Research Article

Forecasting Crude Oil Price with Multiscale Denoising Ensemble Model

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

[1] *School of Earth Science and Resources, Chang'an University, Xi'an 710054, China*
[2] *School of Economics and Management, Beijing University of Chemical Technology, Beijing 100029, China*
[3] *International Business School, Shaanxi Normal University, Xi'an 710062, China*
[4] *Department of Management Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon Tong, Kowloon, Hong Kong*
[5] *College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China*

Correspondence should be addressed to Kaijian He; paulhekj@gmail.com

Crude oil price becomes more volatile and sensitive to increasingly diversified influencing factors with higher level of deregulations worldwide. Current methodologies are being challenged as they have been constrained by traditional approaches assuming homogeneous time horizons and investment strategies. Approximations they provided over the long term time horizon no longer satisfy the accuracy requirement at shorter term and more microlevels. This paper proposes a novel crude oil price forecasting model based on the wavelet denoising ARMA models ensemble by least square support vector regression with the reduced forecasting matrix dimensions by independent component analysis. The proposed methodology combines the multi resolution analysis and nonlinear ensemble framework. The wavelet denoising based algorithm is introduced to separate and extract the underlying data components with distinct features, corresponding to investors with different investment scales, which are modeled with time series models of different specifications and parameters. Then least square support vector regression is introduced to nonlinearly ensemble results based on different wavelet families to further reduce the estimation biases and improve the forecasting generalizability. Empirical studies show the significant performance improvement when the proposed model is tested against the bench-mark models.

1. Introduction

Given its role as one of the most important industry inputs, the accurate and reliable forecasting of the crude oil price movement has profound impacts throughout the national economy. The forecasting of crude oil price has developed through different stages, including structural approach, time data series approach, artificial intelligence (AI) approach, and other recent developments. Earlier attempts were mainly linear, parametric, structural, and economic theory based. They have been effective in the understanding and forecasting of the crude oil price movement over the medium to long term time horizon with acceptable approximation accuracy and computational efficiency, where the aggregated price behavior is comparably stable and stationary. These include models in the much simplified macroeconomic equilibrium framework to analyze the economic relationships among participants in the crude oil markets and derive analytic equations to model them. Typical models include the supply and demand equilibrium model and more recently the cointegration model, and so forth. For example, Chevillon and Rifflart [1] use cointegration OPEC related relationships in the market and develop a forecasting equation based on it [1]. Ye et al. [2] incorporate further the cumulative excess production capacity as the ratcher variable into forecasting models and show that it results in improved performance [2].

Time series models became popular in recent years. It uses historical information to characterize and forecast movement of the series itself. RW and autoregressive moving average (ARMA) models are two representative time series models [3]. Till now RW and ARMA still serve as important benchmark in crude oil price forecasting because

there is no conclusive evidence that any econometric models consistently beat these two benchmark models [4]. For example, Alquist and Kilian [4] find RW model to be by far the best models available [4]. Meanwhile, another trend is the emergence of AI approach such as the artificial neural network (ANN) and support vector regression (SVR). For example, Yu et al. [5] find that ANN outperforms ARMA model but has room for further improvement using ensemble algorithms. Azadeh et al. [6] propose a flexible oil price forecasting algorithm based on ANN and fuzzy regression (FR), incorporating the economic indicators during the forecasting process. They have achieved improved forecasting performance [6].

Some recent development includes the wavelet analysis and ensemble algorithm. Wavelet analysis possesses the capability to project data into time-scale domain and conducts multiscale analysis [7]. In crude oil forecasting literature, only limited efforts can be observed. Mostly these attempts use wavelet analysis to preprocess data before time series techniques and machine learning techniques are used to analyze the data and make forecasts. Yousefi et al. [8] use wavelet analysis to decompose crude oil price and extend them directly to make forecasts [8]. Meanwhile, ensemble algorithm represents another interesting development. It attempts to incorporate partial information set captured by individual forecasters to produce more accurate forecasts [9]. Traditionally linear based ensemble approaches have achieved only moderate level of accuracy, while the recently emerging nonlinear ensemble approaches based on intelligence techniques such as neural network have achieved much improved performance. For example, Xiong et al. [10] propose a revised hybrid model built upon empirical mode decomposition (EMD) based on the feed-forward neural network (FNN) modeling framework incorporating the slope based method (SBM) and have achieved improved prediction performance with accredited computational load [10]. He et al. [11] propose a wavelet decomposed ensemble model to improve the forecasting accuracy of crude oil price with deeper understanding of the market microstructure [11].

This paper explores the potential of multiscale signal processing techniques, more specifically the denoising technique, in forecasting price movement in crude oil markets. The main research problem addressed is the empirical evidence of multiscale heterogeneous data characteristics distinguishable by sizes and the incorporation of this stylized fact in the construction of forecasting algorithm. More specifically, the wavelet denoising algorithms based on diverse wavelet families are used to construct the forecasting matrixes, whose members are individual forecasts based on different time series specifications capturing different data features. Wavelet denoising algorithm could serve as the potential source separation tool to analyze its underlying structure. Experiment results confirmed that the model specification and parameters vary with different choices of parameters for the wavelet denoising algorithm. The forecasting accuracy also improves when noises are filtered out during the modeling process.

