

RESEARCH

Open Access



A hybrid Bayesian-network proposition for forecasting the crude oil price

Babak Fazelabdolabadi

Correspondence: fazelb@ripi.ir
Center for Exploration and
Production Studies and Research,
Research Institute of Petroleum
Industry (RIPI), 14665-1998, Tehran,
Iran

Abstract

This paper proposes a hybrid Bayesian Network (BN) method for short-term forecasting of crude oil prices. The method performed is a hybrid, based on both the aspects of classification of influencing factors as well as the regression of the out-of-sample values. For the sake of performance comparison, several other hybrid methods have also been devised using the methods of Markov Chain Monte Carlo (MCMC), Random Forest (RF), Support Vector Machine (SVM), neural networks (NNET) and generalized autoregressive conditional heteroskedasticity (GARCH). The hybrid methodology is primarily reliant upon constructing the crude oil price forecast from the summation of its Intrinsic Mode Functions (IMF) and its residue, extracted by an Empirical Mode Decomposition (EMD) of the original crude price signal. The Volatility Index (VIX) as well as the Implied Oil Volatility Index (OVX) has been considered among the influencing parameters of the crude price forecast. The final set of influencing parameters were selected as the whole set of significant contributors detected by the methods of Bayesian Network, Quantile Regression with Lasso penalty (QRL), Bayesian Lasso (BLasso) and the Bayesian Ridge Regression (BRR). The performance of the proposed hybrid-BN method is reported for the three crude price benchmarks: West Texas Intermediate, Brent Crude and the OPEC Reference Basket.

Keywords: Bayesian networks, Random Forest, Markov chain Monte Carlo, Support vector machine

Introduction

The price of crude oil has a pivotal role in the global economy and remains at the core of energy markets. As such, its fluctuations have the potential to impact economic developments worldwide. The ability to forecast the price of crude oil is therefore a useful tool in the management of most industrial sectors (Shin et al. 2013). Nevertheless, crude oil price forecasting has been a challenging task, owing to its complex behavior resulting from the confluent influence of several factors on the crude oil market. In specific, the nonlinear features exhibited in the dynamics of oil price volatilities present a quandary for predictive techniques, making the issue of (long-term) crude price forecasting open to finance research.

A wealth of literature exists on the topic of forecasting crude oil prices. These articles are myriad, both in terms of the types of models and the number of methods being used concurrently. Some studies use an approach with a single method (non-hybrid) and some are defined by several methods (hybrid). In this regard, the generalized autoregressive

conditional heteroskedasticity (GARCH) was amongst the first methods used because of its ability to capture time-varying variance or volatility (Agnolucci 2009; Arouri et al. 2012; Cheong 2009; Fan et al. 2008a; Hou and Suardi 2012; Kang et al. 2009; Mohammadi and Su 2010; Narayan and Narayan 2007; Sadorsky 2006; Wei et al. 2010). We attempted to perform the GARCH model as a hybrid method by combining with other models, such as the stochastic volatility (SV) model, the implied volatility (IV) model and the support vector machine (SVM) model.

The neural network (NNET) method has been another approach for crude price forecasting (Azadeh et al. 2012; Ghaffari and Zare 2009; Movagharnejad et al. 2011; Shin et al. 2013; Wang et al. 2012; Yu et al. 2008; Zhang et al. 2008). However, it reportedly bears the disadvantage of over-fitting, local minima and weak generalization capability (Zhang et al. 2015). For this sake, its hybrid usage has been recommended for the purposes of crude price forecasting.

Some authors opted to use the SVM model for price prediction, taking advantage of its suitability for modeling small-sized data samples with nonlinear behavior (Guo et al. 2012). Others have reported on the merits of the wavelet technique for crude price forecasting (Yousefi et al. 2005) with one major shortcoming being its sensitivity to the sample size. However, the recent literature advocates for the use of hybrid methods to improve on the accuracy of price forecasting. The use of the best of each technique in a hybrid framework has been enhanced by combining the soft-computing or econometric method or both (Fan et al. 2008b; Xiong et al. 2013). The reader is referred to the excellent review by Zhang et al. (2015) for a more complete assessment of past research on oil price forecasting.

The motivation behind the present work was to exploit the potential of Bayesian network (BN) theory, in the context of crude price prediction, by constructing a network over decomposed price components. As such, the present article contributes to the existing literature in this field by proposing a novel hybrid method within a Bayesian network framework. In addition, this article reports results on other devised hybrid methods, using Random Forest (RF), Markov Chain Monte Carlo (MCMC), NNET and SVM. The rest of the article is organized as follows. The next section will detail the methods being used. A description of the results is provided in the third section, which will be followed by some concluding remarks.

Methodology

The hybrid methodology followed in this article takes advantage of several concepts, which are introduced in this section.

Bayesian network

A Bayesian network is an implementation of a graphical model, in which nodes represent (random) variables and arrows represent probabilistic dependencies between the nodes (Korb and Nicholson 2004). The BN's graphical structure is a directed acyclic graph (DAG) that enables estimation of the joint probability distribution. For each variable, DAG defines a factorization of the joint probability distribution into a set of local probability distributions, where the form of factorization is given by the BN's Markov property, assuming a variable to be solely dependent on its parents (Scutari 2010). In this

sake, the BN methodology initially seeks to find a DAG structure amongst the variables being considered. The two classifications of the BN-structure-learning process either treat the issue by analyzing the probabilistic relationships supervised by the Markov property of Bayesian networks with conditional independence tests and subsequently constructing a graph that satisfies the corresponding d-separation statements (constraint-based algorithms), or by assigning a score to each BN candidate and maximizing it with a heuristic algorithm (score-based algorithms) (Scutari 2010).

By taking advantage of the fundamental properties of the Bayesian networks, approximate inference (on an unknown value) is attainable. This approach should be able to avoid the curse of dimensionality due to its usage of local distributions (Nagarajan et al. 2013). Given the established BN network structure, the stochastic simulation can be applied to generate a large number of cases from the distribution network, from which the posterior probability of a target node is estimated. In this regard, the two prominent algorithms are logic sampling (LS) and likelihood weighting (LW). The LS algorithm generates a case by selecting values for each node, weighted by the probability of the values occurring at random. The nodes are traversed from the parent (root) nodes down to children (leaf) nodes. As a consequence, at each step, the weighting probability is either the prior or the Conditional Probability Table entry for the sampled parent values. A representation of all the nodes in the BN is created later on, once the full structure is visited. The collection of instantiation data enables estimation of the posterior probability for node X given evidence E (Appendix 1). The LW algorithm is similar to the former with a slight modification: adding the fractional likelihood of the evidence combination to the run count, instead of one (Appendix 2).

Empirical mode decomposition (EMD)

As a signal develops over time, a time-series may possess several temporal features. As such, exploring its characteristic behavior at different time scales should be informative. The Empirical Mode Decomposition (EMD) (Huang et al. 1998) separates the original signal into two parts: a fast-varying symmetric oscillation and a slow-varying local mean. The former constitutes the Intrinsic Mode Functions (IMF) of the original signal while the latter captures its residue. A repetitive sifting process is then implemented in order to extract the residue and the IMFs (Appendix 3). The process terminates once no more oscillations can be separated from the last residue. The EMD method has a stronger local representative capability compared to the wavelet transform and is more effective in processing non-linear or non-stationary signals (Hong 2011). In this work, the EMD implementation was made using the EMD package (Kim and Oh 2009; Kim and Oh 2014).

Quantile regression with lasso penalty (QRL)

Consider a problem of regression on a data set $\{(x_i, y_i); 1 \leq i \leq n; x_i, y_i \in \mathcal{R}^p\}$ of size N , with predictors x and response values y . The conditional ξ th quantile function, $f(x; \xi)$, is defined such that $P(Y \leq f(x; \xi) | X = x) = \xi$, for $0 < \xi < 1$. Additionally, the absolute loss function, $(r)\rho_\xi$, can be defined as (Koenker and Bassett 1978; Wu and Liu 2009):

$$\rho_\xi(r) = \begin{cases} \xi r & r > 0 \\ -(1-\xi)r & \text{otherwise} \end{cases} \tag{1}$$

The ξ th conditional quantile function can be obtained by minimizing the following (Koenker 2004):

$$\min_{f_\xi} \sum_{i=1}^n \rho_\xi(y_i - f_\xi(x_i)) + \lambda \Xi(f_\xi) \tag{2}$$

, where $\lambda \geq 0$ is a regularization parameter and $\Xi(f_\xi)$ is a roughness penalty of the f_ξ function. Assuming the conditional quantile function to be a linear function of the regressor x – as it is in the case of linear quantile regression – the function can be written as $f_\xi(x) = x^T \beta_\xi$ with $\beta_\xi = (\beta_{\xi, 1}, \beta_{\xi, 2}, \dots, \beta_{\xi, p})^T$. There are several recommendations for the penalty function in eq. 2. The Least Absolute Shrinkage and Selection Operator (Lasso) (Tibshirani 1996; Tibshirani 2011) treats this minimization case by constraining the L1-norm of the coefficients. In other words, the (classical) Lasso formulation considers a form of $\sum_{i=1}^p |\beta_{\xi,i}|$ as the penalty function. Alternatively, the functional form considered for the penalty function may account for an L2 norm constraint, the Ridge Regression method. The solution should yield a set of $\{\beta\}$ regression parameters for the problem of interest. The rqPen package (Sherwood and Maidman 2016) was used for the QRL implementation in this work.

