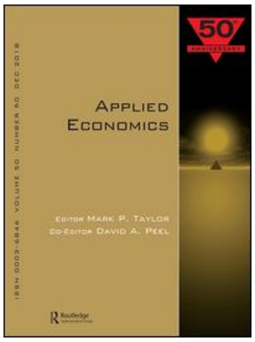

Applied Economics

ISSN: 0003-6846 (Print) 1466-4283 (Online) Journal homepage: https://www.tandfonline.com/loi/raec20

Volatility forecasting: long memory, regime switching and heteroscedasticity

Feng Ma, Xinjie Lu, Ke Yang & Yaojie Zhang

APPLIED ECONOMICS
https://doi.org/10.1080/00036846.2019.1589645

Volatility forecasting: long memory, regime switching and heteroscedasticity

Feng Ma[a], Xinjie Lu[a], Ke Yang[b] and Yaojie Zhang[c]

[a]School of Economics & Management, Southwest Jiaotong University, Sichuan, China; [b]School of Economics and Commerce, South China University of Technology; [c]School of Economics and Management, Nanjing University of Science and Technology, Nanjing, China

ABSTRACT

In this article, we account for the first time for long memory, regime switching and the conditional time-varying volatility of volatility (heteroscedasticity) to model and forecast market volatility using the heterogeneous autoregressive model of realized volatility (HAR-RV) and its extensions. We present several interesting and notable findings. First, existing models exhibit significant nonlinearity and clustering, which provide empirical evidence on the benefit of introducing regime switching and heteroscedasticity. Second, out-of-sample results indicate that combining regime switching and heteroscedasticity can substantially improve predictive power from a statistical viewpoint. More specifically, our proposed models generally exhibit higher forecasting accuracy. Third, these results are widely consistent across a variety of robustness tests such as different forecasting windows, forecasting models, realized measures, and stock markets. Consequently, this study sheds new light on forecasting future volatility.

KEYWORDS
Volatility forecasting; realized volatility; long memory; regime switching; heteroscedasticity

JEL CLASSIFICATION
G15; C22

I. Introduction

Modeling and forecasting of stock market volatility are of great interest to scholars, financial market participants and policymakers, as the volatility is a core input of derivative pricing, hedging, portfolio selection, and risk management (see, e.g., Andersen, Bollerslev, and Diebold 2007; Bollerslev et al. 2017). Although there are a large number of papers on volatility forecasting, accurately forecasting volatility is still a daunting task.

The seminal work of Bollerslev (1986) proposes generalized autoregressive conditional heteroskedasticity (GARCH) to model and forecast volatility. This model and its various extensions have become popular econometric models that describe and predict financial market volatility (e.g., Hansen and Lunde 2005; Harrison and Moore 2012; Wei, Wang, and Huang 2010). It is worth noting that those models usually employ daily or lower-frequency data, which could result in a substantial loss of intraday trading information (Carnero, Peña, and Ruiz 2004; Corsi 2009; Ma et al. 2018). Additionally, the GARCH-class models have lower predictive power compared to volatility models based on high-frequency data such as Koopman, Jungbacker, and Hol (2005), Martens and Zein (2004), Wei, and Wang, and Huang (2010).

Interestingly, with the availability of abundant high-frequency (intraday) data, modeling and forecasting of financial market volatility have recently taken a new direction. Specifically, high-frequency data contain a wealth of information that can help market participants make quicker decisions. More important, an influential study by Andersen and Bollerslev (1998) constructs volatility using high-frequency data, dubbed realized volatility (RV), [1] which is robust to market microstructure effects from both a theoretical and an empirical point of view. The chief advantage of RV is that it is directly observable, thereby enabling researchers to measure it and understand its dynamics. To the best of our knowledge, many scholars have built a large number of volatility models to describe and forecast realized volatility (see, e.g., Andersen et al. 2003; Corsi, 2009; Hansen, Huang, and Shek 2012; Ghysels, Santa-Clara, and Valkanov 2006, 2006; Shephard and Sheppard 2010; Engle and Gallo 2006).

In this paper, we use the heterogeneous autoregressive model of realized volatility (HAR-RV)

CONTACT Ke Yang yangkdc@scut.edu.cn School of Economics and Commerce, South China University of Technology

[1]We will use the terms realized volatility and realized variation (variance) interchangeably, as do Andersen, Bollerslev, Diebold (2007), and Ma et al. (2018).

and its extended model – named the HARQ-RV – that accounts for measurement errors and was proposed by Bollerslev, Patton, and Quaedvlieg (2016). There are two primary reasons for this choice. On the one hand, the HAR-RV model was proposed by Corsi (2009) based on the hypothesis of heterogeneous markets, which capture such 'stylized facts' of financial markets as long memory and multi-behavior. On the other hand and more importantly, this model can be easily estimated and has good forecasting performance. In recent years, this model has been prevalently applied to forecast realized volatility because of its related advantages (see, for example, Andersen, Bollerslev, and Diebold 2007; Bekaert and Hoerova 2014; Bollerslev, Patton, and Quaedvlieg 2016; Clements and Liao 2017; Duong and Swanson 2015; Patton and Sheppard 2015; Wang, Ma, Wei, and Wu 2016).

To the best of our knowledge, a large body of papers have used the linear HAR-RV and its extended models to predict future volatility, whereas previous studies (Goldman et al. 2013; Ma et al. 2017; Raggi and Bordignon 2012; Tian, Yang, and Chen 2017) have found that the coefficients of those linear models are not constant. This study thus makes three contributions to the literature on modeling and forecasting realized volatility. The first contribution is to consider the nonlinear and highly persistent dynamics of realized volatility and then to introduce regime switching to the HAR-RV and HARQ-RV models. Cipollini, Gallo, and Otranto (2017), McAleer and Medeiros (2008), Tian, Yang, and Chen (2017), among others, indicate that there exists a higher level of persistence when volatility is low, implying the presence of nonlinearities. It is worth noting that because of business cycles, major events, economic policy, and other factors, the statistical property of volatility (e.g., volatility persistence) always undergoes structural breaks (e.g., Banerjee and Urga 2005) or switches between different regimes (Wang, Ma, Wei, and Wu, 2016; Ma et al. 2017). Consequently, introducing regime switching to realized volatility models is justified.

The second contribution of this paper is to take into account the conditional time-varying volatility of realized volatility (heteroscedasticity) using the FIGARCH model. Notably, only a small amount of literature has considered the volatility of realized volatility (VOV) when forecasting future realized volatility; see, for example, Beltratti and Morana (2005), Corsi et al. (2008). These studies find that accounting for time-varying volatility of realized volatility substantially improves the goodness-of-fit as well as the model's predictive power. Thus, in line with Beltratti and Morana (2005), we use the FIGARCH model (Baillie, 1996) to describe the time-varying volatility clustering of residual errors of the HAR-RV and HARQ-RV models.

The third and most important contribution of this paper is to introduce for the first time regime switching and heteroscedasticity (VOV) to the HAR-RV and its extended models and thereby construct new forecasting models. Our new models not only capture long memory but also include regime switching and heteroscedasticity. To evaluate whether these new models outperform the existing models in forecasting volatility, we used the 5-minute rule-of-thumb sampling frequency of the CSI 300 index futures to carry out our analysis.

