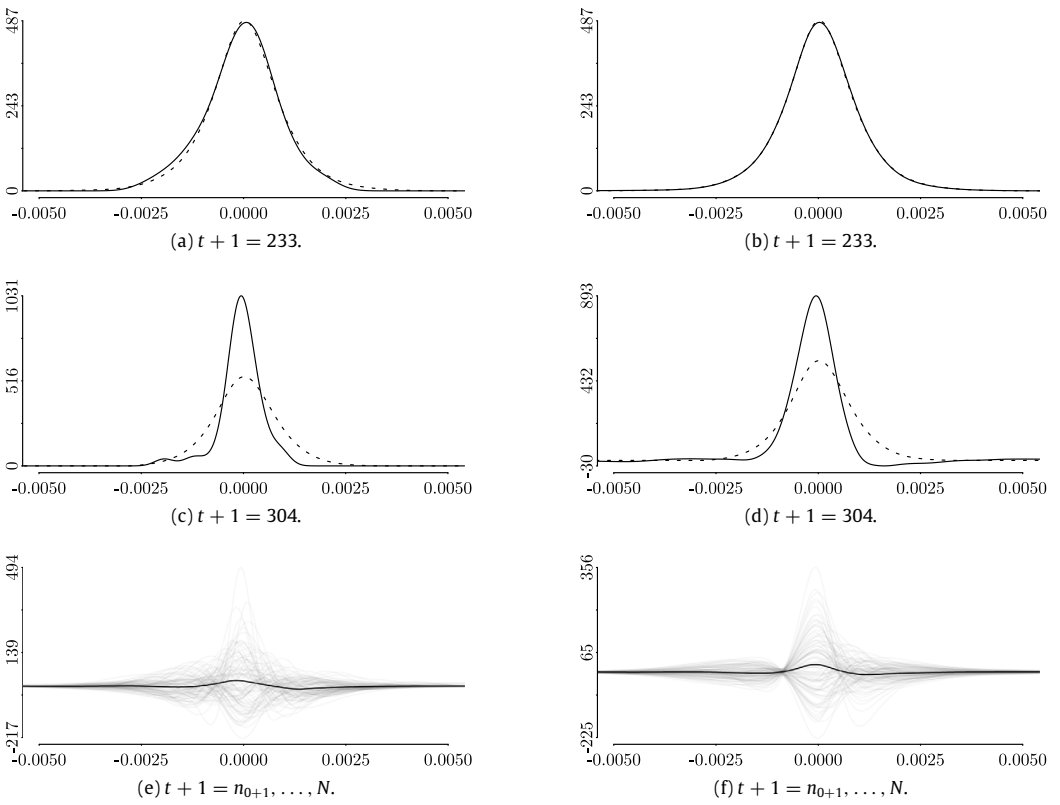


Fig. 10. Panels (a) and (c) (resp. (b) and (d)): (dotted) PDF forecasts $\widehat{f}_{t+1|t}$ and (solid) ex post densities g_{t+1} (resp. $\widehat{f}_{t+1|t+1}$). Panels (a) and (b) (resp. (c) and (d)) correspond to the best (resp. worst) L^2 prediction. Panel (e) (resp. (f)): whole set of forecasting errors (gray) and forecasting bias (black) w.r.t. g_{t+1} (resp. $f_{t+1|t+1}$).

Letting $\sigma_t^2 := \mathrm{Var}\,(r_{1t}|\mathscr{F})$, and assuming that intraday returns are serially uncorrelated (conditional on $\mathscr{F}$), we have

$$\begin{aligned}\varsigma_t^2 &= \mathrm{Var}\left(\sum_{i=1}^{n_t} r_{it}\,\middle|\,\mathscr{F}\right)\\ &= \sum_{i=1}^{n_t}\mathrm{Var}\,(r_{it}|\mathscr{F})\\ &= n_t\sigma_t^2.\end{aligned} \tag{18}$$

Now, via Eqs. (1) and (5), a straightforward calculation yields the identity

$$\sigma_t^2 = \beta_0 + \beta_1\eta_{t1} + \beta_2\eta_{t2} + \beta_3\eta_{t1}^2 + \beta_4\eta_{t2}^2 + \beta_5\eta_{t1}\eta_{t2}, \tag{19}$$

where the coefficients β_j are functions of the first and second moments of μ, ψ_1 and ψ_2. Therefore, the stochastic dynamic behavior of σ_t^2 in the population is determined entirely by $\boldsymbol{\eta}_t$. Also, if the innovations $\boldsymbol{u}_t$ are assumed to be stationary, it is easy to establish, via Eq. (16) (and again assuming $q=1$ for simplicity), that the equality

$$\begin{aligned}\sigma_t^2 &= \beta_0' + \beta_1\alpha_1\left(\boldsymbol{\eta}_{t-1}\right) + \beta_2\alpha_2\left(\boldsymbol{\eta}_{t-1}\right)\\ &\quad+\beta_3\alpha_1\left(\boldsymbol{\eta}_{t-1}\right)^2 + \beta_4\alpha_2\left(\boldsymbol{\eta}_{t-1}\right)^2\\ &\quad+\beta_5\alpha_1\left(\boldsymbol{\eta}_{t-1}\right)\alpha_2\left(\boldsymbol{\eta}_{t-1}\right) + v_t\end{aligned} \tag{20}$$

holds, with $v_t := \sigma_t^2 - \mathbb{E}\left(\sigma_t^2\,|\,\boldsymbol{\eta}_{t-1}\right)$ being such that $\mathbb{E}\left(v_t\,|\,\boldsymbol{\eta}_{t-1}\right) = 0$, and where β_0' is a function of the first and second moments of μ, ψ_1, ψ_2 and $\boldsymbol{u}_t$.

