Technological Forecasting & Social Change xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

An ICA-based support vector regression scheme for forecasting crude oil prices

Liwei Fan [a,*], Sijia Pan [b], Zimin Li [c], Huiping Li [b]

[a] Business School, Hohai University, 8 Focheng West Road, Nanjing 211100, China
[b] College of Economics and Management, Nanjing University of Aeronautics and Astronautics, 29 Jiangjun Avenue, Nanjing 211106, China
[c] Offshore Oil Engineering (Qingdao) Company, 492 Lianjiang Road, Qingdao 266520, China

ARTICLE INFO

Article history:
Received 25 December 2015
Received in revised form 3 April 2016
Accepted 29 April 2016
Available online xxxx

Keywords:
Crude oil price
Forecasting
Independent component analysis
Support vector regression

ABSTRACT

The fluctuations of crude oil prices affect the economic growth of importing and exporting countries as well as regional security and stability. The intrinsic complex features of oil prices and the uncertainty in economic policy pose challenge on the accurate forecasting of crude oil prices. This paper employs independent component analysis (ICA) to analyze crude oil prices which are decomposed into several independent components corresponding to different types of influential factors affecting oil price. We also propose a novel ICA-based support vector regression scheme, namely ICA-SVR^2, for forecasting crude oil prices. The ICA-SVR^2 starts from the use of ICA to decompose oil price series into three independent components, which are respectively forecasted by SVR models. The forecasted independent components are then integrated together by developing a new SVR model with independent components as inputs for forecasting crude oil prices. Our experimental results show the usefulness of ICA in identifying the driving factors behind the fluctuations of crude oil prices. A comparative study between ICA-SVR^2 and other two models shows that ICA-SVR^2 is an effective tool in forecasting crude oil prices.

1. Introduction

Crude oil is one of the most actively traded commodities in the world (Alvarez-Ramirez et al., 2012). The large fluctuations of crude oil prices affect the economic growth of importing and exporting countries as well as regional security and stability (Wu and Zhang, 2014). Recent decades have seen the more frequent fluctuation of crude oil prices, which attracted the concern from both market participants and governmental regulators (Zhang, 2013). Undoubtedly, accurate oil price forecasting is of strategic significance in multiple aspects such as determining the timing for crude oil importing and ensuring economic security (Zhang and Wang, 2013).

The intrinsic complex features of oil prices and the uncertainty in economic policy pose big challenge on the accurate forecasting of crude oil prices (Bekirosa et al., 2015). Many scholars have thus contributed to develop novel methods and models for improving the accuracy of crude oil price forecasting. Fan and Li (2015) provided a relatively comprehensive review of major crude oil price forecasting models and found that artificial intelligence models (e.g. neural networks and support vector machines) had received increased attention. Recent methodological developments of artificial intelligence-based forecasting models can be found in Zhu and Wei (2013), Yu et al. (2014), Azadeh et al. (2012; 2015), Barunik and Krehlik (2016), Chen and Chen (2016), Mostafa and El-Masry (2016), and Oztekin et al. (2016).

In crude oil price forecasting, Jammazi and Aloui (2012) combined wavelet decomposition and artificial neural network to achieve better forecasting performance. He et al. (2012) showed the effectiveness of a wavelet decomposed ensemble model. Tang and Zhang (2012) developed a multiple wavelet recurrent neural network simulation model to analyze crude oil prices. Guo et al. (2012) proposed an improved support vector machine (SVM) model by using genetic algorithm to optimize the parameters. Zhang et al. (2015) proposed a hybrid method by combining SVM with ensemble empirical mode decomposition and particle swarm optimization models to improve the forecasting performance. Wang et al. (2016) recently proposed a Markov switching multifractal volatility model to forecast crude oil return volatility.

Of the existing methodological developments in crude oil price forecasting, the hybrid models, e.g. the Divide-and-Conquer (DAC) methodology, are found to be particularly useful owing to its potential for capturing the intrinsic features of oil price series (Fan and Li, 2015). The DAC methodology initiated by Professor Shouyang from Chinese Academy of Sciences follows the "decomposition and ensemble" principle (Lai, 2005), which is usually implemented by integrating several types of data analysis and forecasting techniques. Examples of some newly developed DAC models for forecasting crude oil prices are Tang et al. (2015) and Yu et al. (2014, 2015, 2016) and Tang et al. (2015). In the DAC models, the first step is to decompose the crude oil price series into several new data series, whose intrinsic features may be identified more easily than the original series. In this process, an issue is that the new data series may not be independent from each other, which brings difficulty in interpreting their economic implications and identifying the intrinsic features behind the fluctuation in crude oil prices.

* Corresponding author.
E-mail address: hhufanlw@163.com (L. Fan).

http://dx.doi.org/10.1016/j.techfore.2016.04.027
0040-1625/© 2016 Elsevier Inc. All rights reserved.

To resolve the issues, in this paper we propose a novel DAC scheme for forecasting crude oil prices by integrating independent component analysis (ICA) with support vector regression (SVR) techniques. ICA, a relatively new multivariate statistical analysis technique, is used to extract the key features (or components) embedded in crude oil price series. An attractive feature of ICA is that it will make the resulting components be independent from each other to a certain degree. While ICA as a feature extraction technique has been successfully applied in other fields like microarray classification (e.g. Fan et al., 2009, 2010), none of previous studies ever explored the use of ICA in analyzing crude oil prices series which should be treated as one main contribution of this paper. Through ICA, we decompose crude oil prices into three independent series which may reflect the influences of three types of factors on crude oil prices. Then we build individual SVR models for forecasting the three components, with which an integrated SVR model is constructed for forecasting crude oil prices. Our empirical analysis shows not only the usefulness of the proposed forecasting scheme but also the effectiveness of ICA in identify the driving factors behind the fluctuations in crude oil prices.

The rest of the paper is organized as follows. Section 2 introduces our ICA-based support vector regression scheme for forecasting crude oil prices. In Section 3, we propose a case study based on four datasets on weekly oil prices. Our modeling results show the effectiveness of ICA and the proposed scheme in forecasting crude oil prices. Section 4 concludes this study with possible future research directions.

