Research Article

Day-Ahead Crude Oil Price Forecasting Using a Novel Morphological Component Analysis Based Model

Qing Zhu,[1,2] Kaijian He,[3] Yingchao Zou,[3,4] and Kin Keung Lai[2,5]

[1] *School of Economics and Finance, Xi'an Jiaotong University, Xi'an 710049, China*
[2] *International Business School, Shaanxi Normal University, Xi'an 710062, China*
[3] *School of Economics and Management, Beijing University of Chemical Technology, Beijing 100029, China*
[4] *College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China*
[5] *Department of Management Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon Tong, Kowloon, Hong Kong*

Correspondence should be addressed to Kaijian He; kaijian.he@my.cityu.edu.hk

As a typical nonlinear and dynamic system, the crude oil price movement is difficult to predict and its accurate forecasting remains the subject of intense research activity. Recent empirical evidence suggests that the multiscale data characteristics in the price movement are another important stylized fact. The incorporation of mixture of data characteristics in the time scale domain during the modelling process can lead to significant performance improvement. This paper proposes a novel morphological component analysis based hybrid methodology for modeling the multiscale heterogeneous characteristics of the price movement in the crude oil markets. Empirical studies in two representative benchmark crude oil markets reveal the existence of multiscale heterogeneous microdata structure. The significant performance improvement of the proposed algorithm incorporating the heterogeneous data characteristics, against benchmark random walk, ARMA, and SVR models, is also attributed to the innovative methodology proposed to incorporate this important stylized fact during the modelling process. Meanwhile, work in this paper offers additional insights into the heterogeneous market microstructure with economic viable interpretations.

1. Introduction

With the technological advancement and development of global economic integration, both the demand and the supply of crude oil are influenced by increasingly complex and diverse market participants around the world. This, together with numerous influential factors, such as weather conditions, political stability, economic prospects, consumer expectations, and business indicators, has led to the fluctuating price movement in the crude oil market. This situation has exacerbated in recent years accompanying the wave of liberalization and globalization, which are beyond the explanatory abilities provided by the current models and methodologies. Meanwhile, since crude oil is traded less frequently than equities, this results in higher levels of market imperfection and relatively lower levels of market efficiency, which leads to theoretically valuable and challenging research problems. Therefore, there has been arising interest from both academics and industries for more accurate modeling and forecasting of its price movement [1, 2].

Traditionally the structural and econometric models, mostly linear in nature, have been the mainstream approach in the crude oil forecasting field. ARMA model represents the typical time series approach while regression model as well as vector autoregressive (VAR) model represents the typical multivariate approaches. Mostly they offer satisfactory performance over the medium to long time horizon but fail over the short time horizons. This indicates that the characteristics of prices in the crude oil market contain unknown nonlinear features in the case of time series models, as well as nonlinear interrelations with other macroeconomic factors in the case of multivariate models. Current approaches alone can only offer insufficient level of explanatory and forecasting power for the price movement.

The nonparametric nonlinear artificial intelligence based approaches such as neural network, support vector regression, and empirical mode decomposition largely rely on the data mining exercises to extract nonlinear data patterns [3, 4]. They have shown some promising performance improvement. However, the performance improvement is not consistent for all test cases [5]. Meanwhile, arguments often arise for their results as they risk overfitting the data. Results solely that relied on these approaches also suffer from their powerful but "black box" approaches as limited insights into the underlying influencing factors with economic rationale can be inferred [3, 4, 6]. Therefore, better understanding of the underlying DGP and accurate forecasting in the crude oil market remains one of the most difficult problems in the field [1, 2].

Recent empirical researches have increasingly revealed the significance of multiscale data behaviors, where the mainstream approaches had failed to explain and incorporate during the modeling process. Therefore, the semiparametric paradigm based on computational harmonic analysis has emerged as the preferred alternative method [7]. This is based on the notion that the seemingly unstationary nonlinear data consist of different stationary data of both linear and nonlinear characteristics. This is consistent with HMH; that is, the market is heterogeneous in nature, and exploitation of this stylized fact leads to better understanding of the underlying DGPs and forecasting accuracy. Thus it is inappropriate to abandon the use of linear models in favor of nonlinear models, which could lead to ignorance of important deterministic patterns. However, different underlying DGPs also need to be incorporated during the modeling process. Linear models, however, capture only parts of them. For example, there are ample empirical evidences suggesting the transient data features, which lead to the extreme value theory in the forecasting field.

Thus, the multiscale decomposition method provides an important alternative but only a handful of works have explored it in this field. Some preliminary findings using the wavelet analysis to analyze the multiscale structure of DGP have led to some positive performance improvement in areas such as crude oil, electricity, equities, and exchange rate markets. These empirical studies have used one single family of wavelets to extract data features of interest. For example, wavelet analysis has been widely used to preprocess or denoise the data. For derivatives market, Haven et al. [8] and Almeida and Moriconi [9] use the wavelet analysis to denoise the option data and find that it significantly improves the option valuation accuracy [8, 9]. For crude oil market, Jammazi and Aloui [10] use wavelet analysis to denoise the data while shifting the focus of the paper on the choice of different activation functions on the performance improvement [10]. Yousefi et al. [11] use the wavelet analysis to decompose the crude oil price and extend the decomposed data components directly to make forecasts [11]. For electricity forecasting purpose, Amjady and Keynia [12], Meng et al. [13], and Aggarwal et al. [14] based their neural network forecasting model on the wavelet preprocessed data series and have obtained positive performance improvement [12–14]. Xu et al. [15] have attempted to combine the wavelet analysis with the support vector machine and have obtained positive performance improvement in the empirical studies in Australian electricity market [15]. Conejo et al. [16] have combined the wavelet analysis and ARMA model to analyze the Spanish electricity market and obtained positive results [16]. Meanwhile, another emerging trend is to identify the causal relationship or interrelationship among different financial markets. For example, recently, Benhmad [17] has conducted the empirical studies to test for Granger causality between oil price and US GDP and finds it varying across time scale domain [17]. Work in Reboredo and Rivera-Castro [18] and Reboredo and Rivera-Castro [19] has identified the contagion and interdependence relationship between crude oil and exchange rate, as well as the crude oil and the stock markets in Europe, respectively [18, 19].