The above expressions are useful because they provide explicit functional relationships between ς_t^2 and $\boldsymbol{\eta}_t$, and between ς_t^2 and $\boldsymbol{\eta}_{t-1}$, via Eq. (18), which can be useful when it comes to modeling these time series jointly. The scatterplots in Fig. 11 illustrate some of these relationships. Panel (a) displays the well-known linear relationship between $\log\left(\widehat{\varsigma}_t^2\right)$ and $\log\left(\widehat{\varsigma}_{t-1}^2\right)$, for the sake of exposition. Panel (b) shows the scatterplot of the points $\left(\log(\widehat{\varsigma}_t^2),\ \widehat{\eta}_{t1}\right)$, indicating a slight deviation from a linear relationship. Panel (c) shows what appears to be a linear relationship between $\log\left(\widehat{\varsigma}_t^2\right)$ and $\widehat{\eta}_{t-1,1}$. Panel (d) displays the scatterplot of $\widehat{\eta}_{t,2}$ vs. $\widehat{\varsigma}_t^2$, which indicates a nonlinear dependence between these variables. The reader should also contrast the ACF and PACF of $\log(\widehat{\varsigma}_t^2)$, which are displayed in Fig. 12, with the corresponding plots of $\widehat{\eta}_{t1}$.

We now evaluate the performances of one-step-ahead volatility forecasts in the model confidence set (MCS) framework. Given a set of candidate models $\mathfrak{M}_0$, the MCS procedure seeks to obtain a subset $\widehat{\mathfrak{M}} \subset \mathfrak{M}_0$ such that the best candidate model lies in $\widehat{\mathfrak{M}}$ with a certain confidence level. Here, the term *best* is understood in terms of a user-specified loss function; refer to Hansen, Lunde, and Nason (2011) for details. Now, based on the preceding discussion, it is legitimate to run regressions of $\widehat{\varsigma}_t^2$

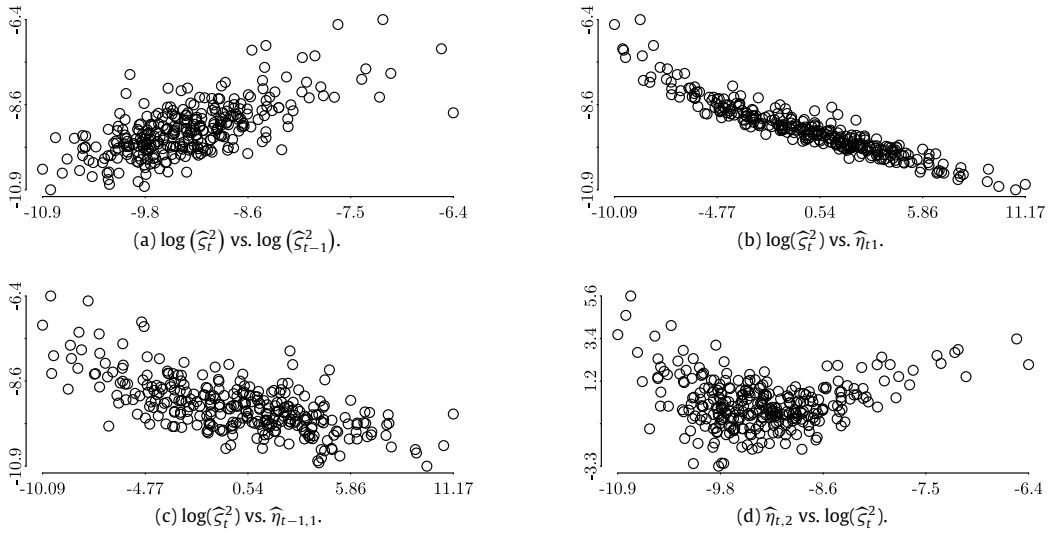


Fig. 11. Scatterplots of $\widehat{\varsigma}_t^2$ and $\widehat{\boldsymbol{\eta}}_t$.