We present several noteworthy findings. First, the residuals of the HAR-RV and HARQ-RV models exhibit significant clustering, which supports our choice to model the volatility of realized volatility (heteroscedasticity). Additionally, the likelihood-ratio statistic test indicates that existing models obviously reject the linear hypothesis, implying that accounting for regime switching of the two models is justified. The high volatility regime of lagged RV has more powerful impacts on future volatility than the low volatility regime. Second, out-of-sample results show that introducing regime switching and heteroscedasticity can substantially increase forecasting accuracy. More specifically, the MS-HARQ-FIGARCH model has more power to forecast volatility. Third, these results are widely consistent across a variety of robustness tests, such as different forecasting windows, forecasting models, realized measures, and stock markets.

The rest of the paper is organized as follows. Section 2 describes the realized measures and the econometrics models. Section 3 provides the data and some preliminary analysis. The evaluated method, empirical results, and various robustness tests are presented in Section 4. Section 5 concludes.

II. Econometric models

Realized measures

We begin by introducing some important concepts, namely, realized variance/volatility (RV) proposed by Andersen and Bollerslev (1998) and realized bipower variation (BPV) proposed by Barndorff-Nielsen and Shepherd (2004). For a given day t, we divide the time interval (e.g., [0, 1]) into n equal subintervals, where $M = 1/\Delta$ and Δ is the sampling frequency. Consequently, RV can be defined as the sum of all available intraday high-frequency squared returns:

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

where $r_{t,j}$ represents the intraday returns. Barndorff-Nielsen and Shephard (2004) theoretically prove that when $\Delta \to 0$, RV converges to

$$\mathrm{RV}_t \to \int_0^t \sigma^2(s)ds + \sum_{0<s\leq t} \kappa^2(s), \tag{2}$$

where $\int_0^t \sigma^2(s)ds$ is the integrated variance and can be calculated by BPV as follows:

$$\mathrm{BPV}_t = u_1^{-2} \sum_{j=2}^{1/\Delta} |r_{t,j}||r_{t,j-1}|, \tag{3}$$

$u_1 \simeq 0.7979$ and $\sum_{0<s\leq t} \kappa^2(s)$ are the discontinuous jump segment of the quadratic variation process.

The HAR-RV and HARQ-RV models

The HAR-RV model

Recently, the heterogeneous autoregressive model of realized volatility (HAR-RV) proposed by Corsi (2009) has become the popular model in modeling and forecasting the dynamics of RV. There are two primary reasons for the popularity of Corsi's model. First, this model can be easily estimated using the OLS method so that it can be conveniently employed. Second, Corsi (2009) notes that the model is capable of reproducing the same volatility persistence observed in the empirical data and capturing main stylized facts of financial data (e.g., long memory and multi-scaling behavior). This model contains three components: daily lagged realized volatility, weekly average of realized volatility, and monthly average of realized volatility, which represent the behavior of short-, medium-, and long-term volatilities, respectively. The specification of the HAR-RV model is

$$\mathrm{RV}_{t+1} = \beta_0 + \beta_1\mathrm{RV}_t + \beta_2\mathrm{RVW}_t + \beta_3\mathrm{RVM}_t + \varepsilon_{t+1}, \tag{4}$$

where RVW_t is the average of RV from day t-4 to day t, RVM_t is the average of RV from day t-21 to day t, and ε_{t+1} is a disturbance error.

The HARQ-RV model

Bollerslev, Patton, and Quaedvlieg (2016) take into account the temporal variation in the errors (the variance of the measurement error) and construct a new model, which is called the HARQ-RV model. This model is also straightforward to implement and can easily be tailored to any autoregressive specification for RV defined as

$$\mathrm{RV}_{t+1} = \beta_0 + (\beta_1 + \beta_{1Q}\mathrm{RQ}_t^{1/2})\mathrm{RV}_t + \beta_2\mathrm{RVW}_t + \beta_3\mathrm{RVM}_t + \varepsilon_{t+1}, \tag{5}$$

where RQ represents the realized quarticity and $\mathrm{RQ}_t = (M/3)\sum_{j=1}^{1/\Delta} r_{t,j}^4$. It is clear that this model only allows the coefficient on the daily lagged RV to vary as a function of $\mathrm{RQ}_t^{1/2}$.

Extended models including the regime switching and heteroskedasticity

To the best of our knowledge, there is a large body of literature on modeling and forecasting volatility using the HAR-RV and its various extensions; see, for example, Andersen, Bollerslev, and Diebold (2007), Bekaert and Hoerova (2014), Clements and Liao (2017), Patton and Sheppard (2015), Wang, Wu, and Xu (2015), Gong and Lin (2017, 2018), among others. However, from theoretical and empirical perspectives, existing studies (e.g., Cipollini, Gallo, and Otranto 2017; Ma et al. 2017; McAleer and Medeiros 2008; Raggi and Bordignon 2012; Tian, Yang, and Chen 2017) have found that long memory will be overestimated ('spurious' high persistence) when neglecting regime or structural breaks. Therefore, we propose a new model that incorporates the regimes and forecasts the volatility of financial markets in the framework of the HAR-RV model.

Corsi et al. (2008) first find that the residuals of the HAR-RV model have been inflicted by strong volatility clustering. They describe this heteroscedasticity (volatility-of-volatility, VOV) using the generalized autoregressive conditional heteroskedasticity (GARCH) and indicate that the improvement obtained by considering the VOV can generate higher forecasting accuracy and better density forecasts. However, there is only a limited strand of literature that focuses on VOV when forecasting volatility; see, for example, Caporin and Velo (2015), Ceylan (2014), Cipollini, Gallo, and Otranto (2017), Louzis, Xanthopoulos-Sisinis, and Refenes (2012), and Ma, Liu, Huang, and Chen (2017).

More importantly, we verified that no study has modeled and forecasted volatility accounting for regime switching and the VOV at the same time. Consequently, we are the first to introduce long memory, regime switching and heteroscedasticity (VOV) into the HAR-RV and HARQ-RV models.

Following Louzis, Xanthopoulos-Sisinis, and Refenes (2012), Beltratti and Morana (2005), among others, we used the FIGARCH model (Baillie, Bollerslev, and Mikkelsen 1996) to describe heteroscedasticity in this study. We call these two new models MS-HAR-FIGARCH and MS-HARQ-FIGARCH. A brief description of each model follows.

Model 1: MS-HAR-FIGARCH. We introduce regime switching and heteroscedasticity to the HAR-RV model.

$$\begin{aligned}\mathrm{RV}_{t+1} &= \beta_0 + \beta_{1,s_t}\mathrm{RV}_t + \beta_2\mathrm{RVW}_t \\ &\quad + \beta_3\mathrm{RVM}_t + \varepsilon_{t+1},\end{aligned} \tag{6}$$

$$\begin{aligned}\sigma_{t+1}^2 &= \omega \\ &\quad + \beta\sigma_{t-1}^2 + [1-(1-\beta L)^{-1}(1-\varphi L)(1-L)^d]\varepsilon_{t+1}^2,\end{aligned} \tag{7}$$

where S_t is an unobserved state variable. In line with Raggi and Bordignon (2012), Goldman et al. (2013), Ma et al. (2017), we chose two regimes in our models. S_t= 0 indicates the low volatility regime with a smaller conditional variance, reflecting a stable market, whereas S_t= 1 indicates the high volatility regime with a larger conditional variance, reflecting a highly fluctuating market. The unobserved state variable, S_t, is assumed to follow a two-state Markov process with a transition probability matrix given by

$$P = \begin{bmatrix} p^{00} & 1-p^{00} \\ 1-p^{11} & p^{11} \end{bmatrix}, \tag{8}$$

where

$$p^{00} = p(s_t = 0 | s_{t-1} = 0), \tag{9}$$

$$p^{11} = p(s_t = 1 | s_{t-1} = 1), \tag{10}$$

The parameter d is a fractional integration parameter, and L is a lag operator. Additionally, the parameter d characterizes the long-memory property of hyperbolic decay in volatility because it allows autocorrelations to decay at a slow hyperbolic rate.

