416

RELATÓRIOS COPPEAD

Fevereiro 2015

EGARCH-RR: **Realized Ranges Explaining** EGARCH **Volatilities**

Victor Bello Accioly
Beatriz Vaz de Melo Mendes

Relatórios COPPEAD é uma publicação do Instituto COPPEAD de Administração da Universidade Federal do Rio de Janeiro (UFRJ)

Editora
Leticia Casotti

Editoração
Lucilia Silva

Ficha Catalográfica
Cláudia de Gois dos Santos

A171c Accioly, Victor Bello.
EGARCH-RR: realized ranges explaining EGARCH volatilities / Victor Bello Accioly, Beatriz Vaz de Melo Mendes. – Rio de Janeiro: UFRJ /COPPEAD, 2015.

25 p.; 27 cm. – (Relatórios COPPEAD; 416)

ISBN 978-85-7508-106-8
ISSN 1518-3335

1. Mercado de ações - Brasil. I. Mendes, Beatriz Vaz de Melo. II. Título. III. Série.

CDD: 332.63222

EGARCH-RR: Realized Ranges explaining EGARCH volatilities

Victor Bello Accioly
COPPEAD - Federal University at Rio de Janeiro, Brazil.

Beatriz Vaz de Melo Mendes
IM/COPPEAD - Federal University at Rio de Janeiro, Brazil.

Abstract

The purpose of this paper is to investigate whether the inclusion of a realized measure of volatility as external regressor on the GARCH and EGARCH variance equation would result in more accurate fits. The estimation of the model is performed by maximum likelihood with fifteen daily volatility series incorporated in the variance equation one at a time. The results show that the realized volatility measures add information to the EGARCH process; particularly, the realized range estimators that seem to outperform the realized volatility one. The scaling approach appears to be superior to the others that include squared overnight returns. GARCH is the most-adopted volatility model, which in itself justifies any improvement attempting. Besides, to the best of our knowledge, this is the first work to include range estimators to the EGARCH variance equation.

Keywords: GARCH; EGARCH; Realized Volatility; Realized Range; Brazilian stock market.

JEL subject classifications: $C22$, $C58$, $C32$.

Resumo

O propósito do artigo é investigar se a inclusão de medidas de volatilidade realizada como regressor externo na equação de variância dos modelos GARCH e EGARCH resultaria em melhores ajustes. A estimação é realizada pelo método de máxima verossimilhança com quinze séries de volatilidade incorporadas na equação de variância uma de cada vez. Os resultados mostram que medidas de volatilidade realizada adicionam informações para o processo EGARCH, principalmente os estimadores *realized range* que parecem ter melhor desempenho do que os *realized volatility.* A abordagem de um fator parece ser superior à que incluí o retorno *overnight* ao quadrado. O GARCH é o modelo de volatilidade mais adotado, o que por si só já justifica qualquer tentativa de melhoria. Além disso, de acordo com o nosso conhecimento, esse é o primeiro trabalho a incluir estimadores de intervalo na equação de variância EGARCH.

Palavras-Chave: GARCH; EGARCH; Volatilidade Realizada; *Realized Range*; Bovespa

1 Introduction

Volatility is a key piece in finance with significant role in investment, security valuation, risk management, and monetary policy making. Thus, most of the activity in the research area of a financial institution is devoted to modeling and forecasting of an asset volatility.

There are different volatility modeling approaches. In a comprehensive review of the methodologies and empirical findings of 93 papers, Poon and Granger (2003) classified the volatility forecasting methods in two broad categories, time series models and option implied standard deviation. The first is composed of three well known large classes of models: Historical volatility models, Generalized Autoregressive Conditional Heteroscedasticity (GARCH) family, and Stochastic Volatility (SV) models.

Since Engle's (1982) seminal paper, which motivation was to provide a tool for measuring the dynamics of inflation uncertainty, a long list of articles have empirically proved the usefulness of the GARCH family members, specially in-sample, in areas of economics and social sciences (e.g., the works of Bollerslev (1986, 1987), Engle et al. (1987), Nelson (1991), Glosten et al. (1993), and Hansen and Lunde (2005a)). Nonetheless, its out-of-sample predictive ability was cast in doubt after some unsatisfactory empirical results (e.g., Akgiray (1989), Kat and Heynen (1994), Franses and Van Dijk (1995), Brailsford and Faff (1996) and Figlewski (1997)), which were accredited to the disturbance caused by the inherent noise of the squared innovation used as a proxy for the ex-post volatility in the comparison with the predictions in Andersen and Bollerslev (1998), whom, therefore, used a new measure based on high-frequency returns, later called "realized variance", to show that daily GARCH volatility models perform well and accounts for around half the variability.

Chou (2005) appointed the flexibility of GARCH's volatility dynamics and its easier estimation procedure as its strengths against SV models. Brandt and Jones (2006) resumed the volatility forecasting difficulties in three main causes: sensitivity to the volatility model specification; difficulty of correctly estimate the parameters; and, volatility forecasts anchored at noisy proxies or at current level volatility estimates.

The literature related to the GARCH models is such enormous, as well as the list of acronyms resulted, that Bollerslev (2009) wrote a type of reference guide. Some of the models, the "range-based" ones, are related to the use of log price ranges, instead of closing prices in the model estimation (e.g., Alizadeh et al. (2002), Bali and Weinbaum (2005), Shu and Zhang (2006), Brandt and Diebold (2006), Brandt and Jones (2006) and Chou (2005)). The log price ranges are better volatility proxies than the log squared, or absolute, returns, for adding more information, particularly in turbulent periods; besides, the log price ranges distribution can be very well approximated by a Gaussian distribution, facilitating the maximum likelihood estimation.

The work of Mandelbrot (1971) was the first attempt to use range in the finance field; however, the work of Parkinson (1980) is the breakthrough for having established a relation between the range and the constant diffusion, which resulted in a volatility measure far more efficient than the classical closing prices only. Thereafter, many studies have followed, as Garman and Klass (1980), Beckers (1983), Rogers and Satchell (1991), Wiggins (1992), Kunitomo (1992), Yang and Zhang (2000), Chou (2005), Martens and van Dijk (2007), Christensen and Podolskij (2007), and Chou et al. (2010).

