

Bivariate Volatility Modeling with High-Frequency Data

Marius Matei ^{1,2,3,*}, Xari Rovira ⁴ and Núria Agell ⁴

¹ Department of Economics, Macquarie Business School, Macquarie University, Sydney, NSW 2109, Australia

² Systemic Risk Monitoring Division, Financial Stability Department, National Bank of Romania, Bucharest 030031, Romania

³ Centre for Macroeconomic Modelling, National Institute of Economic Research ‘Costin C. Kirişescu’, Romanian Academy, Bucharest 050711, Romania

⁴ Department of Operations, Innovation and Data Sciences, ESADE Business School, Ramon Llull University, Sant Cugat E-08172, Spain; xari.rovira@esade.edu (X.R.); nuria.agell@esade.edu (N.A.)

* Correspondence: mateimar@gmail.com; Tel.: +61-466-484-775 or +40-744-380-012

Received: 6 August 2018; Accepted: 6 September 2019; Published: 15 September 2019

Abstract: We propose a methodology to include night volatility estimates in the day volatility modeling problem with high-frequency data in a realized generalized autoregressive conditional heteroskedasticity (GARCH) framework, which takes advantage of the natural relationship between the realized measure and the conditional variance. This improves volatility modeling by adding, in a two-factor structure, information on latent processes that occur while markets are closed but captures the leverage effect and maintains a mathematical structure that facilitates volatility estimation. A class of bivariate models that includes intraday, day, and night volatility estimates is proposed and was empirically tested to confirm whether using night volatility information improves the day volatility estimation. The results indicate a forecasting improvement using bivariate models over those that do not include night volatility estimates.

Keywords: high-frequency; volatility; forecasting; realized measures; bivariate GARCH

JEL Classification: C32, C53, C58

1. Introduction

We aim to improve volatility modeling by adding information that exists on latent volatility processes while the markets are closed and no transactions occur. We build upon the observation that the price at market closing usually differs from the price at market opening, despite no transactions occurring between the two recordings. Models previously proposed usually estimate volatility by including information on past day and intraday volatility, estimated from day-recorded prices and sampled at various time intervals. Some papers have proposed methods to address overnight returns. The latent volatility component apparent in periods when markets are closed, highlighted by the difference between the two prices, may be the effect of events that occurred during the market closing, both domestic or international, or may be due to other latent factors that usually influence the financial markets, and may prove useful in volatility modeling. We propose an estimation of this night latent volatility and suggest a new model that uses day, intraday, and night volatility information to model day volatility. What distinguishes our contribution from other papers published on similar topics is that we propose a two-factor structure in a realized generalized autoregressive conditional heteroskedasticity (GARCH) setting that takes advantage of the natural

relationship between the realized measure and the conditional (day and night) variance. The mathematical structure is thus elegant, facilitates volatility estimation, and allows the inclusion of return-volatility dependence. We call the structure bivariate because it uses both day and night volatility information, as opposed to the univariate ones that only use day information. To strengthen the robustness of our empirical research, we further extended this idea to a number of realized GARCH models that use day and intraday volatility information, creating an equivalent set of bivariate models that additionally use night volatility information. We obtained a class of realized GARCH models that incorporate day, night, and intraday volatility measures; they were assessed against their counterparts that did not include night volatility information using an extended set of 10 stock prices. Empirical results of the forecasting performance assessment show a degree of improvement of the newly proposed models over those that do not include night volatility measures. This finding suggests the potential of our method for volatility forecasting problems for financial assets and other assets with night latent volatility information.

Financial volatility modeling has benefited significantly from the availability of high-frequency data. The main interest in modeling using frequently sampled information and integrating it into models built to estimate day conditional variance was initiated by Andersen and Bollerslev (1998), who used realized volatility estimates extracted from intraday data (realized variance) as better estimates of conditional volatility than squared returns. They proved that by adding up squared intraday returns, the forecasted volatility would correlate closely to the future latent volatility factor.

Engle (2002) was among the first econometricians who extended the standard GARCH model to include an exogenous realized measure (the realized variance) in the conditional variance (GARCH) equation. In this model, the realized measures' variation is not explained; thus, such models (GARCH-X) are considered incomplete. Engle and Gallo (2006) proposed the multiplicative error model (MEM), which was the first attempt to contain a separate GARCH structure equation for the realized measure. A similar complete model nested in a MEM setting is the high frequency based volatility (HEAVY) model of Shephard and Sheppard (2010). Both MEM and HEAVY models are difficult to use as they work with multiple latent processes—for every realized measure used, there is a corresponding latent volatility process. The Realized GARCH model proposed by Hansen et al. (2012) combines a GARCH structure for returns with realized measures of volatility. Compared with MEM and HEAVY models, the Realized GARCH model takes advantage of the natural relationship between the realized measure and the conditional variance. Instead of introducing additional latent factors, it proposes a single measurement equation in which the realized measure is a consistent estimator of the integrated variance. Besides its elegant mathematical structure, the Realized GARCH model is easy to estimate, captures the return-volatility dependence (leverage effect), and has been empirically shown to outperform conventional GARCH. A more robust version of the Realized GARCH model was introduced by Banulescu-Radu et al. (2019), suggesting a variant that is less sensitive to outliers and minimizes the impact on volatility of days with extreme negative volatility shocks. A realized exponential GARCH model that can use multiple realized volatility measures for the modeling of a return series, using a similar framework, has also been proposed (Hansen and Huang 2016). Finding that the Realized GARCH model was insufficient for capturing the long memory of underlying volatility, Huang et al. (2016) developed a parsimonious variant of the Realized GARCH model by introducing Corsi's (2009) heterogeneous autoregressive (HAR) specification in the volatility dynamics. A multivariate GARCH model that incorporates realized measures of variances and covariances was also introduced by Hansen et al. (2014), but it did not suggest the introduction of night volatility information. Bollerslev et al. (2018) proposed asymmetric multivariate volatility models that exploit estimates of variances and covariances based on the signs of high-frequency returns to allow for more nuanced responses to positive and negative return shocks than the threshold leverage effect. Hansen et al. (2019) proposed a multivariate GARCH model that incorporates realized measures for the covariance matrix of returns.

Overnight (close-to-open) volatility is usually higher than the five-minute realized volatility estimated during trading hours, and the close-to-open price differential may trigger a distorting effect

on the realized volatility. Thus, the inclusion of overnight returns when constructing the realized conditional covariance matrix of the daily returns has been empirically documented to reduce information loss and consequently improve volatility forecasting. A common approach to account for volatility during the market's closing hours has been to calculate a close-to-open return from the price change recorded between the trading day closing and the next trading day opening, and then add its squared value to the sum of intraday returns (Bollerslev et al. 2009; Martens 2002; Blair et al. 2001). Hansen and Lunde (2005) compounded optimal weights corresponding to overnight returns and to the sum of intraday returns, and Fleming and Kirby (2011) and Fuertes and Olmo (2013) further applied it. De Pooter et al. (2008) and Fleming et al. (2003) computed it in matrix form by incorporating the cross-product of the vector of overnight returns in the summation of the matrix that provided the covariance matrix of the daily returns, acknowledging that the outer product of the vector of overnight returns is an inaccurate estimator of the integrated covariance matrix for the period when markets were closed (Fleming et al. 2003). Koopman et al. (2005); Martens (2002); and Angelidis and Degiannakis (2008) excluded the noisy overnight returns to compute an estimate of volatility during trading hours, instead of daily volatility; then, they scaled up the sum of intraday returns to cover the whole 24-h day. The literature has not yet reached a consensus on the best method of accounting for overnight returns; however, Ahoniemi and Lanne (2013) suggested that the weighted sum of the squared overnight return and the sum of intraday squared returns was the most accurate measure of realized volatility for the Standard&Poor's S&P 500 index.

This paper suggests a method of capturing and incorporating night volatility into the day conditional volatility equation of one low-frequency as well as a number of high-frequency GARCH models. We propose a two-factor structure of the conditional variance, one for night and one for day variance, in a realized GARCH setting that takes advantage of the natural relationship between the realized measure and the conditional (day and night) variance. The mathematical structure is thus elegant, facilitates volatility estimation, and allows the inclusion of the return-volatility dependence. A general framework is formulated; based on it, a set of GARCH models is adapted such that it uses the estimation of night latent volatility to model day conditional volatility. This approach enabled us to document, in an empirical context, whether the introduction of the night volatility component, in the two-factor structure and realized GARCH setting we propose, improved the volatility modeling for each of the models discussed. The new models are called bivariate as they use both night and day volatility information and are defined to work in typical financial settings, such as volatility modeling of stock and commodity prices. We assessed the performance of the bivariate models by comparing the error functions of the forecasts of the bivariate models with those obtained when the simple versions of the models, which do not use night volatility information, were used. We call the latter models univariate models. The scope of this study was thus to analyze whether the use of night volatility information in the forms proposed improves the modeling of day volatility.

