DYNAMICS OF FINANCIAL RETURNS DENSITIES: A FUNCTIONAL APPROACH APPLIED TO THE BOVESPA INTRADAY INDEX

Eduardo de Oliveira Horta[1] and Flavio Augusto Ziegelmann[2]

July, 2011

Resumo. Neste artigo, seguimos a metodologia desenvolvida em Bathia et al. (2010) para modelar a dinâmica das funções densidade de probabilidade de retornos intradiários do índice BOVESPA. Com isso, obtemos estimativas filtradas das densidades desses retornos, no sentido de que são removidos quaisquer ruídos provenientes de sua estimação por métodos não-paramétricos. Ademais, encontramos evidência de que o comportamento dinâmico dessas funções se reduz a um processo no $\mathbf{R}^2$, o qual é bem representado por um modelo VAR(1). Mostramos ainda que esse processo vetorial tem um efeito de deformação sobre a dispersão e simetria das densidades. Por fim, ao incorporar no modelo a estrutura dinâmica do processo dessas curvas, geramos previsões um passo à frente para as mesmas.

Palavras-chave. Autocovariância. Redução de dimensionalidade. Séries temporais funcionais. Dados financeiros intradiários. Expansão de Karhunen-Loève.

Abstract. In this paper, we follow the methodology developed in Bathia et al. (2010) to model the dynamics of the probability density functions (pdf's) of IBOVESPA intraday returns over business days. As a byproduct, we obtain filtered estimates of these intraday return pdf's, in a sense that whatever noise that may occur in previously estimating non-parametrically the pdf's is removed. Furthermore, we find that the pdf's dynamic behavior reduces to a $\mathbf{R}^2$-valued process, and that this process is well represented by a VAR(1) model. Its effect on the pdf's is shown to be a dispersion-symmetry deforming. Moreover, by taking into account the dynamics of the curve process, we generate one-step-ahead forecasts of upcoming pdf's.

Keywords and phrases. Autocovariance. Dimension reduction. Functional time series. High frequency financial data. Karhunen-Loève expansion.

JEL Classifications. C14, C22

[1]Department of Economics, Federal University of Rio Grande do Sul, Porto Alegre, RS 90040-000, Brazil, email: 00134934@ufrgs.br

[2]Department of Statistics, Federal University of Rio Grande do Sul, Porto Alegre, RS 90040-000, Brazil, email: flavioz@ufrgs.br and Department of Statistics, London School of Economics, London, WC2A 2AE UK.

1. Introduction

Adequate specification of the probability density functions (hereafter pdf's) of asset returns is a most relevant topic in statistical modelling of financial data. Many of the well known models in the literature present drawbacks in the way they restrain – or even fail to incorporate – some of the empirical stylized features of financial data. In the last couple of decades, development of more sophisticated models seeking to provide better adjustment to these features has undergone a steep growth. Nonetheless, some of these features which have been shown to be particularly difficult to encompass regard the distribution (that is, the pdf) of asset returns. This is so because most models focus on specifying a mathematical formulation capable of describing the behavior of the asset prices, or of their returns, rather than to model their pdf's directly. In this sort of scenario, some assumptions must be made on the family to which the pdf's pertain, and such restrictions can prove difficult to reconcile with the mentioned stylized empirical features. The approach that we shall take here is distinct in that we focus the analysis on the dynamics of returns distributions, seeing the pdf's of intraday returns as a sequence of random variables taking values on a function space, that is, as a functional time series. The serial dependence existing between these curves allows one to obtain filtered estimates of the pdf's and even to forecast upcoming densities. We therewith aim to provide a richer modelling setting for the pdf's of asset returns.

Much academic research and debate has been directed in the last decades to acknowledging the empirical features of financial data and in particular of the distribution of asset returns. Rydberg (2000) provides a survey on that matter; most of the features therein reported are academic consensuses. Some of them refer to properties of asset returns time series, as volatility clustering, quasi long range dependence and seasonality. In the present context, however, it is of more interest to concentrate on the features of returns pdf's, the most notable being that these densities tend to show fat tails and to be asymmetric, and that they display aggregational Gaussianity. Fat tails means that returns distributions tend to have tails which are heavier than that of a Gaussian pdf; that is, the tails decrease at a rate slower than $\exp\left(-u^2\right)$. One of the most widely known models in the literature, and also one of the most used by market participants, the Black and Scholes model, is based on the geometric Brownian motion, which assumes normally distributed logarithmic returns[3]; this is however a long disputed topic. Distinguishedly, much debate has gone through on how many finite moments actually exist. It is generally accepted that the pdf's of asset returns have at least finite variance, but Müller et al. (1998) find evidence that the fourth moment may not exist. The empirical evidence also points towards the pdf's of stock returns being slightly negatively skewed, the usual explanation being that traders react more strongly to negative than to positive returns. Many models have been proposed so as to incorporate this feature; the EGARCH outstands as one of the most well known among them. Aggregational Gaussianity for returns over lengthier intervals (i.e. weeks, months, etc.) is in turn a mere consequence of the Central Limit Law: since decreasing sampling frequencies is the same as aggregating (i.e., summing) the data, the distribution of returns over lengthier time intervals is the same distribution of a sum of returns over narrower time intervals. By the Central Limit Theorem, this distribution tends to a Gaussian law.

Most of the literature on the aforementioned stylized features deals with financial data sampled at a daily or even lower rate, despite the fact that since the 1990's, with development of computer technology, the collection, storage and retrieval of

[3]See Duffie (2001) for instance.

financial markets data at higher levels of detail has become widespread. On these higher frequency data, much research has been focused on microstructure noise – the fact that observed prices seldom correspond to efficient prices (see for instance Aït-Sahalia et al. (2011), Bandi and Russell (2011) and also Ghysels and Sinko (2011), and Hansen and Lunde (2006)). Apparently less attention has been directed to describing their stylized features, although Ghysels et al. (1998) do present some of these. In particular, by studying the log returns of FX market foreign exchange rate bid and ask quotes for the USD/DEM, the USD/JPY and the DEM/JPY, they find no particular evidence towards the sign of the pdf's skewness; some of these display positive skewness, some negative, and this depends on whether one is looking at the ask or the bid quotes; yet, Rydberg and Shephard (2000) find evidence of negatively skewed tails for the IBM share log returns pdf. Regarding the kurtoses of high frequency returns distributions, Ghysels et al. (1998) report very large values for the mentioned exchange rate returns. Their findings also point towards a sampling frequency effect – changing the time interval over which the returns are measured appears to alter the distribution moments.

With this characteristics in mind, we shall adopt an approach on modelling the densities of asset returns which is innovative in that we take into account the serial dependence existing between intraday returns pdf's throughout business days. Specifically, our approach consists on estimating the pdf's of intraday returns separately for each day, and to assume that these functions present a dynamic structure. Namely, the densities are seen as a functional time series, i.e., a sequence of – possibly non-independent – random variables taking values on a function space. This approach allows for greater generality since no constrains are imposed on the shape of the functions (aside of course that they must be nonnegative and integrate to unity). Likewise, the assumption of serial dependence between the curves permits one to obtain filtered estimates of the pdf's, in the sense that whatever noise that may occur in previously estimating these functions non-parametrically is removed. One of the strengths of this methodology is that along the filtering procedure it is also possible to obtain forecasts for the upcoming densities.

Statistical inference on objects pertaining to function spaces has come to be known in the literature as functional data analysis (FDA)[4]. In recent years, FDA has received growing attention from researchers of a wide spectrum of academic disciplines; see for instance the collection edited by Dabo-Niang and Ferraty (2008) for a discussion on recent development and many applications. Unfortunately this blossom is not yet as widespread in the fields of economics and finance – not to say merely incipient. Still, an application to implied volatility estimation can be found in Benko et al. (2009). Ramsay and Silverman's (1998) cornerstone monograph presents a thorough treatment on the topic, albeit from an applied perspective. From a theoretical point of view, functional data are to be seen as realizations of function-valued random variables – a formal treatment on random variables taking values in Hilbert and Banach spaces can be found in Bosq (2000). An in-between approach may be found in Ferraty and Vieu (2006). Interestingly, most of this literature on FDA deals with the case where the functional data are supposed to be independent realizations from a random function, only very recently coming to be considered the case where the random functions display a dynamic dependence – that is, the case of a sequence of function-valued, non-independent random variables. There is a good presentation on the theory of linear processes of such objects in Bosq (2000).

[4]Actually, since observation and recording of data are discrete by nature, it would be more accurate to say that it is the underlying structure of the data that is believed to be continuous, i.e. functional.

One technique that is central to FDA is that of principal components analysis. At short, such methodology – whose foundation lies in the Karhunen-Loève Theorem – seeks a decomposition of the observed functions as orthogonal projections onto a suitable orthonormal basis which corresponds to the eigenfunctions of a covariance operator. Notwithstanding, as pointed out by Hall and Vial (2006), if the observed functional data are imprecise – due to roundings, experimental measurement errors, etc. – then so is the estimator of the covariance operator, and this poses a major methodological problem in applying principal components analysis to these observed data. The authors suggest solving this issue by letting the measurement errors vanish when the sample size is sufficiently large. The methodology developed in Bathia et al. (2010), which we shall follow throughout this paper, relies instead on the dynamic structure of the curve process as a means to filter the noise from the observed functions, and to find an appropriate orthogonal basis which spans the linear space to which the pdf's pertain. This is an entirely original approach in that it needs not to rely on the stronger assumption that the measurement errors would vanish in face of a large sample. Besides, one of its benefits is that it allows for reducing the analysis to finite-dimensional objects. In terms of its implementation, the method reduces to the eigenanalysis of a finite-dimensional matrix, and the modelling of the dynamic structure of the pdf's, as well as prediction procedures, can be carried out through traditional – and computationally less expensive – multivariate time series methods.