Bayesian lasso (BLasso) and Bayesian ridge regression (BRR)

Recall the original problem of finding the regression parameters β for the model in eq. 3:

$$y = \mu 1_n + X\beta + \varepsilon \tag{3}$$

where y is the $n \times 1$ vector of responses, μ is the overall mean, X is the $n \times p$ matrix of standardized regressors and ε is the $n \times 1$ vector of independent and identically distributed normal errors with mean 0 and unknown variance 2σ . Lasso achieves a solution for β by minimizing eq. 4 through L1-penalized least squares. As such, the method should have a Bayesian interpretation, viewing the lasso estimate as the mode of the posterior distribution of β (Hans 2009).

$$\min_{\beta} (\tilde{y} - X\beta)^T (\tilde{y} - X\beta) + \lambda \sum_{i=1}^p |\beta_i| \quad | \tilde{y} = y - \bar{y} 1_n \tag{4}$$

Assuming a conditional Laplace prior on β , $\pi(\beta | \sigma^2) = \prod_{i=1}^p \frac{\lambda}{2\sqrt{\sigma^2}} e^{-\lambda|\beta_i|/\sqrt{\sigma^2}}$ and a scale invariant prior on σ^2 , $\pi(\sigma^2) = \frac{1}{\sigma^2}$ a hierarchical representation of the full model is suggested as (Park and Casella 2008):

$$\begin{aligned} y | \mu, X, \beta, \sigma^2 &\sim N_n(\mu 1_n + X\beta, \sigma^2 I_n) \beta | \sigma^2, \tau_1^2, \tau_2^2, \dots, \tau_p^2 \sim N_p(0_p, \sigma^2 D_\tau) D_\tau \\ &= \text{diag}(\tau_1^2, \dots, \tau_p^2) \sigma^2, \tau_1^2, \dots, \tau_p^2 \sim \pi(\sigma^2) d\sigma^2 \prod_{i=1}^p \frac{\lambda^2}{2} e^{-\lambda^2 \tau_i^2 / 2} d\tau_i^2 \end{aligned} \tag{5}$$

Consequently, a basis is formed for an efficient Gibbs sampler from the Bayesian posterior distribution, updating each parameter one at a time conditioned on all other parameters, with the block updating of the regression

parameters. The Bayesian concept can be also extended to the ridge regression—the Bayesian Ridge Regression (BRR) method—with an altered formulation for the priors. The reader is, however, referred to the seminal work of Park and Casella (2008) for comprehensive details of the techniques. The monomvn package (Gramacy 2016) was used to implement the BLasso/BRR methods in this work.

Markov chain Monte Carlo (MCMC)

Based on an assumption of a multivariate Gaussian () NK prior on β , and an inverse Gamma (IG) prior on the conditional error variance ε (eq. 6), the MCMC method uses Gibbs sampling to evaluate the posterior distribution of a linear regression model, enabling Bayesian inference on the regression parameters:

$$\begin{aligned} y &= X\beta + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2) \\ \beta &\sim N_K(b_0, B_0^{-1}) \\ \sigma^2 &\sim IG\left(\frac{c_0}{2}, \frac{d_0}{2}\right) \end{aligned} \quad (6)$$

In this work, the parameters used during MCMC implementation were (0.001, 0.001) $c_0 = d_0 =$ for shape factor/scale parameter of the inverse gamma prior on σ^2 , and (0, 0) $b_0 = B_0 =$ for the mean/precision of the prior on β , respectively. The latter choice corresponds to a case of putting an improper uniform prior on β . A comprehensive treatment of the MCMC method can be found in Robert and Casella (2004). The MCMC implementation was made using the R Language package, MCMCpack (Martin et al. 2011).

Random Forest (RF)

Developed upon the seminal work of Breiman (2001), the Random Forest (RF) method is an extension of classification and regression trees with a modified leaning algorithm, that is, selecting a random subset of the features at each candidate split during the learning process. The algorithm exploits trees that use a subset of the observations through bootstrapping techniques. For each tree grown on a bootstrap sample, the error rate for observations left out of the bootstrap sample is monitored as the out-of-bag (OOB) error rate, the accuracy of which indicates the RF predictor accuracy. The Random Forest algorithm seeks to improve on bagging by de-correlating the trees, implementing random feature selection at each node for the set of splitting variables (Meyer et al. 2003). As such, the RF algorithm works with two main input parameters: the number of trees and the number of variables. However, an in-depth description of the method can be found in other useful literature (Breiman 2001). For the development of results presented herein, the number of trees to grow was set to 500, in accordance with the large-value recommendation in the literature (Breiman 2001; Micheletti et al. 2014). The choice for the number of trees, however, was rendered after conducting a series of runs over a grid in the range of [100–1100] (for the number of trees) and [10–32] (for the number of variables randomly sampled) to select the optimum values with minimum mean squared residuals. This resulted in the number of variables randomly sampled as candidate at each split to be set to a 32, corresponding to the maximum number of variables available. The RF implementation was accomplished using the randomForest package (Liaw and Wiener 2002).

The proposed hybrid forecasting methodology

The hybrid methodology proposed herein exploits the characteristics of the IMFs as its mainstream. In other words, the predicted crude oil price at any future point in time is assessed based on the summation of the corresponding IMFs and the residue. Hence, forecasts of the IMFs/residue are required in the time step(s) ahead. The regression forecast is attempted based on the two types of regressors, namely internal and external, for each IMF/residue. The internal regressors were considered to be those previous values of an IMF/residue, to which the current value depends, which is determined through the Partial Auto-correlation of the decomposed signal, whereas the external regressors were considered to be the technical indicators, the volatility index (VIX), and the implied oil volatility index (OVX). In this regard, the technical indicators taken into account have been the Aroon indicator (aroon), the Commodity Channel Index (CCI), the Double Exponential Moving Average (DEMA), the Exponential Moving Average (EMA), the Moving Average Convergence/Divergence (MACD), the Relative Strength Index (RSI), the Simple Moving Average (SMA), the Traders Dynamic Index (TDI) and the Triple Exponentially Smoothed Moving Average (TRIX).

The proposed methodology is hybrid in two different ways: classification and regression. The initial classification step involves the determination of the set of significant regressors, from the pool of regressors described using BN (constrained/scored), QRL, BLasso and as its methods, where the criterion of significance differs for each method. In the BLasso/BRR classification, for instance, a regressor is considered significant when the estimated posterior probability of the individual component's regression coefficient is returned as nonzero. In the BN scenario, on the other hand, significance is recognized in the strength of the corresponding arch—an arch strength coefficient of equal to or less than 0.05 when a conditional independence test is applied (constrained BN), or a strength coefficient of less than zero threshold, when network scores are applied (scored BN) (Scutari 2010). The final set of significant regressors is selected as the whole set of significant parameters, determined by the BN-QRL-BLasso-BRR methods, for each IMF/residue.

