Evaluation of Realized Volatility Predictions from Models with Leptokurtically and Asymmetrically Distributed Forecast Errors

Degiannakis, Stavros and Livada, Alexandra

Department of Economics and Regional Development, Panteion University, Department of Statistics, Athens University of Economics and Business, Postgraduate Department of Business Administration, Hellenic Open University

November 2013

Evaluation of Realized Volatility Predictions from Models with Leptokurtically and Asymmetrically Distributed Forecast Errors

Stavros Degiannakis[1,3*], Alexandra Livada [2,3]

[1]Department of Economics and Regional Development,
Panteion University,
136 Syngrou Avenue, 17 671, Greece
[2]Department of Statistics,
Athens University of Economics and Business,
76 Patission str., 10 434, Greece
[3]Postgraduate Department of Business Administration,
Hellenic Open University,
Aristotelous 18, 26 335, Greece

[*]Corresponding author, email: stavros.degiannakis@gmail.com

Abstract

Accurate volatility forecasting is a key determinant for portfolio management, risk management and economic policy. The paper provides evidence that the sum of squared standardized forecast errors is a reliable measure for model evaluation when the predicted variable is the intra-day realized volatility. The forecasting evaluation is valid for standardized forecast errors with leptokurtic distribution as well as with leptokurtic and asymmetric distribution. Additionally, the widely applied forecasting evaluation function, the predicted mean squared error, fails to select the adequate model in the case of models with residuals that are leptokurtically and asymmetrically distributed. Hence, the realized volatility forecasting evaluation should be based on the standardized forecast errors instead of their unstandardized version.

Keywords: integrated volatility, intra-day, predicted mean squared error, realized volatility, standardized prediction error criterion, simulating forecast errors, ultra-high frequency, volatility forecasting evaluation.

JEL Classifications: G17; G15; C15; C32; C53.

1. *Introduction*

The methods of models' evaluation can be categorized into three groups: i) The evaluation or loss functions that measure the distance between the predicted and actual values of the variable under investigation[1]. ii) The information criteria, which are based on the estimation of the Kullback and Leibler (1951) discrepancy[2]. iii) The loss functions which are dependent upon the aims of a specific application[3].

The paper investigates a method of models' evaluation that belongs to the first group; the sum of squared standardized forecast errors. The performance of the evaluation function is explored in case that the predicted variable is the realized volatility. The importance of volatility forecasting has been long established in financial literature. Volatility forecasting is essential for investors in predicting portfolio's future uncertainty, forming suitable hedging strategies, pricing volatility indices and other derivative products, estimating their capital requirements, the Value-at-Risk, etc. The computation of intra-day realized volatility is based on the sum of squared log returns of an underlying asset over a trading day. As Andersen and Bollerslev (1998) first noted, the intra-day realized volatility generates the most accurate volatility measures.

We compute the forecast errors for the most widely known specifications for modelling and forecasting intra-day realized volatility, the ARFIMA and HAR frameworks, and investigate whether the data-generated model achieves the lowest value of the sum of squared standardized forecast errors. We investigate the properties of the sum of squared standardized forecast errors under model specifications with i) symmetric ii) leptokurtic and iii) leptokurtic and asymmetric distributions.

The most widely applied loss function in forecast evaluation is the predicted mean squared error, or PMSE. The PMSE evaluation function fails to provide the lowest value to the data-generated model in the case of leptokurtically and asymmetrically distributed innovations. However, its standardized version, named SPEC (standardized prediction error criterion) by Degiannakis and Xekalaki (2005), picks the correct model as the most accurate. Thus, the sum of squared standardized forecast errors is a reliable criterion for evaluating predictability for realized volatility models with leptokurtically and asymmetrically distributed residuals as well.

---

[1] The most known evaluation functions for volatility forecasts are the heteroskedasticity adjusted absolute error (Andersen et *al*., 1999) and the logarithmic error (Pagan and Schwert, 1990).
[2] The most widely applied information criterion is the Schwarz's (1978) Bayesian criterion.
[3] For example, Granger and Pesaran (2000) linked forecast evaluation with the decisions made based on the predictions. Engle et *al*. (1993) and Xekalaki and Degiannakis (2005) developed an evaluation function that measures the profitability of trading options.

The rest of the paper is organised as follows: Section 2 illustrates the theoretical framework of integrated variance and its estimator the realized volatility, while Section 3 describes the most widely known specifications for forecasting realized volatility, the ARFIMA and HAR frameworks. Section 4 presents the steps of constructing simulated forecasts from symmetric, leptokurtic and/or asymmetric distributions. Section 5 investigates the distributional properties of the standardized forecast errors, whereas Section 6 provides evidence that the sum of squared standardized forecast errors is an evaluation function that provides information about the forecasting accuracy of models with residuals that are either leptokurtically or leptokurtically and asymmetrically distributed. Section 7 tests whether the distribution function of the forecast error is stochastically equal to the distribution function of the simulated stochastic process, and, Section 8 concludes the paper.

2. *Integrated and Realized Volatility*

Financial literature assumes that the instantaneous logarithmic price, $\log p(t)$, of a financial asset follows a simple diffusion process $d\log p(t) = \sigma(t)dW(t)$, where $\sigma(t)$ is the volatility of the instantaneous log-returns process and $W(t)$ is the standard Wiener process. Over a trading day with opening and closing times denoting as $[a,b]$, the aggregated, or *integrated*, volatility $\sigma_{[a,b]}^{2(IV)} = \int_a^b \sigma^2(t)dt$ is a latent variable which can be consistently estimated by the realized volatility (i.e. theory of quadratic variation of semi-martingales, Barndorff-Nielsen and Shephard, 2002 and 2004). For the trading day $[a,b]$ being partitioned into τ equidistance points $t_1, t_2, ..., t_\tau$, the realized volatility converges in probability to the integrated volatility, as $\tau \to \infty$,

$$p\lim_{\tau \to \infty}\left(RV_t^{(\tau)}\right) = \sigma_{[a,b]}^{2(IV)}. \tag{1}$$

Therefore, as $\sigma_{[a,b]}^{2(IV)} = \int_{t_1}^{t_2} \sigma^2(t)dt + \int_{t_2}^{t_3} \sigma^2(t)dt + ... + \int_{t_{\tau-1}}^{t_\tau} \sigma^2(t)dt$, and the length of each sub-interval is tending to zero, or $dt \approx t_j - t_{j-1}$, the $RV_t^{(\tau)}$ is computed as:

$$RV_t^{(\tau)} = \sum_{j=1}^{\tau}\left(\log P_{t_j} - \log P_{t_{j-1}}\right)^2, \tag{2}$$

and is asymptotically distributed, $\sqrt{\tau}\left(RV_t^{(\tau)} - \sigma_{[a,b]}^{2(IV)}\right) \xrightarrow{d} N\left(0, \sigma_{[a,b]}^{2(IQ)}\right)$, where $\sigma_{[a,b]}^{2(IQ)} = \int_a^b 2\sigma^4(t)dt$ is termed *integrated quarticity*. For Barndorff-Nielsen and Shephard's (2005) realized power variation of order 2q defining as $RV_t^{(\tau)[2q]} = \sum_{j=2}^{\tau}\left(\log P_{t_j} - \log P_{t_{j-1}}\right)^{2q}$, the logarithmic realized variance is asymptotically normally distributed:

$$\frac{\log\left(RV_t^{(\tau)}\right) - \log\left(\sigma_{[a,b]}^{2(IV)}\right)}{\sqrt{\frac{2}{3}RV_t^{(\tau)[4]}RV_t^{-2}}} \xrightarrow{d} N(0,1). \tag{3}$$

3. Modelling Realized Volatility

3.1. ARFIMA(k,d,l)-GARCH(p,q) Model

The Autoregressive Fractionally Integrated Moving Average with time varying Heteroscedastic Errors, or ARFIMA(k,d,l)-GARCH(p,q) model, initially developed by Baillie et *al*. (1996), is defined as:

$$\begin{gathered}
(1 - C(L))(1 - L)^d\left(\log\left(RV_t^{(\tau)}\right) - \beta_0\right) = (1 + D(L))\varepsilon_t \\
\varepsilon_t = h_t z_t \\
h_t^2 = a_0 + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} b_i h_{t-i}^2 \\
z_t \sim F(\nu, 0, 1),
\end{gathered} \tag{4}$$

where $\beta_0, d, a_0, a_1, ..., a_q, b_1, ..., b_p$ are parameters for estimation, $C(L) = \sum_{i=1}^{k} c_i L^i$ and $D(L) = \sum_{i=1}^{l} d_i L^i$ are polynomials with parameters $c_1, ..., c_k, d_1, ..., d_l$ for estimation, and z_t is a vector process with zero mean, unit variance and its density function, $F(.)$ is defined by a vector of parameters ν.

Due to the long memory property of volatility, the ARFIMA framework is suitable for estimating the logarithmic of the realized volatility. The volatility of volatility also exhibits time-variation and volatility clustering; see Corsi et *al*. (2008). The time varying estimate h_t^2 approximates the $\frac{2}{3}RV_t^{(\tau)[4]}RV_t^{-2}$. The GARCH component accounts for the volatility clustering of realized volatility. Degiannakis (2008) showed that the ARFIMA(k,d,l)-

GARCH(p,q) model is able to provide statistically superior one trading day ahead realized volatility forecasts.

3.2. HAR-RV- GARCH(p,q) Model

The Heterogeneous Autoregressive realized volatility GARCH, or HAR-RV-GARCH(p,q), model relates the current trading day's realized volatility with the daily, weekly and monthly realized volatilities. The autoregressive structure of the volatilities realized over different interval sizes replicates the different perspectives that market participants have on their investment horizon, which is Müller's et *al*. (1997) basic idea of the heterogenous market hypothesis.

Corsi et *al*. (2008) introduced the model as:

$$\log\left(RV_t^{(\tau)}\right) = w_0 + w_1 \log\left(RV_{t-1}^{(\tau)}\right) + w_2\left(5^{-1}\sum_{j=1}^{5}\log\left(RV_{t-j}^{(\tau)}\right)\right) + w_3\left(22^{-1}\sum_{j=1}^{22}\log\left(RV_{t-j}^{(\tau)}\right)\right) + \varepsilon_t,$$

$$\begin{aligned} \varepsilon_t &= h_t z_t, \\ h_t^2 &= a_0 + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} b_i \sigma_{t-i}^2, \\ z_t &\sim F(v,0,1), \end{aligned} \tag{5}$$

where $w_0, w_1, w_2, w_3, a_0, a_1, ..., a_q, b_1, ..., b_p$ are the unknown parameters to be estimated.

