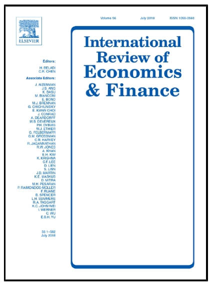
Accepted Manuscript



Volatility forecasting of crude oil market: Can the Regime Switching GARCH model beat the single-regime GARCH models?

Yue-Jun Zhang, Ting Yao, Ling-Yun He, Ronald Ripple

PII:	S1059-0560(16)30384-7
DOI:	10.1016/j.iref.2018.09.006
Reference:	REVECO 1687
To appear in:	*International Review of Economics and Finance*
Received Date:	23 December 2016
Accepted Date:	21 September 2018

Please cite this article as: Yue-Jun Zhang, Ting Yao, Ling-Yun He, Ronald Ripple, Volatility forecasting of crude oil market: Can the Regime Switching GARCH model beat the single-regime GARCH models?, *International Review of Economics and Finance* (2018), doi: 10.1016/j.iref.2018.09.006

Volatility forecasting of crude oil market: Can the Regime Switching GARCH model beat the single-regime GARCH models?

Yue-Jun Zhang [a,b*], Ting Yao [a,b], Ling-Yun He [c*], Ronald Ripple [d]

[a] Business School, Hunan University, Changsha 410082, China
[b] Center for Resource and Environmental Management, Hunan University, Changsha 410082, China
[c] Institute of Resource, Environment and Sustainable Development Research, Jinan University, Guangzhou 510632, China
[d] Collins College of Business, University of Tulsa, 800 South Tucker Drive, Tulsa, Oklahoma, USA

Biographical notes

Prof. Dr. Yue-Jun Zhang is a Professor at Business School, Hunan University, China, and the Director of Center for Resource and Environmental Management, Hunan University. He got his PhD degree in Energy Economics from Chinese Academy of Sciences (CAS) in 2009. His research interests include oil market pricing, forecast and risk management. Dr. Zhang has published more than 70 articles in peer-reviewed journals, such as Energy Economics. Dr. Zhang was a visiting scholar at Energy Studies Institute (ESI) of National University of Singapore (NUS).

Ms. Ting Yao is a PhD candidate at Business School, Hunan University, China. Her research interests focus on oil pricing mechanisms and modeling.

Prof. Dr. Ling-Yun He is a Professor at Institute of Resource, Environment and Sustainable Development Research, Jinan University, China. His research interests focus on crude oil modeling, commodity futures market analysis.

Prof. Dr. Ronald Ripple is a Mervin Bovaird Professor of Energy Business and Finance at Collins College of Business, University of Tulsa, USA. He has worked in the field of oil and gas market for several decades. Now he is also the vice-President of International Association of Energy Economics (IAEE).

Corresponding authors

Yue-Jun Zhang, PhD, Professor

Ling-Yun He, PhD, Professor

* Corresponding author. Tel.: 86-731-88822899. E-mail: zyjmis@126.com (Yue-Jun Zhang), lyhe@amss.ac.cn (Ling-Yun He).

Acknowledgments

We gratefully acknowledge the financial support from the National Natural Science Foundation of China (nos. 71322103, 71431008 and 71774051), National Program for Support of Top-notch Young Professionals (no. W02070325), Changjiang Scholars Program of the Ministry of Education of China (no. Q2016154), China Scholarship Council (no. 201606135020) and Hunan Youth Talent Program. Meanwhile, we would like to thank Dr. Klein and Walther for their technical support. We would also like to thank the editor, anonymous reviewers, and the seminar participants at the Center for Resource and Environmental Management of Hunan University for their insightful comments.

Volatility forecasting of crude oil market: Can the regime switching GARCH models beat the single-regime GARCH models?

Abstract: GARCH-type models are frequently used to forecast crude oil price volatility, and whether we should consider multiple regimes for the GARCH-type models is of great significance for the forecasting work but does not have a final conclusion yet. To that end, this paper estimates and forecasts crude oil price volatility using three single-regime GARCH (i.e., GARCH, GJR-GARCH and EGARCH) and two regime-switching GARCH (i.e., MMGARCH and MRS-GARCH) models. Furthermore, the Model Confidence Set (MCS) procedure is employed to evaluate the forecasting performance. The in-sample results show that the MRS-GARCH model provides higher estimation accuracy in weekly data. However, the out-of-sample results show the limited significance of considering the regime switching. Overall, our results indicate that the incorporation of regime switching does not perform significantly better than the single-regime GARCH models. The findings are proved to be robust to both daily and weekly data for WTI and Brent over different time horizons.

Keywords: Crude oil market; Volatility forecasting; GARCH; Regime switching; MCS

JEL Classification: G15; E17

1. Introduction

The vital role of crude oil price in macro-economic development is conclusive. In particular, crude oil is a crucial upstream product in the supply chain, and sudden and huge volatility of crude oil prices may often lead to shocks in productive capacity and further bring economic fluctuations. As a result, the oil-importing and oil-exporting countries suffer economic instability due to the changes in purchasing power. Moreover, crude oil is a peculiar commodity with evident political and financial properties, and some non-fundamental factors (such as speculation, geopolitics, and US dollar exchange rate) also influence crude oil price movement. Hence, modeling and forecasting crude oil market volatility is a vital and yet complex issue in both commodity and financial markets (Kilian and Vigfusson, 2011; Mi et al., 2015; Tang et al., 2015; Zhang and Wang, 2013; Fan et al., 2008; Zhang et al., 2015; Zhang and Yao, 2016; Bouri et al., 2017).

In order to forecast crude oil price volatility, the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model proposed by Bollerslev (1986) is widely used due to its good performance in capturing the time-varying features of the data (Wang and Wu, 2012; Agnolucci, 2009; Hou and Suardi, 2012; Marzo and Zagaglia, 2010; Mohammadi and Su, 2010; Wang and Nishiyama, 2015; Lux et al., 2016). However, the standard GARCH model is intrinsically symmetric, and the forecasting results with the standard GARCH model may be biased when skewed time series are considered (Franses and Dijk, 1996). To address this problem, some nonlinear and asymmetric GARCH models are proposed for forecasting crude oil price volatility, like GJR-GARCH model by Glosten et al. (1993) and EGARCH model by Nelson (1991).

However, several studies present empirical evidence that crude oil price features a prominent long memory in its volatility (Chkili et al., 2014). It is claimed that when the model has high persistence the forecasts suffer, so that we can expect an improvement in forecasting performance when model persistence is reduced by properly accounting for structural changes (Hillebrand, 2005; Teterin et al., 2016). Unfortunately, the GARCH-type models above basically focus on one regime of crude oil price changes, and often fit the in-sample and out-of-sample data with the same pattern but ignore potential structural changes (Lamoureux and Lastrapes, 1990; Timmermann 2000).

To address this problem, regime switching models are employed to forecast asset price volatility. Cai (1994) and Hamilton and Susmel (1994) introduce the regime switching process (Hamilton, 1988, 1989) into the GARCH model, in order to consider potential structural changes. In particular, the Markov Regime[1] Switching GARCH (MRS-GARCH) model permits the regimes in the Markov chain to have different GARCH behaviors, i.e., different volatility structures, so as to extend the GARCH model to the dynamic forms and realize better estimating and forecasting performance (Klaassen, 2002; Haas et al., 2004; Marcucci, 2005; Zhang and Wang,

[1] The regimes, which are identified by the Markov regime switching model, can be viewed as unobservable states. A regime may last for a random period, and when a switching takes place, the current regime may be replaced by another regime (Ang and Timmermann, 2012). In empirical estimates, regimes identified by econometric models often respond to different periods in policy or secular changes, such as the business cycle expansion and recession (Hamilton, 1989), or the high and low volatility regimes (Ang and Timmermann, 2012).

2015; Zhang and Zhang, 2015).

Unlike the Markov Regime Switching models, there is another kind of regime switching GARCH models, i.e., the mixture models, where the switching probabilities depend only on the lagged observed variables. For instance, Li et al. (2013) propose a regime switching model named the Mixture Memory GARCH (MMGARCH). The MMGARCH model mixes the standard GARCH and Fractionally Integrated GARCH (FIGARCH) models by interpreting a weight coefficient as a regime indicator. As a result, the MMGARCH is able to reveal the existence of different volatility structures. Klein and Walther (2016) investigate the variance and Value-at-Risk forecasting using the MMGARCH model and show that MMGARCH outperforms the other discrete volatility models due to its dynamic approach in varying the volatility level and memory of the process.

It should be noted that although the Markov Regime Switching models are effective in capturing potential state transition and nonlinearity in crude oil price volatility, regarding crude oil price volatility forecasting, the role of regime switching is not conclusive. For instance, Chang (2012) finds that the regime switching plays a decisive role in forecasting volatility by incorporating the regime switching into the EGARCH model. However, Sévi (2014) claims that the model seems to perform quite well in-sample, but the potential out-of-sample performance is weak for the difficulties of Markov regime switching models for forecasting.

Meanwhile, since the forecasting accuracy of GARCH-type models often appears sensitive to data frequency and time horizon (Manera et al., 2007; Zhang et al., 2015), it is interesting to examine whether and how the results change when crude oil price volatility is forecasted at different data frequencies, i.e., daily and weekly data[2], and different time horizons.

Furthermore, among the literature on the forecasting performance of the MRS-GARCH model, the results are mainly evaluated based on a few statistical loss functions (Fong and See, 2002; Chang, 2012; Nomikos and Pouliasis, 2011). However, the satisfactory statistical results are likely to appear by chance rather than due to the inherent superiority of the model (White, 2000; Hansen, 2005). Under this circumstance, the model confidence set (MCS) procedure proposed by Hansen et al. (2011) is employed, which infers the set of best models from the initial model set statistically. Compared to the conventional tests, the advantages of the MCS procedure are attractive. First, the MCS test does not need a pre-set benchmark or a pairwise comparison, and it is useful when there is not an obvious benchmark. Second, the MCS approach acknowledges the limitations of the data so that uninformative data yield an MCS with many models while informative data yield an MCS with only a few models. Finally, the MCS procedure generates a model confidence set, which allows for the possibility of more than one best model.

Based on the consideration above, this paper extends related research in the following ways. First, it considers both MRS-GARCH and MMGARCH models, while in related literature one regime-switching model is often considered, such as Klein and Walther (2016). Second, it conducts the evaluation based on a broader sample, i.e.,

[2] In fact, we also consider the monthly data; however, as for MMGARCH, the size of monthly data is limited, so the optimization cannot converge. Therefore, here, we only report the results based on the daily and weekly data, but the detailed monthly results of the other four models can be obtained upon request.

both daily and weekly WTI and Brent crude oil price data during the period ranging from 01/02/1986 to 06/30/2017, compared with previous related literature, such as Chang (2012) for weekly WTI crude oil prices. Finally, based on the results of six loss functions for the forecasting performance, it further uses the MCS procedure to examine whether the differences are statistically significant while the related research (such as Fong and See (2002)) often only show the results based on loss functions.

