Songklanakarin J. Sci. Technol.
40 (1), 75-78, Jan. - Feb. 2018

http://www.sjst.psu.ac.th

Original Article

Predicting prices of agricultural commodities in Thailand using combined approach emphasizing on data pre-processing technique

Thoranin Sujjaviriyasup*

Department of Logistics Engineering, School of Engineering,
University of the Thai Chamber of Commerce, Din Daeng, Bangkok, 10400 Thailand

Received: 5 May 2016; Accepted: 14 October 2016

Abstract

In this research, a combined approach emphasizing on data pre-processing technique is developed to forecast prices of agricultural commodities in Thailand. The future prices play significant role in decision making to cultivate crops in next year. The proposed model takes ability of MODWT to decompose original time series data into more stable and explicit subseries, and SVR model to formulate complex function of forecasting. The experimental results indicated that the proposed model outperforms traditional forecasting models based on MAE and MAPE criteria. Furthermore, the proposed model reveals that it is able to be a useful forecasting tool for prices of agricultural commodities in Thailand.

Keywords: combined approach, agricultural commodity price, discrete wavelet transform, ARIMA, support vector regression

1. Introduction

Thailand's agricultural sector (Board of Investment of Thailand, 2014) plays a crucial role in Thailand's economy, which not only generates several billion baht a year in economic value but also is a pivotal part of the Thai way of life. Furthermore, Thailand is one of the world's top ten both producer and exporter (Food and Agricultural Organization of the United Nations, 2015) of agricultural commodities. Consequently, Thai agriculturists play important role in Thailand's agricultural market as a primary investor. Regarding investment of those agriculturists, the future prices (Brooks *et al.*, 2013; Chen & Chang, 2015; Xiong *et al.*, 2015; Yang & Zhang, 2013) of agricultural commodities are needed to realize before making a critical decision on investment as well as risk management. Thus, an accuracy of future prices is very challenging in order to support a decision making on whether to invest this commodity or other competitive crops. In recent years, numerous combined approaches (Panapakidis & Dagoumas, 2016; Shrivastava & Panigrahi, 2014; Zhu & Wei, 2013) have been proposed to forecast future prices in many fields of science. A combined approach, including data pre-processing techniques, is an interesting approach, due to high performance and easy to find literature as well. This approach emphasizes on a preliminary process on datasets by decomposing an original time series into more stationary and regular subseries that are more explicit to analyze by filtering out the irrelevant feature of the dataset. The empirical results (Joo & Kim, 2015; Kriechbaumer *et al.*, 2014; Shayeghi *et al.*, 2015; Yu *et al.*, 2016; Zhang *et al.*, 2015) of combined approaches including data pre-processing techniques indicated that more stable subseries and the most informative training data are able to improve the quality of the data as well as accuracy of prediction. In this research, a combined approach of Maximal Overlap Discrete Wavelet Transform (MODWT) and SVR is proposed to forecast future prices of Thailand's agricultural commodities. Furthermore, the proposed model is compared with traditional forecasting models based on mean absolute percentage error (MAPE), mean absolute error (MAE), and relative improvement (RI).

*Corresponding author
Email address: kwannan@yahoo.com

2. Materials and Methods

Data used in this research are prices (Office of Agricultural Economics, 2016) of agricultural commodities in Thailand, which are obtained from Office of Agricultural Economics (OAE) as shown in Figure 1.

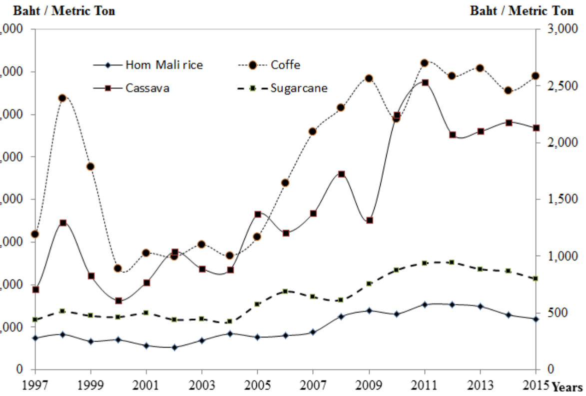

Figure 1. Prices of agricultural commodities in Thailand.