In the following section, we present the aforementioned methodology in detail. In Section 3, we apply this methodology to intraday IBOVESPA data. In particular, we find that the dynamic behavior of IBOVESPA returns pdf's reduces to a $\mathbf{R}^2$-valued process, and that this process is well represented by a VAR(1) model, whose dynamics affect the dispersion and symmetry of the distribution of returns at each day. Section 4 concludes.

2. Methodology

Let $\mathcal{L}$ denote the Hilbert space of square integrable functions[5] defined on an interval $\mathcal{I} \subset \mathbf{R}$ and equipped with the inner product $\langle f, g \rangle = \int_{\mathcal{I}} f(u)\, g(u)\, du$ for all $f, g \in \mathcal{L}$. Say the curve sequence of interest is $(X_t,\ t \in \mathcal{T})$ for some countable set $\mathcal{T}$, where, for each $t \in \mathcal{T}$, X_t is a $\mathcal{L}$-valued random variable. In our case, X_t represents the pdf of a financial asset k-minute return at day t, for some suitable k. The observed densities $Y_t,\ t \in \{1, \ldots, n\}$ – obtained, for example, through some non-parametric method – are taken to be

$$Y_t(u) = X_t(u) + \varepsilon_t(u), \tag{1}$$

where ε_t is white noise in the sense that $E(\varepsilon_t(u)) = 0$ for all $t \in \mathcal{T}$ and for all $u \in \mathcal{I}$, and that $\mathrm{Cov}(\varepsilon_t(u), \varepsilon_s(v)) = 0$ for all $u, v \in \mathcal{I}$ provided $t \neq s$. Thus one can not observe the real pdf's directly, but rather a noisy approximation. There are a number of causes that make reasonable the assumption that the X_t are not perfectly observable; see for example Bathia et al. (2010), and Hall and Vial (2006). We may assume as well that $\mathrm{Cov}(X_t(u), \varepsilon_{t+k}(v)) = 0$ for all integer k.

Now suppose that the mean-function and the k-th lag covariance function of X_t are time-independent, and let them be defined respectively by

$$\mu(u) \equiv E(X_t(u)) \tag{2}$$

and

$$M_k(u, v) \equiv \mathrm{Cov}(X_t(u), X_{t+k}(v)). \tag{3}$$

[5]We shall stick to this more compact notation, instead of the usual $\mathcal{L}^2(\mathcal{I})$.

We shall say that μ is a *quasi*-distribution in the sense that, albeit it being the expectation of random pdf's, there is nothing in principle to grant that μ itself should satisfy the conditions of being positive and integrating to unity. Further, denote by T_F the linear operator from $\mathcal{L}$ into itself defined by $T_F g(u) = \int_{\mathcal{I}} F(u,v)\, g(v)\, dv$ for all square integrable kernel F over $\mathcal{I}^2$ and for all $g \in \mathcal{L}$. By Mercer's Lemma it holds that associated to T_{M_0} there exists a sequence (λ_j) of strictly decreasing eigenvalues and a sequence (φ_j) of corresponding eigenfunctions which forms an orthonormal basis of the linear subspace $\mathcal{M} \subset \mathcal{L}$, and such that

$$M_0(u,v) = \sum_{j=1}^{\infty} \lambda_j \varphi_j(u)\, \varphi_j(v). \tag{4}$$

Moreover, if the condition that $\int_{\mathcal{I}} E\left\{X_t(u)^2 + \varepsilon_t(u)^2\right\} du < \infty$ holds, then by the Karhunen-Loève Theorem it can be shown that

$$X_t(u) - \mu(u) = \sum_{j=1}^{\infty} \xi_{tj} \varphi_j(u), \tag{5}$$

where $\xi_{tj} = \langle X_t - \mu, \varphi_j \rangle$ is a real-valued, zero-mean random variable with $\mathrm{Var}(\xi_{tj}) = \lambda_j$ and $\mathrm{Cov}(\xi_{ti}, \xi_{tj}) = 0$ if $i \neq j$. A formal statement and proof to these theorems can be found in Bosq (2000).

It is assumed that X_t is d-dimensional, meaning that $\lambda_d > 0$ and $\lambda_{d+1} = 0$, such that the sums in (4) and (5) are truncated at d, and $Y_t(u) = \mu(u) + \sum_{j=1}^{d} \xi_{tj}\varphi_j(u) + \varepsilon_t(u)$. The serial dependence of (Y_t) is thus entirely induced by that of the d-vector process $\boldsymbol{\xi}_t \equiv (\xi_{t1}, \ldots, \xi_{td})$. The primary goal of the methodology is to identify d and to estimate the dynamic space $\mathcal{M}$ spanned by the eigenfunctions $\varphi_1, \ldots, \varphi_d$.

Unfortunately, it is not as straightforward to obtain estimates for $\boldsymbol{\xi}_t$ and $\varphi_1, \ldots, \varphi_d$. To see why, consider the estimator

$$\widehat{M}_k(u,v) = \frac{1}{n-p} \sum_{t=1}^{n-p} \left(Y_t(u) - \widehat{\mu}(u)\right)\left(Y_{t+k}(v) - \widehat{\mu}(v)\right), \tag{6}$$

where p is some prescribed integer[6] and

$$\widehat{\mu}(u) = \frac{1}{n} \sum_{t=1}^{n} Y_t(u). \tag{7}$$

A conventional approach would be to perform an eigenanalysis on $T_{\widehat{M}_0}$, letting $\widehat{d}$ be the number of its non-zero eigenvalues, and $\widehat{\mathcal{M}}$ the space spanned by the corresponding eigenfunctions. This approach however suffers from complications due to the fact that $\widehat{M}_0$ is an illegitimate estimator for M_0, since $\mathrm{Cov}(Y_t(u), Y_t(v)) = M_0(u,v) + \mathrm{Cov}(\varepsilon_t(u), \varepsilon_t(v))$ and consequently the noise term would have to be removed before the eigenanalysis could be performed. This is not a trivial matter, as X_t and ε_t are unobservable. Hall and Vial (2006) suggest solving this issue by assuming the variance of ε_t decays to zero as $n \to \infty$.

Bathia et al. (2010) propose an alternate approach that does not depend on such an assumption, but rather relies on the serial dependence of (X_t) as a means to filter the noise of the observable functions Y_t. In fact, for any $k \neq 0$, it holds that $\mathrm{Cov}(Y_t(u), Y_{t+k}(v)) = M_k(u,v)$. Yet, as of estimation it is not possible to use $\widehat{M}_k$ directly since T_{M_k} is not necessarily a non-negative operator; see Remark 1

[6]The truncation of the sum at $n-p$ instead of $n-k$ ensures a duality operation which simplifies the computation for eigenfunctions. See Remark 2.

below. Instead, the eigenanalysis is performed on the operator T_K with K defined in the following proposition:

Proposition 1. *Let $\mathbf{\Sigma}_k \equiv E\left(\boldsymbol{\xi}_t \boldsymbol{\xi}_{t+k}'\right)$ denote the k-th lag autocovariance matrix of $\boldsymbol{\xi}_t$, and define*

$$K(u,v) = \sum_{k=1}^{p} \int_{\mathcal{I}} M_k(u,z)\, M_k(v,z)\, dz \tag{8}$$

for some $p \in \mathbf{N}$. If $\mathbf{\Sigma}_{k_0}$ is full-ranked for some integer $k_0 \geq 1$, then it holds that, for any $p \geq k_0$, T_K has exactly d non-zero eigenvalues, and $\mathcal{M}$ is the linear space spanned by the corresponding eigenfunctions.

Proof. See Bathia et al. (2010, Appendix B). □

Remark 1. *(i)* The condition that $\operatorname{rank}(\mathbf{\Sigma}_{k_0}) = d$ for some $k_0 \geq 1$ is a direct implication of the assumption that X_t is d-dimensional. If this were not the case, than the component with no serial correlations in X_t should be absorbed into the white noise ε_t.

(ii) Note that $T_K \zeta = T_{M_k} \zeta = 0$ for all $\zeta \in \mathcal{M}^{\perp}$. The choice of introducing the operator T_K is motivated by the fact that it is a self-adjoint, non-negative operator, whilst these properties do not necessarily hold for T_{M_k}; i.e. it may happen that $T_{M_k} g = 0$ even if $g \in \mathcal{M}$. Furthermore, setting $p > 1$ may help avoiding spurious choices of $\widehat{d}$ when of estimation.

