Entropy **2015**, *17*, 7167-7184; doi:10.3390/e17107167

OPEN ACCESS
entropy
ISSN 1099-4300
www.mdpi.com/journal/entropy

Article

Wavelet Entropy Based Analysis and Forecasting of Crude Oil Price Dynamics

Yingchao Zou [1,2], Lean Yu [1] and Kaijian He [1,*]

[1] School of Economics and Management, Beijing University of Chemical Technology, Beijing 100029, China; E-Mails: evangeline1203@outlook.com (Y.Z.); yulean@amss.ac.cn (L.Y.)
[2] College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China

* Author to whom correspondence should be addressed; E-Mail: kaijian.he@my.cityu.edu.hk; Tel.: +86-10-6443-8793.

Academic Editor: Raúl Alcaraz Martínez

Received: 24 May 2015 / Accepted: 15 October 2015 / Published: 22 October 2015

Abstract: For the modeling of complex and nonlinear crude oil price dynamics and movement, wavelet analysis can decompose the time series and produce multiple economically meaningful decomposition structures based on different assumptions of wavelet families and decomposition scale. However, the determination of the optimal model specification will critically affect the forecasting accuracy. In this paper, we propose a new wavelet entropy based approach to identify the optimal model specification and construct the effective wavelet entropy based forecasting models. The wavelet entropy algorithm is introduced to determine the optimal wavelet families and decomposition scale, that will produce the improved forecasting performance. Empirical studies conducted in the crude oil markets show that the proposed algorithm outperforms the benchmark model, in terms of conventional performance evaluation criteria for the model forecasting accuracy.

Keywords: wavelet entropy; wavelet analysis; crude oil forecasting; Autoregressive Moving Average (ARMA) model

---

1. Introduction

In recent years, rapid development of electronic technology and the increasing level of global economic integration have fundamentally changed the crude oil markets, both in terms of market structure and market risk exposure. Higher levels of price fluctuations are witnessed in crude oil markets, accompanied by more competitive and risky environments, increasingly dominated by nonlinear multi scale dynamics. Besides, due to the unique characteristics of crude oil markets, such as high storage costs, *etc.*, they exhibit unique features that deserve research attention during the modeling process.

With the volatile crude oil price movement observed in the market, the modeling and forecasting of daily crude oil price movement remains one of the most important and difficult research issues in the energy research field. It attracts significant research interests as its resolution is fundamentally important to some important theoretical issues such as the crude oil derivatives and the energy risk management.

Over the years, numerous approaches have been developed to incorporate nonlinearity, auto correlation and heteroscedasticity data features into the modeling process, aiming at improving the forecasting accuracy further. These models include structural and econometric models, artificial intelligence models, and ensemble models. Equilibrium models analyze the economic relationships among participants in crude oil market and derive analytic equations to model them. For example, Bekiros *et al.* [1] used the time-varying Vector Autoregressive (VAR) model to model the impact of the economic policy uncertainty on the oil price movement and found the improved forecasting performance compared to other more standard univariate models [1]. Deng and Sakurai [2] used the multiple kernel learning regression method to forecast the crude oil spot price. They found that information from different time frame is useful in improving the forecasting accuracy of the model [2]. Chen [3] found the oil sensitive stock index to be significant predictors for the oil price movement [3]. Cuaresma *et al.* [4] derived a simple unobserved component model incorporating asymmetric cycles and found it to be superior in performance than the symmetric counterparts and benchmark Auregressive (AR) models [4]. Interestingly Alquist and Kilian [5] showed contradictory results: they found Random Walk (RW) model to be by far the best models available [5]. Meanwhile artificial intelligence models such as the traditional neural network and the more recent support vector regression have achieved significant progress. Empirical work utilizing the power of these models is on the rise, but with mixed results. For example, Godarzi *et al.* [6] found that the proposed dynamic Artificial Neural Network model achieves the improved forecasting accuracy than the time series and static neural network model [6]. Yu *et al.* [7] found that Artificial Neural Network (ANN) outperforms Autoregressive Moving Average (ARMA) model, but has room for further improvement using ensemble algorithms. Shin *et al.* [8] proposed a semi-supervised learning method to predict the directional movement of oil price. They have found the improved accuracy with the proposed method [8]. Work by Bildirici and Ersin [9] show that the multilayer perception type neural network contributes significantly the performance improvement in the proposed model [9]. Ensemble algorithm aims at combining individual forecasters to produce forecasts which are based on more complete information [10]. Since the seminal work by Bates and Granger [10], ensemble forecasts from different models to further reduce forecasting errors have attracted much research attention [10]. For example, Yu *et al.* [7] proposed the adaptive neural network to ensemble individual forecasts using neural network to model components extracted by Empirical Mode Decomposition (EMD) and observed significant performance improvement [7].

Recently, data driven computational approaches have emerges to take advantage of the nonlinear data characteristics during the modelling process. Typical nonlinear data features include chaotic, fractal and the multi scale data characteristics revealed by accumulating empirical evidence in the financial literature [11]. For example, Alvarez-Ramirez *et al.* [12] revealed that the autocorrelation of the crude oil price is sensitive to the price asymmetry and different time scales, which are one form of embodiment of price nonlinearity [12]. Barkoulas *et al.* [13] used the correlation dimensions test and recurrent plot to test the data generating mechanisms of the crude oil price. They found that the crude oil price contains non nomral nonlinar dynamics, which the current ARMA and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models provide insufficient modeling capability [13]. It becomes increasingly clear that the incorporation of the nonlinear and complex dynamics during the modeling process provide the promising alternative to deeper understanding of the market dynamics and higher level of forecasting accuracy, especially when the modeling and forecasting exercises are conducted at the daily frequency level, higher than three weekly and monthly frequency in the literature. Rodriguez *et al.* [14] showed that the level of market efficiency varies across different time scales.