Model 2: MS-HARQ-FIGARCH. The HARQ-RV model takes into account regime switching and heteroscedasticity, as follows:

$$\begin{aligned}\mathrm{RV}_{t+1} &= \beta_0 + (\beta_{1,s_t} + \beta_{1Q}\mathrm{RQ}_t^{1/2})\mathrm{RV}_t \\ &\quad + \beta_2\mathrm{RVW}_t + \beta_3\mathrm{RVM}_t + \varepsilon_{t+1},\end{aligned} \tag{11}$$

$$\begin{aligned}\sigma_{t+1}^2 &= \omega \\ &\quad + \beta\sigma_{t-1}^2 + [1-(1-\beta L)^{-1}(1-\varphi L)(1-L)^d]\varepsilon_{t+1}^2.\end{aligned} \tag{12}$$

The MS-HAR-FIGARCH and MS-HARQ-FIGARCH can be estimated by the maximum likelihood function using the filtering procedure of Hamilton (1990) followed by the smoothing algorithm of Kim (1994).[2] The key issue of this study is to examine whether introducing regimes switching and heteroscedasticity(VOV) can improve the predictive performance compared to the competing models.

III. Data and preliminary analysis

We utilized the CSI 300 index future contract with a maturity of one month and chose the 5-min sampling frequency as our research object. The

[2]We are very thankful to Prof. Davidson, who shared the TSM packages (Time Series Modelling v4.38) on this website: http://www.timeseriesmodelling.com/. In particular, Prof. Davidson provided a detailed help file on how to estimate and forecast volatility using the regime switching model.

CSI 300 index futures were launched on 16 April 2010 on the China Financial Futures Exchange, and after 5 years of development, they have become one of the most actively traded financial instruments of Chinese financial markets (Sohn and Zhang 2017). The trading time of the CSI 300 index futures is from 9:15am to 11:30am and from 1:00pm to 3:15pm (Beijing time). Our data were collected from Wind Financial Terminal in Chinese markets from 1 January 2010 to 31 December 2015.[3]

Figure 1 depicts the 5-min prices of the CSI 300 index futures. We found that the large fluctuation in the CSI 300 index futures occurred after 2015. From early 2014 to middle 2015, we can clearly see that the CSI 300 index futures has increased. Table 1 describes those key variables in the HAR-RV and HARQ-RV models. Using the empirical results of Table 1, we determine that all realized measures (RV, RVW, RVM, BPV and RQ) reject the null hypothesis ('skewness = 0' and 'kurtosis-3 = 0') at the 1% significance level, implying that those series exhibit significant right skew and high kurtosis. Additionally, the Jarque-Bera statistics test (Jarque and Bera 1987) further indicates that those variables are non-Gaussian at the 1% significance level. Moreover, the results of the Augmented Dickey–Fuller test (Cheung and Lai 1995) show that all realized measures are stationary at the 99% confidence level because of the rejection of the null hypothesis of a unit root. Finally, the RV, RVW, RVM, BPV and RQ series have significant autocorrelations up to the 5th order using the Ljung–Box test (Ljung and Box 1978).

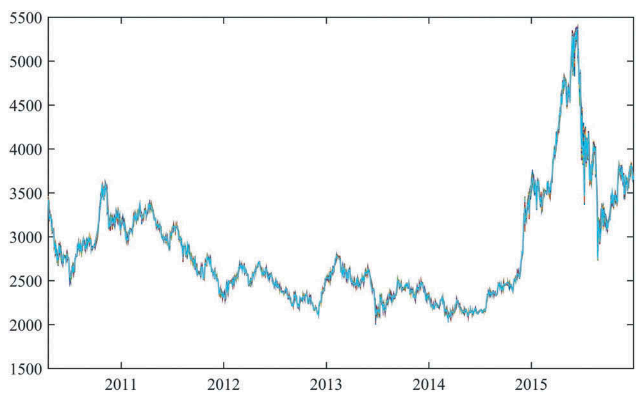

Figure 1. The 5-min prices of the CSI300 index futures over the sample period.

IV. Empirical results

All-of-sample estimated results

Table 2 reports the estimated results of the HAR-RV, HARQ-RV, MS-HAR-FIGARCH and MS-HARQ-FIGARCH models over the entire sample period. The lagged RV, RVW and RVM of the HAR-RV and HARQ-RV models are significant at the 99% confidence level, implying that they reject the null hypothesis (the parameters are equal to 0) and are statistically significant. Compared to the adjusted R-squared, we find that the HARQ-RV model is slightly larger than the HAR-RV model and that both of those models are greater than 0.59. We use Engle's LM ARCH test (Engle 1982) to investigate the residuals of the HAR-RV and HARQ-RV models and find that the HAR-RV and HARQ-RV models have significant ARCH effect at the 95% confidence level. Evidently, modeling and forecasting volatility accounting for the volatility of volatility (heteroscedasticity) is reasonable. Additionally, we use the likelihood-ratio statistic (LR-test) to test the linearity of

Table 1. Descriptive statistics of realized volatility, jump and signed jump variation series.

Obs.	Mean	St. dev.	Skewness	Kurtosis	Jarque-Bera	$Q(5)$	ADF
RV	0.254	0.566	7.869***	82.393***	400477.528***	2626.419***	−6.903***
RVW	0.254	0.480	6.054***	44.140***	119236.080***	5171.635***	−3.916***
RVM	0.255	0.393	4.305***	20.717***	28646.932***	6562.856***	−3.723***
BPV	0.222	0.506	7.451***	67.375***	271003.275***	2720.663***	−6.713***
RQ	0.001	0.007	20.917***	536.654***	16491478.866***	202.608***	−8.940***

We tested the null hypothesis, 'Skewness = 0' and 'Excess Kurtosis = 3'. The Jarque-Bera statistic (Jarque and Bera 1987) tests are for the null hypothesis of normality for the distribution of the series. $Q(n)$ is the Ljung-Box statistic proposed by Ljung and Box (1978) for up to 5th order serial correlation. ADF is the Augmented Dickey–Fuller statistic (Cheung and Lai 1995) based on the least AIC criterion. Asterisk *** denotes rejections of the null hypothesis at the 1% significance level. All realized measures (RV, RVW, RVM, BPV and RQ) of Table 1 are multiplied by 1000.

[3]We choose this sample period because the trading time changed after 31 December 2015. For more details of the CSI 300 index futures, the interested reader may browse the homepage of China Financial Futures Exchange (http://www.cffex.com.cn/index.html) and Yang, Yang, and Zhou (2012), Sohn and Zhang (2017), among other websites.

Table 2. In-sample estimated results of individual models discussed in this article.