The remainder of this paper is organized as follows. Section 2 describes the data and methods used for the empirical analysis. Section 3 introduces the empirical results and discussion, and Section 4 concludes this paper.

2. Data and methods

2.1 Data description

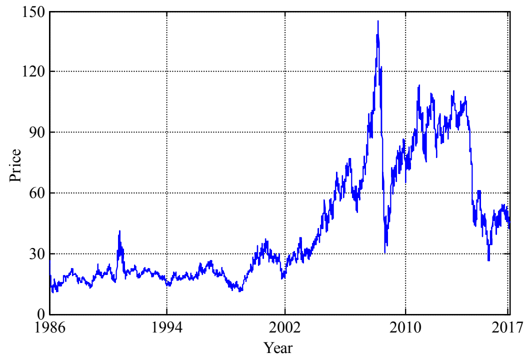
This paper uses daily and weekly spot price data for West Texas Intermediate (WTI) and Brent crude oil, which are obtained from the U.S. Energy Information Administration (EIA). The full sample ranges from 01/02/1986 to 06/30/2017 and specific sample periods for in-sample estimation and out-of-sample forecasting are shown in Table 1.

Table 1. Sample period for different frequencies

	Full sample	Obs.	In-sample	Obs.	Out-of-sample	Obs.

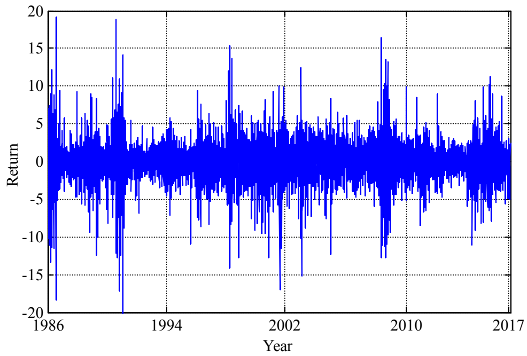
	Full sample	Obs.	In-sample	Obs.	Out-of-sample	Obs.
Panel A: Daily data						
WTI	01/02/1986-06/30/2017	7945	01/02/1986-12/31/2009	6056	01/04/2010-06/30/2017	1889
Brent	05/20/1987-06/30/2017	7645	05/20/1987-12/31/2017	5751	01/04/2010-06/30/2017	1894
Panel B: Weekly data						
WTI	01/03/1986-06/30/2017	1644	01/03/1986-12/25/2009	1252	01/01/2010-06/30/2017	392
Brent	05/15/1987-06/30/2017	1573	05/15/1987-12/25/2009	1181	01/01/2010-06/30/2017	392

In this paper, we define crude oil price returns at time t as $r_t = 100 * (\log p_t - \log p_{t-1})$, where p_t denotes crude oil price at time t, and define the squared returns as the actual volatility of WTI crude oil prices (Wang et al., 2016). The daily crude oil prices, returns, and actual volatility are shown in Figure 1, which indicate that crude oil prices experience a high uncertainty over time. Hence, it is of utmost importance to forecast crude oil price volatility using appropriate approaches. Besides, we can also find that driven by the economic boom in emerging economies, crude oil prices increased from 2003 to 2008, and due to the global economic recession triggered by the subprime crisis in the U.S., crude oil prices dropped quickly in the second half of 2008. Therefore, we select the period from 1998 to 2009 as the sub-sample period for robustness test.

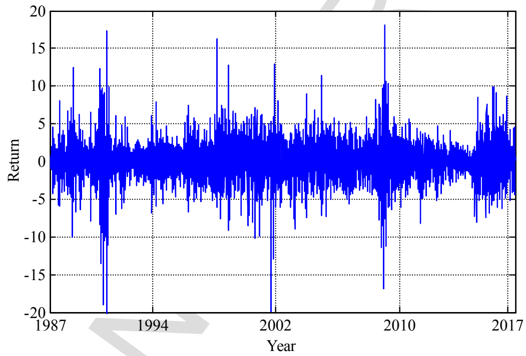


(a) WTI prices

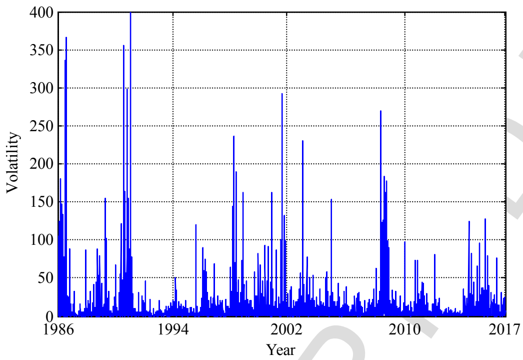
(b) Brent prices



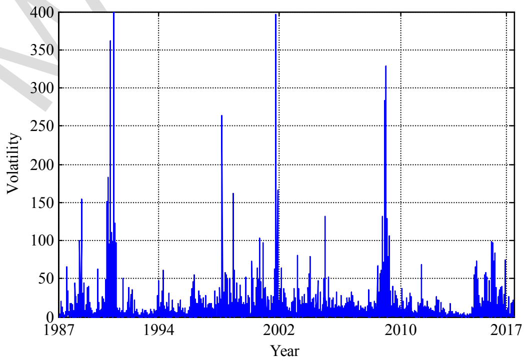
(c) WTI returns



(d) Brent returns



(e) WTI volatilities



(f) Brent volatilities

Fig. 1. Daily crude oil prices, returns and volatilities

Table 2 shows the descriptive statistics of crude oil returns. The descriptive statistics of WTI and Brent returns are similar, and it can be explained by the highly integrated world crude oil markets (Wang et al., 2016). Meanwhile, the descriptive statistics of daily and weekly returns are similar in general, although the values are not that close. Specifically, the mean values of weekly returns are larger than those of daily returns, but both are close to zero. The statistical results of Jarque–Bera test (1980) show that the null hypothesis of a normal distribution is rejected at the 1% significance level. Moreover, the negative skewness and positive excess kurtosis indicate a fat-tailed distribution. The Ljung and Box (1978) Q statistics for serial autocorrelation reject the null hypothesis of no autocorrelation up to the 10th or 20th orders at the 1% significance level, indicating the existence of autocorrelations. The F statistics of ARCH test imply the existence of heteroscedasticity in crude oil returns at the 1% significance level. Table 2 also presents the results of unit root tests. Specifically, the results of the augmented Dickey and Fuller (1979) and Phillips and Perron (1988) (PP) tests reject the null hypothesis of a unit root in crude oil returns at the 1% significance level, which indicate that crude oil returns are stationary during the sample period.

Table 2. Descriptive Statistics for crude oil spot returns

	WTI daily	Brent daily	WTI weekly	Brent weekly
Mean	0.0074	0.0012	0.0334	0.0579
Max	19.1507	18.1297	25.1247	23.3615
Min	-40.6396	-36.1214	-19.2338	-23.2604
Std. Dev.	2.5339	2.2907	4.3665	4.2286
Skewness	-0.6515	-0.5385	-0.1352	-0.1828
Kurtosis	16.5720	16.6515	6.2419	5.9056
Jarque-Bera	61532.2300 [0.0000]	59726.1600 [0.0000]	724.5232 [0.0000]	561.7275 [0.0000]
$Q(10)$	47.8973 [0.0000]	25.8671 [0.0039]	54.2576 [0.0000]	72.8953 [0.0000]
$Q(20)$	61.3075 [0.0000]	65.3778 [0.0000]	78.3307 [0.0000]	90.5191 [0.0000]
$Q^2(10)$	739.8938 [0.0000]	811.6057 [0.0000]	545.2180 [0.0000]	290.8717 [0.0000]
$Q^2(20)$	1006.7486 [0.0000]	953.5571 [0.0000]	794.1073 [0.0000]	499.0485 [0.0000]

	WTI daily	Brent daily	WTI weekly	Brent weekly
ARCH(10)	44.5777 [0.0000]	56.7116 [0.0000]	24.1414 [0.0000]	14.5059 [0.0000]
ARCH(20)	24.3565 [0.0000]	29.6985 [0.0000]	14.1532 [0.0000]	10.5251 [0.0000]
ADF	-65.9913 [0.0001]	-84.6565 [0.0001]	-21.1512 [0.0000]	-32.6579 [0.0000]
PP	-90.7724 [0.0001]	-84.6127 [0.0001]	-36.1172 [0.0000]	-32.7741 [0.0000]

Note: *P*-values are reported in the square brackets. Std. Dev. represents the standard deviation. Jarque-Bera is the statistics from Jarque and Bera (1980) test with the null hypothesis of normal distribution. ADF and PP denote statistics from the augmented Dickey and Fuller (1979) and Phillips and Perron (1988) unit root tests, respectively. The optimal lag length of the ADF test is determined by the Schwarz information criterion (SIC) (Schwarz, 1978). $Q(10)$, $Q(20)$ and $Q^2(10)$, $Q^2(20)$ denote the Ljung and Box (1978) statistics of the return and squared return series for up to 10th and 20th order serial autocorrelation, respectively.

2.2 Methods

(1) The single-regime GARCH models

According to Bollerslev (1986) and Sadorsky (1999), the standard linear GARCH model for WTI crude oil returns can be specified as follows[3]:

$$r_t = \delta + \varepsilon_t\,; \quad \varepsilon_t = \eta_t \sqrt{h_t}\,; \quad h_t = \omega_1 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \tag{1}$$

where α_0, α_1 and β_1 must be positive to guarantee a positive conditional variance; ε_t represents the residual series; $\alpha_1 + \beta_1 < 1$ is set to ensure the stationarity, and $\alpha_1 + \beta_1$ denotes the persistence of shocks to volatility when $\alpha_1 + \beta_1$ converge to 1. As the descriptive statistics indicate a fat-tailed distribution, we adopt the *student's t* distribution with a degree of freedom ν for it allows for fat tails. However, according to the symmetricity of GARCH, the responses of volatility to positive and negative returns are the same, which is inconsistent with the facts.

In order to consider the asymmetric effect of crude oil price volatility

[3] Here, we select the GARCH(1,1) specification in this paper. In fact, as Bollerslev et al. (1994) claim that the GARCH(1,1) model performs well in most applied situations. Similarly, Sadorsky (2006) also demonstrates that the GARCH(1,1) specification fits for crude oil volatility well.

