Accepted Manuscript

Modeling returns volatility: Realized GARCH incorporating realized risk measure

Wei Jiang, Qingsong Ruan, Jianfeng Li, Ye Li

PII: S0378-4371(18)30094-3
DOI: https://doi.org/10.1016/j.physa.2018.02.018
Reference: PHYSA 19138

To appear in: *Physica A*

Received date : 1 February 2017
Revised date : 29 August 2017

Please cite this article as: W. Jiang, Q. Ruan, J. Li, Y. Li, Modeling returns volatility: Realized GARCH incorporating realized risk measure, *Physica A* (2018), https://doi.org/10.1016/j.physa.2018.02.018

Highlights

Modeling volatility by realized GARCH incorporating realized risk measure.

The risk information embedded into realized GARCH provides better volatility estimation and forecasting.

The realized expected shortfall performs best for all of the alternative realized measures.

The future volatility may be more attributable to risk measure.

Modeling Returns Volatility: Realized GARCH Incorporating Realized Risk Measure

Wei Jiang[a, *], Qingsong Ruan[a], Jianfeng Li[a], Ye Li[b]

[a]*School of Economics & Management, Tongji University, Shanghai 200092, China*
[b]*College of Transportation Engineering, Tongji University, Shanghai 201804, China*

Abstract:

This study applies realized GARCH models by introducing several risk measures of intraday returns into the measurement equation, to model the daily volatility of E-mini S&P 500 index futures returns. Besides using the conventional realized measures, realized volatility and realized kernel as our benchmarks, we also use generalized realized risk measures, realized absolute deviation, and two realized tail risk measures, realized value-at-risk and realized expected shortfall. The empirical results show that realized GARCH models using the generalized realized risk measures provide better volatility estimation for the in-sample and substantial improvement in volatility forecasting for the out-of-sample. In particular, the realized expected shortfall performs best for all of the alternative realized measures. Our empirical results reveal that future volatility may be more attributable to present losses (risk measures). The results are robust to different sample estimation windows.

Key words: High-frequency data; Volatility; Realized GARCH; Realized risk measures

1. Introduction

The estimation and forecasting of financial assets volatility are important in terms of both their theoretical and practical applications. Few would dispute that the autoregressive conditional heteroscedasticity (ARCH) model of Engle [1] and the generalized ARCH (GARCH) model of Bollerslev [2], together with their various extensions, are excellent tools for modeling and forecasting the dynamic features of condition volatility. The conventional GARCH-type models, which use the daily close price to infer conditional volatility, may omit certain useful intraday information; that is, the information set of the conventional GARCH-type models insufficiently accounts for informative $\mathscr{F}_{t\text{-}1}$. More recently, as the availability of high-frequency intraday data increases, a growing number of studies have introduced more accurate

[*] Corresponding author.
Email: 280760678@qq.com (W. Jiang).

realized measures of integrated volatility [3–8]. Meanwhile, the literature has introduced many models that incorporate these realized measures in order to improve the estimation of the in-sample data and the out-of-sample predictive ability of conditional volatility [9–15].

Hansen et al. [17] recently introduced realized GARCH (RG) model, a new framework which incorporates the realized measures calculated by intraday high-frequency returns within the modeling of conditional volatility. This application of this approach has been expanded to cover a considerable amount of financial literatures (such as [18–20] and the references therein). The RG model offers several advantages regarding volatility estimates and forecasts. First, RG is parsimonious, and an ARMA structure for conditional volatility and realized measures has been introduced. Second, RG easily estimates the parameters of returns, volatility and measurement equations simultaneously and improve the modeling and forecasting of future volatility. Third, RG may adjust the potential bias caused by non-trading hours and microstructure noise [18].

In fact, a crucial feature of RG is the selection of realized measure. The existing related studies of RG mainly employ realized variance (RV) by Andersen and Bollerslev [3], microstructure noise robust realized variance (e.g., realized kernel (RK) by Barndorff-Nielsen et al. [8]), and jump robust realized variance (e.g., realized bipower variation (BV) by Barndorff-Nielsen and Shephard [7]). It is widely accepted that realized variance and its improved versions provide more efficient nonparametric estimators of the latent volatility and are associated with future volatility. Consequently, these realized measures of volatility are usually employed in the previous research as the exogenous variable embedded into the RG.

However, not all of the realized measures of volatility are unbiased estimator of the latent volatility and the RG framework does not necessarily require the realized measure to be an unbiased estimate of volatility. More generally, the realized measure denotes a vector of an exogenous variable, which is associated with future volatility. In addition to realized variance, other realized risk measures may also affect future volatility. With respect to the measurement of risk in financial markets, except for the widely-used variance (volatility), absolute deviation and tail risk measures, value-at-risk (VaR) and expected shortfall (ES, also known as conditional VaR, CVaR), also constitute extremely popular tools. In particular, volatility may be more attributable to tail information in realistic financial markets. In fact, extreme price movements may lead to more extensive changes in volatility of the financial markets, and tail risk measures can capture extreme price movements.

In this study, we contribute to the empirical literature by applying the more general realized risk measures to replace the conventional realized variance as the realized measure of realized GARCH models. The realized risk measures include realized absolute deviation (RAD) and two realized tail risk measures, realized value at risk (RVaR) and realized expected shortfall (RES). We examine the fitness of these models using normal inverse Gaussian. Although the results are not reported in this paper, we also estimated the RG with the generalized distributions and the main empirical finding does not change regarding the inverse Gaussian distribution specification. Finally, we use the MCS procedure of Hansen et al. [21] to compare empirically the forecasting power delivered by the competing model specifications, the G and EG based on daily returns, RG with realized volatility (RG(RV)) and realized kernel (RG(RK)), and RG with three realized risk measures (RG(RAD), RG(RVaR), and RG(RES)).