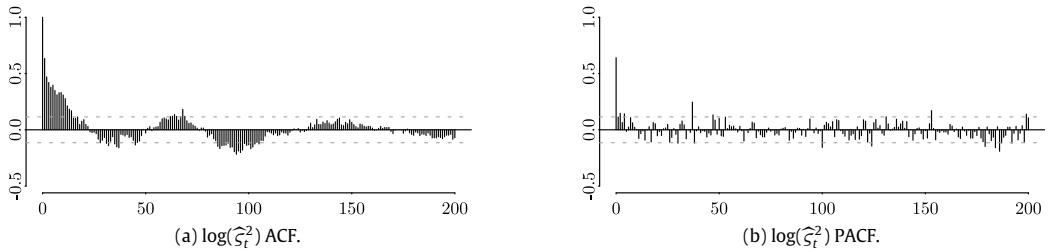


Fig. 12. The ACF and PACF of log$(\widehat{\varsigma}_t^2)$.

on (lagged values of) $\widehat{\boldsymbol{\eta}}_t$. The underlying regression model may be linear, polynomial, fully nonparametric, etc., and choosing any particular one might be difficult to justify. Thus, we adopt an exploratory approach, inspired partly by Eqs. (19) and (20) and partly by the scatterplots in Fig. 11, and propose a few strategies for generating the volatility forecasts.

These strategies can be divided into three groups. Group `Ben` consists of three benchmark strategies: two variations of the HAR-RV model (see Corsi, 2009; Corsi, Audrino, and Renò, 2012; and Corsi and Renò, 2012) and the well-established APARCH model of Ding, Granger, and Engle (1993). The strategies in Group `Reg` proceed by regressing $\log(\widehat{\varsigma}_t)$ directly on the lagged values $(\widehat{\eta}_{t-1,1}, \widehat{\eta}_{t-1,2})$, in both nonparametric and polynomial specifications, so that no forecasting is required. Group `Fcst` is based on a two-step approach: the volatility forecasts (second step) are based on forecasts $\widehat{\boldsymbol{\eta}}_{n+1|n}$ of the latent coefficient process (first step), obtained from one of two distinct models: the VAR(3) and an adaptation of the HAR model. In each of these strategies, the one-step-ahead forecasts are computed sequentially using the restricted data sets $g_1, \ldots, g_n$, with n running from $n_0 := N - 100$ to $N - 1$. That is, we use an expanding window forecast design. The details are given below.

Strategy `Ben1`. The underlying data generating process for the conditional variance is assumed to satisfy a log version of the HAR-RV model:

$$\log\left(\varsigma_t^2\right) = a_0 + a_1 \log\left(\varsigma_{t-1}^2\right) + a_2 \log\left(\varsigma_{t-1,\mathrm{w}}^2\right) + a_3 \log\left(\varsigma_{t-1,\mathrm{m}}^2\right) + \text{error}, \tag{21}$$

where $\varsigma_{t,\mathrm{w}}^2 = \left(\varsigma_t^2 + \varsigma_{t-1}^2 + \cdots + \varsigma_{t-4}^2\right)/5$ is a weekly component and $\varsigma_{t,\mathrm{m}}^2 = \left(\varsigma_t^2 + \varsigma_{t-1}^2 + \cdots + \varsigma_{t-21}^2\right)/22$ is a monthly component. The parameter estimates $\widehat{a}_{j|n}$ are obtained through a median regression fit to the data $\widehat{\varsigma}_1^2, \ldots, \widehat{\varsigma}_n^2$, and the forecast for date $n+1$ is defined as

$$\widehat{\varsigma}_{n+1|n,\mathtt{Ben1}}^2 := \exp\left\{\widehat{a}_{0|n} + \widehat{a}_{1|n} \log\left(\widehat{\varsigma}_n^2\right) + \widehat{a}_{2|n} \log\left(\widehat{\varsigma}_{n,\mathrm{w}}^2\right) + \widehat{a}_{3|n} \log\left(\widehat{\varsigma}_{t-1,\mathrm{m}}^2\right)\right\}, \tag{22}$$

where $\widehat{\varsigma}_{n,\mathrm{w}}^2$ and $\widehat{\varsigma}_{n,\mathrm{m}}^2$ are defined in the obvious manner.

Strategy `Ben2`. This is a generalization of Strategy Ben1 that includes jump and leverage components in the model in Eq. (21). See Corsi and Renò (2012) for details. The

estimation procedure was carried using the R package highfrequency.

Strategy Ben3. Here, $\widehat{\varsigma}^2_{n+1|n,\texttt{Ben3}}$ is a one-step-ahead volatility forecast obtained from an APARCH(1,1) model fitted to the data $r_1, \ldots, r_n$.

