



UNIVERSIDAD CARLOS III DE MADRID

working
papers

Working Paper 09-28
Statistics and Econometrics Series 09
March 2009

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

MODELLING INTRA-DAILY VOLATILITY BY FUNCTIONAL DATA ANALYSIS: AN EMPIRICAL APPLICATION TO THE SPANISH STOCK MARKET.

Kenedy Alva^{*}, Juan Romo[†] and Esther Ruiz[‡]

Abstract

We propose recent functional data analysis techniques to study the intra-daily volatility. In particular, the volatility extraction is based on functional principal components and the volatility prediction on functional AR(1) models. The estimation of the corresponding parameters is carried out using the functional equivalent to OLS. We apply these ideas to the empirical analysis of the IBEX35 returns observed each $_ve$ minutes. We also analyze the performance of the proposed functional AR(1) model to predict the volatility along a given day given the information in previous days for the intra-daily volatility for the firms in the IBEX35 Madrid stocks index.

Keywords: Market microstructure; Ultra-high frequency data; Functional data analysis; Functional AR(1) model.

^{*} Corresponding author. Dpto. Estadística, Universidad Carlos III de Madrid, Avda. Universidad 30, 28911 Leganés (Spain); Tel.: 34 916246201; Fax: 34 916249849; e-mail address: kenedypedro.alva@uc3m.es

[†] Financial support from the Spanish government project SEJ2005-06454 acknowledged by the second author. The usual disclaims apply.

[‡] Financial support from the Spanish government project SEJ2006-03919 is acknowledged by the third author. The usual disclaims apply.

Modelling intra-daily volatility by functional data analysis: An empirical application to the Spanish Stock Market.

Kenedy Alva^{*}, Juan Romo[†] and Esther Ruiz[‡]
Dpto. de Estadística, Universidad Carlos III de Madrid

February 2009

Abstract

We propose recent functional data analysis techniques to study the intra-daily volatility. In particular, the volatility extraction is based on functional principal components and the volatility prediction on functional AR(1) models. The estimation of the corresponding parameters is carried out using the functional equivalent to OLS. We apply these ideas to the empirical analysis of the IBEX35 returns observed each five minutes. We also analyze the performance of the proposed functional AR(1) model to predict the volatility along a given day given the information in previous days for the intra-daily volatility for the firms in the IBEX35 Madrid stocks index.

KEY WORDS: Market microstructure; Ultra-high frequency data; Functional data analysis; Functional AR(1) model.

^{*}Corresponding author. Dpto. Estadística, Universidad Carlos III de Madrid, Avda. Universidad 30, 28911 Leganés (Spain); Tel.: 34 916246201; Fax: 34 916249849; e-mail address: kenedypedro.alva@uc3m.es

[†]Financial support from the Spanish government project SEJ2005-06454 acknowledged by the second author. The usual disclaims apply.

[‡]Financial support from the Spanish government project SEJ2006-03919 is acknowledged by the third author. The usual disclaims apply.

1 Introduction

Recently, in the context of financial markets, there has been an increasing availability of intra-daily and ultimately transaction data called by Engle (2000) ultra-high-frequency (UHF) data. Modelling this kind of data may have important implications for market microstructure theory. Consequently, the development of new methods to describe the dynamics of UHF data is receiving a great amount of attention; see, for example, Engle (2001) and Bollerslev (2001) who include these methods among the hot topics in the analysis of financial time series. Potentially, these models should yield valuable information for market designers and risk management. One particular characteristic of UHF that has attracted a great deal of attention in the literature is intra-daily volatility. There are several procedures proposed with this goal. First, several authors extend the GARCH-type models, which has been very popular to model daily volatility, to the analysis of UHF second order dynamic dependencies; see, for example, Baillie and Bollerslev (1990), Andersen and Bollerslev (1997, 1998), Engle (2000), Gramming and Wellner (2002) and Manganelli (2005) who assume GARCH(1,1) models and, more recently, Baillie *et al.* (2007) and Bordignon *et al.* (2007, 2009) who model intra-daily volatilities by assuming long memory GARCH processes. However, because UHF data can be irregularly spaced, traditional GARCH and Stochastic Volatility (SV) models are ill suited to represent its dynamics. Consequently, several authors have suggested the use of market point processes or continuous time methods. For example, Eraker *et al.* (2003) proposed using Monte Carlo Markov Chain (MCMC) methods to estimate continuous time stochastic volatility models with jumps. They fit their model to estimate the underlying volatility of daily returns but their procedure can also be implemented to other intraday frequencies. An alternative methodology in the related literature is based on fitting function wavelets to approximate the underlying intra-daily volatilities as in Nielsen and Frederiksen (2007).

Another problem faced when modelling UHF data is that financial markets usually exhibit strong periodic dependencies across the trading day. Typically, the volatility is highest at the open and toward the close of the day; see, for example, Andersen and Bollerslev (1997) and Andersen *et al.* (2001) among many others. To represent this seasonal dependence, Baillie and Bollerslev (1990) include a daily lag in the GARCH model for hourly data. Later, Bollerslev and Ghysel (1996) proposed a GARCH model with periodically varying parameters which allows for a greater degree of flexibility when modelling the periodicity in volatilities. Bisaglia *et al.* (2003) have used k -factor Gegenbauer ARMA (GARMA) models to describe intra-daily volatilities. However, this approach may be ineffective when the goal is the prediction of future returns. Arteche (2004) proposes to treat long and periodic persistence in the context of SV models. Recently, Bordignon *et al.* (2009) introduces a new class of models called Periodic Long-Memory GARCH which introduced a fractional seasonal difference in the GARCH specification.