The paper proceeds as follows. Section 2 proposes the new set of bivariate realized models. Section 3 describes the data and methodology, and Section 4 summarizes the results. The paper concludes with Section 5, where final remarks are presented, and some future lines of research are proposed.

2. Bivariate Realized Models

2.1. Base Model

Existing high-frequency GARCH models estimate day conditional variance using day and intraday volatility information. We developed a class of realized models that allow constructing day volatility estimates with day, intraday, and night volatility information. Models previously proposed use return and volatility information estimated from trades that occurred during the trading day to estimate next-day volatility. However, latent volatility existing between the trading periods (called night volatility) has scarcely been considered in the day volatility estimation problem. The idea emerged from an observation on financial stock time series; prices at market closing differ from those

at market opening the following trading day, although during the night the market is closed and thus no transactions occur, so no intranight information exists. Despite the lack of night trades, latent (night) volatility still occurs, causing a price mismatch. We examined whether this latent night volatility can be modeled and whether, if incorporated into the conditional volatility modeling, it would help to provide better estimates of day volatility. Compared to other researchers that also modeled overnight returns, we proposed a two-factor structure in a realized GARCH setting with a GARCH equation that links day/night volatility to returns, night/day volatility, and intraday volatility of the previous day. This allowed us to retain the benefits of the Realized GARCH model of Hansen et al. (2012), namely, to take advantage of the natural relationship between the realized measure and the conditional day (and night for the models we proposed in the current paper) variance in an elegant structure that facilitates volatility estimation, allowed us to capture the return-volatility dependence, and was previously proved to outperform traditional GARCH. Below, we presented a method to capture this volatility and to insert it into the day conditional volatility equation.

The starting model is a reduced form Bivariate Realized GARCH model, which is a Realized GARCH model with night volatility information and exogenous realized measures, defined as follows:

$$r_t = r_t^\bullet + r_t^\circ, z_t^\circ = \frac{r_t^\circ - \mu^\circ}{\sqrt{h_t^\circ}}, z_t^\bullet = \frac{r_t^\bullet - \mu^\bullet}{\sqrt{h_t^\bullet}}, \quad (1)$$

$$\log h_t^\circ = \omega^\circ + \tau^{(\circ 1)}(z_{t-1}^\bullet) + \tau^{(\circ 2)}(z_{t-1}^\circ) + \beta^\circ \log h_{t-1}^\circ + \gamma^\circ \log x_{t-1}, \quad (2)$$

$$\log h_t^\bullet = \omega^\bullet + \tau^{(\bullet 1)}(z_{t-1}^\bullet) + \tau^{(\bullet 2)}(z_{t-1}^\circ) + \beta^\bullet \log h_{t-1}^\bullet + \gamma^\bullet \log x_{t-1}, \quad (3)$$

where \bullet denotes the night information, \circ denotes the day information of the vector, r_t is the return, $z_t \sim iid(0,1)$, $u_t \sim iid(0, \sigma_u^2)$, $h_t^\circ = var(r_t^\circ | \mathcal{F}_{t-1})$, $h_t^\bullet = var(r_t^\bullet | \mathcal{F}_{t-1})$, $\mathcal{F}_t = \sigma(r_t, x_t, r_{t-1}, x_{t-1}, \dots)$, $r_t^\circ = 100 \times (\log(price)_{t_{close}} - \log(price)_{t_{open}})$, and $r_t^\bullet = 100 \times (\log(price)_{t_{open}} - \log(price)_{t-1_{close}})$. As such, r_t is the sum between night r_t^\bullet and day r_t° returns, z_t° represents the standardized day returns, and z_t^\bullet represents the standardized night returns, whereas μ° is the means of day returns and μ^\bullet is the means of night returns. All τ 's are coefficients of the standardized returns that follow to be estimated through the maximum log-likelihood function (MLE). If marked by \circ , τ represents the coefficients of the standardized returns in the equation of conditional day volatility, and if marked by \bullet , τ represents the coefficients of the standardized returns in the equation of conditional night volatility. The numbers next to \circ or \bullet are for indexing purposes: For example, $\tau^{(\circ 1)}$ and $\tau^{(\circ 2)}$ are two coefficients of the standardized returns in the equation of conditional day volatility that follow to be estimated through MLE.

Thus, the base model is formed of three equations: The return equation, which is the sum between day (open-to-close) returns and night (close-to-open) returns, and two conditional volatility equations, as follows: The first expresses day volatility as a function of previous day (z_{t-1}°) and night (z_{t-1}^\bullet ; standardized) returns, conditional day variance (h_{t-1}°), and a realized measure of volatility (x_{t-1} ; realized kernel, high-low, realized variance, etc.). The second defines night volatility as a function of previous day (z_{t-1}°) and night (z_{t-1}^\bullet ; standardized) returns, conditional night variance (h_{t-1}^\bullet), and a realized measure of volatility (x_{t-1}). Notably, in this model (called reduced form for this reason), the realized measure is not endogenized nor linked to the day volatility measure through a measurement equation, but rather is treated as an exogenous variable. We added this equation to the complete form of the model that was documented in the next section. The realized measure was compounded from intraday prices recorded throughout the day.

2.2. Extended Models

We used the base model structure and extended its idea to a class of best-known GARCH-type models. We used this approach as all models used share the same structure and thus similar properties, which enabled us to set up a similar bivariate configuration. The aim was to construct a

group of models that takes advantage of night volatility estimation, and also defines the existing natural relationship between the realized measures and the conditional day and night variance. As such, we proposed four new realized models and one non-realized model: Bivariate Realized GARCH (1,1), with an endogenous component of realized measure and therefore a separate measurement equation, which we will call a complete version model; Bivariate Exponential GARCH-X (Bivariate EGARCH-X), that is a bivariate exponential generalized autoregressive conditional heteroskedastic model with an exogenous realized measure; Bivariate Realized EGARCH (1,1); Bivariate Realized GARCH (2,2); and Bivariate EGARCH (1,1). The detailed specifications of the bivariate models we propose are provided in Table 1.

Next, we summarized the main features of each model. All share similar return equations as in the case of the base model—the daily return r_t is the sum between open-to-close return (day return) r_t° and close-to-open return (night return) r_t^\bullet . The GARCH equations share distinct properties but they have unique features as well. All define the day (open-to-close) volatility h_t° as a function of day z_t° and night z_t^\bullet standardized returns as defined above, and also as a function of the previous day (open-to-close) volatility. Except for the Bivariate EGARCH (1,1) and the reduced form Bivariate Realized GARCH models, all other models also include the relationship between day volatility h_t° and intraday volatility x_{t-1} in the GARCH equation. Since Bivariate EGARCH (1,1) is not a realized model, it does not contain intraday information. In our Bivariate EGARCH-X model, intraday volatility x_{t-1} is treated as an exogenous variable and is thus not linked to any other variable. However, all other realized models incorporate a third equation, the measurement equation, which defines the joint dependence between r_t and x_t . x_t is thus “endogenized” by being formulated as a function of day (open-to-close) volatility, night (close-to-open) volatility, and day and night standardized returns (z_t° and z_t^\bullet , respectively).

Table 1. Summary of the bivariate realized generalized autoregressive conditional heteroskedasticity (GARCH) models proposed.