2. Methodology

2.1. Independent component analysis

ICA was originally proposed for isolating independent source signals from linearly mixed signal (Jutten and Herault, 1991). Theoretically, it is a multivariate statistical technique to estimate independent components from the observed data by using high-order statistics (Comon, 1994). In the past two decades, many scholars have contributed to examine both theoretical and application aspects of ICA. In application, the use of ICA covers different areas such as microarray data classification (Fan et al., 2009, 2010), groundwater pumping analysis (Liu et al., 2015), power system disturbance identification (Ferreira et al., 2015), financial time series forecasting (Lu et al., 2009) and portfolio selection (Hitaj et al., 2015). Although ICA has gained popularity in different areas, its application to analyze and forecast crude oil price are still rare.

The fluctuation of crude oil prices could be driven by different factors such as economic situation and extreme events. Since these factors are likely to generate different influences on the crude oil prices, it is reasonable to separate them from each other to identify the main driving forces behind oil price fluctuations as well as achieve better forecasting performance. Since the process is akin to the decomposition of mixed signal, in this paper we propose to apply ICA to analyze crude oil prices and derive independent components which are linked to different categories of influential factors. Technically, the basic ICA model can be written as

$$X = AS \tag{1}$$

where $X = (x_1, x_2, \cdots x_n)^T$ denotes n by $m$observations on mixed signal (e.g. crude oil prices), $S = (s_1, s_2, \cdots s_n)$ denotes independent n by m unknown independent source signals (or influential factors), and A denotes a n by n mixing matrix. The purpose of ICA is to obtain the de-mixing matrix W (or A^{-1}) such that

$$y = WX \tag{2}$$

where y denotes the independent components estimated from the observed data.

The computation process in ICA is implemented through setting appropriate estimation principles and solving the resulting optimization models. The estimation principles help to ensure the derived source signals as independent as possible, and three commonly used ones are maximum likelihood, nongaussianity maximization, and mutual information minimization (Hyvärinen et al., 2001). Each principle will generate a specific objective function whose optimization enables the estimation of independent components. Various algorithms may be used to solve the optimization model, among which the FastICA algorithm is a popular and effective one owing to its theoretic strengths (Hyvärinen and Oja, 2000). In this paper, we follow the mutual information minimum principle and adopt FastICA algorithm to derive the independent components from the observations on crude oil prices.

2.2. Support vector regression model

Support vector machine as a novel machine learning technique has successfully been used to deal with high-dimensional nonlinear classification and regression (Vapnik and Cortes, 1995). When facing linearly inseparable situation in low dimensional space, support vector machine can be used to map the original dataset in low dimensional space to a new dataset in high dimensional space so that they become linearly separable (Gunn, 1998; Tay and Cao, 2001). In crude oil price forecasting, several scholars ever explored the usefulness of support vector regression (SVR), e.g. Xie et al. (2006), Khashman and Nwulu (2011) and Zhao et al. (2015). The process of constructing a SVR model is briefly described below.

Suppose that we have a training dataset $T = \{(x_1, y_1), \cdots (x_n, y_n)\}$, where x_i and y_i are respectively the observations on input and output variables. In SVR, the goal is to find a $f(x)$ that is as flat as possible while has most ε deviation from the observed targets y_i for all the training data (Smola and Scholkopf, 2004). To reach the purpose while ensure the feasibility of the resulting optimization problem, Vapnik and Cortes (1995) provide the following formulation:

$$\begin{aligned} &\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \zeta_i) \\ &\text{s.t.} \quad (w^T x_i + b) - y_i \le \varepsilon + \xi_i \\ &\quad y_i - (w^T x_i + b) \le \varepsilon + \zeta_i \\ &\quad \xi_i, \zeta_i \ge 0, i = 1, 2, \cdots, N \end{aligned} \tag{3}$$

where $\|w\|^2$ in objective function is the confidence range reflecting the generalization ability, ξ_i and ζ_i are slack variables that represent the upper and lower limits of allowable error, $\sum_{i=1}^{N}(\xi_i + \zeta_i)$ denotes the experimental risk reflecting the learning capacity of function, $\varepsilon > 0$ is an insensitive loss coefficient, and parameter $CC \ge 0$) is a penalty factor. Eq. (3) is a convex optimization problem which can be easily solved. More detailed explanation on the model can be found in Smola and Scholkopf (2004).

In SVR, the dual problem of Eq. (3) is often derived by using the Lagrange multiplier method, based on which a linear regression function can finally be constructed as

$$f(x) = w^T x + b = \sum_{i=1}^{N}(\hat{\alpha}_i - \alpha_i)K(x_i, x) + b \tag{4}$$

where $\alpha_i, \hat{\alpha}_i$ are Lagrange multipliers. Usually only a portion of the parameters$\alpha_i, \hat{\alpha}_i, b$ are nonzero and their corresponding samples are called support vectors.

The next step in constructing a satisfactory SVR model is to choose an appropriate kernel function that determines the algorithm form used (Gunn, 1998). When seeking the optimal function of the dual

Please cite this article as: Fan, L., et al., An ICA-based support vector regression scheme for forecasting crude oil prices, Technol. Forecast. Soc. Change (2016), http://dx.doi.org/10.1016/j.techfore.2016.04.027

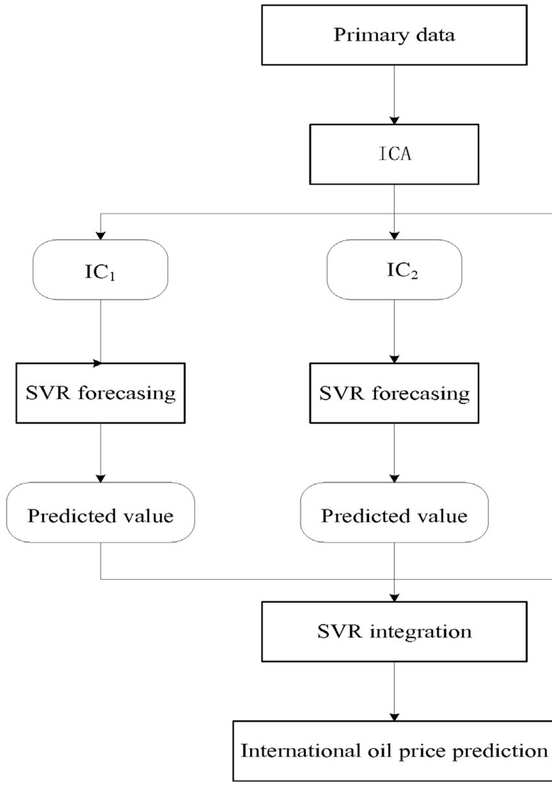

Fig. 1. Schematic description of ICA-SVR^2 forecasting model.

problem, we can replace the inner product operation in dual function by kernel function as follows:

$$K(x_i, x) = \langle \phi(x_i), \phi(x) \rangle. \tag{5}$$

The nonlinear regression function can then be translated as follows:

$$K(x_i x_j) = \exp\left\{ -\left\| (x_i - x_j)^2 \right\| / \sigma^2 \right\} \tag{6}$$

where σ is the width of kernel function. While there are different types of kernel function, in this paper we choose the radial basis kernel function as adopted by many earlier studies. The grid search method is then used to optimize the parameters of the SVR model constructed.

