Bootstrap Test Mean Change Point in Long Memory Time Series

Y H Xing
School of Mathematics and Statistics, Qinghai Normal University, Xining 810008, China
Email: 296613570@qq.com

Abstract. Bootstrap plays an important role in change point analysis for it is a data driving method and can avoid estimate some redundant parameters. In this paper, we applied three well known bootstrap methods, the sieve AR bootstrap, the fractional differencing sieve bootstrap and the fractional differencing block bootstrap to test the mean change point in the stationary long memory time series. We use the self normalized ratio statistic as the test statistic and approximate its critical values via these three bootstrap methods. We evaluate the empirical size and power performances of three bootstrap methods. Simulations show that the sieve AR bootstrap undergoes serious size distortions when the long memory parameter nears to 0.5, and the fractional differencing block bootstrap always too conservative compared to the other two bootstrap. The fractional differencing sieve bootstrap, in general, has the best finite sample performance. Finally, we illustrated the method via a set annual discharge data in the Nile River and a set of temperature data in the northern hemisphere.

Keywords. Change point; long memory time series; bootstrap.

1. Introduction
Since Hidalgo and Eobinson (1996) [1] studied the change point problem in the long memory time series, detecting change point in long memory time series becomes a popular topic in statistics and economics. Some well known studies about this topic include Wright (1998) [2], Wang (2008) [3], Wenger et al. (2018) [4]. How to get the critical values of test statistic plays an important issue in change point analysis for most test statistic only has asymptotic distribution which depends on unknown long memory parameter and other redundant parameters. Approximating the critical values of asymptotic distribution by bootstrap method is a nice choice. Kirch (2008) [5] have showed that the bootstrap critical values are more robust than the actual distribution of the change point test statistic and provide consistency in size and power. Chen et al. [6-8] applied the sieve AR bootstrap (SARB) method to sequentially detect changes in persistence in long memory time series.

Although many paper have used various bootstrap methods detecting change point in the long memory time series, there are lack comparison studies about which bootstrap method has better finite sample performance. This paper aims to compare the size and power performance of three well used bootstrap methods: the SARB [9], the fractional differencing sieve bootstrap (FDSB) [10, 11] and the fractional differencing block bootstrap (FDBB) [12], when using them to approximate the critical values of a well known mean change point test statistic in long memory time series. Peng et al. (2020) [13] have showed that these three bootstrap methods work well when using them to approximate the long memory process.
2. Model and Statistics

We consider the following long memory time series model \{Y_t\},
\[ Y_t = \mu + X_t, \Phi(L)(1 - L)^d X_t = \Psi(L) \epsilon_t, \quad t = 1, 2, \ldots, n \]  
(1)
where \( \mu = E(Y_t) \) is a deterministic component, \( X_t \) is the random component, and the \( n \) denotes the sample size. The random component \( X_t \) follows an ARFIMA\((p,d,q)\) model, in which \( \epsilon_t \) is an i.i.d random variable with mean zero and variance \( \sigma^2 \), and the \( L \) denotes the lag operator. The AR- and MA-polynomials \( \Phi(L) \) and \( \Psi(L) \) are assumed to have all roots outside the unit circle. The long memory parameter \( d \) is restricted to \( 0 \leq d < 0.5 \) so that the process \( \{Y_t\} \) is a stationary long memory process.

We are interested in detecting whether the mean \( \mu \) in model (1) is a constant throughout the sample there is a change in it, that is, we test the following null and alternative hypothesis.
\[ H_0: E(Y_t) = \mu, \quad t = 1, 2, \ldots, n \]
(2)
\[ v.s. \]
\[ H_1: E(Y_t) = \mu, \quad t = 1, \ldots, k^* \]
\[ E(Y_t) = \mu + \Delta, \quad t = k^* + 1, \ldots, n \]  
(3)
where \( k^* = \lfloor n\tau \rfloor \) is the unknown change point, and the change size \( \Delta \) is a fixed non-zero constant.

