Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER Applied Mathematics and Computation 164 (2005) 885–912
APPLIED MATHEMATICS AND COMPUTATION
www.elsevier.com/locate/amc

Model identification of ARIMA family using genetic algorithms

Chorng-Shyong Ong [a], Jih-Jeng Huang [a], Gwo-Hshiung Tzeng [b,c,*]

[a] *Department of Information Management, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 106, Taiwan*
[b] *Institute of Management of Technology and Institute of Traffic and Transportation College of Management, National Chiao Tung University, 1001 Ta-Hsueh Road, Hsinchu 300, Taiwan*
[c] *Kai Nan University, No. 1, Kai-Nan Road, Luchu, Taoyuan 338, Taiwan*

Abstract

ARIMA is a popular method to analyze stationary univariate time series data. There are usually three main stages to build an ARIMA model, including model identification, model estimation and model checking, of which model identification is the most crucial stage in building ARIMA models. However there is no method suitable for both ARIMA and SARIMA that can overcome the problem of local optima. In this paper, we provide a genetic algorithms (GA) based model identification to overcome the problem of local optima, which is suitable for any ARIMA model. Three examples of times series data sets are used for testing the effectiveness of GA, together with a real case of DRAM price forecasting to illustrate an application in the semiconductor industry. The results show that the GA-based model identification method can present better solutions, and is suitable for any ARIMA models.

Keywords: ARIMA; Stationary; SARIMA; Genetic algorithms; Model identification

* Corresponding author.
E-mail address: ghtzeng@cc.nctu.edu.tw (G.-H. Tzeng).

0096-3003/$ - see front matter
doi:10.1016/j.amc.2004.06.044

1. Introduction

ARIMA is the method first introduced by Box–Jenkins [1] to analyze stationary univariate time series data, and has since been used in various fields. The generalized form of ARIMA can be described as

$$\phi(B)\Phi(B^s)(1-B)^d(1-B)^D Y_t = \theta(B)\Theta(B^s)Z_t, \tag{1}$$

where B denotes the backward shift operator; d and D denote the non-seasonal and seasonal order of differences taken, respectively; $\phi(B)$, $\theta(B)$, $\Phi(B)$ and $\Theta(B)$ are polynomials in B and B^s of finite order p and q, P and Q, respectively, and usually abbreviated as SARIMA $(p,d,q)(P,D,Q)_s$. When there is no seasonal effect, a SARIMA model reduces to pure ARIMA(p,d,q), and when the time series data set is stationary a pure ARIMA reduces to ARMA(p,q).

The original assumptions and limitations of ARIMA include weak stationarity, equally spaced observation intervals, and a length of about 50–100 observations [1,2]; in addition, it provides better formulation for incremental than for structural change [2]. As we know, there are three main stages in building an ARIMA model: (1) model identification, (2) model estimation and (3) model checking. Although many previous papers have concentrated on model estimation [3–10], model identification is actually the most crucial stage in building ARIMA models [11], because false model identification will cause the wrong stage of model estimation and increase the cost of re-identification. The stages of building an ARIMA model are described in Fig. 1.

The first method uses the sample partial autocorrelation function (PACF) and the sample autocorrelation function (ACF), as proposed by Box and Jenkins [1] to identify the models in AR and MA, respectively. However, when the time series data sets have mixed ARMA effect, the plot cannot provide clear

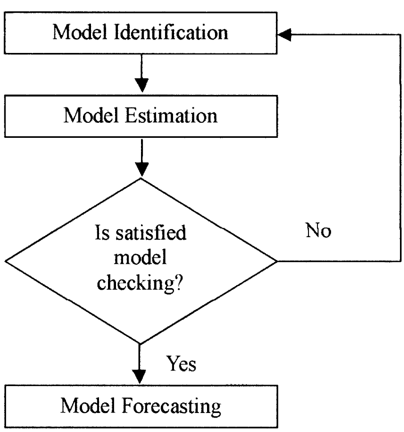

Fig. 1. The stages of building ARIMA models.