2.3. ICA-based SVR model for forecasting crude oil prices

An integrated forecasting model often combines two or more different forecasting methods to make use of their strengths and improve the forecasting performance (Lai, 2005). In view of the merits of ICA and SVR, in this paper we propose an ICA-based SVR scheme for analyzing and forecasting crude oil prices. Its key idea is to use ICA to decompose the original crude oil price series into several independent components, which are separately predicted by using SVR technique. The individual forecasting results are finally integrated by building another SVR model. As such, our model is referred to as ICA-SVR^2. Fig. 1 provides a schematic description of the ICA-SVR^2 model.

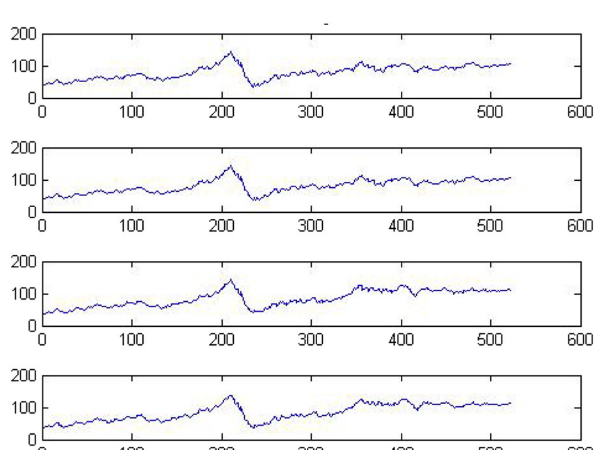

Fig. 2. Four international crude oil price series (unit: US $).

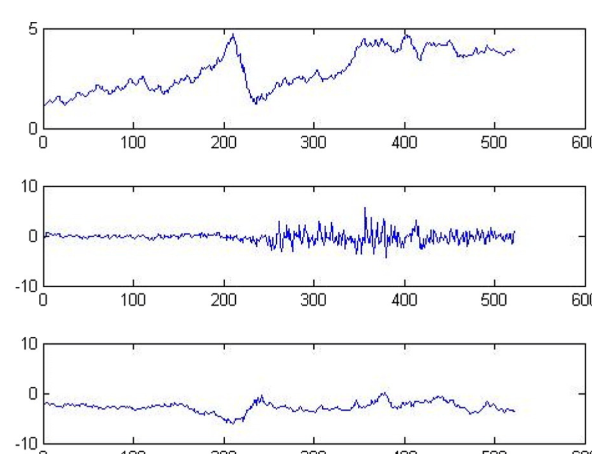

Fig. 3. Three independent components of international oil price series.

Please cite this article as: Fan, L., et al., An ICA-based support vector regression scheme for forecasting crude oil prices, Technol. Forecast. Soc. Change (2016), http://dx.doi.org/10.1016/j.techfore.2016.04.027

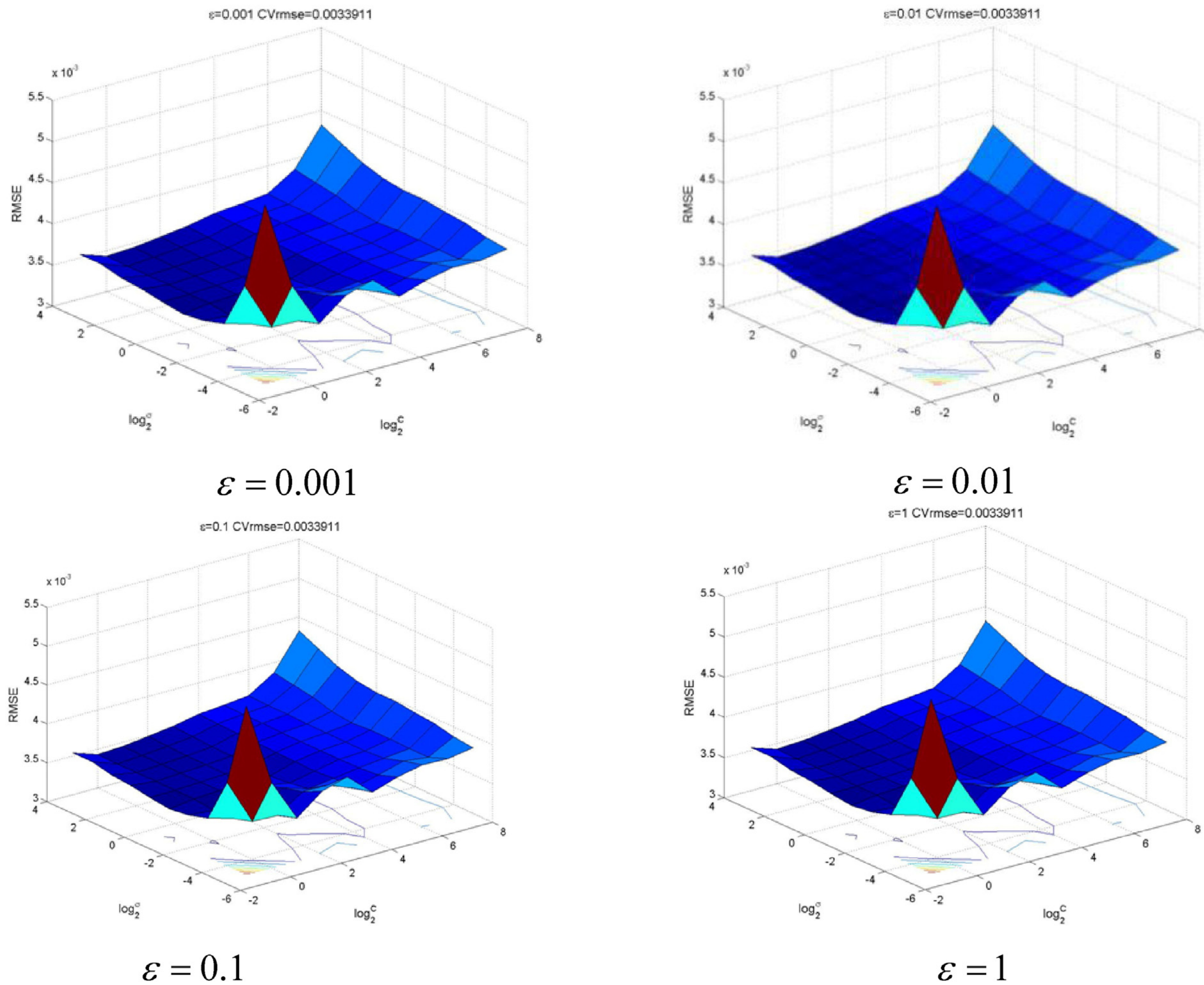

Fig. 4. Root mean square error of SVR forecasting model under different ε.