4. Simulating the Forecast Errors

4.1. Symmetric Distribution

For $RV_{t+1|t}^{(\tau)}$ denoting the *one-day-ahead realized volatility forecast for day* $t+1$, *which was made on previous trading day* t, the distance between actual and forecasted volatility, or *forecast error*, is:

$$\varepsilon_{t+1|t} = \left(\log\left(RV_{t+1}^{(\tau)}\right) - \log\left(RV_{t+1|t}^{(\tau)}\right)\right). \tag{6}$$

In order to generate forecast errors from the ARFIMA(0,d,1)-GARCH(1,1) model under the assumption of conditional normally distributed innovations, we proceed as follows:

1. Generate random numbers from the standard normal distribution; $\{z_t\}_{t=1}^{T} \sim N(0,1)$.
2. Generate the $\log\left(RV_t^{(\tau)}\right)$ from the ARFIMA(0,d,1)-GARCH(1,1), for $\hat{d} = 0.59, \hat{d}_1 = -0.22, \hat{\beta}_0 = -8.92, \hat{a}_0 = 0.048, \hat{a}_1 = 0.088, \hat{b}_1 = 0.720$[4]. The conditional mean equation is computed as[5]:

[4] The values of the parameters are based on the estimation of the model for the realized volatility of the CAC40 index; see Degiannakis and Floros (2012).

$$\log\left(RV_t^{(\tau)}\right) = \hat{\beta}_0 + \sum_{j=0}^{\infty}\left(\frac{\Gamma\left(j+\hat{d}\right)}{\Gamma\left(\hat{d}\right)\Gamma(j+1)}L^j\right)\varepsilon_t + \sum_{j=0}^{\infty}\left(\frac{\Gamma\left(j+\hat{d}\right)}{\Gamma\left(\hat{d}\right)\Gamma(j+1)}L^j\right)\hat{d}_1\varepsilon_{t-1} \quad (7)$$

having computed the innovation term in its time varying heteroscedastic formation:

$$\varepsilon_t = \sqrt{\hat{a}_0 + \hat{a}_1\varepsilon_{t-1}^2 + \hat{b}_1 h_{t-1}^2}\, z_t \,. \quad (8)$$

3. Estimate the forecast of conditional volatility, $h_{t+1|t}^2$, the one-step-ahead logarithmic realized volatility, $\log\left(RV_{t+1|t}^{(\tau)}\right)$, the forecast errors, $\varepsilon_{t+1|t}$:

$$h_{t+1|t}^2 = \hat{a}_0 + \hat{a}_1\varepsilon_{t|t}^2 + \hat{b}_1 h_{t|t}^2 , \quad (9)$$

$$\log\left(RV_{t+1|t}^{(\tau)}\right) = \hat{\beta}_0 + \sum_{j=1}^{\infty}\left(\frac{\Gamma\left(j+\hat{d}\right)}{\Gamma\left(\hat{d}\right)\Gamma(j+1)}L^{j-1}\right)\varepsilon_{t|t} + \sum_{j=0}^{\infty}\left(\frac{\Gamma\left(j+\hat{d}\right)}{\Gamma\left(\hat{d}\right)\Gamma(j+1)}L^j\right)\hat{d}_1\varepsilon_{t|t} \quad (10)$$

$$\varepsilon_{t+1|t} = \log\left(RV_{t+1}^{(\tau)}\right) - \log\left(RV_{t+1|t}^{(\tau)}\right) \quad (11)$$

and the standardized forecast errors, $z_{t+1|t}$, from the data-generated model:

$$z_{t+1|t} = \frac{\varepsilon_{t+1|t}}{\sqrt{h_{t+1|t}^2}} \,. \quad (12)$$

The standardized forecast error, $z_{t+1|t}$, approximates asymptotically the logarithmic realized variance formulation of equation (3).

4.2. Leptokurtic Distribution

For Student t distributed innovations with v degrees of freedom, $z_t \sim t(0,1;\nu)$, the density function is

$$f_{(t)}(z_t;\nu) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi(\nu-2)}}\left(1+\frac{z_t^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}, \text{ for } \nu > 2 , \quad (13)$$

where $\Gamma(.)$ is the gamma function.

In order to generate forecast errors from the ARFIMA(0,d,1)-GARCH(1,1) model under the assumption of conditional Student t distributed innovations, we proceed as follows:

[5] The infinite expansions of the fractional differencing operator, for $d > 0$, are defined as $(1-L)^{-d} = \sum_{j=0}^{\infty}\left(\frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)}L^j\right) = 1 + \frac{1}{1!}dL + \frac{1}{2!}d(1+d)L^2 - \ldots$, see Xekalaki and Degiannakis (2010, p.113) and Baillie (1996, p.18).

1. Generate random numbers from the Student t distribution; $\{z_t\}_{t=1}^{T} \sim t(0,1;v)$. We assume that $\hat{v} = 5.9$, based on the actual data of the CAC40 realized volatility.

2. Generate the $\log\left(RV_t^{(\tau)}\right)$ from the ARFIMA(0,d,1)-GARCH(1,1), for $\hat{d} = 0.57, \hat{d}_1 = -0.22, \hat{\beta}_0 = -8.95, \hat{a}_0 = 0.040, \hat{a}_1 = 0.097, \hat{b}_1 = 0.742, \hat{v} = 5.9$[6]. The conditional mean and variance equations are computed as in the previous section.

3. We estimate the one-step-ahead forecasts of conditional volatility, $h_{t+1|t}^2$, the one-step-ahead logarithmic realized volatility, $\log\left(RV_{t+1|t}^{(\tau)}\right)$, the forecast errors, $\varepsilon_{t+1|t}$, and the standardized forecast errors, $z_{t+1|t}$, as in the previous section.

In Appendix I the process is also implemented for the ARFIMA(0,d,1)-GARCH(1,1) model with conditional GED (Generalized Error Distribution or Exponential Power distribution) distributed innovations.

4.3. Leptokurtic and Asymmetric Distribution

The skewed Student *t* distribution has been introduced by Fernandez and Steel (1998). For $z_t \sim skT(0,1;v,g)$, the density function is:

$$f_{(skT)}(z_t;v,g) = \begin{cases} \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{\pi(v-2)}}\left(\frac{2s}{g+g^{-1}}\right)\left(1+\frac{sz_t+m}{v-2}g\right)^{-\frac{v+1}{2}} & if \quad z_t < -ms^{-1}, \\ \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{\pi(v-2)}}\left(\frac{2s}{g+g^{-1}}\right)\left(1+\frac{sz_t+m}{v-2}g^{-1}\right)^{-\frac{v+1}{2}} & if \quad z_t \geq -ms^{-1}, \end{cases} \tag{14}$$

where g and v are the asymmetry and tail parameters of the distribution, respectively, and $m = \Gamma((v-1)/2)\sqrt{(v-2)}\left(\Gamma(v/2)\sqrt{\pi}\right)^{-1}\left(g-g^{-1}\right)$, and $s = \sqrt{g^2+g^{-2}-m^2-1}$.

In order to generate forecast errors from the ARFIMA(k,d,l)-GARCH(p,q) model, for k=0, l=p=q=1, with skewed Student *t* conditionally distributed innovations, we proceed as follows:

1. Generate random numbers from skewed Student *t* distribution; $\{z_t\}_{t=1}^{T} \sim skT(0,1;v,g)$. We assume that $\hat{v} = 5.84$ and $\hat{g} = 0.056$, based on the actual data of the CAC40 realized volatility. As random numbers from skewed Student *t* distribution are not available, we follow Degiannakis et *al*. (2014) in order to relate the inverse CDF of the skewed Student *t* with the inverse CDF of the skewed Student *t*. Thus, we generate random numbers $\{\rho_t\}_{t=1}^{T}$

[6] The values of the parameters are based on the estimation of the model for the realized volatility of the CAC40 index.

from the standard uniform distribution, and then compute the $\breve{z}_t$ random draw as $\breve{z}_t = F^{-1}_{(skT)}(\rho_t; v, g)$:

$$F^{-1}_{(skT)}(\rho_t; v, g) = \begin{cases} \dfrac{g^{-1}F^{-1}_{(t)}\left(\dfrac{\rho_t}{2}(1+g^2), v\right) - m}{s} & if \quad \rho_t < \dfrac{1}{1+g^2} \\ \dfrac{-gF^{-1}_{(t)}\left(\dfrac{1-\rho_t}{2}(1+g^{-2}), v\right) - m}{s} & if \quad \rho_t \geq \dfrac{1}{1+g^2}, \end{cases} \tag{15}$$

where $F^{-1}_{(t)}(\rho_t; v)$ denotes the inverse CDF of the Student *t* distribution with v degrees of freedom. The $F^{-1}_{(skT)}(\rho_i; v, g)$ corresponds to the inverse CDF of the skewed Student-*t* distribution with v and g denoting the tail and asymmetry parameters of the distribution, respectively.

2. Generate the $\log\left(RV_t^{(\tau)}\right)$ from the ARFIMA(0,d,1)-GARCH(1,1) with $\{z_t\}_{t=1}^{T} \sim skT(0,1; v, g)$, for $\hat{d} = 0.58, \hat{d}_1 = -0.22, \hat{\beta}_0 = -8.88, \hat{a}_0 = 0.042, \hat{a}_1 = 0.094, \hat{b}_1 = 0.739,$ $\hat{v} = 5.84, \hat{g} = 0.056$. The values of the parameters are based on the estimation of the model for the actual data of the CAC40 realized volatility.

3. We estimate the one-step-ahead forecast values of $h^2_{t+1|t}$, $\log\left(RV^{(\tau)}_{t+1|t}\right)$, $\varepsilon_{t+1|t}$, $z_{t+1|t}$, as in the previous section.

In total 11000 values are simulated for each time series but the first 1000 values are discarded, due to convergence reasons, and we keep $T = 10000$ values of each simulated series.

5. *Investigating the Standardized Forecast Errors*

We proceed to the estimation of ARFIMA(0,d,1)-GARCH(1,1) model under the assumption that the standardized innovations are i) symmetrically; standard normal, $z_t \sim N(0,1)$, ii) leptokurtically; Student *t*, $z_t \sim t(0,1; v)$ or GED, $z_t \sim Ged(0,1; v)$, and iii) leptokurtically and asymmetrically; skewed Student *t*, $z_t \sim skT(0,1; v, g)$, distributed.

For the conditional normally distributed innovations, Table 1 presents the relative descriptive statistics, which are in line with the theoretical evidence. The one-step-ahead standardized forecast errors, $z_{t+1|t}$, are normally distributed (kurtosis 2.95, skewness -0.03, and p-value of the Jarque Bera statistics equal to 0.21). The forecast errors are normally distributed with kurtosis 3.03, skewness -0.04, and p-value of the Jarque Bera statistics equal to 0.14. The forecast of logarithmic realized volatility is approximately normally distributed but with a kurtosis of 3.35 and a p-value of the Jarque Bera statistic equal to zero. Degiannakis and Xekalaki (2005) proved that from a model with $z_t \sim N(0,1)$ innovations, the one-step-ahead forecast errors, $z_{t+1|t}$, under the assumption of constancy of model's parameters over time[7], are asymptotically independently standard normally distributed, $z_{t+1|t} \overset{i.i.d.}{\sim} N(0,1)$.

[Insert Table 1 about here]

The above is generalized for any model framework with consistent estimators of the parameters' vector $\mathbf{\theta}^{(T)}$, for T denoting the sample size that has been used to estimate $\mathbf{\theta}^{(t)}$. If $\mathbf{\theta}^{(T)}$ is strongly consistent for $\mathbf{\theta}^{(t)}$ and asymptotically normal with mean $\mathbf{\theta}$, then, $p\lim\left(\mathbf{\theta}^{(T)}\right) = \mathbf{\theta}^{(t)}$. By Slutsky's theorem (Greene, 1997, p.118), for any continuous function $g(x_T)$ that is not a function of T, $p\lim g(x_T) = g(p\lim x_T)$. Using Slutsky's theorem $p\lim\left(z_{t+1|t}\right) = z_t$. As convergence in probability implies convergence in distribution: $z_{t+1|t} \xrightarrow{p} z_t \Rightarrow z_{t+1|t} \xrightarrow{d} z_t \overset{i.i.d.}{\sim} f(0,1)$. Hence, we can support that $z_{t+1|t}$ are asymptotically Student *t* distributed, since, from the definition of convergence in probability

$$P\left(\left\|\left(X_{1T}, X_{2T}, \ldots, X_{nT}\right) - \left(W_1, W_2, \ldots, W_n\right)\right\| > \varepsilon\right) \leq$$
$$P\left(\left|X_{1T} - W_1\right| > \sqrt{\varepsilon^2/n}\right) + P\left(\left|X_{2T} - W_2\right| > \sqrt{\varepsilon^2/n}\right) + \ldots + P\left(\left|X_{nT} - W_n\right| > \sqrt{\varepsilon^2/n}\right),$$

which asserts that component wise convergence in probability always implies convergence of vectors, i.e., $z_{t+1|t} \xrightarrow{d} z_t \overset{i.i.d.}{\sim} t(0,1;v)$.