However, the procedures just described are not able to cope simultaneously with irregular data and seasonality. The prevailing estimation methods for

continuous time models, which are able to cope with irregularly spaced observations, remain relatively complicated and, in general, have not been extended to deal with the seasonal pattern of intra-daily volatilities.

In this paper, we propose to use Functional Data Analysis (FDA) to model the dependence of the volatility along a given day with respect to the volatility during the previous day. FDA procedures are easy to implement and allow to cope simultaneously with irregularly spaced data and periodic dependencies. With this purpose, the stock prices observed along a given day, say day i , are considered as a unique observation which is denoted by $p_i(t_j)$ where $i = 1, \dots, n$, with n the sample size, i.e., the number of days in the sample, and t_j are the moments within the day at which we observe the prices with $j = 1, \dots, J$ and J the total number of data points within the day. Although FDA can be easily implemented for nonregular observations, in this paper, we will focus on regularly observed data and, consequently, $t_j = t_1 + \Delta j$, where t_1 is the moment when the first observation in the day is obtained and Δ is the time span between observations. Our objective is to predict the volatility along day i using information of the volatility up to, and including, day $i - 1$.

FDA has been previously implemented by other authors for the analysis of financial data with different objectives. As far as we know, the first paper proposing to use FDA in the context of financial returns was Anderson and Newbold (2002) who model the time-varying unconditional distribution of intra-daily returns of the Swiss-Franc-Dollar exchange rates. More recently, Müller, Stadtmüller and Yao (2006) and Müller, Sen and Stadtmüller (2006) propose FDA to estimate and predict the volatility during the second part of the day using information contained in the first part of the day, respectively. Many of the procedures implemented in this paper are based on these two last works. However, we introduce a functional AR(1) model in order to represent the dependence of the volatility during one particular day on the volatility functions of previous days.

The rest of the paper is organized as follows. Section 2 describes the FDA techniques implemented in order to extract the volatility and to predict future volatilities. In particular, the volatility extraction is based on functional principal components and the volatility prediction on functional AR(1) models. The estimation of the corresponding parameters is carried out using the functional equivalent to OLS. Section 3 contains the empirical analysis of the IBEX35 returns observed each 5 minutes. Finally, Section 4 concludes the paper.

2 Functional data analysis for volatility extraction and prediction

In this section, we describe the idea of functional principal components implemented to extract the volatility, as well as the functional AR(1) model considered to predict future volatilities.

2.1 Functional principal components

The volatility of returns is a latent variable that cannot be directly observed. Functional principal components can be used to extract the underlying volatilities; see Müller, Stadtmüller and Yao (2006). Functional principal components, proposed by Ramsay and Silverman (2005), is based on the Karhunen-Loève theorem which establishes that any centered process of order two can be expressed as a combination of orthonormal functions. Given that the price functions are not stationary, first we transform the prices plotted in Figure 1 into returns as follows:

$$r_i(t_j) = \log \left(\frac{p_i(t_j)}{p_i(t_{j-1})} \right), \quad i = 1, \dots, n. \quad (1)$$

Then, we consider the absolute values $y_i(t_j) = |r_i(t_j)|$. These absolute returns functions can be seen in Figure 2. Now, volatility $V_i(t)$ is defined as smooth functional of absolute logarithmic returns and can be decomposed as

$$V_i(t) = \mu_V(t) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(t), \quad (2)$$

where $\mu_V(t)$ is the mean function of the volatility and $\phi_k(t)$ are orthogonal eigenfunctions associated to the corresponding principal components, i. e.,

$$\int_{\mathcal{T}} \phi_k(t) \phi_l(t) dt = \delta_{kl}, \quad \delta_{kl} = \begin{cases} 1, & k = l \\ 0, & k < l. \end{cases} \quad (3)$$

and ξ_{ik} are the principal components scores defined by

$$\xi_{ik} = \int_{\mathcal{T}} (V_i(t) - \mu_V(t)) \phi_k(t) dt. \quad (4)$$

These principal components have zero mean, variance λ_k and

$$\int_{\mathcal{T}} \text{cov}(V_i(s), V_i(t)) \phi_k(s) ds = \lambda_k \phi_k(t), \quad (5)$$

Volatility is estimated through the estimation of $\mu_V(t)$, $\phi_k(t)$, and ξ_{ik} . We first estimate the mean volatility function $\mu_V(t)$ by local linear smoothing, finding $\hat{\alpha}_0$ and $\hat{\alpha}_1$ that minimize

$$\sum_{i=1}^n \sum_{j=1}^J \kappa_1 \left(\frac{t_j - t}{b_V} \right) \{y_i(t_j) - \alpha_0 - \alpha_1(t - t_j)\}^2, \quad (6)$$

where $\kappa_1(x) = 0.75(1 - x)x$ is a Epanechnikov kernel function ($\|x\| \leq 1$) and b_V is the smoothing bandwidth of volatility. Finally, we take $\hat{\mu}_V(t) = \hat{\alpha}_0(t)$. See Müller, Sen and Stadtmüller (2006).