Strategy Reg1. The underlying data generating process is assumed to satisfy the regression equation

$$\log\left(\varsigma_t^2\right) = m\left(\boldsymbol{\eta}_{t-1}\right) + \text{error}, \tag{23}$$

where m is the regression function of ς_t^2 on $\boldsymbol{\eta}_{t-1}$. The regression function estimate, $\widehat{m}_{|n}$, is obtained via the npreg function in the R package np, using the data $\left(\widehat{\varsigma}_t^2,\ \widehat{\boldsymbol{\eta}}_{t-1|n}\right)$, $t = 2, \ldots, n$. The forecast for date $n+1$ is defined as

$$\widehat{\varsigma}^2_{n+1|n,\texttt{Reg1}} := \exp\left\{\widehat{m}_{|n}\left(\widehat{\boldsymbol{\eta}}_{n|n}\right)\right\}. \tag{24}$$

Strategy Reg2. The underlying data generating process is taken to satisfy the model in Eq. (23), with m now denoting a second-degree polynomial on two variables (in the same spirit as in Eq. (19)). Estimates of the coefficients of the polynomial m (yielding the estimate $\widehat{m}_{|n}$) are obtained via a median regression fitted to the data $\left(\widehat{\varsigma}_t^2,\ \widehat{\boldsymbol{\eta}}_{t-1|n}\right)$, $t = 2, \ldots, n$. The forecast $\widehat{\varsigma}^2_{n+1|n,\texttt{Reg2}}$ for date $n+1$ is defined in the same way as in Eq. (24).

Group Fcst. As has been mentioned, we shall consider two distinct models for obtaining forecasts $\widehat{\boldsymbol{\eta}}_{n+1|n}$ of the latent coefficient process. In addition to the VAR(3) model (which yields one-step-ahead forecasts $\widehat{\boldsymbol{\eta}}_{n+1|n,\texttt{V}}$), we also consider an adaptation of the HAR model, with the aim of obtaining a better description of the underlying dynamics. This is justified by the fact that, in many contexts, the (unobservable) processes ς_t^2 and $\boldsymbol{\eta}_t$ are essentially tied together. This is a consequence of Eq. (19), and is apparent from the scatterplots in Fig. 11, as well from a comparison of the ACF plots of both processes. With this observation in mind, we assume that the evolution of the process $\left(\boldsymbol{\eta}_t\right)$ is described by the following equations:

$$\begin{aligned}\eta_{t+1,1} &= \gamma_{01} + \gamma_{11}\eta_{t,1} + \gamma_{21}\eta^{\text{w}}_{t,1} \\ &\quad + \gamma_{31}\eta^{\text{m}}_{t,1} + \gamma_{41}\eta_{t,2} + \text{error} \\ \eta_{t+1,2} &= \gamma_{02} + \gamma_{12}\eta_{t,1} + \gamma_{42}\eta_{t,2} + \text{error},\end{aligned} \tag{25}$$

where $\eta^{\text{w}}_{t,1} := \left(\eta_{t,1} + \eta_{t-1,1} + \cdots + \eta_{t-4,1}\right)/5$ is a weekly component and $\eta^{\text{m}}_{t,1} := \left(\eta_{t,1} + \eta_{t-1,1} + \cdots + \eta_{t-21,1}\right)/22$ is a monthly component. We obtain estimates $\left\{\widehat{\gamma}_{ij|n}\right\}$ of the parameters $\left\{\gamma_{ij}\right\}$ through median regression fits to the data $\left(\widehat{\boldsymbol{\eta}}_{1|n}, \ldots, \widehat{\boldsymbol{\eta}}_{n|n,}\right)$. The corresponding one-step-ahead forecasts $\widehat{\boldsymbol{\eta}}_{n+1|n,\texttt{H}}$ are then obtained via a sample analogue of Eq. (25).

Strategies Fcst1V and Fcst1H. Motivated by Eq. (18), we set $\widehat{\varsigma}^2_{n+1|n,\texttt{Fcst1V}} := n_{n+1} \times \texttt{variance}\left(\widehat{f}_{n+1|n,\texttt{Fcst1V}}\right)$, where $\widehat{f}_{n+1|n,\texttt{Fcst1V}}$ is the one-step-ahead density forecast obtained using Eq. (17) with the VAR(3) forecasts. We define $\widehat{\varsigma}^2_{n+1|n,\texttt{Fcst1H}}$ in the obvious manner.

Strategies Fcst2V and Fcst2H. The underlying data generating process is assumed to follow the regression equation

$$\log\left(\varsigma_t^2\right) = m\left(\boldsymbol{\eta}_t\right) + \text{error}, \tag{26}$$

where m is the regression function of ς_t^2 on $\boldsymbol{\eta}_t$. The regression function estimate, $\widehat{m}_{|n}$, is obtained via the npreg function in the R package np, using the data $\left(\widehat{\varsigma}_t^2,\ \widehat{\boldsymbol{\eta}}_{t|n}\right)$, $t = 1, \ldots, n$. The one-step-ahead forecast for date $n+1$ from Strategy Fcst2V is then defined as

$$\widehat{\varsigma}^2_{n+1|n,\texttt{Fcst2V}} := \exp\left\{\widehat{m}_{|n}\left(\widehat{\boldsymbol{\eta}}_{n+1|n,\texttt{V}}\right)\right\}. \tag{27}$$

The forecast $\widehat{\varsigma}^2_{n+1|n,\texttt{Fcst1H}}$ is defined in a similar fashion.

