Test for Structural Change in ARIMA Models

Sangyeol Lee[1] and Siyun Park[2]

Abstract

In this paper we consider the problem of testing for structural changes in ARIMA models based on a cusum test. In particular, the proposed test procedure is applicable to testing for a change of the status of time series from stationarity to nonstationarity or vice versa. The idea is to transform the time series via differencing to make stationary time series. We propose a graphical method to identify the correct order of differencing.

Key words : Testing for parameter change, cusum test, ARIMA model, autocovariance function.

1 Introduction

The problem of testing for a parameter change in time series has been an important issue among statisticians and econometricians. There are a large number of articles as to the change point analysis in iid samples, linear models and time series models. See, for example, Brown, Durbin and Evans (1975), Wichern, Miller and Hsu (1976), Picard (1985), Inclán and Tiao (1994), Bai (1994), Csörgő and Horváth (1997), and Lee and Park (2001), and the papers cited therein. Recently, Lee et al. (2001) proposed a cusum test aimed at testing for a parameter change in time series models. The cusum test not only deals with the classical mean and variance change problem, but also covers general parameter cases, such as the coefficients in RCA and ARCH models. The cusum method turned out to perform adequately in a variety of time series models and to be useful for detecting the locations of changes. However, despite its wide applicability, the attention was paid only to stationary time series models. This motivated us to consider the change point problem in nonstationary models, particularly, the most well-known ARIMA models.

In dealing with the structural change problem in stationary ARMA models,

[1]Ph.D, Associate Professor, Dep. of Statistics, Seoul National University.

[2]Ph.D, Customer Information Analysis Team, SK Corporation.

This research was supported (in part) by KOSEF through the Statistical Research Center for Complex Systems at Seoul National University.

it is natural to employ a test based on the ACF (autocovariance function), since the ACF characterizes the ARMA models. Therefore, the structural change problem in this case is converted to the problem of testing for the change of the ACF, so that a test can be conducted via using the cusum method introduced by Lee et al. (2001). However, if the underlying model is not stationary, the method is no longer applicable since its blind usage will be likely to be in favor of the existence of a change in the ACF. Therefore, in order to construct the method to ARIMA processes, one has to transform given time series to form a union of stataionary processes. A simple method for the transformation is to take differencing consecutively until the time series appear to have stationary characters. However, a question arises as to determining the correct number of differencing. Of course, if we deal with this problem for an ARIMA process with no changes, it is nothing but a part of model selection procedure, and existing methods are already available. However, those methods would not work appropriately when one handles the time series suffering from structural changes. In this case, it may not be easy to derive a suitable estimator in a formal manner. Therefore, here we designate a graphical method to determine the order of differecing that is valid for the time series with structural changes.

The idea to select the number of differencing is simple as seen later in details. We examine the plot of the averaged partial sum of squares of observations. For example, if given time series is stationary, then the averaged partial sum converges to its second moment by a law of large numbers. Furthermore, if the time series is random walk, the partial sums show a hyperbolic trend. Therefore, the partial sum at lag t divided by t^2 lies in certain boundary. A similar reasoning is applicable to other ARIMA processes, and even to the time series with structural changes. Once the order is determined, the cusum test based on the differenced time series can be conducted immediately. In Section 2, we present a modified version of the cusum test for the ACF, and explain the visual method to determine the order of differecing.

2 Test for structural change

Let $\{X_t\}$ denote a time series, and suppose that one wishes to test the following hypotheses:

$H_0 : X_t,\ t = 1, \ldots, n$, follow an ARIMA$(p, d, q)$ model vs.
$H_1 : X_t,\ t = 1, \ldots, m,\ 1 \le m < n$, follow the ARIMA$(p, d, q)$ model
and $X_t,\ t = m+1, \ldots, n$, follow another ARIMA(p', d', q') model.

