9. NON-LINEAR VOLATILITY MODELING OF ECONOMIC AND FINANCIAL TIME SERIES USING HIGH FREQUENCY DATA[1]

Marius MATEI[*]

Abstract

The current work undertakes an overview of the forecasting volatility with high frequency data topic, attempting to answer to the fundamental latency problem of return volatility. It surveys the most relevant aspects of the volatility topic, suggesting advantages and disadvantages of each alternative in modeling. It reviews the concept of realized volatility and explains why forecasting of volatility is more effective when the model contains a measure of intraday data. A discrete and a continuous time model are defined. Sampling methods at different frequencies are reviewed, and the impact of microstructure noise is considered. Details on procedures employed in the literature with respect to modeling and forecasting using realized models are discussed, while an empirical exercise will prove the advantages of using measures of high frequency data.

Keywords: High frequency, Volatility, Modeling, Forecasting, Realized measures, Microstructure noise

JEL Classification: C32, C53, C58

1. General issues on modeling volatility

At the same time with the markets' development in terms of size, complexity and sophistication, there was need for better modeling of volatility in financial and economic time series. Over time, it became also evident that returns at high frequencies are difficult to accurately predict and that the volatility of such returns is better forecastable. This and the remarkable importance of volatility in applied contexts are the central reasons for which financial econometrics dedicated that much attention to modeling financial volatility, playing an essential role in modern pricing

[1] *This paper corresponds to the Chapter 3 of 'Non-Linear Modeling in Economics - Beyond Standard Economics', published at Expert Publishing House, Bucharest 2011*

[*] *ESADE Business School, Department of Finance, Ramon Llull University, Spain and National Institute for Economic Research, Romanian Academy, Romania.*
Email: marius.matei@esade.edu.

and risk management theory. I n the study of financial economics, one of the most researched topics that appears to be essential in describing the fluctuation of any financial or economic time series seems to be the distributional pattern of returns, no matter it is about exchange rates, stocks, bonds, etc. Conclusions that may be grasped from the study of conditional distributions may say a lot on how to price a specific instrument, how to allocate funds according to a specific portfolio, how to measure risk, performance and, finally, how to undertake the management decision process. The distributional pattern is highly connected to other features of a portfolio, like conditional return fractiles that determine the probability that extreme jumps occur in portfolio value.

The most important characteristic of a distribution of a conditional return series is the structure of the second moment. As such, in modeling volatility the frequency at which the data was selected became important, over time an increasing preoccupation for sampling data at lower intervals being observed. This allowed that more information be incorporated into the models, becoming obvious that sampling at shorter intervals may be beneficial to both short and long term horizon forecasting.

The serial correlation in the volatility of the financial asset returns paved the way to an extremely rich literature and research that have been written on the topic of modeling and forecasting volatility. Such volatility is typically modeled in empirical contexts with daily data starting from GARCH[3]-type models or stochastic volatility processes that consider volatility as a latent, unobserved variable. Although returns may be measured with minimal measurement error, since they are constructed from face prices of assets, and according to that they may be analyzed with ordinary time series techniques, volatility needs more careful and complex computational modeling due to the latent property of volatility. In this regard Merton (1980) observed that the conditional variance over a fixed period can be expressed, arbitrarily but still satisfactorily accurate, under a sum of squared realizations, when data are available at such high sampling frequency. Andersen and Bollerslev (1998) also provided an argument for the same conclusion, saying that ex post daily foreign exchange volatility can be optimally estimated by aggregating 288 squared five-minute returns. Actually, the latent character of conditional variance is the one that makes it unobservable and impossible to directly measure, being one of the most important obstacles in modeling financial volatility. As such, this variance needs to be rather estimated from past data, which opens the way towards a competition of models that better describe such measure. Among methodologies that estimate such conditional variance, the ARCH[4] Model proposed by Engle (1982), the GARCH Model proposed by Bollerslev (1986), and the stochastic volatility model proposed by Taylor (1986) are the most known and most applied ones.

A common approach to deal with the fundamental latency problem of return volatility is to conduct inferences on volatility through strong parametric assumptions. Another option is to employ models designed for option pricing in order to transform back the prices of derivatives into forecasts of implied latent volatility over a specific horizon. One important drawback of such methods is that they rely heavily on the models chosen to forecast volatility; this means that forecasts may vary significantly according

---

[3] *GARCH - Generalized AutoRegressive Conditional Heteroskedastic.*

[4] *ARCH - AutoRegressive Conditional Heteroskedastic.*

to the choice of the model. Another drawback is that they include in the estimated measure a volatility risk premium that fluctuates in time, the effect being that their forecasts on the underlying asset volatility are often biased. Another flaw of such methodologies results from the backward looking methodology employed. This methodology implies that the current and future volatilities are estimated starting from the return standard derivation of backward looking rolling samples, which returns are most often calculated from daily observations. Because of this, models are less prepared to represent volatility shocks that currently happen and even less to anticipate them. However, backward looking models are not meaningless, as volatility is persistent and thus offers some useful information on ongoing patterns, but volatility also has a mean reverting character due to which the unit root type forecasts are not optimal, since they are conditionally biased given the history of the past returns.

Despite the large variety of the models that seek solutions to relatively similar questions, most of the models designed to estimate and forecast latent volatility fail to describe adequately significant issues as regards the fluctuation of financial returns (Bollerslev (1987), Carnero, Peña and Ruiz (2004) and Malmsten and Teräsvirta (2004)). One such important feature of latent volatility that is not satisfactorily encapsulated in the models is the low, though diminishing autocorrelation in the squared returns related to the high excess kurtosis of returns. Adequate modeling of return dynamics is required as accurate forecasting is essential in risk management or decision taking processes. As such, the assumption of the existence of Gaussian standardized returns has been contested in many studies, being replaced by heavy-tailed distributions.

The increasing opportunity to get access to higher frequency data than the daily data one allowed researchers to experiment more straightforward methods to model volatility by constructing daily time series out of intraday data. This step allows treating “volatility” as observed, rather than latent, to which standard time-series techniques may be applied. Then, since the addition of intraday sampled squared returns provides a consistent estimator of the actual daily volatility, the forecasting performance of the estimated models could be evaluated more accurately than in the case of methodologies that employ squared daily returns as volatility measures. Among all the existing models, the Autoregressive Fractionally Integrated Moving Average (ARFIMA) and the Heterogeneous Autoregressive (HAR) models emerged as the most popular models in the whole literature, capable of capturing the observed long-memory pattern of volatility and empirically outperforming more traditional counterparts as GARCH and stochastic volatility models.

Realized volatility notion solves many of the drawbacks mentioned above as regards the traditional methods that forecast volatility by using squared returns. In the presence of no transaction costs, with continuously observed price, the realized return variance may be modeled with no error by using realized returns. As such, when we control for the measurement error, the ex post volatility eventually becomes observable from latent, which allows to be modeled directly rather than being estimated from a latent process. Moreover, the realized variance is correlated with the concept of cumulative expected variation of the returns over a specific horizon for a large set of underlying no- arbitrage diffusive data generating processes. On the contrary, it is not possible to link on short terms the actual realized return to the expected returns if not making supplementary assumptions. There have been extensive studies on the expected return volatility but little on what concerns the expected mean return from high frequency asset prices. This perspective has produced a significant effort from the researchers as regards obtaining and empirically using realized volatility estimates disseminated from high frequency data. As such, in the today markets, the realized volatility field is a well-established practice to use intraday returns to build up ex post volatility measures. Due to larger access to high-quality transaction data over a well-diversified panel of financial assets, it is unavoidable that this topic be further investigated and more tested in wider empirical contexts over the future.

However, sampling at higher frequency has some disadvantages too. It has been proved that such a choice is a trade-off between obtaining higher accuracy in latent volatility description, theoretically optimized when the frequency sampling is the highest possible one, and the microstructure noise that may arise through bid-ask bounce, asynchronous trading, price discreteness, infrequent trading, etc. Some references discussing this issue are Madhavan (2000), Biais, Glosten and Spatt (2005), and Hansen and Lunde (2005).

Some authors [Andersen and Bollerslev (1998), Patton (2005), Hansen and Lunde (2005)] have used such measure of volatility in order to estimate the out-of-sample forecasting performance of GARCH models. Starting from pioneering studies led by Barndorff-Nielsen and Shephard (2002), Meddahi (2002) and Andersen, Bollerslev, Diebold and Labys (2003), more recent papers have discussed and proposed methodologies to isolate information obtained from realized volatility measures constructed from high frequency data, and further using it to better model daily (or lower frequency) returns. Aït-Sahalia, Mykland and Zhang (2005), Zhang, Mykland and Aït-Sahalia (2005), Bandi and Russell (2005, 2006), Hansen and Lunde (2006) tried also to provide solutions to the inconsistency problem.

