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 $f_1, f_2, \dots$ be a sequence of *random densities*, and let $r_{1t}, \dots, r_{n_t,t}$ 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 $(f_t : t = 1, 2, \dots)$. Eq. (1) says that, conditional on $\mathscr{F}$, the financial returns $r_{1t}, \dots, r_{n_t,t}$ 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 $g_1, \dots, g_n$, 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.