



Forecasting value of agricultural imports using a novel two-stage hybrid model



Yi-Shian Lee^{a,b,*}, Wan-Yu Liu^c

^a Department of Industrial Engineering and Management, National Chiao Tung University, Hsinchu 300, Taiwan, ROC

^b Research Center for Psychological and Educational Testing, National Taiwan Normal University, Taipei 106, Taiwan, ROC

^c Department of Tourism Information, Aletheia University, New Taipei City 251, Taiwan, ROC

ARTICLE INFO

Article history:

Received 23 July 2013

Received in revised form 24 February 2014

Accepted 27 March 2014

Keywords:

Value of agricultural imports

GM(1,1)

Genetic programming

Residual signs

Residual series

ABSTRACT

Agricultural imports are becoming increasingly important in terms of their impact on economic development. An accurate model must be developed for forecasting the value of agricultural imports since rapid changes in industry and economic policy affect the value of agricultural imports. Conventionally, the ARIMA model has been utilized to forecast the value of agricultural imports, but it generally requires a large sample size and several statistical assumptions. Some studies have applied nonlinear methods such as the GM(1,1) and improved GM(1,1) models, yet neglected the importance of enhancing the accuracy of residual signs and residual series. Therefore, this study develops a novel two-stage forecasting model that combines the GM(1,1) model with genetic programming to accurately forecast the value of agricultural imports. Moreover, accuracy of the proposed model is demonstrated based on two agricultural imports data sets from the Taiwan and USA.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Since agricultural development is critical to the economic development of every country, agricultural issues are of global concern. Governments must devise viable economic policies to avoid unnecessary costs that are incurred with increasing agricultural imports. For example, after joining the World Trade Organization (WTO) in 2002, Taiwan signed the Economic Cooperation Framework Agreement (ECFA) in 2012 for reducing commercial barriers with China, drastically changing the value of agricultural imports. Since economic forecasting in the agricultural sector is critical to agricultural business planning and economic policy making, a high-precision forecasting approach must be designed to evaluate agricultural imports to enable policy makers to implement effective policies concerning agricultural imports and enhance economic development.

Relevant literature includes using various forecasting approaches to forecast agricultural demand (Lambert and Cho, 2008). Multiple linear regression and Box–Jenkins models (Agrawal, 2003; Lambert and Cho, 2008) are two conventional statistical methods. However, those approaches may be inaccurate

when data sets are small and nonlinear, as well as fail to meet certain statistical assumptions (Lee and Tong, 2011b; Pao, 2009). Hence, the forecasting accuracy of traditional statistical methods often varies under real-life condition (Yang et al., 2009). With the development of advances in machine-learning methods, some algorithms such as artificial neural network (ANN) and genetic algorithms (GAs), have been utilized in agricultural forecasting. For example, Jutras et al. (2009) adopted the ANN to predict the morphological parameters of street trees and found that the ANN can yield robust and precise results. Yang et al. (2009) combined principal component analysis and ANN to predict the population of the paddy stem borer (*Scirpophaga incertulas*), indicating that their proposed model outperformed other models. Ou (2012) proposed an improved forecasting model that combined improved GM(1,1) (IGM(1,1)) applied in modeling original time series and GAs applied in estimating the parameters of IGM(1,1), and demonstrated that the proposed model outperformed other models. Despite yielding satisfactory results for real-world data sets, the above methods have certain limitations. For instance, the hidden layers in ANN are difficult to explain, and the relationship between the independent and dependent variables cannot be expressed as a clear mathematical equation (Lee and Tong, 2011b). Moreover, the high precision of the above approaches depends on the sample sizes and the parameter settings that are determined by a trial and error approach. Using neural network-based models to construct an optimal network model is often criticized, owing to the

* Corresponding author at: Research Center for Psychological and Educational Testing, National Taiwan Normal University, Taipei 106, Taiwan, ROC.

E-mail address: bill.net.tw@yahoo.com.tw (Y.-S. Lee).

lack of openness and shift of emphasis towards training the network model (Srinivasan, 2008). Since data on agricultural imports are generally few and nonlinear, they may not yield accurate forecasting results when conventional statistical methods are applied.

Nonlinear or small-size time-series data sets are handled using approaches such as fuzzy theory, grey model (GM), and genetic programming (GP). The observations (real numbers) of fuzzy time series in a certain period are converted as discrete fuzzy sets (Egrioglu et al., 2011a). The procedure of fuzzy time series consists of three stages: fuzzification, determination of fuzzy relations and defuzzification (Song and Chissom, 1993). Some studies have attempted to increase forecasting accuracy by developing fuzzy-based approaches. For instances, Egrioglu et al. (2011a) determined an appropriate number of fuzzy clusters by using the Gustafson–Kessel fuzzy clustering algorithm and, later, determined the length of intervals of fuzzy time by using an optimization technique (2011b). Despite the applicability of the fuzzy-based approach to small data sets, determining an appropriate length of intervals based on different algorithms may expend a considerable amount of time. As useful in forecasting problems (Ou, 2012; Lee and Tong, 2011a; Yin and Tang, 2013; Pao et al., 2012; Chang et al., 2013), GM is often used in forecasting when data sets contain more than four samples (Wu et al., 2013). GM can generally be represented as $GM(g, h)$, where g and h denote the order and number of variables in constructing the GM, respectively. For example, $GM(1, 1)$ represents the first-order single-variable GM, and has been used to forecast agricultural output (Ou, 2012). To enhance the accuracy of $GM(1, 1)$ in the construction of agricultural demand values (including the value of agricultural imports/exports), some studies have modified $GM(1, 1)$ models (Ou, 2012). Although capable of yielding accurate forecasting results, the modified $GM(1, 1)$ belong to the GM system in order to obtain values of necessary parameters. However, few studies have improved the residual time-series data of the $GM(1, 1)$ with a machine-learning approach. Recently, some hybrid forecasting models have been proposed to improve the performance, which can be achieved using only a single forecasting method (Zhou and Hu, 2008; Pai and Lin, 2005; Aladag et al., 2009; Wang et al., 2012; Yolcu et al., 2013; Khashei and Bijari, 2012). For instances, Khashei and Bijari (2012) forecasted time series data by using probabilistic neural networks with feed-forward neural networks. Yolcu et al. (2013) performed time series forecasting by using linear and nonlinear ANN model. A criticism of ANN is the difficulty to explain the layers and neurons in its hidden-layer. Moreover, those studies have ignored the importance of residual-sign estimator. According to some studies (Hsu and Chen, 2003; Hsu, 2003; Lee and Tong, 2011a), the accuracy of the estimator of residual signs can influence the performance of a forecasting model. Moreover, using a complex residual equation to obtain the forecast residual values makes it difficult to use the hybrid model.