Wavelet analysis as a technique to extract and decompose the multiscale data structure emerged as an important and promising approach to analyze the nonlinear and complex data characteristics in the multiscale domain. There has been more and more empirical evidence of the existence of multi scale data feature in the crude oil price movement and its co-movement with other macroeconomic variables. For example, Shahbaz *et al.* [15] found the relationship between the crude oil price and real exchange rate to be anti-cyclical while work by Jammazi *et al.* [16] suggested that this relationship exists in a asymmetric manner over different time horizons [15,16]. Tiwari *et al.* [17] found some interesting economic relationship between share prices and crude oil price when it is viewed in the multiscale domain [17]. In the meantime, we have found some positive results in wavelet based forecasting exercises for crude oil price movement. For example, He *et al.* [18] proposed a wavelet decomposed ensemble model, which introduces wavelet analysis to analyze the time varying dynamic underlying Data Generating Process, representative of heterogeneous market microstructure at finer time scale domain. Results from empirical studies show the superior performance of the proposed algorithm against the benchmark models [18]. Jammazi and Aloui [19] combined the A trous wavelet analysis and neural network in forecasting the crude oil price, and found the improved forecasting performance [19]. de Souza e Silva *et al.* [20] used the wavelet analysis to remove the high frequency data components for the modeling and forecasting by Hidden Markov Model. Experiment results show positive performance improvement [20]. Yousefi *et al.* [21] used wavelet analysis to decompose crude oil price and extended them directly to make forecasts [21]. However, much more positive forecasting results are accumulating in other areas, most notably in the electricity field. For example, Zhang *et al.* [22] used wavelet analysis to extract different data components of interests, where nonlinear component is modeled by neutral network and volatility component is modeled by GARCH model [22]. Kriechbaumer *et al.* [23] used an improved combined wavelet-autoregressive integrated moving average (ARIMA) to forecast monthly price of aluminum, copper, lead and zinc, and found the improved forecasting accuracy [23].

Gallegati *et al.* [24] used wavelet analysis to analyze the information content of some interest rate spread for future output growth [24].

However, one research issue left intact in the literature is the impact of additional parameters introduced by wavelet analysis on modeling and forecasting accuracy. The wavelet based forecasting algorithm introduces the additional parameters including wavelet families and decomposition scales. Previous researches relied on the arbitrarily selected wavelet families to analyze the historical information, leaving their validity under question. When the wavelet analysis is introduced in the economic and financial field, the ultimate aim is to achieve better understanding of the economic and financial relationship among variables, with the improved forecasting accuracy in the end. The determination of these parameters will critically affect the modeling accuracy and the derived policy implications. The wavelet analysis decomposes the data into the multiscale structure which is assumed to represent the true underlying multiscale data structure. Wavelet analysis represents a redundant representation problem, *i.e.*, there are different wavelet models that can replicate the same market price movement. The accuracy of these representations is limited by the constant governed by the uncertainty principle underlying the multi scale analysis. Therefore, there is no analytic solution to the identifications of the exact representations problem, which can be formulated as the optimization problem. This is a less addressed and important literature gap in the application of wavelet analysis in the forecasting field.

As an important information quantification measure, entropy serves as the potential tool to guide the optimization process. The entropy theory has been used to analyze the information content of the wavelet decomposed multiscale data structure in the other engineering literature. Wavelet entropy, relative wavelet entropy, and many other variants have been proposed in the literature to calculate the entropy of the energy distribution in the typical wavelet decomposition, as well as the cost function for the best basis algorithm to choose the optimal basis for wavelet packet transform [25,26]. For example, Pascoal and Monteiro [27] and Kim *et al.* [28] used the entropy measure and the wavelet analysis to analyze the degree of market efficiency and the dynamic correlations in the market respectively [27,28]. Xu *et al.* [29] used the modified wavelet entropy measure to differentiate between the normal and hypertension states [29]. Samui and Samantaray [30] incorporated the wavelet entropy measure in constructing the measuring index for islanding detection in distributed generation [30]. Wang *et al.* [31] used best basis based wavelet packet entropy to extract feature in the decomposed structure for the follow-up classification algorithm, which performs well in EEG analysis for patient classification [31]. Recently we have identified some recent research endeavors in the financial literature to model the multiscal data structure using the multiscale entropy theory. For example, in the energy economics literature, Martina *et al.* [32] introduced the entropy concept to analyze the efficiency of the crude oil markets [32]. Ortiz-Cruz *et al.* [33] introduced multiscale entropy theory to analyze the multiscale data structure in the crude oil market [33]. In the stock market, Niu and Wang [34] introduced a modified multiscale entorpy algorithm and showed its effectiveness in reducing the estimation error in Chinese stock markets [34]. Yin and Shang [35] introduced the weighted multiscale permutation entropy method to quantify the amplitude information of both US and Chinese stock markets, to analyze their difference and similarity [35]. However, the research attempts are limited to this extent so far. In the literature very few research has been identified to explore and tackle various research issues in the modeling and forecasting of the multiscale crude oil data structure using the multiscale entropy framework. Different research issues such as the determination of the appropriate model specifications in the multiscale anlaysis affect the accuracy and generalizability of the multiscale analysis and forecasting models.

The wavelet denoising and multiscale sample projection serves as the valuable tools to reveal and model the hidden multiscale data structure in the multiscale domain. These constituent data structures correspond to various main influencing factors for the crude oil price movement, such as basic supply and demand for crude oil, the macroeconomic factors, and major events in the market.

In this paper, we are motivated by the fact that the market has a heterogeneous underlying structure, where investors have different investment concerns and strategies. During the modeling process some of the components are more important as the main driving forces while other components have less significant impacts and can be classified as the noises. The separation and modeling of these constituent data structure are critical to more accurate modeling and forecasting of the crude oil price movement. Thus, in this paper, we assume the crude oil price is dominated by one component at one scale. We introduce the wavelet entropy theory to identify it and use it as the main driving factors to forecast the future movement of the crude oil price.

In this paper, we propose the wavelet entropy theory to identify the multiscale model structure and construct the effective forecasting algorithm. The wavelet entropy theory as well as entropy measure are introduced to measure the information energy distribution using the historical data and construct a two stage model selection procedure. Empirical studies in the benchmark crude oil markets confirm the statistically significant performance improvement from using the more appropriate multi scale model specification with the proposed wavelet entropy method.

The main contribution of this paper is the introduction of wavelet entropy based two stage model selection procedure to identify the appropriate model specification. This approach is built on the information theoretic approach other than the traditional MSE minimization. At the macro level, we use the wavelet entropy theory to measure the information energy distribution of the entire wavelet coefficients based on different wavelet families. At the micro level, the entropy is used to measure the information distribution of wavelet coefficients of wavelet decomposed data for different wavelet families at different scales. To the best of our knowledge, the work in this paper represents the first attempt to introduce wavelet entropy for the model specification identification for the construction of effective wavelet based forecasting algorithm.

The rest of this paper is organized as follows. Section 2 proposes the wavelet entropy based approach for estimating VaR. We conducted empirical studies in the benchmark crude oil markets and reported the results in Section 3. Finally, some concluding remarks are drawn in Section 4.

2. Methodology

2.1. Entropy and Wavelet Entropy Theory

To measure quantitatively the randomness of data, the entropy can be defined statistically for a stochastic time series system. Given random variables $X \in R^n$ generated with unknown parameters, the entropy is defined as in Equation (1) [36].

$$S = \int_0^{\infty} -(p)log(p)dx \tag{1}$$

where $p(.)$ refers to the probability density function (PDF). The value of entropy lies between 0 and 1. The higher the entropy is, the higher the level of disorder and uncertainty are.