Table 1 Different predictive strategies tested

Name	Strategy
Method-1	RF + GARCH
Method-2	SVM + GARCH
Method-3	MCMC+GARCH
Method-4	NNET+GARCH
Method-5	BN + GARCH
Method-6	RF
Method-7	SVM
Method-8	MCMC
Method-9	NNET
Method-10	BN

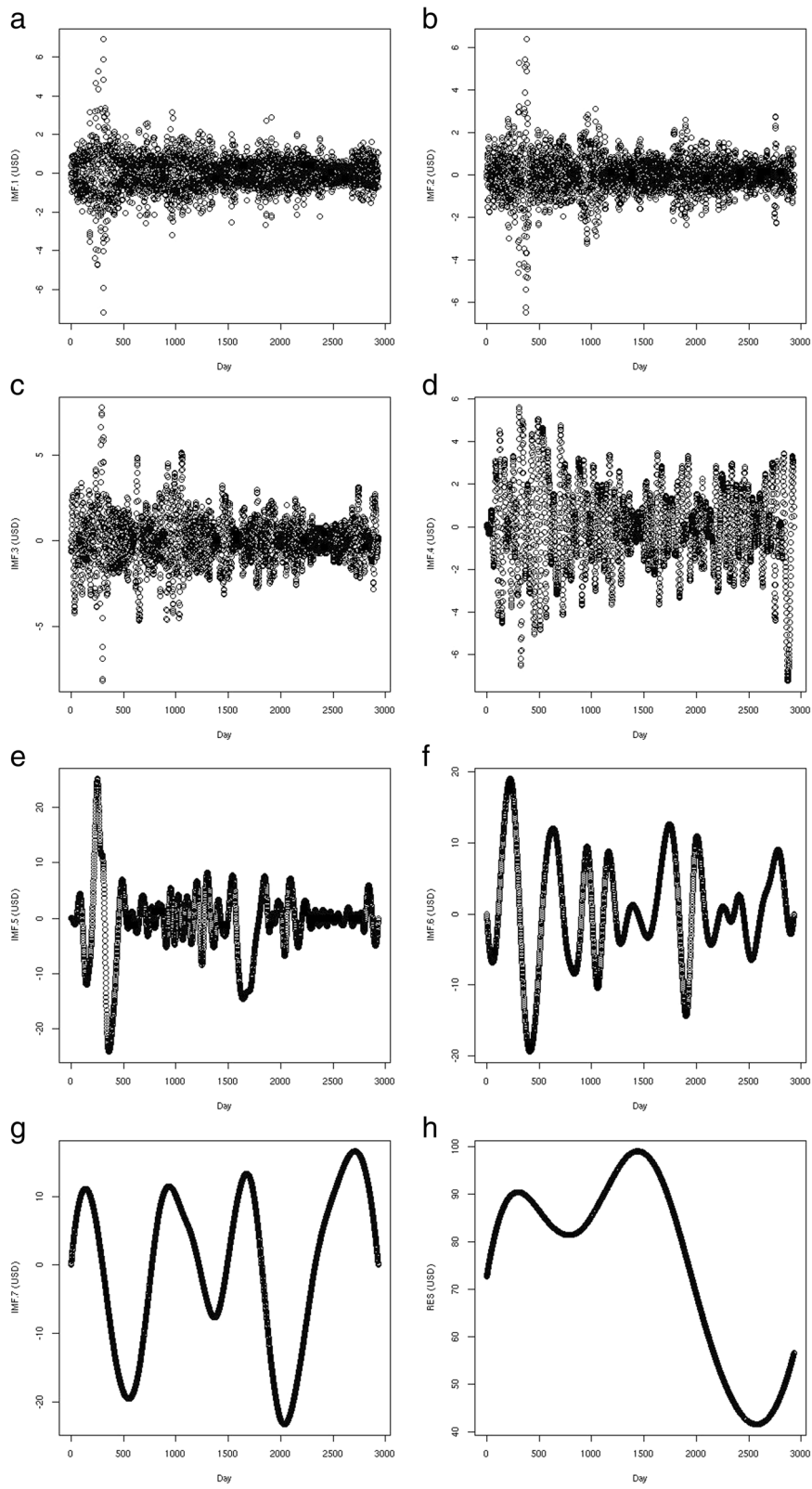


Fig. 1 The Intrinsic Mode Functions- IMF.1 (a), IMF.2 (b), IMF.3 (c), IMF.4 (d), IMF.5 (e), IMF.6 (f), IMF.7 (g) - and the residue (h), decomposed from WTI price data by the EMD method

The second regression step, being hybrid in essence, involves predicting the IMFs/residue from their corresponding regressors, in the time steps ahead. For this reason, several strategies were devised and later tested for efficiency (Table 1). The first five methods take advantage of the presence of time-varying feature in the decomposed IMF signals. Under such circumstances, the GARCH model is used to forecast the time-varying IMF components, while the rest of the IMFs as well as the residue are predicted by the conjugate model listed. In the second five strategies, the IMFs as well as the residue are computed by a candidate model, in the time steps ahead, regardless of the time-varying aspect of some of the IMFs. The procedures of the proposed hybrid method can be summarized as follows:

- (1) Apply the EMD method to the original crude price series. Decompose the series into its IMFs and the residue.
- (2) Extract the significant regressors for each IMF and residue by taking the whole set of significant regressors determined by the BN-QRL-BLasso-BRR methods.
- (3) Under Schemes 1–5 in Table 1, if the corresponding IMF presents the feature of time variation, use the GARCH model to predict its future value. Use the conjugate model in the scheme to predict the future values of other IMFs and the residue based on the regressors determined in stage 2.
- (4) Under Schemes 6–10 in Table 1, use the candidate model prescribed to forecast the future values of the IMFs and residue, based on the regressors determined in stage 2.
- (5) Construct the final forecasted crude oil price by summing the future values IMFs and the residue.

The general specification of the mean/variance in the GARCH model considered in Schemes 1–5 of Table 1 is as follows, respectively:

Table 2 Descriptive statistics of the IMFs/residue of WTI, in the period between 06-July-2007 and 25-February-2019

Name	augmented Dickey–Fuller (ADF) ^a	Kwiatkowski-Phillips-Schmidt-Shin (KPSS) ^b	Jarque–Bera	Kurtosis	mean	Standard Deviation
IMF.1	−15.24 (0.01)	0.22 (0.1)	3781.20 (0.000)	8.567	0.0035	0.92
IMF.2	−16.79 (0.01)	0.03 (0.1)	3917.50 (0.000)	8.66	0.0008	0.97
IMF.3	−13.84 (0.01)	0.022 (0.1)	901.49 (0.000)	5.711	−0.0066	1.567
IMF.4	−8.63 (0.01)	0.062 (0.1)	81.49 (0.184)	3.622	−0.1983	2.266
IMF.5	−7.82 (0.02)	0.180 (0.1)	564.44 (0.000)	5.139	−0.1187	6.445
IMF.6	−3.70 (0.01)	0.276 (0.1)	878.23 (0.000)	5.682	0.2289	7.557
IMF.7	−2.30 (0.45)	1.519 (0.01)	244.24 (0.000)	2.127	−1.2018	10.844
RES	−1.83 (0.64)	20.624 (0.01)	366.29 (0.000)	1.755	76.7498	16.129

^a Alternative hypothesis: Data is stationary

^b Null hypothesis: Data is level stationary

Table 3 Descriptive statistics of the IMFs/residue of BRENT, in the period between 06-July-2007 and 25-February-2019

Name	augmented Dickey–Fuller (ADF) ^a	Kwiatkowski-Phillips-Schmidt-Shin (KPSS) ^b	Jarque–Bera	Kurtosis	mean	Standard Deviation
IMF.1	−14.78 (0.01)	0.204 (0.1)	554.27 (0.000)	5.132	0.0001	0.81
IMF.2	−18.008 (0.01)	0.067 (0.1)	870.12 (0.000)	5.674	0.0263	1.00
IMF.3	−16.059 (0.01)	0.024 (0.1)	1281.5 (0.000)	6.247	−0.0294	1.55
IMF.4	−10.714 (0.01)	0.023 (0.1)	152.74 (0.000)	4.115	0.0302	1.81
IMF.5	−5.140 (0.01)	0.127 (0.1)	511.08 (0.000)	4.914	0.094	5.57
IMF.6	−6.923 (0.01)	0.273 (0.1)	4365.10 (0.000)	8.566	0.0213	7.83
IMF.7	−3.055 (0.13)	2.841 (0.01)	28.49 (0.000)	2.802	0.597	11.88
IMF.8	−2.574 (0.33)	2.841 (0.01)	28.49 (0.000)	2.716	−2.403	7.51
RES	−2.896 (0.19)	14.343 (0.01)	186.48 (0.000)	1.861	82.330	20.49