Let $\psi_1, \ldots, \psi_d$ be the orthonormal eigenfunctions of T_K associated to its d non-zero eigenvalues. From Proposition (1), we have that $\operatorname{span}(\psi_1, \ldots, \psi_d) = \mathcal{M}$. Therefore, it holds that

$$X_t(u) - \mu(u) = \sum_{j=1}^{d} \xi_{tj} \varphi_j(u) = \sum_{j=1}^{d} \eta_{tj} \psi_j(u), \tag{9}$$

where $\eta_{tj} = \langle X_t - \mu, \psi_j \rangle$. Since, as stated above, it is not possible to estimate $\boldsymbol{\xi}_t$ and $\varphi_1, \ldots, \varphi_d$, we estimate $\boldsymbol{\eta}_t$ and $\psi_1, \ldots, \psi_d$ instead. Define an estimator

$$\widehat{K}(u,v) = \sum_{k=1}^{p} \int_{\mathcal{I}} \widehat{M}_k(u,z)\, \widehat{M}_k(v,z)\, dz. \tag{10}$$

The following proposition allows the eigenanalysis on $T_{\widehat{K}}$ to be reduced to a finite-matrix eigenanalysis:

Proposition 2. *Define a scalar sequence $\mathbf{Y}_t$ such that $\mathbf{Y}_t' \mathbf{Y}_s = \langle Y_t - \widehat{\mu}, Y_s - \widehat{\mu} \rangle$, and put $\mathcal{Y}_k = [\mathbf{Y}_{1+k} \ \ldots \ \mathbf{Y}_{n-p+k}]$. Then it holds that the $(n-p) \times (n-p)$ matrix $\mathbf{K}^*$ defined by*

$$\mathbf{K}^* = \frac{1}{(n-p)^2} \sum_{k=1}^{p} \mathcal{Y}_k' \mathcal{Y}_k \mathcal{Y}_0' \mathcal{Y}_0, \tag{11}$$

shares the same non-zero eigenvalues with the operator $T_{\widehat{K}}$, with $\widehat{K}$ defined as in (10)*. Moreover, the associated eigenfunctions of $T_{\widehat{K}}$ are given by*

$$\widetilde{\psi}_j(u) = \sum_{t=1}^{n-p} \gamma_t^{(j)} \left(Y_t(u) - \widehat{\mu}(u) \right), \tag{12}$$

where $\gamma_t^{(j)}$ is the t-th component of the eigenvector $\boldsymbol{\gamma}_j$ associated to the j-th largest eigenvalue of $\mathbf{K}^*$.

Proof. See Bathia et al. (2010, Appendix B). □

Remark 2. *(i)* Observe that the truncation of the sums in (6) at $(n-p)$ ensures that the sum in (11) is well-defined.

(ii) Although the eigenfunctions in (12) span the dynamic space $\mathcal{M}$, nothing grants that they form an orthonormal basis of that space. Therefore a more parsimonious estimator for the ψ_j's may be obtained by applying a Gram-Schmidt algorithm to the obtained eigenfunctions.

Let $\left(\widehat{\psi}_j\right)$ denote the eigenfunctions obtained by applying a Gram-Schmidt algorithm to the functions in (12). This leads to a filtered estimator for X_t defined by

$$\widehat{X}_t(u) = \widehat{\mu}(u) + \sum_{j=1}^{\widehat{d}} \widehat{\eta}_{tj}\widehat{\psi}_j(u), \tag{13}$$

where

$$\widehat{\eta}_{tj} = \left\langle Y_t - \widehat{\mu}, \widehat{\psi}_j \right\rangle, \qquad j = 1, \ldots, \widehat{d}, \tag{14}$$

and where $\widehat{d}$ is an estimate of the dimension of X_t determined through some statistical test, as discussed below.

In spite of T_K having exactly d non-zero eigenvalues, it may happen that the number of non-zero eigenvalues of $T_{\widehat{K}}$ is much larger. This is due mainly to random fluctuations in the sample. Bathia et al. (2010) propose a bootstrap test to determinate $\widehat{d}$, which we shortly present.

Let (θ_j) be the eigenvalues of T_K. Say the true dimension of X_t is d_0, that is, $\theta_1, \ldots, \theta_{d_0} > 0$ and $\theta_{d_0+j} = 0$ for all integer j. 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 hypothesis is

$$H_0 : \ \theta_{d_0+1} = 0, \tag{15}$$

with d_0 being a known integer obtained for example by visual inspecting the graph of the estimated eigenvalues $\widehat{\theta}_1 \geq \widehat{\theta}_2 \geq \cdots \geq 0$ of $T_{\widehat{K}}$, then we reject H_0 if $\widehat{\theta}_{d_0+1} > \ell_\alpha$, where ℓ_α is the critical value at the $\alpha \in (0,1)$ significance level. The following bootstrap procedure allows one to evaluate this critical value:

B1 Define $\widehat{\varepsilon}_t(u) = Y_t(u) - \widehat{X}_t(u)$, with $\widehat{d}$ in (13) set equal to d_0.

B2 Generate bootstrap observations $Y_t^{[b]}$ defined by

$$Y_t^{[b]}(u) = \widehat{X}_t(u) + \varepsilon_t^{[b]}(u), \qquad t = 1, \ldots, n$$

with b varying 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 $\mathbf{K}^*_{[b]}$ the same way as $\mathbf{K}^*$, with $Y_t^{[b]}$ in place of Y_t, and let $\widehat{\theta}^{[b]}_{d_0+1}$ be the (d_0+1)-th largest eigenvalue of $\mathbf{K}^*_{[b]}$.

Now the conditional distribution of $\widehat{\theta}^{[b]}_{d_0+1}$ given observations $\{Y_1, \ldots, Y_n\}$ is taken as the distribution of $\widehat{\theta}_{d_0+1}$ under H_0, such that $B^{-1}\sum_{b=1}^{B} \mathbf{1}_{\left\{x \geq \widehat{\theta}_{d_0+1}\right\}}\left(\widehat{\theta}^{[b]}_{d_0+1}\right)$, is taken to be the probability of obtaining an estimate equal to or grater than $\widehat{\theta}_{d_0+1}$ when θ_{d_0+1} is equal to zero. Thus whenever this magnitude is equal to or less than α, we reject the null.

We briefly state now three relevant theorems regarding the theoretical properties of the presented methodology. For a detailed discussion and proofs to these theorems, as well as for a discussion on numerical properties obtained via simulations, refer to Bathia et al. (2010).

Theorem 1. *Let $\|T\|_{\mathcal{S}}$ denote the Hilbert-Schmidt norm for any linear operator T from $\mathcal{L}$ into itself, and suppose that the following conditions hold:*

C1 *(Y_t) is strictly stationary and ψ-mixing with the mixing coefficient defined as*

$$\psi(\ell) = \sup_{A\in\mathcal{F}_{\infty}^{0}, B\in\mathcal{F}_{\ell}^{\infty}, P(A)P(B)>0} \left|1 - P(B|A)/P(B)\right|,$$

where $\mathcal{F}_i^j = \sigma\{Y_i,\ldots,Y_j\}$ for any $j \geq i$. In addition, it holds that $\sum_{\ell=1}^{\infty} \ell\psi(\ell)^{1/2} < \infty$.

C2 $E\left\{\int_{\mathcal{I}} Y_t(u)^2\, du\right\} < \infty$.

C3 $\theta_1 > \theta_2 > \cdots > \theta_d > 0 = \theta_{d+1} = \cdots$.

C4 *$Cov\{X_t(u), \varepsilon_s(v)\} = 0$ for all s,t and for all $u,v \in \mathcal{I}$.*

Then, as $n \to \infty$,

(i) $\left\|T_{\widehat{K}} - T_K\right\|_{\mathcal{S}} = O_p\left(n^{-1/2}\right)$.

(ii) *For $j = 1,\ldots,d$,* $\left|\widehat{\theta}_j - \theta_j\right| = O_p\left(n^{-1/2}\right)$ *and*

$$\left(\int_{\mathcal{I}} \left\{\widehat{\psi}_j(u) - \psi_j(u)\right\}^2 du\right)^{1/2} = O_p\left(n^{-1/2}\right).$$

(iii) *For $j \geq d+1$,* $\widehat{\theta}_j = O_p\left(n^{-1}\right)$.

(iv) *Let $(\psi_j :\ j \geq d+1)$ be a complete orthonormal basis of $\mathcal{M}^{\perp}$, and put*

$$f_j(u) = \sum_{i=d+1}^{\infty} \left\langle \widehat{\psi}_j, \psi_i \right\rangle \psi_i(u).$$

Then, for any $j \geq d+1$,

$$\left(\int_{\mathcal{I}} \left\{\sum_{i=1}^{d} \left\langle \widehat{\psi}_j, \psi_i \right\rangle \psi_i(u)\right\}^2 du\right)^{1/2} = \left(\int_{\mathcal{I}} \left\{\widehat{\psi}_j(u) - f_j(u)\right\}^2 du\right)^{1/2} = O_p\left(n^{-1/2}\right).$$

Proof. See Bathia et al. (2010, Appendix B). □

Theorem 2. *Let $\widetilde{\mathcal{M}}_{\delta} \equiv span\left(\widehat{\psi}_1,\ldots,\widehat{\psi}_{\delta}\right)$. In particular, $\widetilde{\mathcal{M}}_{\widehat{d}} = \widehat{\mathcal{M}}$. Define the metric*

$$D(\mathcal{N}_1,\mathcal{N}_2) = \sqrt{1 - \frac{1}{\max(d_1,d_2)} \sum_{k=1}^{d_1}\sum_{j=1}^{d_2} \langle \zeta_{2j}, \zeta_{1k}\rangle^2},$$

where $\{\zeta_{i1},\ldots\zeta_{id_i}\}$ is an orthonormal basis of $\mathcal{N}_i \subset \mathcal{L}$, $i = 1,2$. Suppose that d is known. If the conditions of Theorem 1 and Proposition 1 hold, then as $n \to \infty$ it holds that $D\left(\widetilde{\mathcal{M}}_d, \mathcal{M}\right) = O_p\left(n^{-1/2}\right)$.