If the data contain mixture of data features, the entropy may be biased in estimating the uncertainty and disorder levels in the data, potentially underestimating the randomness in data. The Wavelet Entropy offers an important alternative. It calculated the entropy value of the probability density function of the energy distribution of the wavelet coefficients in the wavelet transformed domain, as in (2) [37].

$$WE = \sum_{j=1}^{M} P_j ln P_j P_j = \frac{E_j}{\sum_{j=1}^{M} E_j} = \frac{\sum_{t=1} n(f_j(t))^2}{\sum_{j=1}^{M_i}(\sum_{t=1}^{n}(f_j(t))^2)}, M_i = 1, \ldots\ldots\ldots M \tag{2}$$

The smaller the wavelet entropy value is, the more organized the data are. The higher the wavelet entropy value is, the more disordered and uncertain the data are. However, different from the case of the entropy value, the wavelet entropy value size reflects not the total probability density of the data, but the average level of probability density of the data across different scales. For example, the significant cyclical information at smaller set of scales may be disputed by the noise information. The calculated entropy value may biased towards lower value, ignoring the frequency and cyclical data pretend at particular scales. The wavelet entropy value would take into account the structural distribution of randomness across scales and recognize the preserved of data at different scales.

2.2. Wavelet Entropy based Multiscale Forecasting Methodology

To model the crude oil price movement, we make some simplifying assumptions as follows:

(1) Data Generating Processes can be classified into several main groups with unique features and particular patterns, *etc*.
(2) Different Data Generating Processes are mutually independent across different scales.
(3) Different Data Generating Processes (DGPs) follow the same stochastic processes with different parameters.

With these assumptions, the data structure can be approximated with the combination and mixture of data generating processes at different scales. However, since there are different models of the underlying DGPs, for the observed market price movement, this represents an identification problem of the redundant representation during the modeling process. One approach is to resort to the traditional forecasting error as the criteria to identify the appropriate model specifications, which assumes that the small error corresponds to the maximum level of information or patterns extracted from the historical data. This was traditionally done using the error minimization, as evidenced in the recent researches.

Since the determination of the exact decomposition structure has the bounding limits governed by the uncertainty principle in the multi scale analysis, we would not expect the traditional approaches, such as the Minimization of MSE, to identify the optimal data structure among redundant representations.

In practice, the optimal data structure may be the combinations of decomposition structure with different wavelet families at different scales.

Thus in this paper we resort to the wavelet entropy and the entropy theory in this paper. The entropy value, as the measurement of the disorder in data, is introduced to analyze the historical data at two levels, *i.e.*, both microscope and macroscope. At the macro scale level, the wavelet entropy is used to measure the randomness of the data, taking into account the distribution of randomness across different scales. The measurement calculated with wavelet entropy would more accurately reflect the contribution of some orderly DGPs at some scales revealed in the particular wavelet families. When the wavelet entropy is calculated with different wavelet families at the same maximum scale, the wavelet family with the lowest wavelet entropy valued is retained as it implies the data with the most orderly organization. At the microscale level, the entropy of the individual coefficients at different scales is calculated and used to quantify the information content at different scales and compare their randomness directly. At each scale, the wavelet family with the lowest entropy value is retained as it is assume to contain the most orderly information and is the most suitable for ARMA modeling.

The numerical procedure for the wavelet entropy based forecasting algorithm is laid out as follows.

Firstly we use the wavelet algorithm to decompose the in-sample training data into different sub-data series at different scales up to the maximum scales J, using different wavelet families.

$$r_t = r_{A^J,t} + \sum_{j=1}^{J} r_{D^j,t} \tag{3}$$

Wavelet analysis possesses the ability to project data into time-scale domain and to conduct multiscale analysis [38]. This capability stems from the high energy concentration over a short interval of time in wavelets functions used, which is in direct contrast to the globally time invariant sinusoid functions used in more traditional spectrum analysis tools such as Fourier analysis [39]. Mathematically, wavelets are continuous functions that satisfy admissibility conditions as in Equation (4).

$$C_\psi = \int_0^\infty \frac{|\Psi(f)|}{f} df < \infty \leftrightarrow \int_{-\infty}^{\infty} \psi(t)dt = 0 \tag{4}$$

And unit energy condition as in Equation (5):

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1 \tag{5}$$

where Ψ is the Fourier transform of ψ. Together these two conditions guarantee that the wavelets functions have zero vanishing moments and improved localization in time scale domain during the analysis.

There are different families of wavelets designed, each with their own special characteristics [39]. The Haar wavelet is the simplest symmetric discontinuous wavelet that has characteristics of orthogonality and compact support. It is defined mathematically as in Equation (6):

$$\psi(t) = \begin{cases} +1 \; if \;\; 0 \leq t \leq 0.5 \\ -1 \; if \;\; 0.5 \leq t \leq 1 \\ 0 \quad otherwise \end{cases} \tag{6}$$

Daubechies wavelets are generalizable beyond the Haar wavelets. They are continuous orthogonal wavelets with compact support. Symlet wavelets are continuous orthogonal wavelets with compact support and are designed to be nearly symmetric. Coiflets are also designed to be nearly symmetric wavelets.

The wavelets can be translated over time and dilated by scales as in Equation (7):

$$\psi_{u,s} = \frac{1}{\sqrt{s}}\psi(\frac{t-u}{s}) \tag{7}$$

where $s \in R$ and $u \in R$. Unlike sinusoids used in Fourier transform, wavelets are characterized by two parameters: location u and scale s. Thus, wavelets of different shapes and lengths are formed by adjusting these two parameters.

The original signal can be projected into time scale domain by means of convolving the translated or dilated wavelets to the original signal [39]. Thus, the wavelet transform is a function of these two variables as in Equation (8):

$$W(u,s) = \int_{-\infty}^{\infty} x(t)\psi_{u,s}(t)dt \tag{8}$$

The inverse operation could also be performed as in Equation (9):

$$x(t) = \frac{1}{C_\psi}\int_{0}^{\infty}\int_{-\infty}^{\infty} W(u,s)\psi_{u,s}(t)du\frac{ds}{s^2} \tag{9}$$

Since the original signal can be decomposed by wavelet analysis and reconstructed perfectly by wavelet synthesis, together they form the basis for multi-resolution analysis as in Equation (10):

$$f(t) \approx S_J(t) + D_J(t) + D_{J-1}(t) + ... + D_1(t) \tag{10}$$

where $S_J(t)$ refers to smooth signals and equals $\sum_k s_{j,k}\phi_{j,k}(t)$. $D_i(t)(i = 1...J)$ refers to detail signals and equals $\sum_k d_{j,k}\psi_{j,k}(t)$. The multi-resolution analysis decomposes the complicated data structure into the underlying influencing factors by applying wavelet analysis.

