International Journal of Forecasting 32 (2016) 1–9

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Forecasting crude oil market volatility: A Markov switching multifractal volatility approach

Yudong Wang[a,*], Chongfeng Wu[b], Li Yang[c]

[a] School of Economics and Management, Nanjing University of Science and Technology, China
[b] Antai College of Economics & Management, Shanghai Jiao Tong University, China
[c] School of Banking and Finance, University of New South Wales, Australia

ARTICLE INFO

Keywords:
Markov switching multifractal (MSM) models
Volatility
Crude oil markets
GARCH-class models
Model confidence set

ABSTRACT

We use a Markov switching multifractal (MSM) volatility model to forecast crude oil return volatility. Not only can the model capture stylized facts of multiscaling, long memory, and structural breaks in volatility, it is also more parsimonious in parameterization, after allowing for hundreds of regimes in the volatility. Our in-sample results suggest that MSM models fit oil return data better than the traditional GARCH-class models. The out-of-sample results show that MSM models generate more accurate volatility forecasts than either popular GARCH-class models or the historical volatility model.

1. Introduction

In recent years, large fluctuations in crude oil prices have caused grave concern among both market participants and regulators. One of the reasons for this concern is that the oil price uncertainty has a significant impact on the economy (Elder & Serletis, 2010). Theories of both investment under uncertainty and real options predict that an uncertainty about oil prices can depress current investment (Bernanke, 1983; Brennan & Schwartz, 1985; Henry, 1974; Majd & Pindyck, 1987). In addition, the volatility is a key input in pricing options and a major determinant of the value at risk (VaR). Therefore, the modeling and forecasting of the crude oil return volatility are of considerable interest among academics.

In the literature on the forecasting of volatility, the family of generalized autoregressive conditional heteroscedasticity (GARCH) models (Bollerslev, 1986) has been used widely for capturing the dynamics of oil return volatility (see for example Alizadeh, Nomikos, & Pouliasis, 2008; Giot & Laurent, 2003; Kang, Kang, & Yoon, 2009; Mohammadi & Su, 2010; Narayan & Narayan, 2007; Nomikos & Pouliasis, 2011; Sadorsky, 2006; Wang & Wu, 2012; Wei, Wang, & Huang, 2010). However, several shortcomings of GARCH-class models have been observed. First, most GARCH-class models can only capture the characteristic of short memory, rather than long-range dependence, even though long-range dependence in volatility has been documented commonly in the literature. The fractional integrated GARCH (FIGARCH) of Baillie, Bollerslev, and Mikkelsen (1996) and its extensions seem to capture the long memory in volatility well. However, the unanimous finding of hyperbolic decay of the autocorrelation function of absolute returns or squared returns is more likely to be a fiction due to unaccounted structural breaks, rather than the "genuine" one revealed by FIGARCH. Lamoureux and Lastrapes (1990) argue that the persistence implied by GARCH models becomes much weaker following the incorporation of structural breaks. Specifically, Lee, Hu, and Chiou (2010) show empirically that some sudden events (e.g., the Iraqi invasion of Kuwait and the Gulf Wars) result in an

* Correspondence to: XiaoLinwei Street 200, Xuanwu District, Nanjing, China. Tel.: +86 13681663442.
E-mail addresses: wangyudongnj@126.com (Y. Wang), cfwu@sjtu.edu.cn (C. Wu), l.yang@unsw.edu.au (L. Yang).

http://dx.doi.org/10.1016/j.ijforecast.2015.02.006
0169-2070/

increase in the permanent component of the conditional variance, which is evidence of structural breaks. The simple regime switching process can capture structural shifts in the volatility well, but can also lead to the spurious finding of fractional integration (Granger & Terasvirta, 1999) and exponential rather than hyperbolic decay of the autocorrelation function (Ryden, Terasvirta, & Asbrink, 1998). Baillie and Morana (2009) claim that the proposed adaptive FIGARCH (AFIGARCH) incorporates both long memory and structural breaks by allowing the intercept of FIGARCH to follow a slowly varying function specified by Gallant's (1984) flexible functional form. As was pointed out by Wang, Bauwens, and Hsiao (2013), this parametric model is less efficient if there are no structural breaks in the sample period. In addition, the AFIGARCH model has the problem of needing to determine the order of the trigonometric terms in the Gallant flexible functional form, in addition to the order of the specification of the stationary components in the conditional variance equation. For larger values of the order of the trigonometric terms, the AFIGARCH model has more parameters that need to be estimated, and hence is more likely to result in over-fitting, where a model includes irrelevant explanatory variables that may improve the in-sample fitting but cause a poorer out-of-sample performance.

Second, GARCH-class models cannot accommodate the property of multiscaling (or multifractality) (Lux & Kaizoji, 2007), which is a well-known stylized fact in economic data (Cont, 2001). The scaling property, which is a concept borrowed from statistical physics, defines the behaviors of some forms of volatility measures (e.g., the squared or absolute returns) as a function of the time interval on which the returns are computed.[1] The scaling behavior is characterized by the so-called Hurst exponent and its related index. If q-order moments of the distributions of price increments display different scaling behaviors for different values of q, a multiscaling behavior is revealed. The investigation of scaling behaviors in economic and financial data has expanded considerably since the work of Mandelbrot (1963) (see for example Mandelbrot, 1997, 2001; Mantegna & Stanley, 1995; Muller et al., 1990; Stanley & Plerou, 2001). Multiscaling in crude oil markets, which is what we are interested in, is also found in a few studies (Alvarez-Ramirez, Alvarez, & Rodriguez, 2008; Wang & Liu, 2010; Wang & Wu, 2013). Traditional GARCH-class models are always related to the dynamics of squared returns rather than to another order of moments, and therefore, they do not take into account multiscaling behavior in price movements. The recent empirical study by Wang, Wei, and Wu (2011) also shows the lack of ability of GARCH-class models to capture multiscaling volatility in crude oil markets.