This article is organized as follows: Section 2 reviews the RG framework and the estimation method, including the standard GARCH (G) and EGARCH (EG) as the benchmark, and the loss functions. In addition, we propose the model confidence set (MCS) procedure of Hansen et al. [21] to evaluate the 1-day-ahead volatility forecast performance. Section 3 details the data, the realized variance, and the realized risk measures. Section 4 covers the in-sample estimation results and the volatility forecasting performance. Section 5 concludes our paper.

2. Methodology

2.1. Realized GARCH model

We now briefly describe the RG model of Hansen et al. [17], with its log-linear specification. The RG(1,1) model is specified as follows

$$r_t = \sigma_t z_t, z_t \sim i.i.d.(0,1), \tag{1}$$

$$\log \sigma_t^2 = \omega + \beta \log \sigma_{t-1}^2 + \gamma \log x_{t-1} \tag{2}$$

$$\log x_t = \xi + \varphi \log \sigma_t^2 + \tau(z_t) + u_t, u_t \sim i.i.d.N\left(0, \sigma_u^2\right) \tag{3}$$

where $\{r_t\}$ is a zero-mean return series, $\sigma_t^2 = var(r_t | \mathcal{F}_{t-1})$, $\mathcal{F}_t = \sigma(r_t, x_t, r_{t-1}, x_{t-1}, \cdots)$ is the conditional volatility, and x_t is a realized measure of volatility. In this study, x_t denotes the various realized risk measures. The standardized error z_t and u_t are mutually independent. Then, $\tau(z_t)$ is the leverage function, which can be specified via the following simple quadratic form: $\tau(z_t)=\tau_1 z+\tau_2(z^2-1)$, where $E\tau(z_t)=0$, thus, following Engle and Ng [22]. The news impact curve is defined as $v(z)=E(\log\sigma^2_{t+1}|z_t = z) - E(\log\sigma^2_{t+1})$; that is, $v(z)=\gamma_1\tau(z)$.

The first two equations formulate a GARCH-X model and the X means that x_t is treated as an exogenous variable. In this paper, except for the conventional realized measures, RV and RK, we also use realized absolute deviation, realized value-at-risk, and realized expected shortfall as the realized measures. That is, generalized risk information can affect tomorrow's volatility (see Eq. (2)). At the same time, the condition volatility σ^2_t is driven by yesterday's σ^2_{t-1} and the innovation of return through realized measure x_t. Eq. (3) is the measurement equation, which combines the realized risk measures with conditional volatility. The measurement equation presents a simple way to investigate the joint dependence of r_t and x_t by the leverage function $\tau(z_t)$. Naturally, the logarithmic conditional volatility of RG(1,1) can be shown in an ARMA(1,1) process:

$$\log\sigma_t^2 = \omega + (\beta+\varphi\gamma)\log\sigma_{t-1}^2 + \gamma[\tau(z_{t-1}) + u_{t-1} + \xi] \tag{4}$$

The persistence parameter, π, is given by $\pi = \beta + \varphi\gamma$.

Following Hansen et al. [17], the quasi-maximum likelihood (QML) method is used to estimate the realized GARCH models. The log-likelihood function is

$$\log L(r_t, x_t;\theta\} = \sum_{t=1}^{n} \log f(r_t, x_t | \mathcal{F}_{t-1}) \tag{5}$$

$$f(r_t, x_t | \mathcal{F}_{t-1}) = f(r_t|\mathcal{F}_{t-1})\, f(x_t | r_t, \mathcal{F}_{t-1}) \tag{6}$$

The joint log likelihood of z_t and u_t for the Gaussian specification can be split into the sum

$$\begin{aligned} l(r,x) = &-\frac{1}{2}\sum_{t=1}^{n}[\log(2\pi)+\log(\sigma_t^2)+r_t^2/\sigma_t^2]+ \\ &-\frac{1}{2}\sum_{t=1}^{n}[\log(2\pi)+\log(\sigma_u^2)+u_t^2/\sigma_u^2] \end{aligned} \tag{7}$$

The partial log-likelihood $l(r)$ is given by

$$l(r) = -\frac{1}{2}\sum_{t=1}^{n}[\log(2\pi)+\log(\sigma_t^2)+r_t^2/\sigma_t^2] \tag{8}$$

Then, we can compare the fitness with that of the competing GARCH-type models.

2.2. Standard GARCH-type models

We use G and EG models as the benchmark models. A zero-mean $\{r_t\}$ is specified by

$$r_t = \sigma_t z_t \tag{9}$$

The conditional variance equations are specified as

G(1,1): $\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2$ (10)

EG(1,1): $\log \sigma_t^2 = \omega + \beta \log \sigma_{t-1}^2 + \alpha z_{t-1} + \gamma \left[|z_{t-1}| - E\left(|z_{t-1}|\right) \right]$ (11)

The log-likelihood for the Gaussian specification is

$$l(r) = -\frac{1}{2} \sum_{t=1}^{n} [\log(2\pi) + \log(\sigma_t^2) + r_t^2 / \sigma_t^2] \tag{12}$$

We can compare the fit of the competing models through the partial log-likelihood of RG and the log-likelihood of G and EG.