Regarding the conditional Student *t* distributed innovations, Table 2 provides evidence that the one-step-ahead standardized forecast errors, $z_{t+1|t}$, share the same distributional properties with the simulated draws from the Student *t* distribution. The kurtosis (skewness) for z_t and $z_{t+1|t}$ is 5.6583 (-0.1511) and 5.5874 (-0.1409), respectively[8]. Figure 1 illustrates

---

[7] We assume that the rolling-sample estimated parameters of the ARFIMA-GARCH model do not change across time. In example, for each point t in time and $\mathbf{\theta}^{(t)} \equiv \left(d^{(t)}, d_1^{(t)}, \beta_0^{(t)}, a_0^{(t)}, a_1^{(t)}, b_1^{(t)}, v^{(t)}\right)'$ we assume that $\mathbf{\theta}^{(t)} = \mathbf{\theta}^{(t')}$, for $\forall t \neq t'$.

[8] Descriptive statistics of the simulated variables $\{z_t\}_{t=1}^{10000}$, $\{\varepsilon_t\}_{t=1}^{10000}$, $\{h_t^2\}_{t=1}^{10000}$, $\{\log(RV_t^{(\tau)})\}_{t=1}^{10000}$, $\{RV_t^{(\tau)}\}_{t=1}^{10000}$ and $\left\{\sqrt{252RV_t^{(\tau)}}\right\}_{t=1}^{10000}$ from the Student *t* distribution, are available upon request.

the line graphs and frequency distribution plots of one-step-ahead simulated forecasts $\{z_{t+1|t}\}_{t=1}^{10000}$, $\{\varepsilon_{t+1|t}\}_{t=1}^{10000}$, $\left\{\sqrt{252RV_{t+1|t}^{(\tau)}}\right\}_{t=1}^{10000}$ and $\left\{\log\left(RV_{t+1|t}^{(\tau)}\right)\right\}_{t=1}^{10000}$.

[Insert Table 2 about here]
[Insert Figure 1 About here]

For the conditional GED distributed innovations, Table 3 presents the relative descriptive statistics. The one-step-ahead standardized forecast errors, $z_{t+1|t}$, share the same distributional properties with the simulated draws from the GED distribution. In example the kurtosis (skewness) for z_t and $z_{t+1|t}$ are 4.1541 (-0.0344) and 4.1330 (-0.0323), respectively[9]. For a model with consistent estimators of the parameters' vector $\boldsymbol{\theta}^{(T)}$, with a sample of size T, we know that $p\lim\left(\boldsymbol{\theta}^{(T)}\right)=\boldsymbol{\theta}^{(t)}$. As Slutsky's theorem holds for any distribution and convergence in probability implies convergence in distribution, we can support that $z_{t+1|t}$ are asymptotically GED distributed, i.e., $z_{t+1|t} \xrightarrow{d} z_t \overset{i.i.d.}{\sim} Ged(0,1;v)$.

[Insert Table 3 about here]

Regarding the conditional skewed Student *t* distributed innovations, Table 4 provides evidence that the one-step-ahead standardized forecast errors, $z_{t+1|t}$, share the same distributional properties with the simulated draws from the skewed Student *t* distribution. The kurtosis (skewness) for z_t and $z_{t+1|t}$ are 9.6241 (-1.8996) and 9.5216 (-1.8805), respectively.[10] For the simulated model with conditional skewed Student *t* distributed innovations, Figure 2 depicts the line graphs and frequency distribution plots of one-step-ahead simulated forecasts $\{z_{t+1|t}\}_{t=1}^{10000}$, $\{\varepsilon_{t+1|t}\}_{t=1}^{10000}$, $\left\{\sqrt{252RV_{t+1|t}^{(\tau)}}\right\}_{t=1}^{10000}$ and $\left\{\log\left(RV_{t+1|t}^{(\tau)}\right)\right\}_{t=1}^{10000}$.

[Insert Figure 2 About here]
[Insert Table 4 about here]

6. Sum of Squared Standardized Forecast Errors

Three models, i.e. ARFIMA(1,d,1)-GARCH(1,1), HAR-RV-GARCH(1,1) and HAR-RV-GARCH(0,1), have been chosen to be compared against the data-generated process, the ARFIMA(0,d,1)-GARCH(1,1) one. The three models share very common characteristics with the original data-generated model. The choice of competing models that share common specifications is based on our choice to make the comparison of the models a difficult task.

9 Plots and frequency distributions of the one-step-ahead simulated forecasts $\{z_{t+1|t}\}_{t=1}^{10000}$, $\{\varepsilon_{t+1|t}\}_{t=1}^{10000}$, $\left\{\sqrt{252RV_{t}^{(\tau)}}\right\}_{t=1}^{10000}$, as well as, descriptive statistics of the simulated draws $\{z_t\}_{t=1}^{10000}$, $\{\varepsilon_t\}_{t=1}^{10000}$, $\{h_t^2\}_{t=1}^{10000}$ from the GED distribution, are available upon request.

10 Descriptive statistics of the simulated variables from the skewed Student *t* distribution, are available upon request.

We proceed to the estimation of the 4 models under the assumption that the innovations are i) symmetrically, ii) leptokurtically and iii) leptokurtically and asymmetrically distributed. The density function of z_t is considered as i) the standard normal distribution, $z_t \sim N(0,1)$, ii) the Student t distribution, $z_t \sim t(0,1;\nu)$, and the Generalized Error distribution, $z_t \sim Ged(0,1;\nu)$, and the iii) skewed Student t distribution, $z_t \sim skT(0,1;\nu,g)$. Each one of the 4 models is re-estimated every day, for $\tilde{T}$=4000 days, based on a rolling sample of constant size $\breve{T}$=1000 days. Consider for example the ARFIMA(1,d,1)-GARCH(1,1) model; the parameter vector to be estimated at each point in time t is $\theta^{(t)} = \left(\beta_0^{(t)}, c_1^{(t)}, d^{(t)}, d_1^{(t)}, a_0^{(t)}, a_1^{(t)}, b_1^{(t)}\right)'$. Therefore, for each model the vector $\theta^{(t)}$ is re-estimated every trading day, for $t = \breve{T}, \breve{T}+1, ..., \breve{T}+\tilde{T}-1$ days, based on a rolling sample of constant size $\breve{T}$. Appendix II presents the formulas of computing the one-step-ahead forecasts of logarithmic of realized volatility and its conditional standard deviation from the ARFIMA(0,d,1)-GARCH(1,1), ARFIMA(1,d,1)-GARCH(1,1), HAR-RV-GARCH(1,1) and HAR-RV-GARCH(0,1) models.

The most widely applied loss function in forecast evaluation is the predicted mean squared error, or PMSE. In the case of the *one-day-ahead realized volatility forecast*, the PMSE is the average of squared *forecast errors*:

$$PMSE = \tilde{T}^{-1} \sum_{t=1}^{\tilde{T}} \varepsilon_{t+1|t}^2 \equiv \tilde{T}^{-1} \sum_{t=1}^{\tilde{T}} \left(\log\left(RV_{t+1}^{(\tau)}\right) - \log\left(RV_{t+1|t}^{(\tau)}\right)\right)^2 . \tag{16}$$

The standardized version of the PMSE, named SPEC (standardized prediction error criterion) by Degiannakis and Xekalaki (2005) who investigated its asymptotic distribution for forecast errors from regression models with heteroscedastic residuals, is computed as:

$$SPEC = \tilde{T}^{-1} \sum_{t=1}^{\tilde{T}} z_{t+1|t}^2 \equiv \tilde{T}^{-1} \sum_{t=1}^{\tilde{T}} \left(\frac{\log\left(RV_{t+1}^{(\tau)}\right) - \log\left(RV_{t+1|t}^{(\tau)}\right)}{h_{t+1|t}} \right)^2 . \tag{17}$$

We intend to investigate whether the sum of the $\tilde{T}$ squared standardized forecast errors is an evaluation function that provides information about the forecasting accuracy of models with residuals that are leptokurtically or/and asymmetrically distributed. In the case of normally distributed innovations, we expect the $\sum_{t=1}^{\tilde{T}} z_{t+1|t}^2$ evaluation function to has its lowest value for the data-generated model, i.e. the ARFIMA(0,d,1)-GARCH(1,1).

SPEC criterion defines as most appropriate model for one-step-ahead forecasting, the model with the lowest sum of squared one-step-ahead standardized forecast errors, $\sum_{t=1}^{\tilde{T}} z_{t+1|t}^2$.[11] Hence, under the assumption that $z_{t+1|t(A)}$ and $z_{t+1|t(B)}$ are normally distributed, a solid theoretical background exists for comparing the volatility predictive ability of a set of competing models based on the one-step-ahead standardized forecast errors, $z_{t+1|t}$. Simulations will provide evidence whether the $\sum_{t=1}^{\tilde{T}} z_{t+1|t}^2$ quantity is suitable for evaluating models' predictability leptokurtically and/or asymmetrically distributed forecast errors.

Table 5 panel A provides the descriptive statistics of the standardized one-step-ahead forecast errors from the models with conditionally normally distributed innovations. The skewness, albeit positive, is very close to zero for all the cases. The kurtosis for all the models is almost equal to the normal value of three. The p-values of the Jarque Bera statistic do not reject the null hypothesis of normally distributed standardized one-step-ahead forecast errors, $z_{t+1|t}$.

Panel B of Table 5 provides the descriptive statistics of the standardized one-step-ahead forecast errors from the models with conditionally Student *t* distributed innovations. The skewness of $z_{t+1|t}$ is positive and much higher than in the case of the models with normally distributed innovations. The kurtosis is much close to 4.5 for all the models but the HAR-RV-GARCH(0,1).

Table 5 Panel C provides the descriptive statistics of the standardized one-step-ahead forecast errors from the models with conditionally GED distributed innovations. The skewness of $z_{t+1|t}$ is negative for all the models (the skewness of the simulated forecasts $\{z_{t+1|t}\}_{t=1}^{\tilde{T}}$ is -0.03; see Table 3). For all the models, the kurtosis is much close to 4.1, which is the kurtosis of the simulated forecasts $\{z_{t+1|t}\}_{t=1}^{\tilde{T}}$.

Panel D of Table 5 provides the descriptive statistics of the standardized one-step-ahead forecast errors from the models with conditionally skewed Student *t* distributed innovations**.**

[11] If we denote the realized volatility forecasts produced by models A and B as $RV_{t+1|t(A)}^{(\tau)}$ and $RV_{t+1|t(B)}^{(\tau)}$, respectively, then the forecasts are comparable (in terms of forecasting ability) through testing the null hypothesis that the models produce statistically equivalent predictions against the alternative hypothesis that model A produces more accurate predictions than model B (see also Xekalaki and Degiannakis, 2010). The statistic $\sum_{t=1}^{\tilde{T}} z_{t+1|t(B)}^2 \Big/ \sum_{t=1}^{\tilde{T}} z_{t+1|t(A)}^2$ has known distributional form, the Correlated Gamma Ratio distribution.

The skewness of $z_{t+1|t}$ is much higher, in absolute values, compared to the models with symmetric forecast errors. Noticeable is also the kurtosis of $z_{t+1|t}$ from the models with conditionally skewed Student *t* distributed innovations, which is at least twice higher compared to the models with symmetric forecast errors.

