Na Lv, Zhan-Shou Chen and Jian-Qi Ma

On-line Monitoring Long Memory Parameter Change Point in FARIMA Processes

Abstract: In order to overcome the serious size distortion problem when online monitoring the long memory parameter change point in FARIMA\((p, d, q)\) processes based on the critical values simulated by FARIMA\((0, d, 0)\) process, we propose a sieve bootstrap monitoring procedure. Simulations show that the critical values determined by the sieve bootstrap method can not only control the empirical size well below the test level, but also have satisfied test power.

Keywords: long memory, sieve bootstrap, change point monitoring, MOSUM test

1 Introduction

The research on the change point can be traced back to the 1950s, and it has gradually become a hotspot in statistics and econometrics since the 1970s, and last now. In recent years, the application research of the change point has also got the rapid development, and it is widely used in meteorological, financial, medical, economic, hydrological and other fields.

The detection of change point includes posteriori tests and on-line monitoring. The posteriori tests mainly analyse the observed historical data set, however, change point can appear at any given time and also new data arrives steadily it is necessary to detect a change in persistence as soon as possible in the application. On-line monitoring can find change point much earlier, which is beneficial to adjust decisions faster to reduce the loss. Starting with Bauer and Hackl [1] a strand of literature has emerged that studies on-line monitoring. Horvath et al. [2] studied the unconditional variance change point monitoring in conditionally heteroskedastic time series by CUSUM method, wave-type monitoring method, partial sum residual method and recursive residual method respectively. Berkes [3] researched on-line monitoring change point in GARCH\((p, q)\) process. Chen [4] explored on-line monitoring parameter change in linear regression model. Chen [5-6] extended the kernel-
weighted variance ratio method to the variance-infinite case and used it to monitor
the stability of the heavy tailed random signals and the persistent change point from
the stationary to the non-stationary respectively.

These studies mentioned above all considered short memory of time series
model; in this paper we study online monitoring parameter change point of long
memory time series. Due to plenty of actual data such as financial data is more suit-
able for long memory time series modelling, thus on-line monitoring parameter
change point in long memory time series is also very important. Heine and Willet [7]
constructed moving sums of cumulated residuals (abbreviated as MOSUM) statistics
to monitor the autoregressive fractionally integrating moving-average process,
namely the FARIMA \( (p, d, q) \) of persistent change point. However, when at least one
of the autoregressive order \( p \) and the moving average order \( q \) is not equal to zero, if
the critical value provided in the paper is continued using to monitor, the serious
size distortion problem would occur. We use the sieve bootstrap method proposed
by Buhrmann [8] to solve this problem. Chen [9][10][11][12] monitored variance
change, persistent change in long memory process and change from short to long
memory by sieve bootstrap test respectively. Poskitt [12] studied the properties of the
sieve bootstrap method and pointed out that the sieve bootstrap method is very
useful in the analysis of the fractionally integrating process, but they only consid-
ered fixed sample data. We applied the sieve bootstrap method to the on-line moni-
toring parameter change point of long memory time series, and analysed the value
of this method in online data.

The rest of the paper is organized as follows. Section 2 specifies the monitoring
problem and shows all necessary assumptions and monitoring methods. In Section
3 we first compare the size, power and ARL performance of our monitoring proce-
dures with literatures’ by simulations. Section 4 concludes the paper.

2 Model Assumptions and Monitoring Methods

Consider the FARIMA \((p, d, q)\) process

\[
\Phi(B)(1 - B)^d X_t = \Theta(B) \varepsilon_t
\]

Where \( \Phi(B) \) and \( \Theta(B) \) are the AR- and MA- poly nominals respectively, \( \varepsilon_t \) are i.i.d
random variables with mean zero and variance \( \sigma^2 \), \( t = 1, 2, ..., T \), \( B \) is the lag operator.
The differencing parameter \( d \) is assumed to be \(|d| < \frac{1}{2} \), which means \( X_t \) is always a
stationary process.

**Assumption 1** The first \( T \) samples are assumed to be non-polluting, i.e. the long
memory parameter \( d \) in the model is not changed when observing the historical data
set of size \( T \).
Based on the first $T$ observed samples (historical samples), the change points of the long memory parameter $d$ are monitored from the $T + 1$ new observed samples until the $[T\tau], \tau > 1$, Where $[\cdot]$ denotes the integer part of its argument. Namely, we want to test the following hypothesis test problem continuously

$$H_0: d_t = d_0, \quad d_t = T + 1, T + 2, \ldots [T\tau]$$

$$H_A: d_t = d_1, \quad d_0 \neq d_1$$

To test the above hypothesis test, [7] proposed the following MOSUM monitoring function

$$MS_{T,h,d} = \max_{r+1 \leq k \leq [T\tau]} \sigma^{-1}_{e} T^{-\frac{1}{d}} \sum_{k=j-[\tau h]+1}^{j} x_t - \frac{[\tau h]}{T} \sum_{t=1}^{T} x_t$$

(3)

Where $0 < h \leq 1$ is a bandwidth parameter. In order to choose the appropriate boundary function, [7] found the boundary function $b(t) = z \sqrt{2 \log_2(m)}$ (where $m \in [0,1]$, $z$ is a suitable scaling factor) is most effective under the long memory time series model compared with different boundary functions by simulating, so we still use this boundary function in this paper.