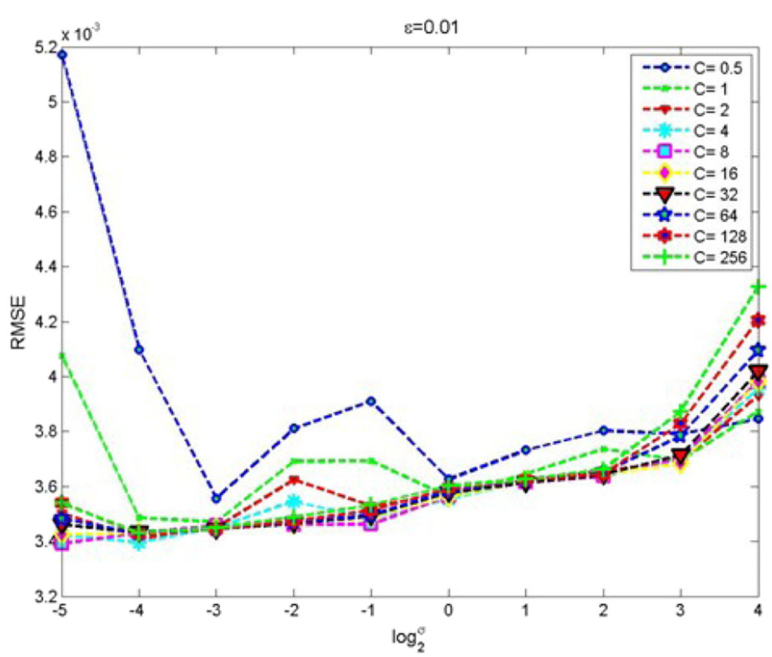

Fig. 5. The effect of σ on forecasting performance ($\varepsilon = 0.01$).

Please cite this article as: Fan, L., et al., An ICA-based support vector regression scheme for forecasting crude oil prices, Technol. Forecast. Soc. Change (2016), http://dx.doi.org/10.1016/j.techfore.2016.04.027

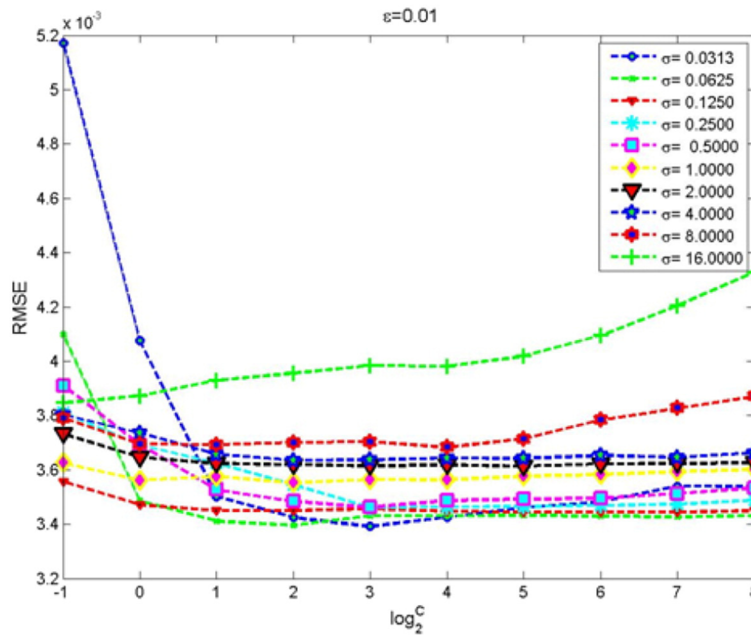

Fig. 6. The effect of C on forecasting performance ($\varepsilon = 0.01$).

In detail, the ICA-SVR[2] model consists of the following three steps:

Step 1. Use ICA to decompose the crude oil price time series into m independent component $\{IC_{1t}\}_{t=1}^{t=n}, \{IC_{2t}\}_{t=1}^{t=n}, \cdots \{IC_{mt}\}_{t=1}^{t=n}$.

Step 2. Develop SVR models for each of the independent components and do the forecasting one by one.

In Step 2, we need to seek the optimal parameters (C,ε,σ) for each independent component according to its data characteristics, based on which we can construct the component-dependent SVR models to forecast the independent components $\{IC_{i(t+1)}\}^{t=n+k}, i = (1,2,\cdots,m), k = (1,2,\cdots,l)$. In the process, the lag order p of oil price series is taken as an input parameter. Repeat the process until we get the estimate of the l-period.

Step 3. Build a SVR model to integrate the forecasting results of independent components.

In this step, we use the estimates of independent components as inputs to build a SVR model which is used to forecast crude oil prices. The final estimate $x_{i(n+k)}$ can be represented by

$$\hat{x}_{i(n+k)} = g\left(I\hat{C}_{1(i+k)}, .I\hat{C}_{2(i+k)}, \cdots, I\hat{C}_{m(t+k)}\right) \tag{9}$$

where $i = (1,2,\cdots,m), k = (1,2,\cdots,l)$.

3. Crude oil price forecasting results

3.1. Data and model evaluation criteria

We apply our proposed scheme to four commonly used crude oil price series, i.e. British Brent crude oil spot, American West Texas Intermediate (WTI) crude oil spot, Brent future exchanged in New York mercantile exchange (NYMEX) and WTI future. The weekly data from July 2, 2004 to June 27, 2014, which consists of 4 × 522 observations, are used in our study. All the data come from the website of US Energy Information Administration. Fig. 2 shows the trends of the four oil price series.