(sometimes referred to as 'leverage effect'), we employ the nonlinear GJR-GARCH model proposed by Glosten et al. (1993) and the Exponential GARCH (EGARCH) model by Nelson (1991), whose variance equations are defined in Eq. (2) and Eq. (3), respectively.

$$h_t = \omega_1 + [\alpha_1 + \xi I(\varepsilon_{t-1} < 0)]\varepsilon_{t-1}^2 + \beta_1 h_{t-1} \tag{2}$$

where $I(\varepsilon_{t-1} < 0)$ is an indicator function, and if $\varepsilon_{t-1} < 0$, $I(\varepsilon_{t-1} < 0) = 1$; otherwise, $I(\varepsilon_{t-1} < 0) = 0$. According to Hentschel (1995), h_t can be guaranteed to be positive, if $\omega_1 > 0, \alpha_1 \geq 0, \beta_1 \geq 0$ and $\xi \geq 0$. The coefficient ξ $(\xi \geq 0)$ captures the asymmetric effect.

$$\log(h_t) = \omega_1 + \alpha_1 \left| \frac{\varepsilon_{t-1}}{h_{t-1}} \right| + \xi \frac{\varepsilon_{t-1}}{h_{t-1}} + \beta_1 \log(h_{t-1}) \tag{3}$$

(2) The Regime switching GARCH models

Apart from the three single-regime GARCH models above, two regime switching GARCH models are also introduced in this paper, i.e., MMGARCH and MRS-GARCH models. According to Li et al. (2013) and Klein and Walther (2016), the MMGARCH model consists of a plain GARCH component and a Fractionally Integrated GARCH [4] (FIGARCH) component (Baillie et al., 1996) in volatility, and can be defined as

$$y_t = \delta + \varepsilon_t\,; \quad \varepsilon_t = \eta_t \sqrt{h_t}\,; \quad h_t = z_t h_{1,t} + (1 - z_t) h_{2,t} \tag{4}$$

where $h_{1,t} = \omega_1 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{1,t-1}$ and $h_{2,t} = \omega_2 + (1 - \beta_2 L - (1-L)^d)\varepsilon_t^2 + \beta_2 h_{2,t-1}$, denoting the GARCH and FIGARCH components, respectively; d $(0 \leq d \leq 1)$ is the fractional differencing parameter and its value depends on the decay rate of a shock to conditional volatility; the FIGARCH nests a GARCH when $d = 0$ and an Integrated

[4] Fractional integration, in which the differencing parameter can be a fraction, is often introduced to standard models by a fractional differencing parameter in order to account for long memory processes. For FIGARCH, the long-term behavior of volatility is modeled by the fractional differencing parameter, while the short-term volatility is modeled by conventional ARCH and GARCH parameters.

GARCH (IGARCH) when $d=1$; $z_t \in [0,1]$ guarantees that the GARCH and FIGARCH components are applied as instantaneous, conditional variance at time t.

Different from MMGARCH, MRS-GARCH model allows the parameters switching between different regimes following the Markov process. Specifically, the regime variable may switch according to a Markov process and the switching probability from regime i at time $t-1$ to regime j at time t is equal to $P(s_t = j \mid s_{t-1} = i) = p_{ji}$.

The conditional mean and conditional variance following the GARCH process as well as the expectation of squared innovations can be specified as Eqs. (5), (6) and (7), respectively.

$$r_t = \delta^{(i)} + \varepsilon_t\,; \quad \varepsilon_t = \eta_t \sqrt{h_t} \tag{5}$$

$$h_t^{(i)} = \alpha_0^{(i)} + \alpha_1^{(i)} \varepsilon_{t-1}^2 + \beta_1^{(i)} E_{t-1}\{h_{t-1}^{(i)} \mid s_t\} \tag{6}$$

$$E_{t-1}\{h_{t-1}^{(i)} \mid s_t\} = \sum_{j=1}^{2} \tilde{p}_{ji,t-1}[(\mu_{t-1}^{(j)})^2 + h_{t-1}^{(j)}] - [\sum_{j=1}^{2} \tilde{p}_{ji,t-1} \mu_{t-1}^{(j)}]^2 \tag{7}$$

where $i, j = 1,2$ denotes the two regimes for the MRS-GARCH model[5], s_t is the regime variable, $\tilde{p}_{ji,t} = \Pr(s_t = j \mid s_{t+1} = i, \varsigma_{t-1}) = \dfrac{p_{ji} \Pr(s_t = j \mid \varsigma_{t-1})}{\Pr(s_{t+1} = i \mid \varsigma_{t-1})} = \dfrac{p_{ji} p_{j,t}}{p_{i,t-1}}$, and ς_{t-1} represents the information at time $t-1$.

Then, according to Klaassen (2002)[6], we sum the actual volatility of WTI crude oil prices during k periods to forecast the k-step-ahead volatility of crude oil prices

[5] The literature focused on regime switching of crude oil price volatility often adopts the two-regime or three-regime switching model. We test the number of regimes by the useful tool proposed by Hansen (1992, 1996) and estimate the samples with both two-regime and three-regime switching models. The results indicate that most of the parameters in the three-regime switching model are not statistically significant, but the significance of the parameters in the two-regime switching model appears acceptable. Hence, we select the two-regime switching model in this paper.

[6] The main reasons for us to adopt the regime switching GARCH models of Klaassen (2002) other than those by Gray (1996) or Haas et al. (2004) are as follows. The model of Klaassen (2002) allows higher flexibility in capturing the persistence of shocks to volatility. Meanwhile, according to Klaassen (2002), the multi-step-ahead volatility forecasts are expressed straightforward, so that they can be calculated recursively as in standard GARCH models.

at time t as Eq. (8).

$$\hat{h}_{t,t+k} = \sum_{\gamma=1}^{k} \hat{h}_{t,t+\gamma} = \sum_{\gamma=1}^{k} \sum_{i=1}^{2} \Pr(s_{\tau} = i \mid \varsigma_{t-1}) \hat{h}_{t,t+\gamma}^{(i)} \tag{8}$$

where $\hat{h}_{t,t+\tau}^{(i)} = \alpha_0^{(i)} + (\alpha_1^{(i)} + \beta_1^{(i)}) E_t\{h_{t,t+\tau-1}^{(i)} \mid s_{t+\tau}\}$ is the γ-step-ahead volatility forecast of regime i at time t.

(3) The evaluation criteria for forecasting performance

In order to evaluate the in-sample estimation and out-of-sample forecasting performance for crude oil price volatility, we apply six widely used statistical loss functions:

$$MAE1 = \frac{1}{T-N} \sum_{t=N+1}^{T} \left| \sqrt{\hat{h}_t} - \sqrt{h_t} \right| \tag{9}$$

$$MAE2 = \frac{1}{T-N} \sum_{t=N+1}^{T} \left| \hat{h}_t - h_t \right| \tag{10}$$

$$MSE1 = \frac{1}{T-N} \sum_{t=N+1}^{T} \left(\sqrt{\hat{h}_t} - \sqrt{h_t} \right)^2 \tag{11}$$

$$MSE2 = \frac{1}{T-N} \sum_{t=N+1}^{T} \left(\hat{h}_t - h_t \right)^2 \tag{12}$$

$$QLIKE = \frac{1}{T-N} \sum_{t=N+1}^{T} \left(\ln(\hat{h}_t) - h_t / \hat{h}_t \right)^2 \tag{13}$$

$$R^2LOG = \frac{1}{T-N} \sum_{t=N+1}^{T} \left(\ln(h_t / \hat{h}_t) \right)^2 \tag{14}$$

where h_t and $\hat{h}_t$ represent the actual and forecasted volatility, respectively; T and N stand for the number of full sample and in-sample observations, respectively; MAE and MSE indicate the mean absolute error and mean squared error, respectively; $QLIKE$ denotes the loss implied by a Gaussian likelihood (Bollerslev et al., 1994); and R^2LOG is the logarithmic loss function (Pagan and Schwert, 1990) and can penalize the volatility forecasting asymmetry for the high and low level volatilities.

Unfortunately, these loss functions cannot distinguish whether the loss differences are statistically significant. To overcome this problem, we employ the Model Confidence Set (MCS) procedure proposed by Hansen et al. (2011) to evaluate the out-of-sample forecasting performance of the models. The MCS procedure proceeds as follows. For a given set M^0, there is a finite number of objects which are indexed by $a = 1,...,m_0$. The objects are determined by a loss function and the loss of object a in the tth iteration is defined as $L_{a,t}$. The relative performance variables can be defined as $d_{ab,t} \equiv L_{a,t} - L_{b,t}$, for all $a,b \in M^0$. The set of superior objects M^* is defined by $M^* \equiv \{a \in M^0 : \mu_{ab} \leq 0 \text{ for all } b \in M^0\}$, where $\mu_{ab} = \mathrm{E}(d_{ab,t})$, and M^* is determined by eliminating the inferior elements in M_0 through a set of significance tests. The null hypothesis is $H_{0,M} : \mu_{ab} = 0 \text{ for all } a,b \in M$, where $M \subset M^0$.

The MCS procedure determines M^* by an equivalence test (ϖ_M) and an elimination rule (e_M). The equivalence test is used to test the hypothesis $H_{0,M}$ at significance level ϑ for any $M \subset M^0$, and in the event $H_{0,M}$ is rejected, the object of M that is to be eliminated is determined by e_M. After repeating the test, we finally identify the model confidence set, i.e., $\hat{M}^*_{1-\vartheta}$, which consists of the surviving objects, with an MCS *p*-value. The MCS *p*-value indicates that the MCS is a random subset of models that contain M^* with a certain probability *p*.

3. Results and discussions

3.1 In-sample estimation results

According to the models introduced in Section 2, we estimate the three single-regime GARCH models, i.e., the standard linear GARCH model, the nonlinear GJR-GARCH and EGARCH models, and the two regime switching GARCH models, i.e., the

MMGARCH and the two-regime MRS-GARCH models, respectively. The estimation results for WTI and Brent are shown in Tables 3 and 4, respectively. For one thing, we see that for the GARCH, GJR-GARCH and EGARCH models, the values of $\alpha+\beta$ are close to 1. This indicates a high degree of volatility persistence in WTI and Brent crude oil markets. For another, the values of ξ for WTI crude oil market are significantly different from zero at the 1% level in the GJR-GARCH and EGARCH models, indicating a leverage effect in crude oil market. However, the estimation results of ξ in WTI crude oil market is mixed, for the coefficient is significant in the GJR-GARCH model but not significant in the EGARCH model. Meanwhile, the standard linear GARCH model satisfies the second and fourth moment conditions provided by Ling and McAleer (2002a, 2002b).