In what it follows there will be offered necessary argumentation in the favor of integration of information compounded from high-frequency data into the measuring, modeling, estimating and forecasting daily and lower frequency volatility of financial and economic time series. The largest part of the literature on the topic of volatility forecasting has concentrated so far on obtaining and using highly restrictive and complex parametric versions of GARCH models or of stochastic volatility models, which soon have found their limitation in terms of predictability, especially at higher frequency distributions of returns.

However, in the last decade, the advancement in volatility modeling has stalled in some aspects. The larger access to high-frequency data moved away the attention from further modeling of daily volatility with daily data, and thus the effect of better data inputs was negligible in what concerns improving model designs. It has been empirically demonstrated that the standard models designed to provide estimates by using daily observations were improper in functioning with intraday values. New models specified for the intraday data failed in capturing the information of the interdaily movements as well, for which reason their daily forecasts were not as precise as expected. In the context of not having an empirically superior proved alternative as concerns modeling day or intraday volatility by using high-frequency data, the standard practice continued to use the traditional modeling tools in order to obtain relatively good estimates of daily values, although intraday data were available. Another important factor is that the emphasis has continuingly been placed on low-dimensional volatility modeling, and mainly univariate. Although multivariate variants of the existing ARCH and GARCH models have been proposed by Bollerslev, Engle and Nelson (1994), Ghysels, Harvey and Renault (1996), and Kroner and Ng (1998), the multitude of constraints and computational problems they raise make them computationally difficult to be applied in empirical contexts. Therefore, few applications deal with more assets at the same time. As a consequence, practitioners have avoided to search for solutions to highly practical relevant multidimensional problems, and continued to rely on simple exponential smoothing methods combined with the assumption of normally distributed conditional returns. An example of such models which found large applicability in business contexts is the RiskMetrics proposed by J.P. Morgan; although it employs counterfactual assumptions and proves to be most of the times suboptimal, it functions sufficiently well, its main quality being the model feasibility, simplicity, short implementation time characteristic of high-dimensional contexts.

In the above described context, the realized volatility methodology has been proposed trying to improve modeling process in two regards: the first one is that it proposes a rigorous methodology that fully exploits the information contained in the high-frequency data and which proves efficient in forecasting daily return values; and second, the model offered distinguishes itself through simplicity and facile implementation in high-dimensional environments. These desiderata are met through such method called realized volatility, which computes useful information from high-frequency intraday return data. All works written by Anderson, Bollerslev, Diebold and Labys (2001) and Barndorff-Nielsen and Shephard (2001) argue that the realized volatility measures are not just unbiased ex-post estimators of daily volatility, but also asymptotically free of any measurement error. Realized volatility provides a comprehensive framework for integration of high-frequency intraday return data into the modeling, measuring and analysis of daily or lower frequency volatility and frequency distributions.

Reviewing articles which discussed the topic of realized volatility are extremely heterogeneous in their scope, but as well in the way they comprised and delivered information in the literature. Some papers comprise extensive reviews of the literature, rather limiting to a general discussion of volatility, like Poon and Granger (2003) and Andersen, Bollerslev, Christoffersen and Diebold (2006). A commonality in the general reviews is not discussing the microstructure noise problem. Instead, this problem has been discussed in Bandi and Russell (2006), paper which places an emphasis on the economic determinant of the noise component. McAleer and Medeiros (2008), in exchange, address to the problem of measurement error. Barndorff-Nielsen and Shephard (2007) reviewed the papers by putting a stronger emphasis on nonparametric estimation of volatility and on the frictionless case with/without jump effects.

In what it follows there will be presented some aspects on modeling and forecasting realized volatility in different settings by using high frequency data.

2. Univariate setting of realized volatility. A discrete time model

The realized volatility theory is highly linked to the accessibility of stock price data at arbitrarily high frequencies. As such, it is natural to model volatility as a continuous-time problem, even if we allow ultimately, for simplicity, sampling at discrete intervals. We focus on one risky asset with a price that moves continuously, but which may be observed at equally-spaced discrete points in time over a certain interval, in which a unit interval relates to the primary time period over which we consider to measure the volatility of the asset return, like, for example, one trading day.

The iterations in sections 2 - 4 closely follow the structure of McAleer and Medeiros (2008). We propose at the beginning the case of a simple discrete time model, following to build upon later extensions of it. In such a model the daily returns of any asset are characterized by the following equation:

$$r_t = \sigma_t^{1/2} z_t,$$

where $\{z_t\}_{t=1}^{T}$ represents a series of independently and normally distributed random variables which have a zero mean and variance one, $z_t \sim NID(0,1)$.

We make the following assumptions: the data is sampled at its highest frequency (tick-by-tick) in a trading day *t*, $\Upsilon_t = \{\varphi_0, \ldots \varphi_{n_t}\}$ is a grid which contains all data points observed, and $P_{t,i}, i = 1, \ldots, n_t$ is the *i*-th price observation in day *t*, with n_t being the total observation number in day *t*. We denote

$$r_{t,i} = \sigma_{t,i}^{1/2} z_{t,i},$$

with $z_{t,i} \sim NID(0, n_t^{-1})$, $r_{t,i} = P_{t,i} - P_{t,i-1}$ is the *i*th intraperiod return of day *t* such that

$$r_t = \sum_{i=0}^{n_t} r_{t,i} \text{ and } \sigma_t = \frac{1}{n_t} \sum_{i=0}^{n_t} \sigma_{t,i}.$$

We further define $\Theta_{t,i} = \Theta\{P_{a,b}\}_{a=-\infty, b=0}^{a=t, b=i}$ as the *sigma*-number generated by all information to the *i*th quote in day *t*. Therefore, $\Theta_{t,0}$ is the information set which exists at the beginning of day *t*. It results then that $Exp(r_t^2|\Theta_{t,0}) = \sigma_t$ and $Var(r_t^2|\Theta_{t,0}) = 2\sigma_t^2$.

We may define then the realized variance as the sum of all high frequency intraday observations (returns) of the form

$$RV_t^{(T)} = \sum_{i=0}^{n_t} r_{t,i}^2.$$

We may write then the squared daily return as

$$r_t^2 = (\sum_{i=0}^{n_t} r_{t,i})^2 = \sum_{i=0}^{n_t} r_{t,i}^2 + 2\sum_{i=0}^{n_t-1} \sum_{j=i+1}^{n_t} r_{t,i} r_{t,j},$$

such that

$$Exp\left(r_t^2 \middle| \Theta_{t,0}\right) = Exp\left(\sum_{i=0}^{n_t} r_{t,i}^2 \middle| \Theta_{t,0}\right) + 2Exp\left(\sum_{i=0}^{n_{t-1}} \sum_{j=i+1}^{n_t} r_{t,i} r_{t,j} \middle| \Theta_{t,0}\right)$$

$$= Exp\left(RV_t^{(T)} \middle| \Theta_{t,0}\right) + 2Exp\left(\sum_{i=0}^{n_t-1} \sum_{j=i+1}^{n_t} r_{t,i} r_{t,j} \middle| \Theta_{t,0}\right).$$

When the correlation between the intraday returns is zero, it results that

$$Exp\left(r_t^2 \middle| \Theta_{t,0}\right) = Exp\left(RV_t^{(T)} \middle| \Theta_{t,0}\right) = \sigma_t.$$

It results that in one day *t* there exist two unbiased estimators of the return variance, namely the squared return and the realized variance.

It can be further derived that

$$Var\left(RV_t^{(T)} \middle| \Theta_{t,0}\right) = \frac{2}{n_t} \sum_{i=0}^{n_t} \frac{\sigma_{t,i}^2}{n_t} < \frac{2}{n_t} \left(\sum_{i=0}^{n_t} \frac{\sigma_{t,i}}{\sqrt{n_t}}\right)^2 = Var\left(r_t^2 \middle| \Theta_{t,0}\right)$$

as

$$Exp\left[\left(\sum_{i=1}^{n_t} \sigma_{t,i} z_{t,i}^2\right)^2 \middle| \Theta_{t,0}\right] = \frac{3}{n_t^2} \sum_{i=0}^{n_t} \sigma_{t,i}^2 + \frac{2}{n_t^2} \sum_{i=1}^{n_t-1} \sum_{j=i+1}^{n_t} \sigma_{t,i}\, \sigma_{t,j}$$

The above calculation shows the core of the realized volatility theory: the summation of the squared intraday returns would provide a better estimation of the latent daily return variance than the sum of the squared daily returns. Furthermore, if returns are sampled at any random frequency, we can formulate the average daily variance without any measurement error, as it follows

$$\lim_{n_t \to \infty} Var\left(RV_t^{(T)} \middle| \Theta_{t,0}\right) = 0$$

The only constraint for the dynamics of the intraday return variance formulation as above is to have the following, for any $0 \le l \le 1$:

$$\sum_{i=1}^{n_t} \sigma_{t,i}^2\, \xi n_t^{1+l}$$

Such result has been explored by Andersen and Bollerslev (1997, 1998), Andersen, Bollerslev and Lange (1999), and Martens (2001, 2002).