[Insert Table 5 about here]

According to Tables 6 and 7, which present the values of $\sum_{t=1}^{\tilde{T}} \varepsilon_{t+1|t}^2$ and $\sum_{t=1}^{\tilde{T}} z_{t+1|t}^2$ evaluation functions, respectively, the data-generated ARFIMA(0,d,1)-GARCH(1,1) model has the lowest value for both loss functions under all the assumptions for the distribution of the standardized innovations except for the skewed Student *t* distribution. In the case of the leptokurtically and asymmetrically distributed innovations, the unstandardized version of the SPEC criterion, i.e. the PMSE evaluation function, has the same value for both the ARFIMA(0,d,1)-GARCH(1,1) and ARFIMA(1,d,1)-GARCH(1,1) models. Hence, the simulation exercises provide evidence in favor to the use of $\sum_{t=1}^{\tilde{T}} z_{t+1|t}^2$ evaluation function as a method of volatility forecasting evaluation for models with standardized residuals that are leptokurtically and asymmetrically distributed. A case that does not hold for the widely applied PMSE evaluation function.

[Insert Table 6 about here]
[Insert Table 7 about here]

Figure 3 plots the time-varying estimates of the vector of parameters, $\boldsymbol{\theta}^{(t)} \equiv \left(d^{(t)}, d_1^{(t)}, \beta_0^{(t)}, a_0^{(t)}, a_1^{(t)}, b_1^{(t)}\right)'$, of the ARFIMA(0,d,1)-GARCH(1,1) model for normally distributed innovations. The values of $\boldsymbol{\theta}^{(t)}$ change over time, although they are close to the theoretical values $\boldsymbol{\theta} = (0.59, -0.22, -8.92, 0.048, 0.088, 0.720)$ of the data-generated process. The time-varying attitude of $\boldsymbol{\theta}^{(t)}$ is in accordance to Degiannakis et *al*. (2008) who provide evidence that the rolling-sampled parameter estimates of volatility models are indeed time varying.

[Insert Figure 3 About here]

Figure 4 plots the time-varying estimates of the vector of parameters, $\boldsymbol{\theta}^{(t)} \equiv \left(d^{(t)}, d_1^{(t)}, \beta_0^{(t)}, a_0^{(t)}, a_1^{(t)}, b_1^{(t)}, v^{(t)}\right)'$, of the ARFIMA(0,d,1)-GARCH(1,1) model for Student *t* distributed innovations. The values of $\boldsymbol{\theta}^{(t)}$ change over time, although they are close to the theoretical values $\boldsymbol{\theta} = (0.57, -0.22, -8.95, 0.040, 0.097, 0.742, 5.9)$ under the data-generated process. We note that the $\boldsymbol{\theta}^{(t)}$ is more volatility (in terms of time-varying attitude) under the leptokurtic distribution (of the data generated process) than under the normal distribution; specially the $a_0^{(t)}, a_1^{(t)}$ and $b_1^{(t)}$ parameters. Regarding the model with the GED distributed innovations, the values of $\boldsymbol{\theta}^{(t)}$ follows a similar pattern; they are time varying but close to the theoretical values $\boldsymbol{\theta} = (0.59, -0.22, -8.92, 0.043, 0.094, 0.735, 1.33)$ under the data-generated process.

[Insert Figure 4 About here]

Figure 5 plots the time-varying estimates of the vector of parameters, $\boldsymbol{\theta}^{(t)} \equiv \left(d^{(t)}, d_1^{(t)}, \beta_0^{(t)}, a_0^{(t)}, a_1^{(t)}, b_1^{(t)}, v^{(t)}, g^{(t)}\right)'$, of the ARFIMA(0,d,1)-GARCH(1,1) model for skewed Student t distributed innovations. The values of $\boldsymbol{\theta}^{(t)}$ change over time, although they are close to the theoretical values $\boldsymbol{\theta} = (0.58, -0.22, -8.88, 0.042, 0.094, 0.739, 5.84, 0.056)$ under the data-generated process. It is worth noting that $\boldsymbol{\theta}^{(t)}$ is slightly less volatility (in terms of time-varying attitude) under the leptokurtic and asymmetric distribution (of the data generated process) than under the normal distribution. For the models with skewed Student t conditionally distributed residuals, Figure 6 plots the time series of the data-generated $RV_{t+1}^{(\tau)}$ against the forecasts $RV_{t+1|t}^{(\tau)}$, and the standardized forecast errors, $z_{t+1|t}$.

[Insert Figure 5 About here]

[Insert Figure 6 About here]

7. Stochastic Equality of Forecast Errors' Simulated and Empirical Distribution

In this section, we test whether the distribution function of the one-step-ahead standardized forecast errors, $\{z_{t+1|t}\}$, is stochastically equal to the distribution function of the simulated stochastic process $\{z_t\}$.[12] According to Table 8, the standardized forecast errors do have the same distribution with the theoretical (simulated) residuals in the case of the i) symmetric and ii) leptokurtic distributions. However, the standardized forecast errors do not follow the same distribution with the simulated residuals when these have been generated by the leptokurtic and asymmetric distribution; i.e. $z_t \sim skT(0,1;v,g)$.

The test of whether the forecast errors have the distribution function of the simulated residuals is repeated for the unstandardized version of the residuals. Hence, we test the null hypothesis that the distribution function of the one-step-ahead forecast errors, $\{\varepsilon_{t+1|t}\}$, is stochastically equal to the distribution function of the simulated stochastic process $\{\varepsilon_t\}$.

[12] The Mann and Whitney (1947) proposed the U statistic for testing the null hypothesis that two random variables with continuous cumulative distribution functions f and g have stochastically equal distributions against the alternative hypothesis that one distribution is stochastically smaller than the other.

However, the functional form of the conditional, on information set at time $t-1$, distribution of $\{\varepsilon_t\}$ is known (the functional form of the unconditional distribution of $\{\varepsilon_t\}$ is unknown). The results are presented in Table 9. The forecast errors do have the same distribution with the theoretical residuals in the case of the i) symmetric and ii) leptokurtic distributions, but not in the case of the iii) leptokurtic and asymmetric distribution. However, in the case of the GED distributed innovations, the level of significance is much lower for $H_0 : f\left(\{\varepsilon_{t+1|t}\}_{t=1}^{\tilde{T}}\right) = g\left(\{\varepsilon_t\}_{t=1}^{\tilde{T}}\right)$ than for $H_0 : f\left(\{z_{t+1|t}\}_{t=1}^{\tilde{T}}\right) = g\left(\{z_t\}_{t=1}^{\tilde{T}}\right)$. For the ARFIMA(1,d,1)-GARCH(1,1), HAR-RV-GARCH(1,1) and HAR-RV-GARCH(0,1) models, the null hypothesis is not rejected at 10% level of significance.

[Insert Table 8 about here]

[Insert Table 9 about here]

8. Conclusion

We generated simulated realized volatility series from an ARFIMA-GARCH framework assuming that the residuals are conditionally i) standard normally distributed, ii) Student *t* distributed, iii) GED distributed and iv) skewed Student *t* distributed. For a model with consistent estimators of the parameters' vector, we can assume that the standardized forecast errors, $z_{t+1|t}$, convergence asymptotically in distribution to the theoretical distribution of the innovations; i.e if $z_t \overset{i.i.d.}{\sim} t(0,1;v)$ then $z_{t+1|t} \overset{d}{\rightarrow} z_t \overset{i.i.d.}{\sim} t(0,1;v)$.

Then, we estimated the ARFIMA(0,d,1)-GARCH(1,1), ARFIMA(1,d,1)-GARCH(1,1), HAR-RV-GARCH(1,1) and HAR-RV-GARCH(0,1) models under the assumption that the innovations are i) normally ii) Student *t* ii) GED and iv) skewed Student *t* distributed. The models were estimated for the data-generated values of the logarithmic realized volatility. Each one of the models is re-estimated every day, for $\tilde{T}$=4000 days, based on a rolling sample of constant size $\breve{T}$=1000 days. In all the cases, the data-generated model, the ARFIMA(0,d,1)-GARCH(1,1), had the lowest value of the sum of squared standardized forecast errors.

Therefore, simulations provide evidence that the SPEC predictability criterion can be applied to the evaluation of models with residuals which are leptokurtically, or even leptokurtically and asymmetrically, distributed. Therefore, the SPEC evaluation function is indeed a framework under which the forecasting evaluation is valid for forecast errors a) with leptokurtic distribution (such as the standardized Student *t* distribution and the generalized error distribution), as well as b) with leptokurtic and asymmetric distribution (such as the skewed Student t distribution). On the contrary, the unstandardized version of the SPEC criterion, the PMSE evaluation function, does not provide the lowest value to the data-generated model in the case of the leptokurtically and asymmetrically distributed innovations.

Additionally, the forecast errors (both standardized and unstandardized) do have the same distribution with the simulated residuals in the case of the i) symmetric and ii) leptokurtic distributions. On the other hand, the forecast errors do not follow the same distribution with the simulated residuals when these have been generated by the skewed Student t distribution. Finally, for GED distributed innovations, the level of significance of the U statistic is much lower for the unstandardized forecast errors than for the standardized forecast errors.

The aim of the paper is to offer evidence that the SPEC criterion is a useful tool for investigating which model provides better forecasts of intra-day realized volatility. Accurate estimate of future volatility is a key determinant in quantitative finance. The Basel Committee has introduced the VaR estimate in order to measure the minimum capital which is required as a protection against the banks' exposure to financial risks; the correct estimation of the VaR requires accurate volatility forecast. Bernake and Gertler (2001) and Rigobon and Sack (2003) central bankers were also interested in volatility prediction as asset price volatility provides information regarding the state of the economy and future levels of inflation.

Acknowledgement

Authors acknowledge the support from the European Community's Seventh Framework Programme (FP7-PEOPLE-2010-RG) funded under grant agreement no. PERG08-GA-2010-276904. Also, we would like to thank the editor Robert Aykroyd and the anonymous referees for their constructive suggestions.

References

1. **Andersen, T. and Bollerslev, T. (1998).** Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review*, 39, 885-905.

2. **Andersen, T., Bollerslev, T. and Lange, S. (1999).** Forecasting Financial Market Volatility: Sample Frequency vis-à-vis Forecast Horizon. *Journal of Empirical Finance,* 6, 457-477.

3. **Baillie, R.T. (1996).** Long Memory Processes and Fractional Integration in Econometrics. *Journal of Econometrics*, 73, 5-59.

4. **Baillie, R.T., Bollerslev, T. and Mikkelsen, H.O. (1996).** Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 74, 3-30.

5. **Barndorff-Nielsen, O.E. and Shephard, N. (2002).** Econometric Analysis of Realised Volatility and its Use in Estimating Stochastic Volatility Models, *Journal of the Royal Statistical Society, Series B*, 64, 253–280.

6. **Barndorff-Nielsen, O.E. and Shephard, N. (2004).** Power and Bipower Variation with Stochastic Volatility and Jumps, *Journal of Financial Econometrics*, 2, 1–37.

7. **Barndorff-Nielsen, O.E. and Shephard, N. (2005).** How Accurate is the Asymptotic Approximation to the Distribution of Realised Volatility? In (eds.) Andrews, D., Powell, J., Ruud, P., and Stock, J., *Identification and Inference for Econometric Models*. Cambridge University Press, Cambridge.

8. **Bernanke, B.S. and Gertler, M. (2001).** Should Central Banks Respond to Movements in Asset Prices? *The American Economic Review*, 91(2), 253-257.

9. **Corsi, F., Mittnik, S., Pigorsch, C. and Pigorsch, U. (2008).** The Volatility of Realised Volatility. *Econometric Reviews*, 27(1-3), 46-78.