Several exogenous variables, as pointed out by Zivot (2008), have been shown to improve the volatility forecasts when added to GARCH's conditional variance formula, such as trading volume, macroeconomic news announcements, overnight returns, after hours volatility, implied volatility from option prices and realized volatility.

Following Day and Lewis (1992) methodology, although adding the realized volatility as exogenous variable in the variance equation instead of the implied volatility, Zhang and Hu (2013) cast some doubt if the former could provide additional information to the volatility process. It is worthy note that previously, Hansen et al. (2010) introduced the called Realized GARCH framework, which combines differently a GARCH structure and a model for realized volatility, finding considerable improvements when compared with the standard GARCH model. Applying this model to quantile forecasts of the S&P 500 index, Watanabe (2012) reached that the Realized GARCH model with the skewed Student's t-distribution outperform the other models in the research.

Some models incorporating daily, or weekly, ranges were found promising, such as the GARCH Parkinson Range of Mapa (2003), the Conditional Autoregressive Range of Chou (2005) and the Range-Based Autoregressive Volatility of Li and Hong (2011). Moreover, Li and Hong (2011) and Watanabe (2012) suggested the use of other realized measures of volatility such as the realized range. The Exponential GARCH (EGARCH) was adopted in Brandt and Jones (2006) for having the ability to capture the most important stylized characteristics of volatility series, a slight superiority of its specification over plain GARCH, the simplicity with which handle volatility asymmetry, and for its familiarity.

In this paper, we investigate whether the inclusion of realized range estimators as exogenous variables in the variance equation of EGARCH, and GARCH, models, according to Day and Lewis (1992) methodology, may result in better volatility forecasts and more accurate one-step-ahead risk measures in the Brazilian market.

The contribution of this paper is threefold. First, we show that the realized range and realized volatility estimators provide further information to EGARCH process. Second, forecasting stock volatility is important, specially for derivative pricing. Third, this study in an important emerging market provides valuable insights for foreign investors interested in investing into such markets.

The remainder of this paper is organized as follows. The next section provides the theoretical framework for the empirical analysis, including a brief discussion on high-frequency data, the definitions of realized volatility and realized range, and also a review of the GARCH class of models. Section 3 briefly describes the Brazilian stock market and the data sets used in the study as well as discusses the empirical results. Concluding remarks are contained in Section 4.

2 Literature Review

2.1 High-Frequency Data

The so called high-frequency data have recently become widely available despite the processing of a massive data amount being very time consuming requiring better hardwares and programming skills. Therefore, it is natural to question whether to use it or not. Andersen et al. (1999) present evidences of significant improvements in inter-daily volatility forecasts for exchange rate returns to justify its use. Another reason may be the usefulness of intraday market risk evaluation to market participants involved in frequent trading. Furthermore, So and Xu (2013) suggest that professionals may include an intraday seasonal indexes in the GARCH models, and incorporate realized variance and time-varying degrees of freedom to capture more intraday information on the volatile market.

Andersen et al. (2012) advocated that the rich information contained in the high-frequency data may be effectively exploited through the use of realized volatility measures. Several researches have been conducted to analyze alternative measures based on different assumptions and information sets, specially the properties of volatility proxies based on intra-daily data sampled at different frequencies (e.g, Andersen and Bollerslev (1998), Engle (2000), Barndorff-Nielsen and Shephard (2002), Meddahi (2002), Maheu and McCurdy (2011), Andersen et al. (2011)).

It is important to highlight that the use of high-frequency data to reveal prices until transaction-to-transaction level increases the market microstructure effects, which theoretical manifestation through the realized variance estimator is that the estimator fails to converge to the "true variance", increasing without bound when the sampling interval approaches the transaction-to-transaction level (Bandi and Russell, 2008).

2.2 Realized Volatility

The term realized volatility was first used in Fung and Hsieh (1991) and Andersen and Bollerslev (1998) to account for the sum of intraday squared returns at short intervals used as a proxy for "realised volatility", or an indicative of future price movements. The main reason for the new estimator was the interest in forecasting the unobserved integrated volatility over time intervals. According with Diebold and Strasser (2012), it has become popular due to being: model free, computationally trivial and, in principle, highly accurate.

The Andersen et al. (2001b) framework assumes that the logarithmic asset price increments evolve continuously through time according to a stochastic volatility diffusion process; and, the sum of the intraday squared returns converges to the integrated volatility through the quadratic variation theory. The estimator over a time interval of 1-day is defined as

$$RV_t(\Delta) = \sum_{j=1}^{n} r^2_{t-1+j\Delta}, \tag{1}$$

where $\Delta = 1/n$ is the sampling interval, n is the interval numbers in 1-day, and $r_{t-1+j\Delta} = p_{t-1+j\Delta} - p_{t-1+(j-1)\Delta}$ defines continuously compounded returns.

Under these assumptions including the absence of jumps and microstructure noise, the ex-post realized volatility is an unbiased volatility estimator providing a consistent measure of the integrated volatility when the sampling frequency theoretically goes to continuous basis.

In his seminal paper, Merton (1980) had already observed that, in the theoretical limit of continuous observation, the variance could in principle be estimated without error for any finite interval; however, the sampling frequency cannot be any higher than transaction by transaction. This sampling issue was addressed for several studies (e.g., Zhou (1996), Andersen et al. (2003) and Meddahi (2002)) by attempting to sample in intervals of five to thirty minutes. Hence, Bai et al. (2004) stated that large amounts of high-frequency data are not necessarily translated into precise estimates due to microstructure noise, but instead there is an optimal sampling interval.

By doing so, defining an optimal sampling interval, potentially valuable informations are lost. Some recent works (e.g., Zhang et al. (2005), Aït-Sahalia et al. (2005), Hansen and Lunde (2006), Bandi and Russell (2008) and Aït-Sahalia et al. (2011)) have used all information to estimate the volatility, what gave rise to a series of *robust realized variance* estimators.