3. Realized variance modeling in a continuous time setting and no microstructure noise

3.1 Main setting

We set up at the beginning some necessary assumptions to describe the context of the model. We operate in a continuous-time diffusive setting, for simplicity ruling out price jumps, and assume a frictionless market. In the beginning, let’s assume that in a trading day *t*, we have the following time diffusion process of a time series formed from the logarithmic prices of any asset:

$$dP(t+\varphi) = \Xi(t+\varphi)d\varphi + \delta(t+\varphi)dW(t+\varphi), 0 \le \varphi \le 1, t = 1,2, \ldots$$

We have called above $P(t+\varphi)$ as the logarithmic price at time $(t+\varphi)$, $\Xi(t+\varphi)$ stands for the drift component, $\delta(t+\varphi)$ represents the instantaneous volatility (standard deviation) and $W(t+\varphi)$ represents the standard Brownian motion. In order not to create a leverage effect, $\delta(t+\varphi)$ is orthogonal to $W(t+\varphi)$.

Andersen, Bollerslev, Diebold and Labys (2003) and Barndorff-Nielsen and Shephard (2002) indicated that for daily returns defined as $r_t = P(t) - P(t-1)$, they are Gaussian conditionally distributed on $\Theta_t = \Theta\{\vartheta(t+\varphi-1), \delta(t+\varphi-1)\}_{\varphi=0}^{\varphi=1}$, with *sigma*-algebra information set generated by sample paths of $\Xi(t+\varphi-1)$ and $\delta(t+\varphi-1), 0 \le \varphi \le 1$ such that

$$r_t|\Theta_t \sim N\left(\int_0^1 \vartheta(t+\varphi-1)d\varphi, \int_0^1 \delta^2(t+\varphi-1)d\varphi\right)$$

McAleer and Medeiros (2008) define further the term $IV_t = \int_0^1 \delta^2(t+\varphi-1)d\varphi$ as the *integrated variance*, an ex post measure of daily volatility. This is actually what is sought to be obtained, as the measure (as precise as possible) of the true latent volatility.

3.2. Modeling with various sampling schemes

The realized volatility approximates the quadratic variation pretty well as the sampling frequency increases. Nevertheless, this simple statement complicates further the problem according to the following two stances. The first one is that even for the most liquid assets a continuous price is not available. This constraint leads to an unavoidable discretization error in the estimates of the realized volatility which determines us to recognize the existence of a measurement error. Although by subsequent reiteration we may estimate the magnitude of such errors, according to the continuous asymptotic theory, this inference is always subject to sampling distortions and is totally true only when price jumps are disregarded. The second issue refers to the large panel of microstructure effects which induces spurious autocorrelations in the high frequency sampled return series. This category includes the effects of rounding, price discreteness, bid-ask bounces, the trades which occur on various markets, the steady (gradual) response of process to a block trade, asymmetric information contained in order of different size, spreads positioning according to the dealer inventory control, strategic order flows and data recording flaws. The spurious autocorrelations emerging from these sources may increase the estimates of the realized variance and thus generate a traditional type of bias-variance trade off. Although the general recommendation is to use the highest sampling frequency as its optimal for efficiency captured signal, this also tends to bias the estimate of the realized volatility.

The above described trade-off may be plotted through the volatility signature diagram which illustrates the sample mean of the realized volatility estimator over a long time period as a function of the sampling frequency. As such, the long time duration diminishes the impact of sampling variability and therefore, when the microstructure noise is not considered, the plot should appear as an approximately horizontal line. Nevertheless, it is observed in empirical applications that in plots with transaction data sampled from highly liquid stocks we will find spikes at high sampling frequencies and more moderate reductions in order to stabilize at frequencies at 5-40 minute range. On the contrary, the reversal occurs for returns built up from bid-ask quote midpoints as asymmetric adjustments of the spread determine positive serial correlation and bias the signature diagram downward at the highest sampling frequency. As such, for the case of the illiquid stocks, the inactive trading produces positive return serial autocorrelation, which induces the signature diagram increase at lower sampling frequencies. Aït-Sahalia, Mykland and Zhang (2006), Bandi and Russell (2007) and Andersen, Bollerslev, Diebold and Labys (2003) have further developed this topic by trading off efficient sampling with bias-inducing noise in order that optimal sampling schemes be obtained.

Another solution proposed in order to deal with the trading-off described above is to use alternative quadratic variance estimators that would be more efficient and less sensitive to the microstructure noise. Huang and Tauchen (2005) and Andersen, Bollerslev and Diebold (2007) are among them, suggesting that staggered returns and realized bipower variation (the latter for non-parametrically measuring the jump component in asset return volatility) be used, effective in noise reduction, while Andersen, Bollerslev, Frederiksen and Nielsen (2006) extended the signature diagrams in order to count also for power and h-skip bipower variation. An alternative realized variance like high-low measure has been used by Brandt and Jones (2006), Alizadeh, Brandt and Diebold (2002), Brandt and Diebold (2006), Gallant, Hsu and Tauchen (1999), Yang and Zhang (2000), Schwert (1990), Parkinson (1980), and Garman and Klass(1980). Moreover, Christensen and Podolskij (2006) and Dobrev (2007) generalized the high-frequency data estimator in various ways, and discussed its link to the realized variance topic. Moreover, Zhou (1996) sought a method to correct the bias of the realized variance estimators by explicitly accounting for the covariance in the lagged squared return observations. Hansen and Lunde (2006) extended the work began by Zhou for the case of non-independent and identically-distributed noise. Aït-Sahalia, Mykland and Zhang (2006) examined the necessary correction when the noise is independent and identically normally distributed, while Zhang, Mykland and Aït-Sahalia (2005) came with a consistent volatility estimator which considers all the data available, averaging realized variances through forming different sub-samples and correcting for the remaining bias. Aït-Sahalia, Mykland and Zhang (2005) extended further this work and proposed a method to account for some serial correlated errors. Barndorff-Nielsen, Hansen, Lunde and Shephard (2006) proposed kernel estimators as realized measures.

In a traditional setting, prices are observed at discrete and unevenly spaced intervals, which reason determines one to look for different sampling schemes. An interval [0,1] is subdivided into n_t sub-periods and the observation times are defined under the form of a set $Y_t = \{\varphi_0, \dots, \varphi_{n_t}\}$ with $0 < \varphi_0 < \varphi_1 < \cdots < \varphi_{n_t} = 1$ where $\rho_{1,n_t} = \varphi_i - \varphi_{i-1}$ is the length of each subinterval. Naturally, such length should decrease while the number of observations in a day increases. Then the intraday variance over each subperiod may be defined as

$$IV_{i,t} = \int_{\varphi_{i-1}}^{\varphi_i} \delta^2(t+\varphi-1)d\varphi$$

McAleer and Medeiros (2008) distinguish four sampling schemes, as follows:

1. *The calendar time sampling* in which the intervals have equal length in calendar time, meaning that $\rho_{1,n_t} = \frac{1}{n_t}$, $\forall i$. One example is sampling prices at each 5, 10 or 15 minutes. A methodology for this type of sampling was offered by Wasserfallen and Zimmermann (1985), Andersen and Bollerslev (1997), and Dacorogna, Gencay, Muller and Pictet (2001), motivated by the fact that intraday data is irregularly spaced, with no fixed period spacing, so that for most of the data sampled observations must be built upon artificially. Hansen and Lunde (2006) found that the *previous tick* method (method which adds values of the last observations in the missing gaps) is a straightforward and competitive method to sample prices according to calendar time. More exactly, this method samples only the first observation of a five-minute interval.
2. Another sampling method is the *transaction time sampling* in which prices are sampled with every transaction made.
3. A third alternative is the *business time sampling* in which sampling times are selected in such a way that $IV_{i,t} = \frac{IV_t}{n_t}$.
4. Finally, there is the *tick time sampling* in which prices are recorded at each change.