10. **Degiannakis, S. (2008).** ARFIMAX and ARFIMAX- TARCH Realized Volatility Modeling. *Journal of Applied Statistics*, 35(10), 1169-1180.

11. **Degiannakis, S. and Floros, C. (2012).** Modeling CAC40 Volatility Using Ultra-high Frequency Data, *Research in International Business and Finance*, 28, 68-81.

12. **Degiannakis, S. and Xekalaki, E. (2005).** Predictability and Model Selection in the Context of ARCH Models. *Journal of Applied Stochastic Models in Business and Industry*, 21, 55-82.

13. **Degiannakis, S., Dent, P. and Floros, C. (2014).** A Monte Carlo Simulation Approach to Forecasting Multi-period Value-at-Risk and Expected Shortfall Using the FIGARCH-skT Specification. *The Manchester School,* 82(1), 71-102.

14. **Degiannakis, S., Livada A. and Panas, E. (2008).** Rolling-sampled parameters of ARCH and Levy-stable models. *Applied Economics*, 40(23), 3051-3067.

15. **Engle, R.F., Hong, C.H., Kane A. and Noh, J. (1993).** Arbitrage valuation of variance forecasts with simulated options. *Advances in Futures and Options Research*, 6, 393-415.

16. **Fernandez, C. and Steel, M.F.J. (1998).** On Bayesian modelling of fat tails and skewness, *Journal of the American Statistical Association*, 93, 359–371.

17. **Granger, C.W.J. and Pesaran, M.H. (2000).** Economic and Statistical Measures of Forecasting Accuracy. *Journal of Forecasting*, 19, 537-560.

18. **Greene, W.H. (1997).** Econometric Analysis, 3rd edition, Prentice Hall, New Jersey.

19. **Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995).** Continuous Univariate Distributions, Vol. 2, 2nd edition, John Wiley & Sons Ltd., New York.

20. **Kullback, S. and Leibler, R.A. (1951).** On Information and Sufficiency. *Annals of Mathematical Statistics*, 22, 79-86.

21. **Mann, H.B. and Whitney, D.R. (1947).** On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18, 50-60.

22. **Müller, U.A., Dacorogna, M.M., Davé, R.D., Olsen, R.B., Pictet, O.V. and VonWeizsäcker, J.E. (1997).** Volatilities of Different Time Resolutions – Analyzing the Dynamics of Market Components. *Journal of Empirical Finance*, 4, 213-239.

23. **Pagan, A.R. and Schwert, G.W. (1990).** Alternative Models for Conditional Stock Volatility. *Journal of Econometrics,* 45, 267-290.

24. **Rigobon, R. and Sack, B. (2003).** Measuring the Reaction of Monetary Policy to the Stock Market. *The Quarterly Journal of Economics*, 118(2), 639-669.

25. **Schwarz, G. (1978).** Estimating the Dimension of a Model. *Annals of Statistics,* 6, 461-464.

26. **Xekalaki, E. and Degiannakis, S. (2005).** Evaluating Volatility Forecasts in Option Pricing in the Context of a Simulated Options Market. *Computational Statistics and Data Analysis,* 49(2), 611-629.

27. **Xekalaki, E. and Degiannakis, S. (2010).** *ARCH Models for Financial Applications*. John Wiley & Sons Ltd., New York.

Figures and Tables

Table 1. Descriptive statistics of the simulated forecasts $\{z_{t+1|t}\}_{t=1}^{10000}$, $\{\varepsilon_{t+1|t}\}_{t=1}^{10000}$, $\{h_{t+1|t}^2\}_{t=1}^{10000}$, $\{\log(RV_{t+1|t}^{(\tau)})\}_{t=1}^{10000}$, $\{RV_{t+1|t}^{(\tau)}\}_{t=1}^{10000}$ and $\{\sqrt{252RV_{t+1|t}^{(\tau)}}\}_{t=1}^{10000}$, from the ARFIMA(0,d,1)-GARCH(1,1) model with conditional normally distributed innovations.

Index	Mean	Median	Maximum	Minimum	Std.Dev	Skewness	Kurtosis
$\{z_{t+1\|t}\}_{t=1}^{10000}$	-0.001	0.008254	2.223529	-2.31394	0.633784	-0.03528	2.95129
$\{\varepsilon_{t+1\|t}\}_{t=1}^{10000}$	-0.0011	0.004094	1.12287	-1.31951	0.31714	-0.04557	3.032832
$\{h_{t+1\|t}^2\}_{t=1}^{10000}$	0.250415	0.237684	0.920853	0.175809	0.052679	1.949053	10.76465
$\{\log(RV_{t+1\|t}^{(\tau)})\}_{t=1}^{10000}$	-8.93211	-8.94201	-6.16747	-11.8029	0.72967	0.061653	3.346143
$\{RV_{t+1\|t}^{(\tau)}\}_{t=1}^{10000}$	0.000197	0.00013	0.003434	4.70E-06	0.000227	4.510531	37.04997
$\{\sqrt{252RV_{t+1\|t}^{(\tau)}}\}_{t=1}^{10000}$	0.201168	0.180794	0.930279	0.034418	0.095331	1.70828	8.219353

Table 2. Descriptive statistics of the simulated forecasts $\{z_{t+1|t}\}_{t=1}^{10000}$, $\{\varepsilon_{t+1|t}\}_{t=1}^{10000}$, $\{h_{t+1|t}^2\}_{t=1}^{10000}$, $\{\log(RV_{t+1|t}^{(\tau)})\}_{t=1}^{10000}$, $\{RV_{t+1|t}^{(\tau)}\}_{t=1}^{10000}$ and $\{\sqrt{252RV_{t+1|t}^{(\tau)}}\}_{t=1}^{10000}$, from the ARFIMA(0,d,1)-GARCH(1,1) model with conditional Student *t* distributed innovations.

Index	Mean	Median	Maximum	Minimum	Std.Dev	Skewness	Kurtosis
$\{z_{t+1\|t}\}_{t=1}^{10000}$	-0.0023	-0.0054	4.9808	-6.8937	0.7884	-0.1409	5.5874
$\{\varepsilon_{t+1\|t}\}_{t=1}^{10000}$	-0.0020	-0.0030	3.6115	-3.2702	0.4549	-0.0343	6.2032
$\{h_{t+1\|t}^2\}_{t=1}^{10000}$	0.3367	0.2779	4.5527	0.1655	0.2072	5.7260	66.8227
$\{\log(RV_{t+1\|t}^{(\tau)})\}_{t=1}^{10000}$	-8.9864	-8.9813	-5.0461	-13.0483	0.9820	-0.0448	3.7699
$\{RV_{t+1\|t}^{(\tau)}\}_{t=1}^{10000}$	0.0003	0.0001	0.0377	0.0000	0.0010	20.8712	596.7707
$\{\sqrt{252RV_{t+1\|t}^{(\tau)}}\}_{t=1}^{10000}$	0.2141	0.1778	3.0828	0.0094	0.1627	4.9274	53.4953

Table 3. Descriptive statistics of the simulated forecasts $\{z_{t+1|t}\}_{t=1}^{10000}$, $\{\varepsilon_{t+1|t}\}_{t=1}^{10000}$, $\{h_{t+1|t}^{2}\}_{t=1}^{10000}$, $\{\log(RV_{t+1|t}^{(\tau)})\}_{t=1}^{10000}$, $\{RV_{t+1|t}^{(\tau)}\}_{t=1}^{10000}$ and $\{\sqrt{252RV_{t+1|t}^{(\tau)}}\}_{t=1}^{10000}$, from the ARFIMA(0,d,1)-GARCH(1,1) model with conditional GED distributed innovations.

Index	Mean	Median	Maximum	Minimum	Std.Dev	Skewness	Kurtosis
$\{z_{t+1\|t}\}_{t=1}^{10000}$	-0.0121	-0.0135	3.2236	-3.0122	0.6349	-0.0323	4.1330
$\{\varepsilon_{t+1\|t}\}_{t=1}^{10000}$	-0.0065	-0.0065	1.9926	-2.0189	0.3200	-0.0262	4.6875
$\{h_{t+1\|t}^{2}\}_{t=1}^{10000}$	0.2530	0.2294	2.0695	0.1667	0.0845	4.7806	54.6941
$\{\log(RV_{t+1\|t}^{(\tau)})\}_{t=1}^{10000}$	-9.0366	-9.0194	-6.6818	-12.4825	0.7313	-0.2072	3.2359
$\{RV_{t+1\|t}^{(\tau)}\}_{t=1}^{10000}$	0.0002	0.0001	0.0030	0.0000	0.0002	4.1283	34.3137
$\{\sqrt{252RV_{t+1\|t}^{(\tau)}}\}_{t=1}^{10000}$	0.1901	0.1743	0.8658	0.0144	0.0870	1.4680	7.0810

Table 4. Descriptive statistics of the simulated forecasts $\{z_{t+1|t}\}_{t=1}^{10000}$, $\{\varepsilon_{t+1|t}\}_{t=1}^{10000}$, $\{h_{t+1|t}^{2}\}_{t=1}^{10000}$, $\{\log(RV_{t+1|t}^{(\tau)})\}_{t=1}^{10000}$, $\{RV_{t+1|t}^{(\tau)}\}_{t=1}^{10000}$ and $\{\sqrt{252RV_{t+1|t}^{(\tau)}}\}_{t=1}^{10000}$, from the ARFIMA(0,d,1)-GARCH(1,1) model with conditional skewed Student *t* distributed innovations.

Index	Mean	Median	Maximum	Minimum	Std.Dev	Skewness	Kurtosis
$\{z_{t+1\|t}\}_{t=1}^{10000}$	-0.2256	-0.0465	0.9439	-7.3900	0.7732	-1.8805	9.5216
$\{\varepsilon_{t+1\|t}\}_{t=1}^{10000}$	-0.1314	-0.0246	0.9225	-5.2528	0.4520	-2.2679	12.5460
$\{h_{t+1\|t}^{2}\}_{t=1}^{10000}$	0.3435	0.2563	6.6065	0.1686	0.3017	6.2399	63.9382
$\{\log(RV_{t+1\|t}^{(\tau)})\}_{t=1}^{10000}$	-11.180	-10.981	-8.549	-17.7158	1.1823	-0.8576	3.8813
$\{RV_{t+1\|t}^{(\tau)}\}_{t=1}^{10000}$	0.00002	0.00002	0.00028	0.00000	0.00003	2.4539	11.8073
$\{\sqrt{252RV_{t+1\|t}^{(\tau)}}\}_{t=1}^{10000}$	0.0676	0.0622	0.2661	0.0006	0.0387	0.8389	3.7846

Table 5. Descriptive statistics of the standardized one-step-ahead prediction errors, $z_{t+1|t}$, from the four models. The dependent variable $\log\left(RV_t^{(\tau)}\right)$ has been generated by the ARFIMA(0,d,1)-GARCH(1,1) under (i) normally distributed innovations, for $\hat{d}=0.59, \hat{d}_1=-0.22, \hat{\beta}_0=-8.92, \hat{a}_0=0.048, \hat{a}_1=0.088, \hat{b}_1=0.720$, (ii) Student t distributed innovations, for $\hat{d}=0.57, \hat{d}_1=-0.22, \hat{\beta}_0=-8.95, \hat{a}_0=0.040, \hat{a}_1=0.097, \hat{b}_1=0.742, \hat{v}=5.9$, (iii) GED distributed innovations, for $\hat{d}=0.59, \hat{d}_1=-0.22, \hat{\beta}_0=-8.92, \hat{a}_0=0.043, \hat{a}_1=0.094, \hat{b}_1=0.735, \hat{v}=1.33$, and (iv) skewed Student t distributed innovations, for $\hat{d}=0.58, \hat{d}_1=-0.22, \hat{\beta}_0=-8.88, \hat{a}_0=0.042, \hat{a}_1=0.094, \hat{b}_1=0.739, \hat{v}=5.84, \hat{g}=0.056$.