Secondly, we calculate the entropy of the coefficients at different scales with different families. We group them at different scales. For each scale, we select one wavelet family with the lowest entropy value. In the end, we determine wavelet family WF_n for n scale. The entropy is calculated as follows. For a stochastic time series system, given random variables $X \in R^n$ generated with unknown parameters, the shannon entropy is defined as in Equation (11) [36].

$$S = \int_{0}^{\infty} -(p)log(p)dx \tag{11}$$

where $p(.)$ refers to the PDF. The value of entropy lies between 0 and 1. The higher the entropy value is, the higher the level of disorder and uncertainty is.

Thirdly we calculate the wavelet entropy for all the chosen wavelet families WF_n at the maximum scales J. Among them, we choose the one particular wavelet family WF_{WE} at the scale j_{WE}, that has the minimum wavelet entropy value. The decomposed data component r_{WE} at the scale j_{WE} using the wavelet family WF_{WE} are supposed to represent the main underlying DGP for the original crude oil data.

Fourthly the decomposed data component r_{WE} is assumed to follow ARMA processes. We estimate the conditional means using the ARMA models.

$$\widehat{\mu}_{t,s} = a_{0,s} + \sum_{i=1}^{m} a_{i,s} r_{t-i,s} + \sum_{j=1}^{n} b_{j,s} \varepsilon_{t-j,s} \tag{12}$$

Fifthly as the main underlying DGP of the crude oil dynamics, there is very strong correlation between the estimated DGP and the original crude oil data estimated. The variance of the decomposed data component r_{WE} is assumed to contribute to the total variance off the original data r. Thus we use the linear regression model to determine the intercept and coefficients with the model tuning data set.

Sixthly with the calculated parameters, we repeat step 1 to step 5 to calculate the forecasts.

3. Empirical Studies

Data Collection

Empirical studies are conducted using the observations of the daily closing prices in both US West Taxes Intermediate (WTI) crude oil market and UK Brent crude oil markets. The data set covers the period from 2 January 2002 to 3 August 2015. Following the machine learning literature, we design the experiment to evaluate the model performance using out-of-sample data. We divide the dataset into three parts, the training set, the model tuning set and the test set. The training set is used to estimate the model parameters. The model tuning set is used to estimate the appropriate model specifications, more specifically, the use of the proposed wavelet entropy based algorithm to determine the specifications for the wavelet model such as the decomposition level. The test set is reserved for the out-of-sample test to evaluate the performance of the proposed model. It needs to be sufficiently large in size to ensure that the test results are statistically valid. The proportion based on which the data set is divided is 60%. The proportion of training set, model tuning set and test set is 36-24-40. One step ahead forecast using rolling-window method is performed. Since Autocorrelation and Partial Autocorrelation function analysis indicates that the original data include trend factors, it is log differenced at the first order as $r_t = ln(\frac{P_t}{P_{t_1}})$ to remove trend factors when the data set is constructed. The returns are transformed to be scale free, which correspond to percentage changes in financial positions and have more attractive statistical properties such as stationarity, *etc*. For the investment size, we assume one dollar equal holding position for initial investment in each market. We also assume one day holding period.

Firstly we calculated some routine descriptive statistics and statistical tests. To determine whether simple linear models suffice for the data being analyzed and whether nonlinear models are necessary, tests for nonlinearity data characteristics are employed, which include Brock-Dechert-Scheinkman (BDS) test [40,41], Bispectrum test [42], and Bicorrelation test [43], *etc*. Among them, the BDS test, since its introduction by Brock, Dechert, and Scheinkman, has become the standard for testing independence in the data. The null hypothesis for the test is that elements of the time series are independently and identically distributed (IID). Experiment results using the in-sample crude oil data are listed in Table 1.

Table 1. Descriptive statistics and statistical tests using the training set data.

Statistics	Mean	Standard Deviation	Skewness	Kurtosis	p_{JB}	p_{BDS}
r_{WTI}	0.0011	0.0239	–0.5162	4.9198	0.0001	0.0043
r_{Brent}	0.001	0.0223	–0.0872	4.5961	0.0001	0.6195

In Table 1, we have found the indication of the nonlinear and nonnormal characteristics. Four statistical moments indicate the significant deviation from the normal distribution, where the skewness deviates from 0 and kurtosis deviates from 3. The rejection of the null hypothesis of Jarque-Bera test of normality confirms the deviation from the normal distribution. The rejection of the BDS test in WTI suggests that the market contains nonlinear dynamics. This is consistent with the observations that both crude oil markets are subject to frequent shocks constantly and major extreme events are prevalent in these markets. For example, Hamilton [44] summarized several major events affecting the crude oil markets [44]. Two most important shocks during the period from 2002 to 2015 are Venezuelan unrest and the second persian Gulf war in 2003 as well as the growing demand and stagnant supply partially due to the financial crisis between 2007 and 2008. Group [45] suggested that oil price plunge between 2003 and 2004 qualifies as another recent significant event [45]. There are numerous occasional minor shocks across the markets. Interestingly, it can be observed that the distribution of these shocks may deviate significantly from the normal distribution. Thus the Gaussian assumptions of noises underlying majority of denoising algorithm underestimate the noise level, resulting in less well behaved data to be modeled. In Brent market, the null hypothesis of BDS test can not be rejected at the statistically significant level, but its p value is not large enough to indicate the linear dependence in the data. Between two markets, Brent market has higher level of efficiency than that in the WTI market. But neither of them is efficient enough to rule out the potential patterns exploitable for improving the forecasting accuracy. These results are consistent with results in Zhang *et al.* [46] in the literature.

Recently there have been numerous researches on the use of wavelet analysis combined with other advanced models such as neural network models, *etc*. It has been shown that they achieved superior performance than the ARMA, ARIMA and random walk models using the sample data. More sophisticated models certainly are more appealing and usually demonstrate the superior performance with the tested data, but their performance are usually sensitive to the data window and choices of initial parameters, as well as risk overfitting the data. Thus in this paper the Random Walk (RW) model and ARIMA model are chosen as the benchmark models because they demonstrate the most robust and consistent performance in the literature. The widespread use of both models as benchmark models has been witnessed in numerous works in the literature such as Alquist and Kilian [5], Yu *et al.* [7].

Following the literature on the model performance evaluation, we adopt Mean Square Error (MSE) as well as the Clark West test of predictive accuracy. We conducted experiments using comprehensive set of wavelet families widely used in the literature, including Daubechies (Db2, Db3, Db4, Db5, Db 6), Coiflet (Coiflet1, Coiflet2, Coiflet3, Coiflet4, Coiflet5), Biorthogonal (Bior11,Bior22, Bior31, Bior39), Reverse Biorthogonal (Rbio11, Rbio22, Rbio31, Rbio39) wavelet families.