Table 3. The estimation results of single-regime GARCH models for WTI

	GARCH daily	GJR daily	EGARCH daily	GARCH weekly	GJR weekly	EGARCH weekly
δ	0.0611***	0.0601**	0.0549**	0.1683*	0.1502	0.1363
	(2.5437)	(2.5052)	(2.2870)	(1.6661)	(1.4445)	(1.3499)
ω_1	0.0722***	0.0094***	0.0918***	0.4944***	0.0995***	-0.0643*
	(6.0154)	(5.2222)	(10.1966)	(3.3181)	(3.1688)	(1.8915)
α_1	0.0655***	0.0524***	0.1455***	0.1106***	0.0645***	0.2096***
	(10.9153)	(7.5942)	(12.1236)	(4.8078)	(3.1617)	(5.5163)
ξ	-	0.0273***	-0.0153*	-	0.0685***	-0.0383*
		(2.9674)	(1.9220)		(2.6046)	(1.9172)
β_1	0.9222***	0.9269***	0.9890***	0.8711***	0.8715***	0.9649***
	(131.7471)	(0.0057)	(494.5000)	(37.8752)	(40.4323)	(74.2231)
ν	6.0380***	6.0331***	5.9791***	7.7462***	7.8955***	8.2806***
	(14.6199)	(14.5375)	(14.7633)	(5.5768)	(5.5523)	(5.4334)
$Q(10)$	23.9922***	23.9957***	24.7911***	36.1256***	35.7769***	34.7541***
	[0.0081]	[0.0080]	[0.0059]	[0.0000]	[0.0000]	[0.0000]
$Q(20)$	44.3887***	44.2092***	45.3367***	51.3892***	50.7936***	48.3179***
	[0.0010]	[0.0010]	[0.0010]	[0.0000]	[0.0000]	[0.0000]
$Q^2(10)$	30.5000***	29.2084***	56.2978***	6.9442	6.9861	6.4543
	[0.0010]	[0.0010]	[0.0000]	[0.7309]	[0.7274]	[0.7761]

	GARCH daily	GJR daily	EGARCH daily	GARCH weekly	GJR weekly	EGARCH weekly
$Q^2(20)$	37.5806***	36.2569***	61.1438***	17.9381	17.3749	16.3762
	[0.0100]	[0.0141]	[0.0000]	[0.5911]	[0.6283]	[0.6931]

Note: The *t*-statistics values and *p*-values are reported in the parentheses and square brackets, respectively. ***, ** and * denote the significance at the 1%, 5% and 10% levels, respectively. $Q(10)$, $Q(20)$ and $Q^2(10)$, $Q^2(20)$ denote the Ljung and Box (1978) statistics of the standardized residuals and squared standardized residuals for up to 10th and 20th order serial autocorrelation, respectively.

Table 4. The estimation results of single-regime GARCH models for Brent

	GARCH daily	GJR daily	EGARCH daily	GARCH weekly	GJR weekly	EGARCH weekly
δ	0.0634***	0.0607**	0.0489**	0.1807*	0.1794	0.1482
	(2.6424)	(2.5275)	(2.1243)	(1.7047)	(1.6459)	(1.3985)
ω_1	0.1353***	0.0054***	0.0903***	0.8690***	0.0560***	-0.0714**
	(6.7650)	(4.6108)	(10.0295)	(3.4783)	(3.1952)	(2.1636)
α_1	0.0879***	0.0476***	0.1467***	0.1050***	0.0798***	0.1968***
	(9.7680)	(7.7628)	(11.2838)	(4.5669)	(6.0916)	(5.6228)
ξ	-	0.0237***	-0.0131	-	0.0449***	-0.0136
		(3.0511)	(1.6375)		(3.1844)	(0.7608)
β_1	0.8856***	0.9373***	0.9870***	0.8466***	0.8851***	0.9710***
	(98.4000)	(188.6665)	(329.0000)	(29.1917)	(73.1488)	(88.2727)
ν	6.1511***	6.1873***	5.9288***	9.6047***	9.6007***	9.5778***
	(12.6306)	(12.6271)	(13.1750)	(5.3182)	(5.3189)	(5.2166)
$Q(10)$	13.3438	13.4252	16.7851*	62.2161***	62.1937***	62.0526***
	[0.2050]	[0.2009]	[0.0789]	[0.0000]	[0.0000]	[0.0000]
$Q(20)$	20.4012	20.4318	23.6171	72.7718***	72.1570***	73.1789***
	[0.4329]	[0.4311]	[0.2600]	[0.0000]	[0.0000]	[0.0000]
$Q^2(10)$	17.3356*	17.0841*	25.1161***	3.8881	3.8561	4.9269
	[0.0668]	[0.0729]	[0.0052]	[0.9521]	[0.9316]	[0.9681]
$Q^2(20)$	24.3861	24.0410	30.1624**	15.3147	15.3709	16.2962
	[0.2257]	[0.2411]	[0.0673]	[0.7584]	[0.7563]	[0.6981]

Note: The *t*-statistics values and *p*-values are reported in the parentheses and square brackets, respectively. ***, ** and * denote the significance at the 1%, 5% and 10% levels, respectively. $Q(10)$, $Q(20)$ and $Q^2(10)$, $Q^2(20)$ denote the Ljung and Box (1978) statistics of the standardized residuals and squared standardized residuals for up to 10th and 20th order serial autocorrelation, respectively.

The estimation results of the MMGARCH are shown in Table 5. We find the existence of different volatility structure, which is consistent with the findings of Klein and Walther (2016). Furthermore, from both the daily and weekly results, we can conclude that the GARCH and FIGARCH components do not only differ in volatility levels but also in the persistence of shocks. This finding is identified by the value of d. If there is no difference in the memory structure (short or long memory), the fractional differencing parameter of MMGARCH would be $d = 0$. However, it is not the case in this paper. Hence, there exist two different variance components in each series. Specifically, for WTI daily, Brent daily and Brent weekly data, the first component has a long memory, with a hyperbolic decline of weights and a high volatility level, while the second component has a short memory, with weights declining exponentially and low unconditional volatility for the value of ω.

Table 5. The estimation results of the MMGARCH models

	WTI daily	WTI weekly	Brent daily	Brent weekly
ω_1	5.8976	0.4024	1.3857	7.6812
ω_2	0.0378	41.7908	0.0247	0.4024
α_1	0.4770	0.0867	0.1935	0.9790
β_1	0.5231	0.8654	0.8065	0.0207
β_2	0.0431	0.7442	0.0402	0.0872
d	0.9329	0.2561	0.9357	0.8797
v_1	6.5321	9.2548	6.7832	9.2846
v_2	6.6452	8.1635	6.9347	7.9824
$Q(10)$	16.0589*	36.6622*	22.3343**	55.4146**
	[0.0981]	[0.0646]	[0.0141]	[0.0258]
$Q(20)$	23.4392	50.2537	44.0259***	65.4743
	[0.2681]	[0.2043]	[0.0021]	[0.9811]
$Q^2(10)$	24.3043***	7.2819	33.6023***	3.9447
	[0.0070]	[0.6986]	[0.0000]	[0.9500]
$Q^2(20)$	28.9626*	13.9482	39.3149***	19.1782
	[0.0891]	[0.8327]	[0.0061]	[0.5100]

Note: The *p*-values are reported in the square brackets. ***, ** and * denote the significance at the 1%, 5% and 10% levels, respectively. $Q(10)$, $Q(20)$ and $Q^2(10)$, $Q^2(20)$ denote the Ljung and Box (1978) statistics of the standardized residuals and squared standardized residuals for up to 10th and 20th order serial correlation, respectively. As claimed by Li et al. (2013), the standard errors of the coefficients can be calculated only when in the normal distribution. As the descriptive statistics indicate the fat-tail, we adopt the *student's t* distribution because it allows for fatter tails than the normal distribution (Fong and See, 2002; Hou and Suardi, 20012). $Q(10)$, $Q(20)$ and $Q^2(10)$, $Q^2(20)$ denote the Ljung and Box (1978) statistics of the standardized residuals and squared standardized residuals for up to 10th and 20th order serial autocorrelation, respectively.

The estimation results of the MRS-GARCH are reported in Table 6. It is noted that instead of $\alpha_0^{(i)}$, we report the values of $\sigma^{(i)}$, which represent the standard deviations of returns conditional on each volatility regime, and can be calculated as $\sigma^{(i)} = (\alpha_0^{(i)} / (1-\alpha_1^{(i)} - \beta_1^{(i)}))^{1/2}$. From the daily results shown in Table 6, we can see that there exists regime switching in both WTI and Brent crude oil price volatility. We can define Regime 1 as the high-volatility regime while Regime 2 as the low-volatility regime according to the statistically significant values of $\sigma^{(i)}$. Specifically, for WTI, the value of $\sigma^{(i)}$ ranges from 1.1830 in Regime 1 to 0.1003 in Regime 2; for Brent, the value of $\sigma^{(i)}$ ranges from 0.0881 in Regime 1 to 0.0722 in Regime 2.

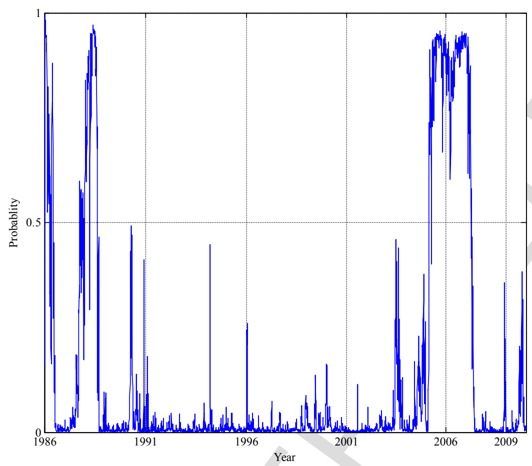
Table 6. The estimation results of the MRS-GARCH models

	WTI daily	WTI weekly	Brent daily	Brent weekly
$\delta^{(1)}$	0.0522	-3.1414***	0.0663**	-2.1837***
	(1.2132)	(-6.874)	(2.2849)	(-4.3327)
$\delta^{(2)}$	0.0570**	0.4150***	-0.0029	0.7410***
	(1.9635)	(3.7728)	(0.0796)	(4.8752)
$\sigma^{(1)}$	1.1830***	1.1455	0.0881***	0.0010
	(3.5208)	(0.6610)	(3.5237)	(0.4280)
$\sigma^{(2)}$	0.1003***	0.3024	0.0722***	0.0020
	(4.1777)	(1.3322)	(3.4357)	(0.2140)
$\alpha_1^{(1)}$	0.1908***	0.0469	0.0636***	0.1679**
	(2.8059)	(0.7688)	(2.8918)	(2.3989)
$\alpha_1^{(2)}$	0.0480***	0.0582***	0.0378***	0.0461*
	(8.0023)	(3.0645)	(6.2940)	(1.8455)
$\beta_1^{(1)}$	0.1028	0.9452***	0.8692***	0.8073***
	(0.4920)	(10.1645)	(34.7671)	(7.9149)
$\beta_1^{(2)}$	0.9357***	0.8839***	0.9495***	0.9005***
	(116.9666)	(33.9962)	(118.6923)	(29.0489)
p_{11}	0.9969***	0.8007***	0.9982***	0.7751***
	(49.8432)	(11.4398)	(99.8234)	(12.5015)
p_{22}	0.9993***	0.9741***	0.9996***	0.9347***
	(25.4464)	(97.4073)	(280.2128)	(38.9467)
$v^{(1)}$	6.6826***	2.7862***	9.3897***	33.3203
	(4.6666)	(6.6337)	(3.1840)	(0.0512)
$v^{(2)}$	6.8139***	16.8047*	7.2838***	7.3472***
	(12.3664)	(1.8209)	(10.3317)	(4.8851)
$Q(10)$	17.5471*	30.7332***	22.1778**	29.7516***
	[0.0628]	[0.0060]	[0.0142]	[0.0091]
$Q(20)$	24.4892	43.9745***	43.7389***	39.3284***
	[0.2217]	[0.0020]	[0.0021]	[0.0062]
$Q^2(10)$	23.3141**	7.9772	34.2706***	4.2096
	[0.0114]	[0.6312]	[0.0000]	[0.9374]
$Q^2(20)$	28.0553	14.8179	39.9818***	17.2721
	[0.1078]	[0.7871]	[0.0052]	[0.6347]

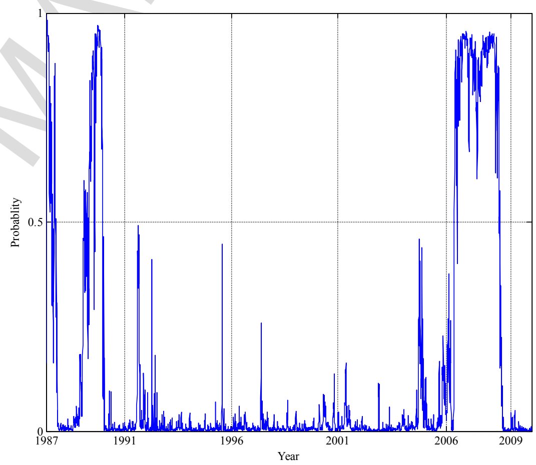
Note: The *t*-statistics values and *p*-values are reported in the parentheses and square brackets, respectively. ***, ** and * denote the significance at the 1%, 5% and 10% levels, respectively. $Q(10)$, $Q(20)$ and $Q^2(10)$,

$Q^2(20)$ denote the Ljung and Box (1978) statistics of the standardized residuals and squared standardized residuals for up to 10th and 20th order serial autocorrelation, respectively.

