



BBR - Brazilian Business Review

E-ISSN: 1807-734X

bbronline@bbronline.com.br

FUCAPE Business School

Brasil

Bello Accioly, Victor; Vaz de Melo Mendes, Beatriz  
Assessing the Impact of the Realized Range on the (E)GARCH Volatility: Evidence from  
Brazil

BBR - Brazilian Business Review, vol. 13, núm. 2, marzo-abril, 2016, pp. 1-26

FUCAPE Business School

Vitória, Brasil

Available in: <http://www.redalyc.org/articulo.oa?id=123044248001>

- How to cite
- Complete issue
- More information about this article
- Journal's homepage in redalyc.org

redalyc.org

Scientific Information System

Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal

Non-profit academic project, developed under the open access initiative

## Assessing the Impact of the Realized Range on the (E)GARCH Volatility: Evidence from Brazil

**Victor Bello Accioly<sup>†</sup>**

*COPPEAD/Universidade Federal do Rio de Janeiro - UFRJ*

**Beatriz Vaz de Melo Mendes<sup>Ω</sup>**

*COPPEAD/IM/Universidade Federal do Rio de Janeiro - UFRJ*

### ABSTRACT

This paper investigates whether the inclusion of the realized range as regressor in the (E)GARCH volatility equation would add information to the process improving out-of-sample forecasts performance and providing more accurate estimates of the volatility persistence. Sixteen range measures at eleven data frequencies are tested using Brazilian stock market data. Several measures for assessing the improvements in the fits were used including the likelihood ratio test, the persistence percentage decrease, and a formal statistical test for comparing forecasts errors from competing models. We found that for both the GARCH and EGARCH models there are always some realized range type at some frequencies bringing information to the volatility process with considerable persistence reduction.

**Key words:** GARCH and EGARCH models. Realized volatility. Realized range. Volatility forecast.

JEL Classification: C22 C58 C32

*Received on 11/18/2014 ; Revised 03/10/2015 ; Accept on 04/26/2015 ; Published on 03/01/2016*

#### \* Corresponding author

<sup>†</sup> Doctor of Sciences in Management Institute COPPEAD by the Federal University of Rio de Janeiro .

**Link:** Visiting Professor of the Federal University of Rio de Janeiro .

**Address:** Rua Senador Vergueiro , no. 272, ap. 1201, Rio de Janeiro / RJ / 22230-001 .

**E- mail:** [me@victoraccioly.com.br](mailto:me@victoraccioly.com.br)

**Phone:** (21) 3259-0563 / 98767-2711

<sup>Ω</sup> Ph.D. in Statistics at the Department of Mathematics, Rutgers The State University of New York

**Link :** Associate Professor at the Institute of Mathematics of the Federal University of Rio de Janeiro

**Address:** Rua Ministro Raul Fernandes 210, ap. 1107 Botafogo , Rio de Janeiro , RJ , 22260040

**E- mail:** [beatriz@im.ufrj.br](mailto:beatriz@im.ufrj.br)

**Phone:** ( 21) 99811-1771

*Note from the Editor: This paper was accepted by Bruno Felix.*



*This work is licensed under a [Creative Commons License - Attribution 3.0 Unported](https://creativecommons.org/licenses/by/3.0/).*

## 1 INTRODUCTION

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

Among all different volatility modeling approaches, the GARCH family is unequivocally the most popular and well known one, probably due to its flexibility, easy to estimate and success when modeling many of the stylized facts about financial returns, in particular the presence of clusters of volatility. For a comprehensive review see Poon and Granger (2003). Since Engle's (1982) seminal paper, a long list of articles have empirically proved the usefulness of the GARCH family members in areas such as economics and social sciences (see, e.g., the works of Bollerslev (1986, 1987), Engle et al. (1987), Nelson (1991), Glosten et al. (1993), and Hansen and Lunde (2005a)). However, their out-of-sample predictive ability was cast in doubt after some unsatisfactory empirical results, see Akgiray (1989), Kat and Heynen (1994), Franses and Van Dijk (1995), Brailsford and Faff (1996) and Figlewski (1997)). It should be noted, though, that the poor out-of-sample performance could be accredited to the inherent noise of the squared returns used in many works as proxies in the evaluation criteria. In fact, Andersen and Bollerslev (1998) using as proxy a new measure based on high-frequency returns were able to show that GARCH volatility daily forecasts do perform well.

To further improve GARCH volatility forecasts one may include exogenous variables in the conditional variance equation. Usual choices of regressors include trading volume, macroeconomic news announcements, overnight returns, after hours volatility, implied volatility from option prices and realized volatility.

However, there is no consensus on the advantages of the use of such exogenous variables. For example, with respect to the addition of trading volume, Lamoureux and Lastrapes (1990) and Brooks (1998) found conflicting results. Using daily equity returns and GARCH models, Flannery and Protopapadakis (2002) investigated seventeen macroeconomic variables and found that six of them (Consumer Prices Index, Producer Price Index, a Monetary Aggregate, Balance of Trade, Employment Report, and Housing Starts) could be strong risk factor candidates. Similarly, Bomfim (2003) found that macroeconomic pre-announcement effects in the stock market could be highly significant.

Gallo and Pacini (1998), using six stock market indices, found evidence of improvements in the forecasting performance of the GARCH and EGARCH models when the overnight returns are inserted into the conditional variance equation. Martens (2002) showed that for the S&P 500 returns, the squared overnight return plus the realized volatility as well as the rescaled realized volatility are statistically significant. Taylor (2007) also found evidence of significant improvements in the conditional volatility forecasts of the S&P 500 index returns when including overnight information revealed in its E-mini futures market. Chen et al. (2012), by breaking up the after-hours period into pre-open, post-close and no-trading, investigate the role of the pre-open period variance in the GARCH conditional volatility.

The implied volatility seems also to bring important additional information to the conditional variance equation of some series. Blair et al. (2001) found that the index VIX implied volatility accounted for almost all relevant information about the conditional variance of the S&P 100 index daily returns. Giot (2003) and Koopman et al. (2005) showed the superiority of implied volatility for options, see also Day and Lewis (1992). We should note that even in mature markets not every stock has an option being traded. Hence, for these cases, model free realized measures of volatility may provide reliable exogenous alternatives. For example, Taylor and Xu (1997) included the 5-minute realized volatility and the daily implied volatility constructed based on exchange rate quotations to explain the dynamics of the Deutschemark/dolar volatility. They found that the former was not quite helpful for daily estimation, although appearing to be more informative than the later when hourly data was used. Zhang and Hu (2013) experimented the realized volatility as a regressor but did not provide a final answer to the question whether the realized measure could actually provide additional information to the volatility process. Thus, there is more to be explored in the world of the realized measures and their additional information to GARCH models.

Accordingly, in this paper we use Brazilian data to investigate whether the inclusion of the realized range as exogenous variables in the variance equation of GARCH and EGARCH models can provide additional information to the volatility process improving the one-step- ahead forecasts and providing more accurate estimates of the persistence. Our methodology is similar to those found in Zhang and Hu (2013) and Day and Lewis (1992), differing from that when we include the log-realized range as regressor, and also when we use formal statistical tests and graphics to evaluate the performance of the augmented model.

Our decision in considering Brazilian data is due to two reasons. First, the inexpressive number of investigations on volatility models based on high-frequency data

devoted to emerging markets. Second, the still open question whether or not an emerging market may provide valuable insights for foreign investors interested in investing in emerging or even in developed markets. Although Merton (1987) suggests that investors holding imperfectly diversified portfolios (those that cannot be fully diversify away firm-specific risk due to limited access to information in the presence of market frictions), and some economic theories suggest that idiosyncratic volatility should be positively related to expected returns (XU; MALKIEL, 2003; EWENS et al., 2013), Ang et al. (2006; 2009) found a negative relation between idiosyncratic volatility and future returns. Measuring the stock return idiosyncratic volatility relative to the Fama and French (1993) model, Ang et al. (2006; 2009) investigated 23 developed markets, with a closer analysis of the U.S. market for which more data were available, concluding that the puzzle of low returns to high idiosyncratic volatility stocks calls for further investigation. Choudhry and Osoble (2015) highlight that although the investigation of interdependence amongst international stock markets has important practical implications for asset allocation, risk management, and economic policy, the existing research on industry or firm level interdependence has been mainly focusing on developed markets.

The contribution of this paper is twofold. Firstly, we empirically show through tests and graphics that the realized range may provide further information to GARCH and EGARCH processes, resulting in better forecasts and persistence reduction. Secondly, although only Brazilian data is analyzed, this study may provide valuable insights for foreign investors interested in investing in emerging markets.

The remainder of this paper is organized as follows. The following two sections provide the theoretical framework for the empirical analysis, including a brief discussion on high-frequency data, the definitions of the realized measures, and a brief review of the (augmented) GARCH class of models. Section 4 describes the Brazilian stock market and the data sets used, carry on the analysis and discuss the empirical findings. Concluding remarks are given in Section 5.

## **2 REALIZED MEASURES**

High-frequency data have recently become widely available, promoting the development of faster and more sophisticated computer resources requiring from researchers beyond basic programming skills. In spite of being very expensive, some of their sub-products, such as the realized measures, indices and proxies, have become very popular.

Andersen et al. (1999) advocated that the rich information contained in the high-frequency data may be effectively exploited through the use of realized volatility measures.

According to varying assumptions made on information sets and frequencies, different realized volatility measures may be extracted from high frequency data. Andersen et al. (1999) showed that the realized volatilities provide a more meaningful and accurate alternative when evaluating the performance of daily or longer-run volatility forecasts. In fact, the two more important applications for these observed series may be (i) their ability for improving or more accurately assess forecasts performance, and (ii) helping quantifying market risk, see Andersen and Bollerslev (1998), Engle (2000), Barndorff-Nielsen and Shephard (2002b), Andersen et al. (2011), among others. In this paper we explore several definitions of realized measures and their usefulness in these two directions.