To be mentioned that in the first sampling choice the observations are latent, while in the last three ones the sampled data are observed, each sampling choice producing effects in the estimated integration variance.

3.3. Distribution of the realized variance

Andersen, Bollerslev, Diebold and Labys (2003) found that the realized variance obtained from intraday data is a consistent estimator of the integrated variance when microstructure noise does not exist:

$$\lim_{p} RV_t^{(T)} = IV_t$$

The asymptotic distribution of the realized variance can be further derived from Jacod and Protter (1998) and Barndorff-Nielsen and Shephard (2002) as follows:

$$n_t^{\frac{1}{2}}(2IQ_t)^{-\frac{1}{2}}\left(RV_t^{(T)} - IV_t\right) \rightarrow N(0.1) \qquad (*)$$

In the above formulation, the IQ_t defines the *integrated quarticity*, which takes the form

$$IQ_t = \int_0^1 \delta^4 (t + \varphi - 1) d\varphi$$

The integrated quarticity is unobserved and is likely to display large period-to-period variation. Therefore, a consistent estimator for the integrated quarticity should be used in relationship to the true realization in order to enable feasible inference. This integrated quarticity may be consistently estimated by the *realized quarticity* (according to Barndorff-Nielsen and Shephard, 2002), defined as

$$RQ^{(T)} = \frac{n_t}{3} \sum_{i=0}^{n_t} r_{t,i}^4$$

and

$$n_t^{\frac{1}{2}} (\frac{2}{3} RQ_t^{(T)})^{-\frac{1}{2}} (RV_t^{(T)} - IV_t) \to N(0,1)$$

In the papers of Meddahi (2002), Gonçalves and Meddahi (2005), Barndorff-Nielsen and Shephard (2005) and Nielsen and Frederiksen (2006) it has been studied the finite sample behavior of the limit theory given in (*). According to it, (*) is poorly seized but

$$n_t^{\frac{1}{2}} (\frac{2}{3} \frac{RQ_t^{(T)}}{(RV_t^{(T)})^2})^{-\frac{1}{2}} [\log(RV_t^{(T)}) - \log(IV_t)] \to N(0,1)$$

performs well.

4. Modeling with microstructure noise

In what it follows we will discuss the effect of microstructure noise in estimating integrated variance. The market microstructure noise may arise from a variety of causes, like discreteness of the price [Harris (1990, 1991)] or properties of the trading mechanisms (Black (1976) and Amihud and Mendelson (1987)). Relevant literature may be found as well in O'Hara (1995), Madhavan (2000), Hasbrouck (2004) and Biais, Glosten and Spatt (2005). Zhang, Mykland and Aït-Sahalia (2005) assert that sampling over longer horizons merely diminishes the effect of microstructure rather than quantifying and correcting its effect for volatility estimation over shorter horizons.

In what follows, the denotation by Zhang, Mykland and Aït-Sahalia (2005) will be kept, so we define $P_{t,i} = P_{t,i}^* + \epsilon_{t,i}$ to be the logarithmic return price process, as observed at the sampling times. $P_{t,i}^*$ represents the latent true, or efficient, return process that follows $P_{t,i} = P_{t,i}^* + \epsilon_{t,i}$. $\epsilon_{t,i}$ is the independent noise (microstructure noise) around the true return. It follows then that

$$r_{t,i} = r_{t,i}^* + \epsilon_{t,i} - \epsilon_{t,i-1} = r_{t,i}^* + \zeta_{t,i}$$

where: $r_{t,i}^* = P_{t,i}^* - P_{t,i-1}^*$ represents the efficient return. We can see that $r_{t,i}$ is an autocorrelated process, and $RV_t^{(T)}$ will be a biased estimator of the true latent daily volatility. Since

$$RV_t^{(T)} = \sum_{i=1}^{n_t} (r_{t,i}^*)^2 + 2\sum_{i=1}^{n_t} r_{t,i}^* \zeta_{t,i} + \sum_{i=1}^{n_t} \zeta_{t,i}^2$$

it is straightforward to assert that, conditional on the efficient returns,

$$Exp\left(RV_t^{(T)} \mid r^*\right) = RV_t^{*(T)} + 2n_t Exp\left(\epsilon_{t,i}^2\right)$$

such that $RV_t^{(T)}$ becomes as well a biased estimator of the integrated variance.

When the microstructure noise is considered as a covariance stochastic process with zero mean and variance $\zeta_{t,i} = \epsilon_{t,i} - \epsilon_{t,i-1}$, Bandi and Russell (2005) found that $RV_t^{(T)} \to \infty$ as $n_t \to \infty$.

Furthermore, if the microstructure noise is a price independent process and an IID random variable with zero mean and variance $\zeta_{t,i} = \epsilon_{t,i} - \epsilon_{t,i-1}$, then according to Zhang, Mykland and Aït-Sahalia (2005), we have

$$n_t^{-1/2}\left[RV_t^{(T)} - IV_t - 2n_t Exp\left(\epsilon_{t,i}^2\right)\right] \to 2\left[Exp\left(\epsilon_{t,i}^4\right)\right]^{\frac{1}{2}} N(0.1)$$

When sampling at high frequency, the intraday observation number becomes finite, while the observations become discrete. This leaves room for another bias due to discretization:

$$RV_t^{(T)} \approx IV_t + 2n_t Exp\left(\epsilon_{t,i}^2\right) + \left[4n_t Exp\left(\epsilon_{t,i}^4\right) + \frac{2}{n_t}\int_0^1 \delta_t^4 dt\right]^{\frac{1}{2}} N(0,1)$$

where "≈" sign represents the fact that, when multiplied by the proper number, convergence occurs in distribution.

Furthermore, when the noise is an independent price process, and the microstructure noise is a stationary, strong, mixing stochastic process, with mixing coefficients decreasing at an exponential rate, and with zero mean and variance $\zeta_{t,i} = \epsilon_{t,i} - \epsilon_{t,i-1}$, if we additionally assume that $Exp\left[\left(\epsilon_{t,i}^{4+k}\right)\right] < \infty$ for some $k > 0$, then Zhang (2006) and Aït-Sahalia, Mykland and Zhang (2006) found that

$$RV_t^{(T)} \approx IV_t + 2n_t Exp\left(\epsilon_{t,i}^2\right) + \left[4n_t\Delta + \frac{2}{n_t^{(sparse)}}\int_0^1 \delta_t^4 dt\right]^{\frac{1}{2}} N(0.1)$$

where:

$$\Delta = Var\left[(\epsilon_{t,1} - \epsilon_{t,0})^2\right] + 2\sum_{i=1}^{\infty}\left[cov\left[(\epsilon_{t,1} - \epsilon_{t,0})^2\right], \left[\epsilon_{t,i+1} - \epsilon_{t,i}{}^2\right]\right]$$

This means that, when n_t becomes large, the realized variance $RV_t^{(T)} = \sum_{i=0}^{n_t} r_{t,i}^2$ will diverge linearly to infinity. Bandi and Russell (2005) and Zhang, Mykland and Aït-Sahalia (2005) showed that the realized variance estimates consistently the microstructure noise variance such that

$$\frac{1}{2n_t} RV_t^{(T)} \to Exp\left(\epsilon_{t,i}^2\right)$$

Andersen, Bollerslev, Diebold and Labys (2000, 2001, 2003) offered a possible solution called "sparse sampling" to this microstructure bias, which consists of return sampling at randomly chosen lower frequencies, like 5 or 15 minute sampling, than tick by tick sampling. However, in the opinion of Zhang, Mykland and Aït-Sahalia (2005), sparse sampling is not very efficient. Their argument included the following: if a $Y_t^{(sparse)}$ grid is defined, with $n_t^{(sparse)}$ equal-length samples, then $Y_t^{(sparse)}$ is a subgrid of Y_t . Defining

$$RV_t^{(sparse)} = \sum_{i=1}^{n_t^{(sparse)}} r_{t,i}^2$$

then by using the results of Barndorff-Nielsen and Shephard (2002), Mykland and Zhang (2006), Zhang, Mykland and Aït-Sahalia (2005), Zhang (2006) and Aït-Sahalia (2006), the sparsed realized variance takes the form

$$RV_t^{(sparse)} \approx IV_t + 2n_t^{(sparse)} Exp(\epsilon_{t,i}^2) + \left[4n_t^{(sparse)} E(\epsilon_{t,i}^4) + \frac{1}{n_t^{(sparse)}} \int_0^1 \delta_t^4 dt\right]^{\frac{1}{2}} N(0.1)$$

5. The conditional return variation and the concept of realized volatility

The following section is dedicated to the natural liaison between quadratic variation and the integrated variance, in order to cover some practical aspects as regards the estimations of the realized volatility and of the variance of conditional return. Assuming an invariable drift and volatility coefficients, both conditional and unconditional variance in returns will equal the quadratic variation in the log price. On the contrary, if assuming volatility as a stochastic process, then precise distinction between conditional variance (which stands for the expected size of the innovations of the squared returns over a specific interval) and the quadratic variation over a specific time horizon is needed. Therefore, the difference may be expressed as an expectation against future realizations of the volatility of stock returns. Theoretically, the realized volatility would express only the actual realizations, and not their previous expectations. However, the realized volatility estimates are efficient in capturing the conditional return variation as one may build up accurate forecasts/conditional expectations of return volatility out of a financial or economic time series formed from past realized volatility.