To evaluate the forecasting performance of the proposed model, we adopt the following three criteria, namely root mean square error (RMSE), mean absolute percentage error (MAPE) and direction change statistics (DS). RMSE and MAPE are often used to assess the forecasting accuracy of each model on average. As discussed in Shin et al. (2013), the direction of oil price is also important for assessing a model's ability in accurately forecasting the direction of crude oil price fluctuations. Therefore, in this paper we follow Shin et al. (2013) to choose the direction forecasting accuracy indicator DS for use. Their computational formula are respectively expressed by

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(T_t - P_t)^2}{n}} \tag{11}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\left(\frac{T_t - P_t}{T_t}\right)\right| \tag{12}$$

$$DS = \frac{100}{n}\sum_{t=1}^{n} d_t, d_t = \begin{cases} 1, (P_t - P_{t-1})(T_t - T_{t-1}) \geq 0 \\ 0, (P_t - P_{t-1})(T_t - T_{t-1}) < 0 \end{cases} \tag{13}$$

where T_t $(t = 1, 2, \cdots, n)$ is the actual value of the crude oil price on time t, P_t $(t = 1, 2, \cdots, n)$ is the predicted value, and n denotes the sample size of prediction values.

Table 1
Parameter settings of SVR forecasting models by independent components.

	C	ε	σ	No. of support vectors
IC_1	8	0.01	0.0313	31
IC_2	2	0.01	2	162
IC_3	128	0.01	0. 5	85

Table 2
Forecasting performance of three independent components.

	RMSE	MAPE	DS
IC_1	0.0108	0.0275	99.99%
IC_2	0.1093	0.2083	87.12%
IC_3	0.0229	0.0131	99.99%

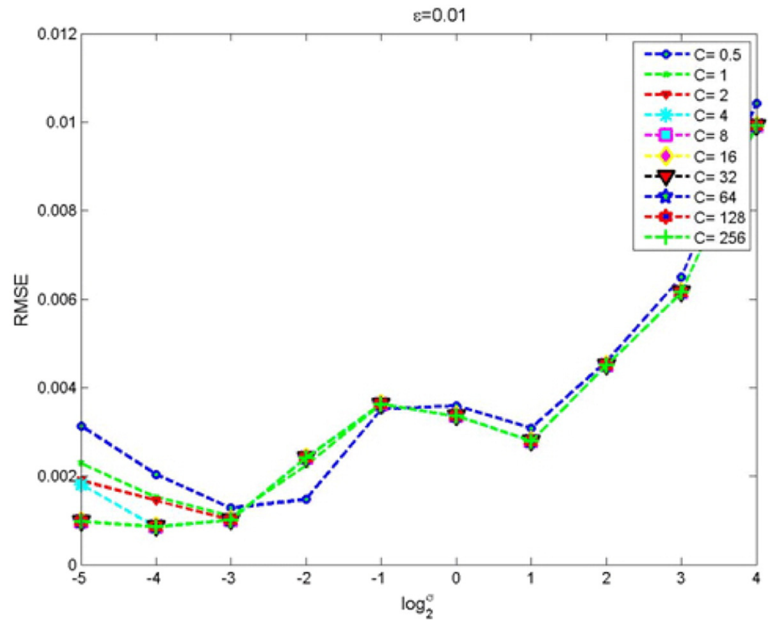

Fig. 7. The effect of σ on forecasting performance ($\varepsilon = 0.01$).

3.2. Estimation of independent components

One primary task in applying ICA to estimate the independent components of crude oil price series is to determine the number of independent components. In this paper, we apply the Kaiser method and the cumulative contribution rate principle to determine it. First, the covariance matrix and the eigenvalues of four international oil price series are computed. Observing the eigenvalues, we find that they decrease quickly and the last eigenvalue is less than unity. Following the cumulative contribute rate principle, we find that the cumulative contribution rate of the first two eigenvalues are 98.8%. However, if the number of independent components is too small, we may ignore some influential factors and weaken the advantage of ICA over conventional multivariate statistical techniques. Based on the considerations, we finally determine the number of independent components as three.

Next, we apply the Fast-ICA algorithm to estimate the three independent components (ICs) hidden in the original oil price series. The results obtained are shown in Fig. 3. Clearly, the three independent components show different volatility characteristics that may be relevant to different influential factors. IC_1 shows the characteristics of low frequency and high fluctuation. It may represent the impact of extreme events like major economic or political events (Zhang et al., 2009). The most prominent parts can be observed in 2008–2009 when oil price fluctuate greatly during the economic crisis. IC_2 displays mild but more frequent fluctuation characteristics. It may represent the impacts of some emergency events like natural disaster and geopolitics conflicts. Since the influence of this sort of factors on international oil price is relatively small, we treat it as high-frequency short-term fluctuation. IC_3 fluctuates slowly but may reflect the long-term trend of oil price series.

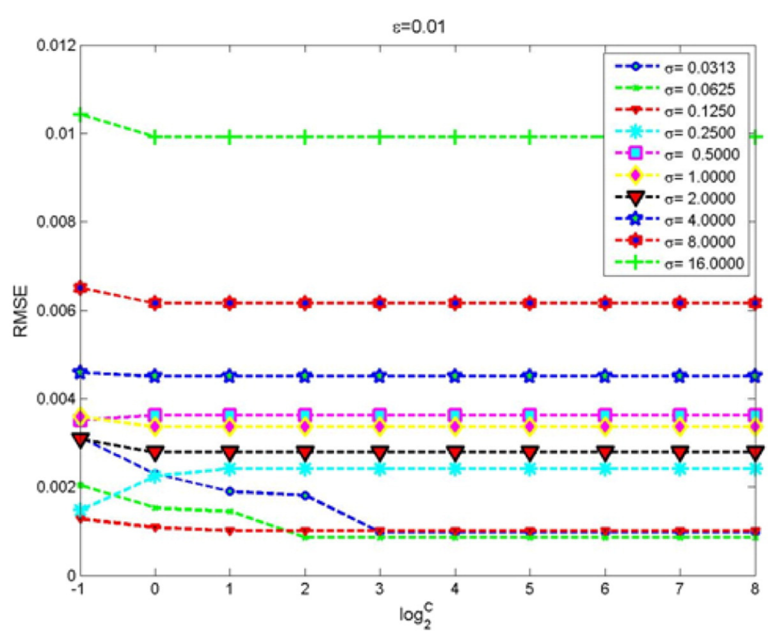

Fig. 8. The effect of C on forecasting performance ($\varepsilon = 0.01$).

