Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2007) 3627 – 3630
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Short communication

An example of a misclassification problem applied to Australian equity data

William K. Bertram, M. Shelton Peiris*

School of Mathematics and Statistics, University of Sydney, Australia

Received 1 July 2004; received in revised form 6 December 2006; accepted 6 December 2006
Available online 9 January 2007

Abstract

This paper examines a misclassification problem that can arise when modelling data which contains low frequency components. We illustrate this problem by fitting GAR(1) and AR(1) models to volume and transaction frequency data from the Australian Stock Exchange.

Keywords: Autoregression; Misclassification; Frequency; Spectrum; Finance; Equity data; Modelling; Index

1. Introduction

The modelling of time series which contain low frequency components is a problem of interest in many fields of study. In particular, low frequency events in financial time series often represent the most important events, i.e. the booms and crashes of the market. An important case of these low frequency components is in the study of long memory processes, see Beran (1994). The motivation for this work arose from the paper by Granger and Teräsvirta (1999) which discussed a possible misclassification between nonlinear models and $I(d)$ processes. In Peiris (2003) it was shown that modelling a time series of varying frequency components with an ARMA process leads to a misclassification problem. In this paper we show that modelling such a time series with a GAR(1) or a AR(1) model yields similar results for the regression parameter, however, for the case of the GAR(1) model we have another parameter that is able to describe the degree of frequency of the data. This result indicates how the ARMA model is unable to distinguish between time series of this type.

2. The model

Let B be the backshift operator and I be the identity operator such that $B^j X_t = X_{t-j}$ and $I = B^0$. The model we consider in this paper is the *generalised autoregressive of order* 1 (or GAR(1)) process given by

$$(I - \alpha B)^{\delta} X_t = Z_t, \tag{1}$$

* Corresponding author.
E-mail address: shelton@maths.usyd.edu.au (M.S. Peiris).

0167-9473/$ - see front matter Crown Copyright © 2007 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2006.12.007

where $|\alpha| < 1$, $\delta > 0$ are constants and $\{Z_t\}$ is a sequence of uncorrelated random variables with mean 0 and variance σ^2 (see Peiris et al., 2004). When $\delta = 1$ the model reduces to an AR(1) process.

The spectral density of the GAR(1) model is given by

$$f_X(\omega) = \frac{\sigma^2}{2\pi}(1 - 2\alpha\cos\omega + \alpha^2)^{-\delta}, \quad -\pi \leqslant \omega \leqslant \pi. \tag{2}$$

Thus it can be seen how the parameter δ controls the degree of frequency of the process. For example, for $\delta > 0$ it is clear that $f_X(\omega)$ is large in neigbourhoods of $(\omega, \alpha) = (0, 1)$ and $(\omega, \alpha) = (\pi, -1)$. The autocovariance of this process has been shown in Peiris (2003) and is given by

$$\gamma_0 = \mathrm{Var}(X_t) = \sigma^2 F(\delta; \delta; 1; \alpha^2), \tag{3}$$

$$\gamma_k = \mathrm{Cov}(X_t, X_{t+k}) = \frac{\sigma^2\alpha^k \Gamma(k+\delta) F(\delta; k+\delta; k+1; \alpha^2)}{\Gamma(\delta)\Gamma(k+1)}, \quad k > 0, \tag{4}$$

where $\Gamma(x)$ is the gamma function and $F(a; b; c, d)$ is the Gauss hypergeometric function, given in Abramowitz and Stegun (1968). Examples shown in Peiris (2003) show that the autocorrelation (acf), partial autocorrelation (pacf) and spectrum are similar for some values of α and δ.

3. Fitting to data

In this section we fit both GAR(1) and AR(1) models to Australian Stock Exchange data in order to demonstrate the misclassification problem that arises when modelling data of varying frequency. We use daily traded volume and transaction frequency data, both of which have been proposed as measures for the stock price *volatility* (Bertram, 2004). The acf, pacf and spectrum shown in Fig. 1 indicate that an AR(1) model may be suitable to describe the data.

We emphasise that we are not suggesting GAR(1) or AR(1) as a plausible model for this process, rather using the data as an example of the problem encountered when dealing with a time series of this type. Empirical studies of this data has shown it to possess slowly decaying correlations and hence a fractionally differenced ARIMA$(1, d, 0)$; $d \in (0, 0.5)$ would be an alternative model. It must be noted that the time series for the volume is bound below by zero, however, since the mean of the process is quite large ($\mu \sim 10^6$) we ignore the effects that this lower bound has on the time series. Table 1 displays the estimated GAR(1) and AR(1) parameters for both the volume and transaction frequency, calculated for a selection of stocks.

Looking at the AR(1) parameters for volume we find that they fall roughly into two categories. The stocks CML, RIO, WMC suggest a regression parameter $\alpha_{\mathrm{AR}} \sim 0.50 \pm 0.1$ while the stocks BHP, BIL, CBA indicate $\alpha_{\mathrm{AR}} \sim 0.80 \pm 0.1$. Examining the GAR(1) parameter values we find that the regression parameter is approximately the same as that of the AR(1) model, $\alpha_{\mathrm{GAR}} \approx \alpha_{\mathrm{AR}}$. However, within the two categories we find that the value of δ varies greatly. This result indicates the inability of the ARMA model to discriminate between the different time series. For instance, Table 1 suggests the same AR(1) model for both WMC and RIO, while the GAR(1) parameters show the extent of their differing frequency components. Turning our attention to the transaction frequency we find results similar to those of the volume. For most of the stocks, the parameter estimates indicate an AR(1) model with $\alpha \sim 0.75 \pm 0.1$. Looking at the GAR(1) estimates we find that although the regression parameter estimates are roughly the same, the frequency parameter values vary wildly, with δ values lying in the range (0.5, 1).

A forecast on standardised CBA volume (mean removed) was performed with both AR(1) and GAR(1) models. The three forecasts from the origin at $t = 1000$, 1001 and 1002 days are displayed in Table 2. These results we can see that for the first two GAR(1) forecasts are closer to the observed values than those of AR(1).

4. Summary

We have shown how the extra parameter of the GAR(1) model is able to control the degree of frequency of the model and thereby describe data that consists of varying frequency components. The GAR(1) and AR(1) model parameters were estimated from volume and transaction data from a selection of equities traded on the Australian Stock Exchange. It is well known that the data in study contain low frequency components that lead to a slowly decaying autocorrelation function. We showed that a misclassification problem can occur whereby the ARMA model is unable to distinguish