^a Alternative hypothesis: Data is stationary

^b Null hypothesis: Data is level stationary

$$\begin{aligned}
 x_{oil,t} = & \eta_1 + \eta_2 x_{oil,t-1} + \eta_3 x_{oil,t-2} + \dots + \eta_{n+1} x_{oil,t-n} + \eta_{n+2} x_{OVX,t-1} + \eta_{n+3} x_{VIX,t-1} \\
 & + \eta_{n+4} x_{aroon,t-1} + \eta_{n+5} x_{CCI,t-1} + \eta_{n+6} x_{DEMA,t-1} + \eta_{n+7} x_{EMA,t-1} \\
 & + \eta_{n+8} x_{MACD,t-1} + \eta_{n+9} x_{RSI,t-1} + \eta_{n+10} x_{SMA,t-1} + \eta_{n+11} x_{TDI,t-1} \\
 & + \eta_{n+12} x_{TRIX,t-1} + \varepsilon_t \\
 \sigma_{oil,t}^2 = & \eta_{n+13} + \eta_{n+14} \sigma_{oil,t-1}^2 + \eta_{n+15} \varepsilon_{t-1}^2
 \end{aligned} \tag{8}$$

,where $x_{i,t}$ refer to the price/value of i at time t , and n refers to the number of previous time steps to which the current oil price data is correlated. The final form of the GARCH specification for the time-varying IMFs may adopt a truncated version in mean, though, as they differ in their type/number of significant regressors, to be incorporated into the equation for mean. The GARCH predictions were made by the rugarch package (Ghalanos 2015). The accuracy of hybrid forecasting methods was evaluated by several statistical criteria, namely, the mean absolute error (MAE), the root mean square error (RMSE) and the mean absolute percentage error (MAPE).

Table 4 Descriptive statistics of the IMFs/residue of ORB, in the period between 06-July-2007 and 25-February-2019

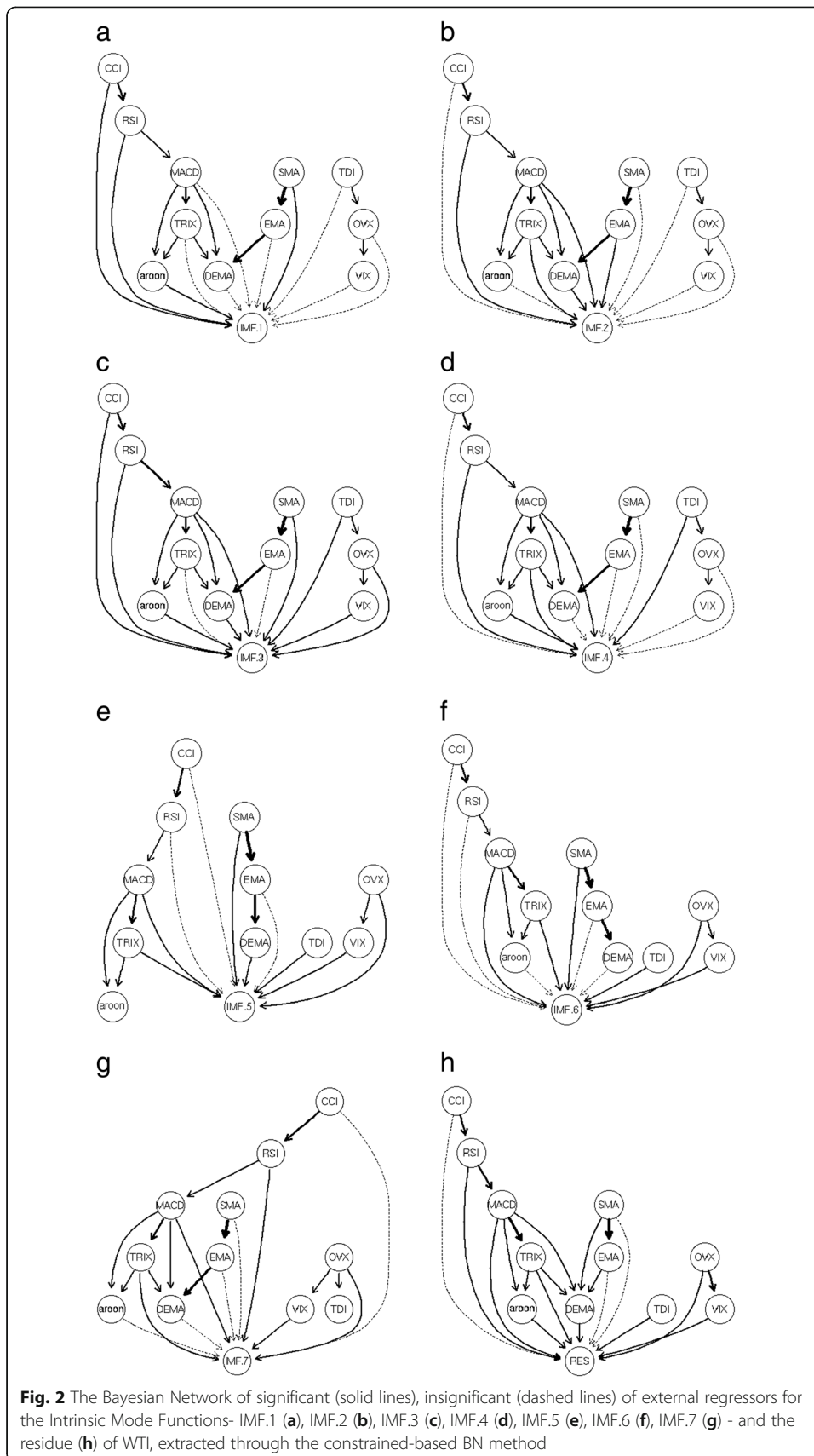
Name	augmented Dickey–Fuller (ADF) ^a	Kwiatkowski-Phillips-Schmidt-Shin (KPSS) ^b	Jarque–Bera	Kurtosis	mean	Standard Deviation
IMF.1	−15.60 (0.01)	0.134 (0.1)	1185.2 (0.000)	6.11	0.0057	0.59
IMF.2	−18.54 (0.01)	0.035 (0.1)	1400.9 (0.000)	6.38	0.0201	0.85
IMF.3	−15.65 (0.01)	0.017 (0.1)	402.6 (0.000)	4.81	−0.0100	1.22
IMF.4	−7.751 (0.01)	0.212 (0.1)	1283.4 (0.000)	6.21	−0.0380	2.53
IMF.5	−7.084 (0.01)	1.806 (0.1)	3266.4 (0.000)	8.05	−1.1880	8.85
IMF.6	−2.993 (0.15)	2.168 (0.01)	1347.2 (0.000)	5.34	1.1660	7.97
IMF.7	−1.914 (0.61)	0.506 (0.04)	86.03 (0.951)	2.42	−0.426	4.50
IMF.8	−3.090 (0.11)	1.694 (0.01)	195.00 (0.000)	1.97	−1.026	8.96
RES	−2.871 (0.20)	14.435 (0.01)	191.58 (0.000)	1.75	79.810	19.36

^a Alternative hypothesis: Data is stationary

^b Null hypothesis: Data is level stationary

Table 5 The set of significant regressors detected for each IMF/residue of WTI, with the BN-QRL-BLasso-BRR techniques

