Liet. matem. rink, **45**, *spec. nr.*, 2005, 385–389

Testing for a changed segment in variance with application

Danas ZUOKAS (VU)
e-mail: danas.zuokas@mif.vu.lt

1. Introduction

Consider a set of observations $X_1, \dots, X_n$, which is driven by the following model

$$X_i = \mu + \sigma_i \varepsilon_i, \quad i = 1, \dots, n, \tag{1}$$

where μ is constant (un)known mean and $\varepsilon_1, \dots, \varepsilon_n$ are i.i.d. random variables with $E\varepsilon_1 = 0$, $E\varepsilon_1^2 = 1$. We wish to test the null hypothesis

$$\mathrm{H}_0\colon\ \sigma_1 = \sigma_2 = \cdots = \sigma_n = \sigma$$

against the following changed segment (epidemic) alternative

H_A: $\exists$ integers $0 < \ell^* < n$ and $0 \leqslant k^* \leqslant n - \ell^*$ and real valued $\tau \neq \sigma$ such that

$$\sigma_i = \begin{cases} \tau, & i = k^* + 1, \dots, k^* + \ell^*, \\ \sigma, & i = \{1, \dots, n\} \setminus \{k^* + 1, \dots, k^* + \ell^*\}. \end{cases} \tag{2}$$

Here ℓ^* stands for the length, k^* the beginning, $m^* = k^* + \ell^*$ the end and $|\tau^2 - \sigma^2|$ the size of a change. If H_0 is rejected, the next step is to estimate the parameters of the change: ℓ^*, k^*, m^*, σ and τ.

It is well known that the variance of a random variable X with known mean μ is a mean of a random variable $(X - \mu)^2$ and so change-point in variance problem can be solved using the methods for change-point in mean problem. Nevertheless testing for a change in variance of observations received particular interest.

If $m^* = n$ we have the usual – one change point problem, which was considered by many authors. Parametric model with ε_i being distributed normally was extencively analyzed in Chapter 2.2 of [1]. Bhatti and Wang [2] by the means of numerical simulations compared empirical power of representative five test statistics for this problem. See also references therein. For the non-parametric model we refer to [3]. The main methods used are likelihood, informational, Bayes in parametric case, in non-parametric case cumulative sums method is common one.

The problem of two change points in variance recieved far less attention by the researchers. Non-parametric model is investigated in [4]. Authors introduce cumulative sum type test statistics. These statistics are built using weight functions to detect 'short' epidemics. The consistency of the estimators for the epidemic length $\ell^* = m^* - k^*$ and the beginning k^* is also proved in that paper.

The problem of changing variance of a time series has its direct application when analyzing daily returns of stock markets composite indexes, or of currency exchange rate or other financial time series. It is observed that usually the variance of such series changes. It is said then that we observe volatility clustering – one of the few well known empirically observed characteristics of the financial time series. The problem is to identify the regions of the constant variance, see for example [5].

In the next section we introduce uniform increments test statistics (UI), state the result of convergence under H_0 and give the critical values of the limiting statistics. Then we introduce consistent statistics under H_A. We apply the proposed method for few financial time series. We end up with conclusions.

2. Test statistics and estimators

We will assume that $0 < \ell^* < n/2$, if $\ell^* \geqslant n/2$ we can treat $n - \ell^* \in (0, n/2]$ as the length of epidemic. For the details see [4]. Set for $k = 1, \ldots, n$

$$V_k^2 = \sum_{j=1}^{k} (X_j - \mu_n)^2, \tag{3}$$

and $V_0^2 = 0$, where

$$\mu_n = \begin{cases} \mu, & \text{when } \mu \text{ is known,} \\ \overline{X} = n^{-1}(X_1 + \cdots + X_n), & \text{when } \mu \text{ is unknown.} \end{cases}$$