2.3 Realized Range Volatility

The works of Martens and van Dijk (2007) and Christensen and Podolskij (2007) provided theoretical support for the integrated variance be estimated with the new realized range estimator, which empirically seems to work better than the realized variance with its efficiency being function of the interval sampling size.

The basis for the arise of this new estimator may, as pointed by Chou et al. (2010), remote to the fact that models based only on closing prices ignores the prices inside of the reference period. Hence, Parkinson (1980) came up with an estimator exploring the price range derived under the assumption that the asset price follows a simple diffusion model without a drift term in a daily periodicity framework, which Martens and van Dijk (2007) later extended to any interval, in particular to the intraday intervals employed by the realized variance.

The "Parkinson" realized range estimator is then defined as,

$$RR_t^{Parkinson}(\Delta) = \frac{1}{4 \ln 2} \sum_{j=1}^{1/\Delta} (u_j - d_j)^2 \tag{2}$$

$$u_j = \ln H_j - \ln O_j \qquad\qquad d_j = \ln L_j - \ln O_j \tag{3}$$

where H_j, L_j and O_j are the highest, the lowest and the opening prices of the jth interval on the tth trading day; and u_j and d_j are a kind of normalized high and low prices.

In their work, Martens and van Dijk (2007) also extended the two best Garman and Klass (1980) volatility estimators to realized range framework, which through simulation experiments were appointed more efficiently than the previous one, defined as:

$$RR_t^{GK1}(\Delta) = \sum_{j=1}^{1/\Delta} [0.511(u_j - d_j)^2 - 0.019[c_j(u_j + d_j) - 2u_j d_j] - 0.383c_j^2] \tag{4}$$

$$RR_t^{GK2}(\Delta) = \sum_{j=1}^{1/\Delta} [0.5(u_j - d_j)^2 - (2 \ln 2 - 1)c_j^2] \tag{5}$$

$$c_j = \ln C_j - \ln O_j \tag{6}$$

where C_j is the close price of the jth interval on the tth trading day; and c_j is a kind of normalized close price; the other variables are defined in Eq.(3).

The (three) previous estimators were derived based on the zero drift price process assumption. A version with a non-zero drift term is proposed in Rogers and Satchell (1991),

$$RR_t^{RS}(\Delta) = \sum_{j=1}^{1/\Delta} [u_j(u_j - c_j) + d_j(d_j - c_j)], \tag{7}$$

where the variables are defined in equations (3) and (6).

2.4 GARCH Family Models

The autoregressive conditional heteroscedastic (ARCH) processes of Engle (1982) was the first model to generalize the implausible assumption of constant one-period forecast variance. The basic idea behind this model is that the asset return innovations are serially uncorrelated and dependent, whose dependence could be described by a function of its lagged quadratic values. A generalized version of this process, allowing for longer memory and a more flexible lag structure, was proposed by Bollerslev (1986).

2.4.1 The GARCH process

The Generalized Autoregressive Conditional Heteroskedasticity process, GARCH(p,q), is given by

$$\varepsilon_t \mid \Psi_{t-1} \sim D(0, h_t), \tag{8}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}, \tag{9}$$

where ε_t denote a real-valued discrete-time stochastic process, Ψ_t the information set (σ-field) of all information through time t, and $D(0, h_t)$ denotes a distribution with mean zero and variance h_t; besides, α_i and β_j are nonnegative parameters with α_0 and β_j also different of zero.

The process defined by (8) and (9) is stationary if and only if $\left(\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1\right)$. Hence, $\mathbb{E}(\varepsilon_t) = 0$, $cov(\varepsilon_t, \varepsilon_s) = 0$ for $t \neq s$, and $\mathbb{V}ar(\varepsilon_t) = \alpha_0 \left(1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j\right)^{-1}$.

The GARCH process innovation at time t may be defined as $\varepsilon_t = r_t - \mu_t$ for a log return series r_t; or, another approach, referred in the literature as GARCH-in-mean, considers that an asset return may depend on its own volatility (Engle and Bollerslev, 1986):

$$r_t = \mu + c h_t + \varepsilon_t, \tag{10}$$

where the mean return μ and the new parameter c, called risk premium parameter, are constants.

2.4.2 The Exponential GARCH

Based on the GARCH model limitation of responding equally to positive and negative innovations, i.e., assuming that only the magnitude of unanticipated excess returns matters, Nelson (1991) proposed a model which responds asymmetrically to positive and negative residuals. The new Exponential GARCH process with weighted innovation $g(\varepsilon_t)$ is given by,

$$\ln(h_t) = \alpha_0 + \frac{1 + \beta_1 B + \cdots + \beta_{q-1} B^{q-1}}{1 - \alpha_1 B - \cdots - \alpha_p B^p}\, g(\varepsilon_{t-1}), \tag{11}$$

$$g(\varepsilon) \equiv \theta \varepsilon_t + \gamma \left[\mid \varepsilon_t \mid - \mathbb{E}(\mid \varepsilon_t \mid) \right], \tag{12}$$

where α_0, θ and γ are real constants; B is the lag operator such that $Bg(\varepsilon_t) = g(\varepsilon_{t-1})$; $1 + \beta_1 B + \cdots + \beta_{q-1} B^{q-1}$ and $1 - \alpha_1 B - \cdots - \alpha_p B^p$ are polynomials with zeros outside the unit circle and have no common factors; ε_t and $|\varepsilon_t| - \mathbb{E}(|\varepsilon_t|)$ are zero-mean i.i.d. random sequences. The $\mathbb{E}(|\varepsilon_t|)$ value depends on the ε_t distribution. It is worthy note that the positiveness of the conditional variance is guaranteed by this model specification differently of the GARCH model.