As the real data naturally have multiple representations and are redundant when using a group of wavelet families to represent them, the single basis approach implies obviously very strong and questionable assumptions. Thus it is of critical importance to represent the DGP by using mixture of different bases, as suggested by recent empirical researches [7, 20, 21]. Since data now have multiple representations in the dictionary of bases with no unique solution, sparsity is proposed as the measure to guide the searching process for the optimal multiple bases representations. The true underlying component should concentrate on its band of influences, which usually represent a few significant points, which in turn correspond to the sparsity definition. Morphological component analysis (MCA) is one emerging technique in the field of sparsity representations of signals. Some positive performance improvements have been witnessed with limited applications in engineering fields where MCA has been used to extract multiple data features of interests [22–25]. For example, in image processing, Liang et al. [26] use the MCA to solve the face hallucination problem, whose resolution relies on the accurate decomposition of the original image into the global high resolution image and an unsharp mask [26]. Grosdidier and Baussard [27] use MCA to extract the target signature from the Range-Doppler images, which contributes to more accurate suppression of noises [27]. Abrial et al. [25] show that MCA can be applied to analysis of spherical data maps [25]. In the field of medical engineering, Gao et al. [28] propose MCA as an effective mammographic mass detection tool that could achieve satisfactory detection performance [28]. Meanwhile, applications of MCA also extend to different fields of engineering. Bobin et al. [29] extend MCA to multichannel case to solve the multichannel inverse problems [29]. Zeng et al. [30] use MCA to separate transient events and stationary noises based on their different morphological characteristics in vibration signals in the Watt experiment [30].

This paper proposes a morphological component analysis based hybrid methodology for modeling and forecasting of crude oil price. Results in this study explore and unveil the complex market structure consisting of data components of different data characteristics modeled using morphological component analysis. Empirical studies have been conducted in the marker benchmark West Taxes Intermediate (WTI)

and Brent markets to investigate the performance improvements of the proposed model, against traditional benchmark models. The main contribution of this paper is the introduction of multiple basis based approach, to recover the underlying constituent component with only the prior information of the signals available to study the heterogeneous market structure without the inside information. This represents the important divergence from the widespread oversimplified single basis approach, inconsistent with market structure, and is only valid at macroscale. The introduction of MCA based approach incorporated the stylized fact that there are redundant forms of representations on the underlying data generating process, which need to be optimized, and contributes to the understanding and forecasting of the evolutions in the market microstructure.

This paper is organized as follows. Section 2 provides a brief account of the sparsity decomposition and MCA theories behind. Section 3 proposes and illustrates MCA based hybrid methodologies and their numerical procedures. Experiment results for empirical studies are reported and analyzed in Section 4, based on which finalizing conclusions are drawn in Section 5.

2. Relevant Theories

2.1. Sparsity Decomposition and Morphological Component Analysis. Over the years, numerous bases have been developed to capture particular data features. Typical bases used include global oscillating discrete sine and cosine transform (DST and DCT) and locally oscillating discrete wavelet transform (DWT). With the availability of a large number of bases to construct an overly complete dictionary, guiding measures are needed to select and distinguish bases for specific data features to provide the most efficient representations. Sparsity is one such measure. Different basis may be more efficient in representing particular data features and thus provide sparse representation better than for other data features. Meanwhile, sparsity provides a measure for bases to discriminate against different data features [23–25]. Therefore, the goal of sparsity decomposition is to search for the appropriate representation of signals based on the morphological diversity criteria. It attempts to represent the signals using a dictionary of overcomplete redundant basis dictionary and model them as the linear combination of different morphologically diversified components [24]. Formally it is defined as follows.

Suppose D is an overcomplete dictionary constructed as the union of orthonormal bases as $D = [\xi_1, \dots, \xi_D]$, from a collection of signal waveforms or atoms $\delta_1, \dots, \delta_K$. The projection coefficient α can be defined for mapping signal $y \in R^n$ into the bases domain as in the following:

$$y = \alpha_i D. \tag{1}$$

The signal is sparse in dictionary D if most entries of projection coefficients α are zero and there exist only a few significant ones. This constraint also ensures an uncertainty principle that states that a component that is sparse in a particular basis δ_k is not sparse in a mutually incoherent basis δ_i, $i \neq k$. This group of basis pairs may include DCF and wavelets and wavelets and Dirac bases.

Suppose signal y can be represented by a linear combination of k components S_k, $k = 1, \dots K$, where each component S_k is sparse in the corresponding unique bases in dictionary δ_k. The goal of sparsity decomposition is to search for components obtained from orthogonal transform with basis functions and provides the optimal linear representation of a data series y. This problem is formulated as the linear optimization problem in the following for a_k:

$$\min_{a_1, \dots, a_k} \sum_{k=1}^{K} \|a_k\|_0$$

$$\text{s.t.} \quad y = \sum_{k=1}^{K} \delta_k a_k, \tag{2}$$

where the function l_o norm refers to $\|x\|_o$, that is, the support of x, counting the number of nonzero components of vector x.

Since it is a typical NP-hard combinatorial discrete optimization problem, algorithms with relaxed condition or approximate accuracy have been developed over the years to reduce computational complexity. These include the greedy matching pursuit (MP) [31], basis pursuit (BP) and basis pursuit denoising (BPDN) [32], LARS [33], and MCA [34]. Compared to BP, MCA represents an alternative efficient approach based on iterative thresholding algorithm.