Secondly, we calculated the entropy value for the wavelet coefficients based on different wavelet families at different scales, as well as MSE of the proposed algorithm using different wavelet families at different scales. Results are listed in Table 2.

Table 2. Performance comparisons of different models using in-sample training data set.

s_1	s_2	s_3	s_4	s_5	s_6	a	
$Entropy_{(MIN,WTI)}$	2.6419	1.1803	0.6232	0.3036	0.1632	0.0813	0.0509
$WF_{(MIN,WTI)}$	sym2	coif5	sym2	bior39	coif3	sym2	coif5
$WE_{MIN,WTI}$	1.2757	1.2737	1.2757	1.2819	1.2759	1.2757	1.2737
$Entropy_{(MIN,Brent)}$	2.3730	1.0827	0.6130	0.3318	0.1971	0.1270	0.0591
$WF_{(MIN,Brent)}$	Sym2	db5	db4	db3	db4	rbio39	dmey
$WE_{MIN,Brent}$	1.4004	1.3547	1.3441	1.3493	1.3441	1.3353	1.3454

Where $S_i, i = 1, 2, \ldots, 7$ refers to the scale i, $WF_{(MIN,i)}, i \in \{WTI, Brent\}$ refers to the chosen wavelet family using the entropy minimization principle in market i. It can be seen from Table 2 that the entropy values vary significantly across different parameters domain. This implies that there are different energy distribution of wavelet coefficients using different wavelet families across scales. Based on the entropy minimization principle for each scale, the optimal wavelet families with their entropy value is listed in Table 2. Based on the wavelet entropy minimization principle, the Coiflet5 at scale 7 is chosen for WTI market while the rbio39 at scale 7 is chosen for the Brent market. This is an interesting result. For WTI market, the chosen Coiflet wavelet family is symmetric in shape. The 5 vanishing moment is chosen for Coiflet wavelet family. This implies that the investment strategy for investors is more symmetrical in WTI market. For Brent market, the chosen reverse biorthogonal 39 wavelet family is not orthogonal and is asymmetric in shape. The vanishing moment for decomposition and reconstruction is 3 and 9. This implies that the investment strategy for investors is more asymmetric in Brent market.

Then we use the model tuning data set to calculate the adjustment ratio. We adopt the robust regression method to reduce the negative impacts of outliers on the estimate parameter accuracy. As for the regression coefficients, the calculated intercept and slope coefficients are -0.00017812 and 0.0723 for WTI market, as well as -0.00020989 and 0.0343 for Brent market.

With the chosen set of parameters, we further conducted experiments using the out-of-sample data set to evaluate the performance of the proposed algorithm, against the benchmark models. The lag order for the ARMA model used to fit the denoised data from the previous stage is determined using the Information Criteria minimization principle. In this paper, we adopt the AIC and BIC information criteria. We further adopt the Clark West test of equal predictive accuracy to test for the statistical significance of the out-of-sample performance gap between the proposed model and the benchmark models [47,48].

The MSE is the simple statistics measuring the deviation of forecasts from actual observations, as defined in Equation (13).

$$MSE = \frac{1}{N}\sum_{i=1}^{N}[\widehat{f_i(x)} - f_i(x)]^2 \tag{13}$$

where N is the number of observations, $\widehat{f_i(x)}$ refers to the forecasted value and $f_i(x)$ refers to the true value.

The Clark-West (CW) test was proposed to adjust Diebold Mariano (DM) statistics since the originally proposed test statistics were upward biased heavily if models tested were nested. The test statistics is defined in Equation (14).

$$\begin{aligned} CW &= \frac{\overline{d}}{P\sqrt{\widehat{V}}} \\ \overline{d} &= \frac{1}{P}\sum_{t=R+h}^{T} d_t^*, d_t^* = e_{1t}^2 - e_{2t}^2 + (\widehat{y}_{1t} - \widehat{y}_{2t})^2 \\ \widehat{V} &= \frac{1}{P}\sum_{t=R+h}^{T}(d_t - \overline{d})^+ \frac{2}{P}\sum_{j=1}^{m}\omega(j,m)\sum_{t=j+r+h}^{T}(d_t - \overline{d})(d_{t-j} - \overline{d}) \end{aligned} \tag{14}$$

where $\overline{d}$ is sample mean while P is the sample size.

The null hypothesis for the test is: $H_0 : Ed_t^* = 0$, *i.e.*, equal predictive accuracy for Mean Square Percentage Error (MSPE). The test statistics have an asymptotically normal distribution.

Results are listed in Table 3.

Table 3. Performance comparisons of different models using out-of-sample test data set.

Market	$MSE_{\times 10^{-4}, RandomWalk}$	$MSE_{\times 10^{-4}, ARMA}$	$MSE_{\times 10^{-4}, WE}$	$p_{CW, RandomWalk}$	$p_{CW, ARMA}$
WTI	7.5435	3.5717	3.5566	0	0.0657
Brent	4.8724	2.6157	2.6020	0	0.0238

Results in Table 3 show that the proposed wavelet entropy based forecasting algorithm has achieved the improved forecasting accuracy in both markets, with lower MSE values. The superior performance of the proposed model is statistically significant against the ARMA model at 95% confidence compared to the benchmark models in both markets. When the performance is compared to the ARMA model, the superior performance is statistically significant at 95% confidence level in Brent market and 93% confidence level in WTI market.

The out-of-sample performance improvement is attributed to more optimal model specification and parameters identified by the proposed wavelet entropy technique. The experiment results show that the proposed model has good level of generalizability. The out-of-sample performance of the proposed model improves as a result of incorporating this important data feature. The proposed model can adapt to different data characteristics in WTI and Brent markets. We have determined different optimal model with different optimal wavelet families and decomposition levels. The wavelet families and the decomposition levels determined reflect different market characteristics.

These results can offer some important lessons to note and have some further implications. Firstly our results provide the empirical evidence that both WTI and Brent markets have multi scale market structure. We have learned that the multiscale data feature is unignorable during the modeling process. It significantly affects the forecasting performance and can contribute significantly to the performance improvement. We learn that different data features should be incorporated with the model assumptions to be relaxed further. Secondly our results show that the performance of the multiscale model critically depends on the specifications and parameters determined. The multiscale models can be misspecifed so that the following-up analysis results derived can not be trusted. There is an important lesson to learn here for the future multiscale models. Our results show that wavelet entropy serves as a nontrivial tool for determining the optimal model specifications and parameters for multiscale analysis, such as wavelet analysis and empirical mode decomposition. There are redundant representation of the data in the multiscale analysis, which corresponds to the fact that wavelet based model using different wavelet families and decomposition level may have the same level of out-of-sample forecasting accuracy. Wavelet transformed data using the arbitrarily chosen wavelet families and decomposition level may risk twisting the underlying structure and provide the significantly biased estimate. Our results provide the initial evidence that the wavelet entropy is an essential tool to help with the model optimization in the multiscale analysis.