## 2.1 REALIZED VOLATILITY

The term realized volatility was first used by Fung and Hsieh (1991) and later on by Andersen and Bollerslev (1998) to denote the sum of intraday squared returns computed at short intervals. The idea was to propose a new estimator for the unobserved integrated variance over time intervals. As such, it was intended to be used as an indicative of future price movements. It is a model free, computationally trivial and, in principle, highly accurate estimator, see Diebold and Strasser (2012).

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

$$RV_t(\Delta) = \sum_{j=1}^n r_{t-1+j\Delta}^2 \quad (1)$$

where  $\Delta = 1/n$  is the sampling interval,  $n$  is the number of intervals within 1 day, and  $r_{t-1+j\Delta} = p_{t-1+j\Delta} - p_{t-1+(j-1)\Delta}$  defines the intraday  $\Delta$ -frequency continuously compounded return.

Assuming the absence of jumps and microstructure noise, the ex-post realized variance is an unbiased and consistent estimator of the integrated variance (IV) when the sampling frequency theoretically goes to the 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.

Therefore, in practice, the optimality of the realized variance as an estimate of the daily integrated variance is attached to smaller sampling intervals  $\Delta$ , which could be as small as those provided by the transaction-to-transaction prices. This sampling issue was addressed in several studies, see, e.g., Andersen et al. (2001a,b, 2003), Barndorff-Nielsen and Shephard (2002a,b), where sampling intervals of five to thirty minutes were considered. Bandi and Russell (2008) derived a mean squared error based optimal sampling frequency resulting in more accurate forecasts and some economic gains. Some recent works (see, e.g., Zhang et al. (2005), Aït-Sahalia et al. (2005), Hansen and Lunde (2006), Aït-Sahalia et al. (2011), and Barndorff-Nielsen et al. (2008)) have used all available intraday information, giving rise to the so called *robust realized variance* estimators.

## 2.2 REALIZED RANGE

A drawback of the realized variance as an estimator of the integrated variance is its strong dependence on the sampling interval. This issue may be better handled by the realized range, whose theoretical properties were derived in Martens and van Dijk (2007) and Christensen and Podolskij (2007).

Parkinson (1980) derived an estimator for the price range under the assumption that the daily asset price follows a simple diffusion model without a drift term. Noting that models based only on closing prices ignore the prices inside the reference period, the definition was later on extended to any interval, in particular to intraday intervals. The Parkinson realized range estimator  $RR_t^{Parkinson}(\Delta)$  is defined as

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

$$u_j = \ln H_j - \ln O_j \quad d_j = \ln L_j - \ln O_j \quad (3)$$

where  $H_j$ ,  $L_j$  and  $O_j$  are the highest, the lowest and the opening prices of the  $j$ th interval in the  $t$ th trading day; and  $u_j$  and  $d_j$  are normalized high and low prices.

Other definitions of realized ranges were then proposed. Extending the definition of the volatility estimator of Garman and Klass (1980), Martens and van Dijk (2007) proposed the  $RR_t^{GK1}(\Delta)$  and the  $RR_t^{GK2}(\Delta)$  estimators:

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

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

$$c_j = \ln C_j - \ln O_j \quad (6)$$

where  $c_j$  is a normalized closing price based on  $C_j$ , the closing price of the  $j$ th interval on the  $t$ th trading day, and where  $u_j$ ,  $d_j$  and  $O_j$  are as previously defined. Through simulation experiments, these range estimates were found to be more efficient than (2).

Assuming a non-zero drift term, Rogers and Satchell (1991) proposed the following range estimator

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

where  $u_j$ ,  $d_j$  and  $c_j$  are as previously defined. The authors proved that  $RR_t^{RS}(\Delta)$  is an unbiased volatility estimator whatever is the drift.

### 2.3 CONSTRUCTION OF THE REALIZED MEASURES

In order to construct the realized measures used in the empirical investigation of Section 4, we discuss two important issues.

The first one is concerned with the sampling frequency  $\Delta$ . Noting that several studies found the optimal choice for the sampling frequency  $\Delta$  to be between one and thirty minutes (see, e.g., Andersen et al. (2001a, 2003), Hansen and Lunde (2006) and Bandi and Russell (2008)), we take the following choices for  $\Delta$  in minutes: **{1/60, 1, 5, 10, 15, 20, 30, 60, 105, 210, 420}**. Since the trading frequency and volume are smaller during the after-market session, when it is observed only up to 2% price variation relative to the regular session closing price, the  $\Delta$ -frequency intervals are computed within the regular session hours only. For example, using  $\Delta = 15$  in a 420 minutes normal trading day results in  $n = 28$  observations of the realized measure.

The second problem is how to extend the definition of a realized measure so that it represents a volatility measure for the whole day. We note that the earliest studies on realized volatility were based on exchange rate data, which are available 24 hours/day (apart from



weekends). However, in the case of financial instruments, usually traded only in the active part of the day, one should somehow include the overnight data. Likewise Hansen and Lunde (2005b), we consider three ways for extending the realized measures of volatility: scaling the estimator or taking the squared overnight returns and either adding to (or combining with) the (daily) realized measure.

Let  $R_t(\Delta)$  represent any previously defined realized range (2), (4), (5) or (7), and let  $r_{1,t}$  denote the overnight return, that is, the close-to-open return, on date  $t$ . The first proposal scales  $R_t(\Delta)$  by a constant so that the resulting estimator has the desired expected value.

$$R_t^{scale}(\Delta) = \delta R_t(\Delta) \quad (8)$$

where  $\delta$  may be estimated (Hansen and Lunde, 2005b) using

$$\hat{\delta} = \sum_{t=1}^T (r_t - \bar{r})^2 / \sum_{t=1}^T R_t \quad (9)$$

where  $T$  is the sample size.

The second proposal just adds the squared overnight return

$$R_t^{+on}(\Delta) = r_{1,t}^2 + R_t(\Delta). \quad (10)$$

Noting that both  $R_t^{scale}(\Delta)$  and  $R_t^{+on}(\Delta)$  are actually linear combinations of  $R_{1,t}^2(\Delta)$  and  $R_t(\Delta)$  with weights  $(0, \delta)$  and  $(1, 1)$  a more general extension may be obtained as

$$R_t^{\omega}(\Delta) = \omega_1 r_{1,t}^2 + \omega_2 R_t(\Delta). \quad (11)$$

where the optimal weights  $\omega_1$  and  $\omega_2$  are those values minimizing the Mean Squared Error (MSE). Hansen and Lunde (2005b) derived the expression for the weights  $(\omega_1, \omega_2)$

$$\omega_1 = (1 - \varphi) \frac{\mu_0}{\mu_1} \quad \omega_2 = \varphi \frac{\mu_0}{\mu_2} \quad (12)$$

where  $\mu_0 = \mathbb{E}[IV]$ ,  $\mu_1 = \mathbb{E}[r_{1,t}^2]$ ,  $\mu_2 = \mathbb{E}[R_t]$ ,  $\varphi = \frac{\mu_2^2 \eta_1^2 - \mu_1 \mu_2 \eta_{12}}{\mu_2^2 \eta_1^2 + \mu_1^2 \eta_2^2 - 2\mu_1 \mu_2 \eta_{12}}$  is a relative importance factor, where  $\eta_1^2 = \text{Var}[r_{1,t}^2]$ ,  $\eta_2^2 = \text{Var}[R_t]$ , and  $\eta_{12} = \text{Cov}[r_{1,t}^2, R_t]$ . The authors also proposed the sample estimates  $\hat{\mu}_0 = (1/T) \sum_{t=1}^T (r_{1,t}^2 + R_t)$ ,  $\hat{\mu}_1 = (1/T) \sum_{t=1}^T r_{1,t}^2$ ,  $\hat{\mu}_2 = (1/T) \sum_{t=1}^T R_t$ ,  $\hat{\eta}_1^2 = (1/T) \sum_{t=1}^T (r_{1,t}^2 - \hat{\mu}_1)^2$ ,  $\hat{\eta}_2^2 = (1/T) \sum_{t=1}^T (R_t - \hat{\mu}_2)^2$ , and  $\hat{\eta}_{12} = (1/T) \sum_{t=1}^T R_t (r_{1,t}^2 - \hat{\mu}_1)$ , where  $T$  is the sample size.

### 3 THE AUGMENTED GARCH CONDITIONAL VOLATILITY

Volatility is not observable and clearly time varying. It shows periods of high volatility and periods of apparent calm, resulting in the well known clustering stylized fact. The Autoregressive Conditionally Heteroskedastic (ARCH) process of Engle (1982) successfully captures the volatility clusters and generalizes the implausible assumption of constant variance forecast. The basic idea behind this model is that the asset returns innovations are serially uncorrelated with a conditional variance which is a function of squared lagged returns. A generalized version of this process, the Generalized Autoregressive Conditionally Heteroskedastic (GARCH) process, with a more flexible lag structure was proposed by Bollerslev (1986).

Let  $P_t$  be the asset price at day  $t$  and  $r_t = \ln P_t - \ln P_{t-1}$  be the continuously compounded return on the asset over the period  $t-1$  to  $t$ . Let  $r_t = \mu_t + \sqrt{h_t} \varepsilon_t$ , where  $\mu_t = \mathbb{E}[r_t | \Psi_{t-1}]$  and  $h_t = \text{Var}[r_t | \Psi_{t-1}]$  are respectively the conditional mean and variance given the information set  $\Psi_{t-1}$  at time  $t-1$ , and  $\varepsilon_t \sim i.i.d.(0,1)$ . The GARCH( $p, q$ ) conditional variance is given by

$$h_t = \alpha_0 + \sum_{i=1}^p \alpha_i r_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j} \quad (13)$$

where  $\alpha_i$  and  $\beta_j$  are nonnegative parameters with  $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$  and  $\alpha_0 > 0$ . The unconditional variance of  $r_t$  is given by  $\alpha_0 (1 - \sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j)^{-1}$ .