Different orthogonal bases have been developed over the years. Typical examples are the traditional discrete cosine function (DCF), wavelets, and so forth. DCF is a global invariance function, well researched in the literature. Wavelets are more complex continuous functions that have high energy concentration over short intervals of time, which is in direct contrast to the globally time invariant sinusoid functions used in more traditional spectrum analysis tools such as Fourier analysis [35]. Wavelets are defined as solutions to the two-scale difference equation as in the following [36]:

$$C_\psi(x) = \sqrt{2} \sum_{k \in z} h_k \psi(2x - k). \tag{3}$$

Therefore, wavelet functions satisfy both admissibility and unit energy conditions that guarantee their time scale localization with zero vanishing moments.

There are different families (or types) of wavelets. Each is capable of adapting to and accentuating certain data characteristics. Typical wavelet families include Haar, Daubechies, and Coiflets, each with different characteristics such as support and vanishing moments [36].

2.2. Support Vector Regression (SVR). Support vector regression (SVR) is an emerging machine learning theory. It adopts the structural risk minimization principle during the data training and learning process and models it as a convex optimization problem to balance between fitting accuracy and model generalizability. Thus it alleviates the overfitting and local minima issues with the traditional supervised learning algorithm, such as neural network models, which

are based on the empirical risk minimization principle [37]. Support vector regression is the extension of the support vector machine theory in regression analysis [37]. This is achieved introduction of some loss function such as the most popular ϵ-insensitive loss function. Typically for a set of data points $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$, where $x_i \in R^n, y_i \in R^n$, the linear regression problem is formulated as

$$f(x) = \omega\phi(x) + b, \qquad \phi: R^n \longrightarrow F, \ \omega \in F, \qquad (4)$$

where $\phi(x_i)$ denotes the transformation in feature F. Given nonlinear data, kernel function $k\langle x_i \cdot x \rangle$ is used for $\phi(x_i)$ to map nonlinear inputs into linear inputs in higher dimensions. w is the weight in feature space F.

Introducing the ϵ loss function, the regression problem is further transformed into the convex optimization problem formulated as in

$$\min \quad R(\omega, \xi_i, \xi_i^*) = \frac{1}{2}\|\omega\|^2 + C_i \sum_{i=1}^{n} (\xi_i + \xi_i^*)$$

$$\text{s.t.} \quad \begin{cases} y_i - \phi(\omega, t) - b \le \varepsilon_i + \xi_i \\ \phi(\omega, t) + b_i - y_i \le \varepsilon_i + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \\ C_i \ge 0 \end{cases}, \qquad (5)$$

where ϵ is the loss function that measures the forecast deviations allowed. Two slack variables ξ_i, ξ_i^* are introduced to measure the size of positive and negative deviations. C is the penalty variable for empirical errors. $w^T w$ is the regularized term.

Applying Lagrangian and Karush-Kuhn-Tucker conditions, the dual problem of the original optimization problem is formulated and solved as in the following to reduce dimensionality and computational costs:

$$\max \quad L = \sum_{i=1} a_i - \frac{1}{2}\sum_{i,j} a_i a_j y_i y_j x_i^T x_j$$

$$\text{s.t.} \quad \begin{cases} \sum_i a_i y_i = 0 \\ a_i \ge 0 \quad \forall i \end{cases}. \qquad (6)$$

As an emerging technique, application of SVR in forecasting literature has been growing in recent years. It is typically viewed as an improvement over the traditional neural network models to avoid the local minima issue. Some very recent development includes its variant named multiple input multiple output SVR. For example, Bao et al. [38], Bao et al. [39], and Xiong et al. [40] have used the MIMO SVR in predicting the time series multiple steps ahead. The application evaluation of the proposed algorithm using the physics time series and stock market data has confirmed the improved predictive accuracy in the stock market [38–40]. Cao and Tay [41] and Ince and Trafalis [42] find separately that the performance of SVR is better than the neural network model [41, 42]. In electricity market, Aggarwal et al. [43] obtain significant performance improvement over traditional models and neural network models [43]. Zhao et al. [44] use the SVR model to forecast price movement and variance and

constructed the prediction interval, claiming the model to be better than the traditional GARCH approach [44]. The SVR model is also being applied in forecasting crude oil price and has achieved superior positive performance. Xie et al. [45] construct a support vector regression based time series forecasting model for crude oil market and found its performance superior to backpropagation neural network and ARMA model [45]. Despite the positive results reported in the literature, SVR suffers from the same problems as neural network; that is, its performance is also sensitive to the chosen parameters, especially the trade-off parameters C, reflecting human preference for the balance between overfitting and generalizability. Recent progress in using the metaheuristics approach to determine the parameters has led to significantly improved performance. For example, Li and Tan [46] and Bao et al. [47] have shown that the evolutionary algorithms such as PSO and memetic algorithm can be used to determine more optimal model specifications for SVR and result in improved performance [46, 47]. Tian et al. [48] have also provided an alternative multiple kernel based framework that emphasizes the inductive approach to determine the parameters in SVR tuning [48]. Although there are debates on the economics insights that SVR based models can offer, SVR model serves as a very good optimization model during the forecasting process, especially when its parameters are fine-tuned with some advanced techniques.