In this paper, we use the Markov switching multifractal (MSM) model of Calvet and Fisher (2001) to forecast the crude oil market volatility. This model is motivated by the stylized fact of multiscaling behavior or multifractality in financial data. The MSM model assumes a hierarchical and multiplicative structure of heterogeneous volatility components, which differs fundamentally from conventional volatility models (such as GARCH-class ones) (Lux & Kaizoji, 2007). The advantage of a MSM model over the conventional regime-switching model is that, while the number of parameters grows quadratically as the number of states increases in a regime-switching model, the MSM model is more parsimonious in parameterization, even after allowing for hundreds of states in order to capture possible structural changes. The MSM model is known to generate outliers and long memory in the volatility and to decompose the volatility into components with heterogeneous decay rates (Calvet & Fisher, 2004). Therefore, it can address the aforementioned problems of traditional volatility models well.

We apply the MSM model to West Texas Intermediate (WTI) and Brent crude oil return data. We compare its in-sample and out-of-sample performances with those of several traditional models, including the popular GARCH-class models and the historical volatility (HV) model. Our in-sample results based on Vuong's (1989) closeness test suggest that MSM models fit the data significantly better than GARCH-class models. For the comparison of out-of-sample performances, we use six loss functions to evaluate the forecast accuracy. An advanced econometric test named the model confidence set (MCS; see Hansen, Lunde, & Nason, 2011) is employed to examine further whether the differences in forecasting losses among different models are statistically significant. We find that MSM models produce more accurate forecasts than either GARCH-class models or the HV model for most of the loss functions employed. The GARCH and HV models are always excluded from MCS at the 90% confidence level, while the MSM models are included in MCS under most loss criteria. Based on the empirical evidence, we conclude that the MSM models outperform the GARCH-class models for forecasting the crude oil market volatility.

The remainder of this paper is organized as follows. Section 2 provides a general description of MSM models for forecasting the volatility. Section 3 describes the data and provides some preliminary analysis. Section 4 reports the empirical results, and Section 5 concludes.

2. Forecasting models

2.1. Markov switching multifractal (MSM) volatility model

We forecast the crude oil return volatility using the MSM volatility model introduced by Calvet and Fisher (2001). The MSM volatility model assumes that the underlying return follows a discrete-time Markov process with multifrequency stochastic volatility.[2]

We denote by ε_t the innovations of crude oil returns, r_t, which can be expressed as $r_t = \mu_t + \varepsilon_t$, where μ_t is the conditional mean. MSM models the innovations ε_t in the

[1] There is also another type of scaling behavior that is studied in the economics literature: the behavior of the tails of the distribution of returns as a function of the size of the price changes, but the interval on which the returns are measured is constant. This type of scaling behavior is measured by a tail index of the distribution.

[2] We use the ML estimator of the MSM model. For a detailed presentation of the ML estimator, see Calvet and Fisher (2004).

following framework of stochastic volatility models (Calvet & Fisher, 2001, 2004):

$$\varepsilon_t = \sigma_t z_t, \qquad \sigma_t = \sigma \left(\prod_{i=1}^{\bar{k}} M_{i,t} \right)^{1/2}, \tag{1}$$

where σ is a positive constant and the random variables $\{z_t\}$ are i.i.d. with standard normal distribution (i.e., $z_t \sim$N(0,1)). For $k \in \{1, 2, \ldots, \bar{k}\}$, the random multipliers or volatility components $M_{k,t}$ are persistent and satisfy the condition $M_{k,t} \geq 0$ and $E(M_{k,t}) = 1$. For the sake of simplicity, we follow Calvet and Fisher (2004) by assuming that the $\bar{k}$ multipliers are independent at any given time. Given the vector of volatility components $M_t = (M_{1,t}, M_{2,t}, \ldots, M_{\bar{k},t})$, for any $m = (m_1, m_2, \ldots, m_{\bar{k}}) \in \mathbf{R}^{\bar{k}}$ we define the function $g(m) = \prod_{i=1}^{\bar{k}} m_i$, and the stochastic volatility process can be written as $\sigma_t = \sigma[g(M_t)]^{1/2}$.

It can be seen that the volatility dynamics are driven by the stochastic process of the vector M_t. Following the suggestion of Calvet and Fisher (2004), we assume that M_t is first-order Markov, for parsimony. The state vector M_t is unobservable, and should be inferred recursively by Bayesian updating. For $k \in \{1, 2, \ldots, \bar{k}\}$, the multiplier $M_{k,t}$ is drawn from a fixed distribution M with probability γ_k, and is equal to its value in the previous period $M_{k,t-1}$ with probability $1 - \gamma_k$. The transition probabilities $\gamma \equiv (\gamma_1, \gamma_2, \ldots, \gamma_{\bar{k}})$ are specified as

$$\gamma_k = 1 - (1 - \gamma_1)^{(b^k - 1)}, \tag{2}$$

where $\gamma_1 \in (0, 1)$ and $b > 1$.[3] Since we have $\gamma_1 < \cdots < \gamma_{\bar{k}} < 1 < b$, $(\gamma_{\bar{k}}, b)$ is used to specify the set of transition probabilities. This feature of transition probabilities is consistent with the definition of multifractality (Calvet, Fisher, & Mandelbrot, 1997; Fisher, Calvet, & Mandelbrot, 1997; Mandelbrot, Fisher, & Calvet, 1997), and is responsible for multifractality in financial data. The only restriction in this multifractal process is: $M > 0$ and $E(M) = 1$. Calvet and Fisher (2004) suggest that M can be taken simply as a binomial random variable with a value of either m_0 or $2 - m_0$, with probability 0.5. In summary, the MSM volatility model has only four parameters $(m_0, \sigma, b, \gamma_{\bar{k}})$ that require estimation, even if the number of states $(2^{\bar{k}})$ is very large.

2.2. GARCH-class models

We compare the forecasting performance of the MSM volatility model with those of several popular GARCH-class models. Five GARCH-class models with different specifications are considered. The first is the GARCH model of Bollerslev (1986), which is the most popular volatility model other than the ARCH model introduced by Engle (1982). The GARCH (1,1) can be written as follows:

$$r_t = \mu_t + \varepsilon_t = \mu_t + h_t^{1/2}\eta_t, \quad \eta_t \sim iid(0, 1),$$
$$h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}, \tag{3}$$

where μ_t denotes the conditional mean and h_t is the conditional variance with the sufficient conditions $\omega > 0, \alpha \geq 0, \beta \geq 0$ to ensure $h_t > 0$.