The key findings and contributions of this paper are threefold. Theoretically this paper proposes the HMH theory to model the crude oil market and provides a more general framework that reveals the hidden data components based on wavelet analysis with different parameters and analyze them incorporating their heterogeneous specifications and parameters. Technically HMH has received less attention in extant literature, due to lack of methodology to test and model it. Wavelet denoising technique has been introduced as an important paradigm shift that could provide the empirical evidence for HMH and help gain further insights. Compared to other blind signal decomposition approaches such as morphological component analysis (MCA) and empirical model decomposition (EMD) methods. The wavelet analysis uses the theoretically well-defined wavelets to capture the matching distinguished data patterns. Results are more stable and offer better interpretations. Practically work in this paper distinguishes from the existing literature by introducing the nonlinear ensemble algorithm to reduce instability and biases introduced by different wavelet families. The performance of wavelet based algorithms is sensitive to the parameters set, as revealed in empirical research. This is a less addressed issue in the literature and could cast doubts on results obtained in the empirical research literature. The nonlinear ensemble algorithm could serve as a promising alternative to reduce instability and biases introduced by different wavelet families. This has significant practical implication that the proposed algorithm adapts better to the practical data and achieves better stability in the estimates.

The rest of the paper is organized as follows. In Section 2, we briefly review the relevant literature of researches on wavelet denoising algorithm as well as ICA-LSSVR based nonlinear ensemble methodology. In Section 3, we propose the integrated methodology WDN-ICA-LSSVR. Major findings and performance evaluations results of empirical studies conducted are reported in Section 4. Section 5 summarizes and concludes.

2. Relevant Theories

2.1. Wavelet Denoising Algorithm. Denoising algorithms have been one of the most important research topics in the engineering field, although they have not received much attention in the field of economics and finance. The critical issue in the denoising algorithm research is to set the right boundary and remove the noises while preserving major data features. There are many approaches developed in the literature. These include moving average filter, exponential smoothing filter, linear Fourier smoothing, simple nonlinear noise reduction and nonlinear wavelet shrinkage, and so forth [12]. During the denoising process, some data features are changed to a varied degree depending on the filters used. For example, simple methods including moving average filter and exponential smoothing filter are known to modify the statistical property of data significantly [13]. There are also some other nonlinear denoising techniques with mixed results in the literature. Yahya Bey [14] and Bey [15] propose both modified and constant frequency extent denoising algorithms and they have shown superior performance compared to the wavelet denoising method in the empirical studies [14, 15]. Verma and

Ganguli [13] propose the nonlinear rational filter with median preprocessor to improve the noise reduction ratio in gas turbin health monitoring applications [13]. Wavelet shrinkage could provide more optimized and localized noise reduction processes. Boto-Giralda et al. [16] use the stationary wavelet based denoising methods to improve the performance of traffic volume prediction models in intelligent transportation system [16]. Gao et al. [17] propose adaptive denoising algorithms and contend it to be superior to wavelet based approaches when applied to analysis of electroencephalogram (EEG) signals contaminated with noises [17]. Kwon et al. [18] point out the problems with existing denoising techniques as assuming homogeneous error structure and they propose wavelet denoising method incorporating a variance change point detection thresholding method to deal with it in protein mass spectroscopy applications [18]. Lotric and Dobnikar [19] and Lotrič [20] integrate the neural network with the wavelet denoising method to optimize the denoising parameters dynamically and found the performance improvement in prediction accuracy [19, 20]. It is interesting to see that wavelet based denoising methods have become the predominant approaches.

The unique feature of wavelet analysis is its localization over both time and scale during the analysis. This is in contrast to the traditional spectrum analysis tool that signifies the patterns at different frequencies at the cost of ignoring details at time scales. This capability stems from the wavelets functions used, which have high energy concentration in a short interval 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 [21]. The forward and inverse operation of wavelet analysis form the multiresolution analysis, which provides unique perspectives into the data structure. As financial data usually exhibit nonstationary time varying characteristics, wavelet analysis is appealing when it comes to financial data analysis. Wavelets are continuous functions in $L_2(\mathbb{R})$ that have vanishing moments and fluctuating characteristics within a certain time period [22]. Wavelets are characterized by their various features including vanishing moments, compact support, regularity, symmetry, and time frequency windows [22]. The typical wavelets used include Haar, Daubechies, Coiflets, discrete Meyers, and Biorthogonal and reverse Biorthogonal [22]. Interested readers are referred to Percival and Walden [22] for more details.

Utilizing the wavelets and their appealing time scale localization characteristics, the original data are projected into the multiscale domain using wavelet transform as in the following:

$$W(u, s) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) dt, \quad (1)$$

where $\psi((t-u)/s)$ refers to the wavelet families translated by location parameter u, which corresponds to time scale localization, and dilated by scale parameter s, which corresponds to the frequency scale localization.

The wavelet transform can be reversed to perfectly reconstruct the original data by inverse wavelet synthesis as in the following:

$$x(t) = \frac{1}{C_\psi} \int_0^\infty \int_{-\infty}^\infty W(u, s) \psi_{u,s}(t) \, du \frac{ds}{s^2}. \quad (2)$$

The wavelet denoising technique utilizes the wavelet transform and synthesis during the data denoising process. It firstly projects the original data series into different scales using wavelet transform. Thus multiresolution analysis (MRA) can be defined as analyzing the time and frequency characteristics of the data, both denoised and noise part, with finer details revealing patterns at more microscales. Then the separation between denoised and noise data components is achieved by applying the threshold chosen specifically at different scales to either suppress or shrink the wavelet coefficients. The processed wavelet coefficients are reconstructed into the unified data series using wavelet synthesis.