2.3. Forecast evaluation

A reliable model should not only compare the fitness for the data but also assess the accuracy of the forecasting power of the model. In the present work, we construct the following nine loss functions as our forecasting evaluation criteria, as Patton [24] and Ma et al. [25] discussed.[1]

MSE-SD: $L(\hat{\sigma}^2, \sigma^2) = (\hat{\sigma} - \sigma)^2$ (13)

MAE-SD: $L(\hat{\sigma}, \sigma) = |\hat{\sigma} - \sigma|$ (14)

MSE: $L(\hat{\sigma}^2, \sigma^2) = (\hat{\sigma}^2 - \sigma^2)^2$ (15)

MAE: $L(\hat{\sigma}^2, \sigma^2) = |\hat{\sigma}^2 - \log \sigma^2|$ (16)

HMSE: $L(\hat{\sigma}^2, \sigma^2) = (1 - \sigma^2 / \hat{\sigma}^2)^2$ (17)

HMAE: $L(\hat{\sigma}^2, \sigma^2) = |1 - \sigma^2 / \hat{\sigma}^2|$ (18)

R^2LOG: $L(\hat{\sigma}^2, \sigma^2) = (\log \hat{\sigma}^2 - \log \sigma^2)^2$ (19)

MAELOG: $L(\hat{\sigma}^2, \sigma^2) = |\log \hat{\sigma}^2 - \log \sigma^2|$ (20)

QLIKE: $L(\hat{\sigma}^2, \sigma^2) = \log \sigma^2 + \hat{\sigma}^2 / \sigma^2$ (21)

where $\hat{\sigma}^2$, and σ^2 is a volatility proxy and a volatility forecast, respectively. Since real volatility is unobservable, we use realized volatility and realized kernel as a proxy

[1] We thank the anonymous referee for suggesting that we employ more loss functions.

for condition volatility, respectively. Here, as the microstructure noise is robust, RK is an unbiased estimator of latent volatility.

2.4.The MSC procedure

In order to identify a model that significantly outperforms the competing models and figure out the 'best fitting model', we apply the model confidence set (MCS) procedure by Hansen et al. [21] to compare the forecasts generated by alternative model specifications and establish the 'superior set models' (SSM), where all models have equal predictive ability (EPA) at a given confidence level α.

We introduce how the MCS procedure is used to obtain the SSM, including the following steps:

(i) set an initial set of models $M^0=M$, containing all competing model specifications at a given confidence level α;

(ii) test EPA-hypothesis: if EPA is not rejected, set $M^*_{1-\alpha}=M$; otherwise, use the elimination rules defined in Hansen et al. [21] to confirm the worst model;

(iii) eliminate the worst model, and go to step 2; and

(iv) obtain the SSM (all models have equal predictive ability in SSM).

For a detailed discussion of the MCS, refer to Hansen et al. [21] and Bernardi and Catania [26].

3. Data Description

3.1. Data

Our sample is the minute-by-minute intraday transaction prices of E-mini S&P 500 futures contracts traded on the Chicago Mercantile Exchange (CME) for the study period from April 12, 2010 to December 31, 2015, including an in-sample period (April 12, 2010 to June 30, 2015) and a six month out-of-sample period (July 1, 2015 to December 31, 2015. We removed the data related to Sundays, since there were only one or two trading hours on that day. In this paper, E-mini S&P 500 index futures are preferred over S&P 500 index data for two reasons. First, since the index futures have almost no overnight and lunchtime effect, and we can measure more precisely the close-to-open return data over the index data. The correct measure of the close-to-open return is of crucial importance to model and forecast the market volatility. Second, the value of the existing S&P 500 contract is too large for many small traders, and the E-mini S&P 500 contract is the most popular equity index futures contract in the world.

3.2. Realized variance and realized tail risk measures

3.2.1. Realized variance and kernel

We adopt the realized variance of Andersen and Bollerslev [3] and the market microstructure noise robust realized kernel of Barndorff-Nielsen et al. [8] as the benchmark. The RV and RK measures for day t are given by

$$RV_t = \sum_{j=1}^{n} r_{t,j}^2 \tag{22}$$

$$RK_t = \sum_{h=-H}^{H} \kappa\left(\frac{h}{H+1}\right)\gamma_h \text{, where } \gamma_h = \sum_{j=|h|+1}^{n} r_{t,j} r_{t,j-|h|} \tag{23}$$

where $r_{t,j}$ is the jth intraday return, n is the total number of intraday returns for day t, H is the bandwidth parameter and $\kappa(x)$ is a kernel weight function. We adopt the Parzen kernel function in this study.

3.2.2. Generalized realized risk measures. Realized absolute deviation

We introduce absolute deviation instead of variance as the realized risk measures for day t. The realized absolute deviation is given by

$$RAD_t = \sum_{j}^{n} \left| r_{t,j} - E_j(r_{t,j}) \right| \tag{24}$$

where $r_{t,j}$ and n are the jth intraday return and the total number of intraday returns for day t, respectively.

Realized value-at-risk: The value-at-risk focus on the tail behavior of the asset can be defined as the maximal loss of a financial position over the time horizon with a given tail probability from the viewpoint of a financial institution, assuming that the returns distribution remains unchanged during both the forecast and the in-sample periods, and that we can use the sample quantile of the intraday returns to calculate VaR. The realized VaR for day t is

$$RVaR_t = \sqrt{n} \cdot x_p^t \tag{25}$$

where n, and x^t_p are the total number of intraday returns and the pth quantile of intraday returns for day t, respectively. We set p=0.05 in this study.