According to the mentioned monitoring function and boundary function, online monitoring parameter change point of long memory time series is carried out by the following method: for any given critical value $C(\alpha)$, the monitoring process is stopped when $MS_{T,h,d} > C(\alpha) b(k)$. It is assumed that the long memory parameter change point appears in the observed data, otherwise it will be monitoring until the maximum monitoring sample size appears.

Because $\sigma_{e}^2$ is unknown in the practical problem, we use the estimator $\hat{\sigma}_{e}^2 = T^{-1} \sum_{i=1}^{T} x_t^2$, and the long memory parameter $d$ can be estimated by extended Whittle method which is based on the first $T$ non-contaminating samples and proposed by Abadir[13], and the critical value $C(\alpha)$ satisfies

$$P(MS_{T,h,d} > C(\alpha) b(k) \mid H_0) = \alpha.$$ 

The limit distribution of the monitoring function under the null hypothesis without change point is given and the consistency of the monitoring method under the alternative hypothesis is proved in [7]. And [7] proposed that the corresponding critical value $C(\alpha)$ could be determined by simulations. We list some critical values in Table 1. From model (1), we have

$$(1 - B)^d X_t = \Phi^{-1}(B) \Theta(B) \xi_t$$
Because $\Phi(B)$ and $\Theta(B)$ are assumed to have all roots outside the unit circle, so process $y_t = \Phi^{-1}(B)\Theta(B)\epsilon_t$ still is a stationary short memory time series, i.e. $y_t \sim I(0)$, so the limit distribution of the monitoring function under the null hypothesis without change point. However, the simulations in the next section will point out that the critical value determined by [7] will present a serious size distortion problem when the autoregressive order $p$ and the moving average order $q$ in model (1) are not equal to 0. To overcome this problem, we propose the following sieve bootstrap test to determine the critical value $C(\alpha)$.

The basic idea of the bootstrap is to invest in computing resources to obtain a resampling based estimate of the control limit, instead of relying on the asymptotic distribution as an approximation. The traditional bootstrap method is quite useful as it overcomes the limitations of insufficient data size or unknown theoretical distribution, but this method has limitations when applied to dependent data such as long memory process. These barriers can be overcome thorough block bootstrap method or sieve bootstrap method. While the block bootstrap method has their advantages, there are some general drawbacks. For instance, the resampled blocks might not represent the actual behaviour of the time series under consideration. As a result of this, the bootstrapped series might have weaker dependency than that of original series. Sieve bootstrap can overcome such disadvantages. In this section we introduce the following fractional difference sieve bootstrap method to approximate the asymptotic critical value of statistic $MS_{T,h,d}^*$. The steps are constructed as follows:

1. Estimate long memory parameter $d$, denote the estimate by $\hat{d}$.
2. Difference $X_t, \ldots, X_t \hat{d}$ times, denote the resulting series $Y_t, \ldots, Y_t$ by $Y_t$.
3. Fit a $AR(pT)$ process on this process, i.e. $Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + \cdots + a_p Y_{t-pT} + e_t$. Denote the residual of this operation by $e_t^{d,pT}$.
4. Resample the fitted residual, we have $e_t^{d,pT} \rightarrow e_t^{dT}$.
5. Calculate the statistic

$$MS_{T,h,d}^* = \max_{t=1}^{T} \left| \sum_{k=-[TH]}^{[TH]} \sum_{i=1}^{T} Y_t - \sum_{i=1}^{T} \sum_{k=-[TH]}^{[TH]} Y_t \right|$$

Step 6: Repeat Steps 4-5 B times, take the $1 - \alpha$ quantile of $MS_{T,h,d}^*/b(k)$ as critical value of significance level $\alpha$.

## 3 Simulations

In this section, we assess the finite sample performance of our monitoring procedures. All the simulations are implemented in the R. Table 1 lists critical value of monitoring function $MS_{T,h,d}$ which is obtained based on FARIMA$(0,d,0)$.
\[ T = 5000, \tau = 2 \] generated long memory time series under 10000 cycles simulation, and \( \alpha = 5\% \), bandwidth parameter \( h \) vary among \{0.25, 0.5, 0.75\}.

According to the critical value in Table 1, we set the historical sample \( T' = 200,500 \), monitoring sample \( n = 2T' \) to generate the long memory time series based on the model \( \text{FARIMA}(1, d, 0) \) and \( \text{FARIMA}(0, d, 1) \) respectively. Table 2 and Table 3 show the empirical size resulted from the two data generation processes less than 1000 cycles, and from Table 2 and Table 3 we can see the empirical size is much higher than 5\% of the test level. Although the size distortion decreases as the sample size increases, the test level is still not well controlled. This shows that when the historical sample size is not large in the real problem, the critical value based on \( \text{FARIMA}(0, d, 0) \) model cannot be directly used to monitor change point.