If the orders d and d' are known, one can test H_0 vs. H_1 applying Lee et al. (2001)'s method to $(1-B)^D X_t$, where D denotes the maximum of d and d'. In the following we describe how the test procedure is conducted. Put $x_t = (1-B)^d X_t$. Then under H_0, we can write

$$x_t = \sum_{i=0}^{\infty} a_i \epsilon_{t-i},$$

where the real sequence a_i are geometrically bounded and ϵ_t are iid random variables with mean 0, variance σ_ϵ^2, and $E|\epsilon_1|^{4\lambda} < \infty$ for some $\lambda > 1$. For $|h| < n$, define

$$\hat{\gamma}_n(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_t - \bar{x}_n)(x_{t+|h|} - \bar{x}_n), \quad \bar{x}_n = \frac{1}{n} \sum_{t=1}^{n} x_t,$$

and let $\{h_n\}$ be a sequence of positive integers, such that as $n \to \infty$,

$$h_n \to \infty \text{ and } h_n = O(n^\beta) \text{ for some } \beta \in (0, (\lambda-1)/2\lambda).$$

Let $\hat{\kappa}_4$ be a consistent estimator of the kurtosis κ_4 of ϵ_1, and set

$$\hat{\Gamma}_{ij} = \hat{\kappa}_4 \hat{\gamma}_n(i) \hat{\gamma}_n(j) + \sum_{r=-h_n}^{h_n} \left(\hat{\gamma}_n(i+r)\hat{\gamma}_n(j+r) + \hat{\gamma}_n(i-r)\hat{\gamma}_n(j+r) \right), \quad i, j = 0, \ldots, m.$$

In view of Theorem 2.2 of Lee et al, we have the following result.

Theorem 2.1 *Let*

$$\mathcal{S}_n(s) = \left(\frac{[ns]}{\sqrt{n}} \left(\hat{\gamma}_{[ns]}(0) - \hat{\gamma}_n(0) \right), \ldots, \frac{[ns]}{\sqrt{n}} \left(\hat{\gamma}_{[ns]}(m) - \hat{\gamma}_n(m) \right) \right), \quad 0 \le s \le 1.$$

Then under H_0*,*

$$\mathcal{S}_n'(s) \hat{\Gamma}^{-1} \mathcal{S}_n(s) \xrightarrow{w} ||\mathbf{W}_{m+1}^\circ(s)||^2,$$

where $\hat{\Gamma}$ *denotes the* $(m+1) \times (m+1)$ *matrix whose (i,j)th component is* $\hat{\Gamma}_{ij}$*, and* $\mathbf{W}_{m+1}^\circ$ *denotes an* $(m+1)$*-dimensional standard Brownian bridge. Therefore,*

$$T_n := \sup_{0<s<1} \mathcal{S}_n'(s) \hat{\Gamma}_n^{-1} \mathcal{S}_n(s) \xrightarrow{w} T := \sup_{0<s<1} ||\mathbf{W}_{m+1}^\circ(s)||^2.$$

We reject H_0 *if* T_n *is large.*

Here we considered the quadratic type test statistic rather than the maximum type test statistic in Theorem 2.2 of Lee et al. In order to conduct the test, one should know the critical values corresponding to significance levels. Since it is not easy to calculate the critical values analytically, we provide the tables through a Monte Carlo simulation. For this task, we generate the random numbers ϵ_t following the standard normal distribution, and compute the empirical quantiles based on the r.v.'s :

$$\mathcal{U}_{n,M} = \sup_{1\leq k\leq n} \sum_{j=1}^{M} \left\{ n^{-1/2} \sum_{j=1}^{k} \epsilon_{ij} - n^{-1/2}(k/n) \sum_{j=1}^{n} \varepsilon_{ij} \right\}^2 .$$

Tables show the significance levels for $\alpha = 0.05, 0.1$ and $M = 1, \ldots, 10$, which are obtained by computing the empirical quantiles using 10000 number of simulated $\mathcal{U}_{1000,M}$'s.

Table 1. $(1-\alpha)\%$ Empirical quantiles of $\mathrm{U}_{1000,M}$, for $M = 1, \cdots 10$.