3. A Morphological Component Analysis (MCA) Based Hybrid Methodology

The proposed MCA based hybrid methodology follows the "divide and conquer" principle. The theoretical basis behind the proposed approach is the proposition of heterogeneous market hypothesis (HMH), relaxing the homogeneous and rational assumption of efficient market hypothesis (EMH) underlying the majority of mainstream models. The rationale is that EMH assumes homogeneous time horizon, frequency, and individual characteristics in the data, which provides the acceptable level of approximations over the medium to long term time horizon when the market structure is relatively simpler due to strict regulations and demand for the forecasting accuracy, are only at moderate level. However, over the shorter interval, there are market imperfections that enable price predictability. Meanwhile, with increasingly complex market structure due to deregulation and technological development, recent empirical evidence in the exchange rate and equity markets suggests that the heterogeneous nature of market is no longer ignorable and could be the key to explain and reconcile EMH and empirical stylized facts that suggest price predictability [49, 50]. Meanwhile, the heterogeneous market hypothesis (HMH) arises to complement the traditional efficient market hypothesis (EMH) [51–57]. The HMH proposes that the market consists of heterogeneous agents with heterogeneous investment strategies and investment time horizons. Compared to the homogeneous reaction to the news shocks in the EMH, HMH states that these agents or investors react to news shocks differently based on their own characteristics. On one hand, their investment time horizon

and dealing frequency are diversely different. For example pensions funds and central banks tend to have low dealing frequency focusing on long time horizon while the market traders tend to have high dealing frequency with short time horizon. On the other hand, these agents or investors employ diversely different investment strategies or measures based on their own characteristics and focus. In the meantime, as recent harmonic analysis research suggests, appropriate recovery of the underlying morphologically diversified components based on sparsity decomposition is important to data trend modeling. The original data series with the mixture of linear and nonlinear data characteristics need to be decomposed into the underlying morphologically diversified DGP, whose distributional characteristics conform to the assumptions of mainstream econometric models. Technically to study the heterogeneous market structure without inside information, innovative algorithms such as MCA that can recover the underlying constituent component are needed since only the prior information of the signals is available.

The proposed MCA approach also represents a significant paradigm shift from traditional approaches, over simplifying assumptions that are inconsistent with market structure, and is only valid at macroscale. The MCA based approach incorporates the stylized fact that there are redundant forms of signal representations. The accuracy of optimal extraction of components with overcomplete dictionary of bases that represent our assumptions has a limit, governed by the uncertainty principle. However, the approximation of accuracy can improve continuously with the development of technology. Based on HMH, MCA assumes that the data are influenced by some underlying components, which have morphologically diverse features, distinguishable with a dictionary of bases, together with additive noises as defined as in

$$r(t) = \sum \xi^{(i)} \alpha^{(i)} + n, \qquad (7)$$

where $r(t)$ and $\alpha^{(i)}$ refer to the original data series and underlying component series of morphologically different characteristics such as permanent and transient data characteristics. Consider $i \in k$ where $\delta_1, \ldots, \delta_k$ refers to the collection of morphologically different dictionaries including undecimated discrete wavelet, discrete cosine transform, and the Dirac basis. $\xi^{(i)}$ refers to the coefficient vectors for different complete dictionaries. n refers to the contaminating noises, possibly of Gaussian white noise of irrelevant nature.

MCA involves several key steps, including components extraction, feature identification, data modeling, and final forecast phases, as illustrated in Algorithm 1.

Firstly in the components extraction phase, the set of dictionaries with mutually incoherent bases is constructed. This is to ensure that the sparest solution can be found as the uncertainty principle ensures that no signal can have sparse representations in mutually incoherent bases simultaneously. Then, the underlying components can be extracted using the constructed dictionary based on MCA algorithm. The process of feature extraction in MCA follows the standard iterative thresholding algorithm as in Algorithm 2 [25, 34].

Secondly the nonlinear statistics test is used to test and identify linear and nonlinear data characteristics.

Thirdly appropriate model specifications from a set of model pools, consisting of linear, nonlinear, and random walk, are chosen and used to model the extracted components. Individual forecasts are made for each component accordingly. The optimal one is chosen based on the trial and error method.

For data components of linear nature, that is, if the price series are serially linearly dependent and not independent and identically distributed (i.i.d), the autoregressive moving average model (ARMA) is used to model the linear serial dependence in data, in which the current price level is linearly related to the past price level, incorporating the errors in previous forecasts as well. Typical ARMA model specification is estimated and forecasts are made as in

$$\widehat{r_t} = \delta_t + \sum_{i=1}^{m} \phi_i r_{(t-i)} + \sum_{j=1}^{n} \theta_j \varepsilon_{(t-j)} + a_t, \qquad (8)$$

where r_t is the conditional mean of the data at time t, $r_{(t-i)}$ is the lag m returns with parameter ϕ_i, and $\varepsilon_{(t-j)}$ is the lag n residuals in the previous period with parameter θ_j. δ_t is the constant coefficient. a_t is the error term at time t.

For nonlinear data, the nonlinear model specification is estimated and forecasts are made as in

$$\widehat{r_t} = \sum_{i=1}^{p} \omega_i r_{(t-i)} + b_t. \qquad (9)$$

If market is efficient, the random walk model is valid in that all past information is reflected in the current price, which is the only needed information to forecast the future movement. The price series is i.i.d and not predictable based on past information. Thus the random walk model remains a very important benchmark model as predictions from most linear models are less robust and are biased due to inappropriate extraction of patterns.

