Knowledge-Based Systems 77 (2015) 92–102

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A combination method for interval forecasting of agricultural commodity futures prices

Tao Xiong [a], Chongguang Li [a], Yukun Bao [b,*], Zhongyi Hu [b], Lu Zhang [a]

[a] College of Economics and Management, Huazhong Agricultural University, Wuhan 430070, PR China
[b] School of Management, Huazhong University of Science and Technology, Wuhan 430074, PR China

ARTICLE INFO

Article history:
Received 12 August 2014
Received in revised form 18 November 2014
Accepted 5 January 2015
Available online 9 January 2015

Keywords:
Interval-valued data
Agricultural commodity futures price forecasting
Vector error correction model (VECM)
Multi-output support vector regression (MSVR)
Econometric modeling

ABSTRACT

Accurate interval forecasting of agricultural commodity futures prices over future horizons is challenging and of great interests to governments and investors, by providing a range of values rather than a point estimate. Following the well-established ''linear and nonlinear'' modeling framework, this study extends it to forecast interval-valued agricultural commodity futures prices with vector error correction model (VECM) and multi-output support vector regression (MSVR) (abbreviated as VECM–MSVR), which is capable of capturing the linear and nonlinear patterns exhibited in agricultural commodity futures prices. Two agricultural commodity futures prices from Chinese futures market are used to justify the performance of the proposed VECM–MSVR method against selected competitors. The quantitative and comprehensive assessments are performed and the results indicate that the proposed VECM–MSVR method is a promising alternative for forecasting interval-valued agricultural commodity futures prices.

1. Introduction

In recent years, agricultural commodity futures markets in China have witnessed a massive growth. With an increasing product variety and deepening liquidity pools, the agricultural commodity futures markets in China are playing an increasing important role in serving the global financial market and the national economy. In 2012, the three most active agricultural futures in the world by contract volume were happening in China—the cotton, sugar, and rubber contracts, which traded over 139 million, 128 million and 104 million, respectively. Out of the top 10 agricultural contracts by volume, 7 were Chinese. Accurately measuring and forecasting the dynamics of agricultural commodity futures prices is inevitably an important component not only in trade risk management but also in price speculation. Agricultural commodity futures price forecasting is considered as a challenging task due to the fact that the prices are highly volatile, complex and dynamic, and is thus of great interests to finance researchers, market practitioners, and policymakers.

An extensive literature investigation reveals that it is not difficult to find that great research efforts have been expended to explore the underlying dynamic of futures prices and develop models suitable for forecasting futures prices, including stock index futures [1–5], oil futures [6–13], gold futures [14,15], and metal futures [16]. But there exists a much smaller body of research examining the predictability of agricultural commodity futures prices [17–19]. Furthermore, an important point to note from past studies [1–19] mentioned above is their preoccupation with point forecasting rather than interval one.

Our focus in this study is on interval forecasting of agricultural commodity futures prices. An interval forecasting of futures prices has the advantage of taking into account the variability and/or uncertainty so as to reduce the amount of random variation relative to that found in classic single-valued futures prices time series (e.g., future closing price) [20]. Since the work of Moore [21] and Moore and Moore [22], interval analysis and forecasting has attracted particular attention in various fields, particularly in finance market [23–26] and energy market [27,28]. As Hu and He [23] pointed out, the interval forecasts of stock price are superior to the traditional point forecasts in terms of the overall lower mean error and higher average accuracy ratio. As shown in García-Ascanio and Maté [27], moreover, the interval-valued time series (ITS) forecasting method as a potential tool will lead to a reduction in risk when making power system planning and operational decisions. To date, a variety of ITS forecasting methods has been developed (see [25] for a recently survey), among which the traditional statistical techniques, including interval exponential smoothing methods [29], vector autoregressive (VAR) model [27], and vector error

* Corresponding author. Tel.: +86 27 87558579; fax: +86 27 87556437.
E-mail addresses: yukunbao@hust.edu.cn, y.bao@ieee.org (Y. Bao).

http://dx.doi.org/10.1016/j.knosys.2015.01.002
0950-7051/

correction model (VECM) [30], have been widely used due to their statistical properties. However, traditional statistical techniques can provide good predictions only when ITS under study are linear and stationary. But ITS may unfortunately appear nonlinear and non-stationary due to the intrinsic complexity and volatility of ITS (e.g., interval-valued agricultural commodity futures prices). In order to overcome the limitations of the traditional statistical techniques and capture the nonlinear pattern exhibited in ITS considered, machine learning techniques have recently attracted much attentions, among which interval multi-layer perceptrons (iMLP) [28] and multi-output support vector regression (MSVR) [24] have shown excellent nonlinear modeling capability for ITS in real world. Although both traditional statistical techniques and machine learning techniques have achieved success in their own linear or nonlinear domains, it is difficult in practice to determine when an ITS is generated by either a linear or nonlinear process, and thus which particular technique is more suitable than another for a specific ITS forecasting task. Actually, ITS in real world rarely appears linear or nonlinear pattern and usually contains both patterns. Thus, a novel method, which can simultaneously handle linear and nonlinear patterns exhibited in ITS, is required for such tough forecasting tasks such as, in this present study, interval-valued agricultural commodity futures prices forecasting.