	HAR-RV	HARQ-RV	MS-HAR-FIGARCH	MS-HARQ-FIGARCH
β_0	−0.883***	−0.833***	−1.2824*	−1.8942*
β_1	0.134***	0.772***		
$\beta_{1,0}$			−0.0126	0.3069***
$\beta_{1,1}$			0.4493**	0.3431***
β_2	0.544***	0.514***	0.5316***	0.5171***
β_3	0.234***	0.223***	0.3418***	0.5665***
β_{1Q}		−0.298***		−0.1380***
ω			0.3411***	0.3508***
d			0.0510	0.0311
ARCH(φ)			0.0866**	0.0933*
GARCH(β)			0.7349***	0.5524***
p^{00}			0.9965***	0.9987***
p^{11}			0.9889***	0.9955***
Normality test	22.956***	21.820***	–	–
ARCH	6.529**	6.542**	–	–
Adj.R^2	0.591	0.595	0.6135	0.6107

Asterisks ***, ** and * denotes rejections of null hypothesis at 1%, 5% and 10% significance levels, respectively. Those models are transformed by logarithmic form.

those models[4] and find that the two models significantly reject the linear hypothesis based on the linear LR-test, implying that introducing regime switching to the two models is feasible.

Compared to the switched coefficients of the MS-HAR-FIGARCH and MS-HARQ-FIGARCH models, we find that the RV of high volatility regime has more powerful impacts on future volatility than the low volatility regime and that its coefficients are significant at the 95% significance level. However, the memory parameter d of the regime switching models seems not to reject the null hypothesis ('d = 0') implying that the long-memory property is very weak in affecting the volatility of RV. Based on the magnitudes of the p^{00} and p^{11}, we find that the high volatility regime is slightly short-lived. From the adjusted R-square of the last column in Table 2, it is clear that our models including regime switching and heteroscedasticity have a better fit to the data compared to the HAR-RV and HARQ-RV models.

Out-of-sample evaluation

Evaluation methods

It is worth noting that the out-of-sample predictability of a model is more useful to market participants and researchers compared to performance during the in-sample period. The primary reason for that is that the latter is more concerned with the model's ability to predict the future than its ability to analyze the past (e.g., Wang, Ma, Wei, and Wu, 2016; Ma et al. 2017). We use the rolling window method to obtain future volatility. The entire sample is divided into two subgroups, in-sample and out-of-sample. The length of the in-sample data is 800, covering the first 800 trading days. The rest of the sample is out-of-sample data, covering the last 566 trading days. Simply stated, the length of the in-sample data (800 days) is used to estimate the models given a fixed length and assuming that the forecasts do not overlap. Using this forecasting method, we obtain 566 forecasting observations and re-estimate each model discussed in this paper 566 times.

To assess the differences between those models, we use the following two loss functions:

$$\text{HMSE} = M^{-1}\sum_{m=1}^{M}\left(1-\hat{\sigma}_m^2/\text{RV}_m\right)^2, \tag{13}$$

$$\text{HMAE} = M^{-1}\sum_{m=1}^{M}\left|1-\hat{\sigma}_m^2/\text{RV}_m\right|, \tag{14}$$

where HMSE and HMAE represent the heteroscedasticity-adjusted mean absolute error and the heteroscedasticity-adjusted mean squared error, respectively. A large body of literature (see, for example, Chen, Yu, and Zivot 2012; Kristjanpoller and Minutolo 2016; Koopman, Jungbacker, and Hol 2005; Gong and Lin 2017a; Ma et al. 2017) uses these two popular loss functions to evaluate forecasting performance. $\hat{\sigma}_m^2$ denotes the out-of-sample volatility forecast obtained by different HAR-type models, RV_m is a proxy for actual market volatility in the out-of-sample period, and M is the length of the out-of-sample data.

Notably, the aforementioned loss functions do not provide any information on whether the differences among the models are statistically significant. Therefore, we utilize the model confidence set (MCS) proposed by Hansen, Lunde, and

[4]Due to space limitation, we do not present the procedure of the Linearity LR-test. More details of this procedure can be found in Oxmetrics 6.21 with PcGive help (§14.2.3 Testing linearity). PcGive reports a test for linearity for all outputs. The test is based on the likelihood-ratio statistic between the estimated model and the derived linear model. It also reports the approximate upper bound for the significance level of the LR statistic derived from Davies (1987) and Garcia and Perron (1996).

Nason (2011) to choose a subset of models that contain all possible superior models from the initial model set. The MCS method has several attractive advantages over conventional tests such as a superior predictive power (Hansen 2005) and 'reality check' tests. For example, this test does not require specifying a benchmark model, which is useful in applications without an obvious benchmark. If the MCS p-values of several volatility models are greater than a critical value α, the corresponding models are 'surviving' models, implying that those models outperform other models that produce a value lower than α. The higher the p-value of the MCS test is, the more likely the corresponding model to be better than other models. Additionally, following Audrino and Hu (2016), Gong and Lin (2017b), Martens, Van, and De (2009), and Rossi and Fantazzini (2014), we use the range statistic ($\mathrm{T_R}$) and the semi-quadratic statistic ($\mathrm{T_{SQ}}$) as the MCS statistics to test whether the null hypothesis of equal predictive power for the remaining models is not rejected. These statistics are defined as follows,

$$T_R = \max_{u,v\in M} \frac{\left|\bar{d}_{i,uv}\right|}{\sqrt{\mathrm{var}(d_{i,uv})}},$$

$$T_{SQ} = \max_{u,v\in M} \frac{\left(\bar{d}_{i,uv}\right)^2}{\mathrm{var}(\bar{d}_{i,uv})}, \tag{15}$$

where $\bar{d}_{i,uv} = \frac{1}{M}\sum_{m=1}^{M} d_{i,uv,m}$. In this paper, we do not give further technical details of the MCS test; more in-depth discussions may be found in Hansen, Lunde, and Nason (2011).

Out-of-sample results of the MCS test

In this paper, we use logarithmic volatility to represent future volatilities. There are two primary reasons for this choice. First, Andersen et al. (2001), Aït-Sahalia and Mancini (2008), Paye (2012), and Ma et al. (2017), among others, have empirically found that the distribution of log-RV can be closer to Gaussian than that of RV. Second, Bekaert and Hoerova (2014) indicate that the logarithmic form may be easier for predicting logarithmic variances compared to linear models, as logarithmic RV tends to have a near Gaussian distribution. To represent the future variances of each model, we transform the logarithmic realized volatility following the works of Bee, Dupuis, and Trapin (2016), Bekaert and Hoerova (2014), Brownlees and Gallo (2010), Giot and Laurent (2004), Louzis, Xanthopoulos-Sisinis, and Refenes (2014) as $\hat{\sigma}^2_{t+1} = \exp(\log \mathrm{R\hat{V}}_{t+1} + 0.5\sigma^2_u)$. The predictive logarithmic RV ($\log \mathrm{R\hat{V}}_{t+1}$) can be forecasted by individual models, where σ^2_u is the variance of the residuals.

Following Martens, Van, and De (2009), Laurent, Rombouts, and Violante (2012), and Ma et al. (2018), among others, we chose the critical value of α to be 25%, implying that a model with a p-value smaller than 0.25 will be excluded from the initial model set. Clearly, the 'surviving' models significantly outperform the removed models. The out-of-sample empirical results of Table 3 show the significant differences of the existing models (HAR-RV, HARQ-RV) and extended models with regimes and heteroscedasticity. Using the range statistic ($\mathrm{T_R}$) and the semi-quadratic statistic ($\mathrm{T_{SQ}}$) of the HMSE loss function, we find that the MCS p-values of our proposed forecasting models, the MS-HAR-FIGARCH and MS-HARQ-FIGARCH, are greater than 0.25, which strongly indicates that those models including the regimes and VOV can generate better forecasts than the existing models. Additionally, the MS-HARQ-FIGARCH model has superior forecasting performance compared to the MS-HAR-FIGARCH model based on the p-values produced by the MCS test. Given the larger critical value and following Wei et al. (2017), we choose the 50% confidence level and find that the MS-HARQ-FIGARCH model exhibits superior forecasting performance compared to the MS-HAR-FIGARCH model. Interestingly, we find that the MS-HAR-FIGARCH and MS-HARQ-FIGARCH models exhibit higher predictive power based on the p-values produced by the MCS test.