Proof. See Bathia et al. (2010, Appendix B). □

Theorem 3. *Define $\widehat{d} = \#\left\{j :\ \widehat{\theta}_j \geq \epsilon(n)\right\}$ with $\epsilon(n) > 0$, and let $\epsilon(n) \to 0$ and $\epsilon(n)^2 n \to \infty$ as $n \to \infty$. Then it holds that $P\left(\widehat{d} \neq d\right) \to 0$.*

Proof. See Bathia et al. (2010, Appendix B). □

In the following section we apply the presented methodology to the 5- and 10-minute returns series of the BOVESPA index.

3. Data Analysis

Our data set consists of intraday BOVESPA stock indexes through 305 business days ranging from September 1st, 2009 to November 6th, 2010[7]. The tick-by-tick series is sampled at each 5 minutes, such that for each day we have a sample $(S_{1t}, \ldots, S_{\bar{m}_t t})$, where S_{1t} is the stock index at the t-th day market opening time, S_{2t} is the index 5 minutes ahead, and so on up to $S_{\bar{m}_t t}$ which is the index $5\,(\bar{m}_t - 1)$ minutes ahead from the t-th day market opening time, $\bar{m}_t$ being the number of index observations within day t. Most days share the same market opening and closing times, but it is not necessarily so. For illustration, Figure 3.1 plots the intraday 5-minute index time series, as well as the corresponding 5-minute return series, of the first four business days in the sample. In addition, we also study the 10-minute return series, which leads us to define the i-th k-minute return at day t as

$$r_{it}^{(k)} = \log S_{i+k/5,t} - \log S_{it},\ i = 1, \ldots, \bar{m}_t - k/5,\ k = 5, 10. \tag{16}$$

For the sake of notational simplicity, let r_{it} be $r_{it}^{(k)}$ for either $k = 5$ or $k = 10$. Also let $m_t = \bar{m}_t - k/5$ for either k. Denote by Y_t the standard kernel method[8] estimate of the k-minute return pdf at day t. These densities are defined by

$$Y_t(u) = (m_t h_t)^{-1} \sum_{i=1}^{m_t} \mathcal{K}\left(\frac{r_{it} - u}{h_t}\right), \qquad t = 1, \ldots, n, \tag{17}$$

where $\mathcal{K}(v) = \left(\sqrt{2\pi}\right)^{-1} \exp\left(-v^2/2\right)$ is a Gaussian kernel and h_t is a bandwidth. Note that the true density X_t is taken as the unconditional distribution of a k-minute return at day t; that is, in obtaining Y_t we assume the r_{it}'s are independent observations of returns over k minutes. This assumption does not pose much of a problem, since kernel density estimators like (17) have the whitening-by-windowing property; refer to Fan and Yao (2003) for a discussion. However, the densities *are* seen as conditional on information from previous days.

We compute Y_t for the 5- and 10-minute returns, letting p in (6) be equal to 5. As the support of Y_t we set the interval $\mathcal{I} = [-\max\{|r_{it}|\}, \max\{|r_{it}|\}]$, and for the bandwidth we use three rescalings of Silverman's rule of thumb bandwidth $\widehat{h}_t = 1.06\widehat{\sigma}_t m_t^{-1/5}$, where $\widehat{\sigma}_t$ denotes the sample standard deviation of $(r_{1t}, \ldots, r_{m_t t})$. Namely, we set h_t in (17) equal to $0.5\widehat{h}_t$, to $\widehat{h}_t$ and to $2\widehat{h}_t$. Figure 3.2 plots the estimated mean-function[9] $\widehat{\mu}$ for both the 5- and 10-minute return, using the three mentioned levels of smoothing.

Hereafter we apply the methodology discussed in the previous section to the obtained pdf's. All computational work throughout this paper was carried out through the R statistical package[10]. The estimated eigenvalues of $\mathbf{K}^*$ are displayed in Figure 3.3. It can be seen that $\widehat{\theta}_1$ is much larger than $\widehat{\theta}_2$ for $\mathbf{K}^*$ built using either the 5- or the 10-minute returns pdf's, and that this disparity does not depend on the selected bandwidth as well. Although this graphical inspection alone could

[7]All data were kindly made available by *CAPSE investimentos*.

[8]See Fan and Yao (2003) for a good exposition on nonparametric density estimation.

[9]Note that, although the theoretical mean-function μ may not represent a distribution, its empirical counterpart $\widehat{\mu}$ is a pdf by construction. With this in mind, in what follows we shall use expressions such as "the mode of $\widehat{\mu}$", "the skewness of the mean-function", etc.

[10]This consists mostly of self-written code. Where pertinent, we mention the other packages used.

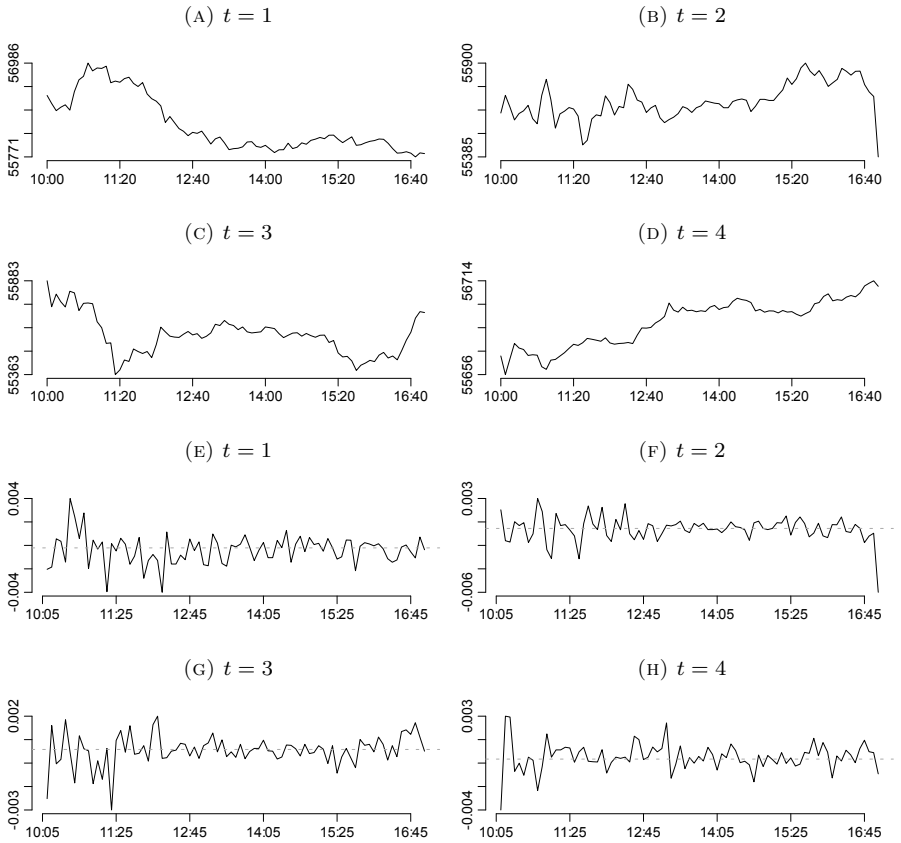

FIGURE 3.1. Intraday BOVESPA index series (panels (A)–(D)) and corresponding 5-minute return series (panels (E)–(H)) for dates $t = 1, \ldots, 4$.

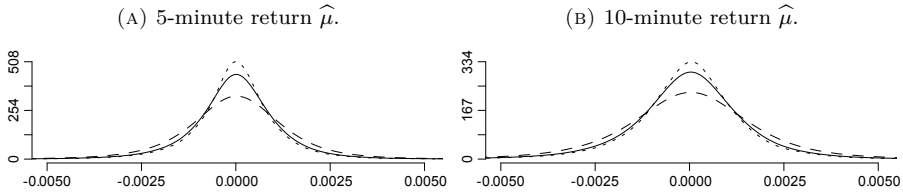

FIGURE 3.2. Estimated mean-function $\widehat{\mu}$ for the (A) 5- and (B) 10-minute return obtained using bandwidths $\widehat{h}_t$ (solid), $0.5\widehat{h}_t$ (dotted) and $2\widehat{h}_t$ (dashed).

recommend setting $\widehat{d} = 1$, we further evaluate the true dimension of X_t by executing the bootstrap test presented in Section 2, with $B = 1000$. We sequentially test the null-hypotheses that $\theta_2 = 0$ and that $\theta_3 = 0$. Table 3.1 reports the obtained p-values. These results suggest that neither the sampling frequency (be it 5 or 10 minutes) nor the particular choice of bandwidth have major impacts on the implied decisions. In particular, it is clear that we systematically reject the null that $\theta_2 = 0$ at the $\alpha = 0.1$ significance level, whilst we cannot reject that $\theta_3 = 0$ at the same level. Further evidence towards the true dimension of X_t being equal to 2 is given

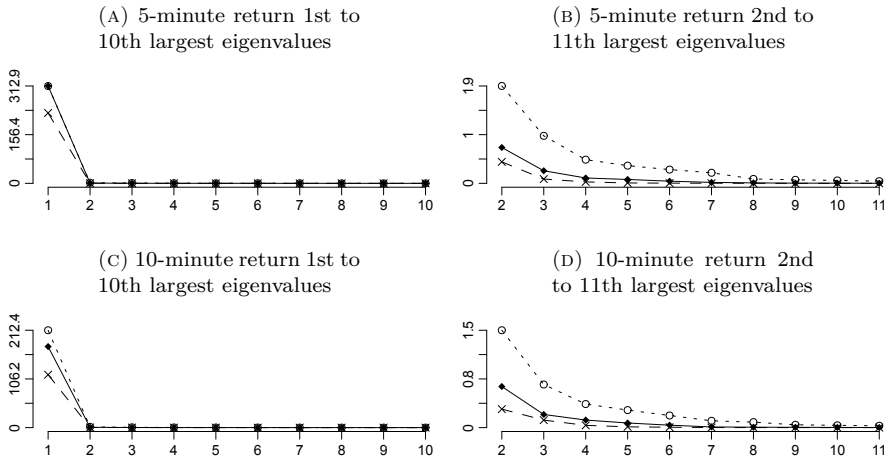

FIGURE 3.3. 5- and 10-minute return largest estimated eigenvalues using bandwidths $\widehat{h}_t$ (solid), $0.5\widehat{h}_t$ (dotted) and $2\widehat{h}_t$ (dashed).