3 Data and Empirical Results

3.1 Data Description and Cleaning

Our dataset was obtained directly from BM&FBOVESPA, which trading scheme is consisted of two trading sessions, regular session and after-market, both preceded with pre-opening auction sessions. The continuous regular session begins at 10:00 and ends at 17:00 with its pre-opening session from 9:45 to 10:00. The after-market is from 17:35 until 19:00 with its pre-opening session from 17:30 to 17:35. These trading hours, in Braslia Time (BRT) time zone, are changed in view of the start of daylight saving time by the addition of one hour.

The tick-by-tick transaction information (time, price and volume), from December 1st 2009 to March 23rd 2012, of the eight most liquid constituent stocks of the Bovespa Index were used; i.e, PETR4 (Petrobras), VALE5 (Vale), TNLP4 (Telemar), USIM5 (Usiminas), BBDC4 (Banco Bradesco), CSNA3 (Companhia Siderúrgica Nacional), OXGP3 (OGX Petróleo) and ITUB4 (Itauunibanco).

3.2 Measure Construction

The earliest studies using realized volatility were based on exchange rate data, which are available for the whole day (apart from weekends). When considering financial instruments that are frequently traded only in part of the day, e.g., six to seven hours daily, the realized measures of volatility have to be extended for accounting the full day. Hansen and Lunde (2005b) considered three ways of doing it: scaling the estimator and adding, or combining, the squared overnight returns.

Since trading frequency and volume are much lower during the after-market session, and there is a 2% price variation boundary relative to the regular session closing price, the realized measures are constructed based on Δ-frequency intervals between the regular session hours only.

Therefore, there are three definitions of daily volatility in this paper, which are functions of the volatility measure and sampling interval chosen:

$$\sigma 1_t^2 = \delta R_t \tag{13}$$

$$\sigma 2_t^2 = r1_t^2 + R_t \tag{14}$$

$$\sigma 3_t^2 = \omega_1 r1_t^2 + \omega_2 R_t \tag{15}$$

where R_t is one of the five realized measures of variance previously defined (i.e., at equations (1), (2), (4), (5) and (7)), and $r1_t$ is the overnight return (i.e., the regular session close-to-open return) on date t; δ is the volatility measure scale factor, ω_1 and ω_2 are optimal linear combination weights, or factors, that minimizes the Mean Square Error (MSE),

$$\hat{\delta} = \sum_{t=1}^{n} (r_t - \bar{r})^2 / \sum_{t=1}^{n} R_t \tag{16}$$

$$\omega_1^* = (1-\varphi)\frac{\mu_0}{\mu_1} \qquad\qquad \omega_2^* = \varphi\frac{\mu_0}{\mu_2} \tag{17}$$

where the *relative factor* $\varphi = \frac{\mu_2^2\eta_1^2 - \mu_1\mu_2\eta_{12}}{\mu_2^2\eta_1^2 + \mu_1^2\eta_2^2 - 2\mu_1\mu_2\eta_{12}}$, $\hat{\mu}_0 = \sum_{t=1}^{n}(r_t - \bar{r})^2/n$, $\hat{\mu}_1 = \mathbb{E}(r1_t^2)$, $\hat{\mu}_2 = \mathbb{E}(R_t)$, $\hat{\eta}_1^2 = \sum_{t=1}^{n}(r1_t^2 - \hat{\mu}_1)^2/n$, $\hat{\eta}_2^2 = \sum_{t=1}^{n}(R_t - \hat{\mu}_2)^2/n$, and $\hat{\eta}_{12} = \sum_{t=1}^{n}(r1_t^2 - \hat{\mu}_1)(R_t - \hat{\mu}_2)/n$.

There are a total of 420 trading minutes in a normal trading day. Several studies found the optimal choice of the sampling frequency Δ for constructing the realized variance estimator to be between one and thirty minutes (Andersen et al. (2001a, 2003), Hansen and Lunde (2006) and Bandi and Russell (2008)). We investigate different choices of $\Delta \in \{1, 5, 10, 15, 20, 30, 60, 105, 210, 420\}$ minutes and 1-second for the construction of the realized range estimator, and some of them to the realized variance.

3.3 Methodology

This paper applies the same methodology applied in Zhang and Hu (2013) and in Day and Lewis (1992), which is the addition of external variables to a GARCH (EGARCH) model variance equation in order to improve its explanatory power. Specifically, each one out of fifteen realized measures of volatility – five types (Parkinson, GK1, GK2, RS and RV) times three definitions ($\sigma1_t^2$, $\sigma2_t^2$ and $\sigma3_t^2$) – is added one at a time, in the logarithmic form, at the following two model specifications:

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} + \sum_{k=1}^{m} \theta_k \ln\left(\sigma_{t-k}^2\right) \tag{18}$$

$$\ln\left(h_t\right) = \alpha_0 + \sum_{i=1}^{q} \alpha_i \frac{|\varepsilon_{t-1}| + \gamma_i \varepsilon_{t-i}}{\sqrt{h_{t-1}}} + \sum_{j=1}^{p} \beta_j \ln\left(h_{t-j}\right) + \sum_{k=1}^{m} \theta_k \ln\left(\sigma_{t-k}^2\right) \tag{19}$$

where σ_t^2 stands for one of the realized measures of volatility and m represents the number of lagged external variance factors. Through the R environment (R Core Team, 2013) and rugarch package (Ghalanos, 2013), used to perform the maximum likelihood estimation across the fifteen daily volatility series for several Δ sampling intervals, m is chosen equal to 3 for PETR4 and ITUB4 cases and equal to 2 for other cases.

The Akaike information criterion (AIC) is used, with the t-statistics for the individual coefficients, to test the hypothesis of improvement by adding the realized measure of volatility.

3.4 Empirical Results

After estimating the best GARCH and EGARCH models the realized measures of volatility (σ_t^2) are added and the model is re-estimated. The estimation results of

GARCH+σ_t^2 models show that information is not added to the simple GARCH. For parsimonious reasons, these results are not reported.