2.1 Autoregressive integrated moving average model

Due to its high performance of linear forecasting, the autoregressive integrated moving average models have exploited in many linear problems of time series forecasting. The model is generally referred to as an ARIMA (p, d, q) model with the that has the form as Equation 1.

$$\left(1-\sum_{i=1}^{p}\varphi_i B^i\right)(1-B)^d(y_t-\mu)=\left(1-\sum_{j=1}^{q}\theta_j B^j\right)\varepsilon_t \tag{1}$$

where y_t and ε_t are the actual value and random error at time period *t*, respectively. *B* is the backward shift operator; *P* and *q* are referred to orders of autoregressive integrated moving average model, which are integers as well. Referencing Equation 1, the future value of a variable is formulated from linear function of several past observations and random error. Since the best ARIMA model relies on past observation update, the ARIMA model is used to stand for ARIMA (p, d, q) in this research. The criterion used to select the proper model is AICc as shown in Equation 2.

$$\mathrm{AICc}=\mathrm{AIC}+\frac{2k(k+1)}{n-k-1} \tag{2}$$

$$\mathrm{AIC}=2k-2\ln(L) \tag{3}$$

where *n* and *k* are sample size and parameters, respectively. *L* is the maximized value of the likelihood function for the model.

2.2 Support vector regression model

The support vector regression model is extended from support vector machine to solve regression problems. The mathematical expression is described as Equation 4.

$$f(x_i)=\sum_{i=1}^{T}(\alpha_i-\alpha_i^*)K(x,x_i)+b \tag{4}$$

where α_i and α_i^* are the so-called Lagrange multipliers, *b* is a scalar threshold, *K(x, x_i)* is kernel function. The kernel functions are the most used for SVR model, which are defined as follows:

Linear: $K(x,x_i)=x^T x_i$,

Polynomial: $K(x,x_i)=(\gamma x^T x_i+r)^p$

Radial basis: $K(x,x_i)=\exp\left(-\gamma\|x-x_i\|^2\right)$,

Sigmoid: $K(x,x_i)=\tanh(\gamma x^T x_i+r)$

where γ, *r*, and *p* are kernel parameters.

2.3 Discrete wavelet transform

Wavelet transform is a signal processing algorithm, which is developed from Fourier transform. With regard to time series, the wavelet transform is adopted to decompose time series into more stationary and regular subseries. The mathematical expression of discrete wavelet transform is described by Equation 5.

$$W_x(m,n,\psi)=a_0^{-m/2}\int_{-\infty}^{+\infty}f(t)\psi^*(a_0^{-m}t-nb_0)dt \tag{5}$$

where $\psi^*(t)$ denotes the complex conjugate of ψ, $a_0>1$, $b_0\in R$, *n* and *m* are integer numbers. One of several discrete wavelet transforms is MODWT that exploited in this research due to its ability to deal with the circular shift effect. Furthermore, the MODWT is also known as the zero-shift or time-invariant DWT. The MODWT expression is illustrated as Equations 6 and 7.

$$\tilde{V}_{j,t}=\sum_{l=0}^{L-1}\tilde{g}_l\tilde{V}_{j-1,t-2^{j-1}l \bmod N},\ t=0,1,2,\ldots,N-1 \tag{6}$$

$$\tilde{W}_{j,t}=\sum_{l=0}^{L-1}\tilde{h}_l\tilde{V}_{j-1,t-2^{j-1}l \bmod N},\ t=0,1,2,\ldots,N-1 \tag{7}$$

where $\tilde{V}_{j,t}$ and $\tilde{W}_{j,t}$ are scaling coefficients and wavelet coefficients, respectively, the $\tilde{h}_l$ and $\tilde{g}_l$ are the MODWT wavelet and scaling filter, letting $\tilde{V}_{0,t}=X_t$. The variable *L* is the length of either wavelet filter ($\tilde{h}_l$) or scaling filter ($\tilde{g}_l$).

2.4 Proposed model

The main objective of the proposed model is to develop complex models by taking ability of MODWT to decompose an original time series data into more stationary and regular subseries, and the SVR models to model the forecasting function. Regarding the MODWT technique of this hybrid model, the original input data is decomposed based on Daubechies wavelet and only the first level of decomposition into more stationary and regular subseries before using SVR models. The forecast value is the sum of all predicted value as shown in Figure 2. The algorithm of the proposed model is presented in Figure 3.