	5-minute return			10-minute return		
	$h_t = \widehat{h}_t$	$h_t = 0.5\widehat{h}_t$	$h_t = 2\widehat{h}_t$	$h_t = \widehat{h}_t$	$h_t = 0.5\widehat{h}_t$	$h_t = 2\widehat{h}_t$
$H_0: \theta_2 = 0$	0.07	0.11	0.01	0.10	0.08	0.05
$H_0: \theta_3 = 0$	0.56	0.52	0.38	0.71	0.54	0.28

TABLE 3.1. p-values from bootstrap test. The reported values correspond to the proportion of $(d_0 + 1)$-th eigenvalues generated – under the null hypothesis – through the bootstrap loops that are greater than the estimated eigenvalue $\widehat{\theta}_{d_0+1}$.

by the p-values obtained from applying the Ljung-Box portmanteau test to the scalar series $\widehat{\eta}_{tj}$. The p-values are virtually zero for the $\widehat{\eta}_{t1}$ and the $\widehat{\eta}_{t2}$ series, and much larger for the remaining $\widehat{\eta}_{tj}$, $j > 2$, no matter the chosen sampling frequency or bandwidth. These results suggest that there is a lot of dynamic structure in the two-dimensional subspace corresponding to the eigenvalues $\widehat{\theta}_1$ and $\widehat{\theta}_2$, but there may be none in the remaining directions. We therefore set $\widehat{d} = 2$. Figure 3.4 displays the estimated eigenfunctions $\widehat{\psi}_1$ and $\widehat{\psi}_2$, and Figure 3.5 plots the time series path of the estimated loadings $\widehat{\eta}_{t1}$ and $\widehat{\eta}_{t2}$ for $t = 1, \ldots, 60$. These figures suggest that the choice of return frequencies and bandwidth also have a weak influence on the overall shape of the eigenfunctions and the paths of the discrete scalar time series. With this in mind, we restrict the remainder of our analysis to the 5-minute return pdf's built with bandwidth set equal to $\widehat{h}_t$.

In Figure 3.6 the non-parametric density estimates Y_t and the filtered estimates $\widehat{X}_t(u) = \widehat{\mu}(u) + \sum_{j=1}^{2} \widehat{\eta}_{tj}\widehat{\psi}_j(u)$[11] corresponding to the first 8 business days in the sample are displayed. It comes into attention the fact that the non-parametric and the filtered estimates are similar in shape, with the filtered ones showing a slightly smoother aspect. Figure 3.7 plots the mean-function $\widehat{\mu}$ and the eigenfunctions $\widehat{\psi}_1$

[11]The estimates were corrected such that the obtained functions would satisfy the conditions required by a pdf, namely that $\widehat{X}_t \geq 0$ and $\int_{\mathcal{I}} \widehat{X}_t = 1$. This was done by setting all negative points of $\widehat{X}_t$ equal to zero and by normalizing the obtained function. In spite of seeming rather arbitrary, these corrections were in fact minimal since the negative points that appeared were close to zero, and the estimated functions integrated nearly 1.

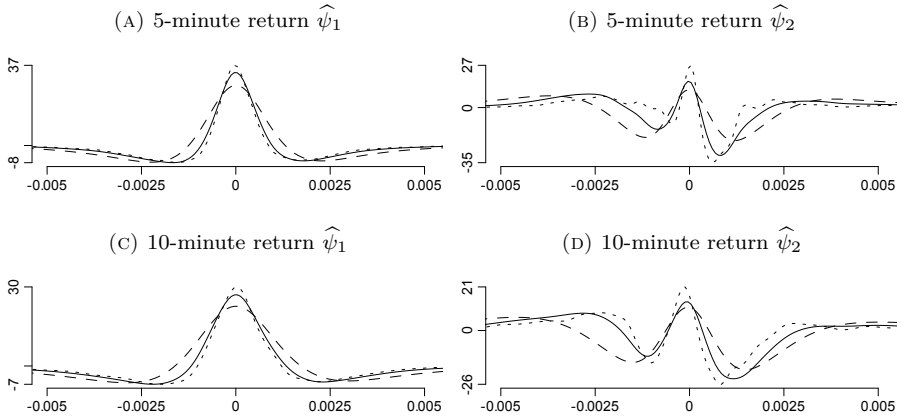

FIGURE 3.4. 5- and 10-minute return estimated eigenfunctions using bandwidths $\widehat{h}_t$ (solid), $0.5\widehat{h}_t$ (dotted) and $2\widehat{h}_t$ (dashed).

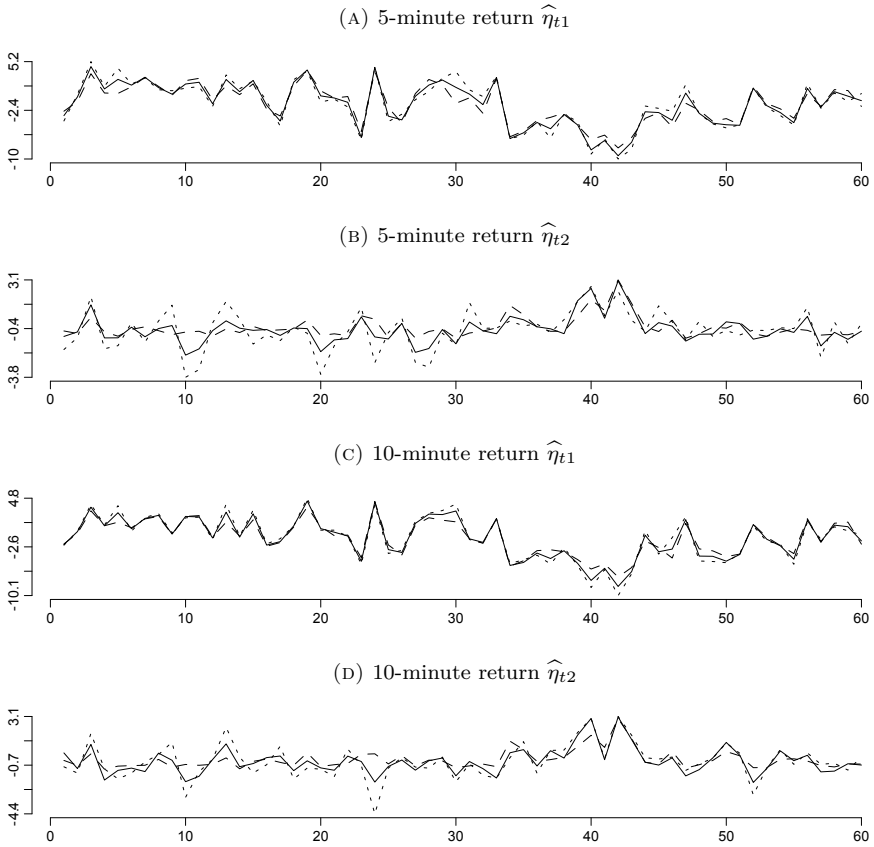

FIGURE 3.5. 5- and 10-minute return estimated loadings $\widehat{\eta}_{t1}$ and $\widehat{\eta}_{t2}$ for $t = 1, \ldots, 60$, using bandwidths $\widehat{h}_t$ (solid), $0.5\widehat{h}_t$ (dotted) and $2\widehat{h}_t$ (dashed).