The Figure 1 graphs AIC versus Δ sampling frequency in minutes for EGARCH with realized range models. There are eight panels, each drawing one particular asset, and three letters – **P**, **G** and **R** – coloured by blue, red and green colour. The first letter represents Parkinson realized range type; the second, Garman and Klass 1[1]; and the third, Rogers and Satchell. The blue dashed line marks the AIC for the best EGARCH specification without external variance.

It is worth mentioning that nor for every Δ, the model estimations give reliable results. The McCullough and Vinod (2003) procedures for nonlinear optimization solution verification[2] were followed. Even when a computer software package returns a solution to an estimation problem, the solution may be inaccurate, or even does not exist; then a verification procedure must be employed. The cases where accurate solutions could not been found, with same model specification, were omitted from the graph.

At a first glance to Figure 1, the red colour catches the eye. The red letters, which represent the definition (14), are well separated from other colour letters, usually concentrated at the panel top area, except for the TNLP4 case. This means that $\sigma 2_t^2$ tends to provide worse fits. The blue and green letters, which represent the definitions (13) and (14) respectively, divide almost equally, and very close, the numbers of winning combinations. The Parkinson realized range type dominates, with five winning combinations, followed by Rogers and Satchell (two) and Garman and Klass (one). The optimal sampling frequency seems to be somewhere between one minute and thirty minutes.

Table 1 reports the estimation results of the winning combination for each asset. The first column identify the case. The second and third columns present the AIC for the EGARCH and EGARCH+RR_t model. The next three columns show the coefficients θ_1, θ_2 and θ_3 with their standard error below. The RR and σ_t^2 columns identify the range and the volatility definition that compose the model. The last three columns give the sampling interval (Δ) and the squared overnight returns ($\omega r1_t^2$) and realized range (ωR_t) weights respectively.

Analysing the results, since at least one of the θ coefficients for each asset is significant at the 5% level, moreover the EGARCH+RR_t model AIC is smaller than the EGARCH one, there are plenty evidences that realized range provides additional

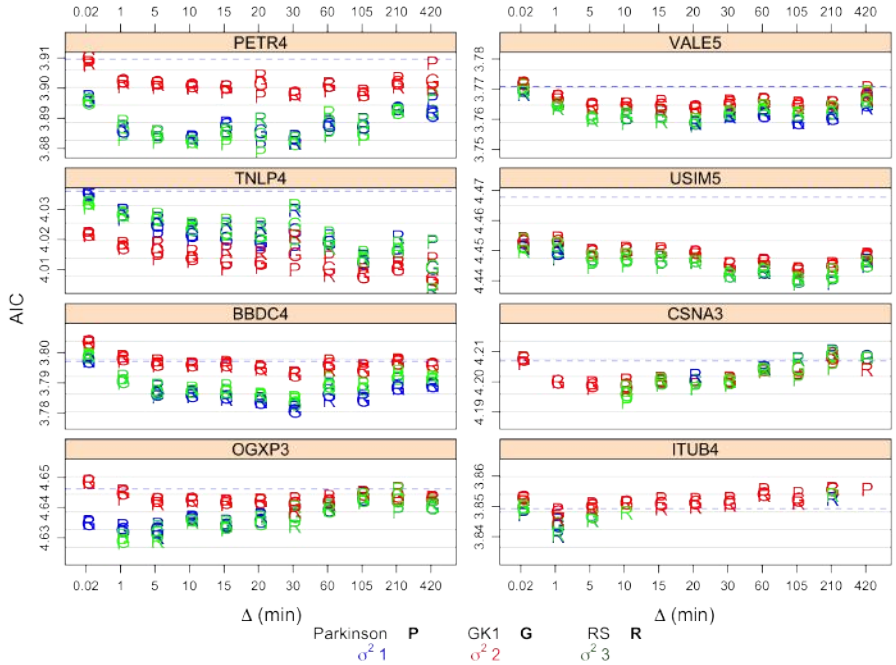

Figure 1: EGARCH+RR Fits: Sampling Frequency in minutes (Δ) × Akaike information criterion (AIC).

Notes: Each of the eight panel represents one stock for several realized range model specifications, where:
i) Parkinson, Garman and Klass 1, and Rogers and Satchell range types are labeled **P**, **G** and **R** respectively;
ii) $\sigma 1_t^2$, $\sigma 2_t^2$, and $\sigma 3_t^2$, defined on (13), (14) and (15), are represented by blue, red and green colours respectively;
iii) The blue dashed line (- - -) represents the Akaike information criterion (AIC) for the best EGARCH specification without external variance for each stock.

information content to EGARCH process.

Regarding to the realized variance, Figure 2 graphs AIC versus Δ sampling frequency in minutes for EGARCH with realized variance models. There are eight panels, one per asset, with circles filled with blue, red or green colour. The blue dashed line remain marking the AIC for the best EGARCH specification without external variance.

At Figure 2, the red circle, which represent the definition (14), still well separated from other colours and concentrated at the panel top area, except for the TNLP4 case. The $\sigma 2_t^2$ seems to provide worse fits in most of the cases; however, for the TNLP4 and ITUB4 cases, it constitutes the winning combination. Perhaps it happens due to some particularity of the TNLP4 data and the scarcity of ITUB4 reliable fits. Once again the blue colour ($\sigma 1_t^2$) excels. Although there are two cases showing one second[3] optimal sampling frequency, it might be related to same partic-