Before estimating the eigenfunctions $\phi_k(t)$, it is necessary to estimate the covariances $\text{cov}(V(s), V(t))$ which are obtained from the empirical covariances

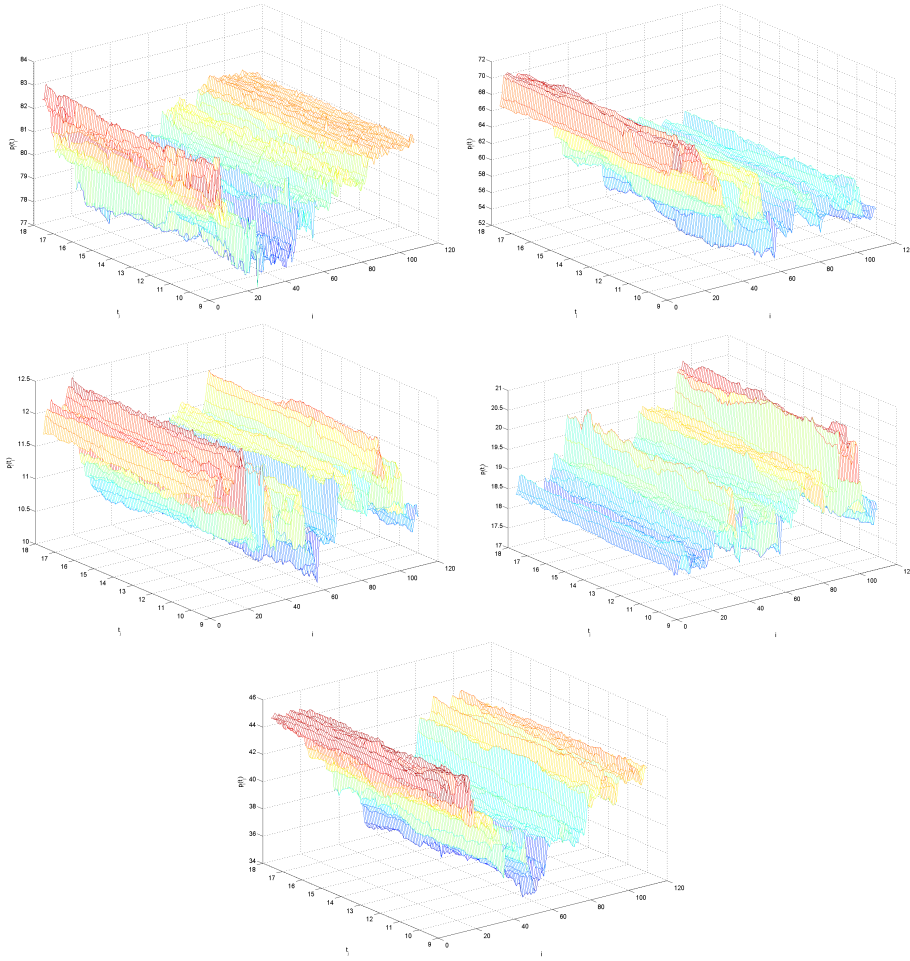


Figure 1: Prices for 110 days of METROVACESA stock (top left), FCC (top right), CINTRA (middle left), INDRA (middle right) and GAS NATURAL (bottom center) from 9:05 to 17:40.

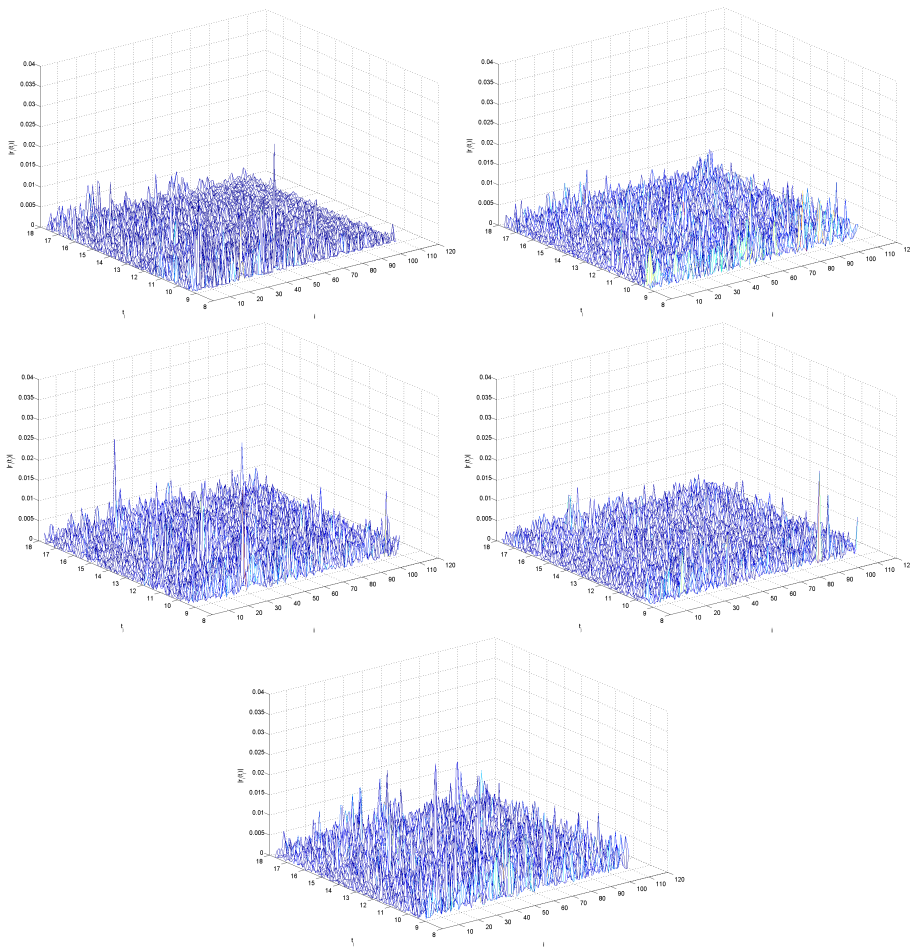


Figure 2: Absolute logarithmic returns for 110 days of METROVACESA stock (top left), FCC (top right), CINTRA (middle left), INDRA (middle right) and GAS NATURAL (bottom center) from 9:10 to 17:40.