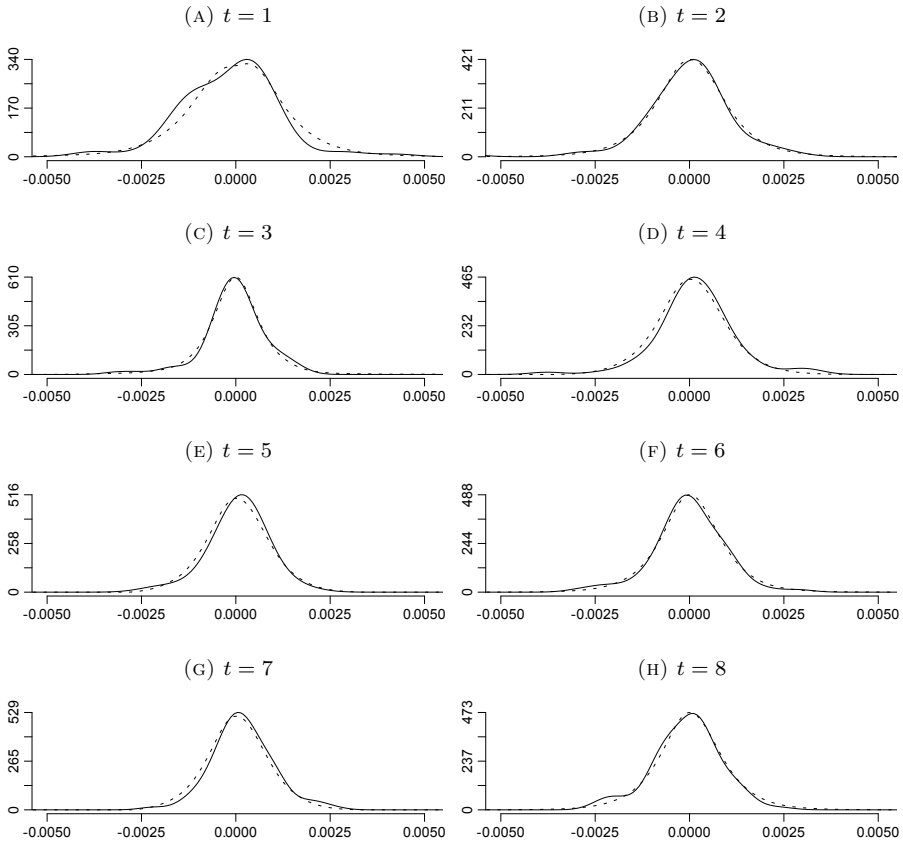

FIGURE 3.6. Non-parametric density estimates Y_t (solid) and filtered estimates $\widehat{X}_t$ (dotted), $t = 1, \ldots, 8$.

and $\widehat{\psi}_2$. This figure allows for a straightforward yet meaningful interpretation. By the definition of $\widehat{X}_t$, for each t the filtered estimate is formed by adding to the mean-function a weighted sum of the two eigenfunctions, whose weights are the random scalars $\widehat{\eta}_{t1}$ and $\widehat{\eta}_{t2}$. Thus the filtered pdf $\widehat{X}_t$ can be seen as a random deformation of $\widehat{\mu}$, with the shape-deforming structure given by the eigenfunctions, and the degree of deformation given by the $\widehat{\boldsymbol{\eta}}_t$ process. In particular, from Figure 3.7 it appears that in face of a positive value of $\widehat{\eta}_{t1}$, the eigenfunction $\widehat{\psi}_1$ adds probability mass for values close to the mode of $\widehat{\mu}$ and symmetrically takes probability mass from the tails, the converse being true in face of a negative value of $\widehat{\eta}_{t1}$. The eigenfunction $\widehat{\psi}_2$, in turn, facing a positive (negative) value of $\widehat{\eta}_{t2}$ seems to add (take) little probability mass for values very close to the mode of $\widehat{\mu}$, to asymmetrically take (add) probability mass from (to) values to the left and to the right of this mode – this asymmetry being very sound towards the right –, and to asymmetrically add (take) probability mass to (from) the tails of the distribution. 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 to distributions $\widehat{X}_t$ showing lower (higher) dispersion; in the same sense, negative (positive) values of $\widehat{\eta}_{t2}$ should be associated to distributions $\widehat{X}_t$ with positive (negative) skewness.

The preceding conjectures are further supported by the following argument: let $\kappa'_{t1} = \int_{\mathcal{I}} u\widehat{X}_t(u)\,du$ and $\kappa_{tj} = \int_{\mathcal{I}} (u - \kappa'_{1t})^j\, \widehat{X}_t du$. That is, κ'_{t1} is the expectation

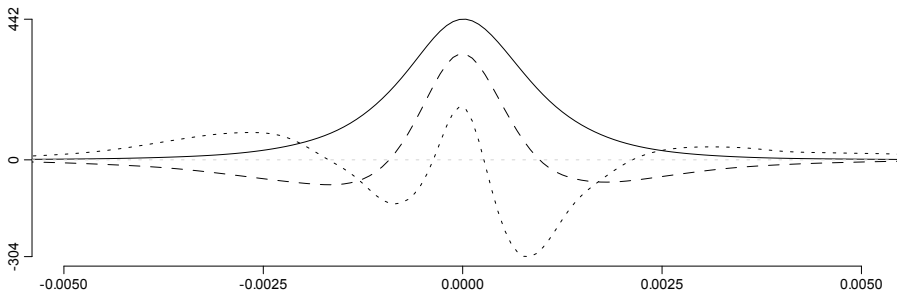

FIGURE 3.7. The estimated mean-function $\widehat{\mu}$ (solid), and the eigenfunctions $\widehat{\psi}_1$ (dashed) and $\widehat{\psi}_2$ (dotted) . The eigenfunctions are rescaled by a $10\times$ factor for better visualization.

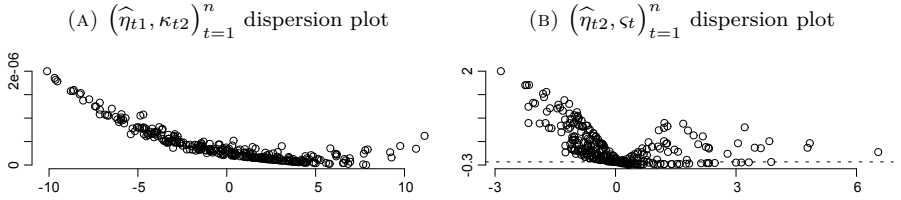

FIGURE 3.8. Dispersion plots of (A) $\widehat{\eta}_{t1}$ vs. κ_{t2}, and (B) $\widehat{\eta}_{t2}$ vs. ς_t. The dotted line in panel (B) indicates the vertical-axis's zero.

of r_{it} under the distribution $\widehat{X}_t$, κ_{t2} is its variance under the same distribution, and so on. In particular, $\varsigma_t = \kappa_{t3}/\kappa_{t2}^{-3/2}$ is the skewness of the distribution $\widehat{X}_t$. Then if the above reasoning should hold, the time series $(\widehat{\eta}_{t1})$ and (κ_{t2}) ought to be related in some manner; the same should happen with $(\widehat{\eta}_{t2})$ and (ς_t). The existence of such relationships is supported by the dispersion plots of $(\widehat{\eta}_{t1}, \kappa_{t2})_{t=1}^{n}$ and of $(\widehat{\eta}_{t2}, \varsigma_t)_{t=1}^{n}$. These are displayed in Figure 3.8. It is very likely that the assumed relations do exist, specially the one between $(\widehat{\eta}_{t1})$ and (κ_{t2}). Yet, we point out that the interpretation of such relations is not as straightforward, as they do not appear to be linear.

Since it is a pertinent topic in the literature, we open a short parenthesis regarding the moments of the distribution $\widehat{\mu}$ – even though it is not our aim to discuss this matter in detail[12]. Table 3.2 reports the values of the first four moments, skewness and kurtosis of $\widehat{\mu}$. In particular, one sees that this distribution is slightly positively skewed (0.68), and that it presents a very large kurtosis (48.86). Figure 3.9 displays the mean-function $\widehat{\mu}$ compared to a Gaussian pdf of same expectation and variance, as well as the log-pdf's of these distributions. The shape of $\log \widehat{\mu}$ is typical to distributions with very large fourth moments. As just said, it is not our aim to discuss this matter in further details; rather, we only want to bring into debate the recurring issue on whether higher-order moments of financial returns distributions do or do not exist. Refer to Müller et al. (1998), for instance.

We further fit a VAR(q) model[13] to the 2-vector process $\widehat{\boldsymbol{\eta}}_t$. We use the `R` function `VAR` from package `vars` (see Pfaff (2008)). Firstly we perform the augmented Dickey-Fuller test to the time series $\widehat{\eta}_{t1}$ and $\widehat{\eta}_{t2}$. The obtained p-values are virtually

[12]Note that $\widehat{\mu}$ is the expected pdf of a 5-minute return at any day, so it brings information regarding the unconditional distribution of this return.

[13]See Hamilton (1994) for a thoroughly exposition of VAR models and their estimation.

Distribution moment	Value
1st uncentered	1.60e-06
2nd centered	1.78e-06
3rd centered	1.61e-09
4th centered	1.64e-10
Skewness	0.68
Kurtosis	48.86

TABLE 3.2. Distribution moments of $\widehat{\mu}$.

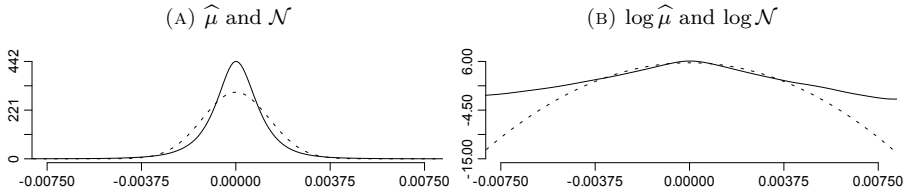

FIGURE 3.9. (A) The pdf $\widehat{\mu}$ (solid) and a Gaussian pdf $\mathcal{N}$ of same mean and variance (dotted); (B) The corresponding log-pdf's.

