Temporal Aggregation Effects on a Mean-Change in Time Series

Bu Hyoung Lee and William W. W. Wei
Department of Statistics, Temple University, 1801 Liacouras Walk, Philadelphia, PA 19122

Abstract
We investigate the effects of temporal aggregation on a structural mean-change of a time series. Even though the aggregation induces substantial information loss, it does not affect the test results of detecting a mean-change point, using the CUSUM test. The results show that the temporal aggregation does not weaken the impact of the structural change in mean.

Key Words: Temporal aggregation, Structural mean-change, CUSUM test, Time series

1. Introduction

Since temporal aggregation induces substantial information loss, it also affects structural changes or breaks of time series. However, we cannot directly apply traditional statistical tests of independent samples, such as the *t*-test for testing a mean-difference, to detecting the structural changes because time series observations are almost certainly dependent and no possibility for randomization exists (Box and Tiao, 1965; Wei, 2006).

Therefore alternative approaches have been proposed and developed by many authors. The issue how to find a change point of structural mean-change or level-shift has been discussed within two theoretical frameworks, i.e., the likelihood ratio (LR) test (see Hinkely, 1970; Hinkely, 1971; Chang, *et al*., 1988; Tsay, 1988; Balke, 1993; Chen and Liu, 1993; Galeano, *et al*., 2006) and the cumulative sum (CUSUM) test (see ; Brown, *et al*., 1975; Hsu, 1977; Krämer *et al*., 1988; Bai, 1994; Incláin and Tiao, 1994; Juhl and Xiao, 2009; Shao and Zhang, 2010).

However both the two methods have some drawbacks. For the LR test, the model parameters should be known or predetermined. The CUSUM test has an issue to select a consistent long-run variance of series. Therefore, we employ a modified CUSUM test proposed by Shao and Zhang (2010), which is not dependent of a model and adopts a new self-normalized estimator for the long-run variance.

In this paper, we investigate the effects of temporal aggregation, due to information loss, on a structural mean-change of a univariate time series, using the modified CUSUM test.

2. Temporal Aggregation

Assume that time series $\{Z_T, T=1,\ldots,N\}$ is the *m*-period nonoverlapping aggregates of series $\{X_t, t=1,\ldots,n\}$ defined as

$$Z_T = \sum_{t=m(T-1)+1}^{mT} X_t = (1+B+\cdots+B^{m-1})X_{mT},$$

where B is the backshift operator of $B^j X_t = X_{t-j}$, m is the order of aggregation, and T is the aggregate time unit. In general, $\{X_t\}$ is called a nonaggregate series and $\{Z_T\}$ an aggregate series. We note that the number of observation of the aggregate is rewritten as the quotient of the size of original series and the aggregation order, *i.e.*, $N = n/m$ (Teles *et al*., 2008; Wei, 2006).

We also express the sample mean of aggregates $\{Z_T\}$ as the form of the sample mean of nonaggregates $\{X_t\}$ such that

$$\overline{Z}_N = \frac{1}{N}\sum_{T=1}^{N} Z_T = \frac{m}{N}\sum_{T=1}^{N}\sum_{t=m(T-1)+1}^{mT} X_t = \frac{m}{n}\sum_{t=1}^{n} X_t = m\overline{X}_n,$$

where $\overline{Z}_N = \frac{1}{N}\sum_{T=1}^{N} Z_T$ and $\overline{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t$.

3. Detecting a Structural Change

3.1 A Change in Mean

Assume that a univariate time series $\{X_t, t=1,\ldots,n\}$ follows the model

$$X_t = \mu_t + \varepsilon_t,$$

where $\{\varepsilon_t\}$ is a white noise with mean zero and variance σ^2. And so $E(X_t) = \mu_t$.