The above assertions may be even strengthened under a simplified setting. If the instant return is a continuous-time process and the return, average and volatility series are low or not correlated processes, then the conditional expectation of the return should be normally distributed, conditional on the cumulative drift and on the quadratic variation. Therefore, the distribution of the return series is mixed Gaussian with the mixture ruled out by the integrated variance realizations, along with their integrated mean. Realization jumps from the integrated variation process make the outliers of the returns become probable while the persistence in the integrated variance process may determine volatility clustering. Furthermore, over short horizons, when the conditional mean is very low as compared to the cumulative absolute return innovations, the integrated variance process may be intrinsically linked to the conditional variance.

Because the realized variance is roughly unbiased for the related unobserved quadratic variation, the realized volatility estimate comes as the natural point of reference against which we estimate the volatility forecasts accuracy. There may be also tests of goodness-of-fit undertaken on the residuals resulted from subtracting the forecast from the realized volatility measures.

The realized volatility topic is also related to the return variation estimated over a discrete time period rather than with the spot (instant) volatility. The distinction appears due to the differentiation between realized volatility concept and a whole range of literature written in the search of spot volatility estimation from discrete observations, mainly in a setting with a constant diffusion coefficient. Although theoretically the measurement of realized volatility can be adapted easily to spot volatility estimation, in practice this is not feasible as frequent sampling over very small intervals may amplify the effects of microstructure noise.

6. Modeling and forecasting realized volatility

A well-known fact in the literature is that when GARCH and SV[5] models are employed, the standardized returns do not exhibit a Gaussian distribution. Instead, the standardized returns present an excess kurtosis, thing that reasons for the employment of heavy-tailed distributions. Andersen, Bollerslev, Diebold and Labys (2000, 2001, 2003) proved that when modeling has been made with the employment of realized variance measures, the distribution of standardized exchange rates approaches the properties of a typical Gaussian. A similar application with stock returns run by Andersen, Bollerslev, Diebold and Labys (2001) arrived to similar conclusions.

The log-realized variance is significantly persistent, but stationary, with long memory properties, traditionally expressed as an ARFIMA(p,d,q) process. Various models have been proposed to catch the properties of such time series. One of them is the Multiplicative Error Model (MEM) proposed by Engle and Gallo (2006) that is consistent and asymptotically normal under a wide range of specifications for the error density function. The MEM model is best suited to model the conditional behavior of positively valued variables choosing a convenient GARCH-type structure when modeling variance and persistence. Another model is the HEAVY model [Shephard and Sheppard (2009)], a high frequency-based volatility model of daily asset return volatility based on measures constructed from high frequency data. The authors proved that such models perform more robust to level breaks in the volatility than conventional GARCH models, adjusting to the new level much faster. Supplementarily, although such model shows mean reversion, it exhibits as well momentum, a feature that misses from classical models.

[5] Stochastic volatility.

Another model that uses realized measures is the Heterogeneous AutoRegressive Realized Volatility (HAR-RV) model proposed by Corsi, Zumbach, Muller and Dacorogna (2001) and Corsi (2003), model that has at its ground the Heterogeneous ARCH (HARCH) model proposed by Müller, Dacorogna, Davé, Olsen, Puctet and von Weizsäcker (1997). The HAR-RV model represents an additive cascade of different volatility components produced by actions of the participants in the market that produces remarkably good out-of-sample forecasting performance. The HAR-RV model is built up in such a way that the additive volatility cascade leads to an AR-type model in the realized volatility, considering volatilities realized over different sampling sizes.

Subsequently, McAleer and Medeiros (2006) offered a multiple regime smooth transition generalization of the HAR-RV model (called Multiple Regime Smooth Transition Heterogeneous Autoregressive HARST), by coming with a flexible model able to capture the non-linearities and long-range dependence in time series dynamics. The model has been designed to describe concurrently long memory and size and sign asymmetries.

In volatility forecasting topic, sources of long memory have been intensively searched, because shorter memory of a model, better forecasting performances may be produced. For example, Hyung, Poon and Granger (2005) revealed that numerous nonlinear short memory models, especially those which present infrequent breaks, may generate long memory patterns. Some of these models are the regime switching model of Hamilton and Susmel (1994), the volatility component model of Engle and Lee (1999), the model proposed by Diebold and Inoue (2001), the break model developed by Granger and Hyung (2004), and the multiple regime-switching model of Medeiros and Veiga (2004). The latter one is developed to describe size and sign asymmetries in financial volatility as well as intermittent dynamics and excess kurtosis. Hilledebrand (2005) and Hilledebrand and Medeiros (2006) revealed the statistical consequences of neglecting structural breaks and regime switches in autoregressive and GARCH models, proposing two solutions to remedy the problem: the identification of those regimes with constant unconditional volatility that use a change point detector and then estimate a separate GARCH model on each of the separate resulting segments, and the estimation of a multiple-regime GARCH model, like that of FCGARCH type (Flexible Coefficient GARCH).

Scharth and Medeiros (2006) came with a new model built up on regression trees that described the realized volatility dynamics for some DJIA stocks. They presented empirical evidence that additive price changes convey meaningful information as regards multiple regimes in the realized volatility of stocks, whereas large rises (falls) in prices are highly dependent on persistent regimes of low (high) variance in stocks. Therefore, past cumulated daily returns incorporated as a source of the regimes' switches accounts for high empirical values of long memory parameter estimates. The nonlinear model has been found to be superior to the other long memory models, ARFIMA and HAR-RV.

In all previously mentioned references, volatility has been assumed to refer only to short memory between breaks in each component of volatility and within each regime. A significant improvement of this approach came from Martens, van Dijk and Pooter

(2004) who considered a model that combined the long memory properties with nonlinearity, particularly relevant in modeling asymmetries and leverage effects. The model they proposed is a nonlinear model for realized volatility which accommodated level shifts, day-of-the-week effects, leverage effects and volatility level effects. Deo, Hurvich and Lu (2006) proposed a long-memory stochastic volatility model (LMSV) which is found as a very good competitor to the method that predicts realized variance by using a long memory stochastic volatility model applied to high frequency return data while accounting for significant gradually varying intra-day seasonality in volatility. Koopman, Jungbacker and Hol (2005) established a model which joined unobserved elements and long-memory, while Hillebrand and Medeiros (2008) documented a model that joined long memory with different features of nonlinearity.

Despite such a rich literature on the emerging field of realized volatility, open questions regard the sources of long memory characteristic in the realized volatility and the extension of benefits in terms of volatility predictability from combining long-memory with nonlinear models (Ohanissian, Russell and Tsay (2004)).

Treating the same topic of long-memory, Lieberman and Philips (2008) offered some analytical explanations on the reason according to which realized volatility series typically display long range dependence with a memory parameter (d) of around 0.4. They found that long-memory properties are an effect of the accumulation of realized variance and offered some solutions to refine the statistical inference as regards the parameter d in ARFIMA(p,d,q) models.

Aït-Sahalia and Mancini (2006) compared the out-of-sample relative capacity of forecasting of realized variance in different contexts. Ghysels and Sinko (2006) assessed the extent to which the correction for microstructure noise improved forecasting future volatility using Mixed Data Sampling (MIDAS) and found that the conditional optimal sampling works reasonably well in practice. They also found that within the class of quadratic variation measures, the subsampling and averaging approach (Zhang, Mykland and Aït-Sahalia (2005)) represents the class of estimators that best predicts volatility at five minute sampling schemes. Furthermore, Corradi, Distaso and Swanson (2006) estimated and forecasted conditional predictive density and confidence intervals for integrated volatility by newly proposed nonparametric kernel estimators, built upon various realized volatility measures constructed using *ex post* variation of asset prices. Corsi, Kretschmer, Mittnik and Pigorsch (2008) showed that the residuals of the commonly used time-series models for realized volatility exhibited non-Gaussianity and volatility clustering, proposing extensions to explicitly account for these properties and assess their relevance when modeling realized volatility. Moreover, they demonstrated that allowing for time-varying volatility of realized volatility leads to significant improvement of model fit and of the predictive performance as well, while the distributional assumption for residuals proved to be crucial in density forecasting.