Name	Method	Significant external regressors detected
IMF.1	BN (Constrained)	aroon, CCI, RSI, SMA
	BN (Scored)	aroon, CCI, RSI, SMA
	QRL	aroon, CCI
	BLasso	aroon, CCI, DEMA, RSI, SMA
	BRR	aroon, CCI, DEMA, RSI, SMA
IMF.2	BN (Constrained)	DEMA, EMA, MACD, RSI, TRIX
	BN (Scored)	DEMA, MACD, RSI, TRIX
	QRL	aroon, CCI, TDI
	BLasso	CCI, MACD, RSI, TRIX, VIX
	BRR	DEMA, EMA, MACD, RSI, SMA, TRIX, VIX
IMF.3	BN (Constrained)	aroon, CCI, DEMA, MACD, OVX, RSI, SMA, TDI, VIX
	BN (Scored)	aroon, CCI, DEMA, MACD, OVX, RSI, SMA, TDI, VIX
	QRL	aroon, CCI, TDI
	BLasso	aroon, CCI, DEMA, EMA, MACD, OVX, RSI, SMA, TDI, TRIX, VIX
	BRR	aroon, CCI, DEMA, EMA, MACD, OVX, RSI, SMA, TDI, TRIX
IMF.4	BN (Constrained)	aroon, MACD, RSI, TDI
	BN (Scored)	aroon, MACD, TDI
	QRL	aroon, CCI, TDI
	BLasso	aroon, DEMA, EMA, MACD, RSI, SMA, TDI, TRIX, VIX
	BRR	aroon, DEMA, EMA, MACD, RSI, SMA, TDI
IMF.5	BN (Constrained)	DEMA, MACD, SMA, TDI, OVX, VIX
	BN (Scored)	DEMA, MACD, SMA, TDI, OVX, VIX
	QRL	aroon, CCI, DEMA, EMA, TDI
	BLasso	aroon, CCI, DEMA, EMA, MACD, OVX, RSI, SMA, TDI, TRIX, VIX
	BRR	DEMA, EMA, MACD, OVX, RSI, SMA, TDI
IMF.6	BN (Constrained)	MACD, OVX, SMA, TDI, TRIX, VIX
	BN (Scored)	MACD, OVX, SMA, TDI, TRIX, VIX
	QRL	CCI, DEMA, VIX
	BLasso	DEMA, EMA, MACD, OVX, RSI, SMA, TDI, TRIX, VIX
	BRR	DEMA, EMA, MACD, OVX, RSI, SMA, TDI, TRIX, VIX
IMF.7	BN (Constrained)	MACD, RSI, TRIX, OVX, VIX
	BN (Scored)	MACD, RSI, TRIX, OVX, VIX
	QRL	aroon, CCI, DEMA, OVX, TDI
	BLasso	CCI, DEMA, EMA, OVX, SMA, TDI, VIX
	BRR	CCI, DEMA, EMA, OVX, SMA, TDI, TRIX, VIX
RES	BN (Constrained)	aroon, MACD, OVX, RSI, TDI, TRIX, VIX
	BN (Scored)	aroon, MACD, OVX, RSI, TDI, TRIX, VIX
	QRL	aroon, CCI, EMA, TDI, VIX
	BLasso	aroon, DEMA, MACD, OVX, SMA, TDI, TRIX, VIX
	BRR	aroon, DEMA, MACD, OVX, SMA, TDI, TRIX, VIX



Data and results

The price data of oil/OVX/VIX were acquired through the Quandl package (McTaggart et al. 2016). As for the crude price, the data related to three benchmarks were collected: West Texas Intermediate (WTI), Brent crude (BRENT), and the OPEC Reference Basket (ORB).

A number of seven distinct IMFs (IMF.1, IMF.2, ..., IMF.7) and one residue (RES) were detected for WTI (Fig. 1), while for BRENT/ORB, eight IMFs were detected. In this process, the S stoppage rule (Huang and Wu 2008) was considered, as for the stopping rule of the sifting process. Tables 2, 3 and 4 report the statistics measured on each IMF/residue in the time span between July 6, 2007 and February 25, 2019. The subscripts in parenthesis indicate the corresponding p -values. The IMFs with leptokurtic characteristic (Kurtosis > 3) were later used in Methods 1–5 (Table 1) as input into the GARCH model. The statistics were measured using the tseries package (Trapletti and Hornik 2015).

The technical indicators were computed using the TTR package (Ulrich 2016). However, it should be noted that because the oil price data used contained the closing price values, the CCI numbers obtained herein essentially receive an altered meaning to their original definition. Table 5 lists the set of significant external regressors, separately detected by the BN-QRL-BLasso-BRR methods, for each IMF/residue of WTI. Figure 2 provides a graphical representation of the extracted Bayesian network of significant/insignificant previous-time external regressors of the IMFs/residue of WTI, obtained through the constraint-based BN concept. The Rgraphviz package (Hansen et al. 2008) was used to plot the graph. The final set of previous-time external regressors considered for each IMF/residue was, however, taken as the whole of the significant detected parameters (Tables 6, 7 and 8). Detection of the significant internal regressors was made by following the standard procedure of considering the partial autocorrelation of the decomposed signal (Fig. 3). The DAG graphs (Fig. 3) are important in the sense that they reveal the influencing parameters on each intrinsic mode function. The CCI, SMA and OVX parameters are not directly connected; the DAG assumes that there is no possibility to revisit a given vertex after starting from that same vertex, following a consistently directed sequence of edges. A connection between these nodes (vertices) would violate this rule for the data involved, which is the reason why they appear unconnected in the reported DAGs. Moreover, the CCI, SMA and OVX can be considered to be the parents for most of intrinsic mode functions as well as the residue.

Table 6 The final set of external regressors for each IMF/residue of WTI

Name	Significant external regressors
IMF.1	aroon, CCI, DEMA, RSI, SMA
IMF.2	aroon, CCI, DEMA, EMA, MACD, RSI, SMA, TDI, TRIX, VIX
IMF.3	aroon, CCI, DEMA, EMA, MACD, OVX, RSI, SMA, TDI, TRIX, VIX
IMF.4	aroon, CCI, DEMA, EMA, MACD, RSI, SMA, TDI, TRIX, VIX
IMF.5	aroon, CCI, DEMA, EMA, MACD, OVX, RSI, SMA, TDI, TRIX, VIX
IMF.6	CCI, DEMA, EMA, MACD, OVX, RSI, SMA, TDI, TRIX, VIX
IMF.7	aroon, CCI, DEMA, EMA, MACD, OVX, RSI, TDI, TRIX, VIX
RES	aroon, CCI, DEMA, EMA, MACD, OVX, RSI, TDI, TRIX, VIX

Table 7 The final set of external regressors for each IMF/residue of BRENT

Name	Significant external regressors
IMF.1	aroon, CCI, EMA, MACD, RSI, SMA, VIX
IMF.2	aroon, CCI, DEMA, MACD, TRIX
IMF.3	aroon, CCI, DEMA, EMA, MACD, OVX, RSI, SMA, TDI, TRIX, VIX
IMF.4	aroon, CCI, MACD, RSI, SMA, TDI, TRIX
IMF.5	aroon, CCI, DEMA, EMA, MACD, OVX, RSI, VIX
IMF.6	CCI, DEMA, EMA, MACD, OVX, SMA, TDI, TRIX, VIX
IMF.7	aroon, CCI, DEMA, MACD, OVX, RSI, TDI, VIX
RES	aroon, CCI, EMA, MACD, OVX, RSI, TRIX, VIX

The hybrid strategy optimization (Ghalanos 2015) was used within the GARCH implementation in Methods 1–5. This ensures that a number of non-linear solvers are called in a sequence in the case that the initial optimization fails. For the SVM implementation (Methods 2 and 7 in Table 1), a grid-search was initially conducted over the parameter ranges so as to calibrate the SVM model (Meyer et al. 2015). This was followed by a kernel-based SVM regression, where the hyperparameters of the kernel were taken as those obtained from the calibration stage. The Laplacian kernel was used within the SVM regression with bound constraint (Karatzoglou et al. 2004). For the NNET predictions (Methods 4 and 9), the k-nearest neighbor method was used without any preprocessing of the predictor data (Kuhn et al. 2016). As for the MCMC (Methods 3 and 8), a number of 1,000 burn-in iterations was elapsed, followed by 10,000 Metropolis iterations for the sampler. Also, a number of 1,000 MCMC samples were collected for the BLasso/BRR outputs. The initial lasso penalty parameter was taken as one and the Rao-Blackwellized samples were used for σ^2 (Gramacy 2016). The selection of the model for the columns of the design matrix for regression parameters in BLasso/BRR was made using the Reverse-Jump MCMC (Gramacy 2016).