Another model employed in this paper is the integrated GARCH (IGARCH) proposed by Engle and Bollerslev (1986). This model assumes an infinite persistence in the conditional volatility. The specification of IGARCH(1,1) is the same as that of GARCH(1,1), but with the parameter restriction $\alpha + \beta = 1$.

In order to take possible asymmetries into account, Glosten, Jagannathan, and Runkle (1993) propose a GJR model. The specification for the conditional variance of GJR(1,1) is given by

$$h_t = \omega + [\alpha + \gamma I(\varepsilon_{t-1} < 0)]\,\varepsilon_{t-1}^2 + \beta h_{t-1}, \tag{4}$$

where $I(\cdot)$ is an indicator function, i.e., $I(\cdot)$ is 1 when the condition $(\cdot)$ is met, and 0 otherwise. The sufficient condition for confirming $h_t > 0$ is $\omega > 0, \alpha \geq 0, \beta \geq 0$ and $\alpha + \gamma \geq 0$. The asymmetric effect is captured by the coefficient γ, with $\gamma \geq 0$.

Another popular model that is capable of dealing with asymmetric effects is the exponential GARCH (EGARCH) proposed by Nelson (1991). The EGARCH is given by

$$\log(h_t) = \omega + \alpha\left(|\eta_{t-1}| - E\,|\eta_{t-1}|\right) + \gamma\eta_{t-1} + \beta\log(h_{t-1}). \tag{5}$$

As was claimed by Nelson (1991), there are no parameter restrictions with EGARCH.

The abovementioned GARCH-class models assume that the volatility autocorrelation decays at an exponential rate. Baillie et al. (1996) propose a fractionally integrated ARCH model (FIGARCH) that allows the volatility autocorrelation to decay at a hyperbolic rate. Interestingly, the FIGARCH$(1, d, 1)$ nests the GARCH(1,1) with $d = 0$. The FIGARCH$(1, d, 1)$ model can be written as

$$h_t = \omega + \beta h_{t-1} + [1 - (1 - \beta L)^{-1}(1 - \phi L)(1 - L)^d]\varepsilon_t^2, \tag{6}$$

where $0 \leq d \leq 1, \omega > 0, \phi, \beta < 1$, d is the fractional integration parameter, and L is the lag operator. The parameter d characterizes the long memory property in the volatility. The advantage of the FIGARCH process is that for $0 < d < 1$, it is sufficiently flexible to allow for intermediate ranges of persistence, between the complete integrated persistence of the volatility shocks that is associated with $d = 1$ and the geometric decay that is associated with $d = 0$.

In summary, in addition to MSM volatility models, we also use five GARCH-class models to describe and forecast WTI and Brent crude oil return volatilities, namely the standard GARCH(1,1), IGARCH, GJR, EGARCH and FIGARCH.

3. Data and preliminary analysis

We use daily spot price data for West Texas Intermediate (WTI) and Brent crude oil. The sample covers the period from January 4, 1993, to September 9, 2013, resulting in 5141 observations. Our data are obtained from the U.S. Energy Information Administration (EIA).

[3] This specification is connected with the discretization of a Poisson arrival process. For a detailed introduction to this specification, see Calvet and Fisher (2001).

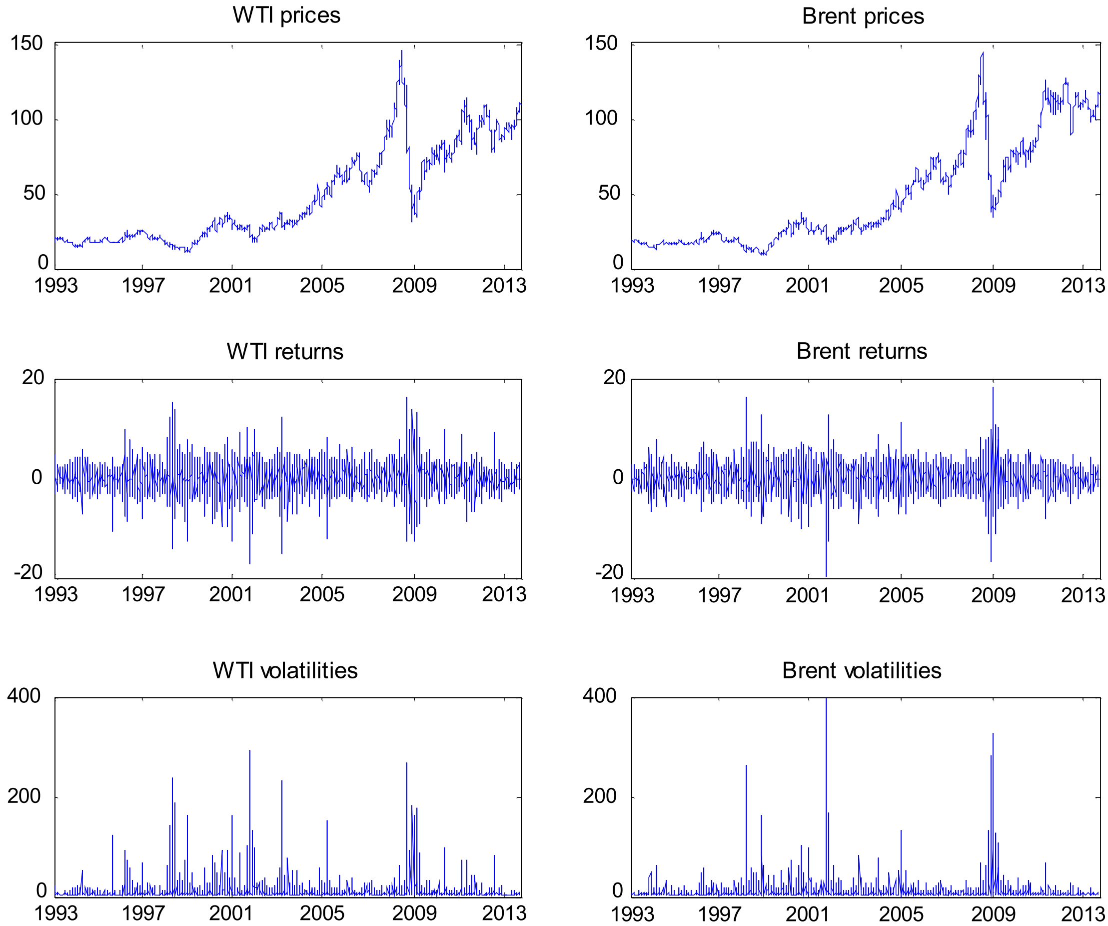

Fig. 1. Prices, returns and volatilities of crude oil.