Another important topic in the context of realized volatility is that, regardless of the microstructure noise presence, the realized volatility is an estimated quantity rather than a true, daily value of volatility or of the integrated variance, while integrated quarticity may be replaced by realized quarticity. This fact opens the perspective of employing generated regressors and generated variables in forecasting exercises, associated with critical questions on the efficient estimation and invalid inferences that may occur when biased (asymptotic) standard errors are used [Pagan (1984, 1986), McKenzie and McAleer, (1997)].

Andersen, Bollerslev and Meddahi (2004, 2005) built up a general model-free adjustment method aimed at estimating the unbiased volatility loss functions starting from practically feasible realized volatility benchmarks. According to them, an efficient measurement error accounting in the evaluations of volatility forecasts may lead to markedly higher estimates for the true degree of return-volatility predictability. Corradi and Distaso (2006) proposed a procedure to test for the correct specification of the functional form of the volatility process within the class of eigen function stochastic volatility models. The procedure starts from the comparison of the moments of realized volatility measures with the corresponding ones of integrated volatility implied by the model under the null hypothesis. They first provided primitive conditions as regards the measurement error associated with the realized measure, which would allow to construct asymptotically valid specification tests. Then, they established those regularity conditions under which the realized measures (realized volatility, bipower variation, and modified subsampled realized volatility) satisfy the given primitive assumptions.

7. Multivariate empirical studies

One of the most cited papers which discussed the topic of realized variance in applications with multivariate models is de Pooter, Martens and van Dijk (2008). This paper investigates the merits of high-frequency intraday data when forming mean-variance efficient stock portfolios with daily rebalancing from the individual stock components of the S&P100 index. They focused on the problem of establishing the optimal sampling frequency as revealed by the performance of these portfolios. Surprisingly, the authors found that the optimal frequency is not the highest frequency one, but it ranges between 30 and 65 minutes, significantly lower than the popular five-minute one, which is typically motivated by the aim of maintaining a balance between the variance and bias in covariance matrix estimates due to market microstructure effects like non-synchronous trading and bid-ask bounce. Another important finding is that bias-correction procedures, based on combining covariance matrix estimates with low-frequency and high-frequency, and on the summing of leads and lags, do not significantly influence the optimal sampling frequency or the portfolio performance. This is also robust to the presence of transaction costs and to the portfolio rebalancing frequency.

Another paper that discusses in multivariate context the functioning of realized variance modeling is Bauer and Vorkink (2006) who propose a new matrix logarithm model of the realized covariance of stock returns, by employing latent factors as functions of both lagged volatility and returns. The model proves advantageous as it is parsimonious, does not require imposing parametric restrictions and yields a positive definite covariance matrix. The model is empirically tested with a covariance matrix of size-sorted stock returns and two factors are isolated as satisfactory to capture most of the dynamics. A new method to track down an index using the model of the realized volatility covariance matrix proposed is also introduced.

8. Empirical exercise

In what follows we consider an empirical exercise to prove the advantages of considering modeling with intraday volatility measures instead of using only daily data. Therefore, we shall consider the AIG stock forming a 3436 long time series sampled daily over January 4, 1995 – September 30, 2008 period. We shall use as measures of intraday volatility a 3436 long time series with daily realized kernels that describe the intraday variance of each day. We shall employ an in sample procedure, trying to maximize the loglikelihod functions over the whole sample of the following two models:

GARCH(1,1) model (simple model): $z_t = \frac{r_t - \mu}{\sqrt{\sigma_t}}$, $\sigma_t = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}$ and

EGARCH-X(1,1) model (realized model): $z_t = \frac{r_t - \mu}{\sqrt{\sigma_t}}$,

$$log\sigma_t = \omega + \beta log\sigma_{t-1} + \gamma log x_{t-1} + \tau_1 z_{t-1} + \tau_2 (z_{t-1}^2 - 1)$$

where: *r* represents the returns, σ the variance, *x* the realized measures (kernels) and $z_t \sim iid(0, \sigma_u^2)$ the studentized returns.

The loglikelihood functions to be maximized are:

For the GARCH(1,1) model: $l(r) = -\frac{1}{2}\sum_{t=1}^{n}(log2\pi) + \log(h_t) + \frac{r_t^2}{h_t})$

For the EGARCH-X(1,1) model: $l(r,x) = -\frac{1}{2}\sum_{t=1}^{n}(log2\pi) + \log(h_t) + \frac{r_t^2}{h_t}) + -\frac{1}{2}\sum_{t=1}^{n}(log2\pi) + \log(\sigma_u^2) + \frac{u_t^2}{\sigma_u^2})$

The parameters obtained and the maximum loglikelihood functions are:

Model		Parameters						Loglikelihood function
	μ	ω	α	β	γ	τ_1	τ_2	
GARCH(1,1)	-0,0018	0,0209	0,1103	0,8937	-	-	-	-6215,1406
EGARCH-X(1,1)	-0,0377	0,0291	-	0,7145	0,3006	-0,0468	0,0151	-6153,3633

As we can see, the loglikelihood function improves when the intraday volatility measures are used. Similarly, if we measure the errors of each model by using the RMSE (root mean squared error) and the MAPE (mean absolute percentage error) error measures, we will obtain:

Model	Errors		Loglikelihood function
	RMSE	MAPE	
GARCH(1,1)	107,3708	79,6350	-6215,1406
EGARCH-X(1,1)	109,8659	81,6018	-6153,3633

Both results (loglikelihood functions and error measurements) indicate that the realized model (EGARCH-X) improves volatility estimation as compared to a simple GARCH(1,1) model.

9. Conclusions

This paper provides a review of the most important developments that occurred in the literature on the topic of volatility forecasting using high frequency data, searching to investigate the existing gaps in what concerns modeling and forecasting realized volatility. It realized an extensive review of the literature on relevant aspects in modeling volatility topic, highlighting the problems posed by the properties of financial time series (like the long memory pattern of volatility or the latent character of the conditional variance) and the way the literature attempted to overcome them, as well the drawbacks of the methods proposed for this purpose. In this context, it has been underlined the step forward made in modeling activity when high frequency data was employed, this type of data allowing the experimentation of more straightforward methods to estimate volatility. It was mentioned that the realized volatility models allowed solving many of the drawbacks of the traditional non-realized models. Disadvantages of sampling at higher frequency were discussed as well, among which the most relevant is the microstructure noise that may arise from bid-ask bounce, asynchronous trading, price discreteness or infrequent trading, when sampling is done at the highest frequency as possible. It was concluded that the literature on modeling volatility has concentrated on obtaining highly restrictive and complex parametric versions of GARCH models or of stochastic volatility models, which soon found their limitations in terms of predictability, especially at higher frequency distributions of returns.

In this context, the modeling activity has aroused concern about improving the modeling process in two regards: to offer a rigorous methodology that would fully exploit the information contained in the available high frequency data, and to offer a simpler, more straightforward and easier implementation of the methods in the highly-dimensional environments.

For the purpose of offering a more comprehensive review of the models which best respond to the two above-mentioned desiderata of the literature, realized models were introduced, first in a univariate setting; so a simple discrete time model was described, revealing a central finding in the literature, that the summation of the squared intraday returns provided a better estimation of latent daily return variance than the sum of the squared daily returns. Then, a continuous time specification was employed, showing the advantages and disadvantages of various sampling schemes that might be used. Thus, it was asserted that although the realized volatility approximated better the quadratic variation as the sampling frequency increased, the continuous price was not always available even for the most liquid assets. Moreover, a large panel of microstructure noise effects induces spurious autocorrelations in the high frequency sampled return series, which may increase the estimates of the realized variance and generate a type of bias-variance trade-off. Various solutions were described in order to overcome such difficulties, all of them involving different sampling schemes.

Thus, the microstructure noise arising from discreteness of price or from properties of the trading mechanisms topic was further discussed; since the microstructure noise was found as a serious problem in modeling in continuous time setting, a separate section has been added. For this purpose, both independent and dependent noise processes were considered. Solutions to the consistency problems were also mentioned.