Motivated by the GARCH model limitation of responding equally to positive and negative innovations, Nelson (1991) proposed a model for the logarithm of the conditional variance which also captures the effect of bad news. The Exponential GARCH (EGARCH) process specifies that

$$\ln h_t = \alpha_0 + \sum_{i=1}^p \alpha_i \left( \frac{|r_{t-i}|}{\sqrt{h_{t-i}}} - \mathbb{E} \left[ \frac{|r_{t-i}|}{\sqrt{h_{t-i}}} \right] \right) + \sum_{j=1}^q \beta_j \ln h_{t-j} + \sum_{i=1}^p \gamma_i \frac{r_{t-i}}{\sqrt{h_{t-i}}} \quad (14)$$

It is worthy noting that taking the logarithm of the conditional variance simplifies greatly the estimation procedure.

We now include the realized range as exogenous variable in the volatility equations (13) and (14). From now on we assume  $\mu_t = 0$  or that it has been estimated and subtracted from the return series. To unify the notations given in Section 2.3, let  $R_t^*(\Delta)$  represent any of

the  $R_t(\Delta)$  measures, or their three extensions  $R_t^{scale}(\Delta)$ ,  $R_t^{+on}(\Delta)$  and  $R_t^{\omega}(\Delta)$ , or their logarithms. The augmented GARCH and EGARCH conditional variance equations estimated in the empirical investigation in Section 4 are:

$$h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 h_{t-1} + \theta_1 R_{t-1}^*(\Delta) \quad (15)$$

and

$$\ln h_t = \alpha_0 + \alpha_1 \left( \frac{|r_{t-1}|}{\sqrt{h_{t-1}}} - \mathbb{E} \left[ \frac{|r_{t-1}|}{\sqrt{h_{t-1}}} \right] \right) + \beta_1 \ln h_{t-1} + \gamma_1 \frac{r_{t-1}}{\sqrt{h_{t-1}}} + \theta_1 R_{t-1}^*(\Delta) \quad (16)$$

In (15) the explanatory variable and its coefficient  $\theta_1$  must be positive, whereas in (16) there are no restrictions. Therefore, in the applications that follow, we consider the logarithm of the realized measures only in (16). We observe that a more general model with several lagged values of the regressor may be considered.

Equation (15) may be rewritten as

$$h_t = \alpha_0 + \alpha_1 (r_{t-1}^2 - h_{t-1}) + (\alpha_1 + \beta_1) h_{t-1} + \theta_1 R_{t-1}^*(\Delta) \quad (17)$$

The term  $(r_{t-1}^2 - h_{t-1})$  in (17) is known as the shock to volatility and possesses zero expected value. Thus, in the long run, the conditional volatility value has two important components. One is the previous value times  $(\alpha_1 + \beta_1)$ , and the other the positive quantity brought in by the explanatory variable. Clearly,  $(\alpha_1 + \beta_1)$  represents the persistence of the volatility, the rate at which the conditional variance reverts to its unconditional value. For the EGARCH model the persistence equals to  $\beta_1$ .

Models are estimated by maximum likelihood using the R environment and the `rugarch` package (GHALANOS, 2013). We test the explanatory power of the regressor using a *t*-test and assess the statistical significance of its contribution applying the standard likelihood ratio test (LRT). To test whether or not the inclusion of the realized range in the volatility equation improves forecasts and reduces persistence we apply formal tests and graphics. These evaluation procedures will be explained in the next section.

#### 4 DATA AND EMPIRICAL ANALYSIS

The dataset was obtained directly from BM&FBOVESPA, which trading scheme is consisted of two trading sessions, regular session and after-market, both preceded by pre-opening auction sessions. The continuous regular session, in Brasília time (BRT) zone, 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. The trading hours may change according to the daylight saving program, with the addition of one hour.

The data are composed by the tick-by-tick transactions information on time, price and volume, from December 1st 2009 to March 23rd 2012, on the eight currently most liquid constituent stocks of the Bovespa Index, namely the PETR4 (Petrobras), VALE5 (Vale), TNLP4 (Telemar), USIM5 (Usiminas), BBDC4 (Banco Bradesco), CSNA3 (Companhia Siderúrgica Nacional), OGXP3 (OGX Petróleo) and ITUB4 (Itaunibanco). Table 7 in the Appendix gathers some summary statistics on the corresponding daily log-returns series of 572 observations. In this table the stocks are ordered according to their standard deviations. Columns 7 and 8 give the values of the asymmetry and excess kurtosis sample coefficients, where the \* indicates which returns distributions are asymmetric with heavier tails than the Gaussian distribution at the 5% level. The larger excess kurtosis were observed for OGXP3 and TNLP4. The Jarque-Bera test for normality rejected the null hypothesis for all stocks.

The 20 realized measures (3 extensions on 5 definitions) are then computed for each  $\Delta$  frequency (a total of 11  $\Delta$  values) using the high frequency data for the 8 stocks. We recall that the main features of the unconditional distribution of stocks realized volatility are much less known for emerging markets. Regarding the realized range, the number of studies is even more limited. We illustrate and give in tables 1 and 2, for all stocks and for  $\Delta = 5$ -minutes, some summary statistics of the unconditional distribution of the logarithm of  $RR_t^{Parkinson}(\Delta)$  and its three extensions, and the logarithm of the  $RV_t(\Delta)$ , respectively.

In tables 1 and 2 we observe that all measures' distributions present large positive skewness and excess kurtosis, a well known stylized fact about volatility measures. The  $RR_t^{+on}(\Delta)$  definition produces, in general, the smallest minimum and the largest maximum, standard deviation and skewness coefficient. This behavior is in line with others previous works which credit this larger variability to the equal weights given to the daily and overnight quantities.

#### 4.1 ESTIMATING THE AUGMENTING VOLATILITY EQUATION

After applying the Ljung-Box test to the daily squared returns, and the Lagrange multiplier test for ARCH effects on the daily returns, and rejecting their null hypotheses for all stocks at the 5% significance level, we proceed in the estimation of the GARCH-type models.

We estimate by maximum likelihood the parameters of the GARCH and EGARCH specifications (13) and (14) with  $p = q = 1$  (the search for the best orders for the large majority of cases indicated the simple GARCH(1, 1) as the best model). All solutions provided by the statistical computer package were carefully checked following suggestions given in McCullough and Vinod (2003). Then, all realized range measures (32 types and 11 frequencies) were included in the volatility equations, and (15) and (16) were estimated. We recall that for the GARCH model the log-realized measures were not considered as exogenous variables.

**Table 1 - Summary Statistics of the Logarithm of the Parkinson Realized Range ( $\Delta = 5$  Minutes) and its Three Extensions**

Asset	$R_t^*$	Mean	Min	Max	StdDev	Skew	Kurt	$\omega_1$	$\omega_2$
BBDC4	$\ln R_t$	0.6748	-0.9166	2.7518	0.5575	0.7330	1.3116	0.0000	1.0000
	$\ln R_t^{\text{scale}}$	0.8381	-0.7533	2.9151	0.5575	0.7330	1.3116	0.0000	1.1774
	$\ln R_t^{+\text{on}}$	0.9556	-0.6516	3.4623	0.6401	0.8348	1.2195	1.0000	1.0000
	$\ln R_t^{\omega}$	1.0228	-0.5726	3.0810	0.5580	0.7407	1.2818	0.0653	1.3895
VALE5	$\ln R_t$	0.3010	-1.2660	3.0443	0.5657	0.6296	1.4521	0.0000	1.0000
	$\ln R_t^{\text{scale}}$	0.8897	-0.6773	3.6330	0.5657	0.6296	1.4521	0.0000	1.8016
	$\ln R_t^{+\text{on}}$	0.7197	-1.2090	3.7136	0.7326	0.8626	1.3426	1.0000	1.0000
	$\ln R_t^{\omega}$	0.8702	-0.7311	3.5769	0.5710	0.6332	1.3528	0.0912	1.7019
ITUB4	$\ln R_t$	0.6636	-1.0988	2.9103	0.5575	0.8249	1.4918	0.0000	1.0000
	$\ln R_t^{\text{scale}}$	0.9185	-0.8440	3.1652	0.5575	0.8249	1.4918	0.0000	1.2903
	$\ln R_t^{+\text{on}}$	0.9527	-0.7906	3.7341	0.6752	1.0293	1.9009	1.0000	1.0000
	$\ln R_t^{\omega}$	1.0654	-0.6934	3.3106	0.5578	0.8248	1.4860	0.0069	1.4914
PETR4	$\ln R_t$	0.5959	-0.8919	2.7741	0.5235	0.5951	1.3263	0.0000	1.0000
	$\ln R_t^{\text{scale}}$	1.0013	-0.4865	3.1795	0.5235	0.5951	1.3263	0.0000	1.4999
	$\ln R_t^{+\text{on}}$	0.9143	-0.7892	3.5600	0.6483	0.8287	1.2875	1.0000	1.0000
	$\ln R_t^{\omega}$	1.0149	-0.4845	3.1827	0.5254	0.5931	1.2787	0.0454	1.4980
TNLP4	$\ln R_t$	0.8740	-0.5705	2.6628	0.5482	0.4266	0.3129	0.0000	1.0000
	$\ln R_t^{\text{scale}}$	1.1589	-0.2856	2.9476	0.5482	0.4266	0.3129	0.0000	1.3296
	$\ln R_t^{+\text{on}}$	1.0955	-0.4776	3.7518	0.6221	0.8475	1.5632	1.0000	1.0000
	$\ln R_t^{\omega}$	1.1732	-0.2698	3.0025	0.5477	0.3563	0.2582	-0.1785	1.4074