Realized expected shortfall: In practice, VaR may underestimate the real loss. To assess the potential loss with greater accuracy, the expected shortfall is introduced and defined as the expected value of the loss if the VaR is exceeded. The realized expected shortfall for day t is

$$RES_t = \sqrt{n} E_j(r_{t,j} \mid r_{t,j} > x_p^t) \quad (26)$$

where $r_{t,j}$, n, and x_p^t are the jth intraday return, the total number of the intraday returns, and the pth quantile of intraday returns for day t, respectively. We set p=0.05.

In this study, we use negative returns for a long financial position in the data analysis. For more details about the extreme values, quantiles, and value at risk, see Tsay [27].

3.3.Descriptive statistics

Table 1 reports the descriptive statistics for the returns and logarithm realized measures time series used in the empirical application. In Table 1, the values for skewness and kurtosis show that all series possess the significant features of auto-correlations, skewed, and leptokurtic, thus supporting the view that a higher number of actual distributional assumptions can be conducive to enhanced model performance. In this paper, we use the normal inverse Gaussian distribution.

Fig. 1 plots the daily series of returns, two conventional realized measures, and three generalized realized risk measures.

Table 1 Descriptive statistics of returns and logarithm realized measures time series

	Mean	Maximum	Minimum	Std. dev.	Skewness	Kurtosis	JB Test	LB Test
returns	0.038	-5.845	6.206	1.001	-0.338	4.385	0.000	0.000
Ln(RV)	-2.522	-7.157	0.949	0.777	0.516	1.832	0.000	0.000
Ln(RK)	-2.808	-7.235	1.071	0.898	0.386	1.180	0.000	0.000
Ln(RAD)	5.372	1.282	7.161	0.418	-0.169	6.772	0.000	0.000
Ln(RVaR)	2.656	1.499	4.327	0.415	0.273	1.055	0.000	0.000
Ln(RES)	3.057	1.517	4.780	0.398	0.599	1.051	0.000	0.000

Notes: Descriptive statistics of returns and logarithm realized measures time series used in the empirical application. “JB test” and “LB test” denotes Jarque Bera test and the10*th* orders Ljung and Box test, with the corresponding p-values in the column.

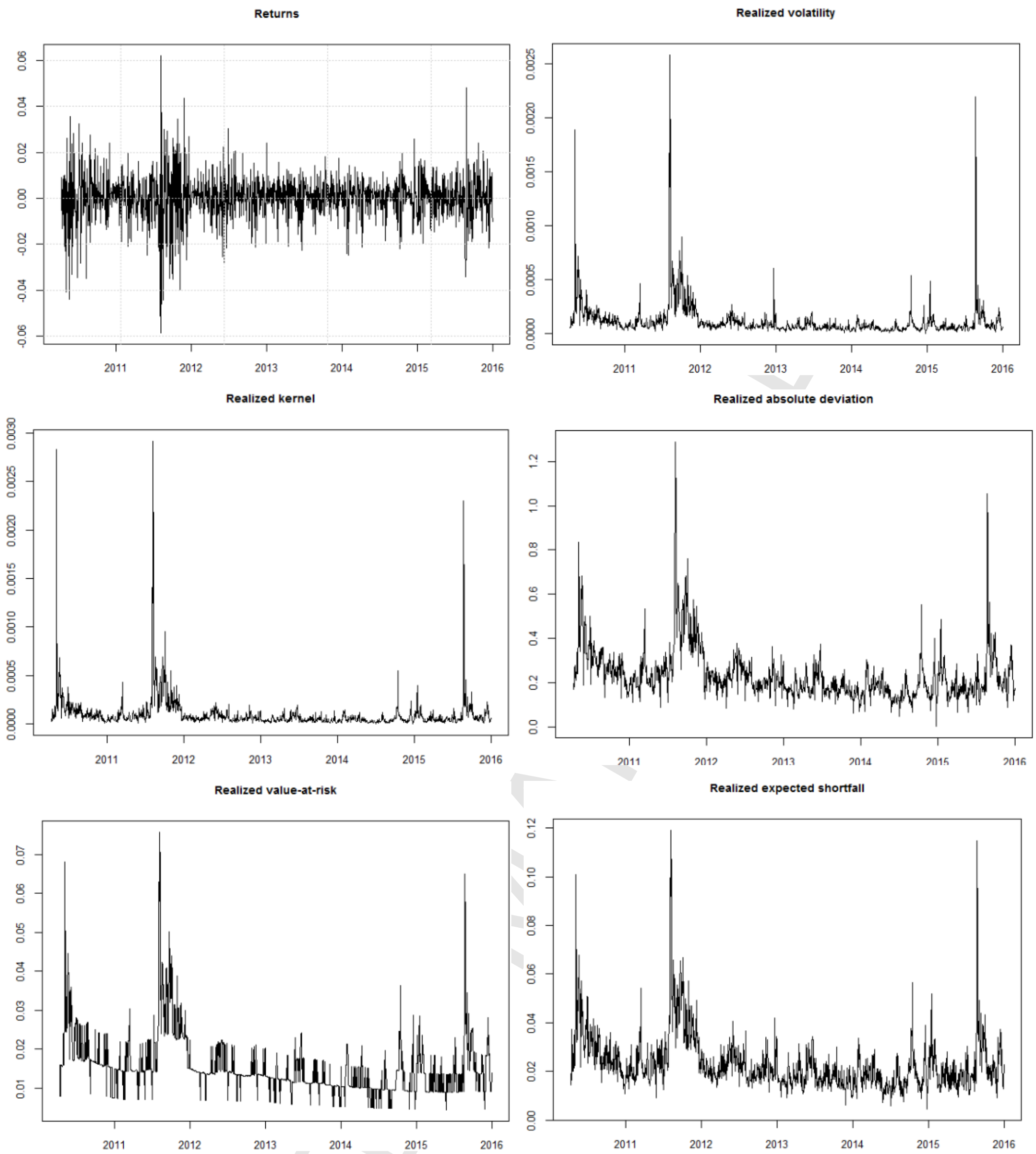

Fig. 1. The daily returns, realized risk measures series

Notes: This figure shows the daily returns, realized volatility, realized kernel, realized absolute deviation, realized value-at-risk, and realized expected shortfall, consecutively.