Asset	AIC EGARCH	AIC $+RR_t$	θ_1	θ_2	θ_3	RR	σ_t^2	Δ	$\omega r1_t^2$	ωR_t
PETR4	3.9096	3.8787	0.6798* (0.1698)	-0.6229* (0.2309)	0.1678 (0.1299)	P	$\sigma 3_t^2$	20	0.2243	1.3772
VALE5	3.7710	3.7580	0.3811** (0.1729)	-0.2322 (0.1724)	–	P	$\sigma 1_t^2$	20	0.0000	1.7618
TNLP4	4.0361	4.0031	0.4288* (0.1223)	0.3214* (0.1129)	–	RS	$\sigma 1_t^2$	420	0.0000	0.1401
USIM5	4.4677	4.4391	0.2014** (0.0822)	0.0001 (0.0051)	–	P	$\sigma 1_t^2$	105	0.0000	1.4009
BBDC4	3.7972	3.7807	0.4096* (0.1541)	-0.2720 (0.1807)	–	GK1	$\sigma 1_t^2$	30	0.0000	1.1427
CSNA3	4.2069	4.1933	0.4358* (0.1616)	-0.4167* (0.1128)	–	P	$\sigma 3_t^2$	10	-0.0016	1.1853
OGXP3	4.6462	4.6283	0.6419* (0.1665)	-0.5955* (0.1696)	–	P	$\sigma 3_t^2$	1	-0.0774	1.6177
ITUB4	3.8492	3.8403	0.5263* (0.1582)	-0.6971* (0.2508)	0.1763 (0.1443)	RS	$\sigma 1_t^2$	1	0.0000	1.3078

Notes: i) Estimation results of EGARCH model with realized range measures in the variance equation.

ii) The report is based on estimation of the following model:

$$\ln(h_t) = \alpha_0 + \sum_{i=1}^{q} \alpha_i \frac{|\varepsilon_{t-1}| + \gamma_i \varepsilon_{t-i}}{\sqrt{h_{t-1}}} + \sum_{j=1}^{p} \beta_j \ln(h_{t-j}) + \sum_{k=1}^{m} \theta_k \ln(\sigma_{t-k}^2)$$

iii) The columns AIC report EGARCH and EGARCH with realized range Akaike information criterion.

iv) The three coefficients θ_1, θ_2 and θ_3 are reported with their standard error below. *, **, *** statistically significant at 1%, 5% and 10% levels respectively.

v) The RR column stands for the realized range estimator type, where P, GK1 and RS are, respectively, Parkinson, Garman and Klass 1, and Rogers and Satchell estimators.

vi) $\sigma 1_t^2$, $\sigma 2_t^2$, and $\sigma 3_t^2$ refer to definitions (13), (14) and (15) respectively.

vii) $\omega r1_t^2$ and ωR_t stand for the weights in the linear combination.

viii) Δ stands for the sampling frequency in minutes.

Table 1: The best EGARCH+Realized Range Fits

ularly from OGXP3 and ITUB4 cases, the optimal sampling frequency seems again to be somewhere between one minute and thirty minutes.

The Figure 2 winning combination results are reported on Table 2, which follows the same composition of the previous table, except for substituting RR_t for RV_t and dropping the range type column.

The finding is confirmed with EGARCH+RV_t model results. At least one of the θ coefficients for each asset is significant at the 10% level[4], as well as the AIC is smaller than the EGARCH without external variable model.

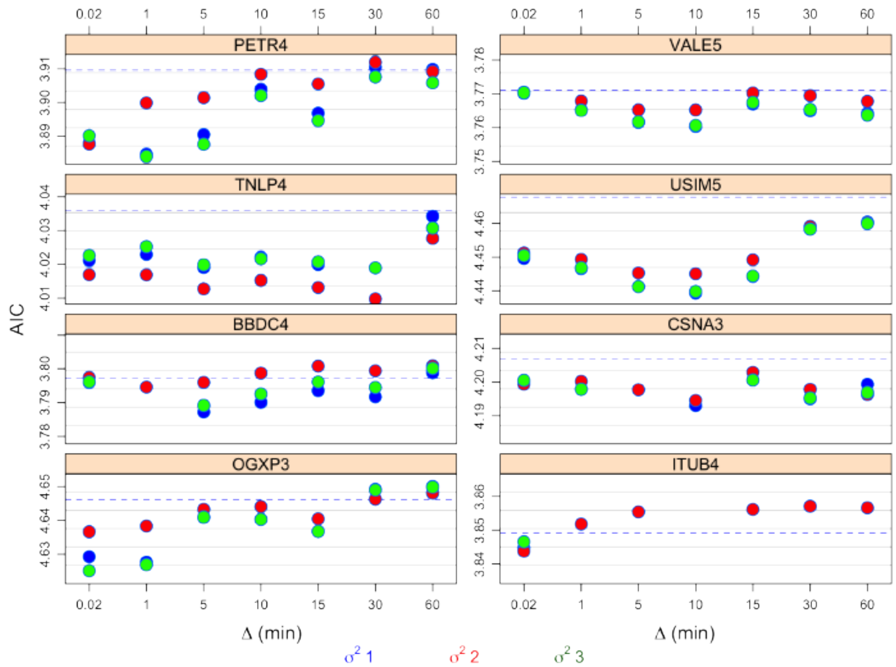

Figure 2: EGARCH+RV Fits: Sampling Frequency in minutes (Δ) × Akaike information criterion (AIC).

Notes: Each of the eight panel represents one stock for several realized volatility model specifications, where: i) $\sigma 1_t^2$, $\sigma 2_t^2$, and $\sigma 3_t^2$, defined on (13), (14) and (15), are represented by blue, red and green colours respectively;
ii) The blue dashed line (- - -) represents the Akaike information criterion (AIC) for the best EGARCH specification without external variance for each stock.

4 Conclusion

In this paper we empirically investigated whether or not the inclusion of realized range and realized volatility estimators into the GARCH and EGARCH models variance equation provide additional information to the process. According to the literature, the range estimator tends to be more efficient than the squared return. Our empirical findings are supported by the theory, the combinations between EGARCH model and realized range estimator outperform the realized variance ones, except for the OGXP3 and CSNA3 cases, which the squared overnight return factors are negative.