The paper ends with four sections dedicated solely to realized volatility topic. They present the concept of conditional return variation and its natural liaison to the integrated variance, showing how the realized modeling succeeded to enclose it in its formulation. After presenting the advantages of this type of modeling, an extensive review of the literature that covered this topic was undertaken, presenting the relevant points of view each proposed model tried to address to. Critical issues on modeling and forecasting realized volatility were considered also in multivariate empirical contexts, and some papers that approached volatility modeling with realized methods in multivariate settings being reviewed. Finally, the paper closed with a short empirical exercise that underlined the gain in accuracy when realized measures were employed.

The subject of volatility forecasting proves to be of high relevance to various theoretical or applied contexts, like modern pricing, investment or risk management fields, market timing decisions, portfolio selection, in point forecasting, interval forecasting, probability forecasting including sign forecasting and density forecasting. Given the large interest in such a large panel of possible applications, it is obvious that we need to correctly formulate the variance forecasting models, so that they would prove useful eiher applied to investment or risk management field, security valuation and pricing, or to monetary policy making. That is why, in this context, the step towards using higher frequency data enables us to provide better modeling to the benefit of all these related activities.

References

Aït-Sahalia, Y. and Mancini, L. (2006), “Out of sample forecasts of quadratic variation”, *Working Paper*, Princeton University and University of Zürich.

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

Aït-Sahalia, Y. Mykland, P.A. and Zhang, L. (2006), “Ultra-high frequency volatility estimation with dependent microstructure noise”, *Working Paper*, Princeton University.

Alizadeh, S. Brandt, M.W. and Diebold, F.X. (2002), “Range-based estimation of stochastic volatility models”, *Journal of Finance,* 57: 1047-1091.

Amihud, Y. and Mendelson, H. (1987) “Trading mechanisms and stock returns: An empirical investigation”, *Journal of Finance,* 42: 533–553.

Andersen, T.G. and Benzoni, L. (2008), “Realized volatility, Federal Reserve Bank of Chicago”, *Working Paper,* 2008-14.

Andersen, T.G. and Bollerslev, T. (1997), “Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns”, *Journal of Finance,* 52: 975-1005.

Andersen T,G. and Bollerslev, T. (1998), “Answering the skeptics: yes, standard volatility models do provide accurate forecasts”, *International Economic Review,* 39: 885-905.

Andersen, T.G. Bollerslev, T. Christoffersen, P. and Diebold, F.X., (2006), “Volatility and correlation forecasting”, In: Elliott G, Granger CW J, Timmermann A (eds.), *Handbook of Economic Forecasting,* Amsterdam: North-Holland, pp. 778–878.

Andersen, T.G. Bollerslev, T. and Diebold, F.X. (2007), “Roughing it up: Including jump components in measuring, modeling and forecasting asset return volatility”, *Review of Economics and Statistics*, 89(4): 701-720.

Andersen, T.G. Bollerslev, T. Diebold, F.X. and Labys, P. (2000), “Exchange rate returns standardized by realized volatility are (nearly) Gaussian”, *Multinational Finance Journal, “*Roughing it up: including jump components in measuring, modeling and forecasting asset return volatility”, *Review of Economics and Statistics,* 4: 159-179.

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

Andersen, T.G., Bollerslev, T. Diebold, F.X. and Labys, P. (2003) “Modeling and forecasting realized volatility”,*Econometrica,* 71, pp. 529-625.

Andersen, T.G., Bollerslev, T. Frederiksen, P.H. and Nielsen, M.Ø. (2006), “Comment on realized variance and market microstructure noise”, *Journal of Business & Economic Statistics,* 24: 173-179.

Andersen, T.G., Bollerslev, T. and Lange, S. (1999), “Forecasting financial market volatility: sample frequency vis-a-vis forecast horizon”, *Journal of Empirical Finance,* 6: 457-477.

Andersen, T.G., Bollerslev, T. and Meddahi, N. (2004), “Analytic evaluation of volatility forecasts”, *International Economic Review,* 45: 1079-1110.

Andersen, T.G., Bollerslev, T. and Meddahi, N. (2005), “Correcting the errors: Volatility forecast evaluation using high-frequency data and realized volatilities”, *Econometrica,* 73: 279-296.

Bandi, F. and Russell, J.R. (2005), “Realized covariation, realized beta, and microstructure noise”, *Working Paper*, University of Chicago.

Bandi, F. and Russell, J.R. (2006), “Separating market microstructure noise from volatility”, *Journal of Financial Economics,* 79: 655–692.

Bandi, F. and Russell, J.R. (2007), “Microstructure noise, realized volatility, and optimal sampling”, *Review of Economic Studies*, 75(2): 339-369

Barndorff-Nielsen, O.E. and Shephard, N. (2001), “Non-Gaussian OrnsteinUhlenbeck based models and some of their uses in Financial Economics”, *Journal of the Royal Statistical Society*, Series B, 63: 167-241.

Barndorff-Nielsen, O.E. and Shephard, N. (2002), “Econometric analysis of realized volatility and its use in estimating stochastic volatility models”, *Journal of the Royal Statistical Society*, Series B, 64: 253-80.

Barndorff-Nielsen, O.E. and Shephard, N. (2005), "How accurate is the asymptotic approximation to the distribution of realised volatility?", In: Andrews D W K and Stock J H, (eds), *Identification and Inference for Econometric Models. A Festschrift for Tom Rothenberg,* Cambridge: Cambridge University Press, pp. 306–331.

Barndorff-Nielsen, O. and EandShephard, N. (2007), "Variation, jumps and high frequency data in financial econometrics", In: Blundell R., Torsten P. and Newey W. K. (eds.), *Advances in Economics and Econometrics: Theory and Applications*, Ninth World Congress, Econometric Society Monograph, Cambridge University Press.

Bauer, G.H. and Vorkink, K. (2006), "Multivariate realized stock market volatility", *Paper presented to CIREQ Conference on Realized Volatility*, Montreal, Canada.

Biais, B. Glosten, L. and Spatt, C. (2005), "Market microstructure: A survey of microfoundations, empirical results, and policy implications", *Journal of Financial Markets,* 8: 217–264.

Black, F. (1976), "Noise", *Journal of Finance,* 41: 529–543.

Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics,* 21: 307–328.

Bollerslev, T. (1987), "A conditionally heteroskedastic time series model for speculative prices and rates of return", *Review of Economics and Statistics,* 69: 542–547.

Bollerslev, T. Engle, R.F. and Nelson, D.B. (1994), "ARCH Models", In: Engle R. F. and McFadden D. (eds.), *Handbook of Econometrics*, Vol.4: 2959-3038, Amsterdam: North-Holland.

Brandt, M. and Wand Diebold, F.X. (2006), "A no-arbitrage approach to range-based estimation of return covariances and correlations", *Journal of Business,* 79: 61-73.

Brandt, M.W. and Jones, C.S. (2006), "Volatility forecasting with range-based EGARCH models", *Journal of Business & Economic Statistics,* 24: 470-486.

Carnero, M.A. Peña, D. and Ruiz, E. (2004), "Persistence and kurtosis in GARCH and stochastic volatility models", *Journal of Financial Econometrics,* 2: 319–342.

Christensen, K. and Podolskij, M. (2007), "Realised range-based estimation of integrated variance", *Journal of Econometrics*, 141(2): 323-349.

Corradi, V. Distaso, W. (2006), "Semiparametric comparison of stochastic volatility models via realized measures", *Review of Economic Studies,* 73: 635–677.

Corradi, V. Distaso, W. and Swanson, N.R. (2006), "Predictive inference for integrated volatility", *Working Paper*, Rutgers University.

Corsi, F. (2003), "A simple long memory model of realized volatility", *Working Paper*, University of Southern Switzerland.

Corsi, F. Kretschmer, U. Mittnik, S. and Pigorsch, C. (2008), “The volatility of realized volatility”, *Econometric Reviews*, 27: 46–78

Corsi, F. Zumbach, G. Müller, U. and Dacorogna, M (2001), “Consistent high-precision volatility from high frequency data”, *Economic Notes,* 30: 183–204.

Dacorogna, M.M. Gencay, R. Müller, U. Olsen, R.B. and Pictet, O.V. (2001), “*An introduction to high-frequency finance”,* Academic Press, San Diego, CA.

Deo, R. Hurvich, C. and Lu, Y. (2006), “Forecasting realized volatility using a long memory stochastic volatility model: Estimation, prediction and seasonal adjustment”, *Journal of Econometrics,* 131: 29-58.

Diebold, F. X. and Inoue, A. (2001), “Long memory and regime switching”, *Journal of Econometrics,* 105: 131–159.

Dobrev, D. (2007), “Capturing volatility from large price moves: Generalized range theory and applications”, *Working Paper*, Federal Reserve Board.

Engle, R.F. (1982), “Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation”, *Econometrica,* 50: 987–1007.