Based on the basic assumption in MCA analysis as in (7), the final forecasts are simply linear summation of individual forecasts made for different individual components as in

$$\widehat{r(t)} = \widehat{r_{\text{PT}}(t)} + \widehat{r_{\text{TF}}(t)}. \qquad (10)$$

4. Empirical Studies

4.1. Data and Descriptive Statistics. As two representative benchmark marker markets were considered by US Energy Information Administration, both the US West Taxes Intermediate (WTI) crude oil market and the European Brent (Brent) are used as the testing fields for empirical studies in this paper. The experiments are designed following the convention in the literature. For both datasets, the performance evaluations of different models are conducted covering the time period from January 2, 2002, to February 13, 2009, when the latest event and data are incorporated while the impact of previous direct market disruptions, such as the Gulf war, is reduced to the minimum. This includes 1790 daily observations for WTI dataset and 1868 daily observations for Brent dataset. The data source is Energy Information Administration (EIA) of US Department

$N \leftarrow$ the initial decomposition scale

$r_t = \sum\limits_{i}^{N} r_{i,t}$ {Component extraction}

while $i < N$ **do**

    $s_i \leftarrow$ Statistical tests of the null hypothesis that the DGP is i.i.d or linear {Feature identification}

    if $s_i \geq 0$ **then**

        $\widehat{r_t} = \delta_t + \sum\limits_{i=1}^{m} \phi_i r_{(t-i)} + \sum\limits_{j=1}^{n} L\theta_j \varepsilon_{(t-j)} + a_t$ {Linear data modeling}

    else

        $\widehat{r_t} = \sum\limits_{i=1}^{P} \omega_i r_{(t-i)} + b_t$ {Nonlinear data modeling}

    end if

end while

$\widehat{r(t)} = \sum\limits_{i}^{N} \widehat{r_{i,t}}$ {Forecast reconstruction}

ALGORITHM 1: A MCA based hybrid methodology.

$I_{\max} \leftarrow$ number of iterations

$\lambda^{(0)} \leftarrow$ the initial threshold

$\widehat{y}^{(0)} \leftarrow 0.$

$\lambda_{\min} \leftarrow$ lower bound

while $\lambda^{(t)} > \lambda_{\min}$ **do**

    $\widehat{r(t)} = \widehat{r_{PT}(t)} + \widehat{r_{TF}(t)}$ {Compute the residual}

    $\alpha^{(t)} = \phi_{\lambda^{(t)}}(\xi^{-1}(r^{(t)}))$ {Thresholding}

    $\widehat{y}^{(t)} = \xi\alpha^{(t)}$ {Reconstruction}

    $\lambda^{(t+1)} \leftarrow$ Decreased threshold $\{\lambda^{(t+1)} < \lambda^{(t)}$ following a given strategy$\}$

end while

ALGORITHM 2: Iterative thresholding algorithm for MCA.

of Energy. The datasets are conventionally divided on the 60–40 basis, which ensures sufficient samples for statistical significance of results. The first 60% of the dataset serves as the training set for estimating model specifications and parameters. The remaining 40% of dataset is reserved as the test set for evaluating performance of different models [58, 59]. The directional predictive accuracy of different models is evaluated using Pesaran-Timmermann test of directional predictive accuracy [60]. The original daily price series is log differenced at the first order to remove trend factors. The statistical predictive accuracy of different models is evaluated using mean squared error (MSE), for measuring the deviation of the predictive values from the actual observations, and the Clark-West statistical predictive accuracy test for nested models, for measuring the predictive accuracy between two nested models [61, 62]. The directional predictive accuracy of different models is evaluated using Pesaran-Timmermann test of directional predictive accuracy [60].

The calculated mean, standard deviation, skewness, and kurtosis for both WTI and Brent markets are 0.0011, 0.0230, −0.4326, and 4.8282 and 0.0011, 0.0218, −0.0763, and 4.3898, respectively. This suggests that on the aggregated level,

crude oil market is relatively efficient and roughly normally distributed. However, the Jarque-Bera test of normality is rejected while the BDS test of independence is accepted at a low confidence level, that is, 70.8% in Brent market, which indicates that crude oil data deviate from normal and independent distribution. Therefore, this paper uses MCA techniques to extract morphologically distinct data components. In this paper, three different bases have been attempted, including DCT, Daubechies 4, and Symlet 4 wavelet family, where the number 4 is the order for the wavelet family.

The DCT and DST fail to extract the deterministic underlying components, which suggests the absence of long term periodic deterministic trends in the market. This is an interesting and provoking observation as it suggests that the market is characterized by salient data features and violation of the stationary character over long term in traditional modeling technique based on single time framework.

For components extracted using Daubechies 4 and Symlet 4, in WTI market, the calculated mean, standard deviation, skewness, and kurtosis are 0, 0.0041, −11.3492, and 273.8995 and 0, 0.0001, 0, and 1.5076, respectively. In Brent market, the calculated mean, standard deviation, skewness, and kurtosis

are 0, 0.0022, 15.2061, and 329.3004 and 0, 0.0001, 0, and 1.5015, respectively. Thus distributions of all components extracted are significantly different from the normal distribution. This is also confirmed by the rejection of the Jarque-Bera test and BDS tests.

4.2. Experiment Results. Following conventions in the literature, three benchmark models, random walk (RW), ARMA, and SVR model, are used in the model evaluation process [63]. The lag order for benchmark ARMA(r,m) during the forecasting process is set to ARMA(1,1). The lag order p for benchmark SVR model is set to 2. The lag order for ARMA(r,m) in the forecasting process is determined based on the Akaike information criteria (AIC) and Bayesian information criteria (BIC) minimization principle.

Since very little guidance is available in the extant literature, parameters for MCA are determined using trial and error method. The rolling window is set to 512 to cover the relevant information set. Wavelet families used in the MCA analysis process include Daubechies 4 and Symlets 4 for WTI market, as well as Daubechies 6 and Symlets 6 for WTI market.

The ϵ-insensitive SVR is chosen with the radial basis function (RBF) kernel. The parameters for SVR model, including cost, gamma, epsilon, and tolerance of termination, are determined using the standard grid search method. They are listed in Table 1.

Predictive accuracy of the proposed algorithm against alternative benchmark models is as in Table 2.