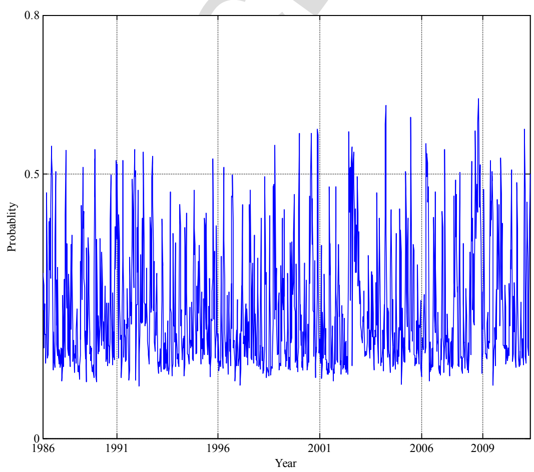
Meanwhile, Figure 2 presents the time-varying probabilities staying in Regime 1. If the probability at time *t* is larger than 0.5, we can say that the volatility stays in Regime 1 (high-volatility regime) at that time, and the volatility regime switches to Regime 2 (low-volatility regime) if the probability is smaller than 0.5. The results in Figure 2 confirm the existence of regime switching due to the probabilities larger than 0.5. This finding is in line with Chang (2012) and Nomikos and Pouliasis (2011), who find different volatility regimes by using different Markov regime switching models.

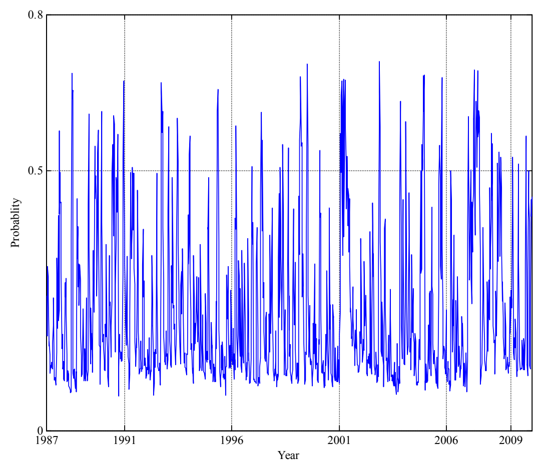


(a) WTI daily



(b) Brent daily





(c) WTI weekly (d) Brent weekly

Figure 2 The time-varying probabilities of staying in Regime 1

From the weekly estimation results in Table 6 and the time-varying probabilities in Figure 2, we can find that the regime switching in weekly data is more frequent than that in daily data. In weekly data, the conditional mean estimates are all significant while the conditional variance parameters are not significant. Hence, we define the two regimes according to the values of conditional mean. Specifically, the value of $\delta^{(i)}$ ranges from -3.1414 in Regime 1 to 0.4150 in Regime 2 for WTI, and from -2.1837 in Regime 1 to 0.7140 in Regime 2 for Brent. Consequently, we can define Regime 1 as the bearish regime while Regime 2 as the bullish regime. Besides, the Ljung and Box (1978) Q statistics on the standardized residuals and squared standardized residuals indicate that all the GARCH-type models effectively reduce the serial autocorrelation.

Furthermore, we evaluate the in-sample estimation performance of the candidate models for WTI and Brent using the MCS procedure introduced in Section 2, and the results are shown in Tables 7 and 8, respectively. From the results for WTI in Table 7, we can see that for daily data, GJR-GARCH performs the best; specifically, it is always included in the MCS. Besides, except for GJR-GARCH and MMGARCH, all the models are included in the MCS under five of the six loss functions. For weekly data, only GJR-GARCH and MRS-GARCH are always included in the MCS, while other models are excluded under some loss functions. The results for Brent in Table 8 indicate that, for daily data, the regime-switching GARCH models perform worse than the single-regime GARCH models, but we can find a significant improvement in weekly data after incorporating the Markov regime switching. For daily data, MMGARCH and MRSGARCH are only included in the MCS under one of the six loss functions while the single-regime switching GARCH models are included in the MCS under most loss functions. For weekly data, only MRS-GARCH model is always included in the MCS. It is noted that all the other models are always excluded.

Table 7. The results of the in-sample estimation evaluation for WTI

	MAE1	*MAE2*	*MSE1*	*MSE2*	*QLIKE*	R^2LOG
Panel A1: Daily results						
GARCH	1.6459	5.8498	3.5990	92.0141	**1.7769**	12.8163
GJR-GARCH	**1.6426**	**5.6634**	**3.4113**	**74.5010**	1.7806	**12.7688**
EGARCH	1.6464	5.8567	3.6040	92.6621	1.7770	12.8169
MMGARCH	1.7141	6.4297	4.0869	134.4231	1.8182	13.1552
MRS-GARCH	1.6537	5.7127	3.4634	77.3661	1.7910	12.8908
Panel B1: Weekly results						
GARCH	2.9460	18.0256	10.5819	592.3534	2.9992	13.1162
GJR-GARCH	2.8471	**16.5870**	**9.5256**	**430.6503**	2.9723	12.8575
EGARCH	2.9305	17.8809	10.4668	581.0140	2.9941	13.0838
MMGARCH	2.8322	16.7019	9.6403	483.2697	2.9627	12.8733
MRS-GARCH	**2.8278**	17.2794	10.0785	439.6190	**2.9477**	**12.7475**

Note: The numbers in the table refer to the values of loss functions. Numbers in bold indicate that the corresponding models have the lowest forecasting losses under a pre-specified criterion. Underlined numbers indicate that the corresponding models are included in the MCS. We perform 10,000 block bootstraps to generate p-values for the MCS test. The confidence level for MCS is 90%.

Table 8. The results of the in-sample estimation evaluation for Brent

	MAE1	*MAE2*	*MSE1*	*MSE2*	*QLIKE*	R^2LOG
Panel A1: Daily results						
GARCH	**1.5056**	**4.6824**	**2.8744**	55.3825	1.6479	12.3723
GJR-GARCH	1.5343	4.7446	2.8759	**45.1242**	1.6669	12.4647
EGARCH	1.5068	4.7032	2.8908	58.2080	**1.6473**	**12.3656**
MMGARCH	1.5733	5.2194	3.3129	81.3376	1.6878	12.7156
MRS-GARCH	1.7158	5.8184	4.0008	84.0068	1.8698	14.0032
Panel B1: Weekly results						
GARCH	2.7465	15.2449	8.7984	337.5372	2.9411	12.7601
GJR-GARCH	2.7990	15.7675	9.0927	352.1442	2.9552	12.7775
EGARCH	2.7477	15.2620	8.8109	338.7961	2.9414	12.7622
MMGARCH	2.7788	15.8826	9.3577	402.4160	2.9440	12.8651
MRS-GARCH	**2.4877**	**13.2358**	**7.3413**	**266.4160**	**2.8160**	**11.8565**

Note: The numbers in the table refer to the values of loss functions. Numbers in bold indicate that the corresponding models have the lowest forecasting losses under a pre-specified criterion. Underlined numbers indicate that the corresponding models are included in the MCS. We perform 10,000 block bootstraps to generate p-values for the MCS test. The confidence level for MCS is 90%.

Based on the analysis above, we can find that the MRS-GARCH model provides a significant improvement in estimating weekly crude oil price volatility. The different performance of MRS-GARCH in daily and weekly data may be sourced from the frequency of regime switching. As shown above, regime switching occurs more frequently in weekly data than in daily data.

3.2 Out-of-sample volatility forecasting results

A model that can well fit the in-sample data does not necessarily imply the model can accurately forecast the volatility for the out-of-sample data. Moreover, as accurate crude oil price volatility forecasting can effectively contribute to portfolio allocation along with risk investment and measurement, investors are likely to focus more on the out-of-sample forecasting performance of models (Wei et al., 2010). Hence, we forecast crude oil price volatility using three single-regime GARCH models and two regime switching GARCH models based on different data frequencies (i.e., daily and weekly data). We also evaluate their forecasting performance using the MCS procedure as mentioned above.

Table 9 shows the forecasting results for WTI crude oil volatility. From the daily results, we can find that the single-regime EGARCH model result performs the best at all the forecasting horizons. Specifically, the single-regime EGARCH model results have lower forecasting losses than the regime-switching GARCH and other single-regime GARCH models, indicating a higher forecasting accuracy. The MRS-GARCH model performs the worst at the 1-day ahead horizon, and the single-regime GJR-GARCH model performs the worst at other forecasting horizons.

Specifically, the 1-day ahead forecasting results indicate that the regime-switching GARCH models are excluded in the MCS under most criteria, while the non-linear single-regime GARCH models are included in the MCS under most criteria. The 5-day and 10-day ahead forecasting results are similar. Specifically, for the 5-day ahead forecasting results, the single-regime EGARCH model performs the best, since it survives in the MCS under five of the six loss functions. MRS-GARCH model is excluded from the MCS under all the six loss functions, and MMGARCH model is included in the MCS under two of the six loss functions. For the 10-day ahead forecasting results, MMGARCH is included in the MCS under four of the six loss functions, but its performance is still worse than that of EGARCH (included under five loss functions). Meanwhile, the MRS-GARCH model is included in the MCS under one of the six loss functions, and its performance is only superior to that of GJR-GARCH. The 22-day ahead forecasting results indicate that except for GJR-GARCH model, all models are included in the MCS under all the six loss functions.

In addition, the weekly results, as shown in Table 9, indicate that the regime-switching GARCH forecasts display higher losses than the single-regime GARCH-type ones. EGARCH model displays the best performance at 1-week ahead horizon and GJR-GARCH performs the best at other time horizons. Specifically, MRS-GARCH is always excluded from the MCS at 1-week ahead forecasting horizon, while MMGARCH is included under four of the six loss functions. From the 2-week and 3-

week ahead forecasting results, we can see that MMGARCH and MRS-GARCH are always excluded from the MCS, while EGARCH is included in the MCS under the *MSE2* criterion and the GJR-GARCH is included in the MCS under six and five loss functions, respectively. It indicates that GJR-GARCH and EGARCH models provide a higher forecasting performance than the regime switching GARCH models. At 4-week ahead horizon, the MMGARCH is never included in the MCS, and MRS-GARCH is included in the MCS under three of the six loss functions. However, GJR-GARCH and EGARCH are included in the MCS under six and three loss functions, respectively.