4. Empirical analysis

4.1. Estimation results

Table 2 reports the in-sample estimation results for the alternative models with a normal inverse Gaussian distribution for the E-mini S&P 500 index futures returns. Although the results are not reported in this paper, we also estimated the RG model with the generalized distributions and the main empirical finding remained unchanged.[2] Therefore, we explain the results of the inverse Gaussian distribution in what follows. From the empirical results (see, Table 2), we see that all of the parameters were highly significant at the 1% confidence level. The persistence parameter of volatility, $\pi=\beta+\varphi\gamma$, exceeds 0.92 no matter which model is used, which shows a strong persistence in volatility and covariance stationarity in returns. Judging from the partial log-likelihood function, $l(r)$, the RG model clearly dominates the standard G and EG models. In terms of joint log-likelihood values, $l(r, x)$, three RG models using generalized realized risk measures clearly outperform the RG with RV and RK. The AIC and BIC also suggest that the RG models with realized risk measures are appropriate. In particular, the RG(RES) model fits the data best, because the additional tail innovation of the intraday return embodied in the intraday measures improves the fitness for the daily data.

In Table 2, we see that the estimates of φ are close to 1 for both RV and RK, the conventional realized measures. This suggests that the realized measure, RV or RK, is roughly proportional to the conditional variance. For the three generalized risk measures, φ are further away from 1 compared with RV and RK. This is normal, because RAD/RVaR/RES is not the realized measure of volatility. However, the estimates of γ for the risk measures are close to 1 which is greater than those for both RV and RK. This suggests that the present realized risk measure is a larger proportion of the future volatility with respect to the realized variance and the realized kernel. This implies that today's greater losses are likely to lead to tomorrow's greater volatility. Moreover, the negative values of parameters τ_1 and the positive ones for τ_2 confirm the leverage effect in the futures markets.

Fig. 2 shows the news impact curve from the RG(RES) model defined in Section 2.1, where the vertical axis is the standard deviation σ_t, and the horizontal axis is standard innovation z_{t-1}. The news impact curve is strikingly unsymmetrical, indicating that volatility is more strongly impacted by negative price shocks.

[2] Thanks to an anonymous reviewer for the suggestion that employing more types of distribution dilutes the contribution of the current paper. Consequently, we only report the results of inverse Gaussian distribution, that provides the best fitness for the in-sample across all distributions. The generalized distributions include the standard normal, standard skew normal, Student's t, skew Student's t, generalized error, skew generalized error, and Johnson SU distributions. Interested readers are referred to Ghalanos [23] for more details.

Table 2. Empirical results for the competing models

Model			RV	RK	RAD	RVaR	RES
	G(1,1)	EG(1,1)	RG (1,1)	RG (1,1)	RG (1,1)	RG(1,1)	RG(1,1)
ω	0.03	-0.01	3.19	2.94	-3.24	-0.53	-0.94
	(0.01)	(0.01)	(0.3)	(0.26)	(0.39)	(0.05)	(0.08)
α	0.18	-0.28					
	(0.03)	(0.03)					
β	0.80	0.94	0.38	0.39	0.49	0.52	0.52
	(0.03)	(0.02)	(0.05)	(0.05)	(0.08)	(0.04)	(0.04)
γ		0.16	0.71	0.62	1.00	1.00	1.00
		(0.04)	(0.07)	(0.06)	(0.11)	(0.09)	(0.09)
ξ			-4.51	-4.77	3.22	0.51	0.93
			(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
φ			0.79	0.87	0.46	0.43	0.44
			(0.04)	(0.04)	(0.02)	(0.02)	(0.02)
τ_1			-0.11	-0.13	-0.05	-0.09	-0.11
			(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
τ_2			0.06	0.09	0.04	0.02	0.02
			(0.01)	(0.01)	(0.01)	(0)	(0.00)
skew	-0.26	-0.33	-0.41	-0.39	-0.36	-0.36	-0.41
	(0.05)	(0.05)	(0.07)	(0.07)	(0.06)	(0.06)	(0.07)
shape	1.83	2.24	4.56	4.2	2.67	3.53	4.35
	(0.48)	(0.56)	(1.65)	(1.45)	(0.78)	(1.13)	(1.59)
σ_u			0.38	0.47	0.24	0.22	0.19
			(0.02)	(0.02)	(0.04)	(0.01)	(0.01)
$l(r,x)$			-2195.03	-2503.62	-1472.98	-1474.43	-1232.36
$l(r)$	-1660.27	-1607.42	-970.38	-970.83	-989.76	-985.81	-970.47
π	0.976	0.939	0.939	0.928	0.943	0.954	0.955
AIC	2.465	2.388	3.264	3.721	2.195	2.198	1.840
BIC	2.485	2.412	3.303	3.760	2.234	2.236	1.878

Notes: This table presents the empirical results for the competing models with a normal inverse Gaussian distribution for E-mini S&P 500 index futures returns. G, EG, RG denote the GARCH, EGARCH, and realized GARCH model, respectively. RV, RK, RAD, RVaR, RES denote RG using realized variance, realized kernel, realized absolute deviation, realized value-at-risk, and realized expected shortfall, respectively, which are calculated from the intraday high-frequency E-mini S&P 500 index futures returns. The robust standard errors are shown in parentheses. *l(r), l(r, x)* and π denote the partial likelihood, joint log-likelihood, and persistence parameter, respectively. AIC and BIC denote the Akaike and Bayesian information criteria, respectively.