Experiment results in Table 2 show the superior performance of the proposed MCA hybrid methodology, which outperforms individual benchmark models, random walk (RW), ARMA, and SVR model, in terms of predictive accuracy measured by MSE. Meanwhile, result of the Clark-West test of equal predicative accuracy suggests that the superior performance against benchmark models is statistically significant. The performance superiority is significant at 95% confidence level for WTI market and at 90% for Brent market.

Directional predictive accuracy of the proposed model against benchmark alternatives is listed in Table 3.

Experiment results in Table 3 show that the proposed MCA based methodology achieves a higher level of directional forecasting accuracy than the individual benchmark models, ARMA and SVR model, in terms of the ratio of correct predictions. Meanwhile, Pesaran-Timmermann test of directional predictive accuracy suggests that the superior performance against benchmark models is statistically significant. The performance superiority is significant at 90% confidence level for WTI market and at 95% for Brent market.

The optimal performance for WTI market is achieved when components extracted using Daubechies wavelet are modeled by the Random Walk model, while components extracted using Symlets wavelets are modeled by the SVR model. Interestingly the optimal performance for Brent market is achieved when components extracted using Daubechies wavelet are modeled by ARMA model while components extracted by using Symlets are modeled by the SVR model.

Table 1: Parameters chosen for SVR models.

Models	Cost	Gamma	Epsilon	Tolerance of termination
Original$_{WTI}$	32	1	0.01562500	0.00050000
Daubechies 4$_{WTI}$	8	0.06250000	0.06250000	0.00000010
Symmlet 4$_{WTI}$	8	0.06250000	0.06250000	0.00000010
Original$_{Brent}$	2	0.00781250	0.03125000	0.00040000
Daubechies 6$_{Brent}$	32	0.06250000	0.00390625	0.00000600
Symmlet 6$_{Brent}$	8	0.06250000	0.06250000	0.00000001

Table 2: Performance comparison of different models.

Models	MSE$_{\times 10^{-4}}$	CW$_{ARMA}$ $_{P\ value}$	CW$_{RW}$ $_{P\ value}$
RW$_{WTI}$	8.8140	N/A	0.6641
ARMA$_{WTI}$	8.8739	0.6641	N/A
SVR$_{WTI}$	13.4749	0.5076	0.4304
MCA-RW-SVR$_{WTI}$	8.7337	0.0077	0.0274
RW$_{Brent}$	7.2035	N/A	0.4485
ARMA$_{Brent}$	7.2153	0.4485	N/A
SVR$_{Brent}$	12.0060	0.5838	0.6407
MCA-ARMA-SVR$_{Brent}$	7.1106	0.0702	0.0568

Table 3: Pesaran-Timmermann directional test.

Models	Dstat$_{\%}$	PT
ARMA$_{WTI}$	51.96	0.1227
SVR$_{WTI}$	48.74	0.7499
MCA-RW-SVR$_{WTI}$	52.65	0.0768
ARMA$_{Brent}$	50.48	0.4524
SVR$_{Brent}$	46.52	0.9707
MCA-ARMA-SVR$_{Brent}$	54.30	0.0102

Therefore, these observations support the argument that crude oil prices are complicated processes with a mixture of underlying DGPs of different natures. Their modeling and analysis are tricky issues which require detailed investigation of the underlying structure in the time scale domain. Thus appropriate recovery of the underlying structure is critical to further performance improvement during the modeling process. Meanwhile, experiment results also show that the proposed algorithm is generalizable to different datasets with the flexibility to model nonlinear dynamics characterized by mixture of data of time varying natures.

5. Conclusions

Based on the HMH, this paper proposes a hybrid modeling methodology to incorporate multiscale market structure information into the modeling process, which provides a view of the microstructure of the underlying DGP, besides finer modeling accuracy. The proposed algorithm introduces the MCA techniques to analyze the multiscale market structure. Empirical studies on two major benchmark crude oil

markets in the world suggest their effectiveness in analyzing the heterogeneous market structure and demonstrate significant positive performance improvement, as a result.

The proposed methodology reveals that the sparsity decomposition based methodology using MCA offers more complete and accurate representation of data features than the more widely used single basis methodology. The conventional single basis methodology provides only a partial and twisted view of data features by imposing strict assumptions in the inappropriate time domain. However, the proposed algorithm can capture more accurately the underlying DGPs of diverse natures, in both linear and nonlinear domain, by incorporating both multitime scale information and the multiple bases frequency feature information during the modeling process. Original results reported in this paper also merit further research on constructing redundant basis transform to explore the underlying GDPs in the seemingly fractal and chaotic financial market, whose characteristics can be revealed more clearly only in appropriate time and frequency settings. Meanwhile, research results in this paper also pave the way for further research into two largely overlooked assumptions behind mainstream wavelet bases research: the selection of appropriate basis to represent economic and financial data and the atomic decomposition of the underlying data features. The performance of the proposed algorithm is sensitive to the introduced bases parameter.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] C. W. Yang, M. J. Hwang, and B. N. Huang, "An analysis of factors affecting price volatility of the US oil market," *Energy Economics*, vol. 24, no. 2, pp. 107–119, 2002.

[2] A. Plourde and G. C. Watkins, "Crude oil prices between 1985 and 1994: how volatile in relation to other commodities?" *Resource and Energy Economics*, vol. 20, no. 3, pp. 245–262, 1998.

[3] L. Yu, S. Wang, and K. K. Lai, "Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm," *Energy Economics*, vol. 30, no. 5, pp. 2623–2635, 2008.

[4] X. Zhang, K. K. Lai, and S.-Y. Wang, "A new approach for crude oil price analysis based on empirical mode decomposition," *Energy Economics*, vol. 30, no. 3, pp. 905–918, 2008.

[5] M. P. Clements, P. H. Franses, and N. R. Swanson, "Forecasting economic and financial time-series with non-linear models," *International Journal of Forecasting*, vol. 20, no. 2, pp. 169–183, 2004.