Table 3. MCS test results of out-of-sample forecasting performance.

	HMSE		HMAE	
Models	$\mathrm{T_R}$	$\mathrm{T_{SQ}}$	$\mathrm{T_R}$	$\mathrm{T_{SQ}}$
HAR-RV	**0.7380**	0.0018	0.0104	0.0000
MS-HAR-FIGARCH	**0.7380**	**0.2916**	**0.2560**	**0.2560**
HARQ-RV	**0.7380**	0.0038	0.0104	0.0000
MS-HARQ-FIGARCH	**1.0000**	**1.0000**	**1.0000**	**1.0000**

MCS p-values are calculated according to the test statistics $\mathrm{T_R}$ and $\mathrm{T_{SQ}}$. The MCS $p > 0.25$ are indicated in bold. The length of out-of-sample is 566.

Both of the range (T_R) and semi-quadratic (T_{SQ}) statistics are consistent with the result of the HMSE loss function. Judging by the empirical results associated with those two loss functions, we can clearly determine that introducing regime switching and heteroscedasticity to the HAR-RV and HARQ-RV models significantly improves their predictive power. Furthermore, we find that the MS-HARQ-FIGARCH model generally exhibits higher forecasting accuracy than the MS-HAR-FIGARCH model. In conclusion, this paper proposes two superior forecasting models, especially the MS-HARQ-FIGARCH, and provides a new insight into forecasting future volatility.

Robustness analysis

Different forecasting windows

Rossi and Inoue (2012) argue that different estimations and forecasting windows maybe produce different empirical results. Therefore, the length of the out-of-sample data plays an important role in assessing the predictive power of the models discussed in this paper. We therefore further reallocate the length of the in-sample and out-of-sample data and demonstrate the empirical results in Table 4 in choosing 466 and 666 as our alternative forecasting windows. Using the HMSE and HMAE loss functions with range and semi-quadratic statistics, we find that our proposed model, MS-HARQ-FIGARCH, substantially outperforms the original models in forecasting volatility from a statistical perspective. However, the MCS *p*-values of the MS-HAR-FIGARCH model are lower than 0.25 using the range and semi-quadratic statistics of HMSE and HMAE loss functions, implying that this model has inferior performance compared to the MS-HARQ-FIGARCH. More importantly, the MS-HARQ-FIGARCH model can also reflect more accurate forecasts, which can further support our conclusions and verifies that our findings are robust and reliable.

Different forecasting models

Several works have forecasted the volatility of financial markets using the HAR-RV model as well as modeled heteroscedasticity by the GARCH and EGARCH models; see, for example, Ceylan (2014), Cipollini, Gallo, and Otranto (2017), Ma, Liu, Huang, and Chen (2017), Cipollini, Gallo, and Otranto (2017). We therefore employ the GARCH and EGARCH models to model the residuals of the HAR-RV and HARQ-RV models and introduce regimes to those models, labelled MS-HAR-GARCH, MS-HAR-EGARCH, MS-HARQ-GARCH and MS-HARQ-EGARCH, respectively. Subsequently, we evaluate those models using our proposed models and the existing models using the MCS test. Table 5 shows the empirical results of the MCS test. The results imply that the MS-HARQ-FIGARCH model can yield more accurate forecasts based on the *p*-values produced by the MCS test with range and semi-quadratic statistics. More importantly, we confirmed that the MCS *p*-values of those models including regimes and heteroscedasticity are slightly larger (or equal to) than the HAR-RV and HARQ-RV models. To some extent, therefore, those models have superior forecasting performance compared to the existing models. The findings of Table 5 provide additional convincing evidence that accounting for regimes and heteroscedasticity of the realized volatility models can increase forecasting accuracy.

Table 4. MCS test results on different forecasting windows.

	HMSE		HMAE	
Models	T_R	T_{SQ}	T_R	T_{SQ}
The length of out-of-sample is 666.				
HAR-RV	**0.9511**	0.0000	0.0038	0.0000
MS-HAR-FIGARCH	**0.9511**	0.0000	0.0038	0.0000
HARQ-RV	**0.9511**	0.0000	0.0032	0.0000
MS-HARQ-FIGARCH	**1.0000**	**1.0000**	**1.0000**	**1.0000**
The length of out-of-sample is 466.				
HAR-RV	**0.2652**	0.0614	**0.2666**	0.0007
MS-HAR-FIGARCH	**0.2652**	0.0614	**0.2666**	0.0007
HARQ-RV	**0.2652**	0.0614	**0.2666**	0.0007
MS-HARQ-FIGARCH	**1.0000**	**1.0000**	**1.0000**	**1.0000**

MCS *p*-values are calculated according to the test statistics T_R and T_{SQ}. Values of $p > 0.25$ are indicated in bold.

Table 5. The MCS test results on the predictive power of individual models during the forecasting windows.

	HMSE		HMAE	
Models	T_R	T_{SQ}	T_R	T_{SQ}
HAR-RV	0.0039	0.0246	0.0019	0.0043
MS-HAR-GARCH	**0.2879**	0.0328	0.0539	0.0408
MS-HAR-EGARCH	0.0039	0.0246	**0.2554**	0.1805
MS-HAR-FIGARCH	**0.2879**	0.0328	**0.2560**	**0.2560**
HARQ-RV	**0.2879**	0.0246	0.0002	0.0009
MS-HARQ-GARCH	**0.2879**	0.0328	**0.2554**	0.0524
MS-HARQ-EGARCH	0.0039	0.0246	0.0539	0.0408
MS-HARQ-FIGARCH	**1.0000**	**1.0000**	**1.0000**	**1.0000**

MCS *p*-values are calculated according to the test statistics T_R and T_{SQ}. Values of $p > 0.25$ are indicated in bold.

Different realized measures

We use the realized bi-power variation (BPV) to replace the RV in MS-HAR-FIGARCH and MS-HARQ-FIGARCH models. Based on the theory of Barndorff-Nielsen and Shephard (2004), RV can be divided into BPV and Jump components when the sampling interval is close to zero, so that BPV has no noise and can be continuously compared to RV. We report the empirical results in Table 6 using the MCS test, which involves the same forecasting window as in Table 3. Importantly, Table 6 indicates that compared to the existing models proposed by Corsi (2009) and Bollerslev, Patton, and Quaedvlieg (2016), the MS-HARQ-BPV-FIGARCH model has greater predictive power in forecasting volatility based on the HMSE and HMAE with range and semi-quadratic statistics. This empirical result strongly indicates that our findings are indeed robust and provide a new insight on the benefit of introducing regime switching and realized volatility of volatility (heteroscedasticity) for volatility forecasting.