Taking P_t to be the oil price on day t, we use the percentage daily price return, r_t, which can be written as $r_t = 100 \times (\log(P_t) - \log(P_{t-1}))$. The squared daily returns are taken as a proxy for the actual volatilities.

Fig. 1 plots the prices, returns and volatilities of crude oil. It is evident that some aggregate demand and supply shocks can lead to large fluctuations in crude oil markets. For example, as a typical case of an oil demand shock, the Asian financial crisis in 1998 resulted in large crashes in crude oil prices. The largest oil price change in our sample period occurred in September 2001, shortly after the "9/11 terrorist attack". The second Gulf War in 2003 led to decreases in oil production, thus causing sharp increases in oil prices in a short period of time. Driven by the economic boom in emerging economies, the oil price experienced persistent increases from 2003 to 2008. Due to the global economic recession triggered by the subprime crisis in the U.S., the oil price dropped quickly in the second half of 2008. From this general review of the history of oil prices, we can see that oil prices demonstrate a high uncertainty over time. Considering the essential role played by crude oil in the world economy, the problem of managing the oil price risk is of great interest for economists, market participants and policy makers.

Table 1 reports the descriptive statistics of crude oil returns. We can see that the WTI and Brent returns display similar statistical properties. This can be explained by the high degree of integration in world crude oil markets (see, e.g. Adelman, 1984; Fattouh, 2010). Specifically, the mean values of crude oil returns are close to zero, while the standard deviations are tens of times larger. The Jarque–Bera statistics reject the null hypothesis of a Gaussian distribution at the 1% significance level. This stylized fact of a fat-tailed distribution is also evidenced by the negative skewness and positive excess kurtosis. The Ljung and Box (1978) Q statistics for serial correlation show that the null hypothesis of no autocorrelation up to the 10th or 20th orders is rejected at the 1% significance level for both returns and volatilities, implying the existence of autocorrelations. The F-statistics of the ARCH test consistently indicate the existence of ARCH effects in oil returns. The presence of strong ARCH effects may be the reason why GARCH-class models are used widely in financial time series analysis and are appreciated in risk management. Table 1 also gives the results of unit root tests. The augmented Dickey and Fuller (1979) and Phillips and Perron (1988) statistics both show rejections of the null hypothesis of a unit root in oil returns. The test statistic of Kwiatkowski, Phillips, Schmidt, and Shin (1992) cannot reject the null hypothesis of stationarity. Overall, the evidence from unit root tests reveals the crude oil return process to be stationary.

4. Empirical results

4.1. In-sample performance

We describe and forecast oil return volatility dynamics using a Markov switching multifractal (MSM) volatility model. Table 2 reports the estimation results of MSM($\bar{k}$) for WTI and Brent oil price returns, where the hierarchical level $\bar{k}$ varies from 1 to 10. We find that MSM(7) (128 regimes) and MSM(10) (1024 regimes) fit WTI oil returns almost equally well, and better than the other MSM models, as is evidenced by the higher log-likelihoods of these two models. Meanwhile, MSM(6) (64 regimes) fits Brent returns best.

We compare the in-sample performances of MSM with those of several popular GARCH-class models.[4] Table 3 shows the results based on Vuong's (1989) test. We choose MSM(10) and MSM(6) as benchmarks for modeling the WTI and Brent oil return volatilities, respectively, because they fit the data better than the other MSM models. The test statistics indicate that the selected MSM models fit the oil return data better than the GARCH-class models at the 1% significance level.

4.2. Out-of-sample performance

Market participants are generally more interested in models' out-of-sample performances than their in-sample performances, because they are more concerned about how well they can do using these volatility models in the future. Therefore, we now compare the forecasting abilities of the MSM and GARCH models.

Our forecasting procedure divides the sample period into two subsamples: the in-sample period for parameter estimation, covering the first 1000 trading days, and the out-of-sample period for forecast evaluation, covering the last 4141 days. We compute 1- and 20-day-ahead forecasts. For the sake of computational convenience, we re-estimate the model parameter every 20 days using the rolling window method, where the window length is 1000 observations. Thus, the parameters are fixed within the 20-day window, and only the data are updated. This fixed parameter and rolling window forecast scheme follows the work of Laurent, Rombouts, and Violante (2012) to satisfy the assumptions of the MCS test (Hansen et al., 2011) for comparing the forecasting abilities of nested models.

As has been pointed out by researchers (see, e.g. Bollerslev, Engle, & Nelson, 1994; Diebold & Lopez, 1996; Lopez, 2001), it is difficult to determine which is the best loss function for evaluating forecasting performances. Therefore, we employ the following six popular loss function criteria, rather than making a single choice:

$$\mathrm{MAE}_1 = \frac{1}{T-N}\sum_{t=N+1}^{T}\left|\hat{h}_t - h_t\right|, \tag{7}$$

[4] To save space, the parameter estimates of these GARCH-class models are not given here, but are available upon request.

Table 1
Descriptive statistics of crude oil price returns.