$G_i(t_{j_1}, t_{j_2}) = (y_i(t_{j_1}) - \hat{\mu}_V(t_{j_1}))(y_i(t_{j_2}) - \hat{\mu}_V(t_{j_2}))$ as follows. Fit a two-dimensional smoother to obtain the nonparametric regression of $G_i(t_{j_1}, t_{j_2})$ versus (t_{j_1}, t_{j_2}) by finding the functions $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2$ that minimize

$$\sum_{i=1}^n \sum_{1 \leq j_1 \neq j_2 \leq J} \kappa_2 \left(\frac{t_{j_1} - s}{h_V}, \frac{t_{j_2} - t}{h_V} \right) \{G_i(t_{j_1}, t_{j_2}) - [\alpha_0 - \alpha_1(s - t_{j_1}) - \alpha_2(t - t_{j_2})]\}^2, \quad (7)$$

where $\kappa_2(\cdot, \cdot)$ is a bivariate kernel function. The estimated covariances are given by $\widehat{cov}(V(s), V(t)) = \hat{\alpha}_0(s, t)$.

The next step is the estimation of the eigenfunctions $\phi(t)$ and the eigenvalues λ , replacing the covariance surface $cov(V(t), V(s))$ by $\widehat{cov}(V(t), V(s))$. We obtain $(\hat{\lambda}_k, \hat{\phi}_k(t))$ by numerical eigenanalysis (Yao et al., 2005). The estimations of the functional principal components (FPC) scores are given by numerical integration,

$$\hat{\xi}_{ik} = \sum_{j=2}^J (y_i(t_j) - \hat{\mu}_V(t_j)) \hat{\phi}_k(t_j), \quad i = 1, \dots, n, \quad k = 1, 2, \dots, \quad (8)$$

and the estimate of the volatility is

$$\hat{V}_i(t) = \hat{\mu}_V + \sum_{k=1}^K \hat{\xi}_{ik} \hat{\phi}_k(t). \quad (9)$$

2.2 Prediction of Volatility

Once the volatility has been extracted, our goal is to predict the volatility along a given day using the information contained in the volatility functions of previous days. With this goal, we fit the following functional AR(1) model

$$V_i(t) = \mu_V(t) + \int_{\mathcal{T}} \beta(s, t) (V_{i-1}(s) - \mu_V(s)) ds + \eta_i(t) \quad (10)$$

where η_i are i.i.d second order random variables assuming values in the Hilbert space H ; see Bosq (1991). The regression parameter function β can be represented by

$$\beta(s, t) = \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{\gamma_{km}}{\lambda_k} \phi_k(s) \phi_m(t), \quad (11)$$

where $\gamma_{km} = E(\xi_k \xi_m)$ and $\lambda_k = E(\xi_k^2)$. Note that the functional AR(1) model in (11) resembles the Stochastic Volatility model for daily observations proposed by Harvey *et al.* (1994), where the log-volatility was assumed to follow an AR(1) process.

The estimation of $\beta(s, t)$ is according to (12). We construct estimates $\hat{\lambda}_k$ for λ_k and estimates

$$\hat{\gamma}_{km} = \int \int \hat{\phi}_k(s) \widehat{cov}(V_{i-1}(s), V_i(t)) \hat{\phi}_m(t) ds dt, \quad (12)$$

for the γ_{km} , where $\widehat{cov}(V_{i-1}(s), V_i(t))$ is a local linear smoother for the cross-covariance function $cov(V_{i-1}(s), V_i(t))$; see Yao et al. (2005). Then, the estimate for the regression parameter function is

$$\hat{\beta}(s, t) = \sum_{k=1}^K \sum_{m=1}^K \frac{\hat{\gamma}_{km}}{\hat{\lambda}_k} \hat{\phi}_k(s) \hat{\phi}_m(t), \quad (13)$$

where K is the number of components that represent functional data $V_{i-1}(s)$ and $V_i(t)$.

On the other hand, the coefficient of determination R^2 plays an important role in applications of regression analysis. Yao et al. (2005) propose a extension to functional linear regression. In our case to measure the global association between the functional predictor $V_{i-1}(t)$ and the functional response $V_i(t)$, we obtain

$$R^2 = \frac{\int_{\mathcal{T}} var(E[V_i(t)|V_{i-1}(t)])dt}{\int_{\mathcal{T}} var(V_i(t))dt} = \frac{\sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \gamma_{km}/\lambda_m}{\sum_{k=1}^{\infty} \lambda_k}, \quad (14)$$

where R^2 always satisfies $0 \leq R^2 \leq 1$. A estimate \hat{R}^2 for the functional coefficient determination R^2 is

$$\hat{R}^2 = \frac{\sum_{k=1}^K \sum_{m=1}^K \hat{\gamma}_{km}/\hat{\lambda}_m}{\sum_{k=1}^K \hat{\lambda}_k}, \quad (15)$$

where $\hat{\gamma}_{km}$ are as in (13).

The confidence interval $\beta(s, t)$ for a confidence level $100(1-\alpha)\%$ and hypothesis testing of parameter function $H_0 : R^2 = 0$ is based on bootstrapped sampling, returns decision binary and p-value (see PACE package¹) by means the following steps:

- a) Generate N random sample of $\hat{\xi}_{1k}, \dots, \hat{\xi}_{Nk}$ from $\hat{\xi}_{1k}, \dots, \hat{\xi}_{nk}$ calculated of (9).
- b) Calculate $\hat{V}_{i-1}(s), \hat{V}_i(t)$ from (10) and $\hat{\gamma}_{km}$ from (13) with each one of the random sample generated in a).
- c) Calculate $\hat{\beta}_1(s, t), \dots, \hat{\beta}_N(s, t)$, accumulate frequency estimated \hat{F} , lower $\hat{F}^{-1}(\alpha/2)$ and upper confidence $\hat{F}^{-1}(1-\alpha/2)$ by means of quantiles. Then calculate $\hat{R}_1^2, \dots, \hat{R}_N^2$.