Realized kernel as actual market volatility

A large body of literature indicates that market microstructure noise plays a negative role in the measurement of realized volatility (e.g., Andersen, Bollerslev, and Meddahi 2011; Awartani, Corradi, and Distaso 2009; Bandi and Russell 2008; Liu, Patton, and Sheppard 2015). Several studies (e.g., Liu, Patton, and Sheppard 2015; Sévi 2014; Wang, Wu, and Xu 2015; Wang, Ma, Wei, and Wu, 2016) have used the 5-minute rule-of-thumb to mitigate these noise effects and thereby balance the estimated accuracy and errors, whereas Zhang, Mykland, Ait-Sahalia (2005) find that this sampling frequency does not seem to be an optimal solution to the problem. Hence, to re-examine our findings, we choose the realized kernel (RK) as a measure that is robust to market microstructure noise (Barndorff-Nielsen et al. 2008) as our alternative actual volatility. The specification of RK is defined as follows:

$$RK_t = \sum_{j=-H}^{H} k\left(\frac{j}{H+1}\right)\gamma_j,$$

$$\gamma_j = \sum_{i=|j|+1}^{n} r_{t,i} r_{t,i-|j|} \tag{16}$$

where$k(x)$is the Parzen kernel function given by:

$$k(x) = \begin{cases} 1 - 6x^2 + 6x^3, & 0 \le x \le 1/2 \\ 2(1-x)^3, & 1/2 \le x \le 1 \\ 0, & x > 1. \end{cases} \tag{17}$$

It is necessary for H to increase with the sample size to estimate the increments of quadratic variation consistently in the presence of noise. We precisely followed the bandwidth choice of H (Barndorff-Nielsen et al. 2009). Table 7 shows the empirical results of individual models using the MCS test as well as realized kernel (RK) as the actual market volatility. Under both the HMSE and the HMAE loss functions, we found that the p-values produced by the MCS test of the MS-HAR-FIGARCH and MS-HARQ-FIGARCH models are greater than those in the HAR-RV and HARQ-RV models, which strongly indicates that adding regimes and heteroscedasticity to the simple realized volatility models can statistically and significantly improve forecasting performance.

Different stock markets

Could our findings be caused by a special financial market (i.e., CSI300 index futures market)? To answer this key question, we further apply the two important stock market indexes, the S&P 500 and CAC400 index, to re-examine whether our proposed model, MS-HAR-FIGARCH and MS-HARQ-

Table 6. The empirical results on the predictive differences of each model basing on the MCS test.

	HMSE		HMAE	
Models	T_R	T_{SQ}	T_R	T_{SQ}
HAR-RV	**0.6322**	0.0006	**0.6343**	0.0000
MS-HAR-BPV-FIGARCH	**0.6322**	0.0006	**0.6343**	0.0000
HARQ-RV	**0.6322**	0.0006	**0.6343**	0.0000
MS-HARQ-BPV-FIGARCH	**1.0000**	**1.0000**	**1.0000**	**1.0000**

MCS p-values are calculated according to the test statistics T_R and T_{SQ}. Values of $p > 0.25$ are indicated in bold. We use the BPV to replace the RV of the MS-HAR-FIGARCH and MS-HARQ-FIGARCH, and dub as MS-HAR-BPV-FIGARCH and MS-HAR-BPV-FIGARCH, respectively.

Table 7. MCS test results of out-of-sample forecasting performance as realized kernel.

	HMSE		HMAE	
Models	T_R	T_{SQ}	T_R	T_{SQ}
HAR-RV	0.0310	0.0000	0.0000	0.0000
MS-HAR-FIGARCH	**0.3988**	**0.3988**	**0.3090**	**0.3090**
HARQ-RV	0.0310	0.0000	0.0000	0.0000
MS-HARQ-FIGARCH	**1.0000**	**1.0000**	**1.0000**	**1.0000**

MCS p-values are calculated according to the test statistics T_R and T_{SQ}. Values of $p > 0.25$ are indicated in bold.

Table 8. The empirical results of different stock markets using the MCS test.

Models	HMSE		HMAE	
	T_R	T_{SQ}	T_R	T_{SQ}
S&P 500				
HAR-RV	0.0000	0.0000	0.0000	0.0000
MS-HAR-FIGARCH	0.0688	0.0688	0.1080	0.1080
HARQ-RV	0.0000	0.0000	0.0000	0.0000
MS-HARQ-FIGARCH	**1.0000**	**1.0000**	**1.0000**	**1.0000**
CAC400				
HAR-RV	0.0221	0.0242	0.0000	0.0000
MS-HAR-FIGARCH	**1.0000**	**1.0000**	**1.0000**	**1.0000**
HARQ-RV	0.0026	0.0196	0.0000	0.0000
MS-HARQ-FIGARCH	**0.2511**	**0.2511**	**0.4162**	**0.4162**

MCS p-values are calculated according to the test statistics T_R and T_{SQ}. Values of $p > 0.25$ are indicated in bold.

FIGARCH, can obtain higher forecasts accuracy. Our data were collected by Thomson Reuters Tick History Database (TRTH) from 1 January 2010 to 31 December 2015. Table 8 reports the empirical results of each model using the MCS test. We obtain the forecasting values by using the rolling windows and find that those results are consistent with the aforementioned findings that account for the regimes and heteroscedasticity and can significantly increase the predictive power of the existing models based on the MCS p-values.

V. Concluding remarks

With the increasing availability of high-frequency data in financial markets, volatility models based on intraday data have received widespread attention from academics and practitioners. In this article, we modeled and forecasted realized volatility using the framework of HAR-RV and HARQ-RV models that capture a long memory feature. More important, by introducing regimes and realized volatility of volatility (heteroscedasticity) to those models, we proposed two new models for forecasting the volatility and evaluate their predictive power compared to the existing models. The empirical results include several noteworthy findings. First, all-of-sample results show that existing models exhibit significant clustering and heteroscedasticity, which support our choice to model the volatility of realized volatility. Additionally, the likelihood-ratio statistic test indicates that the existing models obviously reject the linear hypothesis at the 99% confidence level, which implies that accounting for regime switching in the two new models is feasible. The RV of a high volatility regime has more powerful impacts on future volatility than a low volatility regime, and its coefficients are significant. Second, our out-of-sample results provide convincing evidence that introducing regime switching and heteroscedasticity can significantly increase forecasting accuracy. In particular, the MS-HARQ-FIGARCH model generally exhibits greater predictive power than other models discussed in this paper. Third, our findings are robust to various robustness tests, such as different forecasting windows, forecasting models, realized measures, and stock markets. Consequently, this study sheds new light on forecasting future volatility.

Acknowledgments

We are thankful to the editor-in-chief, Prof. Mark P. Taylor, who provided us the opportunity to revise this paper, and to two anonymous referees for providing valuable comments that helped us to substantially improve the quality of our novel paper. Feng Ma is grateful for the financial support from the Natural Science Foundation of China [71671145, 71701170], the humanities and social science fund of the Ministry of Education [17YJC790105, 17XJCZH002], and the Fundamental Research Funds for the Central Universities [2682017WCX01, 2682018WXTD05]. Ke Yang is supported by the financial support from the Natural Science Foundation of China [71673089], the Fundamental Research Funds for the Central Universities [2017MS114], and the National Social Science Foundation of China (15CJY004).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Feng Ma is grateful for the financial support from the Natural Science Foundation of China [71671145, 71701170, 71701171], the humanities and social science fund of the Ministry of Education [17YJC790105, 17XJCZH002], and the Fundamental Research Funds for the Central Universities [26816WCX02, 2682017WCX01]. Ke Yang is supported by the financial support from the Natural Science Foundation of China [71673089], the Fundamental Research Funds for the Central Universities, and the National Social Science Foundation of China (15CJY004); National Social Science Foundation of China [15CJY004]; the humanities and social science fund of the Ministry of Education [17YJC790105, 17XJCZH002]; the Fundamental Research Funds for the Central Universities [2017MS114, 26816WCX02, 2682017WCX01].