Although many test statistics have been proposed to test above hypothesis, we focus on the self normalized ratio statistic proposed by Shao (2011) [14]. The major advantage of this ratio statistic is that we do not need to estimate any scale parameter. This will be very useful in practice for scale parameter can hardly be well estimated. The self normalized ratio statistic is defined as
\[ G_n(k) = \frac{\sqrt{n}|\bar{Y}_{1,k} - \bar{Y}_{k+1,n}|}{\sqrt{n^{-1}(\sum_{i=1}^{k} s_i^2(1, k) + \sum_{i=k+1}^{n} s_i^2(k+1, n))^{1/2}}} \]  
(4)
where \( \bar{Y}_{j,k} = (k - j + 1)^{-1} \sum_{i=j}^{k} Y_t \) and \( s_i = (\sum_{h=j}^{k} (Y_h - \bar{Y}_{j,k}))^2, \quad t = j, \ldots, k, \quad 1 \leq k \leq n. \)

Let
\[ T_n(\tau_1, \tau_2) = \sup_{k \in [n\tau_1, n\tau_2]} G_n(k), 0 < \tau_1 < \tau_2 < 1 \]
where \( \tau_1, \tau_2 \) denote the trimming proportion. For a given critical value \( c(\alpha) \), the ratio statistic (4) detected a mean change point at nominal level \( \alpha \) if \( T_n(\tau_1, \tau_2) > c(\alpha) \).

Shao (2011) [14] has proved that the null distribution of ratio statistic (4) is functional of the type I fractional Winner process, and listed some asymptotic critical values of this distribution. Since the critical value depends on the long memory parameter \( d \), we have to estimate the long memory parameter before choosing the right critical value. However, we can hardly give a precise estimation for \( d \) especially under the alternative hypothesis, which leads the determined critical value may wrong. Compared to use the asymptotic critical value, use bootstrap approximated critical value may not need to estimate long memory parameter or need lower precision require when estimating it. We are interested in checking whether the SARB, FDSB and FDBB well work when using them to approximate the critical values of test statistic (4).

3. Simulation

In this section, we compare the size and power performances of SARB, FDSB and FDBB methods when using them detecting mean change point test. We use the ARFIMA\((0, d, 0)\) model to generate the data and vary \( d \) varying among \{0, 0.2, 0.4\}. The sample size is varying among \{100, 400\}. The change point set to be \( \tau = 0.25, \quad 0.5, \quad 0.8 \) and change size to be \( \Delta = 2 \). All simulations are obtained at the nominal level \( \alpha = 5\%, \quad 10\% \) via 1000 replications.

Table 1 shows the empirical sizes of test statistic (4) under the null hypothesis using approximated
critical values by three bootstrap methods as well as asymptotic critical values (ACV). The results under the alternative hypothesis are listed in table 2. From table 1 we can see that ACV controls the empirical size very well. It is because we set d to be known when use ACV. Three bootstrap methods have similar size performance and size distortions become decrease as sample size increase in general. The empirical size computed by SARB method heavily depends on long memory parameter d when n = 100. From table 2 we can conclude that SARB methods gives highest empirical power in all cases and FDSB is the follower. The FDBB method, however, always has lowest empirical power and the performance is very poor compared to other three methods in many times.

| d  | n = 100 |   |   |   | n = 400 |   |   |   |
|----|---------|---|---|---|---------|---|---|---|
| 0  | 10%     | 9.2 | 11.8 | 13.1 | 13.1 | 9.54 | 12.1 | 11.9 | 11.8 |
| 5% | 4.2     | 6.0 | 5.9 | 6.9 | 5.1 | 4.7 | 5.9 | 6.1 |
| d = 0.2  | 9.74 | 14.9 | 11.7 | 10.5 | 9.66 | 11.0 | 11.4 | 9.1 |
| 5% | 4.56 | 7.3 | 6.2 | 5.6 | 5.08 | 5.6 | 5.4 | 4.5 |
| d = 0.4  | 10.2 | 7.6 | 11.9 | 10.1 | 10.3 | 9.0 | 9.7 | 7.41 |
| 5% | 5.02 | 4.1 | 5.8 | 5.1 | 5.24 | 5.5 | 6.2 | 4.5 |

Table 2. Empirical powers (%) when there is a mean change at τ with change size ∆ = 2.