Implementation of the graphical structure-learning of the Bayesian networks was attempted using the bnlearn package (Scutari 2017). Both types of constraint/score-based algorithms were tested in this work. For the constraint-based type, the Monte Carlo permutation test was used for the conditional

Table 8 The final set of external regressors for each IMF/residue of ORB

Name	Significant external regressors
IMF.1	aroon, CCI, DEMA, EMA, RSI, SMA, TRIX
IMF.2	aroon, CCI, DEMA, MACD, SMA, TDI, TRIX
IMF.3	aroon, CCI, DEMA, EMA, MACD, OVX, RSI, SMA, TDI, TRIX, VIX
IMF.4	aroon, CCI, EMA, MACD, OVX, RSI, TRIX, VIX
IMF.5	aroon, CCI, DEMA, MACD, OVX, RSI, TDI, TRIX, VIX
IMF.6	CCI, DEMA, EMA, MACD, OVX, RSI, SMA, TDI, TRIX, VIX
IMF.7	CCI, DEMA, OVX, TDI, VIX
RES	aroon, CCI, DEMA, EMA, MACD, OVX, SMA, TDI, TRIX, VIX

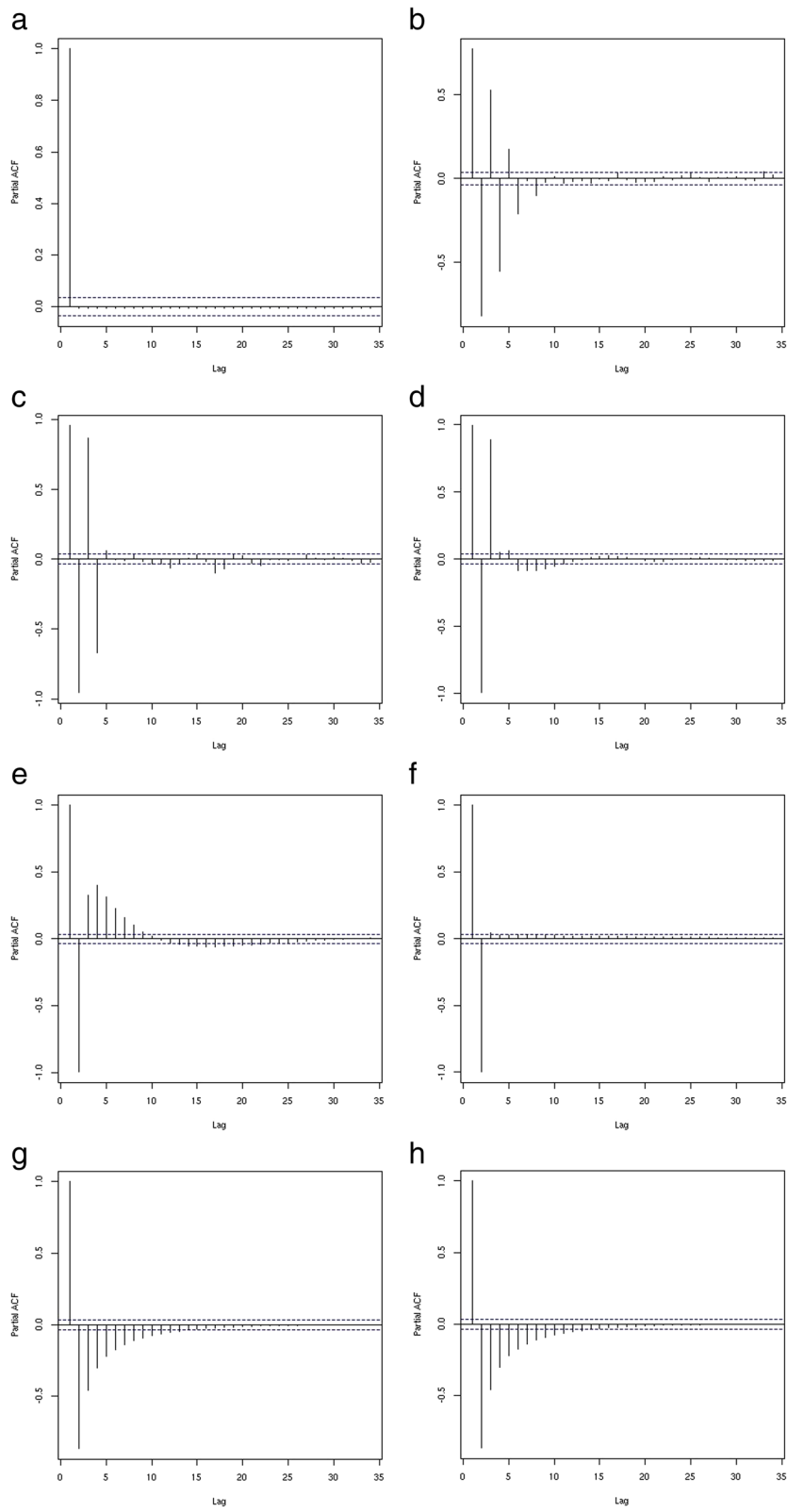


Fig. 3 The partial autocorrelation of the Intrinsic Mode Functions- IMF.1 (a), IMF.2 (b), IMF.3 (c), IMF.4 (d), IMF.5 (e), IMF.6 (f), IMF.7 (g) - and the residue (h) of WTI

Table 9 The average errors of the 10-days-ahead forecasts of WTI

Name	MAE	RMSE	MAPE (%)
Method-1	1.75	2.13	0.03
Method-2	1.79	2.16	0.03
Method-3	2.58	3.32	0.04
Method-4	3.02	3.37	0.05
Method-5	1.87	2.34	0.03
Method-6	0.95	1.17	0.01
Method-7	1.28	1.52	0.02
Method-8	7.26	7.88	0.12
Method-9	1.92	2.14	0.03
Method-10	0.75	0.81	0.015

independence test. While in the score-based case, a score equivalent Gaussian posterior density criterion was applied. For a BN inference, predictions were obtained by applying the LW algorithm and extracting the expected value of the conditional distribution of 500 simulation results. All the available nodes in the structure were taken as evidence in that situation except the node related to the variable being predicted.

The crude price forecasting was attempted from periods of time ten days ahead. To test the performance of the methods, three random training sets were used with a common beginning date of July 6, 2007 and ending dates of January 3, 2017, March 27, 2017 and February 8, 2019, respectively. In each case, the hybrid methods were used to predict the crude prices for the ten out-of-sample days immediately ahead. The average of the errors incurred is reported in Tables 9, 10 and 11. According to the results, Method-10 shows conspicuously better performance when compared with the other methods, in all three statistics measured (MAE, RMSE and MAPE). The superior performance of Method-10 is not merely bounded to a single market, rather it is demonstrated over the three price types (WTI, BRENT and ORB). Furthermore, the predictive accuracy of Method-10 is compared by using a Diebold-Mariano test against other techniques. Table 12 lists the Diebold-Mariano statistics for the 10-day forecast of WTI/BRENT/ORB, assessing the alternative hypothesis that Method-10 is more accurate than the method

Table 10 The average errors of the 10-days-ahead forecasts of BRENT

Name	MAE	RMSE	MAPE (%)
Method-1	18.59	32.08	0.28
Method-2	18.79	32.14	0.29
Method-3	18.16	31.86	0.28
Method-4	20.17	33.23	0.32
Method-5	19.05	32.62	0.29
Method-6	0.96	1.19	0.01
Method-7	1.19	1.47	0.01
Method-8	19.98	23.91	0.30
Method-9	4.39	5.17	0.06
Method-10	0.96	1.20	0.01

Table 11 The average errors of the 10-days-ahead forecasts of ORB

Name	MAE	RMSE	MAPE (%)
Method-1	0.74	0.89	0.01
Method-2	1.29	1.42	0.01
Method-3	3.07	3.17	0.04
Method-4	2.06	2.25	0.03
Method-5	0.68	0.82	0.01
Method-6	0.55	0.63	0.00
Method-7	0.91	1.06	0.01
Method-8	23.60	26.96	0.36
Method-9	2.97	3.56	0.04
Method-10	0.49	0.58	0.00

of choice, clearly demonstrating the superior accuracy of the proposed hybrid technique. A close inspection of the results also indicates that, overall, Methods 6–10 achieve a better predictive success than Methods 1–5, which incorporate the GARCH model. However, this should not rule out the application of GARCH in hybrid price forecasting models.

The extracted Bayesian networks indicate that both OVX and VIX are influential on different layers of IMF or the residue, which accounts for their impact on the value of future crude prices. In addition, the established BN structure shows the previous-time technical indicators to affect different layers of IMF and the residue of the decomposed price signal.

The proposed hybrid methodology was chosen for short-term prediction of crude prices, owing to the short-term viability of the regressors employed. The method, however, deserves further investigation and merits being tested for its long-term forecasting capability, for example incorporating regressors with longer life spans.