ORCID

Yaojie Zhang http://orcid.org/0000-0002-4220-1623

References

Aït-Sahalia, Y., and L. Mancini. 2008. "Out of Sample Forecasts of Quadratic Variation." *Journal of Econometrics* 147 (1): 17–33.

Andersen, T. G., and T. Bollerslev. 1998. "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts." *International Economic Review* 39: 885–905.

Andersen, T. G., T. Bollerslev, and F. X. Diebold. 2007. "Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility." *The Review of Economics and Statistics* 89: 701–720.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens. 2001. "The Distribu- Tion of Realized Stock Return Volatility." *Journal of Financial Economics* 61: 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys. 2003. "Modeling and Forecasting Realized Volatility." *Econometrica* 71 (2): 579–625.

Andersen, T. G., T. Bollerslev, and N. Meddahi. 2011. "Realized Volatility Forecasting and Market Microstructure Noise." *Journal of Econometrics* 160 (1): 220–234.

Audrino, F., and Y. Hu. 2016. "Volatility Forecasting: Downside Risk, Jumps and Leverage Effect." *Econometrics* 4 (1): 8.

Awartani, B., V. Corradi, and W. Distaso. 2009. "Assessing Market Microstructure Effects via Realized Volatility Measures with an Application to the Dow Jones Industrial Average Stocks." *Journal of Business & Economic Statistics* 27 (2): 251–265.

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen. 1996. "Fractionally Integrate Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 74 (1): 3–30.

Bandi, F. M., and J. R. Russell. 2008. "Microstructure Noise, Realized Variance, and Optimal Sampling." *The Review of Economic Studies* 75 (2): 339–369.

Banerjee, A., and G. Urga. 2005. "Modelling Structural Breaks, Long Memory and Stock Market Volatility: An Overview." *Journal of Econometrics* 129 (1): 1–34.

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

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard. 2008. "Designing Realized Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise." *Econometrica* 76: 1481–1536.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard. 2009. "Realized Kernels in Practice: Trades and Quotes." *The Econometrics Journal* 12: 1–32.

Bee, M., D. J. Dupuis, and L. Trapin. 2016. "Realizing the Extremes: Estimation of Tail-Risk Measures from a High-Frequency Perspective." *Journal of Empirical Finance* 36: 86–99.

Bekaert, G., and M. Hoerova. 2014. "The VIX, the Variance Premium and Stock Market Volatility." *Journal of Econometrics* 183: 181–192.

Beltratti, A., and C. Morana. 2005. "Statistical Benefits of Value-At-Risk with Long Memory." *The Journal of Risk* 7 (4): 1–45.

Bollerslev, T. 1986. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31: 307–327.

Bollerslev, T., A. J. Patton, and R. Quaedvlieg. 2016. "Exploiting the Errors: A Simple Approach for Improved Volatility Forecasting." *Journal of Econometrics* 192 (1): 1–18.

Bollerslev, T., B. Hood, J. Huss, and L. H. Pedersen, 2017. "Risk Everywhere: Modeling and Managing Volatility." 21 March 2017. Available at SSRN: https://ssrn.com/abstract=2722591

Brownlees, C. T., and G. M. Gallo. 2010. "Comparison of Volatility Measures: A Risk Management Perspective." *Journal of Financial Econometrics* 8 (1): 29–56.

Caporin, M., and G. G. Velo. 2015. "Realized Range Volatility Forecasting: Dynamic Features and Predictive Variables." *International Review of Economics & Finance* 40: 98–112.

Carnero, M. A., D. Peña, and E. Ruiz. 2004. "Persistence and Kurtosis in GARCH and Stochastic Volatility Models." *Journal of Financial Econometrics* 2 (2): 319–342.

Ceylan, Ö. 2014. "Time-Varying Volatility Asymmetry: A Conditioned HAR-RV (CJ) EGARCH-M Model." *The Journal of Risk* 17 (2): 21.

Chen, C. H., W. C. Yu, and E. Zivot. 2012. "Predicting Stock Volatility Using After-Hours Information: Evidence from the NASDAQ Actively Traded Stocks." *International Journal of Forecasting* 28 (2): 366–383.

Cheung, Y. W., and K. S. Lai. 1995. "Lag Order and Critical Values of the Augmented Dickey–Fuller Test." *Journal of Business & Economic Statistics* 13 (3): 277–280.

Cipollini, F., G. M. Gallo, and E. Otranto, 2017. "On Heteroskedasticity and Regimes in Volatility Forecasting." Available at SSRN: https://ssrn.com/abstract=3037550

Clements, A., and Y. Liao. 2017. "Forecasting the Variance of Stock Index Returns Using Jumps and Cojumps." *International Journal of Forecasting* 33 (3): 729–742.

Corsi, F., S. Mittnik, C. Pigorsch, and U. Pigorsch. 2008. "The Volatility of Realized Volatility." *Econometric Reviews* 27 (1–3): 46–78.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. J. Financ. Econometr. 7 (2): 174–196.

Davies, R. B. 1987. "Hypothesis Testing When a Nuisance Parameter Is Present Only under the Alternative." *Biometrika* 74 (1): 33–43.

Duong, D., and N. R. Swanson. 2015. "Empirical Evidence on the Importance of Aggregation, Asymmetry, and Jumps for Volatility Prediction." *Journal of Econometrics* 187 (2): 606–621.

Engle, R. F. 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica: Journal of the Econometric Society* 50: 987–1007.

Engle, R. F., and G. M. Gallo. 2006. "A Multiple Indicators Model for Volatility Using Intra-Daily Data." *Journal of Econometrics* 131 (1): 3–27.

Garcia, R., and P. Perron. 1996. "An Analysis of the Real Interest Rate under Regime." *Review of Economics & Statistics* 78 (1): 111–125.

Ghysels, E., P. Santa-Clara, and R. Valkanov. 2006. "Predicting Volatility: Getting the Most Out of Return Data Sampled at Different Frequencies." *Journal of Econometrics* 131 (1): 59–95.

Giot, P., and S. Laurent. 2004. "Modelling Daily Value-At-Risk Using Realized Volatility and ARCH Type Models." *Journal of Empirical Finance* 11 (3): 379–398.

Goldman, E., J. Nam, H. Tsurumi, and J. Wang. 2013. "Regimes and Long Memory in Realized Volatility." *Studies in Nonlinear Dynamics and Econometrics* 17 (5): 521–549.

Gong, X., and B. Lin. 2017a. "Forecasting the Good and Bad Uncertainties of Crude Oil Prices Using a HAR Framework." *Energy Economics* 67: 315–327.

Gong, X., & Lin, B. (2018). Structural breaks and volatility forecasting in the copper futures market. *Journal of Futures Markets, 38*(3), 290-339.

Hamilton, J. D. 1990. "Analysis of Time Series Subject to Changes in Regime." *Journal of Econometrics* 45: 39–70.

Hansen, P. R. 2005. "A Test for Superior Predictive Ability." *Journal of Business & Economic Statistics* 23 (4): 365–380.