¹Principal Analysis by conditional Expectation package in this web page <http://anson.ucdavis.edu/~btliu/PACE/download.html>

d) For the hypothesis test, we have

$$\mathbf{p} - \mathbf{value} = \sum_{i=1}^N I(\hat{R}_i^2 > \hat{R}^2)/N \quad (16)$$

where $I(\cdot)$ is equal to 1 when the condition within the parenthesis (\cdot) is true and zero otherwise. We reject the null hypothesis when $\mathbf{p} - \mathbf{value} < \alpha$ and \hat{R}^2 is according to (16).

Finally, the estimated prediction of a new trajectory $V_{i+1}(t)$ given $V_i(t)$ with FPC scores ξ_{ik} is obtained as

$$\hat{V}_{i+1}(t|V_i) = \hat{E}(V_{i+1}(t)|V_i) = \hat{\mu}_V(t) + \int \hat{\beta}(s, t)(V_i(s) - \hat{\mu}_V(s)) ds \quad (17)$$

or alternatively

$$\hat{V}_{i+1}(t|V_i) = \hat{\mu}_V(t) + \sum_{k=1}^K \sum_{m=1}^K \hat{\xi}_{ik} \frac{\hat{\gamma}_{km}}{\hat{\lambda}_k} \hat{\phi}_m(t), \quad (18)$$

where $\hat{\xi}_{ik}$ is according to (9).

2.3 Clustering K-means

We can form clusters for the intradaily volatility of the stocks listed in the IBEX35 index of the Madrid Stock Market, using the regression parameter function. MacQueen (1967) proposes the K -means as one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The main idea is to define G centroids, one for each cluster. These centroids should be placed in a cunning way since location greatly affects the results. The algorithm is composed of the following steps:

- a) Place G points into the space represented by the objects that are being clustered. These points represent initial group centroids.
- b) Assign each object to the group that has the closest centroid.
- c) When all objects have been assigned, recalculate the positions of the G centroids.
- d) Repeat Steps b) and c) until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

Minimizing the objective function, in this case the sum of squared error for the regression parameter function (SSERP), we have

$$\min SSERP(G) = \min \sum_{g=1}^G \sum_{l=1}^{n_g} d_{lg}^2, \quad (19)$$

where

$$d_{lg}^2 = \sum_{j_2=2}^J \sum_{j_1=2}^J |\hat{\beta}_{lg}(t_{j_1}, t_{j_2}) - \bar{\beta}_g(t_{j_1}, t_{j_2})|^2, \quad (20)$$

is the squared Euclidean distance, $\hat{\beta}_{lg}(t_{j_1}, t_{j_2})$ is the estimated regression parameter function of the l -th stock listed in the IBEX35 index for the cluster g , where

$$\bar{\beta}_g(t_{j_1}, t_{j_2}) = \frac{1}{n_g} \sum_{l=1}^{n_g} \hat{\beta}_{lg}(t_{j_1}, t_{j_2}) \quad (21)$$

is the average of estimated regression parameter function and n_g is the numbers of stocks for the g -th cluster.

In the usual application for the algorithm k-means, we must determine the number of groups G . This number can not be determined with a criterion of homogeneity. The way to get very homogeneous groups and minimize the *SSERP* is to make as many groups as observations, so you always have *SSERP* = 0.

Have been proposed various methods to select the number of groups. A procedure that is used is to do the F-test for variability reduction, for G and $G + 1$ groups, and calculating the relative reduction of the variability with increasing an additional group. The test is:

$$F = \frac{SSERP(G) - SSERP(G + 1)}{SSERP(G + 1)/(n - G - 1)} \quad (22)$$

and compares the decrease of the variability to increasing a group with average variance. The value obtained is compared to F with p and $p(n - G - 1)$ degrees of freedom where p are the total times that have been taken at (s, t) . If the p -value $< \alpha$ then we need to increase a cluster, otherwise we keep the clusters. But in this rule the data do not have to verify the assumptions required to implement the distribution F . A rule suggested by Hartigan(1975) is introducing a group if this ratio is greater than 10.

3 Empirical application to IBEX35 stocks

In this section, we extract and predict the pattern of intradaily volatilities of the stocks listed in the IBEX35 index of the Madrid Stock Exchange. Our data, described in the previous section, has been divided in two subperiods, one for the extraction of volatilities from July 2nd until October 19th 2007 (80 days) and the rest from October 22th until November 30th 2007 (30 days) for out-of sample prediction.

Figure 1 and 2 show the prices evolution and absolute logarithmic returns of METROVACESA stock, FCC, CINTRA, INDRA and GAS NATURAL. These graphs are represented from stock listed in the IBEX35 index that are closed to centroid for each cluster. The evolution of prices for the companies is not the same. However METROVACESA and GAS NATURAL have a similar behavior due to lower prices until the mid-term and increase until the end of period.