It is of interest to test a single mean-change in series $\{X_t\}$. The problem can be written as testing the null hypothesis

$$H_0 : \mu_1 = \cdots = \mu_n \equiv \mu$$

against the alternative

$$H_a : \mu_1 = \cdots = \mu_k \neq \mu_{k+1} = \cdots = \mu_n,$$

which says that one mean-change occurs at time point $l = k+1$ for $1 \le k < n$ (Aue and Horváth, 2012).

Let $k = \lfloor nr \rfloor$ for $r \in [0,1]$, where $\lfloor \cdot \rfloor$ denotes the integer part. The cumulative sum (CUSUM) process is given by

$$C_n(k) = \frac{1}{\sqrt{n}}\sum_{t=1}^{k}(X_t - \overline{X}_n) = \frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr \rfloor}(X_t - \overline{X}_n)$$
$$= \frac{1}{\sqrt{n}}\left(\sum_{t=1}^{\lfloor nr \rfloor} X_t - \frac{\lfloor nr \rfloor}{n}\sum_{t=1}^{n} X_t\right),$$

(Brown, *et al*., 1975).

Assume that a sequence $\{X_t - \mu_t, t = 1,2,\ldots,\infty\}$ satisfies the conditions of appropriate moment and weak dependence, shown below (Phillips, 1987):

1. $E(X_t - \mu_t) = 0$ for all t;
2. $\sup_t E\,|X_t - \mu_t|^{\beta} < \infty$ for some $\beta > 2$;
3. The long-run variance
$$\sigma^2 = \lim_{n\to\infty} \frac{1}{n} E\left(\left\{\sum_{t=1}^{n}(X_t - \mu_t)\right\}^2\right) = \lim_{n\to\infty} nVar(\overline{X}_n)$$
exists and $\sigma^2 > 0$.

With these assumption, we have
$$\frac{1}{n}\sum_{t=1}^{\lfloor nr \rfloor}(X_t - \mu_t) \Rightarrow \sigma W(r),$$
where the symbol "$\Rightarrow$" signifies the weak convergence as $n \to \infty$ (see Billingsley, 1995). Therefore, under the null hypothesis, the limiting distribution of the CUSUM process is shown as
$$C_n(k) \Rightarrow \sigma B(r),$$
where a standard Brownian bridge $B(r) = W(r) - rW(1)$ with a standard Brownian motion $W(r)$ (see Perron, 2006; Shao and Zhang, 2010; Aue and Horváth, 2012).

One possible approach for the mean-change test is to examine the maximum value of the standardized CUSUM process, *i.e.*, $\sup_{1\le k<n} |C_n(k)/\sigma|$. In practice, the long-run variance σ^2 is unknown and so a consistent estimator $\hat{\sigma}_n$ has to be substituted for σ.

Lobato (2001) presented a good alternative called the self-normalization (SN) estimator,
$$\hat{\sigma}_n^2 = \frac{1}{n^2}\sum_{s=1}^{n}\left\{\sum_{t=1}^{s}(X_t - \overline{X}_n)\right\}^2.$$
Then we can derive the limiting null distribution of the CUSM test statistic,
$$\sup_{1\le k<n}\left|\frac{C_n(k)}{\hat{\sigma}_n}\right| \xrightarrow{d} \sup_{r\in[0,1]} \frac{|B(r)|}{\sqrt{\int_0^1 B^2(r)dr}}$$
(see Lobato, 2001; Shao, 2010; Shao and Zhang 2010). However, the Lobato test meets a serious zero-power problem. That is, as the denominator of the test statistic gets large with respect to the increase of the mean-change magnitude, the power of test decreases to zero (Vogelsang, 1999; Juhl and Xiao, 2009; Shao and Zhang 2010).