Hansen, P. R., and A. Lunde. 2005. "A Forecast Comparison of Volatility Models: Does Anything Beat A GARCH (1, 1)?" *Journal of Applied Econometrics* 20 (7): 873–889.

Hansen, P. R., A. Lunde, and J. M. Nason. 2011. "The Model Confidence Set." *Econometrica* 79 (2): 453–497.

Hansen, P. R., Z. Huang, and H. H. Shek. 2012. "Realized Garch: A Joint Model for Returns and Realized Measures of Volatility." *Journal of Applied Econometrics* 27 (6): 877–906.

Harrison, B., and W. Moore. 2012. "Forecasting Stock Market Volatility in Central and Eastern European Countries." *Journal of Forecasting* 31 (6): 490–503.

Hillebrand, E., and M. Medeiros. 2008. "Estimating and Forecasting GARCH Models in the Presence of Structural Breaks and Regime Switches." In *Forecasting in the Presence of Structural Breaks and Model Uncertainty. Emerald*, edited by M. Wohar and D. Rapach, 303–327.

Jarque, C. M., and A. K. Bera. 1987. "A Test for Normality of Observations and Regression Residuals." *International Statistical Review/Revue Internationale De Statistique* 163–172.

Kim, C. J. 1994. "Dynamic Linear Models with Markov-Switching." *Journal of Econometrics* 60: 1–22.

Koopman, S. J., B. Jungbacker, and E. Hol. 2005. "Forecasting Daily Variability of the S&P 100 Stock Index Using Historical, Realised and Implied Volatility Measurements." *Journal of Empirical Finance* 12 (3): 445–475.

Kristjanpoller, W., and M. C. Minutolo. 2016. "Forecasting Volatility of Oil Price Using an Artificial Neural network-GARCH Model." *Expert Systems with Applications* 65: 233–241.

Laurent, S., J. V. Rombouts, and F. Violante. 2012. "On the Forecasting Accuracy of Multivariate GARCH Models." *Journal of Applied Econometrics* 27: 934–955.

Liu, L. Y., A. J. Patton, and K. Sheppard. 2015. "Does Anything Beat 5-Minute RV? A Comparison of Realized Measures across Multiple Asset Classes." *Journal of Econometrics* 187: 293–311.

Ljung, G. M., and G. E. Box. 1978. "On a Measure of Lack of Fit in Time Series Models." *Biometrika* 65 (2): 297–303.

Louzis, D. P., S. Xanthopoulos-Sisinis, and A. P. Refenes. 2012. "Stock Index Realized Volatility Forecasting in the Presence of Heterogeneous Leverage Effects and Long Range Dependence in the Volatility of Realized Volatility." *Applied Economics* 44 (27): 3533–3550.

Louzis, D. P., S. Xanthopoulos-Sisinis, and A. P. Refenes. 2014. "Realized Volatility Models and Alternative Value-at-Risk Prediction Strategies." *Economic Modelling* 40: 101–116.

Ma, F., J. Liu, D. Huang, and W. Chen. 2017. "Forecasting The Oil Futures Price Volatility: a New Approach." *Economic Modelling* 64: 560-566. doi:10.1016/j.econmod.2017.04.020.

Ma, F., M. I. M. Wahab, D. Huang, and W. Xu. 2017. "Forecasting the Realized Volatility of the Oil Futures Market: A Regime Switching Approach." *Energy Economics* 67: 136–145.

Ma, F., Y. Wei, L. Liu, and D. Huang. 2018. "Forecasting Realized Volatility of Oil Futures Market: A New Insight." *Journal of Forecasting*, 37(4): 419–4

Martens, M., D. D. Van, and P. M. De. 2009. "Forecasting S&P 500 Volatility: Long Memory, Level Shifts, Leverage Effects, Day-Of-The-Week Seasonality, and Macroeconomic Announcements." *International Journal of Forecasting* 25: 282–303.

Martens, M., and J. Zein. 2004. "Predicting Financial Volatility: High-Frequency Time-Series Forecasts Vis-À-Vis Implied Volatility." *Journal of Futures Markets* 24 (11): 1005–1028.

McAleer, M., and M. C. Medeiros. 2008. "A Multiple Regime Smooth Transition Heterogeneous Autoregressive Model for Long Memory and Asymmetries." *Journal of Econometrics* 147 (1): 104–119.

Patton, A.J., and K. Sheppard. 2015. "Good Volatility." *Bad Volatility: Signed Jumps and The Persistence Of Volatility. Rev. Econ. Stat.* 97 (3): 683–697.

Paye, B. S. 2012. "'Déjà Vol': Predictive Regressions for Aggregate Stock Market Volatility Using Macroeconomic Variables." *Journal of Financial Economics* 106 (3): 527–546.

Raggi, D., and S. Bordignon. 2012. "Long Memory and Nonlinearities in Realized Volatility: A Markov Switching Approach." *Computational Statistics & Data Analysis* 56 (11): 3730–3742.

Rossi, B., and A. Inoue. 2012. "Out-Of-Sample Forecast Tests Robust to the Choice of Window Size." *Journal of Business & Economic Statistics* 30 (3): 432–453.

Rossi, E., and D. Fantazzini. 2014. "Long Memory and Periodicity in Intraday Volatility." *Journal of Financial Econometrics* 13 (4): 922–961.

Sévi, B. 2014. "Forecasting the Volatility of Crude Oil Futures Using Intraday Data." *European Journal of Operational Research* 235: 643–659.

Shephard, N., and K. Sheppard. 2010. "Realising the Future: Forecasting with High-Frequency-Based Volatility (HEAVY) Models." *Journal of Applied Econometrics* 25 (2): 197–231.

Sohn, S., and X. Zhang. 2017. "Could the Extended Trading of CSI 300 Index Futures Facilitate Its Role of Price Discovery?" *Journal of Futures Markets* 37 (7): 717–740.

Tian, F., K. Yang, and L. Chen. 2017. "Realized Volatility Forecasting of Agricultural Commodity Futures Using the HAR Model with Time-Varying Sparsity." *International Journal of Forecasting* 33 (1): 132–152.

Wang, X., C. Wu, and W. Xu. 2015. "Volatility Forecasting: The Role of Lunch-Break Returns, Overnight Returns, Trading Volume and Leverage Effects." *International Journal of Forecasting* 31 (3): 609–619.

Wang, Y., F. Ma, Y. Wei, and C. Wu. 2016. "Forecasting Realized Volatility in a Changing World: a Dynamic Model Averaging Approach." *Journal Of Banking & Finance* 64: 136–149. doi:10.1016/j.jbankfin.2015.12.010.

Wei, Y., J. Liu, X. Lai, and Y. Hu. 2017. "Which Determinant Is the Most Informative in Forecasting Crude Oil Market Volatility: Fundamental, Speculation, or Uncertainty?" *Energy Economics* 68: 141–150.

Wei, Y., Y. Wang, and D. Huang. 2010. "Forecasting Crude Oil Market Volatility: Further Evidence Using GARCH-class Models." *Energy Economics* 32 (6): 1477–1484.

Yang, J., Z. Yang, and Y. Zhou. 2012. "Intraday Price Discovery and Volatility Transmission in Stock Index and Stock Index Futures Markets: Evidence from China." *Journal of Futures Markets* 32 (2): 99–121.

Zhang, L., P. A. Mykland, and Y. Ait-Sahalia. 2005. "A Tale of Two-Time Scales." *Journal of the American Statistical Association* 100: 1394–1411.