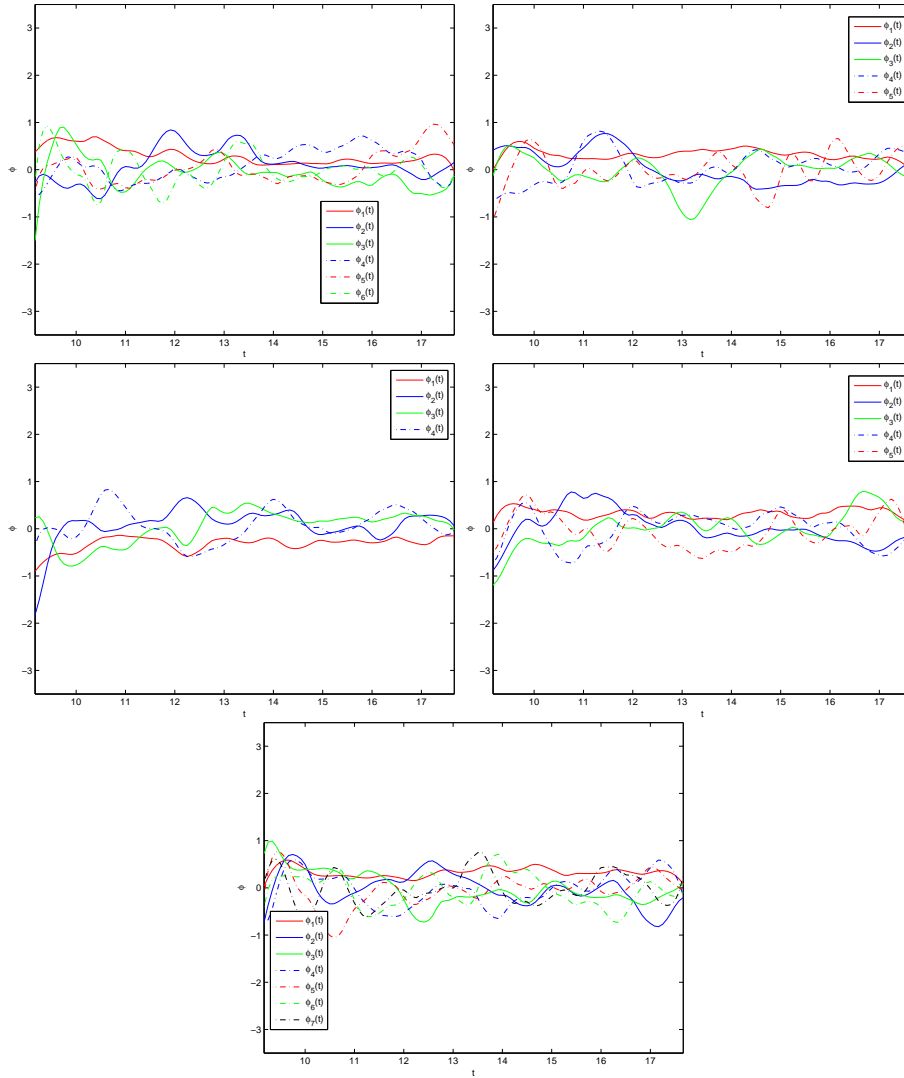


Figure 3: Eigenfunctions for METROVACESA stock (top left), FCC (top right), CINTRA (middle left), INDRA (middle right) and GAS NATURAL (bottom center) from 9:10 to 17:40.