Threshold selection rules and shrinkage algorithms are critical to effective extraction of features during the denoising process. Typical threshold selection strategies include universal, minimaxi, and Steins' unbiased risk estimate threshold. The shrinkage strategies include the hard and soft shrinkage algorithm [23]. The differences among different threshold selection rules lie in the trade-off set between smoothness and accuracy to obtain denoised data. For example, the maximal level of reduction of noises is achieved with universal threshold selection rule at the cost of lower goodness-of-fit and risk of potential disruption of valuable information in the original data. The minimaxi threshold selection rule, on the other hand, attempts to achieve the best fit approximation to the original data by minimizing the function fitness criteria such as mean squared error (MSE), at the cost of lower function smoothness. The shrinkage strategies include the hard and soft shrinkage algorithms, both widely used in mainstream literature over the years [23, 24]. The hard shrinkage rule is the high pass filter which suppresses coefficients in the wavelet transform below the chosen threshold values and leaves the remaining coefficients intact. The soft shrinkage rules suppress coefficients in the wavelet transform below the set threshold value and subtract the threshold value from the remaining wavelet coefficients

2.2. Nonlinear Ensemble Algorithm. Since the seminal work by Bates and Granger [9], empirical researches have shown that the ensemble algorithm can improve the generalizability of the model by incorporating the partial information set [9]. Given the forecasting set pair $(x_i, f_i(x))$ $(i = 1, 2, \ldots, n)$, where x_i is the input and $f_i(x)$ is the output, the goal of the ensemble model as in (3) is to find the hidden pattern that has the minimal generalization error and generalizes well out-of-sample:

$$\text{Minimizes} \quad \varepsilon = \widehat{f(x)} - d(x) = \sum_{i=1}^{n} \widehat{f_i(x)} - d(x), \quad (3)$$

where $\widehat{f(x)}$ is the ensemble forecasts and equal the combination of individual forecasts $\widehat{f_i(x)}$ as $\sum_{i=1}^{n} w_i \widehat{f_i(x)}$. $d(x)$ is

the expected ensemble forecasts. The performance improvement from ensemble forecasts is determined by two factors: the forecasting accuracy of individual models and the level of independence of individual models. The greater the level of independence and heterogeneity of individual models, the smaller the level of correlations among individual forecasters, and thus the higher the level of generalizability of models estimated. These two issues are taken into consideration during the modeling process in this paper. On one hand, the forecasting accuracy issue is addressed by multiscale denoising based time series methodology. On the other hand, independent component analysis has been proposed in this paper to reduce the dimensionality of forecasting matrix and improve the independence of ensemble members. Traditionally principal components analysis (PCA) has been proposed to achieve the uncorrelated ensemble members. But they are independent only in the Gaussian framework. For example, forecasts can have zero covariance and yet be interdependent at higher moments. In the meantime, ensemble members transformed using ICA are statistically independent, that is, not just zero covariance, but independence at higher moments. ICA, also known as blind source separation (BSS), attempts to find the optimal linear projection of the original data to reduce the problem dimensionality and noise level. It recovers latent variables from the original mixed data without prior information except the statistical properties, as in the following:

$$x_t^T = As_t^T, \tag{4}$$

where x is the $[m \times t]$ observation matrix, s is the $[n \times t]$ mutually independent source matrix, and A is the real mixing coefficients $a_{i,j}$ $(i = 1, \ldots, m; \ j = 1, \ldots, n)$.

ICA attempts to estimate A, whose inverse W recovers the original source s from observations x as $s = A^{-1}x = Wx$.

Among different approaches developed over the years, FastICA based on the fixed point algorithm is the most widely used and computationally efficient one [25]. Independent components can be estimated either in parallel or one after another following the Gram-Schmidt orthogonal procedure. During the optimization process, higher order statistics rather than the second moment are utilized as further guidance on the degree of statistical independence of components calculated.

Least square support vector regression (LSSVR) is the emerging nonlinear ensemble algorithm. Compared to traditional neural network based approach, its solution is more stable and global optimum. 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 the fitting accuracy and model generalizability. Thus it would alleviate the overfitting and local minima issue with the traditional supervised learning algorithm such as neural network models, which are based on the empirical risk minimization principle. Interested readers are referred to [26] for details.

The basic notion of LSSVR is as follows. Firstly using kernel functions that satisfy the Mercer condition, the nonlinear data are mapped into higher dimensions using kernel tricks

and a typical regression problem is formulated accordingly as in the following:

$$y = f(x) = \text{sign}\left[\omega^T \phi(x) + b, \phi : R^n \longrightarrow F, \omega \in F\right], \tag{5}$$

where $\phi(x)$ is the mapping function that maps the variable x in the nonlinear input space into the variables in the linear feature space. ω and b are coefficients.