Strategies Fcst3V and Fcst3H. We assume that the data generating process is described by the model in Eq. (26), with m now denoting a second-degree polynomial on two variables. Estimates of the coefficients of the polynomial m (yielding the estimate $\widehat{m}_{|n}$) are obtained via a median regression that is fitted to the data $\left(\widehat{\varsigma}_t^2,\ \widehat{\boldsymbol{\eta}}_{t|n}\right)$, $t = 1, \ldots, n$. The forecasts for date $n+1$ that correspond to Strategies Fcst3V and Fcst3H are then defined analogously to Eq. (27).

In what follows, we use the realized volatility $\widehat{\varsigma}_t^2$ as a proxy for the conditional (daily) variance, as the latter is not observable *ex post*. Fig. 13 displays the boxplots of forecasting errors, corresponding to each of the strategies discussed, across the forecasting window $n_0 + 1, \ldots, N$. As confirmed by the MCS procedure (see below), the Strategy Ben1 appears to dominate the alternative strategies, whereas Ben3, Fcst1V and Fcst1H are seen to perform worse than the competing strategies.

We now construct 1% significance MCSs, using both the mean squared error (MSE) and QLIKE loss functions. Our choice of these loss functions reflects a focus on robustness, as was discussed by Patton (2011). If L denotes a loss function on two variables, the corresponding losses are computed (e.g. for Strategy Ben1), as

$$\frac{1}{N - n_0}\sum_{n=n_0}^{N-1} L\left(\widehat{\varsigma}_{n+1|n,\texttt{Ben1}},\ \widehat{\varsigma}_{n+1}\right).$$

Table 2 reports the outputs of the MCS procedure. The following quantities are reported: the MSE losses of each forecasting strategy relative to that of strategy Ben1 (which is equal to 1.074574e−09); the QLIKE losses of each strategy; the bias of each strategy (relative to Ben1), and their MCS ranks according to the two loss criteria under consideration (models excluded from the MCS are marked with an "a" symbol). The strategies from Group Reg are excluded from both MCSs. The same is true of Fcst1V and Fcst1H, which is interesting since these correspond to an approach that could be considered to be, in a sense, the most intuitive one, namely 'forecasting the coefficient time series and then recovering the variance through the resulting density forecasts'. The remaining strategies from Group Fcst performed nearly as well as the Ben1 strategy. Interestingly, the more general HAR-RV model with jumps and leverage (strategy Ben2) did not perform well, and was even excluded from the MCS according to the QLIKE criterion. We infer that this poor performance is due to the fact that the R package highfrequency estimates the parameters of this model using OLS, whereas we have proceeded via LAD estimation for strategy Ben1.

We conclude by pointing to the important fact that the latent time series $\widehat{\boldsymbol{\eta}}_t$ may be used to forecast not only the

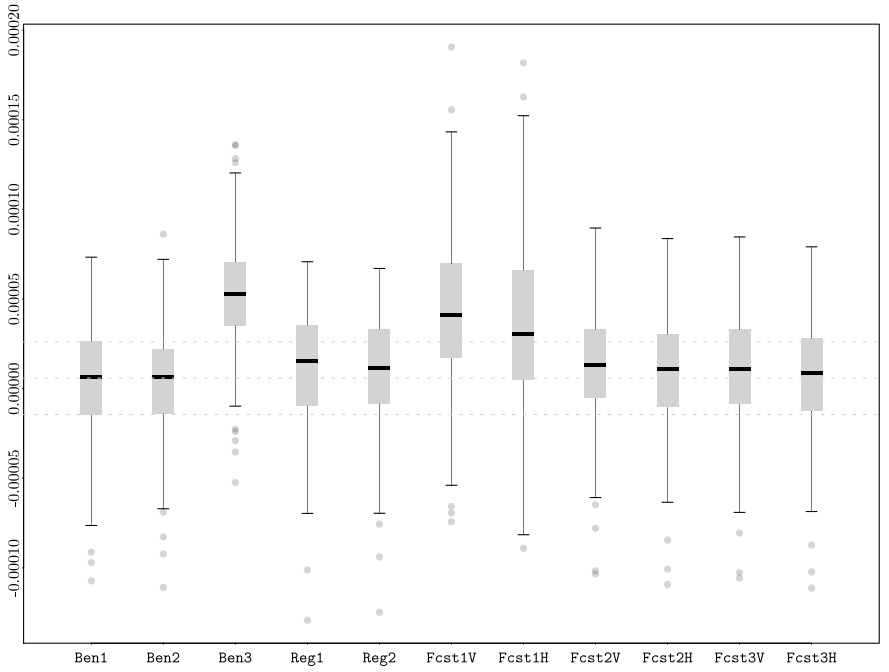


Fig. 13. Boxplot of the forecasting errors $\widehat{\varsigma}^2_{n+1|n} - \widehat{\varsigma}^2_n$ that correspond to each strategy, along the forecasting window $n_0 + 1, \ldots, N$.

Table 2
Mean squared errors (relative to Strategy `Ben1`), QLIKE losses, biases (relative to Strategy `Ben1`), and MCS ranks of the different forecasting strategies.