2.5 Cross – validation

All forecasting models are evaluated its performance based on MAE and MAPE for cross-validation. All criteria equations are shown as Equations 8 and 9.

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n} \tag{8}$$

$$\mathrm{MAPE} = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i| / y_i}{n} \times 100 \tag{9}$$

3. Results and Discussion

Due to heteroskedasticity problem, the problem often is seen in the prices of agricultural commodities. Subsequently, McLeod-Li test is used to determine the presence of conditional heteroskedascity. Given statistical results, they can conclude that heteroskedasticity problem is not seen in these prices of Thailand's agricultural commodities. Thus, the ARIMA model is an appropriate model to forecast these prices due to constant unconditional variance. Based on MAE and MAPE criteria, the summary of all forecasting performances is presented in Table 1.

Table 1. Summary of all forecasting performances.

Datasets	ARIMA		SVR		Proposed model	
	MAE	MAPE	MAE	MAPE	MAE	MAPE
Rice	1475.99	11.48%	393.60	10.39%	1174.07	9.72%
Cassava	0.38	19.03%	0.42	22.06%	0.25	13.34%
Sugarcane	68.88	8.57%	77.38	9.48%	63.72	7.61%
Coffee	11.56	17.75%	12.47	19.42%	9.24	14.70%

2.6 Statistical analysis

Data are shown as mean±SE. Data were analyzed by one-way analysis of variance (ANOVA) followed by Turkey posthoc test or Mann-Whitney U test (GraphPad Prism version 5.01, GraphPad Software, USA). A p-value ≤ 0.05 was considered statistically significant.

Referencing Table 1, ARIMA model outperforms SVR model approximately 75% of all cases, this indicate that ARIMA model should to be the first tool for predicting prices of Thailand's agricultural commodities (e.g., cassava, sugarcane, and coffee). Nonetheless, the SVR model is still to be a useful tool in order to forecast the Thailand's rice price. Although the ARIMA model provides good accuracy, the proposed model reveals that it is able to provide the lowest errors based on MAE and MAPE criteria for all cases. This evidence demonstrates that data pre-processing technique can enhance forecasting performance of SVR model that is a traditional forecasting model to achieve more accuracy.

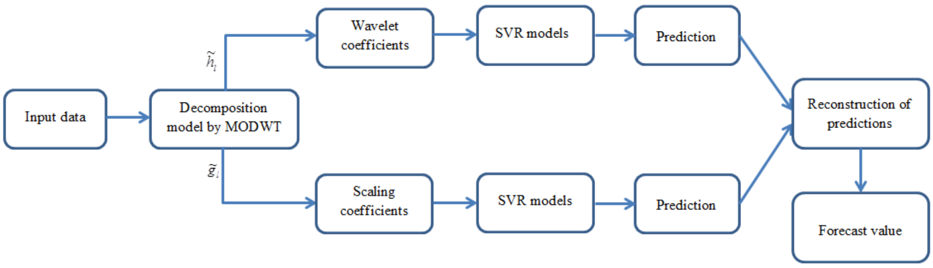

Figure 2. Flowchart of hybridization of MODWT and SVR.

For *m* equal to 2 **to** a termination criterion **do**
  For *n* equal to 2 **to** a termination criterion **do**
    For 70% of the past observations **to** the observation before the last observation **do**
      1. Decompose original input data into subseries by using equation (6) – (7).
      2. Rearrange the scaling coefficients into *m* columns of the scaling coefficients.
      3. Rearrange the wavelet coefficients into *n* columns of the wavelet coefficients.
      4. The first column to the column before the last column of each coefficient and the last column of each coefficient are exploited as input and target data, respectively
      5. Select the kernel function to formulate the model formulation of ε – SVR to predict the future coefficients as equation (4).
      6. Predict the future coefficients of wavelet and scaling coefficients.
      7. Sum the predicted wavelet and scaling coefficients as forecast value.
    End
  End
End

Figure 3. Algorithm of the proposed model.

In other words, the subseries obtained from MODWT provide the most informative training data and more explicit to forecast by SVR model. In order to confirm the forecasting performance improvement of the proposed model, the RI is used to evaluate its improvement that is described as Equation 10.

$$\mathrm{RI} = \frac{\left(y - y_{\mathrm{reference}}\right)}{y_{\mathrm{reference}}} \times 100 \tag{10}$$

The summary of the relative improvement of the proposed model compared with other forecasting models is presented in Table 2. Referencing the relative improvement results in Table 2, they indicate that the proposed model is able to improve accuracy of forecast value at least 6.45% to 40.48% of relative improvement in the cases of MAE and MAPE.