	WTI	Brent
Mean	0.031	0.036
Std. dev.	2.414	2.251
Maximum	16.41	18.13
Minimum	−17.09	−19.89
Skewness	−0.181	−0.103
Excess kurtosis	4.892	5.126
Jarque–Bera	5140.5***	5636.3***
Q(10)	31.84***	23.73***
Q(20)	43.17***	45.54***
$Q^2(10)$	1104.8***	521.5***
$Q^2(20)$	1927.2***	963.5***
ARCH(10)	55.15***	31.70***
ARCH(20)	35.23***	21.99***
ADF	−72.37***	−70.15***
PP	−72.88***	−70.14***
KPSS	0.048	0.050

Note: Std. Dev. is the daily standard deviation. The Jarque and Bera (1980) statistic tests for the null hypothesis of a Gaussian distribution. ADF, PP and KPSS denote statistics from the augmented Dickey and Fuller (1979), Phillips and Perron (1988) and Kwiatkowski et al. (1992) unit root tests, respectively. The optimal lag length of the ADF test is chosen based on the Schwarz information criterion (SIC) (Schwarz, 1978), and the optimal bandwidths of the PP unit root test and the KPSS stationarity test are determined based on the Newey–West criterion (Newey & West, 1994). The null hypothesis of the ADF and PP tests is a unit root, and that of the KPSS test is stationarity. $Q(l)$ and $Q^2(l)$ are the Ljung and Box (1978) statistics of the return and squared return series for up to lth order serial correlation, respectively.

* Denotes rejection of the null hypothesis at the 10% significance level.

** Denotes rejection of the null hypothesis at the 5% significance level.

*** Denotes rejection of the null hypothesis at the 1% significance level.

$$\mathrm{MAE}_2 = \frac{1}{T-N}\sum_{t=N+1}^{T}\left|\sqrt{\hat{h}_t} - \sqrt{h_t}\right|, \tag{8}$$

$$\mathrm{MSE}_1 = \frac{1}{T-N}\sum_{t=N+1}^{T}\left(\hat{h}_t - h_t\right)^2, \tag{9}$$

$$\mathrm{MSE}_2 = \frac{1}{T-N}\sum_{t=N+1}^{T}\left(\sqrt{\hat{h}_t} - \sqrt{h_t}\right)^2, \tag{10}$$

$$\mathrm{QLIKE} = \frac{1}{T-N}\sum_{t=N+1}^{T}\left(\ln(\hat{h}_t) + h_t/\hat{h}_t\right), \quad \text{and} \tag{11}$$

$$R^2\mathrm{LOG} = \frac{1}{T-N}\sum_{t=N+1}^{T}\left(\ln(h_t/\hat{h}_t)\right)^2, \tag{12}$$

where h_t and $\hat{h}_t$ are the actual volatility and volatility forecasts, respectively. T is the length of the full sample ($T = 5140$), and N is the length of the in-sample dataset ($N = 1000$). MAE and MSE are the mean absolute error and the mean squared error, respectively. QLIKE corresponds to the loss implied by a Gaussian likelihood, and R^2 LOG is similar to the R^2 of the Mincer–Zarnowitz regressions.

However, one major limitation of these loss functions is that they cannot tell us whether the differences in forecasting accuracy among different models are statistically significant. To address this issue, we use the model confidence set (MCS) method, a test that was proposed by Hansen et al. (2011). The idea behind this test is that the data available may be not informative enough to

Table 2
Estimation results of MSM models.

	$\bar{k}=1$	$\bar{k}=2$	$\bar{k}=3$	$\bar{k}=4$	$\bar{k}=5$	$\bar{k}=6$	$\bar{k}=7$	$\bar{k}=8$	$\bar{k}=9$	$\bar{k}=10$
WTI oil market										
$\hat{b}$	1.500	9.453*	7.329***	11.06***	8.697***	5.256***	3.980***	20.69***	3.844***	3.235***
	(1.500)	(1.881)	(3.021)	(3.398)	(3.551)	(4.785)	(6.111)	(439.7)	(5.890)	(6.438)
$\hat{m}_0$	1.682***	1.576***	1.478***	1.434***	1.389***	1.344***	1.314***	1.443***	1.311***	1.288***
	(101.6)	(85.58)	(94.23)	(90.95)	(103.9)	(96.53)	(103.1)	(145.7)	(93.23)	(95.01)
$\hat{\gamma}_{\bar{k}}$	0.039***	0.056***	0.096***	0.266***	0.758***	0.752***	0.733***	0.432***	0.716***	0.826***
	(6.488)	(4.369)	(3.306)	(3.143)	(3.777)	(3.359)	(4.286)	(83.20)	(3.692)	(4.500)
$\hat{\sigma}$	3.192***	2.705***	2.906***	3.207***	3.010***	2.861***	2.722***	2.484***	2.087***	2.074***
	(32.86)	(39.61)	(28.87)	(29.72)	(32.05)	(20.56)	(25.56)	(33.43)	(20.67)	(24.86)
Log (L)	−11316	−11228	−11188	−11185	−11179	−11178	−11175	−11188	−11176	−11175
Brent oil market										
$\hat{b}$	1.500	7.617*	4.178***	7.941***	6.133**	8.121***	6.108**	5.980*	15.56***	20.87***
	(1.500)	(1.932)	(2.326)	(2.916)	(2.197)	(3.705)	(2.042)	(1.843)	(4.498)	(23.25)
$\hat{m}_0$	1.569***	1.520***	1.421***	1.376***	1.358***	1.333***	1.358***	1.358***	1.356***	1.389***
	(109.3)	(87.22)	(100.6)	(83.42)	(68.20)	(95.91)	(64.04)	(68.69)	(95.50)	(128.8)
$\hat{\gamma}_{\bar{k}}$	0.020***	0.035***	0.029***	0.156**	0.145**	0.791***	0.144**	0.142**	0.969***	0.430***
	(4.137)	(3.882)	(2.881)	(2.447)	(2.287)	(4.430)	(2.196)	(1.996)	(4.132)	(4.660)
$\hat{\sigma}$	2.627***	2.534***	2.681***	2.599***	2.215***	2.170***	2.371***	2.035***	2.999***	2.234***
	(45.30)	(42.28)	(37.55)	(28.47)	(37.88)	(35.18)	(33.15)	(36.48)	(28.10)	(45.67)
Log (L)	−11102	−10998	−10971	−10964	−10963	−10962	−10964	−10964	−10965	−10968

Notes: The numbers in parentheses are t-statistics.
* Denotes rejections at the 10% significance level.
** Denotes rejections at the 5% significance level.
*** Denotes rejections at the 1% significance level.