Another concluding remarks linked to the extant literature that usually the optimal linear estimated weights of the realized measure of volatility are disproportionately larger when compared with the squared overnight return ones, is that $\sigma 2_t^2$, definition (14), constantly shows the worse results. However, as expected, when

Asset	AIC EGARCH	AIC $+RV_t$	θ_1	θ_2	θ_3	σ_t^2	Δ	$\omega r1_t^2$	ωR_t
PETR4	3.9096	3.8840	0.7684* (0.2102)	-0.8192* (0.3155)	0.2307 (0.1707)	$\sigma 3_t^2$	1	0.0225	1.3764
VALE5	3.7710	3.7603	0.2521 (0.1543)	-0.1056 (0.1514)	–	$\sigma 1_t^2$	10	0.0000	1.9399
TNLP4	4.0361	4.0098	0.3761* (0.1103)	-0.2561** (0.1227)	–	$\sigma 2_t^2$	30	0.5000	0.5000
USIM5	4.4677	4.4393	0.1925*** (0.1092)	0.0562 (0.1617)	–	$\sigma 1_t^2$	10	0.0000	1.4762
BBDC4	3.7972	3.7872	0.2590*** (0.1463)	-0.1110 (0.1762)	–	$\sigma 1_t^2$	5	0.0000	1.1985
CSNA3	4.2069	4.1930	0.3852* (0.1285)	-0.3376** (0.1316)	–	$\sigma 1_t^2$	10	0.0000	1.1882
OGXP3	4.6462	4.6251	0.6524* (0.1635)	-0.6070* (0.1669)	–	$\sigma 3_t^2$	0	-0.0826	0.3260
ITUB4	3.8492	3.8438	0.454* (0.1527)	-0.6864* (0.2474)	0.236 (0.1521)	$\sigma 2_t^2$	0	0.5000	0.5000

Notes: i) Estimation results of EGARCH model with realized volatility measures in the variance equation.
ii) The report is based on estimation of the following model:

$$\ln(h_t) = \alpha_0 + \sum_{i=1}^{q} \alpha_i \frac{|\varepsilon_{t-1}| + \gamma_i \varepsilon_{t-i}}{\sqrt{h_{t-1}}} + \sum_{j=1}^{p} \beta_j \ln(h_{t-j}) + \sum_{k=1}^{m} \theta_k \ln(\sigma_{t-k}^2)$$

iii) The columns AIC report EGARCH and EGARCH with RV Akaike information criterion.
iv) The three coefficients θ_1, θ_2 and θ_3 are reported with their standard error below. *, **, *** statistically significant at 1%, 5% and 10% levels respectively.
v) $\sigma 1_t^2$, $\sigma 2_t^2$, and $\sigma 3_t^2$ refer to definitions (13), (14) and (15) respectively.
vi) $\omega r1_t^2$ and ωR_t stand for the weights in the linear combination.
vii) Δ stands for the sampling frequency in minutes, where 0 symbolize 1 second interval.

Table 2: The best EGARCH+Realized Volatility Fits

the squared overnight return contains some valuable information, $\sigma 3_t^2$ tends to give better fits than $\sigma 1_t^2$, definitions (15) and (13) respectively.

This methodology may be very useful at scenarios where high-frequency data are not available, specially because realized volatility estimators are unfeasible at daily intervals. At these scenarios the first choice to model an asset's volatility tends to be one across the GARCH family models, whose fits, as shown in the study, may be improved with realized range estimators. the improvement happens even in absence of high-frequency data, since the realized range estimators may be computed over daily sampling intervals, and the information required – open, high, low and close daily prices – are usually available.

These findings contrast with the doubt raised by Zhang and Hu (2013) on the notion that realized volatility could have additional information than the GARCH and EGARCH model. It might be explained by the former have included the square root of $\sigma 2_t^2$ definition into the EGARCH variance equation instead of the logarithmic form. The difference between both emerging markets, Brazilian and Chinese, can not be discarded as well.

It is important to highlight that these findings were based on an in-sample estimation. The out-of-sample performance needs to be investigated, in addition this study should be expanded to other markets with larger series. A comparison against option-implied volatility would also be interesting.

Notes

[1]The Garman and Klass 2 realized range type is not reported due to its observed similarity with Garman and Klass 1.

[2]i.e., vary the default options of the nonlinear solver by decreasing the tolerance, switching the convergence criterion, changing the algorithms and starting values.

[3]represented in the graph by its decimal minute approximation.

[4]The θ_1 coefficient of VALE5 is significant as the 10.5% level.

References

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(2):351–416.

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

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B. N. and Csaki, F., editors, *Proceedings of the 2nd International Symposium on information theory*, pages 267–281, Budapest. Akademiai Kiado.

Akgiray, V. (1989). Conditional heteroscedasticity in time series of stock returns: Evidence and forecasts. *The Journal of Business*, 62(1):55–80.

Alizadeh, S., Brandt, M. W., and Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance*, 57(3):1047–1091.

Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4):885–905.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., and Diebold, F. X. (2012). Financial risk measurement for financial risk management. NBER Working Papers 18084, National Bureau of Economic Research, Cambridge, MA, USA.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001a). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61(1):43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001b). 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(2):579–625.

Andersen, T. G., Bollerslev, T., and Lange, S. (1999). Forecasting financial market volatility: Sample frequency vis-à-vis forecast horizon. *Journal of Empirical Finance*, 6(5):457–477.

Andersen, T. G., Bollerslev, T., and Meddahi, N. (2011). Realized volatility forecasting and market microstructure noise. *Journal of Econometrics*, 160(1):220–234.

Bai, X., Russel, J., and Tiao, G. (2004). Effects of non-normality and dependence on the precision of variance estimates using high-frequency financial data. *Working papers, University of Chicago.*

Bali, T. G. and Weinbaum, D. (2005). A comparative study of alternative extreme-value volatility estimators. *Journal of Futures Markets*, 25(9):873–892.

Bandi, F. M. and Russell, J. R. (2008). Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies*, 75(2):339–369.

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, 17(5):457–477.

Beckers, S. (1983). Variances of security price returns based on high, low, and closing prices. *The Journal of Business*, 56(1):97–112.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 69(3):542–547.

Bollerslev, T. (2009). Glossary to ARCH (GARCH). In Bollerslev, T., Russell, J. R., and Watson, M., editors, *Volatility and Time Series Econometrics: Essays in Honour of Robert F. Engle*, pages 137–163. Oxford University Press, Oxford, UK.