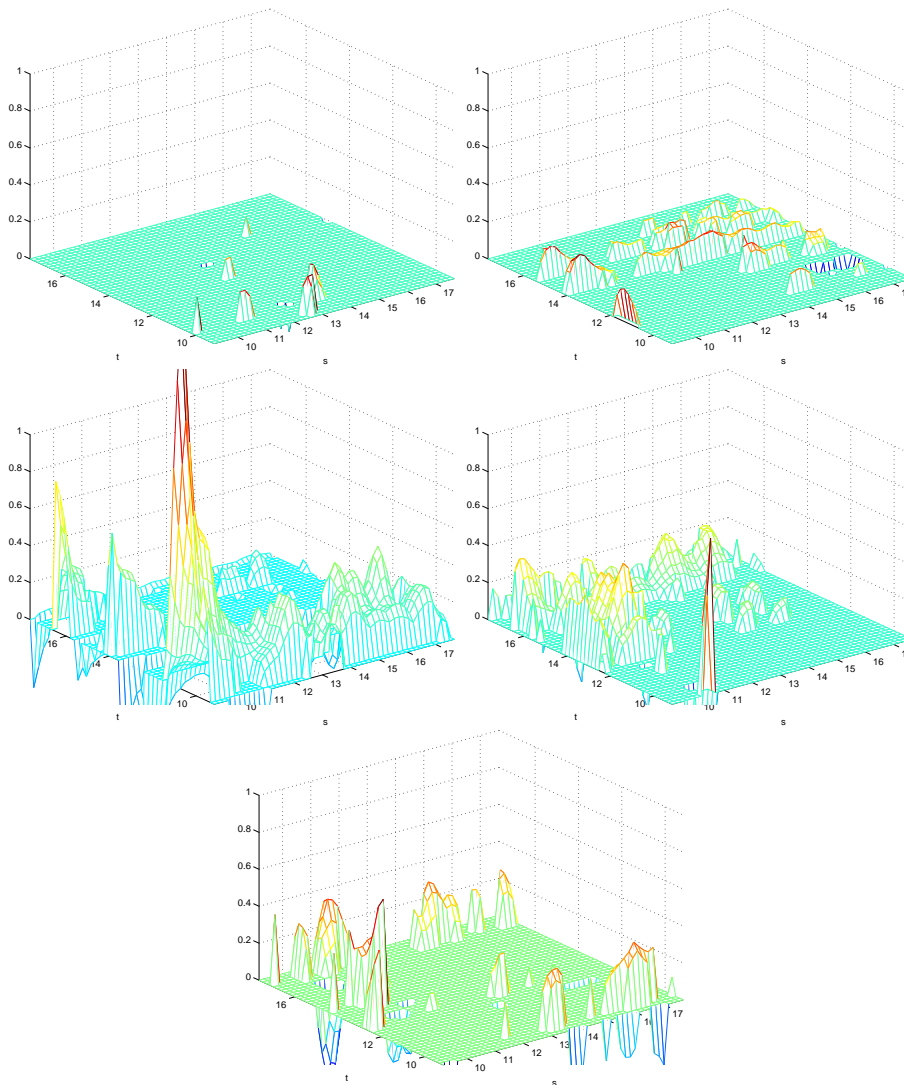


Figure 4: Significant $\beta(s, t)$ for of METROVACESA stock (top left), FCC (top right), CINTRA (middle left), INDRA (middle right) and GAS NATURAL (bottom center) from 9:10 to 17:40.

Figure 3 shows the eigefunctions to calculate the functional principal component scores and volatility intradaily given in (9) and (10), where we select 6, 5, 4, 5 and 7 eigenfunctions. So with this information, we estimate $\beta(s, t)$ and in Figure 4 shows the graphic of regression parameter function for these stock volatility intradaily with significant areas. For METROVACESA, there are few significant areas to predict the volatility. GAS NATURAL has significant areas throughout the day to predict the second part of the next day.

Using the coefficient of determination and significant level $\alpha = 0.05$, we reject the null hypothesis of $R^2 = 0$ for 27 of the 33 stocks listed in the IBEX35 index of the Madrid Stock Market, where we reject the null hypothesis of no regression relation. The intradaily volatilities $V_{i-1}(t)$ and $V_i(t)$ are significantly related in these 27 stock listed. We will use significantly stocks to build clusters. Using the K-means algorithm and squared Euclidean distance, we have the Table 1, where we don't reject the null (p-value>0.05) for increase 5 clusters and it doesn't need to increase one more cluster (see Table 2), being the number of clusters 4. The rule of Hartigan is not applicable because the value of F doesn't exceed 10 to reject the null hypothesis and the degrees of freedom of $p = 103 \times 103$ and $p(n - G - 1) = 103 \times 103 \times 22$, giving values of $F_{1-0.05}$ close to 1.

Table 1: Clusters for many groups

Clusters G	SSERP	F	p-value
2	945,61	0	0
3	844,38	2,8773	0
4	769,42	2,2409	0
5	737,77	0,94381	0,9629

In figure 5, we show the absolute returns functions and predictions of GAS NATURAL with the purpose of comparing the 30 last days, 9 randomly select. We can see that the intradaily volatility predicted are within the empirical volatility. We can predict the intradaily volatility from 9:10 to 10:00 and 12:30 to 17:40 using intradaily volatility taken from previous day.

Table 2: Clustering for the Stock listed of IBEX 35 index

g-th cluster	N.	Stock
Not sig.	01	ANTENA 3 TV.
	20	INDITEX
	21	MAPFRE
	22	METROVACESA
	25	R.E.E.
	29	SOGECABLE
1	02	ABERTIS
	04	ACERINOX
	05	ALTADIS
	07	BBVA
	08	BANKINTER
	10	ENDESA
	11	ENAGAS
	13	FCC
	14	FERROVIAL
	15	GAMESA
	18	IBERIA
	23	BA.POPULAR
	24	NH. HOTELES
	27	BA.SABADELL
	32	AGS. BARCELONA
35	UNION FENOSA	
2	09	CINTRA
3	06	ACCIONA
	17	IBERDROLA
	19	INDRA A
	26	REPSOL YPF
	28	BSCH
	31	TELEFONICA
	33	TELECINCO
4	03	ACS
	16	GAS NATURAL
	30	SACYR VALLE.

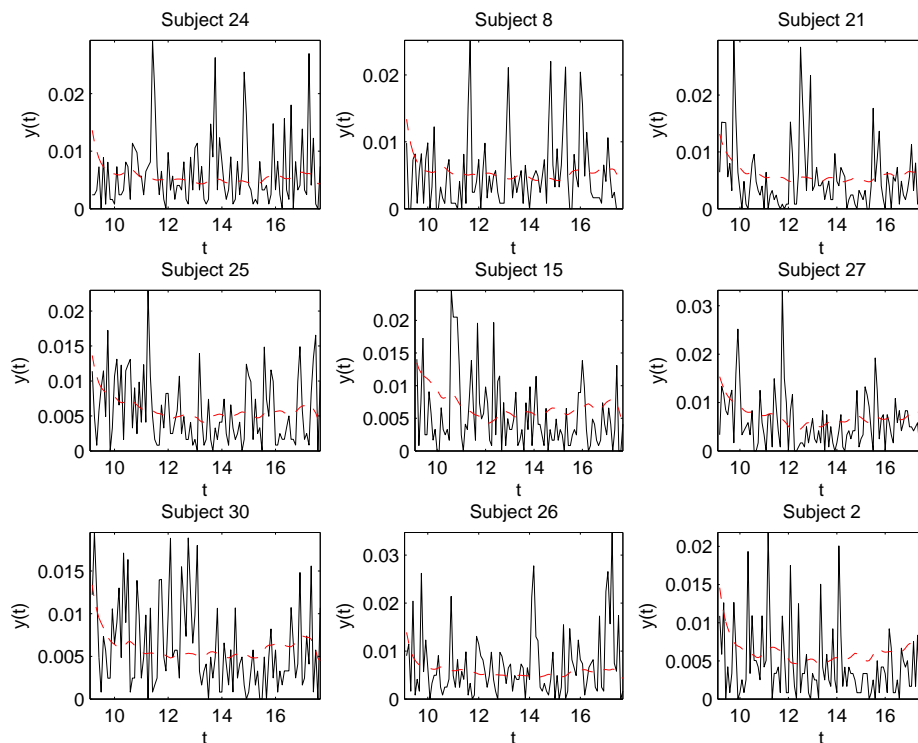


Figure 5: Absolute returns $y_i(t)$ (black line) versus predicted volatility $\hat{V}_i(t)$ (red line) for GAS NATURAL

4 Conclusions

A problem faced when dealing with ultra-high frequency data is that they are not regularly spaced. In this paper, we have transformed the data into regularly spaced data. However, FDA can be easily extended to deal with non-equally spaced observations.