Table 3
Results of Vuong's (1989) closeness test.

	WTI		Brent	
	Likelihoods	Statistics	Likelihoods	Statistics
GARCH	−11293	−4.164***	−11022	−3.146***
IGARCH	−11299	−4.241***	−11024	−2.981***
GJR	−11292	−4.048***	−11012	−2.628***
EGARCH	−11300	−3.818***	−11018	−2.790***
FIGARCH	−11283	−4.092***	−11019	−3.460***
MSM(1)	−11316	−8.068***	−11102	−7.572***
MSM(2)	−11228	−5.220***	−10998	−3.360***
MSM(3)	−11188	−1.888*	−10971	−1.185
MSM(4)	−11185	−1.705*	−10964	−0.596
MSM(5)	−11179	−1.146	−10963	−0.424
MSM(6)	−11178	−1.119	−10962	0
MSM(7)	−11175	−0.290	−10964	−0.544
MSM(8)	−11188	−2.200**	−10964	−0.609
MSM(9)	−11176	−1.729*	−10965	−0.872
MSM(10)	−11175	0	−10968	−1.433

* Denotes rejections at the 10% significance level.
** Denotes rejections at the 5% significance level.
*** Denotes rejections at the 1% significance level.

yield a single model that can dominate all of its competitors significantly. In such cases, one can only obtain a smaller set of models, called the model confidence set, which contains the best forecasting model at a given level of confidence. Therefore, the models in the MCS perform equally well at the given confidence level. To save space, we do not provide a detailed description of the MCS test, but refer interested readers to Hansen et al. (2011).[5]

Table 4 shows the 1-day-ahead forecasting results for the WTI return volatility. We also include the historical volatility (HV) model in our comparison. After all, all efforts at volatility forecasting would be futile if none of the models could beat the historical volatility. We find that MSM models have lower forecasting losses than the GARCH-class models or the HA model under five of the six loss functions, suggesting an improved forecasting accuracy. Although the standard GARCH(1,1) has lower forecasting losses than the MSM models under the QLIKE criterion, the MCS results indicate that their forecasting accuracies are not significantly different. Moreover, we can see that, under four of the six losses, MCS only contains MSM(10). This indicates that the MSM(10) forecasts are significantly more accurate than their competitors, including the GARCH-class models, HA and the other MSM models.

[5] To save space, we do not provide a detailed description of the MCS procedure, but refer interested readers to Hansen et al. (2011). Some recent studies also include applications of MCS (e.g. Laurent et al., 2012).

Table 4
One-day-ahead forecasting results for WTI oil volatility.

	MSE_1	MAE_1	MSE_2	MAE_2	QLIKE	R^2LOG
MSM(2)	246.50	6.655	3.065	1.344	2.672	6.401
MSM(4)	245.33	6.596	3.034	1.332	2.666	6.353
MSM(6)	245.28	6.614	3.035	1.332	2.665	6.328
MSM(8)	**244.50**	6.626	3.043	1.336	2.667	6.361
MSM(10)	245.52	**6.572**	**3.017**	**1.326**	2.667	**6.301**
GARCH	245.84	6.914	3.194	1.384	**2.663**	6.603
EGARCH	253.28	6.831	3.191	1.375	2.694	6.559
GJR	251.22	6.936	3.225	1.384	2.669	6.581
IGARCH	249.93	7.106	3.319	1.402	2.669	6.626
FIGARCH	246.77	6.933	3.208	1.384	2.666	6.582
Historical volatility	275.43	7.385	3.693	1.521	2.872	7.355

Notes: The numbers in this table are 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 MCS. We perform 10,000 block bootstraps to generate *p*-values for the MCS test. The confidence level for MCS is 90%.

Table 5
One-day-ahead forecasting results for Brent oil volatility.

	MSE_1	MAE_1	MSE_2	MAE_2	QLIKE	R^2LOG
MSM(2)	205.84	5.774	2.667	1.266	2.568	6.872
MSM(4)	205.37	5.739	2.647	1.260	2.565	6.838
MSM(6)	**204.63**	**5.724**	**2.628**	**1.254**	2.560	**6.779**
MSM(8)	205.73	5.760	2.655	1.261	2.565	6.823
MSM(10)	206.04	5.727	2.639	1.257	2.568	6.805
GARCH	205.58	5.947	2.736	1.285	2.554	6.914
EGARCH	209.40	6.004	2.799	1.308	2.597	7.048
GJR	205.17	5.933	2.710	1.282	**2.544**	6.894
IGARCH	208.63	6.123	2.838	1.309	2.558	6.982
FIGARCH	206.19	5.930	2.731	1.281	2.562	6.871
Historical volatility	218.59	6.226	3.049	1.389	2.733	7.646

Notes: The numbers in this table are 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 MCS. We perform 10,000 block bootstraps to generate *p*-values for the MCS test. The confidence level for MCS is 90%.

Table 6
Twenty-day-ahead forecasting results for WTI oil volatility.

	MSE_1	MAE_1	MSE_2	MAE_2	QLIKE	R^2LOG
MSM(2)	284.7	6.921	3.183	1.323	4.442	6.124
MSM(4)	269.5	6.758	3.079	1.319	4.435	6.202
MSM(6)	275.8	6.842	3.119	1.315	4.437	**6.121**
MSM(8)	269.9	**6.748**	**3.064**	**1.308**	4.430	6.123
MSM(10)	276.5	6.878	3.142	1.327	4.445	6.206
GARCH	268.1	7.095	3.266	1.371	**4.419**	6.422
EGARCH	275.2	7.123	3.236	1.345	4.425	6.154
GJR	**267.4**	7.138	3.322	1.381	4.493	6.412
IGARCH	280.9	7.659	3.629	1.445	4.466	6.618
FIGARCH	330.1	8.634	4.337	1.603	4.714	7.130
Historical volatility	305.6	7.824	3.850	1.523	4.601	7.076

Notes: The numbers in this table are 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 MCS. We perform 10,000 block bootstraps to generate *p*-values for the MCS test. The confidence level for MCS is 90%.