CSNA3	$\ln R_t$	1.0657	-1.2157	3.2014	0.5783	0.4962	1.0713	0.0000	1.0000
	$\ln R_t^{\text{scale}}$	1.2730	-1.0084	3.4087	0.5783	0.4962	1.0713	0.0000	1.2303
	$\ln R_t^{+\text{on}}$	1.3591	-1.1102	4.1433	0.6811	0.7853	1.3278	1.0000	1.0000
	$\ln R_t^{\omega}$	1.4552	-0.8217	3.5968	0.5783	0.4904	1.0704	-0.0189	1.4849
USIM5	$\ln R_t$	1.1869	-0.9518	3.2429	0.5840	0.3590	0.8030	0.0000	1.0000
	$\ln R_t^{\text{scale}}$	1.5026	-0.6360	3.5587	0.5840	0.3590	0.8030	0.0000	1.3713
	$\ln R_t^{+\text{on}}$	1.4933	-0.9474	4.0022	0.6516	0.5590	1.1916	1.0000	1.0000
	$\ln R_t^{\omega}$	1.5551	-0.6180	3.5821	0.5807	0.3807	0.8302	0.1065	1.3959
OGXP3	$\ln R_t$	1.3365	0.1080	3.9453	0.5850	0.8484	1.4516	0.0000	1.0000
	$\ln R_t^{\text{scale}}$	1.7599	0.5314	4.3687	0.5850	0.8484	1.4516	0.0000	1.5271
	$\ln R_t^{+\text{on}}$	1.6615	0.1808	5.7882	0.7137	1.0218	2.5203	1.0000	1.0000
	$\ln R_t^{\omega}$	1.8201	0.6035	4.4514	0.5822	0.8393	1.4192	-0.0668	1.6588

Summary statistics of the logarithm of the Parkinson realized range and its three extensions across the 8 stocks.  $R_t^{\text{scale}}$ ,  $R_t^{+\text{on}}$  and  $R_t^{\omega}$  refer to definitions (8), (10) and (11), respectively. Kurt stands for the excess kurtosis, and  $\omega_1$  and  $\omega_2$  are the optimal weights. The sampling frequency is  $\Delta = 5$  minutes.

**Table 2 - Summary Statistics of the Logarithm of the Realized Variance ( $\Delta = 5$  Minutes) and its Three Extensions**

Asset	$R_t^*$	Mean	Min	Max	StdDev	Skew	Kurt	$\omega_1$	$\omega_2$
BBDC4	$\ln R_t$	0.6268	-0.9036	2.9614	0.6054	0.6486	0.8816	0.0000	1.0000
	$\ln R_t^{\text{scale}}$	0.8079	-0.7225	3.1425	0.6054	0.6486	0.8816	0.0000	1.1985
	$\ln R_t^{+\text{on}}$	0.9262	-0.6641	3.4068	0.6697	0.7096	0.8933	1.0000	1.0000
	$\ln R_t^{\omega}$	0.9812	-0.5724	3.2530	0.6045	0.6666	0.8258	0.2056	1.3370
VALE5	$\ln R_t$	0.2220	-1.5360	3.1365	0.6443	0.3861	0.8440	0.0000	1.0000
	$\ln R_t^{\text{scale}}$	0.8438	-0.9143	3.7582	0.6443	0.3861	0.8440	0.0000	1.8622
	$\ln R_t^{+\text{on}}$	0.6684	-1.3683	3.7133	0.7816	0.6709	1.0139	1.0000	1.0000
	$\ln R_t^{\omega}$	0.8062	-0.9527	3.6539	0.6468	0.4267	0.7929	0.1538	1.6755
ITUB4	$\ln R_t$	0.6260	-1.0956	3.1499	0.6143	0.6731	1.0466	0.0000	1.0000
	$\ln R_t^{\text{scale}}$	0.8827	-0.8389	3.4066	0.6143	0.6731	1.0466	0.0000	1.2927
	$\ln R_t^{+\text{on}}$	0.9303	-1.0878	3.7513	0.7095	0.8669	1.4999	1.0000	1.0000
	$\ln R_t^{\omega}$	1.0273	-0.7219	3.5235	0.6160	0.6853	1.0136	0.0871	1.4525
PETR4	$\ln R_t$	0.4941	-1.1343	2.9070	0.6092	0.4376	0.8731	0.0000	1.0000
	$\ln R_t^{\text{scale}}$	0.9494	-0.6789	3.3623	0.6092	0.4376	0.8731	0.0000	1.5768

TNLP4	$\ln R_t^{+on}$	0.8442	-1.0051	3.4904	0.7088	0.6014	0.8301	1.0000	1.0000
	$\ln R_t^{\omega}$	0.9299	-0.7490	3.2912	0.6126	0.4422	0.7509	0.1908	1.4438
	$\ln R_t$	1.0385	-1.5381	2.7331	0.5617	0.1083	0.5038	0.0000	1.0000
	$\ln R_t^{scale}$	1.1596	-1.4170	2.8541	0.5617	0.1083	0.5038	0.0000	1.1287
CSNA3	$\ln R_t^{+on}$	1.2380	-0.6182	3.7592	0.6177	0.5945	0.9228	1.0000	1.0000
	$\ln R_t^{\omega}$	1.2961	-1.3890	3.0000	0.5620	0.0838	0.6210	-0.0481	1.3076
	$\ln R_t$	1.1711	-1.2944	3.3629	0.5854	0.3345	0.8945	0.0000	1.0000
	$\ln R_t^{scale}$	1.2742	-1.1913	3.4660	0.5854	0.3345	0.8945	0.0000	1.1086
USIM5	$\ln R_t^{+on}$	1.4489	-1.1807	4.1167	0.6686	0.6678	1.2335	1.0000	1.0000
	$\ln R_t^{\omega}$	1.5281	-0.9455	3.7088	0.5845	0.3485	0.9047	0.0367	1.4131
	$\ln R_t$	1.2014	-0.6665	3.4633	0.6068	0.2882	0.6738	0.0000	1.0000
	$\ln R_t^{scale}$	1.4901	-0.3779	3.7520	0.6068	0.2882	0.6738	0.0000	1.3346
OGXP3	$\ln R_t^{+on}$	1.5127	-0.5686	4.0156	0.6570	0.4835	0.9581	1.0000	1.0000
	$\ln R_t^{\omega}$	1.5645	-0.2731	3.7642	0.5977	0.3376	0.6717	0.1947	1.3472
	$\ln R_t$	1.2551	-0.3509	3.9702	0.6417	0.6759	1.0293	0.0000	1.0000
	$\ln R_t^{scale}$	1.7232	0.1172	4.4383	0.6417	0.6759	1.0293	0.0000	1.5969
	$\ln R_t^{+on}$	1.6055	-0.1914	5.7936	0.7502	0.8838	2.1122	1.0000	1.0000
	$\ln R_t^{\omega}$	1.7540	0.1660	4.4937	0.6429	0.6508	0.9602	-0.0653	1.6880

Summary statistics of the logarithm of the realized volatility and its three extensions across the 8 stocks.  $R_t^{scale}$ ,  $R_t^{+on}$  and  $R_t^{\omega}$  refer to definitions (8), (10) and (11), respectively. Kurt stands for the excess kurtosis, and  $\omega_1$  and  $\omega_2$  are the optimal weights. The sampling frequency is  $\Delta = 5$  minutes.

The GARCH(1,1) and EGARCH(1,1) models based on either the Normal or on the  $t$ - student conditional distribution resulted in excellent fits for all stocks, in the sense that they were able to capture the dynamics observed in the second moment of the returns distributions. Using a fatter tail asymmetric conditional distribution did not further improve the fits. The augmented models were then estimated keeping the winning specification.

**Table 3 - Some Selected GARCH(1,1) and GARCH(1,1)+ $R_t^*$  Fits**

Stock ( $R_t$ , $\Delta$ , $R_t^*$ Cond.Distribution)	GARCH(1,1)	GARCH(1,1) with $R_t^*$		PPR	LRT
	$\alpha_1 + \beta_1$	$\alpha_1 + \beta_1$	$\theta_1$		
BBDC4 (RS, 420min, $R^{+on}, t$ )		0.8421	0.0815*	-12.7%	0.0003
BBDC4 (RS, 210min, $R^\omega, t$ )	0.9646	0.8453	0.1098*	-12.4%	0.0000
BBDC4 (RS, 210min, $R^{+on}, t$ )		0.8463	0.0806*	-12.3%	0.0001
VALE5 (Park, 30min, $R^{+on}, t$ )		0.8011	0.1513*	-18.2%	0.0002
VALE5 (Park, 210min, $R^\omega, t$ )	0.9792	0.8031	0.1431*	-18.0%	0.0002
VALE5 (Park, 420min, $R^{+on}, t$ )		0.8259	0.1238*	-15.7%	0.0002
ITUB4 (GK1, 420min, $R^{+on}, t$ )		0.7870	0.1157*	-18.4%	0.0003
ITUB4 (GK2, 420min, $R^{+on}, t$ )	0.9642	0.7868	0.1160*	-18.4%	0.0003
ITUB4 (RS, 420min, $R^{+on}, t$ )		0.7963	0.1074*	-17.4%	0.0002
PETR4 (RS, 1sec, $R^\omega, N$ )		0.3719	0.4237**	-60.8%	0.0000
PETR4 (Park, 1sec, $R^\omega, N$ )	0.9498	0.3772	0.4751**	-60.3%	0.0001
PETR4 (GK1, 1sec, $R^\omega, N$ )		0.3799	0.4412**	-60.0%	0.0000
TNLP4 (Park, 10min, $R^\omega, t$ )		0.8920	0.1110**	-3.9%	0.0001
TNLP4 (Park, 5min, $R^\omega, t$ )	0.9283	0.8975	0.1057**	-3.3%	0.0002
TNLP4 (RS, 30min, $R_t, t$ )		0.9166	0.1061**	-1.3%	0.0001
CSNA3 (GK1, 1min, $R^{+on}, t$ )		0.8883	0.0897*	-9.4%	0.0000
CSNA3 (GK2, 1min, $R^{+on}, t$ )	0.9805	0.8883	0.0898*	-9.4%	0.0000
CSNA3 (RS, 1min, $R^{+on}, t$ )		0.8899	0.0836*	-9.2%	0.0000
USIM5 (GK1, 1sec, $R_t, N$ )		0.8175	0.4708*	-12.3%	0.0000
USIM5 (GK1, 1sec, $R^{scale}, N$ )	0.9322	0.8175	0.0887*	-12.3%	0.0000
USIM5 (GK2, 1sec, $R_t, N$ )		0.8168	0.4706*	-12.4%	0.0000
OGXP3 (RS, 210min, $R^\omega, t$ )		0.4477	0.3033*	-51.9%	0.0000
OGXP3 (RS, 210min, $R_t, t$ )	0.9310	0.4668	0.4168*	-49.9%	0.0000
OGXP3 (RS, 210min, $R^{scale}, t$ )		0.4668	0.2935*	-49.9%	0.0000

\* and \*\* denote respectively significance at the 5% and at the 1% level. PPR stands for percentage persistence reduction, and LRT stands for likelihood ratio test.