Table 4 and table 5 shows the empirical size resulted based on critical value determined by our sieve boot method, and the historical sample \( T' = 200,500 \), monitoring sample \( n = 2T' \), data generating process is based on the model \( \text{FARIMA}(1, d, 0) \) and \( \text{FARIMA}(0, d, 1) \) respectively. Compared with Table 2 and Table 3, we could see all the empirical sizes are controlled below the test level, but a bit too conservative. Some of the unlisted simulation results show that as the maximum monitoring sample size \( n \) increases gradually, empirical size will close to the selected test level.

**Tab. 1:** Critical values

<table>
<thead>
<tr>
<th>( d )</th>
<th>( h = 0.25 )</th>
<th>( h = 0.5 )</th>
<th>( h = 0.75 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.726</td>
<td>1.075</td>
<td>1.345</td>
</tr>
<tr>
<td>0.2</td>
<td>0.582</td>
<td>0.904</td>
<td>1.164</td>
</tr>
<tr>
<td>0.3</td>
<td>0.458</td>
<td>0.738</td>
<td>0.968</td>
</tr>
<tr>
<td>0.4</td>
<td>0.338</td>
<td>0.548</td>
<td>0.730</td>
</tr>
</tbody>
</table>

**Tab. 2:** Empirical sizes of [7] (%)

<table>
<thead>
<tr>
<th></th>
<th>FARIMA(1, d, 0), AR=0.7</th>
<th>FARIMA(0, d, 1), MA=-0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d )</td>
<td>( h )</td>
<td>( h )</td>
</tr>
<tr>
<td>0.1</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>0.2</td>
<td>94.0</td>
<td>86.2</td>
</tr>
<tr>
<td>0.3</td>
<td>82.2</td>
<td>70.0</td>
</tr>
<tr>
<td>0.4</td>
<td>62.6</td>
<td>50.0</td>
</tr>
<tr>
<td>0.4</td>
<td>40.2</td>
<td>34.4</td>
</tr>
</tbody>
</table>
Tab. 3: Empirical sizes of [7] (%)

<table>
<thead>
<tr>
<th>FARIMA (1, d, 0), AR=0.7</th>
<th>FARIMA (0, d, 1), MA=-0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d )</td>
<td>0.25</td>
</tr>
<tr>
<td>0.1</td>
<td>93.4</td>
</tr>
<tr>
<td>0.2</td>
<td>81.2</td>
</tr>
<tr>
<td>0.3</td>
<td>56.6</td>
</tr>
<tr>
<td>0.4</td>
<td>33.0</td>
</tr>
</tbody>
</table>

Tab. 4: Empirical sizes of sieve boot test (%)

<table>
<thead>
<tr>
<th>FARIMA (1, d, 0), AR=0.7</th>
<th>FARIMA (0, d, 1), MA=-0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d )</td>
<td>0.25</td>
</tr>
<tr>
<td>0.1</td>
<td>2.8</td>
</tr>
<tr>
<td>0.2</td>
<td>1.1</td>
</tr>
<tr>
<td>0.3</td>
<td>1.0</td>
</tr>
<tr>
<td>0.4</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Tab. 5: Empirical sizes of sieve boot test (%)

<table>
<thead>
<tr>
<th>FARIMA (1, d, 0), AR=0.7</th>
<th>FARIMA (0, d, 1), MA=-0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d )</td>
<td>0.25</td>
</tr>
<tr>
<td>0.1</td>
<td>1.6</td>
</tr>
<tr>
<td>0.2</td>
<td>0.6</td>
</tr>
<tr>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>0.4</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Now let the change point be \(k^* = 0.5n\), compare the power and the average run length (ARL) when parameters change. The simulation conditions remain unchanged, Table 6 and Table 7 gives the power and ARL resulted from FARIMA(1, d, 0) and FARIMA(0, d, 1) model data generation processes. Compare the
change of parameter $d$ in Table 6 and Table 7, it can be seen that when the difference between parameters $d_1$ and $d_2$ is larger, that is, the jumping degree is bigger, the power is higher. And the ARL decreases when the jumping degree increases, this is in agreement with the result obtained from the various change point monitoring methods in the literatures. Compare the bandwidth parameter $h$, we can see that the power decreases as $h$ increases, and the ARL increases as $h$ increases. When the bandwidth parameter is too large, the value of the power is too low, which may be due to the fact that the calculated sample size is too small to be stable when calculating the value of the monitoring statistic.

4 Conclusions

The change point monitoring method can not only analyse the change points in online data, but also can find change point earlier than the posteriori tests based on fixed samples in most cases, which is advantageous to adjust the decision in time to reduce the loss. For the serious size distortion problem of long memory time series long memory parameter change point using critical value determined directly by simulation method when the autoregressive order $p$ and the moving average order $q$ are not equal to 0. We proposed determine the critical value by sieve bootstrap resampling method to monitor and simulation results show that the test level can be well controlled and the power is higher as well.

Acknowledgement: This work was supported by the National Natural Science Foundation of China (No. 11661067), Natural Science Foundation of Qinghai Province (No. 2015-ZJ-717).

References