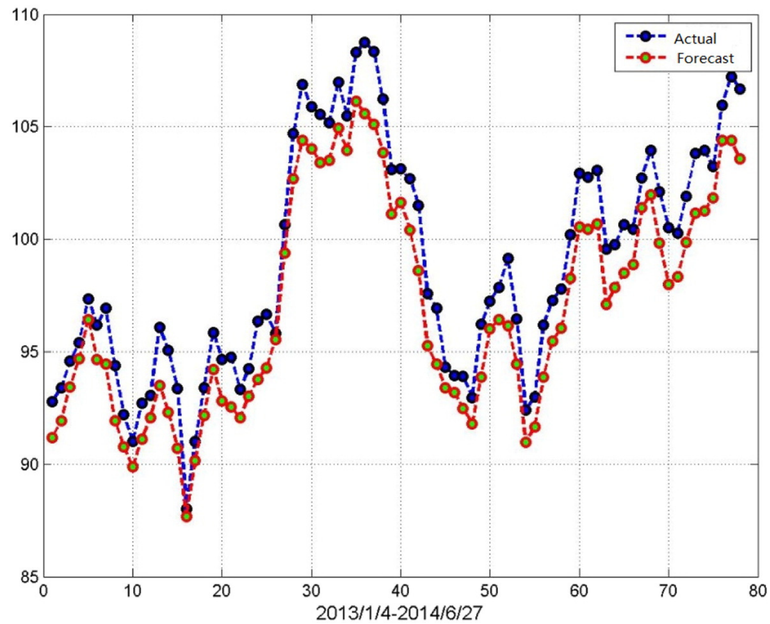

Fig. 9. Forecasting result of WTI crude oil price by ICA-SVR^2 (Dollars/barrel).

3.3. Forecasting independent components

Using the three independent components estimated in Section 3.2, we construct three SVR forecasting models separately. Note that the weekly crude oil price data is often relevant the data in previous period, we set the lag order $p = 1$ in our experiments. That is to say, we use the actual data for period t as input to forecast the independent components for period $t + 1$. Regarding the choice of kernel function, this paper chooses Gaussian radial basis kernel function kernel.

We use IC_1 as an example to show the process of searching for optimal parameters in model development. The weekly observations from 2 July 2004 to 28 December 2012 are used for training and the weekly observations from 4 January 2013 to 27 June 2014 are used for testing. When using grid searching method to determine the parameters, we set the search ranges of ε and C as $\{1,0.1,0.01,0.001\}$ and $(2^{-1},2^{8})$, respectively. The range of kernel parameter σ is $(2^{-5},2^{4})$, and the step size is set as 1. Hence there are $4 \times 10 \times 10$ parameters for (C,ε,σ). The training samples are classified into five equal sections. Four of them are used for training and the remaining one is used for validation.

We first determine the hyper-parameter ε. Fig. 4 shows the values of RMSE under different ε. Comparing the four sub-figures in Fig. 4, we may find that the forecasting performance is not sensitive to ε. Under different ε, the minimum training errors keep no changes, i.e., RMSE $= 0.0039$. Hence, this paper adopts the default setting of $\varepsilon = 0.01$.

Figs. 5 and 6 show the prediction results under different values of C and σ, respectively. It can be seen from Fig. 5 that under a fixed C the model does not show good convergence effect with the increase of σ. Fig. 6 shows that when kernel parameter σ keeps no change, the forecasting performance converge gradually and then diverge with the increase of C. By consolidating the results, we finally select $C = 8$ and $\sigma = 0.0313$ for building the SVR model for forecasting IC_1.

Table 1 provides the settings of parameters in constructing SVR models for forecasting the three independent components. Using the SVR models with the parameter values, we forecast the three independent components respectively and the results of forecasting performance are displayed in Table 2.

It can be seen from Table 2 that the SVR model has better forecasting performance for the items of low-frequency trend and important events (i.e. IC_1 and IC_3). However, the forecasting performance for high-frequency short-term volatility item (i.e. IC_2) is not as good as other two components. This could be explained by the fact that the component may be linked to speculation activity and some unexpected factors like weather and strike that are unpredictable.

3.4. Forecasting crude oil prices

We take the three independent components and WTI crude oil price series for the same period of time to build a SVR model for forecasting crude oil prices. The training sample is used to construct the model. In forecasting crude oil prices, the forecasting results of three independent components are treated as inputs. In selecting model parameters, we still use the grid search procedure to determine the parameter procedure of determining the best parameter values. Figs. 7 and 8 show the effects of σ and C on forecasting performance.

Since the forecasting performance of the SVR model is insensitive to ε, we still takes the default value 0.01 for use. As shown in Fig. 7, the model achieves the best forecasting performance when $\sigma = 0.0625$ for a fixed C. Fig. 8 shows that the forecasting performance shows a convergence trend with the increase of C for a fixed σ. The final optimal parameters of the SVR forecasting model are then set as $C = 4, \sigma = 0.0625, \varepsilon = 0.01$.

Using the ICA-SVR^2 forecasting model constructed, we take the three independent components as predictors to forecast the WTI crude oil prices for the testing period. Fig. 9 shows the actual and forecasted oil price series. It can be found that the forecasting results match the actual oil price series quite well, which shows that the proposed model performs well in forecasting the trend of crude oil prices.

In order to evaluate the effectiveness and robustness of the ICA-SVR^2 model, we change the share of the training dataset in the entire dataset and compares the forecasting performance under different scenarios.

Table 3
Forecasting result of ICA-SVR^2 when the portion of training dataset varies.

Proportion of training set	RMSE	MAPE	DS(100%)
60%	3.3425	0.0351	81.21%
70%	2.1999	0.0289	83.33%
80%	1.6251	0.0167	88.31%
90%	1.5271	0.0150	92.31%

Please cite this article as: Fan, L., et al., An ICA-based support vector regression scheme for forecasting crude oil prices, Technol. Forecast. Soc. Change (2016), http://dx.doi.org/10.1016/j.techfore.2016.04.027