A slack variable e_i is introduced to model the estimation error, thus the LSSVR is formulated as in the following:

$$\min_{w,b,e} \ J_i(w,e) = \frac{1}{2}\omega^T \omega + \lambda \frac{1}{2}\sum_{i=1}^{l} e_i^2 \tag{6}$$

$$\text{s.t.} \quad y_i = \omega^T \phi(x_i) + b + e_i, \quad i = 1, \ldots, l,$$

where $e_i = (e_1; e_2; \ldots; e_l) \in R^l$ and λ is the penalty parameters. It serves to control for two objectives, that is, the minimization of estimation error and the function smoothness, during the optimization process.

The Lagrangian function for the dual problem is formulated as in the following:

$$L(w,b,e;a) = j(w,e) + \sum_{i=1}^{l} a_i\left(\omega^T \phi(x_i) + b + e_i - y_i\right), \tag{7}$$

where $a = (a_1; a_2; \ldots; a_l)$ is the Lagrange multipliers. Optimality is achieved with solving the linear system obtained by differentiating L in (7) with the variable $\omega, b, \varepsilon, \alpha$, as in the following:

$$\frac{\partial L}{\partial \omega} = 0 \longrightarrow \omega = \sum_{i=1}^{l} \alpha_i \phi(x_i)$$

$$\frac{\partial L}{\partial b} = 0 \longrightarrow \omega = \sum_{i=1}^{l} \alpha_i = 0$$

$$\frac{\partial L}{\partial \epsilon_i} = 0 \longrightarrow \omega = \sum_{i=1}^{l} = \lambda \epsilon_i, \quad i = 1, \ldots, N \tag{8}$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \longrightarrow \omega = \omega^T \phi(x_i) + b + \epsilon_i - y_i = 0,$$

$$i = 1, \ldots, N.$$

LSSVR, as the extension and improvement of SVR with more efficiency in the large scale problem, has shown performance improvement similar to the SVR algorithm in many time series forecasting problems. In crude oil market, Bao et al. [27] have proposed a hybrid model combining wavelet analysis and LSSVR in a two-stage process and have demonstrated its superior performance in medium to long term empirical studies [27]. However, there have been few attempts for LSSVR application in the short term crude oil forecasting literature. Despite the positive results accumulated recently in the literature, both LSSVR and SVR suffer issues similar to the neural network model. For example, it is often difficult to provide economic justification for results obtained due to the black box nature of its training process. Its performance is sensitive to the parameters, which are often arbitrarily chosen.

3. A Novel WDN-ICA-LSSVR Algorithm

Recent progress in the mixture modelling literature provides further evidence of heterogeneous data components of different distributions and specifications in the financial markets. For example, work by Thavaneswaran et al. [28] points out that mixture distributions are better in capturing the heteroscedasticity in the financial data. They then derive the Markov models under mixture distribution [28]. Shahbaba [29] points out that ignoring the latent data structure and mixture distributions would lead to wrong inference and they propose the mixture models to effectively capture the underlying DGP [29]. Empirical researches show that the crude oil market has heterogeneous market structure with underlying components corresponding to both normal market conditions and transient events.

Thus this paper proposes the HMH theory as the theoretical basis for modeling the crude oil market. In HMH, the market is assumed to consist of heterogeneous agents with heterogeneous investment strategies and investment time horizons [30]. Compared to the homogeneous reaction to news shocks assumed in efficient market hypothesis (EMH), HMH assumes that agents or investors adopt diversely different investment strategies based on their defining characteristics. Thus their investment time horizon and dealing frequency are intrinsically different; some tend to have lower dealing frequency focusing on long time horizon, while others tend to have high dealing frequency with short time horizon [31]. Although theoretically sound, in practice only limited studies have been conducted to explore the methodology to systematically model the heterogeneous market structure in HMH. When investigating the crude oil markets, there are mainly two major issues addressed in this paper.

Firstly to recognize the heterogeneous market structure, theoretically major influencing factor or criteria need to be chosen to disentangle the hidden market structure. As empirical studies suggest, the microstructure of the crude oil market is very complex, consisting of investors with different investment strategies targeting at different investment time horizons [32]. The market prices reflect the combined influences from the existing traders with different time horizons. This may include fundamentalist trader with long term strategies, short term trader aiming mainly at limited time horizon (e.g., daily), and intraday traders dealing with high frequency shocks throughout the day. The statistical property of the market process is dynamic and constantly changing due to its heterogeneity nature. Therefore, the high frequency shock could only influence part of but not all traders in the market, while the low frequency shock has far deeper impact on the whole market and market makers. The line between the short lived nature of high frequency shock and long term nature of low frequency shock is crucial in the sense that all investors and managers are essentially utility maximizer and should only be concerned with decisions and associated consequences within their own target investment horizon. Among different features that can be used to distinguish the market structure, this paper proposes the investment scales to disentangle the hidden market structure. Thus based on different investment scales, we assume that participants in the crude oil market can be categorized uniquely based on different investment scales into the fundamental investors (FI) and the noise traders (NT) as in the following:

$$r = r_{\mathrm{FI}} + r_{\mathrm{NT}}. \tag{9}$$

Implications from this assumption are that trading activities of these two types of traders characterized by their investment scales would exhibit distinctively different characteristics. The behavior of noise traders restricted by smaller band of frequency is closely related to temporal, stochastic, and nonlinear patterns in nature, usually resulting from highly speculative behavioral patterns and easily arising operational as well as transaction recording errors. The behavior of fundamental investors restricted by larger band of frequency are closely related to more stable, consistent, and long term patterns, which have been the subject for the extensive studies over years. They commonly exhibit characteristics such as autocorrelation, heteroscedasticity, mean reverting, and long memory, described by the mainstream econometric models.