Define

$$\delta_n^2 = n^{-1} \sum_{j=1}^{n} \left[(X_j - \mu_n)^2 - n^{-1} \sum_{j=1}^{n} (X_j - \mu_n)^2 \right]^2.$$

Throughout $0 \leqslant \alpha < 1/2$ and $p = (1/2 - \alpha)^{-1}$. Now define test statistics

$$V_{n,\alpha} = \max_{0<\ell<n/2} \ell^{-\alpha} \max_{0\leqslant k\leqslant n-\ell} \left| V_{k+\ell}^2 - V_k^2 - \frac{\ell}{n} V_n^2 \right|. \tag{4}$$

Let $(W(t), t \in [0,1])$ be a standard Wiener process and let $(B(t), t \in [0,1])$ be a Brownian bridge $B(t) = W(t) - tW(1)$, $t \in [0,1]$. Under the null hypothesis a limiting random variable for the statistic $V_{n,\alpha}$ is

$$V_\alpha = \sup_{0<h<1/2} h^{-\alpha} \sup_{0\leqslant t\leqslant 1-h} \left| B(t+h) - B(t) \right|.$$

In this paper we will state the main results of Rackauskas and Zuokas [4]. Exact formulations and the proofs of the theorems the reader can find in the mentioned paper.

First it is shown that under true H_0

$$n^{-1/p}\delta_n^{-1}V_{n,\alpha}\ \xrightarrow[n\to\infty]{\mathcal{D}}\ V_\alpha. \tag{5}$$

To find critical values of V_α requires numerical simulations. We have chosen several α values and generated 16500 (for every α) random values of the limiting statistic V_α. Empirical quantiles were taken as an approximation for the critical values associated with the fixed significance level α_{s}. Brownian bridge in each replication of V_α was approximated by the partial sum process $\xi_m(t) = (1/\sqrt{m})(\sum_{j\leqslant mt} Y_j - t\sum_{j=1}^m Y_j)$, $t \in [0,1]$, $\xi_m(0) = 0$. Here Y_j, $j = 1,\dots,m$, are independent standard normal random variables and $m = 2^{14}$. Table 1 gives the results.

Now assume that $\mathrm{H_A}$ holds. As an estimator of the epidemic length ℓ^* consider

$$\widehat{\ell^*} = \min\left\{\ell\colon\ V_{n,\alpha}^\circ(\ell) = \max_{0<j<n/2} V_{n,\alpha}^\circ(j)\right\}, \tag{6}$$

where

$$V_{n,\alpha}^\circ(\ell) = \ell^{-\alpha}\max_{0\leqslant k\leqslant n-\ell}\left|V_{k+\ell}^2 - V_k^2 - \frac{\ell}{n}V_n^2\right|.$$

Set

$$V(k,\ell) = \left|V_{k+\ell}^2 - V_k^2 - \frac{\ell}{n}V_n^2\right|,$$

and define

$$\widehat{k^*} = \min\left\{k\colon\ V(k,\widehat{\ell^*}) = \max_{0\leqslant i\leqslant n-\widehat{\ell^*}} V(i,\widehat{\ell^*})\right\}, \tag{7}$$

where $\widehat{\ell^*}$ is an estimator of the length ℓ^* defined by (6).

In [4] it is proved that $V_{n,\alpha}$, $\widehat{\ell^*}$ and $\widehat{k^*}$ are consistent statistics. As the estimators for τ and σ we take empirical standard deviations of the subsets with indexes $\{\widehat{k^*}+1,\dots,\widehat{k^*}+\widehat{\ell^*}\}$ and $\{1,\dots,n\}\setminus\{\widehat{k^*}+1,\dots,\widehat{k^*}+\widehat{\ell^*}\}$ accordingly.