We now look for indicators of whether or not the realized range helps explaining the evolution of the conditional (E)GARCH variance. Typically, for any given stock, all tested realized range measures  $R_t^*(\Delta)$  for some frequencies  $\Delta$  (we experimented 176 specifications in the case of the GARCH model, and 352 in the case of the EGARCH model) were found statistically significant improving the fit as measured by the likelihood ratio test. There were no differences between the GARCH and the EGARCH models with respect to how many realized measures specifications were statistically significant. For both models the winning frequencies at which the realized ranges brought in exploratory power are, ordered by number of cases, 1-second, 210-minutes, 420-minutes, 105-minutes and 30-minutes. The Rogers and Satchell and the Parkinson ranges at the winning  $\Delta$  frequencies were found to be the best choices as exploratory variables in the volatility equation. Results indicate that the Parkinson range should be preferable when analyzing the stocks PETR4, BBDC4, CSNA3, OGXP3, and ITUB4, and that the Rogers and Satchell range should be used in the case of



VALE5, whereas both ranges provided similar performance in the cases of TNLP4 and USIM5. The plain version  $R_t$  and the extension  $R_t^\omega$  are the most successful types.

**Table 4 - Some Selected EGARCH(1,1) and EGARCH(1,1)+ $R_t^*$  Fits**

Stock ( $R_t$ , $\Delta$ , $R_t^*$ Cond.Distribution)	EGARCH(1,1)	EGARCH(1,1) with $R^*$	PPR		
LRT					
	$\beta_1$	$\beta_1$	$\theta_1$		
BBDC4 (Park, 1sec, $\ln R^\omega$ , $\textcircled{S}$ )		0.9414	0.0421 <sup>*</sup>	-4.21%	0.0332
BBDC4 (GK2, 1sec, $\ln R_t$ , $t$ )	0.9828	0.9424	0.0417 <sup>**</sup>	-4.11%	0.0238
BBDC4 (GK1, 1sec, $\ln R_t$ , $t$ )		0.9427	0.0413 <sup>**</sup>	-4.08%	0.0239
VALE5 (Park, 210min, $\ln R_t$ , $N$ )		0.8386	0.1141 <sup>*</sup>	-10.86%	0.0008
VALE5 (Park, 60min, $\ln R_t$ , $N$ )	0.9408	0.8126	0.1451 <sup>*</sup>	-13.63%	0.0005
VALE5 (Park, 105min, $\ln R^{scale}$ , $\textcircled{S}N$ )		0.8428	0.1147 <sup>*</sup>	-10.42%	0.0009
ITUB4 (Park, 1sec, $\ln R_t$ , $N$ )		0.9726	0.0190 <sup>**</sup>	-1.55%	0.0267
ITUB4 (RS, 1sec, $\ln R_t$ , $N$ , $\textcircled{S}$ )	0.9879	0.9660	0.0243 <sup>**</sup>	-2.22%	0.0078
ITUB4 (RS, 15min, $\ln R_t$ , $N$ )		0.9578	0.0352 <sup>**</sup>	-3.05%	0.0198
PETR4 (Park, 210min, $\ln R_t$ , $N$ )		0.5776	0.3433 <sup>**</sup>	-36.62%	0.0000
PETR4 (Park, 210min, $\ln R^\omega$ , $\textcircled{S}N$ )	0.9114	0.5543	0.3409 <sup>**</sup>	-39.18%	0.0000
PETR4 (Park, 210min, $R^{+on}$ , $N$ )		0.6699	0.0440 <sup>*</sup>	-26.50%	0.0002
TNLP4 (RS, 420min, $\ln R_t$ , $t$ )		0.4282	0.4376 <sup>**</sup>	-55.12%	0.0000
TNLP4 (Park, 1sec, $\ln R_t$ , $t$ ) $\textcircled{S}$	0.9541	0.9438	0.0344 <sup>**</sup>	-1.08%	0.0161
TNLP4 (Park, 1sec, $R^\omega$ , $t$ )		0.9455	0.0089 <sup>**</sup>	-0.90%	0.0122
CSNA3 (Park, 1sec, $\ln R^{+on}$ , $t$ )		0.9463	0.0374 <sup>**</sup>	-4.54%	0.0136
CSNA3 (RS, 420min, $\ln R^{scale}$ , $\textcircled{S}t$ )	0.9913	0.9448	0.0527 <sup>**</sup>	-4.69%	0.0011
CSNA3 (RS, 1sec, $\ln R^{+on}$ , $t$ )		0.9567	0.0270 <sup>**</sup>	-3.49%	0.0174
USIM5 (Park, 1sec, $\ln R^{+on}$ , $\textcircled{S}t$ )		0.8227	0.1030 <sup>*</sup>	-13.74%	0.0001
USIM5 (RS, 1sec, $\ln R^{+on}$ , $t$ )	0.9537	0.8429	0.0898 <sup>**</sup>	-11.62%	0.0000
USIM5 (RS, 1sec, $\ln R^\omega$ , $t$ )		0.8271	0.0938 <sup>**</sup>	-13.27%	0.0000
OGXP3 (Park, 210min, $\ln R_t$ , $t$ )		0.6422	0.2901 <sup>**</sup>	-31.54%	0.0000
OGXP3 (RS, 1min, $\ln R_t$ , $t$ ) $\textcircled{S}$	0.9380	0.6699	0.2644 <sup>*</sup>	-28.58%	0.0000
OGXP3 (Park, 210min, $\ln R^{+on}$ , $t$ )		0.7241	0.2114 <sup>**</sup>	-22.80%	0.0000

\* and \*\* denote respectively significance at the 5% and at the 1% level. PPR stands for percentage persistence reduction, and LRT stands for likelihood ratio test.

Since the most important feature of the (E)GARCH volatility is its persistence, one way to assess the goodness of the modeling strategy is computing the persistence percentage decrease, that is, the change in the estimated persistence relative to the simple model. The smaller this (negative) quantity the more responsible is the exogenous variable for explaining the conditional volatility, decreasing the contribution of the past values. Tables 3 and 4 give, respectively for the GARCH and the EGARCH models, for all stocks and for some selected frequencies and realized measures, the estimate of the exogenous variable, the values of the persistence under the simple and extended models, along with the values of the persistence percentage reduction (PPR). We observe a considerable reduction in the persistence due to the

presence of the realized range in the volatility equation. The tables also provide the p-value of the likelihood ratio test (LRT) whose null hypothesis is that the simple model (with no exogenous variable) should be preferable. In table 3, the positive sign of the estimate of  $\theta_1$  is expected indicating that higher values for the range are associated with higher levels of the conditional volatility.

**Table 5 - Mean Absolute Percentage Error (mape) and the Median Absolute Percentage Error (mdape) Used to Compare the GARCH, GARCH+ $R_t^*$ , EGARCH, and EGARCH+ $R_t^*$  Volatility Forecast**

Stock ( $R_t$ , $\Delta$ , $R_t^*$ )	mape				mdape			
	G	G+ $R_t^*$	E	E+ $R_t^*$	G	G+ $R_t^*$	E	E+ $R_t^*$
BBDC4 (Park, 1sec, $\ln R^{\omega}_t$ , $t$ )	18.44	<b>17.89</b>	18.71	18.51	<b>13.12</b>	13.78	16.87	15.08
VALE5 (Park, 105min, $\ln$	43.82	34.75	38.53	<b>33.53</b>	38.62	<b>28.04</b>	35.06	29.72
ITUB4 (RS, 1sec, $\ln R_t$ , $N$ )	23.29	21.08	17.70	<b>17.35</b>	17.25	18.02	15.58	<b>15.35</b>
PETR4 (Park, 210min, $\ln R^{\omega}_t$ , $N$ )	30.80	24.72	25.18	<b>24.59</b>	24.82	<b>20.45</b>	21.11	20.48
TNLP4 (RS, 420min, $\ln R_t$ , $t$ )	20.31	<b>17.70</b>	20.42	18.23	17.98	<b>15.86</b>	18.49	16.35
CSNA3 (RS, 420min, $\ln R^{scale}_t$ , $t$ )	<b>14.34</b>	14.44	16.23	16.02	<b>10.77</b>	12.54	14.22	12.83
USIM5 (Park, 1sec, $\ln R^{+on}_t$ , $t$ )	14.16	16.58	<b>14.04</b>	18.58	11.58	12.70	<b>10.51</b>	14.24
OGXP3 (Park, 210min, $\ln R_t$ , $t$ )	28.72	<b>20.76</b>	22.97	20.85	22.31	<b>17.05</b>	18.11	17.40

Notation in table: Stock ( $R_t$ ,  $\Delta$ ,  $R_t^*$  Cond.Distribution) stands for Asset tick (Realized range, sampling frequency  $\Delta$ , extended realized measure  $R_t^*$  conditional distribution). G and E stands for models GARCH and EGARCH. The **boldface** type indicates the model with the smaller *mape* or *mdape*.