Table 5 reports the 1-day-ahead results for the Brent oil return volatility. We still find that the MSM forecasts display lower losses than the GARCH-class ones under most of the loss functions. MSM(6) generates more accurate forecasts than the other MSM models. MSM models are included in MCS under five of the six loss criteria. HA and the GARCH models are all excluded from MCS under four loss criteria, suggesting that their forecasting accuracies are significantly inferior to those of the MSM models. Tables 6 and 7 report the 20-day-ahead forecasting results for the WTI and Brent oil return volatilities, respectively. Here, we find that MSM models result in lower forecasting losses than their competitors under most of the criteria. MSM models are always included in MCS, while the GARCH and HA models are excluded under most criteria. Overall, the 20-day-ahead forecasting results are consistent with the 1-day-ahead results. Based on this analysis, we can generally conclude that the forecasting abilities of MSM models are superior to those of traditional GARCH-class models and the historical volatility model.

Table 7
Twenty-day-ahead forecasting results for Brent oil volatility.

	MSE_1	MAE_1	MSE_2	MAE_2	QLIKE	R^2LOG
MSM(2)	190.5	**6.574**	**3.030**	**1.369**	4.443	6.947
MSM(4)	188.4	6.634	3.047	1.376	4.419	6.972
MSM(6)	190.3	6.580	3.037	1.371	4.434	**6.942**
MSM(8)	189.7	6.679	3.083	1.388	4.435	7.029
MSM(10)	190.9	6.691	3.101	1.389	4.427	7.053
GARCH	189.6	7.050	3.275	1.452	4.418	7.326
EGARCH	192.0	7.076	3.294	1.450	4.441	7.247
GJR	**186.0**	6.991	3.246	1.445	4.432	7.327
IGARCH	189.2	6.821	3.172	1.398	**4.417**	7.132
FIGARCH	205.6	7.764	3.738	1.537	4.447	7.578
Historical volatility	199.8	7.280	3.496	1.522	4.531	7.724

Notes: The numbers in this table are 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 MCS. We perform 10,000 block bootstraps to generate p-values for the MCS test. The confidence level for MCS is 90%.

5. Conclusions

Multifractality (or multiscaling) is a well-known stylized fact in financial data. However, traditional volatility models such as GARCH-type models do not consider this stylized fact. In this paper, we use a newly developed multifractal Markov switching (MSM) volatility model to capture and forecast the dynamics of the crude oil return volatility. Based on Vuong's (1989) closeness test, we find that the log-likelihoods of MSM models are significantly greater than those of the GARCH-class ones, implying that MSM models fit the oil returns data better.

We compute 1- and 20-day-ahead forecasts of MSM, GARCH-class and historical volatility models. Six loss functions are employed to quantify the forecasting loss. We use an advanced test of the model confidence set (MCS; see Hansen et al., 2011) to examine whether the differences between the loss functions of the different models are statistically significant. Our results indicate that MSM models have lower forecast losses under most criteria. The GARCH-class and HV models are always excluded from MCS at the 90% confidence level. That is, MSM models have greater forecasting abilities than the GARCH or HV models.

We would like to conclude this paper by outlining some points which deserve further investigation. First, one could investigate the performances of MSM models for forecasting the value at risk. The estimation of the value at risk depends on the reasonable assumption of innovations, and is therefore beyond the scope of this paper. Second, as the volatility is a key input in option pricing formulas, it would be interesting to compare the performances of MSM models and other popular models for oil option pricing. Third, the univariate MSM models could be extended to multivariate ones which could model the covariances between two different asset returns, in order to investigate hedging and asset allocations. For some meaningful extensions of univariate to multivariate MSM models, see Calvet, Fisher, and Thompson (2006) and Liu and Lux (2013). Finally, we restrict the competition among MSM models to models with up to 10 hierarchical levels. Lux (2008) has provided evidence that adding additional hierarchical layers could improve the forecasting capacity of MSM further. Lux also proposed a GMM estimator for MSM models where the computational costs for any level k are negligible compared to the time-consuming ML approach. Again, the use of MSM models with still higher values of k might be worthwhile, as the model confidence set would provide evidence on the added explanatory power of such specifications. Overall, at least in this paper, we have found MSM models to provide a more powerful tool in modeling and forecasting the crude oil volatility than traditional volatility models.

Acknowledgments

We would like to express our sincere gratitude to the handling editor (Dick van Dijk) and an anonymous reviewer whose comments and suggestions greatly improved the quality of this paper. This work is supported by the National Science Foundation of China (Nos. 71401077 and 71320107002).

References

Adelman, M. A. (1984). International oil agreements. *The Energy Journal*, *5*, 1–9.
Alizadeh, A. H., Nomikos, N. K., & Pouliasis, P. K. (2008). A Markov regime switching approach for hedging energy commodities. *Journal of Banking and Finance*, *32*, 1970–1983.
Alvarez-Ramirez, J., Alvarez, J., & Rodriguez, E. (2008). Short-term predictability of crude oil markets: a detrended fluctuation analysis approach. *Energy Economics*, *30*, 2645–2656.
Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *73*, 3–20.
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*, 1577–1592.
Bernanke, B. S. (1983). Irreversibility, uncertainty, and cyclical investment. *Quarterly Journal of Economics*, *98*, 85–106.
Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*, 307–327.
Bollerslev, T., Engle, R. F., & Nelson, D. (1994). ARCH models. In R. F. Engle & D. L. McFadden (Eds.), *Handbook of econometrics, vol. IV* (pp. 2961–3038). Amsterdam: Elsevier Science B.V.
Brennan, M., & Schwartz, E. (1985). Evaluating natural resource investment. *Journal of Business*, *58*, 1135–1157.
Calvet, L. E., & Fisher, A. J. (2001). Forecasting multifractal volatility. *Journal of Econometrics*, *105*, 27–58.
Calvet, L. E., & Fisher, A. J. (2004). How to forecast long-run volatility: regime switching and the estimation of multifractal process. *Journal of Financial Econometrics*, *2*, 49–83.
Calvet, L. E., Fisher, A. J., & Mandelbrot, B. B. (1997). *Large deviations and the distribution of price changes*. Cowles Foundation Discussion Paper No. 1165.
Calvet, L. E., Fisher, A. J., & Thompson, S. B. (2006). Volatility comovement: a multifrequency approach. *Journal of Econometrics*, *131*, 179–215.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, *1*, 223–236.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Society*, *75*, 427–431.