Conclusions

The performance of the hybrid Bayesian network proposition was outstanding compared to the other devised hybrid models, in all of the three crude price types (WTI, BRENT and ORB) and against all of the three statistical benchmarks (MAE,

Table 12 The Diebold-Mariano statistics for 10-day forecast of WTI/BRENT/ORB, for the alternative hypothesis that Method-10 is better in terms of accuracy versus the method of choice

	WTI	BRENT	ORB
Method-1	1.93 (0.04)	1.39 (0.09)	1.45 (0.08)
Method-2	1.99 (0.03)	1.39 (0.09)	2.80 (0.01)
Method-3	2.17 (0.02)	1.38 (0.09)	5.89 (0.00)
Method-4	3.32 (0.00)	1.44 (0.09)	3.47 (0.00)
Method-5	1.85 (0.04)	1.41 (0.08)	1.55 (0.07)
Method-6	1.05 (0.01)	0.06 (0.10)	0.33 (0.03)
Method-7	2.20 (0.02)	0.87 (0.20)	1.90 (0.04)
Method-8	4.56 (0.00)	2.35 (0.02)	3.06 (0.00)
Method-9	2.70 (0.01)	2.48 (0.01)	3.14 (0.00)

RMSE and MAPE). The BN demonstrated that the volatility indices (OVX, VIX) are influential on different decomposed signals of the crude price, affecting the level-ahead price values. The predictive power of the hybrid methods adopting GARCH was shown to be inferior to the other methods, which apply regressions to all of the layers of the decomposed signal for crude price forecasting. Since the proposed hybrid method makes use of regressors with short-term life spans (i.e., technical indicators, OVX, VIX and past price values), the method remains a valid option for short-term forecasting. The question of its capability in handling long-term price forecasts is yet to be answered by the future research using parameters with longer-term viability.

Appendix 1

The Logic sampling algorithm

Consider an established Bayesian Network. Assume X to be node in this BN structure, and E as a given evidence. The Logic sampling algorithm for estimation of the posterior probability of node X given evidence $E = e$, is computed by the following procedure (Korb and Nicholson 2004):

Step-1 Initialize

For each value x_i for node X
 Create a count variable $\text{Count}(x_i, e)$
 Create a count variable $\text{Count}(e)$
 Initialize all count variables to zero

Step-2 Repeat

For all the root (parent) nodes
 Choose a value, weighed the choice by the priors, at random
 Loop
 Choose values for children at random, using the conditional probabilities given the known values of the parents
 Until all the BN structure is visited

Step-3 Update

If the case (instantiation) includes $E=e$
 $\text{Count}(e) = \text{Count}(e) + 1$
 If the case includes both $X=x_i$ and $E=e$
 $\text{Count}(x_i, e) = \text{Count}(x_i, e) + 1$

Step-4 Estimate

Obtain an estimate for the posterior probability

$$P(X = x_i | E = e) = \frac{\text{Count}(x_i, E = e)}{\text{Count}(E = e)}$$

Appendix 2

The Likelihood weighing algorithm

Consider an established Bayesian Network. Assume X to be node in this BN structure, and E as a given evidence. The Likelihood weighing algorithm for estimation of the posterior probability of node X given evidence $E = e$, is computed by the following procedure (Korb and Nicholson 2004):

Step-1 Initialize

For each value x_i for node X
 Create a count variable $Count(x_i, e)$
 Create a count variable $Count(e)$
 Initialize all count variables to zero

Step-2 Repeat

For all root nodes
 If a root is an evidence node, E_j
 Choose the evidence value, e_j
 likelihood($E_j = e_j$) = $P(E_j = e_j)$
 Else
 Choose a value for children at random, using the conditional probabilities given the known values of the parents

Until the entire BN structure is visited

Step-3 Update

If the case includes $E=e$

$$Count(e) = Count(e) + \prod_j likelihood(E_j = e_j)$$

If this case includes both $X=x_i$ and $E=e$

$$Count(x_i, e) = Count(x_i, e) + \prod_j likelihood(E_j = e_j)$$

Step-4 Estimate

Obtain an estimate for the posterior probability

$$P(X = x_i | E = e) = \frac{Count(x_i, E = e)}{Count(E = e)}$$

Appendix 3

The Empirical Mode Decomposition algorithm

Consider a time series, $f(t)$, with t referring to time. The EMD fractionation of the original data is made following the below steps (Huang et al. 1998):

Step-1

Find the extrema (maxima and minima) of the input data $m_0(t) = f(t)$

Form the upper-envelope $U_0(t)$, and the lower-envelope $L_0(t)$, by connecting the maxima and minima with spline function, respectively.

$$\text{Calculate the first envelope mean, } m_1(t) = \frac{U_0(t) + L_0(t)}{2}$$

Calculate the first difference between the mean and the data,

$$h_1(t) = m_0(t) - m_1(t)$$

Step-2

Take the previous difference, $h_{i-1}(t)$ as input.

Find the envelopes, using the procedure in step-1.

Calculate the mean and difference.

$$m_i(t) = \frac{U_{i-1}(t) + L_{i-1}(t)}{2}$$

$$h_i(t) = m_{i-1}(t) - m_i(t)$$

Step-3

Repeat step-2 until the stopping condition is satisfied – a condition of small difference between successive iterations, or small value of envelope mean.

Calculate the first IMF and the residue

$$IMF_1(t) = h_n(t)$$

$$r_1(t) = f(t) - IMF_1(t)$$

Step-4

Take $r_1(t)$ as the input and repeat steps 1-3, until the number of extrema is less than 2.

Abbreviations

aroon: Aroon indicator; Blasso: Bayesian Lasso; BN: Bayesian Network; BRENT: Brent crude; BRR: Bayesian Ridge Regression; CCI: Commodity Channel Index; DAG: Directed acyclic graph; DEMA: Double Exponential Moving Average; EMA: Exponential Moving Average; EMD: Empirical Mode Decomposition; GARCH: Generalized autoregressive conditional heteroskedasticity; IMF: Intrinsic Mode Functions; IV: Implied volatility; KSVM: Kernel Support Vector Machine; LS: Logic sampling; LW: Likelihood weighting; MACD: Moving Average Convergence/Divergence; MAE: Mean absolute error; MAPE: Mean absolute percentage error; MCMC: Markov Chain Monte Carlo; NNET: Neural Networks; OOB: out-of-bag; ORB: OPEC Reference Basket; OVX: Implied Oil Volatility Index; QRL: Lasso penalty; RF: Random Forest; RMSE: Root Mean Square Error; RSI: Relative Strength Index; SMA: Simple Moving Average; SV: Stochastic volatility; SVM: Support Vector Machine; TDI: Traders Dynamic Index; TRIX: Triple Exponentially Smoothed Moving Average; VIX: Volatility Index; WTI: West Texas Intermediate

Acknowledgements

Not applicable.

Author's contributions

The work was solely done by the corresponding author, Babak Fazelabdolabadi. The author read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

Not applicable.

Competing interests

The author declares that he has no competing interests.

Received: 18 November 2018 Accepted: 10 June 2019

Published online: 11 July 2019

References

- Agnolucci P (2009) Volatility in crude oil futures: a comparison of the predictive ability of GARCH and implied volatility models. *Energy Econ* 31:316–321. <https://doi.org/10.1016/j.eneco.2008.11.001>
- Arouri MEH, Lahiani A, Lévy A, Nguyen DK (2012) Forecasting the conditional volatility of oil spot and futures prices with structural breaks and long memory models. *Energy Econ* 34:283–293. <https://doi.org/10.1016/j.eneco.2011.10.015>
- Azadeh A, Moghaddam M, Khakzad M, Ebrahimipour V (2012) A flexible neural network-fuzzy mathematical programming algorithm for improvement of oil price estimation and forecasting. *Comput Ind Eng* 62:421–430. <https://doi.org/10.1016/j.cie.2011.06.019>
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32. <https://doi.org/10.1023/A:1010933404324>
- Cheong CW (2009) Modeling and forecasting crude oil markets using ARCH-type models. *Energy Policy* 37:2346–2355. <https://doi.org/10.1016/j.enpol.2009.02.026>
- Fan Y, Liang Q, Wei YM (2008b) A generalized pattern matching approach for multistep prediction of crude oil price. *Energy Econ* 30(3):889–904. <https://doi.org/10.1016/j.eneco.2006.10.012>
- Fan Y, Zhang YJ, Tsai H-T, Wei YM (2008a) Estimating 'value at risk' of crude oil price and its spillover effect using the GED approach. *Energy Econ* 30(6):3156–3171. <https://doi.org/10.1016/j.eneco.2008.04.002>
- Ghaffari A, Zare S (2009) A novel algorithm for prediction of crude oil price variation based on soft computing. *Energy Econ* 31:531–536. <https://doi.org/10.1016/j.eneco.2009.01.006>
- Ghalanos A. 2015. rugarch: Univariate GARCH models. R package version 1.3-6.
- Gramacy RB. 2016. monomvn: Estimation for Multivariate Normal and Student-t Data with Monotone Missingness. R package version 1.9-6. <http://CRAN-R-project.org/package=monomvn>. Accessed 25 Feb 2019
- Guo XP, Li DC, Zhang AH (2012) Improved support vector machine oil price forecast model based on genetic algorithm optimization parameters. *AASRI Procedia* 1:525–530. <https://doi.org/10.1016/j.aasri.2012.06.082>
- Hans C (2009) Bayesian lasso regression. *Biometrika* 96(4):835–845. <https://doi.org/10.1093/biomet/asp047>
- Hansen KD, Gentry J, Long L, Gentleman R, Falcon S, Hahne F, Sarkar D (2008) Rgraphviz: provides plotting capabilities for R graph objects. R package version 2:14.0
- Hong L (2011) Decomposition and forecast for financial time series with high-frequency based on empirical mode decomposition. *Energy Procedia* 5:1333–1340. <https://doi.org/10.1016/j.egypro.2011.03.231>
- Hou A, Suardi S (2012) A nonparametric GARCH model of crude oil price return volatility. *Energy Econ* 34:618–626. <https://doi.org/10.1016/j.eneco.2011.08.004>
- Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zhang Q, Yen N, Tung CC, Liu HH (1998) The empirical mode decomposition and Hilbert Spectrum for nonlinear and nonstationary time series analysis. *Proc R Soc London Ser A* 454:903–995. <https://doi.org/10.1098/rspa.1998.0193>
- Huang NE, Wu Z (2008) A review on Hilbert-Huang transform: method and its applications to geophysical studies. *Rev Geophys* 46:RG2006. <https://doi.org/10.1029/2007RG000228>
- Kang SH, Kang SM, Yoon SM (2009) Forecasting volatility of crude oil markets. *Energy Econ* 31:119–125. <https://doi.org/10.1016/j.eneco.2008.09.006>
- Karatzoglou A, Smola A, Hornik K, Zeileis A (2004) Kernlab – an S4 package for kernel methods in R. *J Stat Softw* 11(9):1–20. <http://www.jstatsoft.org/v11/i09/>
- Kim D, Oh H (2009) A package for empirical mode decomposition and Hilbert Spectrum. *The R Journal* 1:30–46
- Kim D, Oh H, 2014. EMD: Empirical Mode Decomposition and Hilbert Spectral Analysis. R package version 1.5.7.