Figure 1 plots for some selected fits presented in the previous tables the evolution through time of the conditional volatility estimated under the four models. As expected, they are not very close.

Probably the most difficult step when assessing forecasts performance is the choice of the volatility proxy, since the true volatility  $\sigma_t$  is not observed. Recently, the realized volatility has been preferred upon the squares or the absolute values of returns (see, for example, Andersen and Bollerslev (1998)). We use as proxy for  $\sigma_t$  the 1-minute realized volatility, that is, the square root of  $RV_t$  equation (1), which we denote by  $\tilde{\sigma}_t$ . The forecast errors are then defined as  $e_t = \hat{\sigma}_t - \tilde{\sigma}_t$ , where  $\hat{\sigma}_t$  is the forecast under some GARCH-type model.

The out-of-sample one-step-ahead volatility forecasts exercise is carried on as follows. For each selected series we separate the initial 400 observations (data from December 3rd, 2009 to July 18th, 2011), estimate the (E)GARCH models with and without the realized range, and compute the one-step-ahead out-of-sample corresponding forecasts. Then we move the data window incorporating the next observation and keeping the same sample size of 400. Parameters are re-estimated. This results in 4 series of 171 out-of-sample one-step-ahead forecasts errors, denoted by  $e_{i,t}$ ,  $i = 1, 2, 3, 4$ , respectively the forecast errors from

models GARCH, GARCH+ $R_t^*$ , EGARCH, and EGARCH+ $R_t^*$ . The 8 selected model specifications for which we present the forecasts results are marked with an  $\textcircled{S}$  in Table 2.

**Table 6 - Diebold and Mariano Test for Comparing Volatility Forecasts Performances of [Model  $I$  vs Model  $J$ ]. Null Hypothesis is that Model  $J$  Provides Better Out-of-Sample Forecasts. Smaller Values of the Test Statistic Lead to the Rejection of the Null**

Stock ( $R_t$ , $\Delta$ , $R_t^*$ Cond.Distribution)	C1	C2	C3
	GARCH+ $R_t^*$ vs GARCH	EGARCH+ $R_t^*$ vs EGARCH	EGARCH+ $R_t^*$ vs GARCH+ $R_t^*$
BBDC4 (Park, 1sec, $\ln R^\omega$ , $t$ )	<b>-2.4776</b> (0.0071)	<b>-2.8983</b> (0.0021)	0.1033 (0.5411)
VALE5 (Park, 105min, $\ln R^{scale}$ , $N$ )	<b>-6.8324</b> (0.0000)	<b>-3.3370</b> (0.0005)	-0.1828 (0.4276)
ITUB4 (RS, 1sec, $\ln R_t$ , $N$ )	<b>-3.0657</b> (0.0013)	<b>-1.8894</b> (0.0303)	<b>-1.8584</b> (0.0324)
PETR4 (Park, 210min, $\ln R^\omega$ , $N$ )	<b>-3.3720</b> (0.0005)	0.7940 (0.7858)	-0.5717 (0.2841)
TNLP4 (RS, 420min, $\ln R_t$ , $t$ )	<b>-3.9424</b> (0.0001)	<b>-3.6698</b> (0.0002)	-0.2855 (0.3878)
CSNA3 (RS, 420min, $\ln R^{scale}$ , $t$ )	<b>-1.8079</b> (0.0362)	-1.6225 (0.0533)	-0.0902 (0.4641)
USIM5 (Park, 1sec, $\ln R^{+on}$ , $t$ )	0.2325 (0.5918)	0.7854 (0.7833)	2.0903 (0.9810)
OGXP3 (Park, 210min, $\ln R_t$ , $t$ )	<b>-3.3857</b> (0.0004)	-1.5215 (0.0650)	0.4324 (0.6670)

Notation in table: Stock ( $R_t$ ,  $\Delta$ ,  $R_t^*$  Cond.Distribution) stands for Asset tick (Realized range, sampling frequency  $\Delta$ , extended realized measure  $R_t^*$  conditional distribution). The **boldface** type indicates rejection of the null at the 5% significance level.

A simple scale independent measure of performance may be computed using the absolute values of the  $e_{i,t}$ ,  $t = 1, \dots, 171$ , see Hyndman(2006). It is the mean absolute percentage error, defined as  $mape_i = \text{mean}\left(\frac{|e_{i,t}|}{\hat{\sigma}_t}\right) * 100$ . We also report a more robust measure of discrepancy based on the median, namely, the median absolute percentage error, defined as  $mdape_i = \text{median}\left(\frac{|e_{i,t}|}{\hat{\sigma}_t}\right) * 100$ . Table 5 reports the  $mape_i$  and the  $mdape_i$  for models  $i = 1, \dots, 4$ , and for the selected stocks. Overall, we observe smaller values for the augmented models, with occasional disagreement between the two measures. In both cases where the simpler model seems to be providing forecasts closer to the proxy, the result could be a tie if measures' variability would have been taken into account.

We now formally test which model specification produces the better one-step-ahead volatility forecasts. To this end we use an statistic suggested in Diebold and Mariano (1995) and modified in Harvey et al. (1997), which compares the forecasts from two competing models,  $i$  and  $j$ . Consider the differences  $d_t = g(e_{i,t}) - g(e_{j,t})$ ,  $t = 1, \dots, T$ , where  $T$  is the sample size, here  $T = 171$ , and where  $g(x)$  typically is the square or the absolute value of  $x$ . Christoffersen and Diebold (1996) suggested penalizing differently over

and under predictions. However, due to the uncertainty on the best proxy for  $\sigma_t$  we just take  $g(x)$  as  $x^2$ . The test statistic is based on the mean difference  $\bar{d} = \text{mean}(d_t)$ . The null hypothesis is that model  $j$  provides better forecasts. Small negative values of  $\bar{d}$  lead to the rejection of the null. Asymptotically the standardized  $\bar{d}$  follows a standard normal distribution, see details in Diebold and Mariano (1995).

We carry on three out-of-sample forecasts comparisons. The comparison C1: GARCH+ $R_t^*$  versus GARCH, C2: EGARCH+ $R_t^*$  versus EGARCH, and C3: EGARCH+ $R_t^*$  versus GARCH+ $R_t^*$ . Table 6 gives the value of the test statistic and corresponding p-value, for a selection of stocks. The boldface type indicates the rejection of the null hypothesis at the 5% level. Results in the table are a very good summary of what has occurred for the remaining models specifications. That is, typically the GARCH+ $R_t^*$  and the EGARCH+ $R_t^*$  models outperform the simpler GARCH and EGARCH volatility forecasts, whereas there is no clear improvement when moving from the GARCH+ $R_t^*$  to the EGARCH+ $R_t^*$ .

Finally, Figure 2 shows the persistence evolution during the forecast period for the four competing models and for the same stocks of Figure 1. It is clear that the augmented volatility equation consistently attains remarkable reduction in persistence.

## 5 CONCLUSIONS

In this paper we empirically investigated whether or not the realized range provides additional information to the GARCH and EGARCH model. To this end we considered 4 definitions of the realized range. These measures were computed based on 11 frequencies, as well as 3 possible ways for incorporating the overnight return. This resulted in 176 realized range measures. All, along with their logarithm, were tested as exogenous variables in the conditional variance equation.

In summary, we found that the realized range is very successful when explaining part of the (E)GARCH conditional volatility. This modeling strategy reduces the persistence and improves the out-of-sample forecasts. We were able to reach a conclusion on the best frequencies (1- second, 210-, 420-, 105- and 30-minutes) and best realized range definition to be used, Parkinson and Rogers and Satchell, at the plain definition and based on a linear combination with the after hours returns. A possibility not explored here is to obtain the optimal weights of the linear combination in a joint estimation with the (E)GARCH parameters.

These findings are very promising whenever high-frequency data are not available, since the realized ranges may be computed over daily sampling intervals, and the information

required – open, high, low and close daily prices – are usually available. Although only Brazilian data were used, we think that findings might be extrapolated to other emerging markets due to their similarities.

## REFERENCES

- AÏT-SAHALIA, Y.; MYKLAND, P. A.; ZHANG, L. How often to sample a continuous-time process in the presence of market microstructure noise. **Review of Financial Studies**, v. 18, n. 2, p. 351–416, .
- \_\_\_\_\_.; \_\_\_\_\_.; \_\_\_\_\_. Ultra high frequency volatility estimation with dependent microstructure noise. **Journal of Econometrics**, v. 160, n. 1, p. 160–175, 2011.
- AKGIRAY, V. Conditional heteroscedasticity in time series of stock returns: evidence and forecasts. **The Journal of Business**, v.62, n. 1, p. 55–80, 1989.
- ANDERSEN, T. G.; BOLLERSLEV, T. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. **International Economic Review**, v. 39, n. 4, p. 885–905, 1998.
- ANDERSEN, T. G. et al. The distribution of realized stock return volatility. **Journal of Financial Economics**, v. 61, n. 1, p. 43–76, 2001a.
- ANDERSEN, T. G. et al. The distribution of realized exchange rate volatility. **Journal of the American Statistical Association**, v. 96, n. 453, p. 42–55, 2001b.
- ANDERSEN, T. G. et al. Modeling and forecasting realized volatility. **Econometrica**, v.71, n. 2, p. 579–625, 2003.
- ANDERSEN, T. G.; BOLLERSLEV, T.; LANGE, S. Forecasting financial market volatility: Sample frequency vis-à-vis forecast horizon. **Journal of Empirical Finance**, v. 6, n. 5, p. 457–477, 1999.
- ANDERSEN, T. G.; BOLLERSLEV, T.; MEDDAHI, N. Realized volatility forecasting and market microstructure noise. **Journal of Econometrics**, v.160, n. 1, p. 220–234, 2011.
- ANG, A. et al. The cross-section of volatility and expected returns. **The Journal of Finance**, v.61, n. 1, p. 259–299, 2006.
- ANG, A. et al. High idiosyncratic volatility and low returns: International and further us evidence. **Journal of Financial Economics**, v. 91, n. 1, p. 1–23, 2009.
- BANDI, F. M.; RUSSELL, J. R. Microstructure noise, realized variance, and optimal sampling. **Review of Economic Studies**, v. 75, n. 2, p. 339–369, 2008.
- BARNDORFF-NIELSEN, O. E. et al. Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. **Econometrica**, v. 76, n. 6, p. 1481-1536, 2008.