Secondly techniques that can effectively and efficiently model the heterogeneous market structure need to be identified and incorporated to model accurately different market price movements characterized by different investment scales. In this paper, we introduce the wavelet denoising algorithms since they provide rough approximation to the crude oil market characterized by two major forces of influences. The main difficulty to separate data from noises with higher level of accuracy as the separation accuracy affects critically the generalizability of the models fitted. Proper removal of noises could result in more well behaved data. The inappropriate separation twists the original data features. Traditional data denoising techniques include spectrum analysis methods such as simple averages, Kalman filtering, and Fourier transforms. However, these methods do not receive significant attention in the mainstream empirical researches in the literature. Firstly, with the currently prevalent imperfect extraction technique in the literature, the data may be distorted during the denoising process with different assumptions imposed and different parameters used, which reduce forecasting accuracy. Secondly, they are mostly based on spectrum and frequency scale domain while ignoring the multiscale heterogeneous structure in the data and their underlying components, corresponding to different individual investment strategies and time horizon. The wavelet based denoising algorithm serves as the alternative modeling tool to extract data features of different characteristics, which is a more refined methodology. Meanwhile, as the true underlying DGP is unknown, each individual wavelet denoising model attempts to extract restricted information from the data with its unique assumptions or perspectives. These interpretations of data only offer partial insights and are never complete unless unified. Since there is no consensus as to the suitability of parameters to use during the wavelet based denoising process with particular data set, to reduce the biases introduced and further stabilize the estimation results, this paper proposes the LSSVR based nonlinear ensemble algorithm to ensemble partial information set captured by models with different parameters to produce more stabilized

and accurate results. Besides, ICA is also used during the ensemble process to reduce the noises and reduce the computational complexity with the high dimension of forecast matrix consisting of forecasts based on different parameters.

Numerical procedure of the WDN-ICA-LSSVR algorithm involves the following key steps.

(1) Firstly, the wavelet based denoising algorithm is used to separate data from noises using particular wavelet families w, as in the following:

$$r_t = r_{\text{denoised},t,w} + r_{\text{noises},t,w}. \tag{10}$$

By decomposing data into the multiscale domain, data and noises are separated based on their different characteristics across scales with noises smaller in scales. Thus more subtle distinctions between data and noises can be set.

(2) The conditional mean for the denoised data is modeled by ARMA processes as in the following:

$$\widehat{r}_t = \delta_t + \sum_{i=1}^{m} \phi_i r_{t-i} + \sum_{j=1}^{n} \theta_j \varepsilon_{t-j} + \varepsilon_t, \tag{11}$$

where $\widehat{r}_t$ is the conditional mean at time t, r_{t-i} is the lag m returns with parameter ϕ_i, and ε_{t-j} is the lag n residuals in the previous period with parameter θ_j.

(3) The optimal specification for the ARMA model (i.e., model orders) is determined following AIC and BIC minimization principal as in the following:

$$\text{Min} \left\{ \text{AIC}(l) = \frac{-2}{T} \ln(l) + \frac{2l}{T}, \right.$$
$$\left. \text{BIC}(l) = \frac{-2}{T} \ln(l) + \frac{l}{T} \ln T \right\}. \tag{12}$$

(4) With high dimensional forecasts matrix consisting of forecasts based on different wavelet families, the ICA is used to remove noises and reduce dimensions. The original forecast matrix is transformed as in the following:

$$\widehat{\mu}_t^S = A^{-1} \widehat{\mu}_t^G, \tag{13}$$

where S refers to the projected principal components using ICA. G refers to the set of forecasts based on different wavelet parameters. A is the real mixing coefficients.

(5) The nonlinear ensemble algorithm is used to aggregate individual conditional mean forecasts in the ICA transformed forecast matrix as in the following:

$$\widehat{\mu}_t^{\text{WDN-ICA-LSSVR}} = \phi_{i \in S}\left(\widehat{\mu}_t^i\right). \tag{14}$$

The LSSVR algorithm is used during the nonlinear ensemble process to calculate the optimal weights with improved robustness and generalizability for the model.

TABLE 1: Descriptive statistics and statistical tests for WTI market.

Data	r	$r_{\text{denoised Data, Coiflet(5)}}$	$r_{\text{noises, Coiflet(5)}}$
Mean	0.0011	0.0011	0
SD	0.0231	0.0129	0.0192
Skewness	−0.4219	−1.2468	0.0153
Kurtosis	4.7676	16.9730	2.9808
JB test	0	0	0.5000
BDS test	0.0029	0	0.4178

4. Empirical Studies

We conducted the empirical studies using daily observations in the West Taxes Intermediate (WTI) crude oil market. The data source is the Energy Information Administration, Department of Energy, US. The data range covers the time period January 2, 2002, to March 20, 2009. This includes 1814 daily observations. The data set is divided into three subdata sets, that is, the training set for wavelet denoising ARMA model (60%), the training set for ICA-LSSVR based nonlinear ensemble model (24%), and the test set for the out-of-sample test to evaluate the performance of different models (16%). Since there is no consensus on the econometric criteria for sample segmentation, the separation ratios are determined following convention in the machine learning literature, where at least 60% of the dataset is usually reserved as the training set so that the size of the remaining data set for the out-of-sample test is sufficiently large set and the results are statistically valid [33, 34].