Table 9. The forecasting results for WTI crude oil volatility

	MAE1	*MAE2*	*MSE1*	*MSE2*	*QLIKE*	R^2LOG
Panel A1: 1-day ahead volatility forecasting for the daily data						
GARCH	1.1281	4.6385	2.0927	94.2412	2.3207	7.5980
GJR-GARCH	**1.0442**	**4.3408**	2.1696	108.7705	2.7720	**6.5367**
EGARCH	1.0877	4.4386	**1.8627**	**82.3466**	**2.1421**	7.5079
MMGARCH	1.0605	4.4029	1.9721	96.1171	2.3220	7.0354
MRS-GARCH	1.1255	4.6461	2.0976	95.1815	2.2666	7.5949
Panel A2: 5-day ahead volatility forecasting for the daily data						
GARCH	1.1924	11.4885	2.4132	945.0000	3.7881	0.7266
GJR-GARCH	1.7355	16.3692	7.1209	453.3179	5.1085	1.4601
EGARCH	**1.1531**	**10.9601**	**2.1527**	**390.6444**	**3.7645**	0.7131
MMGARCH	1.1557	11.7264	2.6972	583.6107	3.7977	**0.6163**
MRS-GARCH	1.2936	12.7283	3.1199	592.9501	3.8444	0.7892
Panel A3: 10-day ahead volatility forecasting for the daily data						
GARCH	1.2175	16.4463	2.3789	799.0070	4.4797	0.3357
GJR-GARCH	2.3542	30.8438	13.6596	1475.2480	6.0799	1.3932
EGARCH	1.1960	**15.7685**	**2.1750**	**681.1436**	**4.4751**	0.3385
MMGARCH	**1.1814**	17.5348	2.9488	1236.9323	4.4768	**0.2583**
MRS-GARCH	1.3704	19.2937	3.7226	1286.0567	4.5366	0.3895
Panel A4: 22-day ahead volatility forecasting for the daily data						
GARCH	1.5418	**30.1099**	**3.5697**	**2177.3280**	5.2826	0.2316
GJR-GARCH	4.0427	73.3384	32.2480	5426.8530	7.2442	1.6720
EGARCH	1.5841	30.3571	3.6357	2059.7850	5.2865	0.2454
MMGARCH	**1.4687**	33.0403	4.4057	3789.0080	**5.2695**	**0.1448**
MRS-GARCH	1.5128	31.4685	4.6488	3191.9420	5.3016	0.2096
Panel B1: 1-week ahead volatility forecasting for the weekly data						
GARCH	2.0008	13.6367	5.9977	537.1286	3.4665	7.0816
GJR-GARCH	1.8906	12.8108	5.2230	439.2576	**3.3042**	6.8480
EGARCH	1.8913	12.7157	**5.1748**	**433.2288**	3.3121	6.9534
MMGARCH	**1.7837**	**12.1077**	6.1939	634.8069	5.8950	**5.7234**
MRS-GARCH	1.9671	13.5019	5.8237	513.7449	3.9970	6.9919
Panel B2: 2-week ahead volatility forecasting for the weekly data						
GARCH	2.0074	19.7708	6.2245	954.5222	4.0743	2.3895
GJR-GARCH	**1.8039**	**17.6911**	**5.0161**	756.8590	**3.9737**	**2.1217**
EGARCH	1.8806	18.1301	5.2798	**734.0200**	4.0020	2.3216

	MAE1	*MAE2*	*MSE1*	*MSE2*	*QLIKE*	R^2LOG
MMGARCH	2.1362	20.6648	8.4340	1326.0124	7.2130	2.6306
MRS-GARCH	1.9047	18.8581	5.7116	867.1965	4.0198	2.2544
Panel B3: 3-week ahead volatility forecasting for the weekly data						
GARCH	1.9951	24.2441	5.9931	1281.2462	4.4457	1.2106
GJR-GARCH	**1.6987**	**20.7973**	**4.5098**	1027.2608	**4.3634**	**0.9511**
EGARCH	1.8598	22.0109	5.0810	**961.4178**	4.4067	1.1720
MMGARCH	2.3570	28.2766	9.8629	1436.2130	5.1367	1.5935
MRS-GARCH	1.8488	22.6504	5.2089	1126.5590	4.3972	1.0678
Panel B4: 4-week ahead volatility forecasting for the weekly data						
GARCH	1.9604	27.5131	5.7557	1547.2620	4.7161	0.7732
GJR-GARCH	**1.6040**	**23.1743**	**4.1310**	1301.1230	**4.6414**	**0.5248**
EGARCH	1.8375	25.0302	4.9367	**1149.5110**	4.6940	0.7540
MMGARCH	2.7006	37.4753	12.1779	3201.8060	5.2778	1.3018
MRS-GARCH	1.7864	25.4019	4.7471	1325.3540	4.6709	0.6326

Note: The numbers in the table refer to the values of loss functions. Numbers in bold indicate that the corresponding models have the lowest forecasting losses under a pre-specified criterion. Underlined numbers indicate that the corresponding models are included in the MCS. We perform 10,000 block bootstraps to generate p-values for the MCS test. The confidence level for MCS is 90%.

Table 10 reports the forecasting results for Brent crude oil return volatility. From the daily results, we can find that at 1-day ahead horizon, the best forecasting performance is displayed by GJR-GARCH, while at other time horizons, GARCH model presents the highest forecasting accuracy.

Specifically, the 1-day ahead forecasting results indicate that GJR-GARCH model performs the best, with the performance of being included under four of the six loss functions. MRS-GARCH model always results in higher forecasting losses than other models, indicating a lower forecasting accuracy. MMGARCH model is included in the MCS under three of the six loss functions. The 5-day and 10-day ahead forecasting results are consistent, which indicate that the forecasting performance of the regime-switching GARCH models is worse than that of the single-regime GARCH models. At the two forecasting horizons, both MMGARCH and MRS-GARCH models are always excluded from the MCS. The 22-day ahead forecasting results show that GARCH performs the best, for always being included in the MCS. EGARCH and MRS-GARCH are only included in the MCS under the *MSE2* criterion while the GJR-GARCH and MMGARCH are always excluded by the MCS.

Table 10. The forecasting results for Brent oil volatility

	MAE1	*MAE2*	*MSE1*	*MSE2*	*QLIKE*	R^2LOG
Panel A1: 1-day ahead volatility forecasting for the daily data						
GARCH	1.0737	3.9119	1.8146	59.0422	2.0974	9.3653
GJR-GARCH	**0.9167**	3.3889	1.6796	66.3652	2.4531	**7.4662**
EGARCH	1.0389	3.7518	**1.6348**	**51.8472**	**1.9527**	9.2920
MMGARCH	0.9271	**3.3886**	1.7301	67.9726	2.7303	7.6420
MRS-GARCH	1.1598	4.2648	2.0350	59.1989	2.1086	10.0741
Panel A2: 5-day ahead volatility forecast for the daily data						
GARCH	1.0899	9.3436	1.9099	240.9516	3.5972	0.7150
GJR-GARCH	1.5682	13.0415	5.4544	624.7684	4.7238	1.3672
EGARCH	**1.0568**	**8.9301**	**1.7466**	**213.1974**	**3.5815**	**0.7100**
MMGARCH	1.3039	11.4261	3.5645	491.5079	3.9198	0.8992
MRS-GARCH	1.4265	12.3961	3.1439	341.9370	3.7156	1.0550
Panel A3: 10-day ahead volatility forecasting for the daily data						
GARCH	**0.2485**	1.1199	1.8661	401.3439	**2.6124**	**0.0894**
GJR-GARCH	0.4941	2.2261	10.4918	1972.9809	2.8133	0.3413
EGARCH	0.2492	**1.1132**	**1.8041**	**367.6826**	2.6131	0.0922
MMGARCH	0.3671	1.7070	5.3462	1268.6135	2.6597	0.1675
MRS-GARCH	0.3421	1.5802	3.8151	701.4121	2.6395	0.1528
Panel A4: 22-day ahead volatility forecasting for the daily data						
GARCH	**1.3834**	**23.1902**	**2.7674**	**1070.6920**	**5.0890**	**0.2529**
GJR-GARCH	3.6200	58.1982	23.8596	5829.6451	6.6622	1.5459
EGARCH	1.4562	24.3566	2.9933	1136.4209	5.0956	0.2731
MMGARCH	2.7482	48.3562	11.2434	4620.5287	5.3170	0.7176
MRS-GARCH	1.7865	30.3310	4.8741	1610.9041	5.1423	0.3803
Panel B1: 1-week ahead volatility forecasting for the weekly data						
GARCH	1.9946	13.9418	6.3590	729.7866	3.3852	8.0101
GJR-GARCH	1.8835	13.1305	5.4776	606.4918	3.2250	7.6912
EGARCH	1.8974	13.0817	**5.4493**	**580.7444**	3.2355	7.9112
MMGARCH	**1.7783**	**12.3246**	7.0370	875.0829	112.9225	**7.5380**
MRS-GARCH	1.9605	13.7155	6.1672	715.8891	3.3785	7.8581
Panel B2: 2-week ahead volatility forecasting for the weekly data						
GARCH	2.0300	20.4830	6.7137	1230.8944	3.9959	2.6950
GJR-GARCH	**1.8406**	**18.6361**	**5.5318**	1016.3820	**3.8996**	**2.4117**
EGARCH	1.9188	18.9263	5.6782	**918.9046**	3.9299	2.6477
MMGARCH	2.1006	20.5660	9.1760	1672.2520	643.7425	2.9611
MRS-GARCH	1.9621	19.8105	6.2788	1180.6162	3.9755	2.5561
Panel B3: 3-week ahead volatility forecasting for the weekly data						
GARCH	1.9740	24.7368	6.2654	1697.5687	4.3691	1.2652
GJR-GARCH	**1.7213**	**21.9436**	**4.8999**	1439.9096	**4.2893**	**0.9976**
EGARCH	1.8663	22.6493	5.2281	**1220.9994**	4.3342	1.2404
MMGARCH	2.2329	27.4558	10.5581	1846.7122	5.3784	1.4602
MRS-GARCH	1.8465	23.2682	5.6634	1621.8913	4.3447	1.1342
Panel B4: 4-week ahead volatility forecasting for the weekly data						
GARCH	1.9309	27.5419	5.9119	1973.2481	4.6401	0.8870
GJR-GARCH	**1.6290**	**24.2758**	**4.5373**	1816.1623	**4.5665**	**0.6216**
EGARCH	1.8309	25.1101	4.9667	**1393.8593**	4.6205	0.8786
MMGARCH	2.4731	35.2811	12.8713	2139.6486	5.4253	1.2367
MRS-GARCH	1.7669	25.6409	5.1997	1902.5535	4.6129	0.7510

Note: The numbers in the table refer to the values of loss functions. Values in bold indicate that the corresponding models have the lowest forecasting losses under a pre-specified criterion. Underlined numbers indicate that the corresponding models are included in the MCS. We perform 10,000 block bootstraps to generate p-values for the MCS test. The confidence level for MCS is 90%.