Strategy	Rel. MSE	QLIKE	Rel. Bias	MSE rank	QLIKE rank
`Ben1`	1.000	−8.609	1.000	1	1
`Ben2`	1.103	−8.596	1.625	6	[a]
`Ben3`	3.695	−8.475	36.370	[a]	[a]
`Reg1`	1.297	−8.573	7.771	[a]	[a]
`Reg2`	1.154	−8.595	6.457	[a]	[a]
`Fcst1V`	4.065	−8.502	31.382	[a]	[a]
`Fcst1H`	3.484	−8.532	25.225	[a]	[a]
`Fcst2V`	1.076	−8.600	5.654	4	4
`Fcst2H`	1.061	−8.601	4.219	3	2
`Fcst3V`	1.084	−8.600	4.397	5	5
`Fcst3H`	1.060	−8.600	2.376	2	3

Notes:
[a] In the rank columns means that the MCS procedure excludes the strategy from the model confidence set.

volatility, but also other quantities of interest. For instance, if one is interested in predicting the probability of losing money in one's first trade on day $t+1$, a trivial computation, letting $F_t(x) := \mathbb{P}[r_{1t} < x \,|\, \mathscr{F}] = \int_{-\infty}^{x} f_t(z)\, dz$, yields

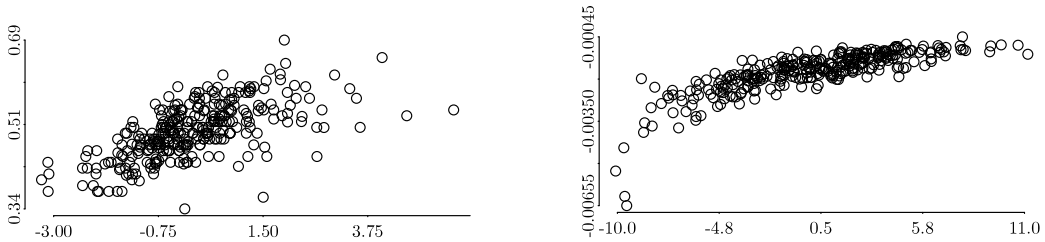
$$F_t(0) \quad = \delta_0 + \delta_1 \eta_{t1} + \delta_2 \eta_{t2}, \tag{28}$$

where $\delta_0 = \int_{-\infty}^{0} \mu(x)\, dx$, $\delta_1 = \int_{-\infty}^{0} \psi_1(x)\, dx$ and $\delta_2 = \int_{-\infty}^{0} \psi_2(x)\, dx$. Panel (a) of Fig. 14 displays the scatterplot of the estimated probability of r_{1t} being negative, $\widehat{F}_t(0)$, versus $\widehat{\eta}_{t2}$, where $\widehat{F}_t$ denotes the empirical cdf of the returns $r_{1t}, \ldots, r_{n_t,t}$. The dependence structure suggested by Eq. (28) is supported further, and this dependence can be exploited to propose forecasting strategies similar to those discussed earlier.

As another example, suppose now that one is interested in forecasting the Value-at-Risk. In this case, the quantity that one wishes to forecast is the τth quantile $Q_t(\tau)$ of $r_{1t}|\mathscr{F}$. Since $Q_t(\tau)$ is determined uniquely by the condition of being the smallest number q such that the equality $\tau = \int_{-\infty}^{q} \mu(x)\, dx + \eta_{t1} \int_{-\infty}^{q} \psi_1(x)\, dx + \eta_{t2} \int_{-\infty}^{q} \psi_2(x)\, dx$ holds, it is clear that the stochastic evolution of $Q_1(\tau), Q_2(\tau), \ldots$ is determined entirely by the 2-dimensional process $(\boldsymbol{\eta}_t)$. Panel (b) of Fig. 14 displays the scatterplot of the sample 0.05th quantiles $\widehat{Q}_t(0.05)$ against $\widehat{\eta}_{t1}$. Again, the dependence structure that exists between $\widehat{\boldsymbol{\eta}}_t$ and $\widehat{Q}_t(0.05)$ can be exploited to obtain forecasts through strategies similar to those discussed earlier.

4. Concluding remarks

This paper has followed the methodology developed by Bathia et al. (2010) for modeling the dynamics of Ibovespa intraday return PDFs over business days. We



(a) $\widehat{F}_t$ (0) vs. $\widehat{\eta}_{t2}$. (b) $\widehat{Q}_t$ (0.05) vs. $\widehat{\eta}_{t1}$.

Fig. 14. Scatterplots: (a) estimated probability of loss $\widehat{F}_t$ (0) vs. $\widehat{\eta}_{t2}$; (b) sample 0.05th quantiles $\widehat{Q}_t$ (0.05) vs. $\widehat{\eta}_{t1}$.

make two main contributions. First, we provide further insights into the finite-dimensional decomposition of the latent density process: it is shown that its evolution can be interpreted as a dynamic dispersion-symmetry shift. In terms of the Karhunen-Loève representation of the PDF curves, we find evidence that a two-dimensional vector process weights the deterministic eigenfunctions dynamically in such a way as to change the dispersion and symmetry of the mean-density sequentially; that is, the eigenfunctions may be seen as parametric shape-deforming structures which, each day, deform the mean distribution of the returns on that day, to a stochastic degree. Second, we provide an application of the methodology to realized volatility forecasting, and show a forecasting ability comparable to those of HAR realized volatility models in the MCS framework.