The one day ahead forecast is performed using rolling-window method. Since Autocorrelation Function (AF) and Partial autocorrelation function (PAF) analyses indicate that the original data include trend factors, it is log differenced at the first order as $r_t = \ln(P_t/P_{t_1})$ 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, and so forth. The holding period is assumed to be 1 day. For each experiment, a portfolio of one asset position worth 1 USD is assumed.

The generalizability of different models in out-of-sample forecast tests is evaluated using the forecasting error measurements and statistical tests. This includes mean square error (MSE), Pesaran-Timmerman directional test (PT), and Clark West test of equal predictive accuracy [35, 36].

Experiment results in Table 1 show some stylized facts. The market exhibits moderate level of fluctuations, as suggested by the moderate level of volatility level. The distribution of the market returns is fat-tail and leptokurtic, as suggested by significant skewness and kurtosis levels. There is also high level of market risk exposure due to extreme events in the market, as reflected in the significant kurtosis level. The market returns also deviate from the normal distribution and exhibit nonlinear dynamics; this is further confirmed by the rejection of both Jarque-Bera test of normality and BDS test of independence at 95% confidence levels. Therefore, both denoised data component and noises data are extracted from the original data series for further analysis. From

experiment results in Table 1, denoised data series exhibit more leptokurtic and fat tailed behavior. The rejection of both Jarque-Bera and BDS test suggests that the data series are characterized by nonlinear dynamics. Thus it is categorized as reflecting the underlying permanent factors. Meanwhile, noises data series are more normally distributed than the original data based on the descriptive statistics, although the Jarque-Bera test of normality is rejected. The BDS test is accepted at the moderate level of statistical significance. Therefore, it is categorized as reflecting the transient factors. However, the moderate significant level of BDS test statistics also suggests that there may be some linear or nonlinear data patterns involved. These analysis suggest that the model specifications for both denoised and noises data are different and the original data are mixture of different underlying DGP.

The wavelet denoising ARMA model is applied to the testing data to investigate the effects of different parameters on the model performance. Different combinations of parameters choices are pooled into the parameters pool, including three threshold selection rules (i.e., Stein's unbiased risk estimate (SURE), MiniMaxi, and Universal), two shrinkage rules (i.e., hard and soft shrinkage rules), 6 decomposition levels up to scale 6, and 7 wavelet families, which includes Haar, Daubechies, Coiflets, Symlets, discrete Meyers, and Biorthogonal and reverse Biorthogonal. The model orders for ARMA (r, m) processes are determined following information criteria (IC) such as Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) minimization. Decomposition level is chosen to be 5, the rolling window is set to 512, the lag order is set to 2, the threshold selection rule chosen is MiniMaxi, and the shrinkage algorithm chosen is hard. Several parameters need to be determined for LSSVR model. The kernel chosen is radial basis function (RBF). Grid search method is used to determine the penalizing parameter λ and parameter σ was chosen for RBF. λ is 0.0518 and σ is 2.8651.

Experiment results in Table 2 list predictive accuracy measured by the MSE for the original data, denoised and the noises data separately in the WTI markets.

Experiment results in Table 2 further confirm that the forecasting accuracy of wavelet denoising ARMA model is sensitive to the wavelet parameters used to denoise the original data. Some of the wavelet parameters can improve the forecasting performance to the level that beats the traditional benchmark models significantly. In the meantime, the forecasting accuracy decreases as a result of improper denoising model used. This result supports the proposed HMH in this paper that the price movement in the markets is joint influences from heterogeneous agent behaviors, including their different beliefs, strategies, and risk appetites [37]. The perception into the separation and modeling of the distinct behaviors of different agents based on different wavelets may be redundant.

Since there is a lack of consensus on the appropriate selection of denoising algorithm and specifications, one approach common in the literature is to follow the trial and error method in the identification process, which is arbitrary and computationally expensive. To tackle this issue, this paper introduces the LSSVR based ensemble algorithm

Table 2: In-sample forecasting accuracy comparison of wavelet models for WTI market.

Models	$\mathrm{MSE}_{\times 10^{-3}}$	$\mathrm{MSE}_{\times 10^{-3}, \mathrm{data}}$	$\mathrm{MSE}_{\times 10^{-3}, \mathrm{noises}}$
Haar	1.4273	1.4189	1.3880
Db(2)	1.3988	1.3890	1.3890
Db(3)	1.4094	1.4140	1.3791*
Db(4)	1.4152	1.3886	1.4083
Db(5)	1.4141	1.3934	1.3987
Dmey	1.3954	1.4166	1.3573*
Coiflet(1)	1.4263	1.4228	1.3844*
Coiflet(2)	1.3907	1.3865	1.3849
Coiflet(3)	1.7108	1.4823	1.6104
Coiflet(4)	1.4031	1.3982	1.3863
Coiflet(5)	1.3570*	1.3677*	1.3688*
Symlet(1)	1.4273	1.4189	1.3880
Symlet(2)	1.4318	1.4233	1.3884
Symlet(3)	1.4094	1.4140	1.3791*
Symlet(4)	1.4037	1.4106	1.3744*
Symlet(5)	1.4510	1.4305	1.4053
Bior(1, 1)	1.4273	1.4189	1.3880
Bior(2, 2)	1.4130	1.4105	1.3863
Bior(3, 1)	1.7564	2.3318	1.8394
Bior(3, 9)	1.4373	1.4133	1.3982
Rbior(1, 1)	1.4273	1.4189	1.3880
Rbior(2, 2)	1.4647	1.4141	1.4294
Rbior(3, 1)	1.4359	1.4248	1.4521
Rbior(3, 9)	1.3790*	1.3816*	1.3789*