Fortunately, these difficult forecasting tasks can be partially solved by using combined "linear and nonlinear" modeling framework. Given the real time series exhibiting both linear and nonlinear patterns, it is very difficult in practice to construct a single model which is best in all situations. In this regard, Zhang [31] proposed a hybrid autoregressive integrated moving average (ARIMA) and artificial neural network (ANN) model for time series forecasting to take advantage of the unique strength of ARIMA and ANN in linear and nonlinear modeling, respectively. Generally speaking, there are three main steps involved in the combined "linear and nonlinear" modeling framework, i.e., linear modeling, nonlinear modeling, and summation. First of all, a linear modeling technique is used to model the linear part, then the residuals from the linear model will contain only the nonlinear relationship. Secondly, by modeling residuals obtained in the previous step using a nonlinear modeling technique, nonlinear relationships can be discovered. Finally, these prediction results from the linear and nonlinear modeling techniques in the previous steps are summed to generate an aggregated output, which can be seen as the final prediction results for the original time series. Since the work of Zhang [31], the combined "linear and nonlinear" modeling framework has been recently well-established in finance market [32,33], tourism management [34], and energy market [35]. However, the aforementioned studies [31–35] focus only on single-valued time series forecasting. Following the thought of Zhang [31], Maia et al. [36] proposed a hybrid method (namely ARIMA–ANN) that combines both ARIMA and ANN for interval-valued time series forecasting. However, the hybrid method in [36] was applied to forecast the lower and upper bound series of interval-valued time series respectively, without considering the possible interrelations (e.g., cointegration between the low and high of the agricultural commodity futures prices in this study) that are presented among themselves, which has been criticized in [25].

Motivated by the thought of Zhang [31], and in particular, to remedy the mentioned above shortcomings in [36], this study proposes an improved combination method (namely VECM–MSVR) for ITS forecasting in agricultural commodity futures market by using VECM and MSVR to model and forecast the low and high prices series of interval-valued agricultural commodity futures prices simultaneously, taking into account the cointegration (the statistical evidences are given in Section 3.1). The MSVR, which is proposed to solve the problem of nonlinear regression estimation for multiple variables [37] and has been successfully used for ITS forecasting [24], is a generalization of the standard SVR. The reason for using MSVR in the proposed combination method is that MSVR can fit and forecast the interval bounds of interval-valued time series simultaneously due to its inherent multi-output structure. For comparison purposes, five counterparts, including VECM [30], single-output support vector regression (SSVR), MSVR [24], ARIMA–MSVR, and ARIMA–ANN developed by Maia et al. [36], are selected as benchmarks. The daily interval-valued futures prices of cotton and corn from Chinese futures market are used as the experimental data series for the purpose of validation. The experimental results are judged as follows. We examine whether the out-of-sample forecasts generated by the VECM–MSVR are more accurate than and preferable to those generated by the benchmark methods for an interval-valued agricultural commodity futures prices, employing statistical criteria such as the goodness of forecast measure and the accuracy competing forecasts test.

Our contributions could be outlined as follows. First, as was mentioned above, although extensive mounts of methods [1–12,14–19] have been developed for futures prices modeling and forecasting, most of them rely only on single-valued futures price series. The interval forecasting of futures prices, particularly agricultural commodity futures prices, has not been widely explored. Therefore, this study attempts to originally investigate the possibility of forecasting the interval-valued agricultural commodity future prices series over short and long horizons using a novel combination method. Second, by introducing VECM and MSVR, the well-established "linear and nonlinear" time series modeling framework can be extended to deal with ITS forecasting, which can fully take the advantages of the unique strength of VECM and MSVR in linear and nonlinear modeling, respectively. Third, the most published literature on agricultural commodity futures prices forecasting focuses on international markets, and on the US market in particular [17–19]. There is limited research on the predictability of Chinese agricultural commodity futures market. So, we hope this study would fill this gap. The fourth contribution is straightforward to provide the empirical evidence on the interval-valued agricultural commodity futures prices forecasting with real-world data from Chinese futures market.

The paper unfolds as follows. Section 2 illustrates the data representation of an interval-valued agricultural commodity futures price series and discusses the VECM, MSVR, and proposed VECM–MSVR method in details. Section 3 illustrates the research design on data processing and preliminary analysis, performance measurement criteria, methodological implementation, input selection, and experimental procedure. Following that, in Section 4, the experimental results are discussed. Finally, Section 5 summarizes and concludes this work.

2. The proposed VECM–MSVR method with interval-valued agricultural commodity futures prices

This section presents the overall formulation process of the proposed VECM–MSVR for interval-valued agricultural commodity futures price series forecasting. First, the data representation of interval-valued agricultural commodity futures price series is illustrated. Then, the VECM, MSVR, and the proposed VECM–MSVR for the obtained ITS are formulated in detail.

2.1. Constructing interval-valued agricultural commodity futures prices

An interval-valued variable, $\mathbf{X}$, is a variable defined for all of the elements i of a set E, where $\mathbf{X}_t = \{[X_t^L, X_t^U]^T : X_t^L, X_t^U \in \mathbb{R}, X_t^L \leqslant X_t^U\}$, $\forall t \in E$. The particular value of $\mathbf{X}$ for the ith element can be denoted by the interval's lower and upper bounds, $\mathbf{X}_t = [X_t^L, X_t^U]^T$. Table 1