Shao and Zhang (2010) and Shao (2011) proposed an idea to avoid the zero-power problem. Their SN estimator $\hat{\sigma}_n^2$ is defined as
$$\hat{\sigma}_n^2 = \frac{1}{n^2}\left[\sum_{t=1}^{k}\left\{S_{1,t} - \left(\frac{t}{k}\right)S_{1,k}\right\}^2 + \sum_{t=k+1}^{n}\left\{S_{t,n} - \left(\frac{n-t+1}{n-k}\right)S_{k+1,n}\right\}^2\right],$$
where

$$S_{t_1,t_2} = \begin{cases} \sum_{t=t_1}^{t_2} X_j, & \text{if } t_1 \le t_2, \\ 0, & \text{otherwise.} \end{cases}$$

Under H_0 the limiting distribution of the CUSM test statistic can be derived as

$$\sup_{1\le k<n} \left| \frac{C_n(k)}{\hat{\sigma}_n} \right| \xrightarrow{d} \sup_{r\in[0,1]} \frac{|W(r) - rW(r)|}{\sqrt{V(r)}},$$

where

$$V(r) = \int_0^r \left\{ W(s) - \left(\frac{s}{r}\right) W(r) \right\}^2 ds + \int_r^1 \left[W(1) - W(s) - \left(\frac{1-s}{1-r}\right) \{W(1) - W(r)\} \right]^2 ds$$

(for more details, see Shao and Zhang, 2010; Shao, 2011).

3.2 A Change in Mean of an Aggregated Series

Now we study a single mean change point of the aggregate $\{Z_T, T = 1, \ldots, N\}$. The problem can be written as the hypothesis test of

$$H_0 : \mu_1^{(A)} = \cdots = \mu_N^{(A)} \equiv \mu^{(A)}$$

against

$$H_a : \mu_1^{(A)} = \cdots = \mu_K^{(A)} \ne \mu_{K+1}^{(A)} = \cdots = \mu_N^{(A)},$$

where $\mu_T^{(A)} = E(Z_T)$ and $K = \lfloor k/m \rfloor$ for $1 \le K < N$.

The CUSUM process of series $\{Z_T\}$ can be expressed as

$$\begin{aligned} C_N^{(A)}(K) &= \frac{1}{\sqrt{N}} \sum_{T=1}^{K} (Z_T - \bar{Z}_N) = \frac{\sqrt{m}}{\sqrt{n}} \sum_{T=1}^{\lfloor k/m \rfloor} \sum_{t=m(T-1)+1}^{mT} (X_t - \bar{X}_n) \\ &= \frac{\sqrt{m}}{\sqrt{n}} \sum_{t=1}^{m\lfloor k/m \rfloor} (X_t - \bar{X}_n) = \sqrt{m}\, C_n\left(m\lfloor k/m \rfloor\right). \end{aligned}$$

Also the long-run variance of series $\{Z_T\}$ can be written as

$$\begin{aligned} \sigma_A^2 &= \lim_{N\to\infty} N \cdot Var(\bar{Z}_N) = \lim_{n\to\infty} \frac{n}{m} \cdot Var(m\bar{X}_n) \\ &= m \lim_{n\to\infty} n \cdot Var(\bar{X}_n) = m\sigma^2, \end{aligned}$$

Let $k' = m\lfloor k/m \rfloor$. Then the CUSUM test statistic of the aggregate has the form of the nonaggregate, *i.e.*,

$$\sup_{1\le K<N} \left| \frac{C_N^{(A)}(K)}{\hat{\sigma}_A} \right| = \sup_{1\le\lfloor k/m \rfloor<n/m} \left| \frac{C_n\left(m\lfloor k/m \rfloor\right)}{\hat{\sigma}_n} \right| = \sup_{k'=m,2m,\ldots,n-m} \left| \frac{C_n(k')}{\hat{\sigma}_n} \right|.$$

Now the limiting null distribution of the test statistic is

$$\sup_{k'=m,2m,\ldots,n-m} \left| \frac{C_n(k')}{\hat{\sigma}_n} \right| \xrightarrow{d} \sup_{r'\in[0,1]} \frac{|W(r') - r'W(r')|}{\sqrt{V(r')}},$$

where $k' = \lfloor nr' \rfloor$ for $r' \in [0,1]$. Therefore we believe that the two limiting null distributions of the nonaggregate test statistic and the aggregate test statistic have a same distribution when *m* is much smaller than *n*.