Model	Mean	Median	Maximum	Minimum	Std.Dev	Skewness	Kurtosis
Panel A: normally distributed innovations							
1	0.000	0.000	3.610	-3.650	1.005	0.032	3.061
2	0.000	0.000	3.630	-3.630	1.007	0.035	3.067
3	0.003	0.010	3.810	-3.470	1.008	0.038	3.057
4	0.002	0.020	4.000	-3.420	1.008	0.022	3.124
Panel B: Student *t* distributed innovations							
1	0.004	0.006	6.697	-4.563	1.005	0.089	4.547
2	0.001	0.007	6.739	-4.566	1.005	0.088	4.550
3	0.017	0.020	6.960	-4.730	1.007	0.075	4.520
4	0.019	0.024	7.062	-5.123	1.009	0.108	4.982
Panel C: GED distributed innovations							
1	0.012	0.015	4.327	-5.031	1.002	-0.023	4.132
2	0.013	0.015	4.346	-5.051	1.003	-0.022	4.125
3	0.015	0.014	4.349	-4.896	1.006	-0.015	4.053
4	0.015	0.015	4.272	-4.928	1.007	-0.016	4.069
Panel D: skewed Student *t* distributed innovations							
1	-0.005	0.229	1.746	-9.544	1.011	-2.002	10.829
2	-0.004	0.230	1.747	-9.263	1.007	-1.971	10.474
3	-0.017	0.221	1.713	-8.684	1.021	-1.897	9.768
4	-0.036	0.209	2.262	-8.761	1.026	-2.158	11.614

Model 1: ARFIMA(0,d,1)-GARCH(1,1), Model 2: ARFIMA(1,d,1)-GARCH(1,1), Model 3: HAR-RV-GARCH(1,1), Model 4: HAR-RV-GARCH(0,1).

Table 6. The sum of the squared one-day-ahead forecast errors, $\sum_{t=1}^{\tilde{T}} \varepsilon_{t+1|t}^{2}$, (PMSE loss function) of the four models for conditionally i) normally ii) Student *t* iii) GED and iv) skewed Student *t* distributed innovations.

Model	(i)	(ii)	(iii)	(iv)
1	1024.0	1846.2	947.9	2010.8
2	1028.2	1849.7	950.1	2010.8
3	1044.9	1877.4	957.1	2054.2
4	1045.1	1878.7	957.1	2077.2

Model 1: ARFIMA(0,d,1)-GARCH(1,1), Model 2: ARFIMA(1,d,1)-GARCH(1,1), Model 3: HAR-RV-GARCH(1,1), Model 4: HAR-RV-GARCH(0,1).

Table 7. The sum of the squared standardized forecast errors, $\sum_{t=1}^{\tilde{T}} z_{t+1|t}^{2}$, (SPEC loss function) of the four models for conditionally i) normally ii) Student *t* iii) GED and iv) skewed Student *t* distributed innovations.

Model	(i)	(ii)	(iii)	(iv)
1	4039.2	4035.5	4011.9	4045.0
2	4053.0	4036.3	4021.2	4052.7
3	4059.9	4057.0	4046.9	4167.9
4	4058.8	4075.2	4055.8	4211.5

Model 1: ARFIMA(0,d,1)-GARCH(1,1), Model 2: ARFIMA(1,d,1)-GARCH(1,1), Model 3: HAR-RV-GARCH(1,1), Model 4: HAR-RV-GARCH(0,1).

Table 8. The p-values for testing $H_0 : f\left(\{z_{t+1|t}\}_{t=1}^{\tilde{T}}\right) = g\left(\{z_t\}_{t=1}^{\tilde{T}}\right)$.

Distribution of z_t	Model of $z_{t+1|t}$	p-values
$N(0,1)$	ARFIMA(0,d,1)-GARCH(1,1)	0.659
	ARFIMA(1,d,1)-GARCH(1,1)	0.657
	HAR-RV-GARCH(1,1)	0.580
	HAR-RV-GARCH(0,1)	0.566
$t(0,1;v)$	ARFIMA(0,d,1)-GARCH(1,1)	0.993
	ARFIMA(1,d,1)-GARCH(1,1)	0.897
	HAR-RV-GARCH(1,1)	0.540
	HAR-RV-GARCH(0,1)	0.547
$Ged(0,1;v)$	ARFIMA(0,d,1)-GARCH(1,1)	0.406
	ARFIMA(1,d,1)-GARCH(1,1)	0.470
	HAR-RV-GARCH(1,1)	0.110
	HAR-RV-GARCH(0,1)	0.178
$skT(0,1;v,g)$	ARFIMA(0,d,1)-GARCH(1,1)	0.000
	ARFIMA(1,d,1)-GARCH(1,1)	0.000
	HAR-RV-GARCH(1,1)	0.000
	HAR-RV-GARCH(0,1)	0.000

Table 9. The p-values for testing $H_0 : f\left(\left\{\varepsilon_{t+1|t}\right\}_{t=1}^{\tilde{T}}\right) = g\left(\left\{\varepsilon_t\right\}_{t=1}^{\tilde{T}}\right)$.

Conditional on I_{t-1} Distribution of ε_t	Model of $\varepsilon_{t+1\|t}$	p-values
$N(0,1)$	ARFIMA(0,d,1)-GARCH(1,1)	0.909
	ARFIMA(1,d,1)-GARCH(1,1)	0.905
	HAR-RV-GARCH(1,1)	0.797
	HAR-RV-GARCH(0,1)	0.799
$t(0,1;v)$	ARFIMA(0,d,1)-GARCH(1,1)	0.937
	ARFIMA(1,d,1)-GARCH(1,1)	0.967
	HAR-RV-GARCH(1,1)	0.464
	HAR-RV-GARCH(0,1)	0.488
$Ged(0,1;v)$	ARFIMA(0,d,1)-GARCH(1,1)	0.102
	ARFIMA(1,d,1)-GARCH(1,1)	0.097
	HAR-RV-GARCH(1,1)	0.078
	HAR-RV-GARCH(0,1)	0.085
$skT(0,1;v,g)$	ARFIMA(0,d,1)-GARCH(1,1)	0.000
	ARFIMA(1,d,1)-GARCH(1,1)	0.000
	HAR-RV-GARCH(1,1)	0.000
	HAR-RV-GARCH(0,1)	0.000

Figure 1. Time series plots and frequency distributions of simulated forecasts $\{z_{t+1|t}\}_{t=1}^{10000}$, $\{\varepsilon_{t+1|t}\}_{t=1}^{10000}$, $\left\{\sqrt{252RV_{t+1|t}^{(\tau)}}\right\}_{t=1}^{10000}$ and $\left\{\log\left(RV_{t+1|t}^{(\tau)}\right)\right\}_{t=1}^{10000}$, from the ARFIMA(0,d,1)-GARCH(1,1) model with conditional Student t distributed innovations.

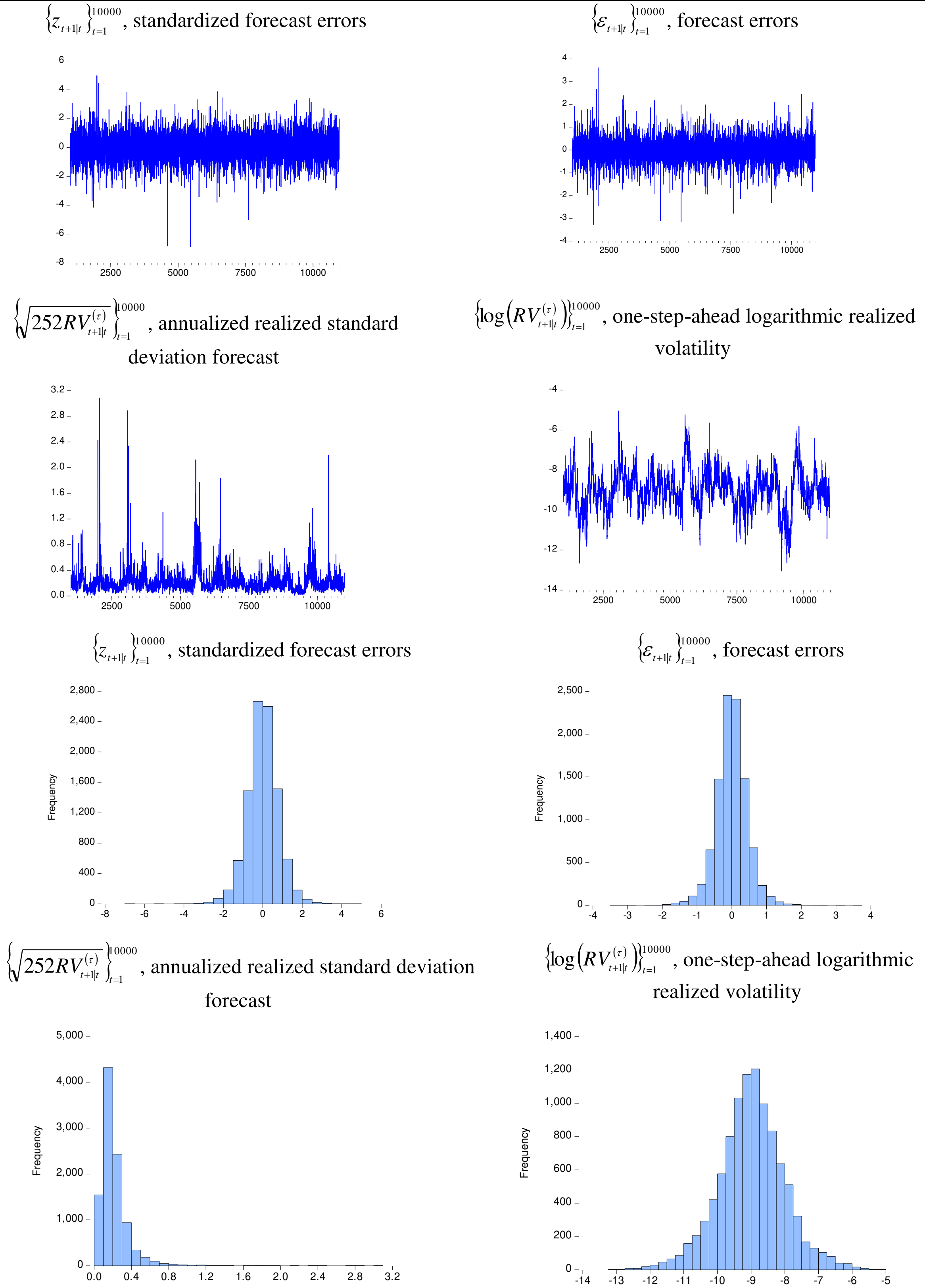


Figure 2. Time series plots and frequency distributions of simulated forecasts $\{z_{t+1|t}\}_{t=1}^{10000}$, $\{\varepsilon_{t+1|t}\}_{t=1}^{10000}$, $\left\{\sqrt{252RV_{t+1|t}^{(\tau)}}\right\}_{t=1}^{10000}$ and $\left\{\log\left(RV_{t+1|t}^{(\tau)}\right)\right\}_{t=1}^{10000}$, from the ARFIMA(0,d,1)-GARCH(1,1) model with conditional skewed Student *t* distributed innovations.