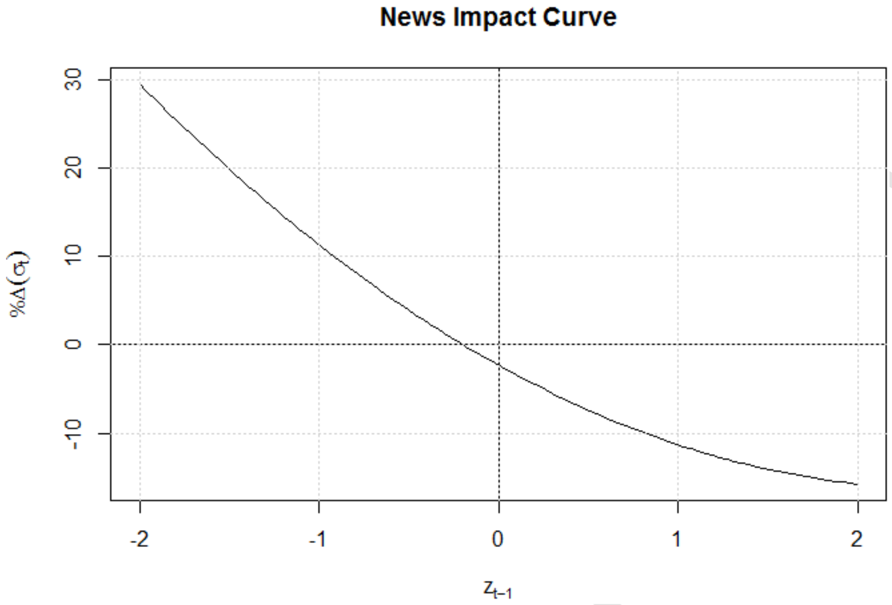

Fig. 2. News impact curve from the RG(RES) model

4.2. Forecasting evaluation

Following the previous estimation results, we apply the rolling estimate method with normal inverse Gaussian distribution to estimate the one-step-ahead conditional volatility for the out-of-sample. Fig. 3 shows the forecasting results in which the realized kernel is used as the forecasting benchmark.

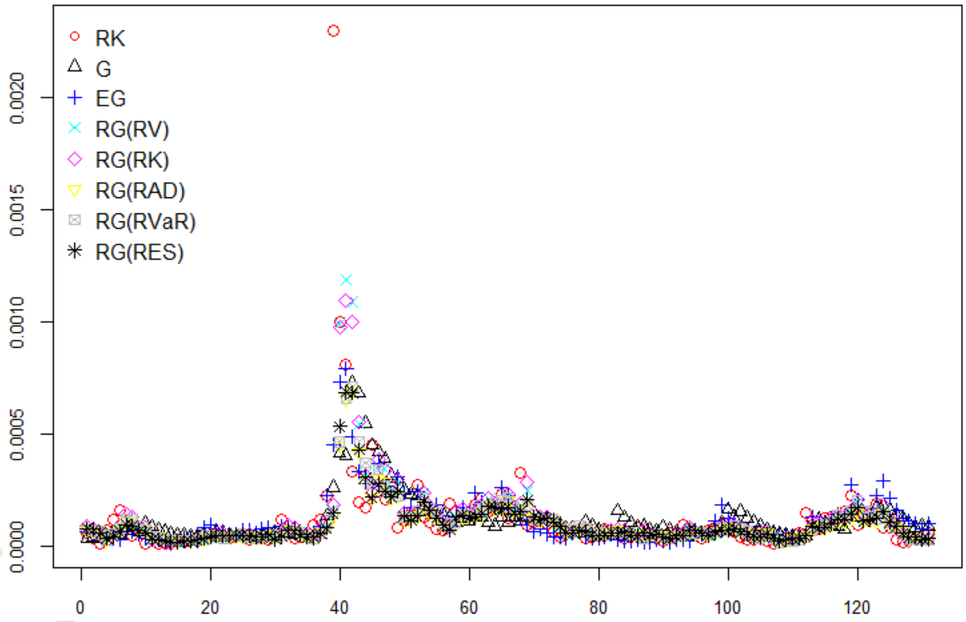

Fig. 3. Realized kernel and one-day ahead volatility forecasts from the competing model

As mentioned in Section 2, the forecasting performance of volatility is evaluated by constructing the loss functions. Then, the MCS procedure is used to test the null hypothesis of equal predictive ability and we can obtain the 'superior set models' (SSM). The EPA hypothesis is at the 10% confidence level. Here, we use 10, 000 bootstrap replications to evaluate the MCS *p*-values. Since it is unclear which loss function is most appropriate for the evaluation of the models' accuracy in forecasting volatility [24–25], the overall results are summarized using a performance measure which considers the percentage of acceptances of the null hypothesis across the different loss functions.