Model	Return Equations	GARCH Equations	Measurement equation
Bivariate EGARCH (1,1)	$r_t = r_t^\bullet + r_t^\circ$ $z_t^\circ = \frac{r_t^\circ - \mu^\circ}{\sqrt{h_t^\circ}}$ $z_t^\bullet = \frac{r_t^\bullet - \mu^\bullet}{\sqrt{h_t^\bullet}}$	$\log h_t^\circ = \omega^\circ + \varepsilon^{(\circ 1)} [\tau^{(\circ 1)}(z_{t-1}^\bullet) + \tau^{(\circ 2)}(z_{t-1}^\circ)]$ $+ \varepsilon^{(\circ 2)} \{ [\tau^{(\circ 1)}(z_{t-1}^\bullet) + \tau^{(\circ 2)}(z_{t-1}^\circ)]^2 - 1 \}$ $+ \beta^\circ \log h_{t-1}^\circ$ $\log h_t^\bullet = \omega^\bullet + \varepsilon^{(\bullet 1)} [\tau^{(\bullet 1)}(z_{t-1}^\bullet) + \tau^{(\bullet 2)}(z_{t-1}^\circ)]$ $+ \varepsilon^{(\bullet 2)} \{ [\tau^{(\bullet 1)}(z_{t-1}^\bullet) + \tau^{(\bullet 2)}(z_{t-1}^\circ)]^2 - 1 \}$ $+ \beta^\bullet \log h_{t-1}^\bullet$	
Bivariate EGARCH-X	$r_t = r_t^\bullet + r_t^\circ$ $z_t^\circ = \frac{r_t^\circ - \mu^\circ}{\sqrt{h_t^\circ}}$ $z_t^\bullet = \frac{r_t^\bullet - \mu^\bullet}{\sqrt{h_t^\bullet}}$	$\log h_t^\circ = \omega^\circ + \varepsilon^{(\circ 1)} [\tau^{(\circ 1)}(z_{t-1}^\bullet) + \tau^{(\circ 2)}(z_{t-1}^\circ)]$ $+ \varepsilon^{(\circ 2)} \{ [\tau^{(\circ 1)}(z_{t-1}^\bullet) + \tau^{(\circ 2)}(z_{t-1}^\circ)]^2 - 1 \}$ $+ \beta^\circ \log h_{t-1}^\circ + \gamma^\circ \log x_{t-1}$ $\log h_t^\bullet = \omega^\bullet + \varepsilon^{(\bullet 1)} [\tau^{(\bullet 1)}(z_{t-1}^\bullet) + \tau^{(\bullet 2)}(z_{t-1}^\circ)]$ $+ \varepsilon^{(\bullet 2)} \{ [\tau^{(\bullet 1)}(z_{t-1}^\bullet) + \tau^{(\bullet 2)}(z_{t-1}^\circ)]^2 - 1 \}$ $+ \beta^\bullet \log h_{t-1}^\bullet + \gamma^\bullet \log x_{t-1}$	

Bivariate Realized EGARCH (1,1)	$r_t = r_t^\bullet + r_t^\circ$	$\log h_t^\circ = \omega^\circ + \varepsilon^{(\circ 1)} [\tau^{(\circ 1)}(z_{t-1}^\bullet) + \tau^{(\circ 2)}(z_{t-1}^\circ)]$	$\log x_t$
	$z_t^\circ = \frac{r_t^\circ - \mu^\circ}{\sqrt{h_t^\circ}}$	$+ \varepsilon^{(\circ 2)} \{ [\tau^{(\circ 1)}(z_{t-1}^\bullet) + \tau^{(\circ 2)}(z_{t-1}^\circ)]^2 - 1 \}$	$= \xi + \phi \log h_t^\circ$
Bivariate Realized GARCH (1,1) complete form	$r_t = r_t^\bullet + r_t^\circ$	$\log h_t^\bullet = \omega^\bullet + \varepsilon^{(\bullet 1)} [\tau^{(\bullet 1)}(z_{t-1}^\bullet) + \tau^{(\bullet 2)}(z_{t-1}^\circ)]$	$= \xi + \phi \log h_t^\bullet$
	$z_t^\bullet = \frac{r_t^\bullet - \mu^\bullet}{\sqrt{h_t^\bullet}}$	$+ \varepsilon^{(\bullet 2)} \{ [\tau^{(\bullet 1)}(z_{t-1}^\bullet) + \tau^{(\bullet 2)}(z_{t-1}^\circ)]^2 - 1 \}$	$+ \vartheta \log h_t^\bullet + \delta^{(\circ 1)} z_t^\bullet$
		$+ \beta^\bullet \log h_{t-1}^\bullet + \gamma^\bullet \log x_{t-1}$	$+ \delta^{(\circ 2)} z_t^\circ + u_t$
Bivariate Realized GARCH (1,1) complete form	$r_t = r_t^\bullet + r_t^\circ$	$\log h_t^\circ = \omega^\circ + \tau^{(\circ 1)}(z_{t-1}^\bullet) + \tau^{(\circ 2)}(z_{t-1}^\circ) + \beta^\circ \log h_{t-1}^\circ$	$\log x_t$
	$z_t^\circ = \frac{r_t^\circ - \mu^\circ}{\sqrt{h_t^\circ}}$	$+ \gamma^\circ \log x_{t-1}$	$= \xi + \phi \log h_t^\circ$
Bivariate Realized GARCH (2,2)	$r_t = r_t^\bullet + r_t^\circ$	$\log h_t^\bullet = \omega^\bullet + \tau^{(\bullet 1)}(z_{t-1}^\bullet) + \tau^{(\bullet 2)}(z_{t-1}^\circ) + \beta^\bullet \log h_{t-1}^\bullet$	$+ \vartheta \log h_t^\bullet + \delta^{(\circ 1)} z_t^\bullet$
	$z_t^\bullet = \frac{r_t^\bullet - \mu^\bullet}{\sqrt{h_t^\bullet}}$	$+ \gamma^\bullet \log x_{t-1}$	$+ \delta^{(\circ 2)} z_t^\circ + u_t$
Bivariate Realized GARCH (2,2)	$r_t = r_t^\bullet + r_t^\circ$	$\log h_t^\circ = \omega^\circ + \tau^{(\circ 1)}(z_{t-1}^\bullet) + \tau^{(\circ 2)}(z_{t-1}^\circ)$	$\log x_t$
	$z_t^\circ = \frac{r_t^\circ - \mu^\circ}{\sqrt{h_t^\circ}}$	$+ \alpha^\circ \log(\max(r_{t-1}^{\circ 2}, 10^{-20}))$	$= \xi + \phi \log h_t^\circ$
Bivariate Realized GARCH (2,2)	$r_t = r_t^\bullet + r_t^\circ$	$+ \beta^{(\circ 1)} \log h_{t-1}^\circ + \beta^{(\circ 2)} \log h_{t-2}^\circ$	$+ \vartheta \log h_t^\circ$
	$z_t^\circ = \frac{r_t^\circ - \mu^\circ}{\sqrt{h_t^\circ}}$	$+ \gamma^{(\circ 1)} \log x_{t-1} + \gamma^{(\circ 2)} \log x_{t-2}$	$+ \varepsilon_1 [\delta^{(\circ 1)} z_t^\bullet$
Bivariate Realized GARCH (2,2)	$r_t = r_t^\bullet + r_t^\circ$	$\log h_t^\bullet = \omega^\bullet + \tau^{(\bullet 1)}(z_{t-1}^\bullet) + \tau^{(\bullet 2)}(z_{t-1}^\circ)$	$+ \delta^{(\circ 2)} z_t^\circ]$
	$z_t^\bullet = \frac{r_t^\bullet - \mu^\bullet}{\sqrt{h_t^\bullet}}$	$+ \alpha^\bullet \log(\max(r_{t-1}^{\bullet 2}, 10^{-20}))$	$+ \varepsilon_2 \{ [\delta^{(\circ 1)} z_t^\bullet$
Bivariate Realized GARCH (2,2)	$r_t = r_t^\bullet + r_t^\circ$	$+ \beta^{(\bullet 1)} \log h_{t-1}^\bullet + \beta^{(\bullet 2)} \log h_{t-2}^\bullet$	$+ \delta^{(\circ 2)} z_t^\circ]^2 - 1 \}$
	$z_t^\bullet = \frac{r_t^\bullet - \mu^\bullet}{\sqrt{h_t^\bullet}}$	$+ \gamma^{(\bullet 1)} \log x_{t-1} + \gamma^{(\bullet 2)} \log x_{t-2}$	$+ u_t$

3. Data and Estimation Methodology

We used tick data sampled along 3537 trading days during the period of 30 August 2004–31 December 2018, corresponding to 10 stocks: AIG (American International Group, Inc.), AXP (American Express Company), BAC (Bank of America Corporation), CSCO (Cisco Systems, Inc.), F (Ford Motor Company (F)), GE (General Electric Company), INTC (Intel Corporation), JPM (JPMorgan Chase & Co.), MSFT (Microsoft Corporation), and T (AT&T Inc.). To avoid the outliers that would result from quiet days, the half trading days around the Christmas and Thanksgiving holidays were removed.