In Figures 1 and 2, we further plot the original crude oil return, the forecasts of the Random Walk model, the forecasts of the ARMA model and the forecasts of the proposed model to analyze the performance off different models.

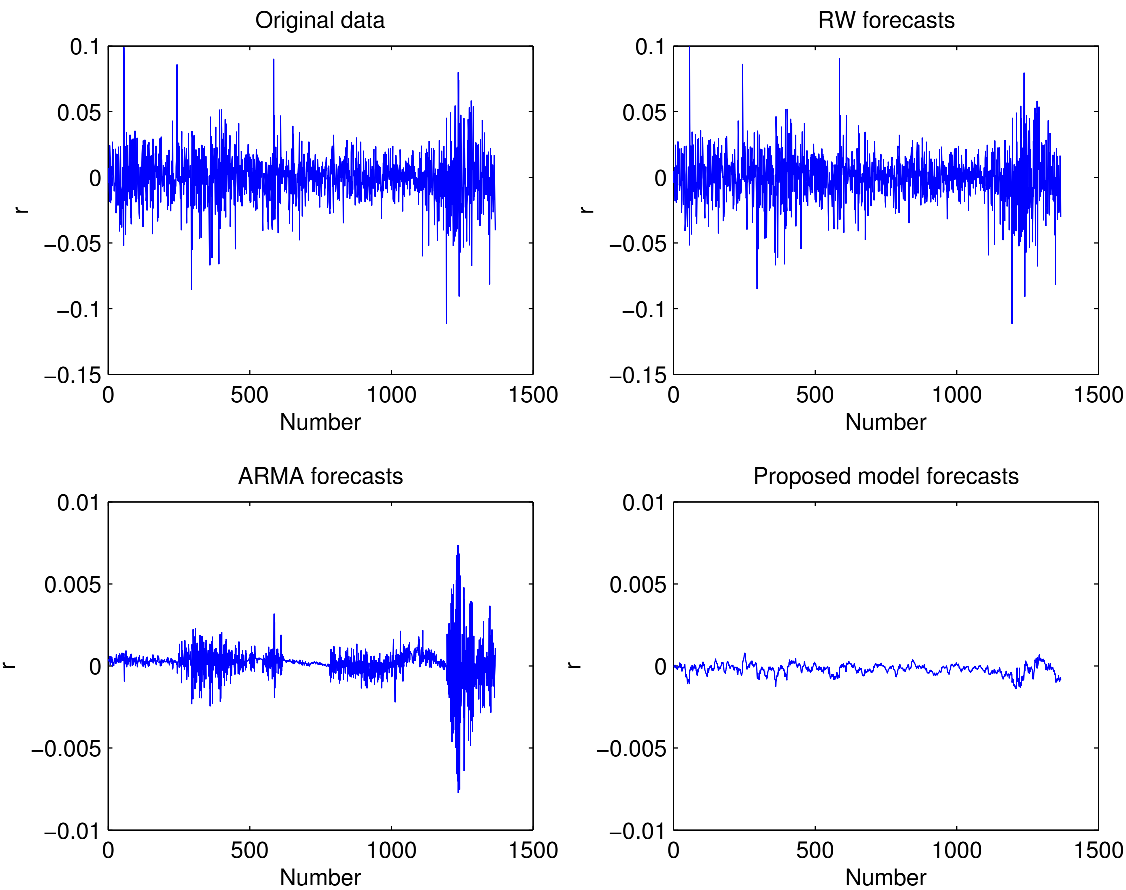

Figure 1. Return forecasts movement of different models.