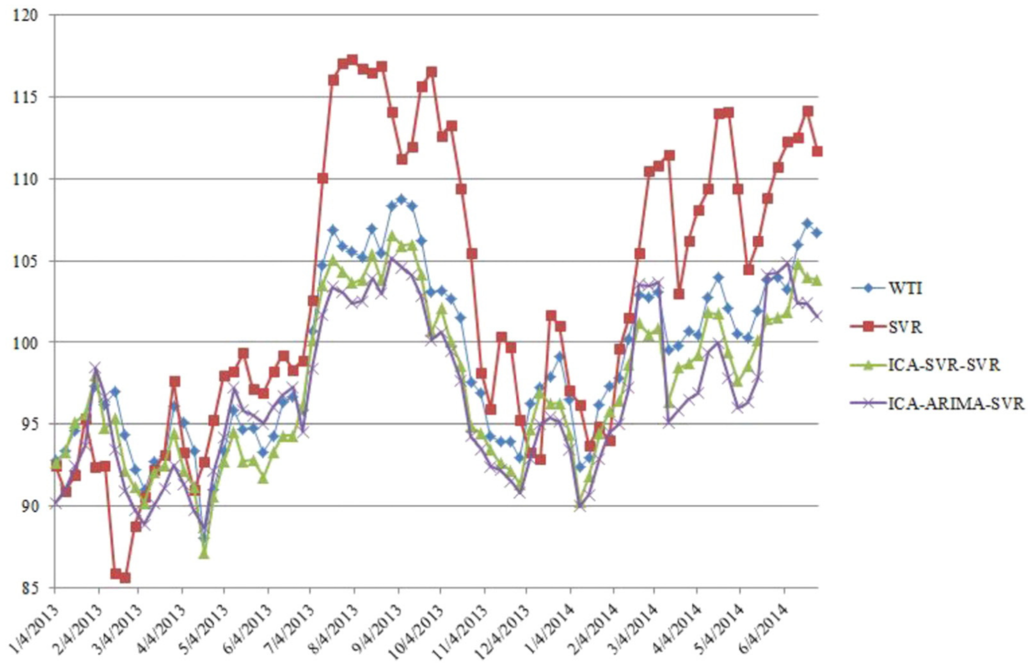

Fig. 10. Crude oil price forecasting results by three methods (Dollars/barrel).

Table 3 shows the results obtained. Not surprisingly, when more data are used to build the model, the forecasting performance gradually improves. When the share of training dataset is as small as 60%, the forecasting performance is still satisfactory. In particular, the accuracy rate of direction forecasting results is still above 80%.

3.5. Comparison with other models

In order to further evaluate the forecasting performance of ICA-SVR^2 in a more comprehensive way, in this paper we also build a single SVR model and ICA-ARIMA-SVR model and compare their forecasting performance with the ICA-SVR^2 model. The single SVR model is directly constructed from the WTI crude oil price series. In the ICA-ARIMA-SVR model, ICA is still used to decompose the oil price series into three independent components, which are respectively forecasted by ARIMA method. Based on the independent components and the WTI crude oil price series, we construct a SVR model to forecast crude oil prices with independent components as model inputs. Fig. 10 shows the forecasting results by three different models. The actual WTI oil price series are also included for comparison purpose. It can be seen that the forecasting results by ICA-ARIMA-SVR and ICA-SVR^2 models match the actual oil price series quite well. However, the single SVR model seems to have poor forecasting performance.

Table 4 shows the results of RMSE, MAPE and DS for different models. Clearly, both ICA-ARIMA-SVR and ICA-SVR^2 have better forecasting performance than singpaore SVR model. It might be an indication that the use of ICA can improve the performance of crude oil price forecasting significantly. Comparing ICA-ARIMA-SVR with ICA-SVR^2, we find that the latter has better forecasting performance.

Table 4
Forecasting deviation comparison of three methods.

	RMSE	MAPE	DS (100%)
SVR	3.9394	0.0537	64.94%
ICA-ARIMA-SVR	1.8855	0.0189	87.01%
ICA-SVR^2	1.6251	0.0167	88.31%

4. Conclusions

This paper proposes to use ICA to analyze crude oil prices and develops an ICA-based SVR (ICA-SVR^2) scheme for forecasting crude oil prices. ICA is helpful to identify the hidden factors behind the fluctuations of international crude oil prices. The proposed ICA-SVR^2 model starts from the use of ICA to decompose four international oil price series into three independent components. Then we build three SVR models to forecast the three independent components, respectively. Based on the three independent components as well as the WTI crude oil price series, we construct one integrated SVR model to forecast crude oil prices. Our experimental results show that the integration of ICA with SVR can improve the forecasting accuracy of SVR significantly. Another advantage of ICA is that the independent components obtained may shed insights on understanding the driving forces behind crude oil prices.

Despite the usefulness of the proposed ICA-SVR^2 model, it has inevitably some limitations. First, this paper only uses the radial basis kernel function for use in constructing the SVR models. Other types of kernel functions may be explored and compared with the radial basis kernel function in the ICA-SVR^2 model. Second, in our experimental study we only compare the ICA-SVR^2 model with ICA-ARIMA and single SVR models. It is therefore worthwhile including more models and conducting a more comprehensive comparison between different models. Third, this paper only explores the usefulness of ICA as a preprocessing procedure for improving the forecasting performance of SVR. Further research may be carried out to investigate the potential for integrating ICA with other newly developed models for forecasting crude oil prices.

Acknowledgements

This work is financially supported by the National Natural Science Foundation of China (nos. 71203055 & 71433003).

References

Alvarez-Ramirez, J., Rodriguez, E., Martina, E., Ibarra-Valdez, C., 2012. Cyclical behavior of crude oil markets and economic recessions in the period 1986–2010. Technol. Forecast. Soc. Chang. 79, 47–58.