We opted for estimating intraday volatility by compounding realized kernels instead of the more widely used realized variance, as it is generally acknowledged that squared daily returns provide a poor estimation of actual intraday volatility. Realized kernels are robust for microstructure errors or frictions, which are known to cause endogenous and dependent noise terms. They are used to estimate the quadratic variation in an efficient price process when the time stamps in every day do not match (non-synchronous, with irregularly spaced observations) and when the high-frequency time series described by the prices are noisy with many microstructure effects. We compounded the realized kernels as measures of intraday volatility (x_t) using the methodology of Barndorff-Nielsen et al. (2009, 2011). The framework is given by Y , a variable that is the sum of a Brownian semi-martingale and a jump process, as follows:

$$Y_t = \int_0^t a_u du + \int_0^t \sigma_u dW_u + J_t. \quad (4)$$

For the purpose of our exercise, we need to find the quadratic variation of Y , $[Y] = \int_0^T \sigma_u^2 du + \sum_{i=1}^{N_T} C_i^2$. Barndorff-Nielsen et al. (2009, 2011) estimated it from the noisy discrete observations X_{τ_j} of Y_{τ_j} , $0 = \tau_0 < \tau_1 < \dots < \tau_n = T$, where $X_{\tau_j} = Y_{\tau_j} + U_{\tau_j}$ and U_{τ_j} represents the

market microstructure effects (noise). Barndorff-Nielsen et al. (2009, 2011) estimated this quadratic variation by proposing realized kernels, a non-negative estimator that is constructed as follows.

The first challenge with the tick data is the non-synchronicity. Non-synchronous trading occurs when the trades or quotes appear at irregularly spaced times across stocks, which is usually the case in stock markets, especially those with low liquidity or stale prices. Barndorff-Nielsen et al. (2011) solved this by suggesting a refresh time when all the stocks are traded. We implemented the same method by recording the prices only when (and immediately after) all of them were traded.

To eliminate start and end effects and their associated errors, which are averaged through this procedure, we proceeded to jittering (averaging) the first and last two prices, as also suggested by Barndorff-Nielsen et al. (2011). Having synchronized and constructed the time series by jittering at the initial and final time points, we defined the semi-definite realized kernels, as follows, according to Barndorff-Nielsen et al. (2009, 2011):

$$K(X) = \sum_{h=-H}^H k\left(\frac{h}{H+1}\right) \gamma_h, \text{ where } \gamma_h = \sum_{j=|h|+1}^n x_j x_{j-|h|}, \quad (5)$$

where $k(x)$ is a kernel weight function that has the $k(0) = 1, k'(0) = 0$ property, and k is twice differentiable with continuous derivatives.

Barndorff-Nielsen et al. (2009) used a Parzen kernel as it satisfies the smoothness conditions through $k'(0) = k'(1) = 0$, and its estimates are positive. We made the same choice, and used the same Parzen kernel function:

$$k(x) = \begin{cases} 1 - 6x^2 + 6x^3, & 0 \leq x \leq 1/2 \\ 2(1-x)^3, & 1/2 \leq x \leq 1 \\ 0, & x > 1 \end{cases} \quad (6)$$

The optimal choice of bandwidth, according to Barndorff-Nielsen et al. (2009), which we chose to use, is $H^* = c^* \xi^{4/5} n^{3/5}$, with $c^* = \left\{ \frac{k''(0)^2}{k_{*,0}^4} \right\}^{1/5}$ and $\xi^2 = \frac{\omega^2}{\sqrt{T} \int_0^T \sigma_u^4 du}$, where $c^* = ((12)^2)^{1/5} = 3.5134$ for

the Parzen kernel. $\int_0^T \sigma_u^4 du$ is called the integrated quarticity, and, in our empirical exercise, it equals RV_{sparse} . This denotes a subsampled realized variance based on 20-min returns. By calculating 1200 realized variances by shifting the first observation recorded time in 1-s increments, we obtained a number of realized variance estimators. We averaged them and obtained RV_{sparse} . ω^2 was estimated by calculating the realized variance using every i th trade. We varied the starting point, and thereby produced i realized variances, namely RV_{dense}^i . Thus, our ω^2 estimator was calculated as:

$$\hat{\omega}_{(j)}^2 = \frac{RV_{dense}^{(j)}}{2n_{(j)}}, j = 1, \dots, i, \quad (7)$$

where $n_{(j)}$ is the number of non-zero returns used to estimate $RV_{dense}^{(j)}$. The estimate of ω^2 is then the average of the j estimates,

$$\hat{\omega}^2 = \frac{1}{i} \sum_{j=1}^i \hat{\omega}_{(j)}^2. \quad (8)$$

By design, the realized kernel is positive semi-definite and the rate of convergence is $n^{1/5}$.

We estimated the in-sample and out-of-sample (3000th day in the sample, 24 November 2016, the cutoff point) in both the univariate and bivariate models with respect to each of the 10 stocks. The univariate models considered are the standard realized versions of the GARCH model (Realized GARCH, Realized EGARCH, EGARCH-X, and Realized GARCH (2,2)), as well as the EGARCH model. The estimated bivariate models are those mentioned in Section 2 (Bivariate EGARCH, reduced and complete forms of Bivariate Realized GARCH, Bivariate Realized EGARCH, Bivariate EGARCH-X, and Bivariate Realized GARCH (2,2)).

The estimation was performed by maximizing the total log-likelihood functions (MLE), namely the sum of partial log-likelihood functions for the returns and for the intraday measures; the ranking criterion with respect to the MLE was the partial log-likelihood function for returns solely. We used

MLE to estimate both the proposed bivariate models and a number of univariate models that do not include night volatility information.

The log-likelihood function used in the estimation of the above models takes the form $l(r_t^\bullet, r_t^\circ, x_t) = L_1$ for Bivariate EGARCH and Bivariate EGARCH-X, or $l(r_t^\bullet, r_t^\circ, x_t) = L_1 + L_2$ for Bivariate Realized GARCH complete version, Bivariate Realized EGARCH (1,1), and Bivariate Realized GARCH (2,2) (Appendix A), where $L_1 = -\frac{1}{2} \sum_{t=1}^n \left\{ 2 \log(2\pi) + \log(1 - \rho^2) + \log h_t^\bullet + \log h_t^\circ + \frac{(r_t^\bullet - \mu^\bullet)^2/h_t^\bullet + (r_t^\circ - \mu^\circ)^2/h_t^\circ}{(1 - \rho^2)} - \frac{2\rho}{(1 - \rho^2)} \frac{(r_t^\bullet - \mu^\bullet)(r_t^\circ - \mu^\circ)}{\sqrt{h_t^\bullet h_t^\circ}} \right\}$ and $L_2 = -\frac{1}{2} \sum_{t=1}^n \{ \log(2\pi) + \log(\sigma_u^2) + u_t^2/\sigma_u^2 \}$.

To evaluate whether introducing night volatility estimations in models' equations improves the day volatility estimation, we calculated two loss functions, root mean squared error (RMSE) and mean absolute error (MAE). Based on these, we documented the number of models for each in-sample and out-of-sample estimation for each of the 10 stocks, at which MAE and RMSE were smaller. This allowed us to draw conclusions about the better performance of the bivariate or univariate models. Based on the size of the loss functions obtained at each estimation, we analyzed the performance of the new models that included night volatility estimates. This contributed to our objective by documenting whether or not night volatility information improves the estimation of day volatility with respect to the main GARCH-type of models proposed in the literature.

The maximized log-likelihood functions in univariate and bivariate estimations are provided in Tables A1 and A2 in Appendix B. As the log-likelihood functions of the bivariate models differ from those of the univariate versions (for the bivariate estimation, we maximized a bi-dimensional vector $\begin{pmatrix} r_t^\bullet \\ r_t^\circ \end{pmatrix}$ with a non-null correlation factor (ρ) between its subvectors), it makes little sense to compare the values of the MLEs across the univariate and bivariate models to document an improvement or loss of performance when introducing night volatility estimates. Specifically, the log-likelihood function for the bivariate models is: $\log l(r_t^\bullet, r_t^\circ) = -\frac{1}{2} \sum_{t=1}^n \left\{ 2 \log(2\pi) + \log(1 - \rho^2) + \log(h_t^\bullet) + \log(h_t^\circ) + \frac{r_t^{\bullet 2}/h_t^\bullet + r_t^{\circ 2}/h_t^\circ}{(1 - \rho^2)} - \frac{2\rho}{(1 - \rho^2)} \frac{r_t^\bullet r_t^\circ}{\sqrt{h_t^\bullet h_t^\circ}} \right\}$, where $\rho = \text{corr}(r_t^\circ, r_t^\bullet)$. In the univariate models' case, the log-

likelihood function is $\log l(r_t) = -\frac{1}{2} \sum_{t=1}^n \left[\log(2\pi) + \log(h_t) + \frac{(r_t - \mu)^2}{h_t} \right]$ for EGARCH and EGARCH-X, and $\log l(r_t) = -\frac{1}{2} \sum_{t=1}^n \left[2 \log(2\pi) + \log(h_t) + \frac{(r_t - \mu)^2}{h_t} + \log(\sigma_u^2) + \frac{u_t^2}{\sigma_u^2} \right]$ for Realized EGARCH, Realized GARCH, and Realized GARCH (2,2). As such, we could not use this method to evaluate the performance of the bivariate models, as we would be comparing the values of estimations of different functions.