Diebold, F. X., & Lopez, J. A. (1996). Forecast evaluation and combination. In G. S. Maddala & C. R. Rao (Eds.), *Handbook of statistics: statistical methods in finance, vol. 14* (pp. 241–268). Amsterdam: North-Holland.

Elder, J., & Serletis, A. (2010). Oil price uncertainty. *Journal of Money, Credit and Banking*, *42*, 1137–1159.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*, 987–1007.

Engle, R. F., & Bollerslev, T. (1986). Modeling the persistence of conditional variances. *Econometric Reviews*, *5*, 1–50.

Fattouh, B. (2010). The dynamics of crude oil price differentials. *Energy Economics*, *32*, 334–342.

Fisher, A. J., Calvet, L. E., & Mandelbrot, B. B. (1997). *Multifractality of Deutschemark/US Dollar exchange rates*. Cowles Foundation Discussion Paper No. 1166.

Gallant, A. R. (1984). The Fourier flexible form. *American Journal of Agricultural Economics*, *66*, 204–208.

Giot, P., & Laurent, S. (2003). Market risk in commodity markets: a VaR approach. *Energy Economics*, *25*, 435–457.

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. *Journal of Finance*, *48*, 1779–1801.

Granger, C., & Terasvirta, T. (1999). A simple nonlinear time series model with misleading linear properties. *Economics Letters*, *62*, 161–165.

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

Henry, C. (1974). Investment decisions under uncertainty: The irreversibility effect. *American Economic Review*, *64*, 1006–1012.

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

Kang, S. H., Kang, S. M., & Yoon, S. M. (2009). Forecasting volatility of crude oil markets. *Energy Economics*, *31*, 119–125.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationary against the alternative of a unit root. *Journal of Econometrics*, *54*, 159–178.

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

Laurent, S., Rombouts, J. V. K., & Violante, F. (2012). On the forecasting accuracy of multivariate GARCH models. *Journal of Applied Econometrics*, *27*, 934–955.

Lee, Y.-H., Hu, H.-N., & Chiou, J.-S. (2010). Jump dynamics with structural breaks for crude oil prices. *Energy Economics*, *32*, 343–350.

Liu, R., & Lux, T. (2013). *Non-homogeneous volatility correlations in the bivariate multifractal model*. Working paper, Deakin University.

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

Lopez, J. A. (2001). Evaluation of predictive accuracy of volatility models. *Journal of Forecasting*, *20*, 87–109.

Lux, T. (2008). The Markov-switching multifractal model of asset returns: GMM estimation and linear forecasting of volatility. *Journal of Business and Economic Statistics*, *26*, 194–210.

Lux, T., & Kaizoji, T. (2007). Forecasting volatility and volume in the Tokyo stock market: Long memory, fractality and regime switching. *Journal of Economic Dynamics and Control*, *31*, 1808–1843.

Majd, S., & Pindyck, R. S. (1987). Time to build, option value, and investment decisions. *Journal of Financial Economics*, *18*, 7–27.

Mandelbrot, B. B. (1963). The variation of certain speculative prices. *Journal of Business*, *36*, 394–419.

Mandelbrot, B. B. (1997). *Fractals and scaling in finance*. New York: Springer Verlag.

Mandelbrot, B. B. (2001). Scaling in financial prices: IV. Multi-fractal concentration. *Quantitative Finance*, *1*, 641–649.

Mandelbrot, B. B., Fisher, A. J., & Calvet, L. E. (1997). *A multifractal model of asset returns*. Cowles Foundation Discussion Paper No. 1164.

Mantegna, R. N., & Stanley, H. E. (1995). Scaling behavior in the dynamics of an economic index. *Nature*, *376*, 46–49.

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

Muller, U. A., Dacorogna, M. M., Olsen, R. B., Pictet, O. V., Schwarz, M., & Morgenegg, C. (1990). Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis. *Journal of Banking and Finance*, *14*, 1189–1208.

Narayan, P. K., & Narayan, S. (2007). Modelling oil price volatility. *Energy Policy*, *35*, 6549–6553.

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. *Review of Economic Studies*, *61*, 631–653.

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

Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regressions. *Biometrica*, *75*, 335–346.

Ryden, T., Terasvirta, T., & Asbrink, S. (1998). Stylized facts of daily return time series and the hidden Markov model. *Journal of Applied Econometrics*, *13*, 217–244.

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

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.

Stanley, E. H., & Plerou, V. (2001). Scaling and universality in economics: Empirical results and theoretical interpretation. *Quantitative Finance*, *1*, 563–567.

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, *57*, 307–333.

Wang, C. S.-H., Bauwens, L., & Hsiao, C. (2013). Forecasting a long memory process subject to structural breaks. *Journal of Econometrics*, *177*, 171–184.

Wang, Y., & Liu, L. (2010). Is WTI crude oil market becoming weakly efficient over time? New evidence from multiscale analysis based on detrended fluctuation analysis. *Energy Economics*, *32*, 987–992.

Wang, Y., Wei, Y., & Wu, C. (2011). Can GARCH-class models capture long memory in WTI crude oil markets. *Economic Modelling*, *28*, 921–927.

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. (2013). Efficiency of crude oil futures markets: New evidence from multifractal detrending moving average analysis. *Computational Economics*, *42*, 393–414.

Wei, Y., Wang, Y., & Huang, D. (2010). Forecasting crude oil market volatility: Further evidence using GARCH-class models. *Energy Economics*, *32*, 1447–1484.