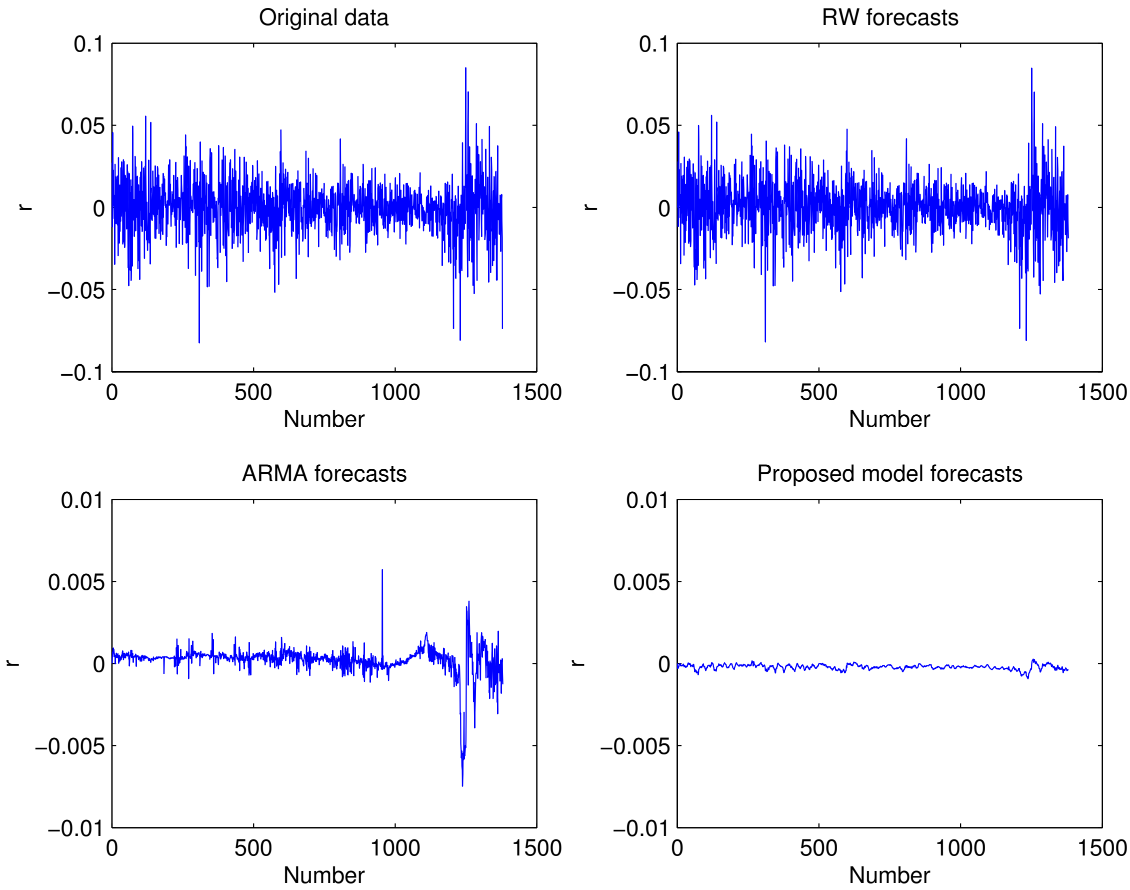

Figure 2. Return forecasts movement of different models

Where r refers to the log differenced price and forecasts, n is the number of the observations. In general the crude oil market is very volatile and is subject to frequent shocks. The forecasting accuracy of the traditional ARMA model is sensitive to the disruptions of many transient and extreme events. The proposed model is more conservative and more effective in suppressing the influence of many transient and extreme events, which take different shapes and have different impacts on different markets across countries. Using Coiflet5 and Reverse Biorthogonal 39 wavelet as the effective filters in both US and UK market, we more effectively separate the transient and extreme events and obtain the denoised crude oil price. More specifically from Figures 1 and 2, it can be seen that forecasts from different models have different level of volatility and forecasting accuracy. Among different forecasts, random walk model tracks very closely the return movement and is the most sensitive to different shocks from the transient and extreme events. The fact that it achieves the lowest performance confirms the transient and temporal nature of these major events, which disrupts the smoothness and robustness of the forecasts. The forecasts from the ARMA model have lower level of fluctuation, suppressing some outliers and extreme events, and achieving the improved forecasting accuracy. The forecasts from the proposed model have the lowest level of fluctuations and volatility among forecasts from different models. It also suppresses majority of transient and extreme events and results in more smooth forecasts. This indicates that during the modeling process, the proposed model more effectively suppresses the short term impacts of transient and extreme events and retains the long term influences of basic macroeconomic factors.

We can also see from Figures 1 and 2 that disruptions from major oil plunges have been kept at the minimum in the forecast from the proposed model. For example, the influence from the oil plunge between 2003 and 2004 has been largely suppressed. This indicates that the major events during this period are transient and temporal in nature. They may have huge impact over a very short period of time, but their influence quickly vanishes outside that period. The forecasts from the proposed model are also much smoother, indicating that the influence from those minor disruptions are reduced to the minimum. Another interesting observation is that in WTI market, the impact of disruptions from those extreme and transient major events takes the symmetric shape, as illustrated by Coiflet wavelet used in WTI market. In Brent market, the impact of disruptions from those extreme and transient major events takes the asymmetric shape, illustrated by Reverse Biorthogonal wavelet used in Brent market.

4. Conclusions

In this paper, we propose wavelet entropy and entropy theory to identify the multiscale model specifications and construct an effective forecasting algorithm for the crude oil price movement. The superior performance of the wavelet entropy based forecasting model is supported by the performance evaluation with the empirical studies using the major crude oil markets data.

The proposed wavelet entropy based model in this paper has far-reaching implications in both theoretical and methodological aspects of multi scale based economic and financial data analysis and modeling such as the wavelet based approach. Firstly, we found that the wavelet parameters have critical impacts on the model performance and accuracy. Due to the inherent constraints on the accuracy of the time and frequency information during the wavelet analysis, there is redundant representations of these parameters. New innovative techniques need to be introduced to determine the best approximation to these parameters. Secondly, the entropy and the wavelet entropy measures, as important information quantification measures, can be used to design suitable criteria to guide the parameter determination during the multi scale analysis. However, the design should take into account the multi scale distribution and structure of there underlying information. The design of appropriate criteria aim at selecting the most generalizable model specifications.

Acknowledgments

This work is supported by the National Science Fund for Distinguished Young Scholars (NSFC No. 71025005), the National Natural Science Foundation of China (NSFC Nos. 71201054, 91224001), the National Program for Support of Top-Notch Young Professionals, the Strategic Research Grant of City University of Hong Kong (No. 7004574), and the Fundamental Research Funds for the Central Universities in Beijing University of Chemical Technology (BUCT).

Author Contributions

All authors participated actively in the discussions and contributed significantly to the experiment and writing. The detailed author contributions are defined as follows: Yingchao Zou: initial draft writing, experiment coding (computing) and analysis; Lean Yu: theory formulation, draft revision; Kaijian He:

idea conceptualization, coding revision and draft revision. All authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Bekiros, S.; Gupta, R.; Paccagnini, A. Oil price forecastability and economic uncertainty. *Econ. Lett.* **2015**, *132*, 125–128.
2. Deng, S.; Sakurai, A. Crude Oil Spot Price Forecasting Based on Multiple Crude Oil Markets and Timeframes. *Energies* **2014**, *7*, 2761–2779.
3. Chen, S.S. Forecasting Crude Oil price Movements with Oil-Sensitive Stocks . *Econ. Inq.* **2014**, *52*, 830–844.
4. Cuaresma, J.C.; Jumah, A.; Karbuz, S. Modelling and Forecasting Oil Prices: The Role of Asymmetric Cycles. *Energy J.* **2009**, *30*, 81–89.
5. Alquist, R.; Kilian, L. What Do We Learn from the Price of Crude Oil Futures? *J. Appl. Econom.* **2010**, *25*, 539–573.
6. Godarzi, A.A.; Amiri, R.M.; Talaei, A.; Jamasb, T. Predicting oil price movements: A dynamic Artificial Neural Network approach. *Energy Policy* **2014**, *68*, 371–382.
7. Yu, L.; Wang, S.; Lai, K.K. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Econ.* **2008**, *30*, 2623–2635.
8. Shin, H.; Hou, T.; Park, K.; Park, C.K.; Choi, S. Prediction of movement direction in crude oil prices based on semi-supervised learning. *Decis. Support Syst.* **2013**, *55*, 348–358.
9. Bildirici, M.; Ersin, O.O. Forecasting oil prices: Smooth transition and neural network augmented GARCH family models. *J. Pet. Sci. Eng.* **2013**, *109*, 230–240.
10. Bates, J.M.; Granger, C.W.J. The Combination of Forecasts. *OR* **1969**, *20*, 451–468.
11. Wen, F.; Li, Z.; Xie, C.; Shaw, D. Study on the Fractal and Chaotic Features of the Shanghai Composite Index. *Fractals* **2012**, *20*, 133–140.
12. Alvarez-Ramirez, J.; Alvarez, J.; Rodriguez, E. Asymmetric long-term autocorrelations in crude oil markets. *Physica A* **2015**, *424*, 330–341.
13. Barkoulas, J.T.; Chakraborty, A.; Ouandlous, A. A metric and topological analysis of determinism in the crude oil spot market. *Energy Econ.* **2012**, *34*, 584–591.
14. Rodriguez, E.; Aguilar-Cornejo, M.; Femat, R.; Alvarez-Ramirez, J. US stock market efficiency over weekly, monthly, quarterly and yearly time scales. *Physica A* **2014**, *413*, 554–564.
15. Shahbaz, M.; Tiwari, A.K.; Tahir, M.I. Analyzing time-frequency relationship between oil price and exchange rate in Pakistan through wavelets. *J. Appl. Stat.* **2015**, *42*, 690–704.
16. Jammazi, R.; Lahiani, A.; Nguyen, D.K. A wavelet-based nonlinear ARDL model for assessing the exchange rate pass-through to crude oil prices. *J. Int. Financ. Mark. Inst. Money* **2015**, *34*, 173–187.