| τ  | n = 100 |   |   |   | n = 400 |   |   |   |
|----|---------|---|---|---|---------|---|---|---|
| 0.25 | 89.5 | 96.8 | 94.6 | 72.7 | 99.9 | 100 | 100 | 99.4 |
| 0.5 | 94.7 | 99.2 | 98.6 | 76.5 | 100 | 100 | 100 | 100 |
| 0.8 | 87.2 | 97.0 | 93.3 | 70.4 | 99.8 | 100 | 100 | 98.6 |
| d = 0.2 | 54.8 | 72.5 | 65.5 | 34.2 | 84.8 | 94.3 | 88.9 | 66.2 |
| 0.5 | 63.0 | 87.3 | 73.1 | 36.8 | 93.3 | 99.6 | 96.1 | 78.2 |
| 0.8 | 50.7 | 68.8 | 59.1 | 38.1 | 81.4 | 93 | 85 | 69.1 |
| d = 0.4 | 22.4 | 40.6 | 28.5 | 14.9 | 29.1 | 47.6 | 29.1 | 19.1 |
| 0.5 | 28.0 | 55.9 | 35 | 19.9 | 37.4 | 63.0 | 41.1 | 26.8 |
| 0.8 | 19.1 | 36.4 | 24.4 | 14.5 | 26.1 | 43.1 | 25.7 | 20.6 |

4. Empirical Application for Change Point Test
In this section, we apply three bootstrap methods detecting mean change point in two sets of real data. The first data set contains 100 observations about annual discharge of the Nile river during the years 1871-1970 (see figure 1). Shao (2011) [14] detected a mean change point in this data set by ratio statistic (4) based on ACV method at 5% nominal level. We found that the ratio statistic (4) reached the maximum value 78.54 at the 26th observation (see the vertical dotted line in figure 1). The SARB method gives approximated critical value 71.23 at 5%, and the critical values obtained by FDSB and FDBB are 49.8 and 64.93 respectively. This indicates that all three bootstrap methods support the conclusion obtained by Shao (2011) [14]. The second data set is 1632 seasonally adjusted monthly deviations of the observed temperature from the northern hemisphere between 1854 and 1989. The maximum value of ratio test statistic is 75.28 at the 826th observation (see the vertical dotted line in figure 2). The critical values obtained by SARB, FDSB and FDBB are 74.48, 38.28 and 58.46 respectively. Obviously, the maximum value of test statistic is greater than the bootstrap approximated critical values, which implies that all three bootstrap methods detected a change point in this data set. Also, this result is the same as that of Shao (2011) [14].
Figure 1. The annual discharge of the Nile river during the years 1871-1970.

Figure 2. The seasonally adjusted monthly deviations of the observed temperature from the northern hemisphere between 1854 and 1989.

5. Conclusions
This paper has studied finite sample performances of three popular used bootstrap methods when using them detecting mean change point in the long memory time series via simulation experiments. From the experiment we concluded that three methods all can control the empirical size well when the long memory parameter $d$ is not too large, but the SARB will lead to serious size distortion when $d$ nears to 0.5. The SARB method has the highest empirical power performance and the FDBB method always too conservative. In general, the FDSB has the best finite sample performance.

Acknowledgements
We thank the editors for their help comments. This work was supported by the National Natural Science Foundation of China (No.12161072) and the Natural Science Foundation of Qinghai Province (No.2019-ZJ-920).

References
[1] Hidalgo J and Robinson P M 1996 Journal of Econometrics 70 159-74.
[2] Wright J H 1998 Journal of Time Series Analysis 19 369-76.
[3] Wang L 2008 Journal of Statistical Computation & Simulation 78 653-68.
[4] Wenger K, Leschinski C and Sibbertsen P 2018 AStA Adv Stat Anal https://doi.org/10.1007/s10182-018-0328-5.
[5] Kirch C 2008 Sequential Analysis 27 33049.
[6] Chen Z, Tian Z and Xing Y 2016 Statistics and Its Interface 9 37-45.
[7] Chen Z, Xing Y and Li F 2016 Economics Letters 140 53-6.
[8] Chen Z, Xiao Y and Li F 2021 Empirical Economics 60 2365-89.
[9] Poskitt D S 2008 Journal of Time Series Analysis 29 224-50.
[10] Davidson J and Rambaccussing D 2015 Journal of Time Series Econometrics 7 115-41.
[11] Kapetanios G A 2010 Quantitative and Qualitative Analysis in Social Sciences 4 19-40.
[12] Papailias F and Kapetanios G 2011 University of London Queen Mary School of EFWP No. 679.
[13] Peng M, Chen Z and He M 2020 Journal of Physics Conference Series https://doi.org/10.1088/1742-6596/1624/3/032003.
[14] Shao X 2011 Journal of Time Series Analysis 32 598C606.