From the weekly results for Brent in Table 10, we can find that EGARCH performs the best at 1-week and 2-week ahead horizons and GJR-GARCH beats other models at other time horizons. The forecasting results of the Regime switching models never display a significant superiority to all the single-regime models.

Specifically, the 1-week ahead results reveal that MMGARCH and MRS-GARCH are included in the MCS under three and one loss functions, respectively. However, EGARCH is always included in the MCS, and GJR-GARCH is excluded only under the *MSE2* criterion. From the 2-week ahead forecasting results, we can find that both the MMGARCH and MRS-GARCH models are included in the MCS under four of the six loss functions, while the GJR-GARCH and EGARCH are included in the MCS under five of the six loss functions. From the 3-week ahead forecasting results, we can see that the performance of MMGARCH displays the worst, with no inclusion in the MCS. Although MRS-GARCH is included in the MCS under five loss functions, GJR-GARCH is always included in the MCS, indicating a worse forecasting performance of MRS-GARCH. The main forecasting results of 4-week ahead horizon are consistent with those of 3-week ahead horizon.

In short, based on the analysis above, we can generally conclude that there is no additional increase in forecasting power when regime switching is considered. Our findings upon the performance of MMGARCH are different from Klein and Walther (2016). They compare the volatility forecasting performance of the MMGARCH model to other discrete volatility models (GARCH, RiskMetrics, EGARCH, APARCH, FIGARCH, HYGARCH, and FIAPARCH), and find that MMGARCH appears superior. The different findings may be sourced from the different data sample, different distribution and different evaluation criteria. First, their study is conducted during the period from 01/01/1995 to 12/31/2014, while we focus on 01/02/1986-06/30/2017. In order to determine whether different sample periods result in different findings, we also reevaluate the findings based on the same period of Klein and Walther (2016) and find that the findings are a bit different, but the main conclusions mentioned above still hold. Consequently, we conclude that the sample period is not the main reason for the different conclusions. Second, they adopt the normal distribution, while we adopt the *student's t* distribution because the data have fat tail feature, which can be better handled by the *student's t* distribution compared to the normal distribution. Finally, they evaluate the forecasting performance based on the SPA test (Hansen, 2005) while we employ the latest MCS procedure for evaluation (Hansen et al., 2011).

Besides, our findings of the performance of MRS-GARCH are inconsistent with Fong and See (2002). Their results indicate that the regime switching model performs noticeably better than non-switching models regardless of evaluation criteria. The selection of MRS variant, sample period, and evaluation criteria of this paper are different from those of Fong and See (2002)[7]. For instance, they adopt the MRS variant according to Gray (1996) and evaluate the forecasting performance based on the daily data from 01/02/1992 to 12/31/1997 by comparing the values of

[7] Actually, in order to further determine the influence of sample period, we also perform robustness test using the same period of Fong and See (2002) to determine the effect of sample period, and the findings are a bit different, but main conclusions are consistent with the former findings. Consequently, we conclude that sample period is not the main driver for the different findings.

MSE, MAE and R^2. However, in recent years, international crude oil markets have evolved a lot, and crude oil price volatility characteristics have experienced significant changes. Meanwhile, some latest methods and appropriate evaluation criteria are used in this paper. Consequently, it is acceptable for the new results in this paper and its inconsistence with previous relevant literature.

What is more, in order to exclude the influence of out-of-sample period selection, we also test the forecasting performance using different sub-samples[8]. In short, the robustness test results are in line with the main results obtained above. We do not identify the improvement when the regime switching is considered, and none of the GARCH-type models in this paper is found to be absolutely superior to others. This result reveals that a model which performs very well in a particular sample period, or under a given loss function, may not provide satisfactory results in other sample periods or under other loss functions. Thus, energy investors and policymakers should be very cautious when using a particular GARCH-type model to forecast crude oil price volatility; in particular, the complicated models do not always have better forecasting results.

In addition, the VaR forecasting results in **Appendix** show that the incorporation of the regime switching does not significantly improve the forecasting performance for crude oil price volatility; meanwhile, it may, but not necessarily, improve the performance of VaR forecasting. Specifically, the MMGARCH generates a more accurate VaR while the MRS-GARCH is beaten by the single-regime GARCH models.

[8] The detailed robustness test results can be obtained upon request.

4. Conclusions and future work

In this paper, we estimate and forecast the volatility for daily and weekly WTI and Brent crude oil prices using three single-regime GARCH models (i.e., the linear GARCH, the nonlinear GJR-GARCH and EGARCH models) and two regime-switching GARCH (i.e., MMGARCH and MRS-GARCH) models.

The empirical results show that the regime switching GARCH models can help us identify the structural changes in volatility. Meanwhile, in the in-sample estimation, the MRS-GARCH model displays superiority in weekly data compared to the single-regime GARCH and MMGARCH models. However, when confronted with the challenge of volatility forecasting in an out-of-sample experiment, the single-regime GARCH models are not beaten by the regime switching GARCH models. Besides, the central results are confirmed by the robustness test.

In brief, the consideration of regime switching fails to provide an evident improvement in forecasting accuracy for crude oil price volatility. This conclusion suggests that energy economists, energy policymakers and energy market practitioners should not simply choose volatility forecasting models with multiple regimes for crude oil prices. In most cases, the forecasting performance of single-regime models outperforms that of regime switching models. In other words, it is not appropriate to prefer the complicated models for crude oil price volatility forecasting. Meanwhile, the commonly-used GARCH-type models and the newly proposed evaluation criteria for forecasting performance in this paper provide important reference for them.

Besides, it should be noted that there is still some interesting work to be explored in the future. First, other MRS models that offer advantages by adjusting underlying GARCH processes can be taken into consideration, such as the models in Chang (2012), Alizadeh (2008), Fong and See (2002), and Nomikos and Pouliasis (2011). Second, other models with time-varying variance levels could be compared to regime switching GARCH models, such as the models in Baillie and Morana (2009), Engle and Rangel (2008), and Shi and Ho (2015). Finally, the flexible Fourier form GARCH (FFF-GARCH) model (Teterin et al., 2016) can be used to determine the presence of structural shifts so that we can further investigate whether the presence of structural shifts influences the forecasting performance of different kinds of models.

Appendix

As the good volatility forecasting does not guarantee good market risk measurement, evaluating the forecasting ability of the competing models in terms of Value-at-Risk (VaR) will help us to investigate whether the consideration of regime switching provides a better measure for market extreme risk. Therefore, we evaluate the forecasting performance not only for volatility but also for VaR, in order to investigate the effect of regime switching in forecasting the risk of crude oil market.

(1) The Value-at-Risk forecasting method

The estimation of VaR using GARCH, GJR-GARCH, EGARCH and MMGARCH for the out-of-sample series can be defined as Eq. (A.1).

$$VaR_{\tau,t}^{k} = \hat{\delta}_{t+k} + Q_{\tau}\sqrt{\hat{h}_{t+k}} \tag{A.1}$$

where Q_{τ} denotes the corresponding quantile of the distribution at the confidence level of $1-\tau$, $\hat{\delta}_{t+k}$ is the forecasted mean, and $\hat{h}_{t+k}$ is the forecasted aggregated volatility. The value of τ is set as 0.01.

In the MRS-GARCH framework (Billio and Pellizzon, 2000), the VaR at time t can be calculated as Eq. (A.2).

$$VaR_{\tau,t}^{k} = \sum_{i=1}^{2} \Pr[s_{t+k} = i \mid \zeta_{t-1}]\left[\hat{\delta}_{t+k} + \Phi(\tau)(\hat{h}_{t+k}^{i})\right]^{1/2} \tag{A.2}$$

where $\Phi(\bullet)$ is the cumulative distribution function, and $\hat{h}_{t+k}^{i}$ is the forecasted aggregated volatility in regime i.

Besides, to evaluate the forecasting performance of different models in terms of VaR, the loss function of Koenker and Bassett (1978) is employed as follows:

$$L[VaR_{t+1}(\tau,\hat{\delta}_{t})] = \{\tau - I[r_{t+1} < VaR_{t+1}(\tau,\hat{\delta}_{t})]\}[r_{t+1} - VaR_{t+1}(\tau,\hat{\delta}_{t})] \tag{A.3}$$

where $L[\bullet]$ is the loss function of $VaR_{t+1}(\tau,\hat{\delta}_{t})$; and $I[\bullet]$ is an indicator function.

(2) The VaR forecasting results

Tables A and B show the MCS test results of the competing GARCH models in terms of the VaR forecasting, for different forecasting horizons. The results indicate that the incorporation of regime-switching may, but not necessarily, improve the VaR forecasting accuracy. Specifically, the MMGARCH passes the MCS test for most forecasting horizons, especially for short-term horizons. However, the MRS-GARCH cannot survive in the MCS test for all horizons.

The VaR forecasting results of WTI, as shown in Table A, indicate that MMGARCH passes all the MCS tests for daily data at all forecasting horizons while MRS-GARCH never survives. For weekly data, MMGARCH is included in the MCS at 1-week and 2-week ahead horizons while MRS-GARCH is still never included in the MCS.

Table A. The loss function values for VaR forecasting results of WTI crude oil prices

Forecasting horizon	GARCH	GJR-GARCH	EGARCH	MMGARCH	MRS-GARCH
Panel A: Daily data					
1-day ahead	0.3161	0.2928	0.3241	**<u>0.2560</u>**	0.3768
5-day ahead	0.7074	0.5974	0.7261	**<u>0.5588</u>**	0.8288
10-day ahead	1.0048	<u>0.7869</u>	1.0336	**<u>0.7738</u>**	1.1719
22-day ahead	1.5085	**<u>1.0304</u>**	1.5461	<u>1.1072</u>	1.7502
Panel B: Weekly data					
1-week ahead	0.5165	0.5317	0.5357	**<u>0.4386</u>**	0.7524
2-week ahead	0.7310	0.7349	0.7613	**<u>0.6645</u>**	1.0769
3-week ahead	0.8999	**<u>0.8733</u>**	0.9338	0.9465	1.3284
4-week ahead	1.0343	**<u>0.9693</u>**	1.072	1.2736	1.5402

Note: The numbers in the table refer to the values of loss functions. Numbers in bold indicate that the corresponding models have the lowest forecasting losses under a pre-specified criterion. Underlined numbers indicate that the corresponding models are included in the MCS. We perform 10,000 block bootstraps to generate p-values for the MCS test. The confidence level for MCS is 90%, while the confidence level for VaR is 99%.

The VaR forecasting results of Brent in Table B indicates that, for daily data, MMGARCH is included in the MCS at 1-day, 5-day, and 10-day ahead horizons while MRS-GARCH never survives in the MCS test. For weekly data, MMGARCH passes the MCS test at all horizons while MRS-GARCH is always excluded by the MCS.