The aim of our approach is to overcome the well-known difficulty of incorporating empirical stylized features of financial asset returns and their distributions into statistical models, in an attempt to describe the dynamic behavior of these assets. Our proposed approach is novel, in that it focuses directly on modeling the dynamic structure of intraday asset return PDFs over business days, and promising, in that it may offer an enhancement for practical implementations relative to other PDF estimation methods and prediction models. Firstly, if the true PDFs actually fall into the proposed context, then the estimates obtained through nonparametric methods, for example, will not be optimal, since the dynamic structure of the curves is not taken into account. Secondly, in a forecasting context, this methodology potentially allows one to obtain richer information about the upcoming PDFs than would be possible when using GARCH models or similar, for example. This is because, by construction, such models restrict the families to which the PDFs may pertain, while our proposed PDF forecasts are not shape-constrained; this feature in turn allows more sophisticated estimates of the population moments and even of specific probabilities of interest to be obtained, which may enrich trading decision-taking procedures.

This introduces many exploratory possibilities. For instance, one potential field of inquiry is that of developing trading strategies based on the methodology applied in this paper. For example, the PDF forecasts described in this paper allow the trader to assess a rich body of information on the distributions of intraday returns of upcoming periods. With such information on hand, it is possible to evaluate the investment risk in a novel manner. Future work is likely to embrace trading simulations using such strategies.

In particular, it would be insightful to filter and standardize the returns process and to model the dynamics of the resulting density process. This approach would have the advantage of capturing higher-order factors better in the spectral representation of the densities, after the variance component (which dominates the dynamics of the density process) has been removed, thus allowing more sophisticated trading strategies to be implemented. It would also be interesting to compare the eigenfunctions ψ_j that correspond to different financial assets, as it may be possible to classify distinct assets according to the underlying number and shape of their eigenfunctions. Another, more academic, research possibility is that of generating simulations in the specific context of asset pricing theory, in order to assess further the properties of the theory presented here. It would be of particular interest to parametrize the distributions and then simulate the price processes according to a certain specification of the stochastic behavior of asset prices, in order to evaluate how the methodology works for estimating such underlying distributions based on observations of generated prices alone.

On theoretical grounds, the incorporation of restrictions on the functions f_t could represent a step forward in applications in which such restrictions are known to hold; for example, our application assumes f_t to lie in the convex subspace of L^2 that consists of those positive square integrable functions which are also integrable and the integral of which equals 1, even though its empirical counterparts $\widehat{f}_t$ could by construction assume negative values and not integrate to unity. The fact that the estimates obtained nearly satisfy the restrictions illustrates the strength of the methodology, but it would still be interesting in principle to restrict some characteristics of the estimators. One step in this direction is taken by Petersen and Müller (2016); a fundamental theoretical difficulty in this context is to overcome the requirement that intra-day sample sizes go to infinity together with the overall sample size. Also, establishing the (finite dimensional) distributions of the estimators $\widehat{\psi}_t$ — possibly via Monte Carlo simulation — is a necessary improvement, as it would permit one to test hypotheses about the parametric eigenfunctions. Another promising area of research is that of developing both theory and applications further to consider random functions with a more general domain and codomain. In principle, all that would be asked is for the f_ts to pertain to a Hilbert space so that a spectral decomposition makes sense; see Horta and Ziegelmann (2016). A natural extension of the work developed herein would be to apply the methodology to joint densities.

Acknowledgments

The author Flavio Ziegelmann wishes to thank CNPq (processes 305290/2012-6 and 485561/2013-1) for financial support.

References

Aït-Sahalia, Y., Mykland, P. A., & Zhang, L. (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*, *160*(1), 160–175 Elsevier.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, *96*(453), 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, *71*(2), 579–625. URL: http://doi.wiley.com/10.1111/1468-0262.00418.

Bandi, F. M., & Russell, J. R. (2011). Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations. *Journal of Econometrics*, *160*(1), 145–159. URL: http://dx.doi.org/10.1016/j.jeconom.2010.03.027.

Barndorff-Nielsen, O. E., & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *64*(2), 253–280.

Barunik, J., Krehlik, T., & Vacha, L. (2016). Modeling and forecasting exchange rate volatility in time-frequency domain. *European Journal of Operational Research*, *251*(1), 329–340. URL: http://linkinghub.elsevier.com/retrieve/pii/S0377221715011194.

Bathia, N., Yao, Q., & Ziegelmann, F. (2010). Identifying the finite dimensionality of curve time series. *The Annals of Statistics*, *38*(6), 3352–3386.

Benko, M., Härdle, W., & Kneip, A. (2009). Common functional principal components. *The Annals of Statistics*, *37*(1), 1–34. URL: http://projecteuclid.org/euclid.aos/1232115926.