A topic of further research is how to extend these methods to analyse the relationship between volatility, volume and durations which is of central interest for market microstructure; see, for example, Engle (2000).

In other extensions, we can use the proposed method to extract the volatility with daily log returns ($\log\{X_i(t)/X_{i-1}(t)\}$) and predict this volatility. We also consider alternatives models to estimate the regression parameter function to make prediction.

References

- [1] Andersen, T.G. and T. Bollerslev (1997), Intraday periodicity and volatility persistence in financial markets, *Journal of Empirical Finance*, 4, 115-158.
- [2] Andersen, T.G., and T. Bollerslev (1998). Answering the sceptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39 (4), pp. 885-905.
- [3] Andersen, T.G., T. Bollerslev and A. Das (2001), Variance-ratio statistics and high-frequency data: testing for changes in intraday volatility patterns, *Journal of Finance*, 56(1), 305-327.
- [4] Andersson, J. and Newbold, P. (2002). Modeling the distribution of financial returns by functional data analysis. Research Report 4. Department of Information Science, Division of Statistics, Uppsala University.
- [5] Arteche, J. (2004), Gaussian semiparametric estimation in long memory in stochastic volatility and signal plus noise models, *Journal of Econometrics*, 119, 79-85.
- [6] Baillie, R.T. and T. Bollerslev (1990), Intra-day and inter-market volatility in foreign exchange rates, *Review of Economic Studies*, 58, 565-585.
- [7] Baillie, R.T. et al. (2007), Long memory models for daily and high frequency commodity future returns, *Journal of Future Markets*, 27, 643-668.
- [8] Bisaglia, L., S. Bordignon and F. Lisi (2003), k -Factors GARMA models for intraday volatility forecasting, *Applied Economics Letters*, 10(4), 251-254.
- [9] Bollerslev, T. (2001), Financial econometrics: Past developments and future challenges, *Journal of Econometrics*, 100, 41-51.
- [10] Bollerslev, T. and E. Ghysel (1996), Periodic autoregressive conditional heteroscedasticity, *Journal of Business and Economic Statistics*, 14(2), 139-151.
- [11] Bordignon et al. (2007), Generalized long-memory GARCH models for intra-daily volatility, *Computational Statistics and Data Analysis*, 51, 5900-5912.
- [12] Bordignon et al. (2009), Periodic long-memory GARCH models, *Econometric Reviews*, 28(1-3), 60-82..
- [13] Bosq, D. (1991). Modelization, non-parametric estimation and prediction for continuous time processes. In Roussas, G., editor, *Nonparametric Functional Estimation and Related Topics*.
- [14] Bosq, D. (2000). *Linear Processes in Function Spaces. Theory and Applications*. Springer.

- [15] Engle, R. (2000), The econometrics of ultra high frequency data, *Econometrica*, 68, 1-22.
- [16] Engle, R. (2001), Financial econometrics - A new discipline with new methods, *Journal of Econometrics*, 100, 53-56.
- [17] Eraker, B., M. Johannes and N. Polson (2003), The impact of jumps in volatility and returns, *Journal of Finance*, 58(3), 1269-1300.
- [18] Gramming, J. and M. Wellner (2002), Modelling the interdependence of volatility and inter-transaction duration processes, *Journal of Econometrics*, 106, 369-400.
- [19] Hartigan, J. A. (1975), *Clustering Algorithms*, New-York, Wiley.
- [20] Harvey, A., Ruiz, E., and Shephard, N. (1994). Multivariate Stochastic Variance Models. *Review of Economic Studies*.
- [21] MacQueen, J. B. (1967), Some Methods for classification and Analysis of Multivariate Observations, *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1:281-297.
- [22] Manganelli, S. (2005), Duration, volume and volatility impact of trades, *Journal of Financial Markets*, 8, 377-399.
- [23] Müller, H.-G., Stadtmüller, U., and Yao, F. (2006). Functional Variance Processes. *Journal of the American Statistical Association*.
- [24] Müller, H.-G., Sen, R., and Stadtmüller, U. (2006). Functional Data Analysis for Volatility. *Manuscript*.
- [25] Nielsen, M.O. and Frederiksen, P. (2007), Finite sample accuracy and choice of sampling frequency in integrated volatility estimation. *Journal of Empirical Finance*.
- [26] Preda, C. (2007). Regression models for functional data by reproducing kernel Hilbert spaces methods. *Journal of Statistical Planning and Inference*, 137, 829-840.
- [27] Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis*. Springer Series in Statistics. Second Edition.
- [28] Sen, R. (2007). Jumps and Microstructure noise in Stock Price Volatility. *Preprint submitted to Elsevier*.
- [29] Yao, F., Müller, H.-G. and Wang, J.L. (2005). Functional linear regression analysis for longitudinal data. *Annals of Statistics*, 33, 2873-2903.