	M									
α	1	2	3	4	5	6	7	8	9	10
0.01	2.59	3.31	3.91	4.48	5.02	5.52	5.90	6.31	6.78	7.12
0.05	1.76	2.43	2.98	3.47	3.93	4.34	4.75	5.14	5.57	5.96
0.1	1.43	2.03	2.55	2.99	3.43	3.85	4.22	4.60	4.95	5.33

3 Graphical method to identify D

In this section we consider the case that d and d' are unknown. As mentioned earlier, if the time series has structural changes, it is not easy to identify the correct orders. Therefore, here we develop a graphical method to estimate the orders. Suppose that δ_t are iid random variables with zero mean and unit variance. Denote

$$y_j(1) = \sum_{i=1}^{j} \delta_i \text{ and } y_j(k) = \sum_{i=1}^{j} y_i(k-1), \quad k \geq 2.$$

Let

$$W_n(u) = n^{-1/2} \sum_{i=1}^{[nu]} \delta_i, \quad 0 \leq u \leq 1,$$

and let $W(u)$ denote a standard Brownian motion. Define

$$W^{(2)}(u) = \int_0^u W(u)du \text{ and } W^{(k)} = \int_0^u W^{(k-1)}(u)du, \quad k \geq 3.$$

¿From Donsker's invariance principle (cf. Billingsley, 1968), we may write that

$$y_j(1) = n^{1/2}W_n(j/n) \overset{d}{\simeq} n^{1/2}W(j/n)$$

for large n, and

$$\begin{aligned} y_j(2) &= n^{3/2}\{\sum_{i=1}^{j} W_n(i/n)/n\} \simeq n^{3/2}\int_0^{j/n} W_n(u)du \\ &\overset{d}{\simeq} n^{3/2}\int_0^{j/n} W(u)du = n^{3/2}W^{(2)}(j/n). \end{aligned}$$

Similarly, we obtain $y_j(k) \overset{d}{\simeq} n^{k-1/2}W^{(k)}(j/n)$, and thus

$$n^{-2k}\sum_{j=1}^{t}(y_j(k))^2 \overset{d}{\simeq} n^{-1}\sum_{j=1}^{t}(W^{(k)}(j/n))^2 \simeq \int_0^{t/n}(W^{(k)}(u))^2du,$$

which implies that for t close to n,

$$t^{-2k}\sum_{j=1}^{t}(y_j(k))^2 \overset{d}{\simeq} \int_0^1 (W^k(u))^2du = O_P(1).$$

The above argument indicates that one can estimate the order d in $X_t = (1-B)^d\delta_t$ via looking at the shape of the function $g_1 : t \to t^{-1}\sum_{i=1}^{t} X_i^2$ and $g_{2k} : t \to t^{-2k}\sum_{i=1}^{t} X_i^2$, $k \geq 1$. For example, if $d = 2$, it is anticipated that g_1 and g_2 explode fast, $g_4(t)$ are within some boundary, and $g_6(t)$ have the values close to 0. From the same reasoning, if g_{2k}, $k < d$, explode, $g_{2d}(t)$ lie in some boundary, and $g_{2(d+1)}(t)$ have values close to 0, then one can select d as the correct order. Since this is still true for δ_t in a class of linear processes including ARMA processes, one can also identify the order d of general ARIMA models. In fact, this visual method is also useful to decide the correct order D even for the time series with structural changes in ARIMA models. The figures below explain this quite well.

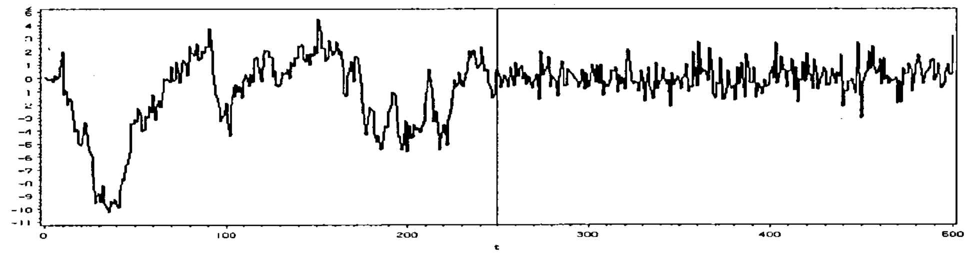

Figure 1: Change from $(1-0.5B)X_t = (1+0.5B)\varepsilon_t$ to $(1-B)(1-0.5B)X_t = (1+0.5B)\varepsilon_t$