	Estimate	Std. Error	t value	Pr(>|t|)
$\widehat{\mathbf{A}}_{1,1}$	0.65	0.04	14.59	0.00
$\widehat{\mathbf{A}}_{1,2}$	-0.35	0.15	-2.36	0.02
$\widehat{\mathbf{A}}_{2,1}$	< 0.01	0.02	0.03	0.97
$\widehat{\mathbf{A}}_{2,2}$	0.34	0.05	6.29	0.00

TABLE 3.3. Estimated coefficients from model $\widehat{\boldsymbol{\eta}}_t = \mathbf{A}\widehat{\boldsymbol{\eta}}_{t-1} + \boldsymbol{\iota}_t$.

zero for both series whatever specification is used, be it with a drift component, a drift and a trend component, or neither. Therefore we take $\widehat{\boldsymbol{\eta}}_t$ to be stationary. Figure 3.10 displays the ACF and PACF plots for $\widehat{\boldsymbol{\eta}}_t$. We then fit the model

$$\widehat{\boldsymbol{\eta}}_t = \sum_{k=1}^{q} \mathbf{A}^{(k)}\widehat{\boldsymbol{\eta}}_{t-k} + \boldsymbol{\iota}_t \tag{18}$$

to the $\widehat{\boldsymbol{\eta}}_t$ series[14], and set $q = 1$ according to the Akaike Information Criterion. Table 3.3 reports the OLS estimates of the parameter matrix $\mathbf{A}^{(1)} \equiv \mathbf{A} = [\mathbf{A}_{i,j}]$. All estimates are highly significant, except for $\widehat{\mathbf{A}}_{2,1}$ which measures the influence of $\widehat{\eta}_{t-1,1}$ on $\widehat{\eta}_{t2}$.

Now let $\widehat{\mu}_{|\bar{n}}$ be the estimated mean-function using a subsample of size $\bar{n} \leq n$. In the same sense, denote respectively by $\widehat{\psi}_{j|\bar{n}}$ and $\widehat{\eta}_{tj|\bar{n}}$, $j = 1, 2$, $t = 1, \ldots, \bar{n}$ the estimated eigenfunctions and the estimated scalar processes, and by $\widehat{\mathbf{A}}^{(k)}_{|\bar{n}}$ the estimated parameter matrices from model (18), obtained through applying the methodology developed in Section 2 using a subsample of size $\bar{n}$. We define the one-step-ahead

[14]There is no intercept in the model, since $E\widehat{\boldsymbol{\eta}}_t = \mathbf{0}$ by construction.

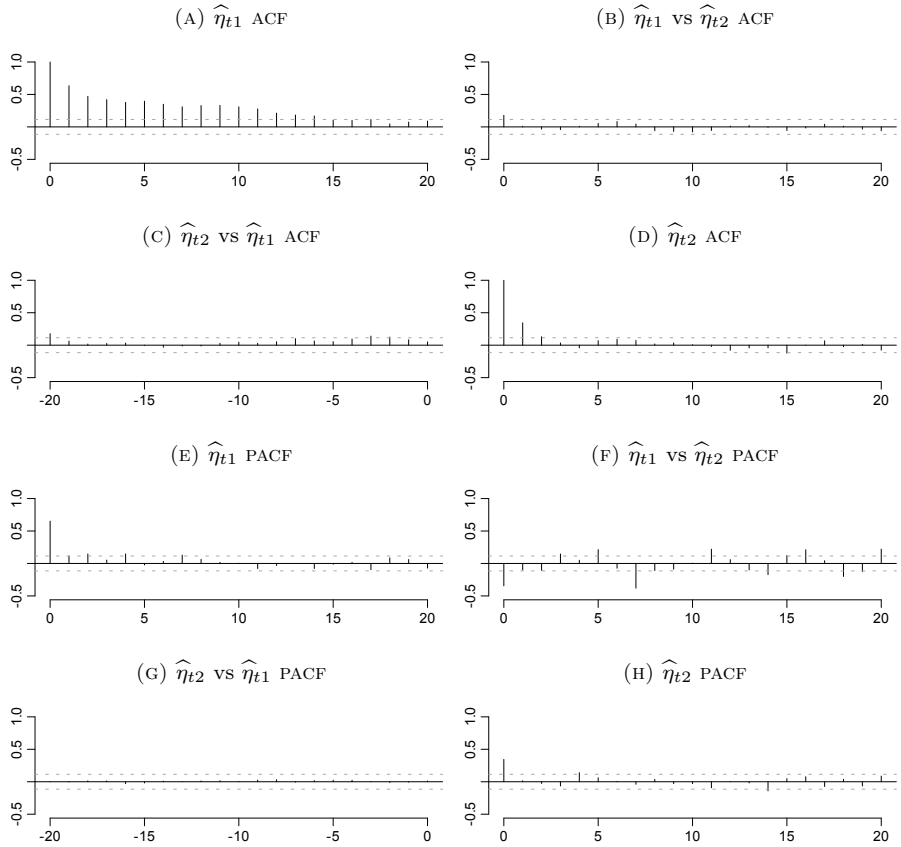

FIGURE 3.10. ACF and PACF of $\widehat{\boldsymbol{\eta}}_t$.

pdf predictor using subsample $\bar{n}$ to be

$$\widetilde{X}_{\bar{n}+1|\bar{n}}(u) = \widehat{\mu}_{|\bar{n}}(u) + \sum_{j=1}^{2} \widetilde{\eta}_{\bar{n}+1,j|\bar{n}} \widehat{\psi}_{j|\bar{n}}(u), \tag{19}$$

where $\widetilde{\eta}_{\bar{n}+1,j|\bar{n}}$ is the j-th component of the vector $\widetilde{\boldsymbol{\eta}}_{\bar{n}+1|\bar{n}}$ defined by

$$\widetilde{\boldsymbol{\eta}}_{\bar{n}+1|\bar{n}} = \sum_{k=1}^{q} \widehat{\mathbf{A}}_{|\bar{n}}^{(k)} \widehat{\boldsymbol{\eta}}_{\bar{n}+1-k|\bar{n}}. \tag{20}$$

In words: we apply the the methodology presented in Section 2 using information available up to $\bar{n}$, and generate a standard VAR one-step-ahead forecast for the vector process obtained using this information. This forecast is in turn used to weight the obtained eigenfunctions so as to predict the upcoming pdf.

In order to evaluate the accuracy of such forecasts, we generate predictions $\widetilde{X}_{\bar{n}|\bar{n}-1}$ with $\bar{n}$ varying in $\{n_0+1,\ldots,n\}$ for some $n_0 < n$, and compare each of them with the filtered pdf's obtainable at $\bar{n}$, $\widehat{X}_{\bar{n}|\bar{n}}(u) = \widehat{\mu}_{|\bar{n}}(u) + \sum_{j=1}^{2} \widehat{\eta}_{\bar{n}j|\bar{n}} \widehat{\psi}_{j|\bar{n}}(u)$[15].

[15]This comparison could also be carried against the filtered pdf estimates obtained using information from the whole sample; in fact, in preliminary tests we found that these two estimates were extremely close one to another. So, we chose to use the restricted sample estimates for comparison, as this iterative evaluation of forecasting performance seems to make more sense in real applications.

f	$f(\Delta_{\bar{n}})$	$f(\nabla_{\bar{n}})$
#	55	55
mean	0.196	0.156
Std.Dev.	0.129	0.093
min	0.038	0.034
Q1	0.101	0.084
median	0.147	0.131
Q3	0.282	0.215
max	0.737	0.425

TABLE 3.4. Summary statistics for $\Delta_{\bar{n}}$ and $\nabla_{\bar{n}}$, $\bar{n} = 251, \ldots, 305$.

Interestingly, the chosen lag q in (18) according to the AIC criterion was always equal to 1, no matter the size of the subsample. The comparison carried out is in terms of the two metrics $\Delta_{\bar{n}}$ and $\nabla_{\bar{n}}$ defined respectively by

$$\Delta_{\bar{n}} = \frac{\max\left\{\left|\widetilde{X}_{\bar{n}|\bar{n}-1}(u) - \widehat{X}_{\bar{n}|\bar{n}}(u)\right|\right\}}{\widehat{X}_{\bar{n}|\bar{n}}(u^*)}, \tag{21}$$

where u^* is the value which maximizes the numerator in (21), and

$$\nabla_{\bar{n}} = \frac{\left\|\widetilde{X}_{\bar{n}|\bar{n}-1} - \widehat{X}_{\bar{n}|\bar{n}}\right\|}{\left\|\widehat{X}_{\bar{n}|\bar{n}}\right\|}, \tag{22}$$

where $\|f\| = \langle f, f\rangle^{1/2}$ is the $\mathcal{L}$-norm induced by the inner-product $\langle , \rangle$. That is, (21) compares the maximum distance between the predicted and the realized (filtered) density with the value of this realized density at the point of maximum distance, and (22) compares the norm of the error function $\widetilde{X}_{\bar{n}|\bar{n}-1} - \widehat{X}_{\bar{n}|\bar{n}}$ with the norm of the filtered pdf – thus $\Delta_{\bar{n}}$ is a pointwise error measure, while $\nabla_{\bar{n}}$ evaluates the prediction error along the whole interval $\mathcal{I}$.

We set $n_0 = 250$ such that we have 55 one-step-ahead forecasts. Table 3.4 displays summary statistics for the prediction-error measures $\Delta_{\bar{n}}$ and $\nabla_{\bar{n}}$ for $\bar{n} = 251, \ldots, 305$. Figure 3.11 plots the forecasts $\widetilde{X}_{\bar{n}|\bar{n}-1}$ together with the realized densities $\widehat{X}_{\bar{n}|\bar{n}}$ for selected values of $\bar{n}$. Table 3.5 displays the values of each metric at these dates. These include the worst and best predictions according to each measure, plus the closest-to-the-mean predictions for either measure, as well as two values of $\bar{n}$ selected at random. Note that the closest-to-the-mean forecasts can give an approximate visual idea of the average size of the respective prediction errors according to each metric.