Table B. The loss function values for VaR forecasting results of Brent crude oil prices

Forecasting horizon	GARCH	GJR-GARCH	EGARCH	MMGARCH	MRS-GARCH
Panel A: The daily data					

1-day ahead	0.2779	0.2468	0.2835	**0.2143**	0.3214
5-day ahead	0.6323	0.5535	0.6427	**0.5146**	0.7237
10-day ahead	0.8902	0.7645	0.9116	**0.7431**	1.0279
22-day ahead	1.3455	**1.0275**	1.3736	1.2145	1.5465
Panel B: The weekly data					
1-week ahead	0.5022	0.5200	0.5209	**0.4577**	0.6697
2-week ahead	0.7027	0.7096	0.7346	**0.4418**	0.9286
3-week ahead	0.8634	0.8405	0.9013	**0.5595**	1.1127
4-week ahead	1.0001	0.9401	1.0427	**0.6672**	1.256

Note: The numbers in the table refer to the values of loss functions. Numbers in bold indicate that the corresponding models have the lowest forecasting losses under a pre-specified criterion. Underlined numbers indicate that the corresponding models are included in the MCS. We perform 10,000 block bootstraps to generate p-values for the MCS test. The confidence level for MCS is 90%, while the confidence level for VaR is 99%.

References

Agnolucci, P. (2009). Volatility in crude oil futures: a comparison of the predictive ability of GARCH and implied volatility models. *Energy Economics,* 31(2), 316-321.

Alizadeh, A.H., Nomikos, N.K., & Pouliasis, P.K. (2008). A Markov regime switching approach for hedging energy commodities. *Journal of Banking & Finance*, 32(9), 1970-1983.

Ang, A., & Timmermann, A. (2012). Regime changes and financial markets. *Annual Review of Financial Economics*, 4(1), 313-337.

Baillie, R.T., Bollerslev, T., & Mikkelsen, H.O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1), 3-30.

Baillie, R.T., & Morana, C. (2009). Modelling long memory and structural breaks in conditional variances: An adaptive FIGARCH approach. *Journal of Economic Dynamics and Control,* 33(8), 1577-1592.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics,* 31, 307-327.

Bollerslev, T., Engle, R.F., & Nelson, B.D. (1994). ARCH models. *Handbook of Econometrics*, 4, 2959-3038.

Bouri, E., Chen, Q., Lien, D., & Lv, X. (2017). Causality between oil prices and the stock market in China: The relevance of the reformed oil product pricing mechanism. *International Review of Economics & Finance*, 48, 34-48.

Cai, J. (1994). A Markov model of switching-regime ARCH. *Journal of Business & Economic Statistics*, 12(3), 309-316.

Chang, K.L. (2012). Volatility regimes, asymmetric basis effects and forecasting performance: An empirical investigation of the WTI crude oil futures market. *Energy Economics*, 34(1), 294-306.

Chkili, W., Hammoudeh, S., & Nguyen, D.K. (2014). Volatility forecasting and risk management for commodity markets in the presence of asymmetry and long memory. *Energy Economics*, 41, 1-18.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a), 427-431.

Engle, R.F., & Rangel, J.G. (2008). The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *The Review of Financial Studies*, 21(3), 1187-1222.

Fan, Y., Zhang, Y.J., Tsai, H.T., & Wei, Y.M. (2008). Estimating 'Value at Risk' of crude oil price and its spillover effect using the GED-GARCH approach. *Energy Economics,* 30, 3156-3171.

Fong, W.M., & See, K.H. (2002). A Markov switching model of the conditional volatility of crude oil futures prices. *Energy Economics*, 24(1), 71-95.

Franses, P.H., & Van Dijk, D., (1996). Forecasting stock market volatility using (nonlinear) GARCH models. *Journal of Forecasting*, 15, 229-235.

Glosten, L.R., Jagannathan, R., & Runkle, D.E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance,* 48, 1779-1801.

Gray, S.F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42(1), 27-62.

Haas, M., Mittnik, S., & Paolella, M.S. (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics,* 2, 493-530.

Hamilton, J.D. (1988). Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates. *Journal of Economic Dynamics and Control,* 12, 385-423.

Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica,* 57, 357-384.

Hamilton, J.D., & Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics,* 64, 307-333.

Hansen, B.E. (1992). The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP. *Journal of applied Econometrics,* 7, S61-S61.

Hansen, B.E. (1996). Erratum: the likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP. *Journal of Applied Econometrics,* 11, 195-198.

Hansen, P.R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, 23, 365-380.

Hansen, P.R., Lunde, A., & Nason, J.M. (2011). The model confidence set. *Econometrica*, 79(2), 453-497.

Hentschel, L. (1995). All in the family nesting symmetric and asymmetric garch models. *Journal of Financial Economics*, 39(1), 71-104.

Hillebrand, E. (2005). Neglecting parameter changes in GARCH models. *Journal of Econometrics,* 129(1), 121-138.

Hou, A., Suardi, S. (2012). A nonparametric GARCH model of crude oil price return volatility. *Energy Economics,* 34, 618-626.

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3), 255-259.

Koenker, R., & Bassett, G. (1978). Regression quantiles. Econometrica, 46, 33-50.

Kilian, L., & Vigfusson, R.J. (2011). Nonlinearities in the oil price–output relationship. *Macroeconomic Dynamics*, 15, 337-363.

Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economic*, 27(2), 363-394.

Klein, T., & Walther, T. (2016). Oil price volatility forecast with mixture memory GARCH. *Energy Economics*, 58, 46-58.

Lamoureux, C.G., & Lastrapes, W.D. (1990). Persistence in variance, structural change, and the GARCH model. *Journal of Business & Economic Statistics,* 8 (2), 225-234.

Li, M., Li, W. K., & Li, G. (2013). On mixture memory GARCH models. *Journal of Time Series Analysis*, 34(6), 606-624.

Ling, S., & McAleer, M. (2002a). Necessary and sufficient moment conditions for the GARCH (r, s) and asymmetric power GARCH (r, s) models. *Econometric Theory*, 18, 722-729.

Ling, S., McAleer, M. (2002b). Stationarity and the existence of moments of a family of GARCH processes. *Journal of Econometrics*, 106, 109-117.

Ljung, G. M., & Box, G. E. (1978). On a measure of lack of fit in time series models. Biometrika, 65(2), 297-303.

Lux, T., Segnon, M., & Gupta, R. (2016). Forecasting crude oil price volatility and value-at-risk: evidence from historical and recent data. *Energy Economics*, 56, 117-133.

Manera, M., Longo, C., Markandya, A., & Scarpa, E. (2007). Evaluating the empirical performance of alternative econometric models for oil price forecasting. Fondazione Eni Enrico Mattei Working Paper.

Marcucci, J. (2005). Forecasting stock market volatility with regime-switching GARCH models. *Stud. Studies in Nonlinear Dynamics & Econometrics,* 9, 1558-3708.

Marzo, M., & Zagaglia, P. (2010). Volatility forecasting for crude oil futures. *Applied Economics Letters*, 17, 1587-1599.

Mi, Z.F., Wei, Y.M., Tang, B.J., Cong, R.G., Yu, H., Cao, H., & Guan, D. (2016). Risk assessment of oil price from static and dynamic modelling approaches. Applied Economics, 49(9), 929-939.

Mohammadi, H., & Su, L. (2010). International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models. *Energy Economics*, 32, 1001-1008.

Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: A new approach. Econometrica, 59, 347-370.

Newey, W. K., & West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, 61(4), 631-653.

Nomikos, N.K., & Pouliasis, P.K. (2011). Forecasting petroleum futures markets volatility: The role of regimes and market conditions. *Energy Economics,* 33(2), 321-337.

Pagan, A.R., & Schwert, G.W. (1990). Alternative models for conditional stock volatility. *Journal of Econometrics,* 45, 267-290.

Phillips, P. C., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika,* 75(2), 335-346.

Sadorsky, P. (1999). Oil price shocks and stock market activity. *Energy Economics,* 21, 449-469.

Sadorsky, P. (2006). Modeling and forecasting petroleum futures volatility. Energy Economics, 28(4), 467-488.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2), 461-464.

Sévi, B. (2014). Forecasting the volatility of crude oil futures using intraday data. *European Journal of Operational Research*, 235(3), 643-659.

Tang, L., Dai, W., Yu, L., & Wang, S. (2015). A novel CEEMD-based EELM ensemble learning paradigm for crude oil price forecasting. *International Journal of Information Technology & Decision Making*, 14(1), 141-169.

Shi, Y., & Ho, K.Y. (2015). Modeling high-frequency volatility with three-state FIGARCH models. *Economic Modelling*, 51, 473-483.

Teterin, P., Brooks, R., & Enders, W. (2016). Smooth volatility shifts and spillovers in US crude oil and corn futures markets. *Journal of Empirical Finance*, 38, 22-36.

Timmermann, A. (2000). Moments of Markov switching models. *Journal of Econometrics,* 96, 75-111.

Wang, C., & Nishiyama, Y. (2015). Volatility forecast of stock indices by model averaging using high-frequency data. *International Review of Economics & Finance*, 40, 324-337.

Wang, Y., & Wu, C. (2012). Forecasting energy market volatility using GARCH models: Can multivariate models beat univariate models?. *Energy Economics,* 34, 2167-2181.

Wang, Y., Wu, C., & Yang, L. (2016). Forecasting crude oil market volatility: A Markov switching multifractal volatility approach. *International Journal of Forecasting*, 32(1), 1-9.

Wei, Y., Wang, Y., & Huang, D. (2010). Forecasting crude oil market volatility: further evidence using GARCH-class models. *Energy Economics*, 32, 1477-1484.

White, H. (2000). A reality check for data snooping. Econometrica, 68(5), 1097-1126.

Zhang, J.L., Zhang, Y.J., & Zhang, L. (2015). A novel hybrid method for crude oil price forecasting. *Energy Economics*, 49, 649-659.

Zhang, Y.J., & Wang, J. (2015). Exploring the WTI crude oil price bubble process using the Markov regime switching model. *Physica A: Statistical Mechanics and its Applications*, 421, 377-387.

Zhang, Y.J., & Wang, Z.Y. (2013). Investigating the price discovery and risk transfer functions in the crude oil and gasoline futures markets: Some empirical evidence. *Applied Energy*, 104, 220-228.

Zhang, Y.J., & Yao, T. (2016). Interpreting the movement of oil prices: driven by fundamentals or bubbles? *Economic Modelling,* 55, 226-240.

Zhang, Y.J., & Zhang, L. (2015). Interpreting the crude oil price movements: evidence from the Markov regime switching model. *Applied Energy*, 143, 96-109.

Highlights

- It compares three single-regime GARCH models and two regime switching GARCH models
- Different data frequencies and time horizons are considered for the forecasting
- The model confidence set procedure is used to evaluate the forecasting performance
- The MRS-GARCH model beats the single-regime GARCH models for in-sample estimation
- The regime switching GARCH models do not outperform in out-of-sample forecasting