Thus, for the purpose of documenting the gain or loss in accuracy, we used the standard method in econometrics for evaluating the models' performance—that of calculating two loss functions (RMSE and MAE)—which would better assess whether adding night volatility information with a two-factor structure in a realized GARCH setting improves estimations of next-day volatility.

4. Results

The standard method used in econometrics to evaluate models' performance is to calculate the size of the loss functions, among which RMSE and MAE are the most common and reliable. We calculated them for both in-sample and out-of-sample estimations, and our results indicate an improvement when night volatility estimations were included in the equations of the day conditional volatility in almost every case.

We worked with a number of models that have different features and for which adding an estimation of night volatility may contribute to the volatility estimation. For example, by inspecting

the results for RMSE (in-sample estimation) in Table 2, the improvement was evident for 55 out of 60 cases (1 loss function result \times 6 models evaluated \times 10 stocks). The cases in which the improvement could not be documented are marked with red (for RMSE) or green (for MAE) numbers in Table 2. In the five cases in which this was not evident, four of them were for Realized GARCH (2,2). This means that Realized GARCH (2,2) only shows some features that did not work better when the night volatility estimates were considered given the way in which the model was designed. This may be because, compared to the other models that model next-day volatility by only using information from the previous day and night, Realized GARCH (2,2) uses information on the previous night volatility as well as information on returns and volatility of the previous two days. We thought that this might be the problem with this model, but it would need to be proven empirically; we left this question for future work.

This conclusion was strengthened by examining the MAE results. When considering MAE as an evaluation tool, the bivariate models produced superior forecasting ability in 59 out of 60 cases, indicating an improvement for the models that included night volatility estimation in the day volatility modeling. However, in only one case out of 60 was the improvement not evident, for the same Realized GARCH (2,2) model. As such, the model itself appears to be problematic, not the evaluation we performed. As mentioned above, we thought that the problem with this model was that it models conditional day volatility by including in the model information on day volatility and returns from the previous two days, instead of one day only as we did for the other models. In Bivariate Realized GARCH (2,2), we considered only one-night volatility information instead of considering the night volatility estimation from the previous two nights.

Table 2. Loss functions in univariate and bivariate estimations; in-sample.

Stock		EGARCH		EGARCH-X		Realized EGARCH		Realized GARCH				Realized GARCH (2,2)	
		Univ	Biv	Univ	Biv	Univ	Biv	Univ	Biv (com)	Univ	Biv (red)	Univ	Biv
AIG	RMSE	203.3	188.6	203.7	195.1	189.8	195.8	254.0	218.2	254.0	219.7	190.6	250.2
	MAE	18.0	15.1	20.1	16.5	17.2	15.2	22.9	21.0	22.9	21.1	17.4	21.4
AXP	RMSE	6.9	6.3	6.7	6.2	6.8	5.5	6.8	5.5	7.0	5.2	7.0	7.1
	MAE	3.1	2.3	3.1	2.6	3.1	2.2	3.2	2.0	3.3	1.9	3.0	2.3
BAC	RMSE	16.7	15.9	16.3	15.4	15.9	15.1	16.2	15.7	16.4	15.3	15.9	15.1
	MAE	4.5	3.5	4.0	3.2	4.1	2.8	4.3	3.0	3.8	3.1	4.2	2.7
CSCO	RMSE	6.5	5.6	6.6	6.4	6.8	5.7	6.4	5.5	6.8	5.9	6.8	6.6
	MAE	3.1	2.2	3.1	2.4	3.3	2.5	3.0	2.2	3.0	1.9	3.2	2.3
F	RMSE	16.9	16.3	16.4	15.1	16.3	15.5	16.1	15.7	16.5	15.3	16.2	14.8
	MAE	4.0	3.2	3.8	3.3	4.2	3.3	4.3	3.3	4.2	2.9	4.0	3.0
GE	RMSE	6.8	6.0	6.9	6.5	6.6	5.8	6.9	5.5	6.7	5.5	6.4	7.0
	MAE	3.3	1.9	2.7	2.2	3.1	1.8	3.2	2.2	3.2	1.8	2.8	2.4
INTC	RMSE	16.6	15.9	16.3	15.4	16.4	15.3	16.5	15.0	16.5	15.1	16.1	15.2
	MAE	4.4	3.1	4.0	3.5	4.3	3.1	4.2	3.0	3.7	3.2	3.9	2.7
JPM	RMSE	11.4	10.4	11.1	9.7	10.7	10.6	11.1	9.9	10.8	10.1	11.1	9.9
	MAE	3.9	2.6	3.5	2.7	3.5	2.7	3.3	3.0	3.6	2.2	3.6	2.7
MSFT	RMSE	6.8	6.4	6.6	6.1	6.8	5.6	6.9	5.6	6.8	5.5	6.6	7.1
	MAE	3.3	1.8	2.8	2.5	3.1	2.1	3.3	2.0	3.4	2.1	3.4	2.9
T	RMSE	16.8	15.9	16.5	15.2	16.7	15.2	16.4	15.4	16.6	15.7	16.2	14.9
	MAE	4.3	3.5	4.2	3.4	4.0	3.0	4.1	3.0	4.5	3.5	4.5	3.1

Univ and *Biv* stand for *Univariate* and *Bivariate*, respectively, while *com* and *red* stand for *complete* and *reduced*, respectively. Red and green numbers indicate the stances in which bivariate models perform worse than the univariate ones (when evaluated according to RMSE or MAE, respectively).

When examining the results for the out-of-sample estimations in Table 3, we found that of 60 evaluations with RMSE, 53 showed forecasting improvement when night volatility information was used. In the seven cases in which the improvement was not evident, three were recorded for the same Realized GARCH (2,2) model. The remaining four belonged to various other models, one for each. However, we observed another pattern. Most of the failures in documenting an improvement were for the same stock: AIG. This suggests that the results were sensitive not only to the model (as we explained earlier with the way in which Realized GARCH (2,2) was built), but were also sensitive to the stock choice. Since AIG persistently failed in showing an improvement when using night volatility information, AIG price recordings should be more carefully examined to understand what makes it less sensitive to this modeling suggestion, including examining the amount of the stock price differential (the difference between the market closing and the market opening prices), and also understanding the roots of the volatility transmission for this stock in particular. Again, we left this as exploratory work for the future paper. When ranked according to MAE, 58 results out of 60 indicated improvement, whereas only two cases (among them, one for Realized GARCH (2,2)) did not. Again, both estimations indicated strong evidence in favor of including night volatility estimation in the modeling problem of day volatility.

Counting the number of cases that fail to show improvement is valuable for two reasons: (1) It is the best tool when comparing models evaluated through MLE given that the log-likelihood functions were not similar for looking at the size of the MLE values; and (2) the cases in which we failed to see improvement indicated some consistency for a specific model and a specific stock. This opens the opportunity for future work in which we might try to understand why the Realized GARCH (2,2) model and AIG stock persistently indicated less evidence compared with other models and stocks, where by adding night volatility information, we produced improved volatility estimation.

Table 3. Loss functions in univariate and bivariate estimations; out-of-sample.