[6] K. He, K. K. Lai, S. M. Guu, and J. Zhang, "A wavelet based multi scale var model for agricultural market," *Computation and Optimization in Information Systems and Management Sciences*, vol. 14, pp. 429–438, 2008.

[7] K. He, C. Xie, and K. Lai, "Estimation of value-at-risk for exchange risk via kernel based nonlinear ensembled multi scale model," in *Advances in Neural Networks—ISNN 2008*, vol. 5263 of *Lecture Notes in Computer Science*, pp. 148–157, 2008.

[8] E. Haven, X. Liu, and L. Shen, "De-noising option prices with the wavelet method," *European Journal of Operational Research*, vol. 222, no. 1, pp. 104–112, 2012.

[9] V. T. X. de Almeida and L. Moriconi, "Option pricing from wavelet-filtered financial series," *Physica A: Statistical Mechanics and Its Applications*, vol. 391, no. 20, pp. 4850–4854, 2012.

[10] R. Jammazi and C. Aloui, "Crude oil price forecasting: experimental evidence from wavelet decomposition and neural network modeling," *Energy Economics*, vol. 34, no. 3, pp. 828–841, 2012.

[11] S. Yousefi, I. Weinreich, and D. Reinarz, "Wavelet-based prediction of oil prices," *Chaos, Solitons and Fractals*, vol. 25, no. 2, pp. 265–275, 2005.

[12] N. Amjady and F. Keynia, "Day ahead price forecasting of electricity markets by a mixed data model and hybrid forecast method," *International Journal of Electrical Power and Energy Systems*, vol. 30, no. 9, pp. 533–546, 2008.

[13] K. Meng, Z. Y. Dong, and K. P. Wong, "Self-adaptive radial basis function neural network for short-term electricity price forecasting," *IET Generation, Transmission and Distribution*, vol. 3, no. 4, pp. 325–335, 2009.

[14] S. K. Aggarwal, L. M. Saini, and A. Kumar, "Electricity price forecasting in ontario electricity market using wavelet transform in artificial neural network based model," *International Journal of Control, Automation and Systems*, vol. 6, no. 5, pp. 639–650, 2008.

[15] Z. Xu, Z. Dong, and W. Liu, "Short-term electricity price forecasting using wavelet and svm techniques," *Dynamics of Continuous Discrete and Impulsive Systems B: Applications & Algorithms*, pp. 372–377, 2003.

[16] A. J. Conejo, M. A. Plazas, R. Espínola, and A. B. Molina, "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 1035–1042, 2005.

[17] F. Benhmad, "Modeling nonlinear Granger causality between the oil price and U.S. dollar: a wavelet based approach," *Economic Modelling*, vol. 29, no. 4, pp. 1505–1514, 2012.

[18] J. C. Reboredo and M. A. Rivera-Castro, "A wavelet decomposition approach to crude oil price and exchange rate dependence," *Economic Modelling*, vol. 32, no. 1, pp. 42–57, 2013.

[19] J. C. Reboredo and M. A. Rivera-Castro, "Wavelet-based evidence of the impact of oil prices on stock returns," *International Review of Economics and Finance*, vol. 29, pp. 145–176, 2014.

[20] K. He, C. Xie, and K. K. Lai, "A wavelet denoising support vector regression ensemble model for exchange rate prediction," in *Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '08)*, pp. 1–4, October 2008.

[21] H. Kaijian, X. Chi, and K. L. Kin, "Multi scale nonlinear ensemble model for foreign exchange rate prediction," in *Proceedings of the 4th International Conference on Natural Computation (ICNC '08)*, vol. 7, pp. 43–47, October 2008.

[22] J. M. Fadili, J.-L. Starck, M. Elad, and D. L. Donoho, "MCALab: reproducible research in signal and image decomposition and inpainting," *Computing in Science and Engineering*, vol. 12, no. 1, pp. 44–62, 2010.

[23] J. Bobin, J.-L. Starck, Y. Moudden, and M. J. Fadili, "Blind source separation: the sparsity revolution," *Advances in Imaging and Electron Physics*, vol. 152, pp. 221–302, 2008.

[24] J. Bobin, J.-L. Starck, J. M. Fadili, Y. Moudden, and D. L. Donoho, "Morphological component analysis: an adaptive thresholding strategy," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2675–2681, 2007.

[25] P. Abrial, Y. Moudden, J.-L. Starck et al., "Morphological component analysis and inpainting on the sphere: application in physics and astrophysics," *Journal of Fourier Analysis and Applications*, vol. 13, no. 6, pp. 729–748, 2007.

[26] Y. Liang, X. Xie, and J.-H. Lai, "Face hallucination based on morphological component analysis," *Signal Processing*, vol. 93, no. 2, pp. 445–458, 2013.

[27] S. Grosdidier and A. Baussard, "Ship detection based on morphological component analysis of high-frequency surface wave radar images," *IET Radar, Sonar and Navigation*, vol. 6, no. 9, pp. 813–821, 2012.

[28] X. Gao, Y. Wang, X. Li, and D. Tao, "On combining morphological component analysis and concentric morphology model for mammographic mass detection," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 266–273, 2010.

[29] J. Bobin, Y. Moudden, J. Fadili, and J.-L. Starck, "Morphological diversity and sparsity for multichannel data restoration," *Journal of Mathematical Imaging and Vision*, vol. 33, no. 2, pp. 149–168, 2009.

[30] W. Zeng, X. Jiang, I. M. Smith, and P. Scott, "Transient signal separation in Watt balance experiments," *Physics Letters A: General, Atomic and Solid State Physics*, vol. 374, no. 11-12, pp. 1301–1306, 2010.