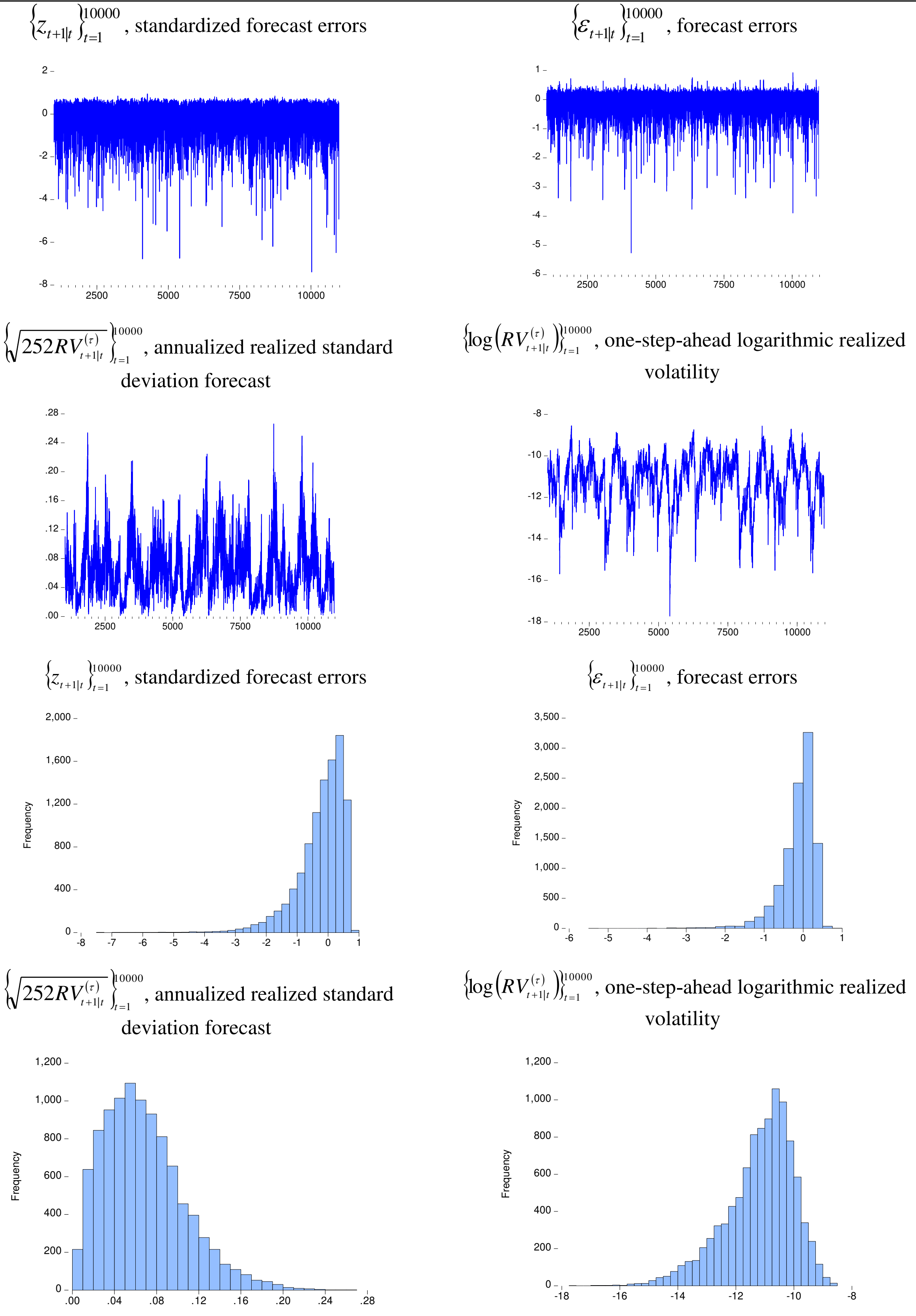


Figure 3. The parameter estimates of the ARFIMA(0,d,1)-GARCH(1,1) model across time. The dependent variable $\log\left(RV_t^{(\tau)}\right)$ has been generated by the ARFIMA(0,d,1)-GARCH(1,1) under normally distributed innovations, for $\hat{d} = 0.59, \hat{d}_1 = -0.22, \hat{\beta}_0 = -8.92, \hat{a}_0 = 0.048, \hat{a}_1 = 0.088, \hat{b}_1 = 0.720$.

Model 1 - ARFIMA(0,d,1)-GARCH(1,1)

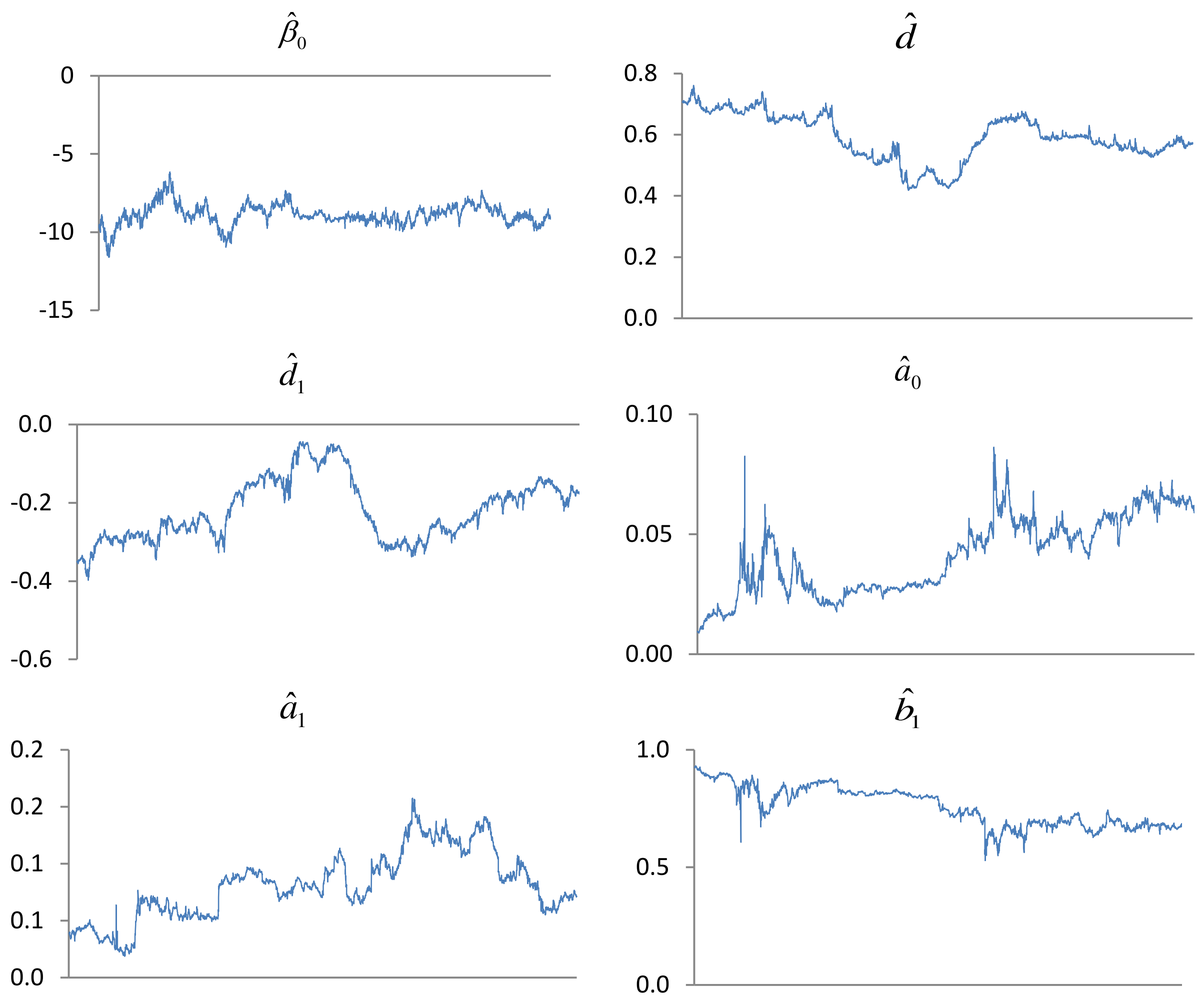


Figure 4. The parameter estimates of the ARFIMA(0,d,1)-GARCH(1,1) model across time. The dependent variable $\log\left(RV_t^{(\tau)}\right)$ has been generated by the ARFIMA(0,d,1)-GARCH(1,1) under Student t distributed innovations, for $\hat{d} = 0.57, \hat{d}_1 = -0.22, \hat{\beta}_0 = -8.95, \hat{a}_0 = 0.040, \hat{a}_1 = 0.097, \hat{b}_1 = 0.742, \hat{v} = 5.9$.

Model 1 - ARFIMA(0,d,1)-GARCH(1,1)

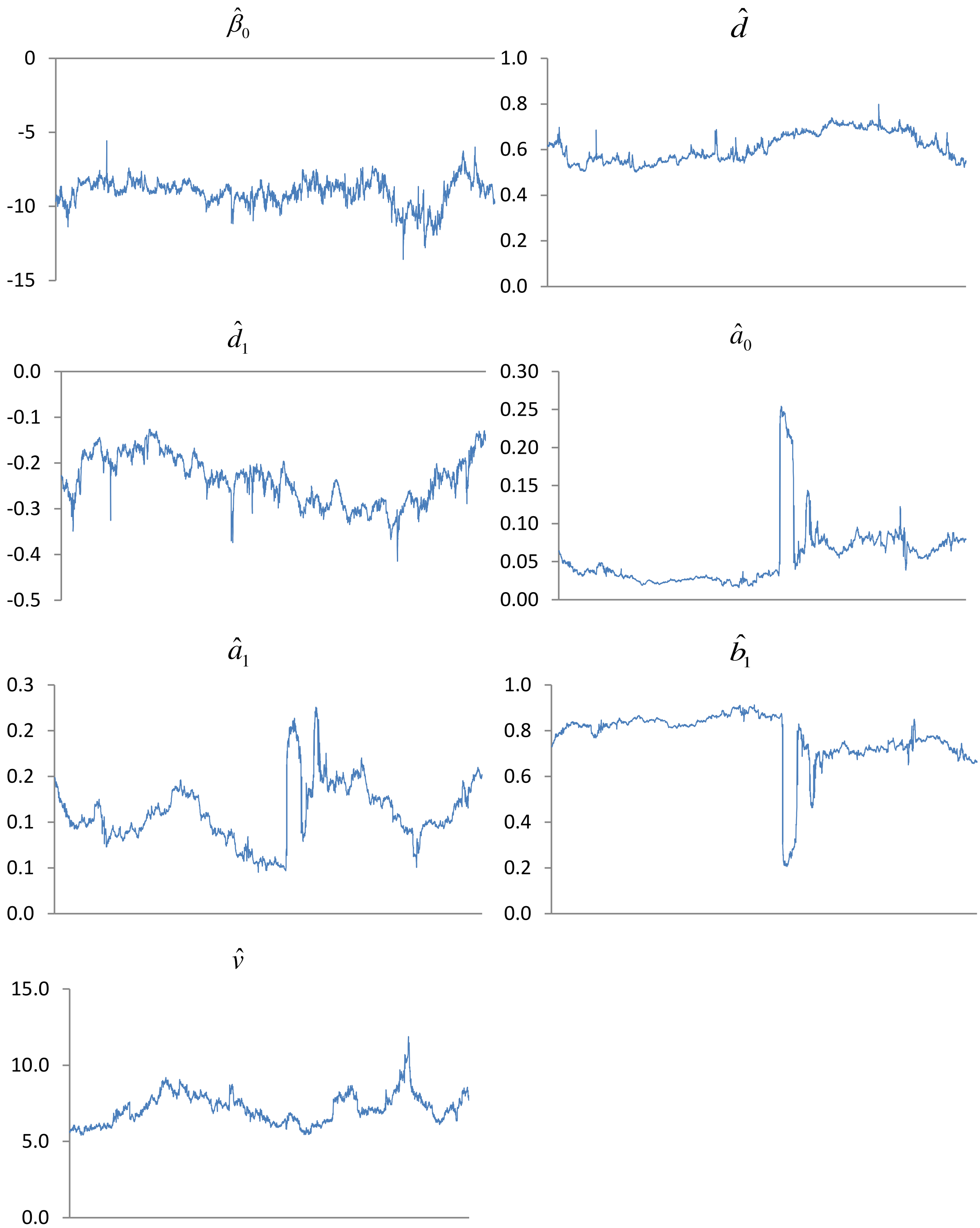


Figure 5. The parameter estimates of the ARFIMA(0,d,1)-GARCH(1,1) model across time. The dependent variable $\log\left(RV_t^{(\tau)}\right)$ has been generated by the ARFIMA(0,d,1)-GARCH(1,1) under skewed Student t distributed innovations, for $\hat{d}=0.58, \hat{d}_1=-0.22, \hat{\beta}_0=-8.88, \hat{a}_0=0.042, \hat{a}_1=0.094, \hat{b}_1=0.739, \hat{v}=5.84, \hat{g}=0.056$.