Table 2. Summary of the relative improvement of the proposed model.

Datasets	ARIMA		SVR	
	MAE	MAPE	MAE	MAPE
Rice	20.46%	15.33%	15.75%	6.45%
Cassava	34.21%	29.90%	40.48%	39.53%
Sugarcane	7.49%	11.20%	17.65%	19.73%
Coffee	20.07%	17.18%	25.90%	24.30%

4. Conclusions

Concerning all empirical results, the proposed model reveals that it provides the lowest errors than traditional forecasting models (i.e., ARIMA model and SVR model) based on MAE and MAPE criteria. Although SVR model cannot provide better results than ARIMA model in almost all cases, the hybridization of MODWT and SVR models outperforms the ARIMA model. In other words, the data pre-processing technique can enhance the SVR model to achieve more high accuracy of forecasting. This evidence indicates that the proposed model can be a promising tool for forecasting prices of agricultural commodities in Thailand in order to support decision making of Thailand's agriculturists. On the other hands, the ARIMA model demonstrates that it is still be a meaningful tool in order to predict prices of Thailand's agricultural commodities (i.e., cassava, sugarcane, and coffee) in the case of traditional forecasting models. Meanwhile, the SVR model illustrates that it is a suitable tool to forecast the Thailand's rice price in the case of single forecasting models.

References

Board of Investment of Thailand. (2014). Thailand: Feeding the world. Retrieved from http://www.boi.go.th/tir/issue/201408_24_8/42.htm

Brooks, C., Prokopczuk, M., & Wu, Y. (2013). Commodity futures prices: More evidence on forecast power, risk premia and the theory of storage. *The Quarterly Review of Economics and Finance, 53*, 73-85.

Chen, Y–L., & Chang, Y–K. (2015). Investor structure and the informational efficiency of commodity futures prices. *International Review of Financial Analysis, 42*, 358-367.

Food and Agricultural Organization of the United Nations. (2015). FAOSTAT. Retrieved from http://faostat3.fao.org/download/T/TP/E

Joo, T. W., & Kim, S. B. (2015). Time series forecasting based on wavelet filtering. *Expert Systems with Applications, 42*, 3868-3874.

Kriechbaumer, T., Angus, A., Parsons, D., & Casado, M. R. (2014). An improved wavelet–ARIMA approach for forecasting metal prices. *Resources Policy, 39*, 32-41.

Office of Agricultural Economics. (2016). Prices of agricultural commodities in Thailand. Retrieved from http://www.oae.go.th/oae_report/price/price_month.php

Shayeghi, H., Ghasemi, A., Moradzadeh, M., & Nooshyar, M. (2015). Simultaneous day-ahead forecasting of electricity price and load in smart grids. *Energy Conversion and Management, 95*, 371-384.

Shrivastava, N. A., & Panigrahi, B. K. (2014). A hybrid wavelet-ELM based short term price forecasting for electricity markets. *International Journal of Electrical Power and Energy Systems, 55*, 41-50.

Panapakidis, I. P., & Dagoumas, A. S. (2016). Day-ahead electricity price forecasting via the application of artificial neural network based models. *Applied Energy, 172*, 132-151.

Xiong, T., Li, C., Bao, Y., Hu, Z., & Zhang, L. (2015). A combination method for interval forecasting of agricultural commodity futures price. *Knowledge-Based Systems, 77*, 92-102.

Yang, L., & Zhang, D. (2013). Can futures price be a powerful predictor? Frequency domain analysis on Chinese commodity market. *Economic Modelling, 35*, 264-271.

Yu, L., Dai, W., & Tang, L. (2016). A novel decomposition ensemble model with extended extreme learning machine for crude oil price forecasting. *Engineering Applications of Artificial Intelligence, 47*, 110-121.

Zhu, B., & Wei, Y. (2013). Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. *Omega, 41*, 517-524.

Zhang, J–L., Zhang, Y–J., & Zhang, L. (2015). A novel hybrid method for crude oil price forecasting. *Energy Economics, 49*, 649-659.