The empirical and asymptotic quantiles of the null distribution based on 100,000 iterations are tabulated below.

Table 1: Empirical and Asymptotic Quantiles of the Null Distribution

	0.250	0.500	0.750	0.900	0.950	0.975	0.990
Q	8.151	10.741	14.316	18.587	21.772	25.253	30.200

NOTE: *Q* is the 100 quantile and $n = 1000$.

4. Temporal Aggregation Effects on a Mean-Change: A Simulation Study

We assume that the first partial series for $t = 1, \ldots, 350$ has mean 0 and the second partial series for $t = 351, \ldots, 1000$ has mean 1. Then we consider three AR(1) process of autoregressive coefficients 0.1,

$$X_t = \begin{cases} 0.1X_{t-1} + e_t, & t = 1, \ldots, 350, \\ 0.9 + 0.1X_{t-1} + e_t, & t = 351, \ldots, 1000, \end{cases}$$

autoregressive coefficients 0.5,

$$X_t = \begin{cases} 0.5X_{t-1} + e_t, & t = 1, \ldots, 350, \\ 0.5 + 0.5X_{t-1} + e_t, & t = 351, \ldots, 1000, \end{cases}$$

autoregressive coefficients 0.9,

$$X_t = \begin{cases} 0.9X_{t-1} + e_t, & t = 1, \ldots, 350, \\ 0.1 + 0.9X_{t-1} + e_t, & t = 351, \ldots, 1000, \end{cases}$$

where $\{e_t\}$ is a white noise with mean 0 and variance 1.

We also consider five different aggregation orders of $m_1 = 10$, $m_2 = 20$, $m_3 = 50$, $m_4 = 100$, and $m_5 = 200$. Then the true *K* values are $K_1 = 35$, $K_2 = 17$, $K_3 = 7$, $K_4 = 3$, and $K_5 = 1$, respectively.

Using the Shao test for the nonaggregate and its extension for the aggregate, we compute the CUSUM test statistics and detect the time point *k* or *K*, as shown in Table 2.

Table 2: The CUSUM Test Statistics and the Change Points

m		1	10	20	50	100	200
coeff.	Test stat.	279.496	270.462	281.509	193.633	118.964	211.560
0.1	*k* or *K*	335	34	17	7	4	2
coeff.	Test stat.	35.018	34.749	31.052	30.921	26.945	40.630
0.5	*k* or *K*	327	33	17	7	3	2
coeff.	Test stat.	5.443	5.436	5.447	5.158	4.199	7.016
0.9	*k* or *K*	720	72	36	15	7	3

Therefore, for the AR(1) series with coefficients 0.1 and 0.5, all the null hypotheses of no mean-change are rejected at significance level $\alpha = 0.05$ and the detected change points are very close to the true values. However, for the coefficient 0.9, the results are not significant at the same α level.

5. Conclusion

We investigated the effects of temporal aggregation on a single mean-change of a time series, using Shao's CUSUM test. Based on the significant results of the simulation study, the detected mean-change points are very similar to the true values of $k = 350$, $K_1 = 35$, $K_2 = 17$, $K_3 = 7$, $K_4 = 3$, and $K_5 = 1$.

Because of the same forms of the test statistics and the null distributions for the nonaggregate series and the aggregate series, the aggregate is only treated as another nonaggregate series with small size in the CUSUM test.

Therefore, the information loss due to temporal aggregation does not seem to effect on detecting the structural mean-change of the series through the test.

References

Aue, A., and Horváth, L. (2012), "Structural Breaks in Time Series," *Journal of Time Series Analysis*, 34, 1–16.