4. Concluding remarks

In this paper, we have followed the methodology developed in Bathia et al. (2010) to model the dynamics of the pdf's of IBOVESPA intraday returns over business days. Therewith we have obtained filtered estimates of these intraday return densities, in the sense that whatever noise occurring in previously estimating them through non-parametric methods is removed. We also found that the dynamic behavior of these pdf's reduces to a $\mathbf{R}^2$-valued process, which is well represented by a VAR(1) model and whose effect on the pdf's shape along business days appears as a dispersion-symmetry shift. In terms of the Karhunen-Loève representation of the pdf's, this 2-vector process is shown to dynamically weight the deterministic eigenfunctions

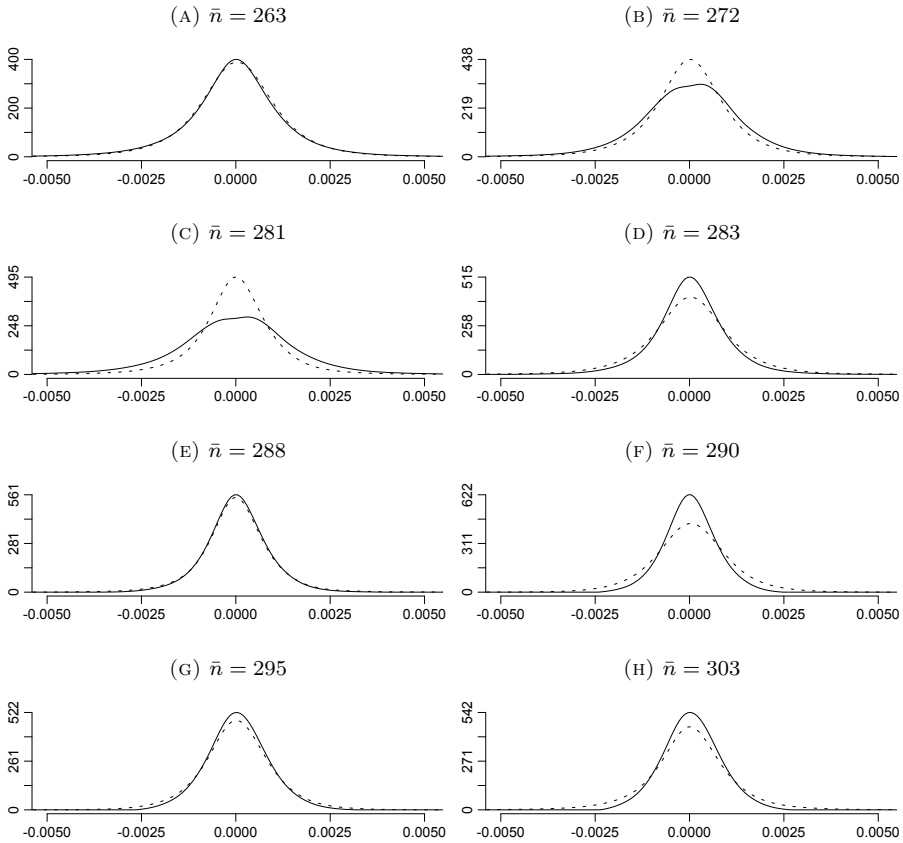

FIGURE 3.11. Pdf forecasts $\widetilde{X}_{\bar{n}|\bar{n}-1}$ (dotted) and realized filtered pdf's $\widehat{X}_{\bar{n}|\bar{n}}$ (solid); (A) $\bar{n} = 263$, best prediction according to $\nabla_{\bar{n}}$; (B) $\bar{n} = 272$, selected at random; (C) $\bar{n} = 281$, worst prediction according to both measures; (D) $\bar{n} = 283$, closest-to-the-mean $\Delta_{\bar{n}}$; (E) $\bar{n} = 288$, best prediction according to $\Delta_{\bar{n}}$; (F) $\bar{n} = 290$, selected at random; (G) $\bar{n} = 295$, selected at random; (H) $\bar{n} = 303$, closest-to-the-mean $\nabla_{\bar{n}}$. See Table 3.5 for the values of $\Delta_{\bar{n}}$ and $\nabla_{\hat{n}}$ at those dates.

$\bar{n}$	$\Delta_{\bar{n}}$	$\nabla_{\bar{n}}$
263	0.063	0.034
272	0.378	0.221
281	0.737	0.425
283	0.207	0.167
288	0.038	0.037
290	0.298	0.256
295	0.083	0.085
253	0.147	0.156

TABLE 3.5. Values of $\Delta_{\bar{n}}$ and $\nabla_{\bar{n}}$ for selected values of $\bar{n}$ corresponding to Figure 3.11.

in a way that it sequentially changes the dispersion and symmetry of the mean-function; that is, the eigenfunctions may be seen as a parametric shape-deforming structure which at each day deforms, at a stochastic degree, the mean distribution of the returns at that day. Moreover, by taking into account the dynamics of the curve process, the methodology allows for obtaining forecasts of upcoming periods pdf's. We have thus constructed one-step-ahead forecasts using subsamples and evaluated these forecasts according to appropriate metrics.

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

Many exploratory possibilities come in sight. For instance, an area which most likely represents a field of exploration 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 have rich information on the distributions of intraday returns of upcoming periods previously from sampling them. With such information at hand, it is possible to evaluate investment risk in a novel manner. Future work will likely embrace trading simulations using such strategies. Another, more academic, possibility of research is that of generating simulations into the specific context of asset pricing theory in order to further assess properties of the presented theory[16]. Of particular interest would be to parametrize the distributions and then simulate the price processes according to a certain specification of the stochastic behavior of asset prices, only to evaluate how the methodology works out on estimation of such underlying distributions based on observation of generated prices alone.

On theoretical grounds, incorporation of restrictions onto the functions $\widehat{X}_t$ could represent a step forward in applications where this restrictions are known to hold for the true functions; for example, in our application the $\widehat{X}_t$'s could by construction assume negative values and integrate not to unity, even though they represent pdf's. The fact that the obtained estimates nearly satisfied the restrictions illustrates the strength of the methodology, but it would still be interesting to restrict in principle some characteristics of the estimators. Another promising area of research is that of further developing the theory to functions defined on some Borel set over $\mathbf{R}^l$ with $l > 1$, or even on some manifold of interest. Guillas and Lai (2010) represent an early development in that direction. A natural extension to the work herein developed would be to apply the methodology to joint densities.

[16]Bathia et al. (2010) do simulate a curve process in their work, but not into the context of probability density functions.

References

Aït-Sahalia, Y., Mykland, P., and Zhang, L. Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*, 160: 160–175, 2011.

Bandi, F. M., and Russell, J. R. Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations. *Journal of Econometrics*, 160:145–159, 2011.

Bathia, N., Yao, Q., and Zieglemann, F. A. Identifying the finite dimensionality of curve time series. *The Annals of Statistics*, 38:3352–3386, 2010.

Benko, M., Härdle, W., and Kneip, A. Common functional principal components. *Ann. Statist*, 37:1–34, 2009.

Bosq, D. *Linear processes in function spaces: theory and applications.* Springer Verlag, 2000.

Dabo-Niang, S., and Ferraty, F., editors. *Functional and operatorial statistics.* Springer Verlag, 2008.

Duffie, D. *Dynamic asset pricing theory.* Princeton University Press, 2001.

Fan, J., and Yao, Q. *Nonlinear time series: nonparametric and parametric methods.* Springer Verlag, 2003.

Ferraty, F., and Vieu, P. *Nonparametric functional data analysis: theory and practice.* Springer Verlag, 2006.

Ghysels, E., Gouriéroux, C., and Jasiak, J. High frequency financial time series data: some stylized facts and models of stochastic volatility. In Dunis, C., and Zhou, B., editors, *Nonlinear Modelling of High Frequency Financial Time Series.* John Wiley & Sons Inc, 1998.

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

Guillas, S., and Lai, M.-J. Bivariate splines for spatial functional. *Journal of Nonparametric Statistics*, 22:477–497, 2010.

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

Hamilton, J. D. *Time series analysis.* Princeton University Press, 1994.

Hansen, P., and Lunde, A. Realized variance and market microstructure noise. *Journal of Business and Economic Statistics*, 24:127–161, 2006.

Müller, U., Dacorogna, M., and Pictet, O. Heavy tails in high-frequency financial data. In Adler, R., Feldman, R., and Taqqu, M., editors, *A practical guide to heavy tails: statistical techniques and applications.* Birkhäuser, 1998.

Pfaff, B. VAR, SVAR and SVEC models: Implementation within R package vars. *Journal of Statistical Software*, 27:1–32, 2008.

Ramsay, J., and Silverman, B. *Functional data analysis.* Springer Verlag, 1998.

Rydberg, T. Realistic statistical modelling of financial data. *International Statistical Review*, 68:233–258, 2000.

Rydberg, T., and Shephard, N. A modelling framework for the prices and times of trades made on the New York stock exchange. In Fitzgerald, W., Smith, R., Walden, A., and Young, P., editors, *Nonlinear and Nonstationary Signal Processing.* Cambridge University Press, 2000.