*The test case with superior performance than that of Random Walk model.

to ensemble forecasts based on different wavelet parameters. Based on the constructed individual forecast matrixes based on different wavelet families, ICA is used to reduce the matrix dimensions to produce the ensemble matrix and the LSSVR is used to ensemble the individual matrix to produce the optimal result. The dimensionality of the forecasting matrices for both denoised data and noises data is reduced to 1. The nonlinearity function used is the cubic function. The parameters for WDN-ICA-LSSVR algorithm are determined using grid search algorithm: for WTI market, λ is 0.3353 and σ is 32.6345.

Experiment results in Table 4 show that the proposed WDN-ICA-LSSVR outperforms the benchmark RW and ARMA model, in terms of MSE and directional accuracy measured by PT directional test statistics.

Experiment results in both Tables 3 and 4 show the superior performance of the proposed multiple wavelet denoising based ensemble against the traditional benchmark RW and ARMA models, at the statistical significant level. The forecasting accuracy is higher in terms of lower MSE value, while the directional forecasting accuracy is better in terms of higher ratio with significant P value at 95% confidence value. Meanwhile, the ensemble forecasts estimated are more stable than individual forecasts as the partial information set from individual forecasters is integrated to produce the optimal and stable forecasts with the proposed approach.

Table 3: Out-of-sample forecasting accuracy comparison of different models.

Models	$MSE_{\times 10^{-3}}$	CW$_{ARMA}$ P value	CW$_{RW}$ P value
ARMA	14.359	N/A	0.6197
RW	13.846	0.6197	N/A
LSSVR	13.960	0.0136	0.7648
WDN-ICA-LSSVR	13.765	0.0010	0.1417

Table 4: Out-of-sample directional predictive accuracy comparison of different models.

Models	Dstat$_\%$	PT
ARMA	52.98	0.0896
LSSVR	52.52	0.0988
WDN-ICA-LSSVR	55.05	0.0095

The proposed algorithm helps produce optimal estimate by combining the partial information set in the constructed forecasting matrix.

Theoretically experiment results in this paper show that the HMH represents important and more realistic alternative theories for the market data structure. Technically experiment results confirm that HMH based modeling would result in the improved forecasting performance when the underlying data components are modeled with the noise removed in the multiscale domain. This also implies that the appropriate optimization of the weights assigned to different forecasts based on different weights would combine the partial information set they capture and lead to more stable and improved forecasting results.

5. Conclusions

This paper proposes a novel wavelet denoising crude oil price forecasting algorithm based on both wavelet based denoising algorithm and ICA-LSSVR based nonlinear ensemble framework. Work in this paper provides further evidence in support of HMH against EMH and predictability of price movement in the crude oil market. The market structure is viewed and analyzed in the emerging HMH perspective, relaxing the assumptions of traditional EMH. Under the framework of HMH, this paper proposes that the underlying heterogeneous market structure can be distinguished by different investment scales; that is, the crude oil price behaviors are affected by noises and main trends, which have different characteristics. Thus the separation of noises and data need to be conducted in a multiscale manner to recover the useful data for further modeling by ARMA model. The more accurate separation of data from noises leads to better behaved data and higher level of model generalizability. This paper also argues that the performance of traditional approaches could be further improved by the appropriate removal of noises from data.

Meanwhile, this paper also finds that the behavior of the denoised data is sensitive to the wavelet parameters chosen and exhibits completely different behavior, which corresponds to different model specifications and performances. Since there is no consensus on the criteria to choose the appropriate parameters in the literature, to reduce the estimation biases and improve the robustness, this paper introduces the LSSVR based nonlinear ensemble algorithm and ICA into the modeling process. The ICA effectively reduces the forecast matrix dimension and recovers important components that contribute most of the data features. The LSSVR based ensemble algorithm achieves higher forecasting accuracy and robustness by achieving faster convergence speed to the global optimal solution in the problem domain. The proposed model shows competent performance and improved robustness.

Conflict of Interests

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

Acknowledgments

References

[1] G. Chevillon and C. Rifflart, "Physical market determinants of the price of crude oil and the market premium," *Energy Economics*, vol. 31, no. 4, pp. 537–549, 2009.

[2] M. Ye, J. Zyren, C. J. Blumberg, and J. Shore, "A short-run crude oil price forecast model with ratchet effect," *Atlantic Economic Journal*, vol. 37, no. 1, pp. 37–50, 2009.

[3] C. Brooks, *Introductory Econometrics for Finance*, Cambridge University Press, Cambridge, UK, 2nd edition, 2008.

[4] R. Alquist and L. Kilian, "What do we learn from the price of crude oil futures?" *Journal of Applied Econometrics*, vol. 25, no. 4, pp. 539–573, 2010.

[5] 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.

[6] A. Azadeh, M. Moghaddam, M. Khakzad, and V. Ebrahimipour, "A flexible neural network-fuzzy mathematical programming algorithm for improvement of oil price estimation and forecasting," *Computers and Industrial Engineering*, vol. 62, no. 2, pp. 421–430, 2012.