BARNDORFF-NIELSEN, O. E.; SHEPHARD, N. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. **Journal Of The Royal Statistical Society Series B**, v. 64, n. 2, p. 253–280, 2002a.

\_\_\_\_\_.; \_\_\_\_\_. Estimating quadratic variation using realized variance. **Journal of Applied Econometrics**, v.17, n. 5, p. 457–477, 2002b.

BLAIR, B. J.; POON, S.-H.; TAYLOR, S. J. Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns. **Journal of Econometrics**, v. 105, n. 1, p. 5 – 26, 2001.

BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. **Journal of Econometrics**, v.31, n. 3, p. 307–327, 1986.

\_\_\_\_\_. A conditionally heteroskedastic time series model for speculative prices and rates of return. **The Review of Economics and Statistics**, v.69, n. 3, p. 542–547, 1987.

BOMFIM, A. N. Pre-announcement effects, news effects, and volatility: Monetary policy and the stock market. **Journal of Banking & Finance**, v. 27, n. 1, p. 133 – 151, 2003.

BRAILSFORD, T. J.; FAFF, R. W. An evaluation of volatility forecasting techniques. **Journal of Banking & Finance**, v.20, n. 3, p. 419–438, 1996.

BROOKS, C. Predicting stock index volatility: can market volume help? **Journal of Forecasting**, v. 17, n. 1, p. 59–80, 1998.

CHEN, C.-H.; YU, W.-C.; ZIVOT, E. Predicting stock volatility using after-hours information: Evidence from the NASDAQ actively traded stocks. **International Journal of Forecasting**, v. 28, n. 2, p. 366– 383, 2012.

CHOUDHRY, T.; OSOBLER, B. N. Nonlinear interdependence between the us and emerging markets' industrial stock sectors. **International Journal of Finance & Economics**, v. 20, n.1, p. 61–79, 2015.

CHRISTENSEN, K.; PODOLSKIJ, M. Realized range-based estimation of integrated variance. **Journal of Econometrics**, v.141, n. 2, p. 323–349, 2007.

DAY, T. E.; LEWIS, C. M. Stock market volatility and the information content of stock index options. **Journal of Econometrics**, v.52, n. 1–2, p. 267–287, 1992.

DIEBOLD, F. X.; MARIANO, R. S. Comparing predictive accuracy. **Journal of Business & Economic Statistics**, v.13, n. 3, p. 253–263, 1995.

DIEBOLD, F. X.; STRASSER, G. H. On the correlation structure of microstructure noise: A financial economic approach. **NBER Working Papers 16469**, National Bureau of Economic Research, 2012.

ENGLE, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. **Econometrica**, v.50, n. 4, p. 987–1007, 1982.

\_\_\_\_\_. The econometrics of ultra-high-frequency data. **Econometrica**, v.68, n. 1, p. 1–22, 2000.

ENGLE, R. F.; LILIEN, D. M.; ROBINS, R. P. Estimating time varying risk premia in the term structure: the arch-m model. **Econometrica**, v. 55, n. 2, p. 391–407, 1987.

EWENS, M.; JONES, C. M.; RHODES-KROPP, M. The price of diversifiable risk in venture capital and private equity. **Review of Financial Studies**, v. 26, n. 8, p. 1854–1889.

FAMA, E. F.; FRENCH, K. R. Common risk factors in the returns on stocks and bonds. **Journal of Financial Economics**, v. 33, n. 1, p. 3 – 56, 1993.

FIGLEWSKI, S. Forecasting volatility. **Financial Markets, Institutions & Instruments**, v. 6, n. 1, p. 1–88, 1997.

FLANNERY, M. J.; PROTOPAPADAKIS, A. A. Macroeconomic factors do influence aggregate stock returns. **Review of Financial Studies**, v.15, n. 3, p. 751–782, 2002.

FRANSES, P.; VAN DIJK, D. Forecasting stock market volatility using (non-linear) GARCH models. **Journal of Forecasting**, v. 15, p. 229–235, 1995.

FUNG, W. K.; HSIEH, D. A. Empirical analysis of implied volatility: stocks, bonds and currencies. **Working papers**, Fuqua School of Business, 1991.

GALLO, G.; PACINI, B. Early news is good news: the effect of market opening on market volatility. **Studies in Nonlinear Dynamics and Econometrics**, v. 2, n. 4, p. 115–131, 1998.

Garman, M. B.; Klass, M. J. On the estimation of security price volatilities from historical data. **The Journal of Business**, v.53, n. 1, p. 67–78, 1980.

GHALANOS, A. **Rugarch**: Univariate GARCH models, R package version 1.2. 7 ed. 2013.

GIOT, P. The information content of implied volatility in agricultural commodity markets. **Journal of Futures Markets**, v.23, n. 5, p. 441–454, 2003.

GLOSTEN, L. R.; JAGANNATHAN, R.; RUNKLE, D. E. On the relation between the expected value and the volatility of the nominal excess returns on stocks. **Journal of Finance**, v. 48, n. 5, p. 1779–1801, 1993.

HANSEN, P.; LUNDE, A. Realized variance and market microstructure noise. **Journal of Business & Economic Statistics**, v.24, n. 2, p. 127–161, 2006.

\_\_\_\_\_.; \_\_\_\_\_. A forecast comparison of volatility models: does anything beat a GARCH(1,1)? **Journal of Applied Econometrics**, v.20, n. 7, p. 873–889, 2005a.

\_\_\_\_\_.; \_\_\_\_\_. A realized variance for the whole day based on intermittent high-frequency data. **Journal of Financial Econometrics**, v. 3, n. 4, p. 525–554, 2005b.

HARVEY, D.; LEYBOURNE, S.; NEWBOLD, P. Testing the equality of prediction mean squared errors. **International Journal of Forecasting**, v. 13, n. 2, p. 281 – 291, 1997.

KAT, H. M.; HEYNEN, R. C. Volatility prediction: a comparison of the stochastic volatility, GARCH(1,1) and EGARCH (1,1) Models. **Journal of Derivatives**, v. 2, n. 2, 1994.

KOOPMAN, S. J.; JUNGBACKER, B.; HOL, E. Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. **Journal of Empirical Finance**, v. 12, n. 3, p. 445–475, 2005.

LAMOUREUX, C. G.; LASTRAPES, W. D. Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects. **The Journal of Finance**, v. 45, n. 1, p. 221–229, 1990.

MARTENS, M. Measuring and forecasting S&P 500 index-futures volatility using high-frequency data. **Journal of Futures Markets**, v.22, n. 6, p. 497–518, 2002.

MARTENS, M.; VAN DIJK, D. Measuring volatility with the realized range. **Journal of Econometrics**, v. 138, n. 1, p. 181–207, 2007.

MCCULLOUGH, B.; VINOD, H. Econometrics and software. **The Journal of Economic Perspectives**, v. 17, n. 1, p. 223–224, 2003.

MERTON, R. C. On estimating the expected return on the market: an exploratory investigation. **Journal of Financial Economics**, v. 8, n. 4, p. 323–361, 1980.

\_\_\_\_\_. A simple model of capital market equilibrium with incomplete information. **The journal of finance**, v.42, n. 3, p. 483–510, 1987.

NELSON, D. B. Conditional heteroscedasticity in asset returns: a new approach. **Econometrica**, v. 59, n. 2, p. 347–370, 1991.

PARKINSON, M. The extreme value method for estimating the variance of the rate of return. **The Journal of Business**, v. 53, n. 1, p. 61–65, 1980.

POON, S.-H.; GRANGER, C. W.J. Forecasting volatility in financial markets: a review. **Journal of Economic Literature**, v. 41, n. 2, p. 478–539, 2003.

ROGERS, L. C. G.; SATCHELL, S. E. Estimating variance from high, low and closing prices. **The Annals of Applied Probability**, v. 1, n. 4, p. 504–512, 1991.

TAYLOR, N. A note on the importance of overnight information in risk management models. **Journal of Banking & Finance**, v. 31, n. 1, p. 161–180, 2007.

TAYLOR, S. J.; XU, X. The incremental volatility information in one million foreign exchange quotations. **Journal of Empirical Finance**, v. 4, n. 4, p. 317–340, 1997.

XU, Y.; MALKIEL, B. G. Investigating the behavior of idiosyncratic volatility. **The Journal of Business**, v. 76, n. 4, p. 613–645, 2003.

ZHANG, J.; HU, W. Does realized volatility provide additional information? **International Journal of Managerial Finance**, v. 9, n. 1, p. 70–87, 2013.

ZHANG, L.; MYKLAND, P. A.; AÏT-SAHALIA, Y. A tale of two time scales: determining in- tegrated volatility with noisy high-frequency data. **Journal of the American Statistical Association**, v. 100, n. 472, p. 1394–1411, 2005.