Bai, J. (1994), "Least Squares Estimation of a Shift in Linear Processes," *Journal of Time Series Analysis*, 15, 453–472.

Balke, N. (1993), "Detecting Level Shifts in Time Series," *Journal of Business & Economic Statistics*, 11, 81–92.

Billingsley, P. (1995), *Convergence of Probability Measures* (3rd ed.), New York: Wiley.

Box, G. E. P., and Tiao, G. C. (1965), "A Change in Level of a Non-stationary Time Series," *Biometrika*, 52, 181–192.

Brown, R. L., Durbin, J., and Evans, J. M. (1975), "Techniques for Testing the Constancy of Regression Relationships over Time," *Journal of the Royal Statistical Society, Ser. B*, 37, 149–192.

Chang, I., Tiao, G. C., and Chen C. (1988), "Estimation of Time Series Parameters in the Presence of Outliers," *Technometrics*, 30, 193–204.

Chen, C., and Liu L. M. (1993), "Joint Estimation of Model Parameters and Outlier Effects in Time Series," *Journal of the American Statistical Association*, 88, 284–297.

Galeano, P., Peña, D., and Tsay, R. S. (2006), "Outlier Detection in Multivariate Time Series by Projection Pursuit," *Journal of the American Statistical Association*, 101, 654–669.

Hinkley, D. V. (1970), "TrustInference About the Change-Point in a Sequence of Random Variables," *Biometrika*, 57, 1–17.

Hinkley, D. V. (1971), "Inference about the Change-Point from Cumulative Sum Tests," *Biometrika*, 58, 509–523.

Hsu, D. A. (1977), "Tests for Variance Shift at an Unknown Time Series," *Journal of the Royal Statistical Society, Ser. C*, 26, 279–284.

Inclán, C., and Tiao, G. C. (1994), "Use of Cumulative Sums of Squares for Retrospective Detection of Change of Variance," *Journal of the American Statistical Association*, 89, 913–923.

Juhl, T., and Xiao, Z. (2009), "Tests for changing mean with monotonic power," *Journal of Econometrics*, 148, 14–24.

Krämer, W., Ploberger, W., and Alt, R. (1998), "Testing for Structural Change in Dynamic Models," *Enonometrica*, 56, 1355–1369.

Lobato, I. N. (2001), "Testing That a Dependent Process Is Uncorrelated," *Journal of the American Statistical Association*, 96, 1066–1076.

Perron, P. (2006), "Dealing With Structural Breaks," in *Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory,* eds. K. Patterson and T. C. Mills, Great Britain: Palgrave Macmillan, pp. 278–352.

Phillips, P. C. B. (1987), "Time Series Regression With Unit Roots," *Econometrica*, 55, 277–301.

Shao, X., and Zhang, X. (2010), "Testing for Change Points in Time Series," *Journal of the American Statistical Association*, 105, 1228–1240.

Shao, X. (2011), "A Simple Test of Changes in Mean in the Possible Presence of Long-Range Dependence," *Journal of Time Series Analysis,* 32, 598–606.

Shao, X., and Zhang, X. (2010), "Testing for Change Points in Time Series," *Journal of the American Statistical Association*, 105, 1228–1240.

Teles P., Wei. W. W. S, and Hodgess, E. N. (2008), "Testing a Unit Root Based on Aggregate Time Series," *Communications in Statistics - Theory and Methods*, 37, 565–590.

Tsay, R. S. (1988), "Outliers, Level Shifts, and Variance Changes in Time Series," *Journal of Forecasting*, 7, 1–20.

Vogelsang, T. J. (1999), "Sources of Nonmonotonic Power when Testing for a Shift in Mean of a Dynamic Time Series," *Journal of Econometrics*, 88, 283–299

Wei, W. W. S. (2006), *Time Series Analysis: Univariate and Multivariate Methods* (2nd ed.), Boston: Addison Wesley.