Stock		EGARCH		EGARCH-X		Realized EGARCH		Realized GARCH				Realized GARCH (2,2)	
		Univ	Biv	Univ	Biv	Univ	Biv	Univ	Biv (com)	Univ	Biv (red)	Univ	Biv
AIG	RMSE	565.5	574.3	552.0	532.9	544.0	566.4	552.9	572.4	552.9	573.3	538.9	584.7
	MAE	109.0	100.4	106.9	104.1	103.8	102.5	122.0	103.1	122.1	103.0	104.6	121.1
AXP	RMSE	14.3	14.0	14.5	13.3	14.2	12.8	14.0	13.0	14.0	12.6	13.9	12.8
	MAE	8.8	8.1	9.0	9.1	8.6	7.8	8.7	7.7	8.7	8.1	9.0	7.7
BAC	RMSE	44.0	43.1	43.2	42.1	44.5	42.6	43.5	42.1	43.3	42.6	43.3	43.1
	MAE	19.4	17.7	18.4	17.6	18.0	17.2	18.3	17.8	18.6	17.2	18.0	17.4
CSCO	RMSE	14.2	14.2	14.4	13.1	14.3	13.2	13.9	12.8	14.2	13.1	13.8	12.7
	MAE	8.9	8.0	9.0	8.9	8.7	7.8	9.0	8.1	9.0	8.0	9.2	8.1
F	RMSE	44.0	43.0	43.0	42.4	44.3	42.3	43.4	42.6	43.5	42.0	43.3	43.4
	MAE	19.1	18.1	18.6	17.4	18.6	16.7	18.5	17.9	18.3	17.0	18.6	17.0
GE	RMSE	14.7	14.4	14.3	13.3	14.0	13.3	13.9	12.7	14.3	12.6	13.9	13.0
	MAE	9.5	7.6	8.8	8.7	8.6	7.8	8.8	8.2	8.4	8.0	9.4	7.8
INTC	RMSE	44.4	43.3	43.2	42.2	44.4	42.0	43.7	42.2	43.4	41.9	43.4	43.3
	MAE	19.6	18.1	18.5	17.1	18.2	17.0	18.4	18.0	18.3	17.1	18.3	17.3
JPM	RMSE	26.7	26.1	26.1	25.3	25.4	24.0	25.2	24.8	25.5	24.6	25.4	25.1
	MAE	12.5	11.9	12.9	11.6	12.1	10.9	12.1	11.7	12.4	11.9	12.3	11.3
MSFT	RMSE	14.6	14.3	14.0	13.5	13.8	13.2	13.7	12.7	14.3	12.7	14.0	12.3
	MAE	8.7	8.0	9.2	9.0	8.6	7.6	8.5	7.6	8.8	8.1	9.0	7.7
T	RMSE	43.8	42.9	43.1	42.3	43.8	42.5	43.3	42.1	43.4	42.0	43.2	43.4
	MAE	19.6	17.9	18.4	17.5	18.0	16.7	18.3	17.6	18.6	17.0	18.3	17.6

Red and green numbers indicate the stances in which bivariate models perform worse than the univariate ones (when evaluated according to RMSE or MAE, respectively).

Thus, we concluded that the proposed bivariate models improved the forecasting performance compared with the univariate models; as such, adding night volatility estimations according to the methodology suggested improves next-day volatility estimates.

5. Conclusions

This paper provided a methodology that captures and integrates night volatility into the modeling of day volatility. In univariate context, this method led to formulating four bivariate realized GARCH models (Bivariate EGARCH-X, Bivariate Realized GARCH, Bivariate Realized GARCH (2,2), and Bivariate Realized EGARCH) and one bivariate non-realized model (Bivariate EGARCH). The novelty of this method is the incorporation of a night measure of volatility into the models, computed from price changes between the closing and opening of the trading market with a two-factor structure of the conditional variance in a realized GARCH setting that takes advantage of the natural relationship between the realized measure and the conditional variance. This captures the leverage effect and maintains an elegant mathematical structure that facilitates the estimation of volatility.

With respect to assessing forecasting performance, the first finding was that rankings were sensitive to the stock and model choice but displayed little sensitivity to the ranking criterion and estimation methodology. However, the bivariate models were proved to perform better in most instances, compared with the univariate models. As such, we concluded that by adding night volatility estimates in the volatility models according to the methodology described, better estimates of next-day volatility could be obtained. This represents a step further from including high-frequency data in the modeling problem of the GARCH models in that estimates of night volatility are added into the equation of the day conditional variance according to the novel methodology we suggest.

The assessment to multivariate assets (e.g., portfolios of stocks) could be extended in future work by documenting a method of forecasting volatility of assets using the principal component (PC) analysis or other statistical procedures that use the orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables, taking advantage of the autoregressive conditional heteroskedastic models we proposed that use estimates of day, intraday, and night volatility. We might refer to these models as PC Bivariate Realized GARCH models and these might be used to formulate the general form of one multivariate asset's conditional variance-covariance matrix expressed in terms of conditional variances of the compounding assets and of their principal components. This would allow the estimation of the volatility of one multivariate asset through estimations of the volatility of principal components using day, intraday, and night volatility information. Then, by reducing the n -multivariate to a $n - k$ stock dimension (n and k positive integers), we could estimate the new models and assess their one-day-ahead forecasting performance. Constructing models that use volatility information from the previous two days and two nights may further improve the modeling of volatility, as we noted by inspecting the results for the current bivariate form of Realized GARCH (2,2). Disseminating among the stocks according to their underlying volatility features may provide a better method of more consistently modeling their volatility patterns.

Integration of volatility estimates of highly interlinked markets that are open during the closing time of the reference market is another suggestion for further research. For example, proposing models for the U.S. market that estimate day volatility using night volatility estimates from the Asian markets open during the non-trading times of the U.S. market would allow for integration in such models of systemic risk and financial contagion related elements, with likely benefits for volatility estimation and forecasting.

Author Contributions: Individual contributions to the current paper were as follows: Conceptualization, M.M., X.R., and N.A.; Methodology, M.M.; Software, M.M.; Validation, M.M.; Formal Analysis, M.M.; Investigation,

M.M.; Resources, M.M, X.R., and N.A.; Data Curation, M.M.; Writing-Original Draft Preparation, M.M.; Writing-Review & Editing, M.M.; Visualization, M.M.; Supervision, M.M., X.R., and N.A.; Project Administration, M. M., X.R., and N.A.; Funding Acquisition, M.M., X.R., and N.A.

Funding: M.M. acknowledges support from the Agency for Management of University and Research Grants (AGAUR) of the Government of Catalonia (Resolution IUE/2681/2008 of 8 August, DOGC no. 5208 of 03.09.08), ESADE Business School (Ramon Llull University), as well as from the University of Tasmania (ARC DP130100168).

Acknowledgments: We acknowledge the contribution of Peter Reinhard Hansen, Zhuo (Albert) Huang and Howard Howan Shek to the definition of the reduced form Bivariate Realized GARCH model described by equations (1) to (3). We are grateful for the comments from Mardi Dungey and participants in the European Economic Association and Econometric Society Meeting 2012, Econometric Society Australasian Meeting 2013, and FIRN 2012 conference.

Disclaimer: The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bank of Romania.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Log-Likelihood Function for the Bivariate Models

The data are bivariate vectors compounded of two univariate vectors that refer to uncorrelated sets of information (we considered first that night volatility was uncorrelated with day volatility): $\begin{pmatrix} r_t^\circ \\ r_t^\bullet \end{pmatrix} | F_{t-1} \sim N(0, \begin{pmatrix} h_t^\circ & 0 \\ 0 & h_t^\bullet \end{pmatrix})$. Accordingly, the random vector $\begin{pmatrix} r_t^\circ \\ r_t^\bullet \end{pmatrix}$ depends solely on the information set available at time $t - 1$, and has a normal distribution with $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ mean and a variance equal to the variance–covariance matrix $\begin{pmatrix} h_t^\circ & 0 \\ 0 & h_t^\bullet \end{pmatrix}$. The latter is equivalent to $var(r_t^\circ) = \sigma_t^\circ$, $var(r_t^\bullet) = h_t^\bullet$ and $cov(r_t^\circ, r_t^\bullet) = 0$. The total volatility is given as $r_t = r_t^\circ + r_t^\bullet$. Theory states that when a random vector (such as $\begin{pmatrix} r_t^\circ \\ r_t^\bullet \end{pmatrix}$) is normally distributed, then its components are also normal. $\begin{pmatrix} r_t^\circ \\ r_t^\bullet \end{pmatrix} | F_{t-1} \sim N(0, \begin{pmatrix} h_t^\circ & 0 \\ 0 & h_t^\bullet \end{pmatrix})$ shows that $r_t^\bullet | F_{t-1} \sim N(0, h_t^\bullet)$ and $r_t^\circ | F_{t-1} \sim N(0, h_t^\circ)$. Since a sum of two normal variables is a normal variable with the average equal to the arithmetic sum of the two component averages, $r_t | F_{t-1} \sim N(0, h_t^\circ + h_t^\bullet)$, then the density function of $r_t | F_{t-1}$ has the form of a normal variable, that is, $f(r_t) = \frac{1}{\sqrt{\sigma_t^* \sqrt{2\pi}}} e^{-\frac{r_t^2}{2h_t^*}}$, where $h_t^* = h_t^\circ + h_t^\bullet$ is the variance of r_t . Since n observations of $t = 1, \dots, n$ are

made, the likelihood function is the $\begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix}$ vector's density, and r_1, \dots, r_n are independent of each other,

so the likelihood function is $l(r_t) = \prod_{t=1}^n f(r_t) = \prod_{t=1}^n \frac{1}{\sqrt{h_t^* \sqrt{2\pi}}} e^{-\frac{r_t^2}{2h_t^*}} = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\prod_{t=1}^n \frac{1}{\sqrt{h_t^*}}\right)^{-\frac{1}{2} \sum_{t=1}^n \frac{r_t^2}{h_t^*}}$. Taking

the log of this expression and using the logarithm properties, the log-likelihood function of the total

returns r_t will become $\log l(r_t) = \log(\prod_{t=1}^n (\frac{1}{\sqrt{2\pi}})^n) + \log(\prod_{t=1}^n \frac{1}{\sqrt{h_t^*}}) - \frac{1}{2} \sum_{t=1}^n \frac{r_t^2}{h_t^*} = -\frac{n}{2} \log(2\pi) -$

$$\frac{1}{2} \sum_{t=1}^n \log(h_t^*) - \frac{1}{2} \sum_{t=1}^n \frac{r_t^2}{h_t^*} = -\frac{1}{2} \sum_{t=1}^n \left[\log(2\pi) + \log(h_t^*) + \frac{r_t^2}{h_t^*} \right].$$