## APPENDIX

Table 7 - Summary Statistics for Daily Stock Returns (%)

Asset	Mean	Median	Min.	Max.	Std. Dev.	Skewness	Kurtosis
BBDC4	0.0244	0.0332	-6.8371	7.7926	1.6663	0.0052	1.2399 *
VALE5	0.0051	0.0439	-8.4009	5.7984	1.7145	-0.2425 *	1.9782 *
ITUB4	-0.0020	-0.0271	-7.9596	7.5120	1.7379	-0.0919	1.4905 *
PETR4	-0.0740	0.0000	-8.1362	4.9402	1.7851	-0.4160 *	1.2992 *
TNLP4	-0.0745	-0.0549	-9.4688	12.0500	1.9379	0.2537 *	4.4382 *
CSNA3	-0.0719	-0.0455	-9.6496	8.5958	2.0771	-0.0447	1.6088 *
USIM5	-0.1090	-0.0922	-6.8066	11.1372	2.3284	0.2951 *	0.9432 *
OGXP3	0.0055	0.0000	-16.6313	11.1156	2.6769	-0.8132 *	5.2743 *

Each sample contains 572 daily observations collected from 12/02/09 to 03/23/12. Kurtosis stands for the excess kurtosis. The \* indicates statistical significance at the 5% level when testing returns distributions asymmetry and kurtosis.

## APPENDIX 2

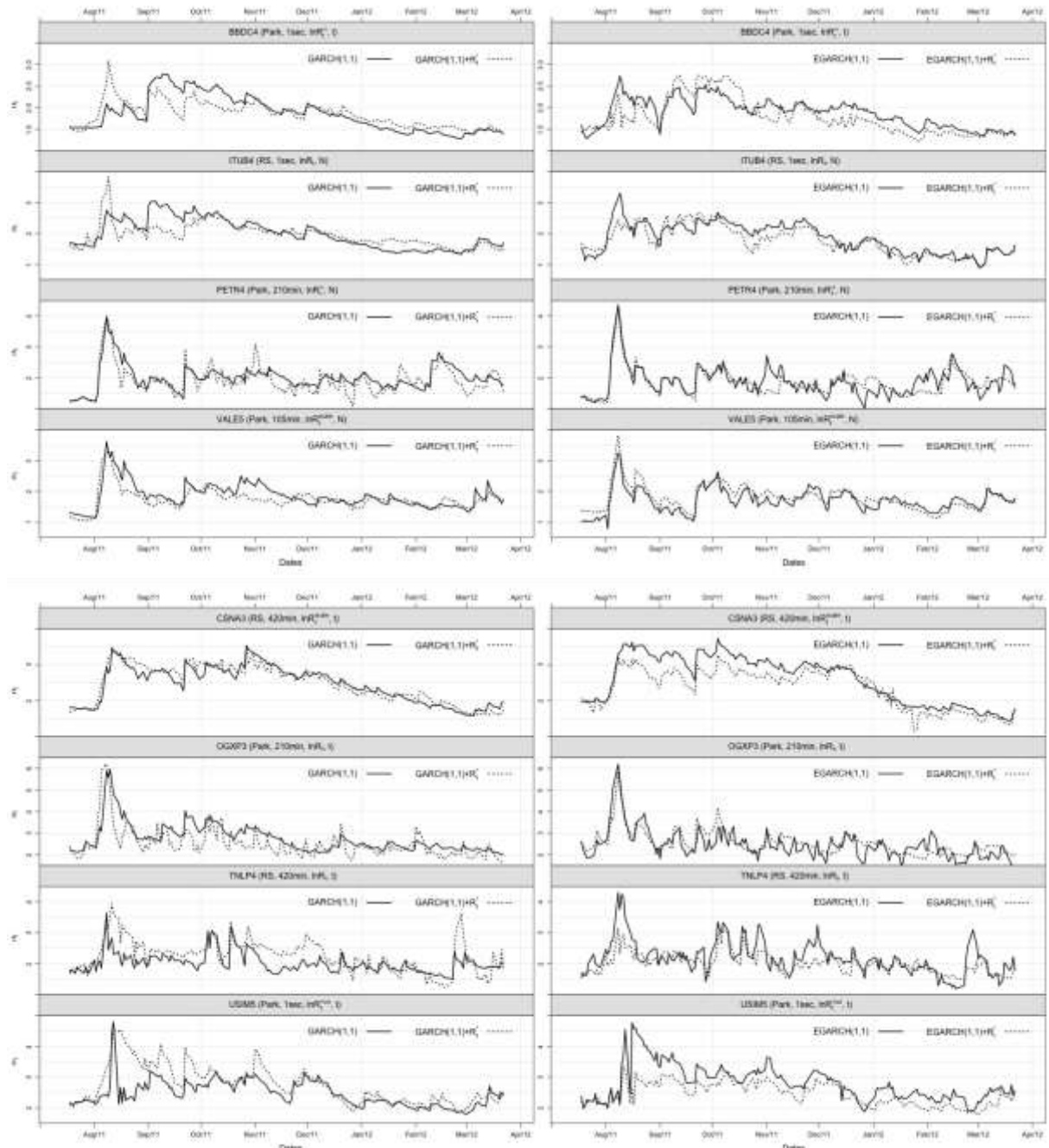


Figure 1: Evolution through time of the estimated conditional volatility, from July 19th 2011 to March 23rd 2012, under the four models across the eight stocks: BBDC4, ITUB4, PETR4, VALE5, CSNA3, OGXP3, TNLP4, and USIM5. The left column compares the GARCH (—) and the  $GARCH+R_t^*$  (·····) models. The right column compares the models EGARCH (—) and  $EGARCH+R_t^*$  (·····).

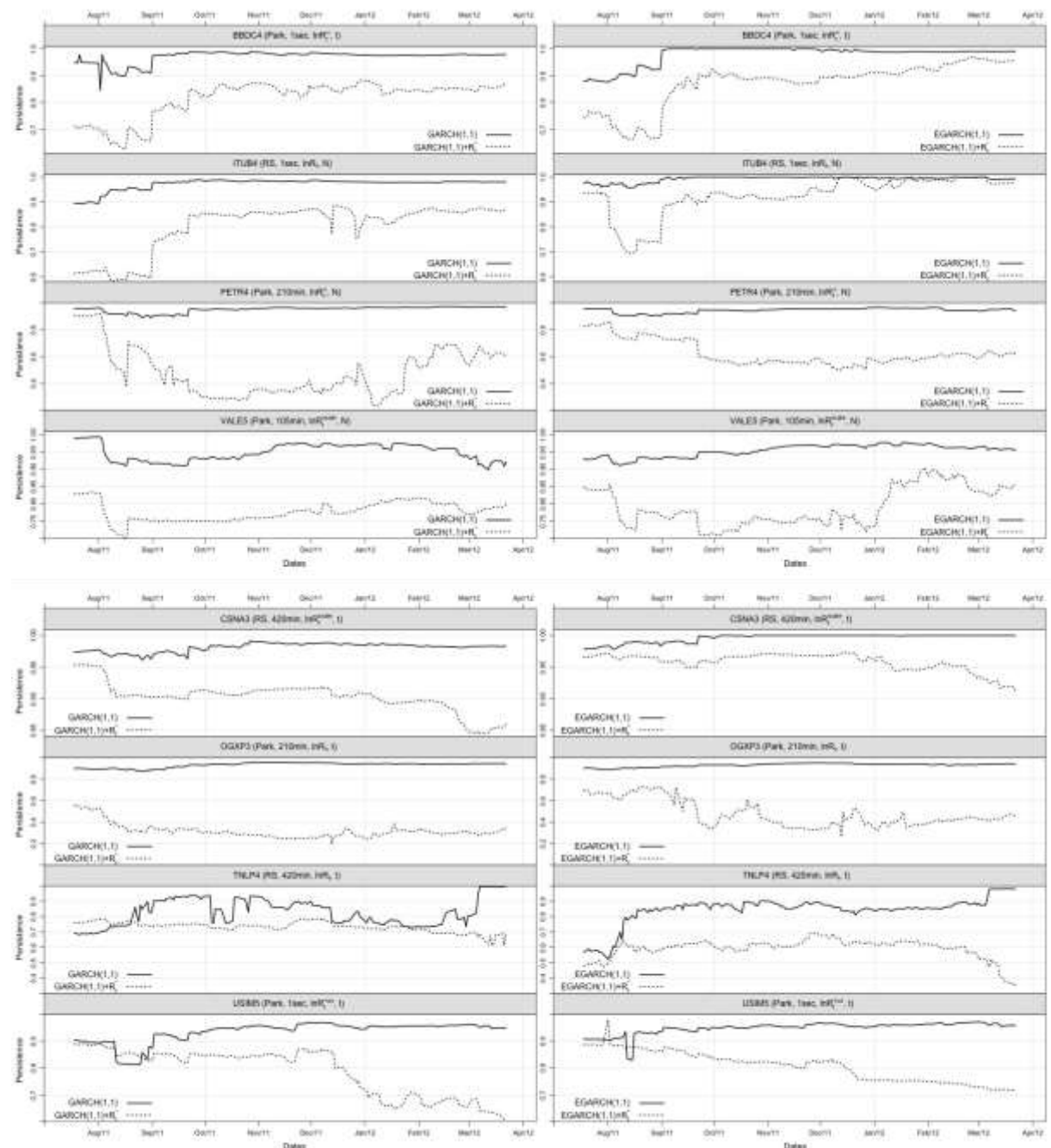


Figure 2: Evolution through time of the estimated persistence, from July 19th 2011 to March 23rd 2012, under the four models across the eight stocks: BBDC4, ITUB4, PETR4, VALE5, CSNA3, OGXP3, TNLP4, and USIM5. The left column compares the GARCH (—) and the  $GARCH+R_t^*$  (·····) models. The right column compares the models EGARCH (—) and  $EGARCH+R_t^*$  (·····).