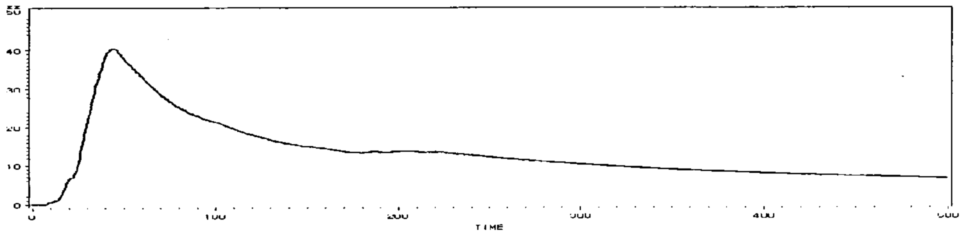

Figure 2: $\frac{1}{t}\sum_{i=1}^{t} X_i^2$ plot for the series presented in Figure 1.

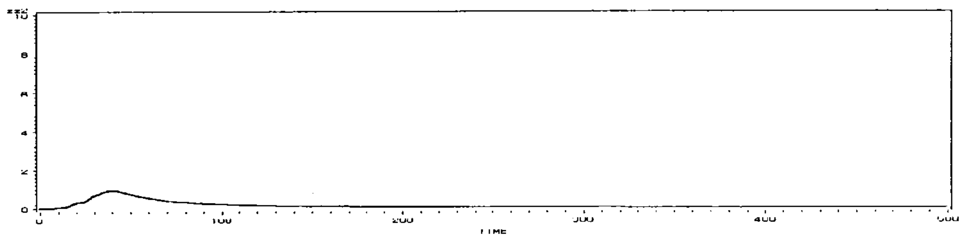

Figure 3: $\frac{1}{t^2}\sum_{i=1}^{t} X_i^2$ plot for the series presented in Figure 1.

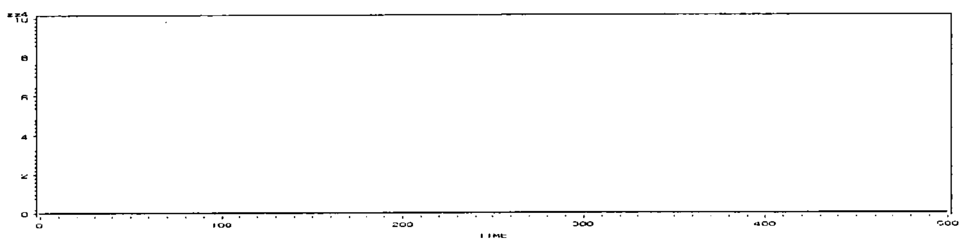

Figure 4: $\frac{1}{t^4}\sum_{i=1}^{t} X_i^2$ plot for the series presented in Figure 1.

REFERENCES

Bai, J. (1994). Weak convergence of the sequential empirical processes of residuals in ARMA models. *Ann. Statist.* **22**, 2051-2061.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Brown, R. L., Durbin, J. and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *J. Roy. Statist. Soc. B.* **37**, 149-163.

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis.* Wiley, New York.

Inclán, C. and Tiao, G. C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variances. *J. Amer. Statist. Assoc.* **89**, 913-923.

Lee, S. and Park, S. (2001). The cusum of squares test for scale changes in infinite order moving average processes. *Scand. J. Statist.* **28**, 625-644.

Picard, D. (1985). Testing and estimating change-points in time series. *Adv. Appl. Prob.* **17**, 841-867.

Wichern, D. W., Miller, R. B. and Hsu, D. A. (1976). Changes of variance in first-order autoregressive time series models - with an application. *Appl. Statist.* **25**, 248-256.