Table 3 reports the results of the MCS evaluation for nine loss functions, with realized volatility and realized kernel as the volatility proxies, respectively. The entries in the first row represent the different loss functions, and the first column shows the volatility models; namely the G, EG, RG(RV), RG(RK), RG(RAD), RG(RVaR), and RG(RES) models. Models with 'ELI' in the column were eliminated under the given confidence level using the elimination rule of Hansen et al. [21]. 'Grade comparison' in the column denotes the percentage of acceptances of the null hypothesis across the different loss functions.

Firstly, we observe that the G model is most frequently eliminated, and that the *p*-value of EG in SSM is the smallest value frequently, whether the proxy is realized volatility or realized kernel, across all loss functions, suggesting that RG models are more appropriate than E and EG models. Therefore, it is proper to consider the RG imbedded within intraday high-frequency returns to provide better volatility forecasting for the out-of-sample.

Furthermore, for the realized volatility as a volatility proxy, RAD and RES as the realized measure of the RG model perform best. Using the realized kernel as a volatility proxy, it becomes clear that the RES as the realized measure has the best predictive ability. Individually, only the MCS *p*-value of the RG(RES) model is equal to 1 both RV and RK as the volatility proxy across all loss function criteria. Therefore, we can conclude that the clear predominance of the RG(RES) model implies that the tail risk information plays a crucial role in forecasting volatility, which suggests that the RG models embedded the realized tail risk measures can improve the forecasting performance of condition volatility. Moreover, using the realized kernel as a volatility proxy, RG(RV) and RG(RK) usually have the smallest *p*-value across all loss functions, implying that considering more general realized risk measures as the realized measure of the RG model can improve the forecasting performance of condition volatility.

Table 3 The MCS *p*-value of competing models for loss function

	MSE-SD	*MAE-SD*	*MSE*	*MAE*	*HMSE*	*HMAE*	R^2LOG	*MAELOG*	*QLIKE*	*Grade comparison*
Volatility proxy: realized volatility										
G	ELI	ELI	ELI	ELI	ELI	ELI	ELI	ELI	ELI	0.000%
EG	1.000	0.326	1.000	1.000	ELI	ELI	0.152	0.147	1.000	77.78%
RG(RV)	0.184	ELI	0.127	0.206	ELI	ELI	1.000	1.000	1.000	66.67%
RG(RK)	1.000	1.000	1.000	1.000	ELI	ELI	1.000	1.000	1.000	77.78%
RG(RAD)	1.000	1.000	0.988	1.000	0.110	0.371	1.000	1.000	1.000	100.0%
RG(RVaR)	1.000	1.000	1.000	1.000	ELI	ELI	1.000	1.000	0.108	77.78%
RG(RES)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	100.0%
Volatility proxy: realized kernel										
G	ELI	ELI	1.000	ELI	ELI	ELI	ELI	ELI	ELI	11.11%
EG	1.000	0.274	1.000	1.000	ELI	ELI	0.143	ELI	1.000	66.67%
RG(RV)	0.205	1.000	0.151	0.275	0.190	ELI	1.000	ELI	1.000	77.78%
RG(RK)	1.000	1.000	1.000	0.213	ELI	ELI	1.000	ELI	1.000	66.67%
RG(RAD)	1.000	1.000	1.000	1.000	ELI	ELI	1.000	ELI	0.332	66.67%
RG(RVaR)	1.000	1.000	1.000	1.000	ELI	ELI	1.000	ELI	0.901	66.67%
RG(RES)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	100.0%

Notes: This table presents the minimum *p*-values of the TR statistics by the MCS procedure evaluation for the different loss functions with realized volatility and realized kernel as the volatility proxies, respectively. G, EG, RG(RV), RG(RK), RG(RAD), RG(RVaR), and RG(RES) denote the standard GARCH, EGARCH, and the realized GARCH model using realized variance, realized kernel, realized absolute deviation, realized value-at-risk, and realized expected shortfall, respectively. 'ELI' denotes that the corresponding model was eliminated by the MCS elimination rule under the given 10% confidence level. The underlined value in the column is the smallest *p*-value across that SSM. 'Grade comparison' in the column denotes the percentage of acceptances of the null hypothesis across the different loss functions.

4.3. The robustness to the length of the estimation window

Since the different lengths of the estimation window could lead to different empirical results, for a robustness check, we use the 'recursive' and 'moving' rolling window method based on a fixed length of 1351 observations (the size of the in-sample). Furthermore, we adjust the fixed length of the estimation window to 500 and 1000 and reevaluated the forecasting performance of the out-of-sample. The results reveal that adjusting the length of the estimation window does not change our main empirical results; that is, it is appropriate to consider that using the generalized realized risk measure as the realized measure of the RG model provides better volatility estimation and forecasting.

5. Conclusion

In this study, we have proposed realized GARCH models with more general realized risk measures to model the E-mini S&P 500 index futures returns. The crucial empirical results are as follows. First, the realized GARCH models with more general realized risk measures constitute an appropriate method for producing a better empirical fit for the in-sample and a substantial improvement in volatility forecasting for the out-of-sample. The models based on realized risk measures have more informative $\mathscr{F}_{t-1}$ than the standard G and EG models based on daily returns. Second, RG models using realized expected shortfall, RG(RES), yield an impressive performance regarding both empirical fit and volatility forecasting. A possible explanation for this is that the tail risk behavior contains useful information about future volatility and provides more accurate volatility forecasting, as RES can capture more tail information. Our empirical results are robust to different sample estimation windows.