GP is an approach for evolving the functions that performs well in the defined problems (Koza, 1992) and constructs a forecasting model by using the symbolic regression method. The intelligence scheme can automatically extract knowledge from data sets and construct the model without defining related problems. The approach used in this paper is based on GP, owing to that GP often performs better than conventional statistical methods, in terms of forecasting accuracy. Although the performances of all forecasting models depend on the quality of the data set, these models differ in the ability to mine the inherent relationships in the data set. Most real-world data sets are nonlinear and time-dependent. GP is a relatively easy means of constructing mathematical models since no specialized knowledge. In some modeling time series applications, GP performs well in small data sets. For instance, based on a multi-level genetic programming (MLGP) approach, Forouzanfar et al. (2012) developed a transport energy demand forecasting model

(training set: 35 samples from year 1968 to 2002; testing set: 3 sample size which from year 2003 to 2005), which is more accurate than other models. By using a GP approach, Lee et al. (1997) designed an electric power demand forecasting model (training set: 20 samples from year 1961 to 1980; testing set: 10 samples from year 1981 to 1990), which is more accurate than the conventional regression model. Moreover, while developing the classification model, Lee and Tong (2012) predict the transfer efficiency of photovoltaic systems by using a GP-based model; the classification model outperforms other models on small photovoltaic data sets. Some studies (Huang et al., 2006; Muttill and Lee, 2005) demonstrated that GP can perform well even in small data sets.

This study develops a novel two-stage forecasting model that first utilizes $GM(1, 1)$ to forecast original data based on the advantage of being applied to small data sets, and then uses GP to forecast the residual signs and residual series of $GM(1, 1)$ based on the advantage of adopting symbolic regression to model complex data sets, to increase its accuracy in forecasting the value of agricultural imports. Analysis results demonstrate that the proposed model is easily applied in practice and performs well in modeling time-series data sets. The rest of this paper is organized as follows. Section 2 examines the feasibility of improving the grey forecasting model, which includes $GM(1, 1)$, to forecast the original data sets. The ability to use GP in order to forecast the residual signs and residual series of $GM(1, 1)$ is examined as well. Section 3 then presents two data sets to demonstrate the application of the proposed model, which is compared with other models. Conclusions are finally drawn in Section 4, along with recommendations for future research.

2. Methodology

2.1. $GM(1, 1)$ forecasting model

The $GM(1, 1)$ has been utilized in agriculture (Ou, 2012) and high-tech industry (Hsu, 2003; Hsu and Wang, 2007; Wang et al., 2011). $GM(1, 1)$ usually requires only four or more data points (Hsu, 2009) to construct a forecasting model. $GM(1, 1)$ is constructed as follows.

The general procedure for constructing a $GM(1, 1)$ is given as follows.

Collect an original non-negative time-series data sequence,

$$w^{(0)} = [w^{(0)}(1), w^{(0)}(2), \dots, w^{(0)}(n)], \quad n \geq 4 \quad (1)$$

where n is the total number of periods, and $w^{(0)}(n)$ is the observation that is associated with the n th time period.

The technique applies the accumulated generating operator (AGO) to $w^{(0)}$ to obtain an accumulated data sequence, as follows.

$$\begin{aligned} w^{(1)} &= \left(w^{(0)}(1), \sum_{m=1}^2 w^{(0)}(m), \dots, \sum_{m=1}^n w^{(0)}(m) \right) \\ &= (w^{(1)}(1), w^{(1)}(2), \dots, w^{(1)}(n)), \end{aligned} \quad (2)$$

where $w^{(0)}(1)$ equals $w^{(1)}(1)$.

$GM(1, 1)$ is constructed using the following grey differential equation.

$$w^{(0)}(k) + a \times s^{(1)}(k) = u, \quad k = 2, 3, \dots, n \quad (3)$$

where a , u , and $s^{(1)}(k)$ represent the development coefficient, grey input, and background value, respectively. Notably, $s^{(1)}(k)$ is obtained by applying the mean operator to $w^{(1)}$, as follows.

$$s^{(1)}(k) = \frac{w^{(1)}(k) + w^{(1)}(k-1)}{2}, \quad k = 2, 3, \dots, n. \quad (4)$$

Download English Version:

<https://daneshyari.com/en/article/84358>

Download Persian Version:

<https://daneshyari.com/article/84358>

[Daneshyari.com](https://daneshyari.com)