- Koenker R (2004) Quantile regression for longitudinal data. *J Multivar Anal* 91:74–89. <https://doi.org/10.1016/j.jmva.2004.05.006>
- Koenker R, Bassett G (1978) Regression quantiles. *Econometrica* 46:33–50. <https://doi.org/10.2307/1913643>
- Korb K, Nicholson A (2004) Bayesian artificial intelligence, CRC Press, London
- Kuhn M, Wing J, Weston S, Williams A., Keefer C., Engelhardt A., Cooper T., Mayer Z, Kenkel B, Benesty M., Lescarbeau R., Ziem A., Scrucca L., Tang Y., Candan C. 2016. Caret: Classification and Regression Training. R package version 6.0-68. URL <http://CRAN.R-project.org/package=caret>. Accessed 25 Feb 2019
- Liaw A, Wiener M (2002) Classification and regression by randomForest. *R News* 2(3):18–22
- Martin AD, Quinn KM, Park J (2011) MCMCpack: Markov chain Monte Carlo in R. *J Stat Softw* 42(9):1–21 <http://www.jstatsoft.org/v42/i09/>
- McTaggart R, Daroczi G., Leung C. 2016. Quandl: API Wrapper for Quandl.com. R package version 2.8.0. <http://CRAN.R-project.org/package=Quandl>. Accessed 25 Feb 2019
- Meyer D, Dimitriadou E., Hornik K., Weingessel A., Leisch F. 2015. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.6-7. <http://www.CRAN.R-project.org/package=e1071>. Accessed 25 Feb 2019
- Meyer D, Leisch F, Hornik K (2003) The support vector machine under test. *Neurocomputing* 55(1–2):169–186. [https://doi.org/10.1016/S0925-2312\(03\)00431-4](https://doi.org/10.1016/S0925-2312(03)00431-4)
- Micheletti N, Foresti L, Robert S, Leuenberger M, Pedrazzini A, Jaboyedoff M, Kanevski M (2014) Machine learning feature selection methods for landslide susceptibility mapping. *Math Geosci* 46:33–57. <https://doi.org/10.1007/s11004-013-9511-0>
- Mohammadi H, Su L (2010) International evidence on crude oil price dynamics: applications of ARIMA-GARCH models. *Energy Econ* 32:1001–1008. <https://doi.org/10.1016/j.eneco.2010.04.009>
- Movagharnajad K, Mehdizadeh B, Banihashemi M, Kordkheili MS (2011) Forecasting the differences between various commercial oil prices in the Persian Gulf region by neural network. *Energy* 36:3979–3984. <https://doi.org/10.1016/j.energy.2011.05.004>
- Nagarajan R, Scutari M, Lèbre S (2013) Bayesian Networks in R with Applications in Systems Biology. Springer-Verlag, New York
- Narayan PK, Narayan S (2007) Modelling oil price volatility. *Energy Policy* 35:6549–6553. <https://doi.org/10.1016/j.enpol.2007.07.020>
- Park T, Casella G (2008) The Bayesian lasso. *J Am Stat Assoc* 103(482):681–686. <https://doi.org/10.1198/01621450800000337>
- Robert C, Casella G (2004) Monte Carlo statistical methods, 2nd edn. Springer-Verlag, New York
- Sadorsky P (2006) Modeling and forecasting petroleum futures volatility. *Energy Econ* 28:467–488. <https://doi.org/10.1016/j.eneco.2006.04.005>
- Scutari M (2010) Learning Bayesian networks with the bnlearn R package. *J Stat Softw* 35(3):1–22 <http://www.jstatsoft.org/v35/i03>
- Scutari M (2017) Bayesian network constraint-based structure learning algorithms: parallel and optimized implementations in the bnlearn R package. *J Stat Softw* 77(2):1–20. <https://doi.org/10.18637/jss.v077.i02>
- Sherwood B, Maidman A. 2016. rqPen: Penalized Quantile Regression. R package version 1.4. <http://CRAN.R-project.org/package=rqPen>. Accessed 25 Feb 2019
- Shin H, Hou T, Park K, Park CK, Choi S (2013) Prediction of movement direction in crude oil prices based on semi-supervised learning. *Decis Support Syst* 55:348–358. <https://doi.org/10.1016/j.dss.2012.11.009>
- Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B* 58:267–288. <https://doi.org/10.2307/2346178>
- Tibshirani R. 2011. Regression shrinkage and selection via the lasso: a retrospective. *73* (3) 273–282. DOI:<https://doi.org/10.1111/j.1467-9868.2011.00771.x>
- Trapletti A., Hornik K. 2015. tseries: Time Series Analysis and Computational Finance. R package version 0.10–34
- Ulrich J. 2016. TTR: Technical Trading Rules. R package version 0.23-1. <http://CRAN.R-project.org/package=TTR>. Accessed 25 Feb 2019
- Wang J, Pan HP, Liu FJ (2012) Forecasting crude oil price and stock price by jump stochastic time effective neural network model. *J Appl Math* 2012:1–15. <https://doi.org/10.1155/2012/646475>
- Wei Y, Wang YD, Huang DS (2010) Forecasting crude oil market volatility: further evidence using GARCH-class models. *Energy Econ* 32:1477–1484. <https://doi.org/10.1016/j.eneco.2010.07.009>
- Wu Y, Liu Y (2009) Variable selection in quantile regression. *Stat Sin* 19:801–817
- Xiong T, Bao YK, Hu ZY (2013) Beyond one-step-ahead forecasting: evaluation of alternative multi-step-ahead forecasting models for crude oil prices. *Energy Econ* 40:405–415. <https://doi.org/10.1016/j.eneco.2013.07.028>
- Yousefi S, Weinreich I, Reinartz D (2005) Wavelet-based prediction of oil prices. *Chaos, Solitons Fractals* 25(2):265–275. <https://doi.org/10.1016/j.chaos.2004.11.015>
- Yu L, Wang S, Lai KK (2008) Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Econ* 30:2623–2635. <https://doi.org/10.1016/j.eneco.2008.05.003>
- Zhang J, Zhang Y, Zhang L (2015) A novel hybrid method for crude oil price forecasting. *Energy Econ* 49:649–659. <https://doi.org/10.1016/j.eneco.2015.02.018>
- Zhang X, Lai KK, Wang SY (2008) A new approach for crude oil price analysis based on empirical mode decomposition. *Energy Econ* 30:905–918. <https://doi.org/10.1016/j.eneco.2007.02.012>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.