Azadeh, A., Asadzadeh, S.M., Mirseraji, G.H., Saberi, M., 2015. An emotional learning-neuro-fuzzy inference approach for optimum training and forecasting of gas consumption estimation models with cognitive data. Technol. Forecast. Soc. Chang. 91, 47–63.
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 (2), 421–430.
Barunik, J., Krehlik, T., 2016. Combining high frequency data with non-linear models for forecasting energy market volatility. Expert Syst. Appl. 55, 222–242.
Bekirosa, S., Gupta, R., Paccagnini, A., 2015. Oil price forecasting and economic uncertainty. Econ. Lett. 132, 125–128.
Chen, T.L., Chen, F.Y., 2016. An intelligent pattern recognition model for supporting investment decisions in stock market. Inf. Sci. 346/347, 261–274.
Comon, P., 1994. Independent component analysis, a new concept? Signal Process. 36 (3), 287–314.
Fan, L., Li, H., 2015. Volatility analysis and forecasting models of crude oil prices: A review. Int. J. Glob. Energy Issues 38, 5–17.
Fan, L., Poh, K.L., Zhou, P., 2009. A sequential feature extraction approach for naïve classification of microarray data. Expert Syst. Appl. 36, 8188–8192.
Fan, L., Poh, K.L., Zhou, P., 2010. Partition-conditional ICA for Bayesian classification of microarray data. Expert Syst. Appl. 37, 9919–9923.
Ferreira, D.D., de Seixas, J.M., Cerqueira, A.S., 2015. A method based on independent component analysis for single and multiple power quality disturbance classification. Electr. Power Syst. Res. 119, 425–431.
Gunn, S.R., 1998. Support vector machines for classification and regression. ISIS technical report (Available at: http://users.ecs.soton.ac.uk/srg/publications/pdf/SVM.pdf).
Guo, X., Li, D., Zhang, A., 2012. Improved support vector machine oil price forecast model based on genetic algorithm optimization parameters. AASRI Procedia 1, 525–530.
He, K., Yu, L., Lai, K.K., 2012. Crude oil price analysis and forecasting using wavelet decomposed ensemble model. Energy 46, 564–574.
Hitaj, A., Mercuri, L., Rroji, E., 2015. Portfolio selection with independent component analysis. Financ. Res. Lett. 15, 146–159.
Hyvärinen, A., Oja, E., 2000. Independent component analysis: algorithms and applications. Neural Netw. 13, 411–430.
Hyvärinen, A., Karhunen, J., Oja, E., 2001. Independent Component Analysis. John Wiley & Sons, New York.
Jammazi, R., Aloui, C., 2012. Crude oil price forecasting: experimental evidence from wavelet decomposition and neural network modeling. Energy Econ. 34, 828–841.
Jutten, C., Herault, J., 1991. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. Signal Process. 24 (1), 1–10.
Khashman, A., Nwulu, N.I., 2011. Support vector machines versus back propagation algorithm for oil price prediction. Lect. Notes Comput. Sci 6677, 530–538.
Lai, K., 2005. Crude oil price forecasting with TEI@ I methodology. J. Syst. Sci. Complex. 18 (2), 145–166.
Liu, H.J., Hsu, N.S., Yeh, W.W.G., 2015. Independent component analysis for characterization and quantification of regional groundwater pumping. J. Hydrol. 527, 505–516.
Lu, C.J., Lee, T.S., Chiu, C.C., 2009. Financial time series forecasting using independent component analysis and support vector regression. Decis. Support. Syst. 47, 115–125.
Mostafa, M.M., El-Masry, A.A., 2016. Oil price forecasting using gene expression programming and artificial neural networks. Econ. Model. 54, 40–53.
Oztekin, A., Kizilaslan, R., Freund, S., Iseri, A., 2016. A data analytic approach to forecasting daily stock returns in an emerging market. Eur. J. Oper. Res. 253, 697–710.
Shin, H., Hou, T., Park, K., Park, C.K., Choi, S., 2013. Prediction of movement direction in crude oil prices based on semi-supervised learning. Decis. Support. Syst. 55, 348–358.
Smola, A.J., Scholkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14, 199–222.
Tang, M., Zhang, J., 2012. A multiple adaptive wavelet recurrent neural network model to analyze crude oil prices. J. Econ. Bus. 64, 275–286.
Tang, L., Dai, W., Yu, L., Wang, S., 2015. A novel CEEMD-based EELM ensemble learning paradigm for crude oil price forecasting. Int. J. Inf. Technol. Decis. Mak. 14, 141–169.
Tay, F.E., Cao, L., 2001. Application of support vector machines in financial time series forecasting. Omega 29, 309–317.
Vapnik, V., Cortes, C., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297.
Wang, Y., Wu, C., Yang, L., 2016. Forecasting crude oil market volatility: a Markov switching multifractal volatility approach. Int. J. Forecast. 32, 1–9.
Wu, G., Zhang, Y.J., 2014. Does China factor matter? An econometric analysis of international crude oil prices. Energ Policy 72, 78–86.
Xie, W., Yu, L., Xu, S., Wang, S., 2006. A new method for crude oil price forecasting based on support vector machines. Lect. Notes Comput. Sci 3994, 444–451.
Yu, L., Dai, W., Tang, L., 2016. A novel decomposition ensemble model with extended extreme learning machine for crude oil price forecasting. Eng. Appl. Artif. Intell. 47, 110–121.
Yu, L., Wang, Z., Tang, L., 2015. A decomposition–ensemble model with data-characteristic-driven reconstruction for crude oil price forecasting. Appl. Energy 156, 251–267.
Yu, L., Zhao, Y., Tang, L., 2014. A compressed sensing based AI learning paradigm for crude oil price forecasting. Energy Econ. 46, 236–245.
Zhang, Y.J., 2013. Speculative trading and WTI crude oil futures price movement: an empirical analysis. Appl. Energy 107, 394–402.
Zhang, Y.J., Wang, Z.Y., 2013. Investigating the price discovery and risk transfer functions in the crude oil and gasoline futures markets: some empirical evidence. Appl. Energy 104, 220–228.
Zhang, X., Yu, L., Wang, S., Lai, K.K., 2009. Estimating the impact of extreme events on crude oil price: An EMD-based event analysis method. Energy Econ. 31 (5), 768–778.
Zhang, J.L., Zhang, Y.J., Zhang, L., 2015. A novel hybrid method for crude oil price forecasting. Energy Econ. 49, 649–659.
Zhao, L., Cheng, L., Wan, Y., Zhang, H., Zhang, Z., 2015. A VAR-SVM model for crude oil price forecasting. Int. J. Glob. Energy Issues 38 (1–3), 126–144.
Zhu, B., Wei, Y., 2013. Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. Omega 41 (3), 517–524.

Liwei Fan is an assistant professor at the School of Business, Hohai University. Her research interest is energy modeling and forecasting, with particular interest in the use of data mining techniques in this field. Dr. Fan has published articles in internationally referred journals such as Expert Systems with Applications, Energy Policy, Journal of Air Transport Management, etc.

Sijia Pan is a postgraduate student at the College of Economics and Management, Nanjing University of Aeronautics and Astronautics, China. Her main research interest is energy economics and policy.

Zimin Li is an engineer in the Offshore Oil Engineering (Qingdao) Company, China National Offshore Oil Corporation. His research interest is crude oil market and oil price forecasting.

Huiping Li was a postgraduate student at the College of Economics and Management, Nanjing University of Aeronautics and Astronautics, China. Her main research interest is oil market analysis.

Please cite this article as: Fan, L., et al., An ICA-based support vector regression scheme for forecasting crude oil prices, Technol. Forecast. Soc. Change (2016), http://dx.doi.org/10.1016/j.techfore.2016.04.027