Engle, R.F. and Gallo, G.M. (2006), “A multiple indicators model for volatility using intra-daily data”, *Journal of Econometrics,* 131: 3-27.

Engle, R.F. and Lee, G.J (1999), “A long-run and short-run component model of stock return volatility”. In: Engle, R and White H (eds.), *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, Chapter 10, Oxford: University Press, pp. 475–497.

Gallant, A.R. Hsu, C. and Tauchen, G.E. (1999, “Using daily range data to calibrate volatility diffusions and extract the forward integrated variance”, *Review of Economics and Statistics,* 81: 617-631.

Garman, M. BandKlass, M.J. (1980), “On the estimation of price volatility from historical data”, *Journal of Business,* 53: 67-78.

Ghysels, E. Harvey, A. and Renault, E. (1996), “Stochastic volatility”, In: Maddala G S and Rao C S (eds.), *Handbook of Statistics: Statistical Methods in Finance*, Vol.14, Amsterdam: Elsevier Science, pp. 119-191.

Ghysels, E. and Sinko A. (2006), “Volatility forecasting and microstructure noise”, *Working Paper*, UNC.

Gonçalves, S. and Meddahi N. (2005), “Boostraping realized volatility”, *Working Paper*, Université de Montréal.

Granger, C.W. and JandHyung, N. (2004), “Occasional structural breaks and long memory”, *Working Paper,* 99-14, San Diego: University of California.

Hamilton, J.D. and Susmel R. (1994), “Autoregressive conditional heteroskedasticity and changes in regime”, *Journal of Econometrics,* 64: 307–333.

Hansen, P.R. and Lunde A. (2005), “A realized variance for the whole day based on intermittent high-frequency data”, *Journal of Financial Econometrics,* 3: 525-554

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

Harris, L. (1990), "Estimation of stock variance and serial covariance from discrete observations", *Journal of Financial and Quantitative Analysis*, 25: 291–306

Harris, L. (1991), "Stock price clustering and discreteness", *Review of Financial Studies,* 4: 389–415

Hasbrouck, J. (2004), "Empirical market microstructure: Economic and statistical perspectives on the dynamics of trade in securities markets", Lecture Notes, Stern School of Business, New York University.

Hillebrand, E. (2005), "Neglecting parameter changes in GARCH models", *Journal of Econometrics,* 129: 121–138.

Hillebrand, E. and Medeiros, M.C. (2006), "Estimating and forecasting GARCH models in the presence of structural breaks and regime switches", In: Rapach, D. and Wohar, M, (eds.) *Forecasting in the Presence of Structural Breaks and Uncertainty. Frontiers of Economics and Globalization,* Elsevier.

Hillebrand, E. and Medeiros, M.C. (2008), "Asymmetries, breaks, and long-range dependence: An estimation framework for times series of daily realized volatility", *Working Paper*, Louisiana State University.

Huang, X. and Tauchen, G. (2005), "The relative contribution of jumps to total price variation", *Journal of Financial Econometrics,* 3: 456-499.

Hyung, N., Poon, S-H. and Granger C. W. J. (2005) "The source of long memory in financial market volatility", *Working Paper*, University of California at San Diego.

Jacod, J. and Protter, P. (1998), "Asymptotic error distributions for the Euler method for stochastic differential equations", *Annals of Probability,* 26: 267–307.

Koopman, S.J. Jungbacker, B. and Hol, E. (2005), "Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements", *Journal of Empirical Finance,* 12: 445–475.

Kroner, K.F. and Ng, V.K. (1998), "Modeling asymmetric comovements of asset returns", *Review of Financial Studies,* 11: 817–844.

Lieberman, O. and Phillips, P.C.B. (2008), "Refined inference on long memory in realized volatility", *Econometric Reviews,* 27: 254–267.

Madhavan A. (2000), "Market microstructure: A survey", *Journal of Financial Markets,* 3: 205–258.

Malmsten, H. and Teräsvirta,T. (2004), "Stylized facts of financial time series and three popular models of volatility", *Working Paper, Series in Economics and Finance,* 563, Stockholm School of Economics.

Martens, M. (2001), "Forecasting daily exchange rate volatility using intraday returns", *Journal of International Money and Finance,* 20: 1–23.

Martens, M. (2002), "Measuring and forecasting S&P 500 index-futures volatility using high-frequency data", *Journal of Futures Market,* 22: 497-518.

Martens, M. van Dijk D. and de Pooter, M. (2004), "Modeling and forecasting S&P 500 volatility: Long memory, structural breaks and nonlinearity", *Discussion Paper*, Tinbergen Institute 04-067/.

McAleer M. and Medeiros, M.C. (2008), "Realized volatility: A review", *Econometric Reviews*, 2(1-3): 10-45.

McAleer M. and Medeiros, M.C. (2008), "A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries", *Journal of Econometrics,* 147(1): 104-119

McKenzie, C.R. and McAleer, M. (1997), "On efficient estimation and correct inference in models with generated regressors: A general approach", *Japanese Economic Review,* 48: 368–389.

Meddahi, N. (2002), "A theoretical comparison between integrated and realized volatilities", *Journal of Applied Econometrics,* 17: 479-508.

Medeiros, M.C. and Veiga, A. (2004), "Modeling multiple regimes in financial volatility with a flexible coefficient GARCH(1,1) model", *Discussion paper* 486, Pontifical Catholic University of Rio de Janeiro.

Merton RC (1980) On estimating the expected return on the market: An exploratory investigation", *Journal of Financial Economics,* 8: 323-361.

Müller, U.A. Dacorogna, M.M. Davé, R.D. Olsen, R.B. Puctet, O.V. and von Weizsäcker, J. (1997), "Volatilities of different time resolutions – Analyzing the dynamics of market components", *Journal of Empirical Finance,* 4: 213–239.

Nielsen, M.Ø. and Frederiksen, P.H. (2006), "Finite sample accuracy and choice of sampling frequency in integrated volatility estimation", *Working Paper*, Cornell University.

O'Hara, M. (1995), *Market microstructure theory*, Oxford: Blackwell.

Ohanissian, A. Russell, J.R. and Tsay, R. (2004), "True or spurious long memory: Does it matter for pricing options?", Unpublished manuscript, Graduate School of Business, University of Chicago.

Pagan, A.R. (1984), "Econometric issues in the analysis of regressions with generated regressors", *International Economic Review,* 25: 221–247.

Pagan, A.R. (1986), "Two stage and related estimators and their applications", *Review of Economic Studies,* 53: 517–538

Parkinson, M. (1980), "The extreme value method for estimating the variance of the rate of return", *Journal of Business,* 53: 61-65.

Patton, A. (2005), "Volatility forecast evaluation and comparison using imperfect volatility proxies", *Research Paper Series*, no. 175, Quantitative Finance Research Centre, University of Technology, Sydney.

Poon, S. and Granger, C.W.J. (2003), "Forecasting volatility in financial markets: A review", *Journal of Economic Literature,* 41: 478–539.

de Pooter, M. Martens, M. and van Dijk, D. (2008), “Predicting the daily covariance matrix for S&P 100 stocks using intraday data - but which frequency to use?”, *Econometric Reviews,* 27: 199- 229.

Scharth, M. and Medeiros, M.C. (2006), “Asymmetric effects and long memory in the volatility of DJIA stocks”, *Working Paper*, Pontifical Catholic University of Rio de Janeiro.

Schwert, G.W. (1990), “Stock volatility and the crash of '87”, *Review of Financial Studies,* 3: 77-102.

Shephard, N. Barndorff-Nielsen, O.E. Hansen P.Hand Lunde, A. (2006), “Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise”, *Economics Series Working Papers*, no. 264, University of Oxford, Department of Economics.

Shephard, N. and Sheppard, K. (2009), “Realising the future: forecasting with high frequency based volatility (HEAVY) models”, *Journal of Applied Econometrics*, 2010, 25(2): 197-231.

Wasserfallen, W. and Zimmermann, H. (1985), “The behavior of intraday exchange rates”, *Journal of Banking and Finance,* 9: 55–72.

Yang, D. and Zhang Q. (2000) “Drift-independent volatility estimation based on high, low, open, and close prices”, *Journal of Business,* 73: 477-491.

Zhang, L. Mykl P.A. and Aït-Sahalia ,Y. (2005), “A tale of two time scales: determining integrated volatility with noisy high-frequency data”, *Journal of the American Statistical Association,* 100: 1394-1411.

Zhou, B. (1996), “High-frequency data and volatility in foreign exchange rates”, *Journal of Business and Economic Statistics,* 14: 45-52