Billingsley, P. (1999). *Wiley series in probability and statistics: Vol. 493. Convergence of probability measures*. John Wiley & Sons.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327. URL: http://linkinghub.elsevier.com/retrieve/pii/0304407686900631.

Bosq, D. (2000). *Lecture notes in statistics: Vol. 149. Linear processes in function spaces: Theory and applications*. New York: Springer-Verlag.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196.

Corsi, F., Audrino, F., & Renò, R. (2012). HAR modeling for realized volatility forecasting. In L. Bauwens, C. Hafner, & S. Laurent (Eds.), *Wiley handbooks in financial engineering and econometrics. Handbook of volatility models and their applications* (pp. 363–382). John Wiley & Sons, (Chapter 15).

Corsi, F., & Renò, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics*, *30*(3), 368–380. URL: http://www.tandfonline.com/doi/abs/10.1080/07350015.2012.663261.

Dabo-Niang, S., & Ferraty, F. (Eds.). (2008). *Contributions to statistics. Functional and operatorial statistics*. Physica-Verlag Heidelberg.

Damon, J., & Guillas, S. (2005). Estimation and Simulation of Autoregressive Hilbertian Processes with Exogenous Variables. *Statistical Inference for Stochastic Processes*, *8*(2), 185–204.

Ding, Z., Granger, C. W., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, *1*(1), 83–106. URL: http://linkinghub.elsevier.com/retrieve/pii/092753989390006D.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987.
URL: http://www.jstor.org/stable/1912773?origin=crossref.

Ferraty, F., & Vieu, P. (2006). *Springer series in statistics. Nonparametric functional data analysis: Theory and practice*. New York: Springer-Verlag.

Ghysels, E., & Sinko, A. (2011). Volatility forecasting and microstructure noise. *Journal of Econometrics*, *160*(1), 257–271.

Hall, P., & Vial, C. (2006). Assessing the finite dimensionality of functional data. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *68*(4), 689–705.

Hansen, P. R., & Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, *24*(2), 127–161.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Horta, E., & Ziegelmann, F. (2016). Identifying the spectral representation of Hilbertian time series. *Statistics & Probability Letters*, *118*, 45–49.

Ledoux, M., & Talagrand, M. (1991). *Classics in mathematics: Vol. 23. Probability in Banach spaces: Isoperimetry and processes*. Berlin Heidelberg: Springer-Verlag.

McAleer, M., & Medeiros, M. (2008). Realized volatility: A review. *Econometric Reviews*, *27*(1–3), 10–45.

Mikosch, T., Kreiß, J.-P., Davis, R. A., & Andersen, T. G. (Eds.). (2009). *Handbook of financial time series*. Berlin, Heidelberg: Springer Berlin Heidelberg,
URL: http://link.springer.com/10.1007/978-3-540-71297-8.

Parthasarathy, K. (2005). *Probability measures on metric spaces, Vol. 352*. Providence, Rhode Island: American Mathematical Society, URL: http://www.ams.org/chel/352.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, *160*(1), 246–256. URL: http://linkinghub.elsevier.com/retrieve/pii/S030440761000076X.

Petersen, A., & Müller, H.-G. (2016). Functional data analysis for density functions by transformation to a Hilbert space. *The Annals of Statistics*, *44*(1), 183–218. URL: http://projecteuclid.org/euclid.aos/1449755961.

Ramsay, J., & Silverman, B. W. (1998). *Springer series in statistics. Functional data analysis*. New York: Springer-Verlag.

Vakhania, N., Tarieladze, V., & Chobanyan, S. (1987). *Mathematics and its applications: Vol. 14. Probability distributions on Banach spaces*. Netherlands: Springer.

van der Vaart, A., & Wellner, J. (1996). *Springer series in statistics. Weak convergence and empirical processes*. New York: Springer-Verlag.

Eduardo Horta is an Adjoint Professor and Early Career Researcher in the Department of Statistics at the Federal University of Rio Grande do Sul, Brazil. He holds a Ph.D. in Economics from the same University (2015), having spent a season as Visitor Graduate Student (Sandwich Doctorate) at Queen Mary University of London (2013). His research interests include functional time series, empirical process theory, Bayesian nonparametrics and quantile regression. Eduardo has a published paper in *Statistics & Probability Letters*.

Flavio Ziegelmann is an Associate Professor in the Department of Statistics and Graduate Programs of Economics and Management at the Federal University of Rio Grande do Sul, Brazil. He holds a PhD in Statistics from the University of Kent at Canterbury (UK, 2002). He is a CNPq researcher and also a research associate at the Center for Quantitative Studies in Economics and Finance of FGV- SP. He is an Associate Editor of the Brazilian Journal of Statistics. His research interests include volatility estimation, dynamic copula modeling, nonparametrics, factor models, functional data analysis, frontier models, MIDAS and LASSO. His publications include articles in the *Annals of Statistics, Econometric Theory, Journal of Forecasting, Insurance: Mathematics & Economics*, among others.