If we considered a more complete model with a non-null correlation between r_t° and r_t^\bullet (meaning that night volatility influences day volatility), that is, $corr(r_t^\circ, r_t^\bullet) = \rho \neq 0$, the formulation of the log-likelihood function slightly changes. Observe first that ρ does not depend on t , that is, the correlation is not time dependent. Then, the covariance will be $(r_t^\circ, r_t^\bullet) = corr(r_t^\circ, r_t^\bullet) \sqrt{var(r_t^\circ) var(r_t^\bullet)} = \rho \sqrt{h_t^\circ h_t^\bullet}$. This means that in the new model (with a non-null correlation), the variance–covariance matrix takes

the form $\begin{pmatrix} h_t^\circ & \rho \sqrt{h_t^\circ h_t^\bullet} \\ \rho \sqrt{h_t^\circ h_t^\bullet} & h_t^\bullet \end{pmatrix}$, having the variances of r_t° and r_t^\bullet on the first diagonal, and the

covariance between r_t° and r_t^\bullet on the second diagonal, that is $cov(r_t^\circ, r_t^\bullet)$ (since $cov(r_t^\circ, r_t^\bullet) = cov(r_t^\bullet, r_t^\circ)$). As such, the $|F_{t-1}$ conditioned distribution of the $\begin{pmatrix} r_t^\bullet \\ r_t^\circ \end{pmatrix}$ vector is $\begin{pmatrix} r_t^\bullet \\ r_t^\circ \end{pmatrix} | F_{t-1} \sim N(0, \begin{pmatrix} h_t^\bullet & \rho\sqrt{h_t^\bullet h_t^\circ} \\ \rho\sqrt{h_t^\bullet h_t^\circ} & h_t^\circ \end{pmatrix})$. The conditional variance of r_t , $var(r_t | F_{t-1})$, is $var(r_t | F_{t-1}) = var(r_t^\bullet | F_{t-1}) + var(r_t^\circ | F_{t-1}) + 2cov(r_t^\bullet | F_{t-1}, r_t^\circ | F_{t-1}) = h_t^\bullet + h_t^\circ + 2\rho\sqrt{h_t^\bullet h_t^\circ}$, that is, $h_t^* = h_t^\bullet + h_t^\circ + 2\rho\sqrt{h_t^\bullet h_t^\circ}$. The log-likelihood function of $r_t = r_t^\bullet + r_t^\circ$ will be the same as the one iterated for the null correlation case, the only difference being that the variance encloses the correlation term $h_t^* = h_t^\bullet + h_t^\circ + 2\rho\sqrt{h_t^\bullet h_t^\circ}$.

However, we want to consider the log-likelihood function of the bivariate vector $\begin{pmatrix} r_t^\bullet \\ r_t^\circ \end{pmatrix}$ and not that of the univariate vector $r_t = r_t^\bullet + r_t^\circ$. As such, to define the new log-likelihood function, we considered the density function of the bi-dimensional normal $\begin{pmatrix} r_t^\bullet \\ r_t^\circ \end{pmatrix}$. The general form of a p -dimensional normal vector $N_p(\mu, \Sigma)$ (a matrix with μ vector average and Σ variance-covariance matrix) takes the form $f(x) = \frac{1}{(\sqrt{2\pi})^p} \frac{1}{\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}$, where x is any vector for which the density function has been calculated with p arguments, $\det(\Sigma)$ is the determinant of the variance-covariance matrix Σ , and $(x-\mu)'\Sigma^{-1}(x-\mu)$ is the matrix product between the transpose of the $(x-\mu)$ vector, the inverse of matrix Σ , and the $(x-\mu)$ vector. As such, with $p = 2$ for the particular case of a bi-dimensional vector $\begin{pmatrix} r_t^\bullet \\ r_t^\circ \end{pmatrix}$, the density function is $f(r_t^\bullet, r_t^\circ) = \frac{1}{(\sqrt{2\pi})^2} \frac{1}{\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(r_t^\bullet, r_t^\circ)\Sigma^{-1}\begin{pmatrix} r_t^\bullet \\ r_t^\circ \end{pmatrix}}$ in which $\mu = 0$ and $\Sigma = \begin{pmatrix} h_t^\bullet & \rho\sqrt{h_t^\bullet h_t^\circ} \\ \rho\sqrt{h_t^\bullet h_t^\circ} & h_t^\circ \end{pmatrix}$. Since $\det(\Sigma) = h_t^\bullet h_t^\circ - \rho^2 h_t^\bullet h_t^\circ = h_t^\bullet h_t^\circ (1 - \rho^2)$, then its log form is $\log(\det(\Sigma)) = \log(h_t^\bullet) + \log(h_t^\circ) + \log(1 - \rho^2)$. The inverse matrix of the variance-covariance matrix is $\Sigma^{-1} = \frac{1}{h_t^\bullet h_t^\circ (1 - \rho^2)} \begin{pmatrix} h_t^\circ & -\rho\sqrt{h_t^\bullet h_t^\circ} \\ -\rho\sqrt{h_t^\bullet h_t^\circ} & h_t^\bullet \end{pmatrix}$. As such, the product $-\frac{1}{2}(r_t^\bullet, r_t^\circ)\Sigma^{-1}\begin{pmatrix} r_t^\bullet \\ r_t^\circ \end{pmatrix}$ becomes $-\frac{1}{2}(r_t^\bullet, r_t^\circ)\Sigma^{-1}\begin{pmatrix} r_t^\bullet \\ r_t^\circ \end{pmatrix} = -\frac{1}{2} \frac{r_t^{\bullet 2} h_t^{\circ 2} + r_t^{\circ 2} h_t^{\bullet 2} - 2r_t^\bullet r_t^\circ \rho \sqrt{h_t^\bullet h_t^\circ}}{h_t^\bullet h_t^\circ (1 - \rho^2)}$. Thus, the log-likelihood function $\log l(r_t^\bullet, r_t^\circ)$ is obtained by multiplying the functions $f(r_t^\bullet, r_t^\circ)$ for the $t = 1, \dots, n$, and by taking the log of the resulting product $\log l(r_t^\bullet, r_t^\circ) = -\frac{1}{2} \sum_{t=1}^n \left\{ 2\log(2\pi) + \log(1 - \rho^2) + \log(h_t^\bullet) + \log(h_t^\circ) + \frac{r_t^{\bullet 2} h_t^{\circ 2} + r_t^{\circ 2} h_t^{\bullet 2} - 2r_t^\bullet r_t^\circ \rho \sqrt{h_t^\bullet h_t^\circ}}{h_t^\bullet h_t^\circ (1 - \rho^2)} \right\}$.

By performing some simple iterations in the expression above, we obtained the final form of the bivariate log-likelihood function as $\log l(r_t^\bullet, r_t^\circ) = -\frac{1}{2} \sum_{t=1}^n \left\{ 2\log(2\pi) + \log(1 - \rho^2) + \log(h_t^\bullet) + \log(h_t^\circ) + \frac{r_t^{\bullet 2} / h_t^\bullet + r_t^{\circ 2} / h_t^\circ}{(1 - \rho^2)} - \frac{2\rho}{(1 - \rho^2)} \frac{r_t^\bullet r_t^\circ}{\sqrt{h_t^\bullet h_t^\circ}} \right\}$.

Appendix B

Table A1. Maximized log-likelihood functions in univariate and bivariate estimations; in-sample.