Model 1 - ARFIMA(0,d,1)-GARCH(1,1)

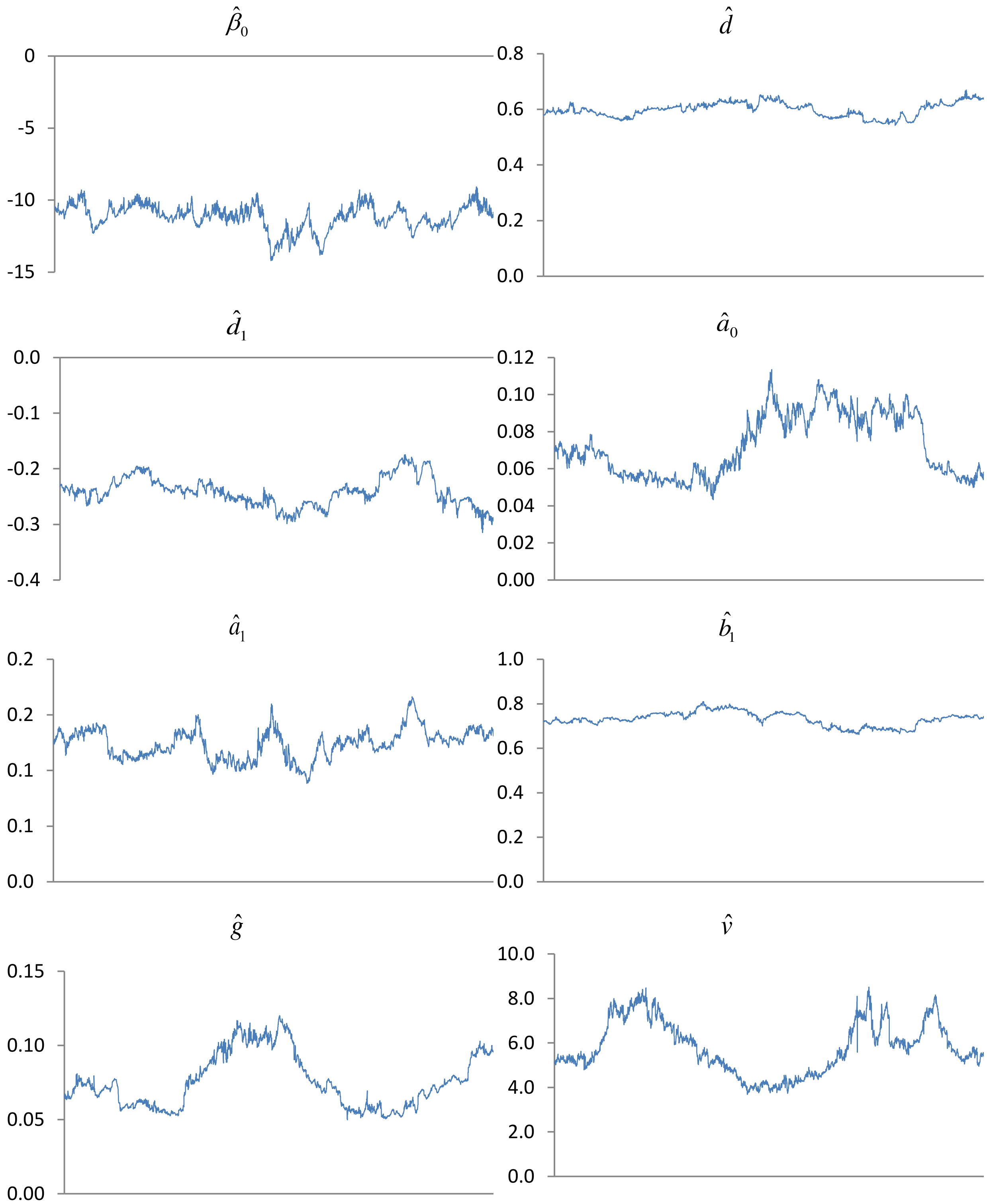
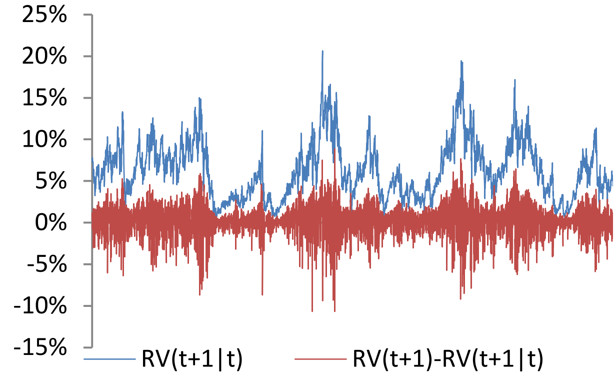


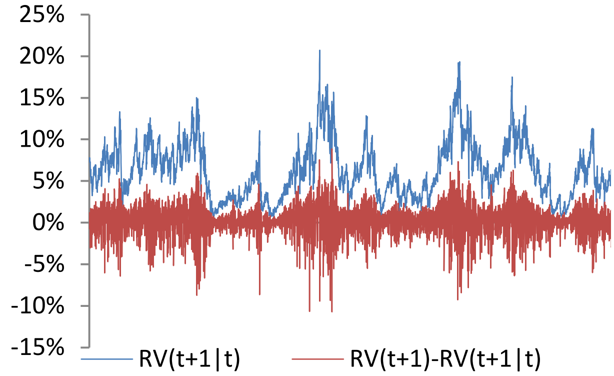
Figure 6. The data generated $RV_{t+1}^{(\tau)}$ against the forecasts $RV_{t+1|t}^{(\tau)}$, and the standardized forecast errors, $z_{t+1\backslash t}$. The models are estimated for conditionally skewed Student t distributed residuals.

Model 1 - ARFIMA(0,d,1)-GARCH(1,1)	Model 2 - ARFIMA(1,d,1)-GARCH(1,1)

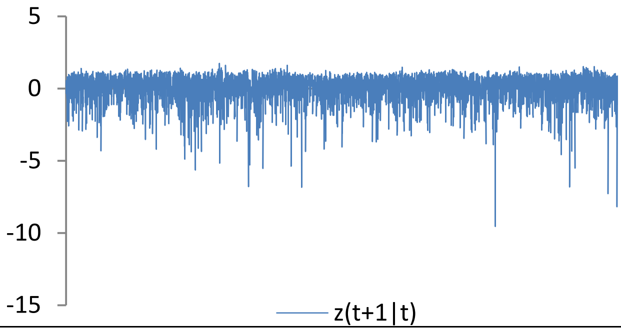
$RV_{t+1|t}^{(\tau)}$ and $\left(RV_{t+1}^{(\tau)} - RV_{t+1|t}^{(\tau)}\right)$



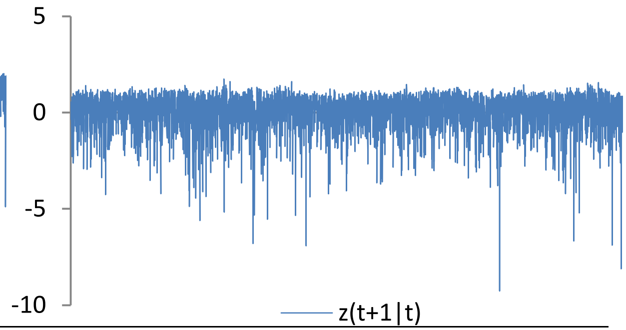
$RV_{t+1|t}^{(\tau)}$ and $\left(RV_{t+1}^{(\tau)} - RV_{t+1|t}^{(\tau)}\right)$



$z_{t+1\backslash t}$

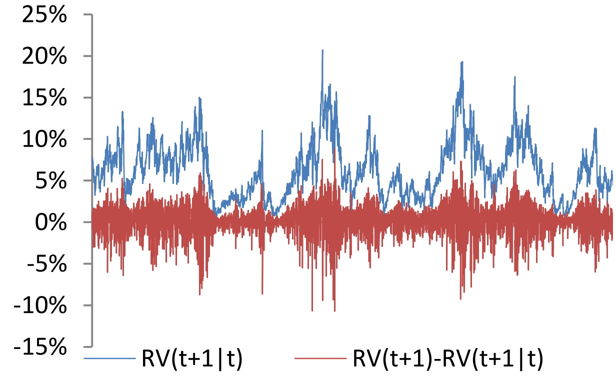


$z_{t+1\backslash t}$

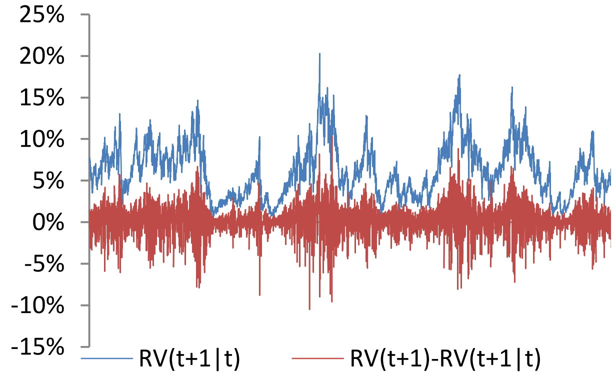


Model 3 - HAR-RV-GARCH(1,1)	Model 4 - HAR-RV-GARCH(0,1)

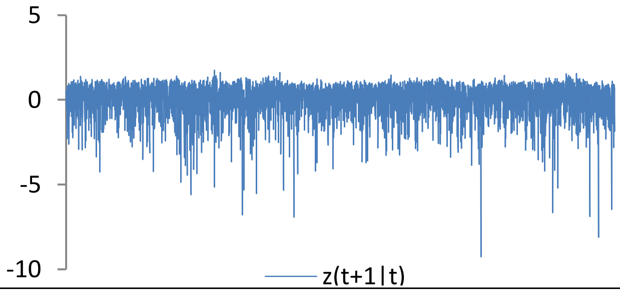
$RV_{t+1|t}^{(\tau)}$ and $\left(RV_{t+1}^{(\tau)} - RV_{t+1|t}^{(\tau)}\right)$



$RV_{t+1|t}^{(\tau)}$ and $\left(RV_{t+1}^{(\tau)} - RV_{t+1|t}^{(\tau)}\right)$



$z_{t+1\backslash t}$



$z_{t+1\backslash t}$