[31] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[32] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[33] B. Efron, T. Hastie, I. Johnstone et al., "Least angle regression," *Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[34] J.-L. Starck, M. Elad, and D. Donoho, "Redundant multiscale transforms and their application for morphological component separation," *Advances in Imaging and Electron Physics*, vol. 132, pp. 287–348, 2004.

[35] R. Gencay, F. Selcuk, and B. Whitcher, *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*, Academic Press, San Diego, Calif, USA, 2002.

[36] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, UK, 2000.

[37] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 2nd edition, 2000.

[38] Y. Bao, T. Xiong, and Z. Hu, "Multi-step-ahead time series prediction using multiple-output support vector regression," *Neurocomputing*, vol. 129, pp. 482–493, 2014.

[39] Y. Bao, T. Xiong, and Z. Hu, "Pso-mismo modeling strategy for multistep-ahead time series prediction," *IEEE Transactions on Cybernetics*, vol. 44, no. 5, pp. 655–668, 2014.

[40] T. Xiong, Y. Bao, and Z. Hu, "Multiple-output support vector regression with a firefly algorithm for interval-valued stock price index forecasting," *Knowledge-Based Systems*, vol. 55, pp. 87–100, 2014.

[41] L. Cao and F. E. H. Tay, "Financial forecasting using support vector machines," *Neural Computing and Applications*, vol. 10, no. 2, pp. 184–192, 2001.

[42] H. Ince and T. B. Trafalis, "Short term forecasting with support vector machines and application to stock price prediction," *International Journal of General Systems*, vol. 37, no. 6, pp. 677–687, 2008.

[43] S. K. Aggarwal, L. M. Saini, and A. Kumar, "Day-ahead price forecasting in ontario electricity market using variable-segmented support vector machine-based model," *Electric Power Components and Systems*, vol. 37, no. 5, pp. 495–516, 2009.

[44] J. H. Zhao, Z. Y. Dong, Z. Xu, and K. P. Wong, "A statistical approach for interval forecasting of the electricity price," *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 267–276, 2008.

[45] W. Xie, L. Yu, S. Xu, and S. Wang, "A new method for crude oil price forecasting based on support vector machines," in *Computational Science—ICCS 2006*, vol. 3994 of *Lecture Notes in Computer Science*, pp. 444–451, 2006.

[46] S. Li and M. Tan, "Tuning SVM parameters by using a hybrid clpsobfgs algorithm," *Neurocomputing*, vol. 73, no. 10–12, pp. 2089–2096, 2010.

[47] Y. Bao, Z. Hu, and T. Xiong, "A PSO and pattern search based memetic algorithm for SVMs parameters optimization," *Neurocomputing*, vol. 117, pp. 98–106, 2013.

[48] X. Tian, G. Gasso, and S. Canu, "A multiple kernel framework for inductive semi-supervised SVM learning," *Neurocomputing*, vol. 90, pp. 46–58, 2012.

[49] U. A. Müller, M. M. Dacorogna, R. D. Davé, R. B. Olsen, O. V. Pictet, and J. E. von Weizsäcker, "Volatilities of different time resolutions—analyzing the dynamics of market components," *Journal of Empirical Finance*, vol. 4, no. 2-3, pp. 213–239, 1997.

[50] T. Lux and M. Marchesi, "Scaling and criticality in a stochastic multi-agent model of a financial market," *Nature*, vol. 397, no. 6719, pp. 498–500, 1999.

[51] W. A. Brock and C. H. Hommes, "A rational route to randomness," *Econometrica*, vol. 65, no. 5, pp. 1059–1095, 1997.

[52] W. A. Brock and A. W. Kleidon, "Periodic market closure and trading volume. A model of intraday bids and asks," *Journal of Economic Dynamics and Control*, vol. 16, no. 3-4, pp. 451–489, 1992.

[53] W. A. Brock and B. D. LeBaron, "A dynamic structural model for stock return volatility and trading volume," *Review of Economics and Statistics*, vol. 78, no. 1, pp. 94–110, 1996.

[54] C. H. Hommes, "Financial markets as nonlinear adaptive evolutionary systems," *Quantitative Finance*, vol. 1, no. 1, pp. 149–167, 2001.

[55] M. M. Dacorogna, R. Genay, U. A. O. Muller, and R. B. V. O. Pictet, *An Introduction to High-Frequency Finance*, Academic Press, San Diego, Calif, USA, 2001.

[56] J. D. Farmer, "Market force, ecology and evolution," *Industrial and Corporate Change*, vol. 11, no. 5, pp. 895–953, 2002.

[57] J. K. Goeree and C. H. Hommes, "Heterogeneous beliefs and the non-linear cobweb model," *Journal of Economic Dynamics and Control*, vol. 24, no. 5–7, pp. 761–798, 2000.

[58] L. Yu, S. Wang, and K. K. Lai, "A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates," *Computers and Operations Research*, vol. 32, no. 10, pp. 2523–2541, 2005.

[59] S. Walczak, "An empirical analysis of data requirements for financial forecasting with neural networks," *Journal of Management Information Systems*, vol. 17, no. 4, pp. 203–222, 2001.

[60] M. H. Pesaran and A. Timmermann, "A simple nonparametric test of predictive performance," *Journal of Business and Economic Statistics*, vol. 10, no. 4, pp. 461–465, 1992.

[61] T. E. Clark and K. D. West, "Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis," *Journal of Econometrics*, vol. 135, no. 1-2, pp. 155–186, 2006.

[62] T. E. Clark and K. D. West, "Approximately normal tests for equal predictive accuracy in nested models," *Journal of Econometrics*, vol. 138, no. 1, pp. 291–311, 2007.

[63] R. A. Meese and K. Rogoff, "Empirical exchange rate models of the seventies. Do they fit out of sample?" *Journal of International Economics*, vol. 14, no. 1-2, pp. 3–24, 1983.