Brailsford, T. J. and Faff, R. W. (1996). An evaluation of volatility forecasting techniques. *Journal of Banking & Finance*, 20(3):419–438.

Brandt, M. W. and Diebold, F. X. (2006). A no-arbitrage approach to range-based estimation of return covariances and correlations. *The Journal of Business*, 79(1):61–74.

Brandt, M. W. and Jones, C. S. (2006). Volatility Forecasting with Range-Based EGARCH Models. *Journal of Business & Economic Statistics*, 24(4):470–486.

Chou, R., Chou, H., and Liu, N. (2010). Range volatility models and their applications in finance. In Lee, C.-F., Lee, A., and Lee, J., editors, *Handbook of Quantitative Finance and Risk Management*, pages 1273–1281. Springer US.

Chou, R. Y. (2005). Forecasting Financial Volatilities with Extreme Values: The Conditional Autoregressive Range CARR Model. *Journal of Money, Credit and Banking*, 37(3):561–582.

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

Day, T. E. and Lewis, C. M. (1992). Stock market volatility and the information content of stock index options. *Journal of Econometrics*, 52(1–2):267–287.

Diebold, F. X. and Strasser, G. H. (2012). On the correlation structure of microstructure noise: A financial economic approach. NBER Working Papers 16469, National Bureau of Economic Research.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.

Engle, R. F. (2000). The econometrics of ultra-high-frequency data. *Econometrica*, 68(1):1–22.

Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, 5(1):1–50.

Engle, R. F., Lilien, D. M., and Robins, R. P. (1987). Estimating time varying risk premia in the term structure: The arch-m model. *Econometrica*, 55(2):391–407.

Figlewski, S. (1997). Forecasting volatility. *Financial Markets, Institutions & Instruments*, 6(1):1–88.

Franses, P. and Van Dijk, D. (1995). Forecasting stock market volatility using (non-linear) GARCH models. *Journal of Forecasting*, 15:229–235.

Fung, W. K. and Hsieh, D. A. (1991). Empirical analysis of implied volatility: Stocks, bonds and currencies. Working papers, Fuqua School of Business.

Garman, M. B. and Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *The Journal of Business*, 53(1):67–78.

Ghalanos, A. (2013). *rugarch: Univariate GARCH models*, R package version 1.2-7 edition.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess returns on stocks. *Journal of Finance*, 48(5):1779–1801.

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

Hansen, P. R., Huang, Z. A., and Shek, H. H. (2010). Realized garch: A complete model of returns and realized measures of volatility. CREATES Research Papers 2010-13, School of Economics and Management, University of Aarhus.

Hansen, P. R. and Lunde, A. (2005a). A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, 20(7):873–889.

Hansen, P. R. and Lunde, A. (2005b). A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics*, 3(4):525–554.

Kat, H. M. and Heynen, R. C. (1994). Volatility Prediction: A Comparison of the Stochastic Volatility, GARCH(1,1) and EGARCH (1,1) Models. *Journal of Derivatives*, 2(2).

Kunitomo, N. (1992). Improving the parkinson method of estimating security price volatilities. *The Journal of Business*, 65(2):295–302.

Li, H. and Hong, Y. (2011). Financial volatility forecasting with range-based autoregressive volatility model. *Finance Research Letters*, 8(2):69–76.

Maheu, J. M. and McCurdy, T. H. (2011). Do high-frequency measures of volatility improve forecasts of return distributions? *Journal of Econometrics*, 160(1):69–76.

Mandelbrot, B. B. (1971). When can price be arbitraged efficiently? a limit to the validity of the random walk and martingale models. *The Review of Economics and Statistics*, 53(3):225–236.

Mapa, D. S. (2003). A Range-Based GARCH Model for Forecasting Volatility. MPRA Paper 21323, University Library of Munich, Germany.

Martens, M. and van Dijk, D. (2007). Measuring volatility with the realized range. *Journal of Econometrics*, 138(1):181–207.

McCullough, B. and Vinod, H. (2003). Econometrics and sofware. *The Journal of Economic Perspectives*, 17(1):223–224.

Meddahi, N. (2002). A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics*, 17(5):479–508.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323–361.

Nelson, D. B. (1991). Conditional heteroscedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *The Journal of Business*, 53(1):pp. 61–65.

Poon, S.-H. and Granger, C. W. J. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, 41(2):478–539.

R Core Team (2013). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rogers, L. C. G. and Satchell, S. E. (1991). Estimating variance from high, low and closing prices. *The Annals of Applied Probability*, 1(4):pp. 504–512.

Shu, J. and Zhang, J. E. (2006). Testing range estimators of historical volatility. *Journal of Futures Markets*, 26(3):297–313.

So, M. and Xu, R. (2013). Forecasting intraday volatility and value-at-risk with high-frequency data. *Asia-Pacific Financial Markets*, 20(1):83–111.

Watanabe, T. (2012). Quantile Forecasts of Financial Returns Using Realized GARCH Models. *Japanese Economic Review*, 63(1):68–80.

Wiggins, J. B. (1992). Estimating the volatility of s&p 500 futures prices using the extreme-value method. *Journal of Futures Markets*, 12(3):265–273.

Yang, D. and Zhang, Q. (2000). Drift-independent volatility estimation based on high, low, open, and close prices. *The Journal of Business*, 73(3):477–91.

Zhang, J. and Hu, W. (2013). Does realized volatility provide additional information? *International Journal of Managerial Finance*, 9(1):70–87.

Zhang, L., Mykland, 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(472):1394–1411.

Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics*, 14(1):45–52.

Zivot, E. (2008). Practical issues in the analysis of univariate garch models. Working Papers UWEC-2008-03-FC, University of Washington, Department of Economics.

COPPEAD
UFRJ

Rua Pascoal Lemme, 355 Ilha do Fundão Rio de Janeiro - RJ CEP:21941-918 Caixa Postal 68514 21941-972
Tel:2598-9898 Fax:2598-9848 e-mail:atendimento@coppead.ufrj.br Home Page:www.coppead.ufrj.br