17. Tiwari, A.K.; Bhanja, N.; Dar, A.B.; Islam, F. Time-frequency relationship between share prices and exchange rates in India: Evidence from continuous wavelets. *Empir. Econ.* **2015**, *48*, 699–714.
18. He, K.; Yu, L.; Lai, K.K. Crude oil price analysis and forecasting using wavelet decomposed ensemble model. *Energy* **2012**, *46*, 564–574.
19. Jammazi, R.; Aloui, C. Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling. *Energy Econ.* **2012**, *34*, 828–841.
20. de Souza e Silva, E.G.; Legey, L.F.L.; de Souza e Silva, E.A. Forecasting oil price trends using wavelets and hidden Markov models. *Energy Econ.* **2010**, *32*, 1507–1519.
21. Yousefi, S.; Weinreich, I.; Reinarz, D. Wavelet-Based Prediction of Oil Prices. *Chaos Solitons Fractals* **2005**, *25*, 265–275.
22. Zhang, J.; Tan, Z.; Li, C. A Novel Hybrid Forecasting Method Using GRNN Combined With Wavelet Transform and a GARCH Model. *Energy Sources B* **2015**, *10*, 418–426.
23. Kriechbaumer, T.; Angus, A.; Parsons, D.; Casado, M.R. An improved wavelet, ÄíARIMA approach for forecasting metal prices. *Resour. Pol.* **2014**, *39*, 32–41.
24. Gallegati, M.; Ramsey, J.B.; Semmler, W. Interest rate spreads and output: A time scale decomposition analysis using wavelets. *Comput. Stat. Data Anal.* **2014**, *76*, 283–290.
25. Blanco, S.; Figliola, A.; Quiroga, R.Q.; Rosso, O.A.; Serrano, E. Time-frequency analysis of electroencephalogram series. III. Wavelet packets and information cost function. *Phys. Rev. E* **1998**, *57*, 932–940.
26. Coifman, R.R.; Wickerhauser, M. Entropy-Based Algorithms for Best Basis Selection. *IEEE Trans. Inf. Theory* **1992**, *38*, 713–718.
27. Pascoal, R.; Monteiro, A.M. Market Efficiency, Roughness and Long Memory in PSI20 Index Returns: Wavelet and Entropy Analysis. *Entropy* **2014**, *16*, 2768–2788.
28. Kim, C.H.; Park, C.H.; Kim, S.Y.; Kim, K.; Scalas, E. Dynamics of avalanche activities in financial markets. *Int. J. Money Phys. C* **2007**, *18*, 119–127.
29. Xu, P.; Hu, X.; Yao, D. Improved wavelet entropy calculation with window functions and its preliminary application to study intracranial pressure. *Comput. Biol. Med.* **2013**, *43*, 425–433.
30. Samui, A.; Samantaray, S.R. Wavelet Singular Entropy-Based Islanding Detection in Distributed Generation. *IEEE Trans. Power Deliv.* **2013**, *28*, 411–418.
31. Wang, S.; Yu, L.; Tang, L.; Wang, S. A novel seasonal decomposition based least squares support vector regression ensemble learning approach for hydropower consumption forecasting in China. *Energy* **2011**, *36*, 6542–6554.
32. Martina, E.; Rodriguez, E.; Escarela-Perez, R.; Alvarez-Ramirez, J. Multiscale entropy analysis of crude oil price dynamics. *Energy Econ.* **2011**, *33*, 936–947.
33. Ortiz-Cruz, A.; Rodriguez, E.; Ibarra-Valdez, C.; Alvarez-Ramirez, J. Efficiency of crude oil markets: Evidences from informational entropy analysis. *Energy Policy* **2012**, *41*, 365–373.
34. Niu, H.; Wang, J. Quantifying complexity of financial short-term time series by composite multiscale entropy measure. *Commun. Nonlinear Sci. Numer. Simul.* **2015**, *22*, 375–382.
35. Yin, Y.; Shang, P. Weighted multiscale permutation entropy of financial time series. *Nonlinear Dyn.* **2014**, *78*, 2921–2939.
36. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.

37. Sang, Y.F.; Wang, Z.G.; Li, Z.L. Discrete Wavelet Entropy Aided Detection of Abrupt Change: A Case Study in the Haihe River Basin, China. *Entropy* **2012**, *14*, 1274–1284.
38. Gencay, R.; Selck, F.; Whitcher, B. Multiscale Systematic Risk. *J. Int. Money Financ.* **2005**, *24*, 55–70.
39. Gençay, R.; Selçuk, F.; Whitcher, B. *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*; Academic Press: San Diego, CA, USA, 2002.
40. Brock, W.A.; Hsieh, D.A.; LeBaron, B.D. *Nonlinear Dynamics, Chaos, and Instability : Statistical Theory and Economic Evidence*; MIT Press: Cambridge, MA, USA, 1991.
41. Panagiotidis, T. Testing the Assumption of Linearity. *Econ. Bull.* **2002**, *3*, 1–9.
42. Hinich, M. Testing for Gaussianity and Linearity of a Stationary Time Series. *J. Time Ser. Anal.* **1982**, *3*, 169–176.
43. Hsieh, D. Implications of Nonlinear Dynamics for Financial Risk Management. *J. Financ. Quant. Anal.* **1993**, *28*, 41–64.
44. Hamilton, J.D. *Routledge Handbook of Major Events in Economic History*; Taylor and Francis Group: New York, NY, USA, 2013; pp. 239–265.
45. Group, W.B. *Global Economic Prospects, January 2015: Having Fiscal Space and Using It*; World Bank: Washington, DC, USA, 2015; pp. 155–182.
46. Zhang, B.; Li, X.M.; He, F. Testing the evolution of crude oil market efficiency: Data have the conn. *Energy Policy* **2014**, *68*, 39–52.
47. Clark, T.E.; West, K.D. Using Out-Of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis. *J. Econom.* **2006**, *135*, 155–186.
48. Clark, T.E.; West, K.D. Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. *J. Econom.* **2007**, *138*, 291–311.