Stock	EGARCH		EGARCH-X		Realized EGARCH		Realized GARCH			Realized GARCH (2,2)	
	Univ	Biv	Univ	Biv	Univ	Biv	Univ	Biv (com)	Biv (red)	Univ	Biv
AIG	-1721.9	-2900.0	-1710.1	-2821.5	-1711.8	-2845.7	-1709.1	-2874.4	-2875.4	-1701.3	-2849.9
AXP	-1668.6	-2742.6	-1637.8	-2790.4	-1645.5	-2842.3	-1642.7	-2855.3	-2857.5	-1638.3	-2904.8

BAC	-1506.6	-2499.9	-1473.0	-2438.8	-1475.5	-2437.7	-1478.5	-2443.0	-2439.7	-1471.5	-2437.1
CSCO	-1722.4	-2886.2	-1709.7	-2820.9	-1712.9	-2841.1	-1711.7	-2876.9	-2876.1	-1702.1	-2845.1
F	-1673.5	-2746.3	-1644.2	-2791.8	-1645.9	-2841.5	-1644.1	-2853.8	-2855.2	-1642.8	-2898.9
GE	-1504.9	-2498.1	-1474.2	-2433.2	-1475.8	-2442.8	-1477.9	-2446.1	-2440.1	-1467.1	-2440.7
INTC	-1505.4	-2497.5	-1471.4	-2434.7	-1478.1	-2439.8	-1475.5	-2445.9	-2439.8	-1468.1	-2437.7
JPM	-1658.1	-2750.6	-1616.4	-2699.0	-1619.6	-2702.0	-1625.9	-2714.3	-2703.1	-1615.4	-2683.7
MSFT	-1668.0	-2743.3	-1639.9	-2792.0	-1639.3	-2840.6	-1642.8	-2851.0	-2855.3	-1639.9	-2903.1
T	-1507.1	-2497.5	-1470.9	-2434.4	-1477.5	-2438.6	-1478.2	-2442.1	-2440.0	-1471.3	-2439.3

Table A2. Maximized log-likelihood functions in univariate and bivariate estimations; out-of-sample.

Stock	EGARCH		EGARCH-X		Realized EGARCH		Realized GARCH			Realized GARCH (2,2)	
	Univ	Biv	Univ	Biv	Univ	Biv	Univ	Biv (com)	Biv (red)	Univ	Biv
AIG	-399.5	-1032.1	-394.2	-795.4	-387.7	-749.6	-372.5	-777.9	-774.4	-383.6	-736.9
AXP	-313.3	-607.7	-309.6	-565.7	-310.7	-576.5	-305.9	-561.3	-561.8	-308.4	-573.1
BAC	-344.6	-654.0	-341.9	-687.9	-352.0	-671.4	-337.0	-673.7	-676.2	-337.6	-670.0
CSCO	-407.3	-1033.4	-386.0	-790.5	-392.5	-752.5	-376.0	-770.1	-777.9	-372.5	-732.2
F	-308.9	-602.1	-308.0	-560.3	-307.6	-566.1	-305.6	-573.0	-563.7	-307.7	-570.1
GE	-348.8	-657.6	-339.0	-687.8	-353.7	-666.6	-348.5	-678.1	-672.0	-339.1	-672.9
INTC	-347.5	-659.5	-345.9	-681.7	-351.3	-678.7	-336.2	-674.3	-674.8	-340.5	-676.8
JPM	-330.9	-607.9	-326.1	-589.3	-324.1	-582.2	-316.0	-579.1	-573.1	-323.2	-584.8
MSFT	-403.5	-1030.5	-393.2	-787.9	-389.4	-745.4	-368.9	-772.7	-780.4	-386.3	-733.3
T	-315.0	-603.8	-305.2	-568.1	-303.8	-569.8	-301.4	-570.6	-572.1	-304.2	-570.6

References

- Ahoniemi, Katja, and Markku Lanne. 2013. Overnight Stock Returns and Realized Volatility. *International Journal of Forecasting* 29: 592–604.
- Andersen, Torben G., and Tim Bollerslev. 1998. Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review* 39: 885–905.
- Angelidis, Timotheos, and Stavros Degiannakis. 2008. Volatility Forecasting: Intra-day versus Inter-day Models. *Journal of International Financial Markets, Institutions and Money* 18: 449–65.
- Banulescu-Radu, Denisa, Peter Reinhard Hansen, Zhuo Huang, and Marius Matei. 2019. Volatility During the Financial Crisis Through the Lens of High Frequency Data: A Realized GARCH Approach. Available online: <https://sites.google.com/site/peterreinhardhansen/research-papers/volatilityduringthefinancialcrisisthroughthelensofhighfrequencydataarealizedgarchapproach> (accessed on 5 May 2019).
- Barndorff-Nielsen, Ole Eiler, Peter Reinhard Hansen, Asger Lunde, and Neil Shephard. 2009. Realized Kernels in Practice: Trades and Quotes. *The Econometrics Journal* 12: C1–C32.
- Barndorff-Nielsen, Ole Eiler, Peter Reinhard Hansen, Asger Lunde, and Neil Shephard. 2011. Multivariate Realised Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices with Noise and Non-Synchronous Trading. *Journal of Econometrics* 162: 149–69.
- Blair, Bevan J., Ser-Huang Poon, and Stephen J. Taylor. 2001. Forecasting S&P 100 Volatility: The Incremental Information Content of Implied Volatilities and High-Frequency Index Returns. *Journal of Econometrics* 105: 5–26.
- Bollerslev, Tim, George Tauchen, and Hao Zhou. 2009. Expected Stock Returns and Variance Risk Premia. *Review of Financial Studies* 22: 4463–92.
- Bollerslev, Tim, Andrew J. Patton, and Rogier Quaadvlieg. 2018. Multivariate Leverage Effects and Realized Semicovariance GARCH Models. Available online: <https://ssrn.com/abstract=3164361> or <http://dx.doi.org/10.2139/ssrn.3164361> (accessed on 10 February 2019).
- Corsi, Fulvio. 2009. A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics* 7: 174–96.

- De Pooter, Michiel, Martin Martens, and Dick van Dijk. 2008. Predicting the Daily Covariance Matrix for S&P 100 Stocks Using Intraday Data—But Which Frequency to Use? *Econometric Reviews* 27: 199–229.
- Engle, Robert F. 2002. New Frontiers of ARCH Models. *Journal of Applied Econometrics* 17: 425–46.
- Engle, Robert F., and Giampiero M. Gallo. 2006. A Multiple Indicators Model for Volatility Using Intra-daily Data. *Journal of Econometrics* 131: 3–27.
- Fleming, Jeff and Chris Kirby. 2011. Long Memory in Volatility and Trading Volume. *Journal of Banking and Finance* 35: 1714–26.
- Fleming, Jeff, Chris Kirby, and Barbara OstDiek. 2003. The Economic Value of Volatility Timing Using “Realized” Volatility. *Journal of Financial Economics* 67: 473–509.
- Fuertes, Ana-Maria, and Jose Olmo. 2013. Optimally Harnessing Inter-day and Intra-day Information for Daily Value-at-Risk Prediction. *International Journal of Forecasting* 29: 28–42.
- Hansen, Peter Reinhard, and Zhuo Huang. 2016. Exponential GARCH Modeling with Realized Measures of Volatility. *Journal of Business & Economic Statistics* 34: 269–87.
- Hansen, Peter Reinhard, and Asger Lunde. 2005. A Realized Variance for the Whole Day Based on Intermittent High-frequency Data. *Journal of Financial Econometrics* 3: 525–54.
- Hansen, Peter Reinhard, Zhuo Huang, and Howard Howan Shek. 2012. Realized GARCH: A Joint Model for Returns and Realized Measures of Volatility. *Journal of Applied Econometrics* 27: 877–906.
- Hansen, Peter Reinhard, Asger Lunde, and Valeri Voev. 2014. Realized Beta GARCH: A Multivariate GARCH Model with Realized Measures of Volatility. *Journal of Applied Econometrics* 29: 774–99.
- Hansen, Peter Reinhard, Pawel Janus, and Siem Jan Koopman. 2019. Realized Wishart-GARCH: A Score-driven Multi-Asset Volatility Model. *Journal of Financial Econometrics* 17: 1–32.
- Huang, Zhuo, Hao Liu, and Tianyi Wang. 2016. Modeling Long Memory Volatility Using Realized Measures of Volatility: A Realized HAR GARCH Model. *Economic Modelling* 52: 812–21.
- Koopman, Siem Jan, Borus Jungbacker, and Eugenie Hol. 2005. Forecasting Daily Variability of the S&P 100 Stock Index Using Historical, Realised, and Implied Volatility Measurements. *Journal of Empirical Finance* 12: 445–75.
- Martens, Martin 2002. Measuring and Forecasting S&P 500 Index-Futures Volatility Using High-frequency Data. *Journal of Futures Markets* 22: 497–518.
- Shephard, Neil, and Kevin Sheppard. 2010. Realising the Future: Forecasting with High-frequency-based Volatility (HEAVY) Models. *Journal of Applied Econometrics* 25: 197–231.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).