[7] R. Gençay, F. Selçuk, and B. Whitcher, "Multiscale systematic risk," *Journal of International Money and Finance*, vol. 24, no. 1, pp. 55–70, 2005.

[8] 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.

[9] J. M. Bates and C. W. J. Granger, "The combination of forecasts," *Journal of the Operational Research Society*, vol. 20, no. 4, pp. 451–468, 1969.

[10] T. Xiong, Y. Bao, and Z. Hu, "Beyond one-step-ahead forecasting: evaluation of alternative multi-step-ahead forecasting models for crude oil prices," *Energy Economics*, vol. 40, pp. 405–415, 2013.

[11] K. He, L. Yu, and K. K. Lai, "Crude oil price analysis and forecasting using wavelet decomposed ensemble model," *Energy*, vol. 46, no. 1, pp. 564–5574, 2012.

[12] D. Lorenz and T. Kohler, *A Comparison of Denoising Methods for One Dimensional Time Series*, Zentrum fr Technomathematik, 2005.

[13] R. Verma and R. Ganguli, "Denoising jet engine gas path measurements using nonlinear filters," *IEEE/ASME Transactions on Mechatronics*, vol. 10, no. 4, pp. 461–464, 2005.

[14] N. Yahya Bey, "Extraction of signals buried in noise Part II: experimental results," *Signal Processing*, vol. 86, no. 10, pp. 2994–3011, 2006.

[15] N. Y. Bey, "Extraction of signals buried in noise. Part I: fundamentals," *Signal Processing*, vol. 86, no. 9, pp. 2464–2478, 2006.

[16] D. Boto-Giralda, F. J. Díaz-Pernas, D. González-Ortega et al., "Wavelet-based denoising for traffic volume time series forecasting with self-organizing neural networks," *Computer-Aided Civil and Infrastructure Engineering*, vol. 25, no. 7, pp. 530–545, 2010.

[17] J. Gao, H. Sultan, J. Hu, and W.-W. Tung, "Denoising nonlinear time series by adaptive filtering and wavelet shrinkage: a comparison," *IEEE Signal Processing Letters*, vol. 17, no. 3, pp. 237–240, 2010.

[18] D. Kwon, M. Vannucci, J. S. Joon, J. Jeong, and R. M. Pfeiffer, "A novel wavelet-based thresholding method for the pre-processing of mass spectrometry data that accounts for heterogeneous noise," *Proteomics*, vol. 8, no. 15, pp. 3019–3029, 2008.

[19] U. Lotric and A. Dobnikar, "Predicting time series using neural networks with wavelet-based denoising layers," *Neural Computing and Applications*, vol. 14, no. 1, pp. 11–17, 2005.

[20] U. Lotrič, "Wavelet based denoising integrated into multilayered perceptron," *Neurocomputing*, vol. 62, no. 1–4, pp. 179–196, 2004.

[21] R. Gençay, F. Selçuk, and B. Whitcher, *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*, Academic Press, 2002.

[22] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*, vol. 4 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, 2000.

[23] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, 1995.

[24] D. L. Donoho and I. M. Johnstone, "Ideal denoising in an orthonormal basis chosen from a library of bases," *Comptes Rendus de l'Académie des Sciences I: Mathématique*, vol. 319, no. 12, pp. 1317–1322, 1994.

[25] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[26] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Statistics for Engineering and Information Science, Springer, New York, NY, USA, 2nd edition, 2000.

[27] Y. Bao, X. Zhang, L. Yu, and S. Wang, "Crude oil price prediction based on multi-scale decomposition," in *Proceedings of the Computational Science (Iccs '07)*, Y. Shi, J. Dongarra, and P. Sloot, Eds., vol. 4489 of *Lecture Notes in Computer Science*, pp. 933–936, Springer, Berlin, Germany, 2007.

[28] A. Thavaneswaran, S. S. Appadoo, and J. Singh, "Random coefficient mixture (RCM) GARCH models," *Mathematical and Computer Modelling*, vol. 42, no. 5-6, pp. 519–532, 2005.

[29] B. Shahbaba, "Discovering hidden structures using mixture models: application to nonlinear time series processes," *Studies in Nonlinear Dynamics & Econometrics*, vol. 13, no. 2, 2009.

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

[31] A. Weron and R. Weron, "Fractal market hypothesis and two power-laws," *Chaos, Solitons and Fractals*, vol. 11, no. 1, pp. 289–296, 2000.

[32] L. Sarno and G. Valente, "Empirical exchange rate models and currency risk: some evidence from density forecasts," *Journal of International Money and Finance*, vol. 24, no. 2, pp. 363–385, 2005.

[33] P. R. Hansen and A. Timmermann, "Choice of sample split in out-of-sample forecast evaluation," Economics Working Papers ECO2012/10, European University Institute, 2012.

[34] H. F. Zou, G. P. Xia, F. T. Yang, and H. Y. Wang, "An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting," *Neurocomputing*, vol. 70, no. 16–18, pp. 2913–2923, 2007.

[35] 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.

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

[37] C. Hommes, "The heterogeneous expectations hypothesis: some evidence from the lab," *Journal of Economic Dynamics & Control*, vol. 35, no. 1, pp. 1–24, 2011.