Finally, in this study, we empirically analyze several realized risk measures but readily extend this to handle almost all risk measures, such as the beta risk, the lower partial moment framework of Fishburn [28], the MiniMax model of Young [29], etc. Similarly, our study may naturally be extended to assess other financial markets. Overall, a variety of intraday risk movements should be considered when financial institutions require more accurate volatility, in order to allocate assets and ensure investment security when faced with extreme market movements.

Acknowledgements

The authors sincerely thank the anonymous reviewers for their many constructive comments and suggestions. The authors also thank the editors.

References

[1] R. F. Engle, Autoregressive Conditional Heteroscedasticity with estimates of the variance of United Kingdom inflation, Econom. 50 (1982) 987–1007.

[2] T. Bollerslev, Generalized autoregressive heteroskedasticity, J. Econom. 31 (1986) 307–327.

[3] T. G. Andersen, T. Bollerslev, Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, Int. Econ. Rev. 39 (4) (1998) 885–905.

[4] T. G. Andersen, T. Bollerslev, F. X. Diebold, P. Labys, Modeling and forecasting realized

volatility, Econom. 71 (2) (2003) 579–625.

[5] T. G. Andersen, D. Dobrev, E. Schaumburg, Jump-robust volatility estimation using nearest neighbor truncation, J. Econom. 169 (1) (2012) 75–93.

[6] OE Barndorff-Nielsen, N. Shephard, Econometric analysis of realised volatility and its use in estimating stochastic volatility models, J. Roy. Stat. Soc.: Ser. B (Stat. Methodol.) 64 (2) (2002) 253–280.

[7] OE Barndorff-Nielsen, N. Shephard, Power and bipower variation with stochastic volatility and jumps, J. Financ. Econom. 2 (1) (2004) 1–37.

[8] OE Barndorff-Nielsen, P. R. Hansen, A. Lunde, N. Shephard, Designing realized kernels to measure the ex post variation of equity prices in the presence of noise, Econom. 76 (6) (2008) 1481–1536.

[9] R. F. Engle, New frontiers for arch models, J. Appl. Econom. 17 (5) (2002) 425–446.

[10] P. Giot, S. Laurent, Modelling daily value-at-risk using realized volatility and ARCH type models. J. Empir. Financ. 11 (3) (2004) 379-398.

[11] R. F. Engle, G. M. Gallo, A multiple indicators model for volatility using intra-daily data, J. Econom. 131 (2006) 3–27.

[12] N. Shephard, K. Sheppard, Realising the future: forecasting with high-frequency-based volatility (HEAVY) models, J. Appl. Econom. 25 (2) (2010) 197–231.

[13] M. Takahashi, Y. Omori, T. Watanabe, Estimating stochastic volatility models using daily returns and realized volatility simultaneously, Stat. Data Anal. 53 (6) (2009) 2404–2426.

[14] D. P. Louzis, S. Xanthopoulos-Sisinis, A. P. Refenes, Realized volatility models and alternative Value-at-Risk prediction strategies. Econ. Model. 40 (2014) 101-116.

[15] S. Shirota, T. Hizu, Y. Omori, Realized stochastic volatility with leverage and long memory, Comput. Stat. Data Anal. 76 (2014) 618–641.

[16] M. Bee, D. J. Dupuis, L. Trapin, Realizing the extremes: Estimation of tail-risk measures from a high-frequency perspective. J. Empir. Financ. 36 (2016): 86-99.

[17] P. R. Hansen, Z. Huang, H. H. Shek, Realized GARCH: A joint model for returns and realized measures of volatility, J. Appl. Econom. 27(6) (2011) 877–906.

[18] T. Watanabe, Quantile forecasts of financial returns using realized GARCH models, Jpn. Econ. Rev. 63 (1) (2012) 68–80.

[19] S. Tian, S. Hamori, Modeling interest rate volatility: A Realized GARCH approach, J. Bank Financ. 61 (2015) 158–171.

[20] Z. Huang, T. Wang, P. R. Hansen, Option pricing with the realized GARCH model: an analytical approximation approach, J. Futures Mark., forthcoming. doi: 10.1002/fut.21821.

[21] P. R. Hansen, A. Lunde, J. M. Nason, The model confidence set, Econom. 79(2) (2011) 453–497.

[22] R. F. Engle, V. K. Ng, Measuring and testing the impact of news on volatility, J. Finance 48 (5) (1993) 1749–1778.

[23] A. Ghalanos, rugarch: Univariate GARCH models. R package version 1.3–6, 2015. Available online at: http://CRAN.R-project.org/package=rugarch.

[24] A. J. Patton, Volatility forecast comparison using imperfect volatility proxies, J. Econom. 160 (1) (2011) 246–256.

[25] F. Ma, Y. Wei, D. Huang, Y. Chen, Which is the better forecasting model? A comparison between HAR-RV and multifractality volatility. Physica A: Statistical Mechanics and its Applications 405 (2014) 171-180.

[26] M. Bernardi, L. Catania, Comparison of Value-at-Risk models using the MCS approach, Comput. Stat. 31 (2) (2016) 579–608.

[27] R. S. Tsay, Analysis of financial time series, 3rd ed., John Wiley, 2010.

[28] P. C. Fishburn, Mean-risk analysis with risk associated with below-target returns, Am. Econ. Rev. 67 (2) (1977) 116-126.

[29] M. R. Young, A minimax portfolio selection rule with linear programming solution, Manage. Sci. 44 (5) (1998) 673-683.