3. Application

In this section we will apply the introduced test for a changed segment in variance and the procedures of locating the changed segment on the well known financial time series – daily closing values of Standard & Poor's composite 500 stock index. To be more

Table 1. The critical values

	V_0	$V_{1/16}$	$V_{1/8}$	$V_{3/16}$	$V_{1/4}$	$V_{5/16}$	$V_{3/8}$	$V_{7/16}$
$\alpha_s = 0.10$	1.606	1.712	1.834	1.984	2.172	2.423	2.793	3.440
$\alpha_s = 0.05$	1.726	1.838	1.969	2.123	2.309	2.563	2.937	3.577
$\alpha_s = 0.01$	1.962	2.080	2.217	2.380	2.591	2.844	3.227	3.878

precise we analyze not the values themselves but take the logarithmic one-day returns $X_i = \log(p_i/p_{i-1})$ where p_i denotes the original series. The time range is taken from the beginning of year 2000 to the beggining of September 2005, which includes 1428 observations.

First we test the series for independancy using Ljung-Box test. The p-value equal 0.319 does not reject the null hypothesis of independancy.

We then take several values of α: 0, 1/16, 1/8, 3/16, 1/4, 5/16, 3/8, 7/16 and test for the presence of a changed segment. As it was expected all test statistics reject the null hypothesis of no change. Next we estimate the length and the beginning of the changed segment. Fig. 1 illustrates the results.

For α equal 0, 1/16, 1/8 and 3/16 we obtain the same values of the parameters of the changed segment $\widehat{\ell^*} = 613$ from $\widehat{k^*} = 815$ to $\widehat{m^*} = 1427$ (dashed lines in Fig. 1). When $\alpha = 1/4$ we get $\widehat{\ell^*} = 596$, $\widehat{k^*} = 832$, $\widehat{m^*} = 1427$. For $\alpha = 5/16$ and 3/8 the same values $\widehat{\ell^*} = 77$, $\widehat{k^*} = 624$ and $\widehat{m^*} = 700$ (dotted lines in Fig. 1). Very similar values were obtained for $\alpha = 7/16$, 74, 627 and 700 accordingly. Time indexes from 624 to 700 corresponds the date range from 27th June to 15th October 2002 and 815 – 1st April 2003.

It is interesting that if we look at the original time series p_i and mark the estimated location of the changed segment in variance of $X_i = \log(p_i/p_{i-1})$ we observe the following: the index value p_i between the dashed lines has an upward trend and is associated with the decreased variance of X_i (see Fig. 1); the variance of X_i between the dotted lines is increased, looking at the original time series in Fig. 2 this time range can be interpreted transitional between the decreasing and increasing character of the index.

The next natural generalization in changed segment problem is to consider several changed segments in observations. This raises the problem of estimating the number of changed segments and showing that such estimate is consistent with the real number of changed segments. The other important problem is to establish data-driven procedure of the choice of parameter α, because now it is chosen arbitrary.

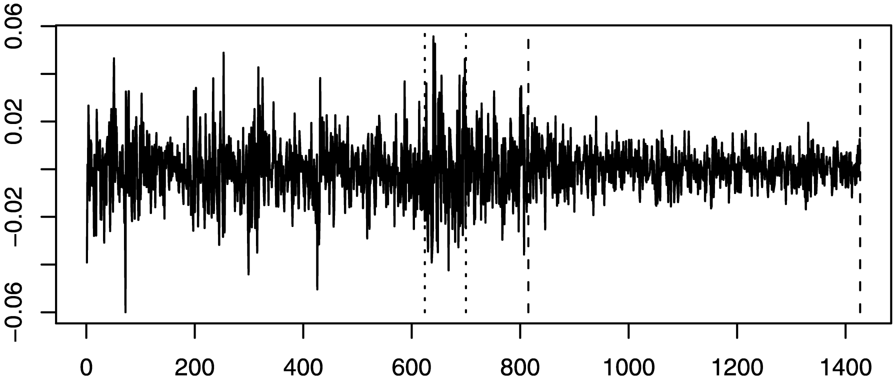

Fig. 1. Log-returns of S&P.

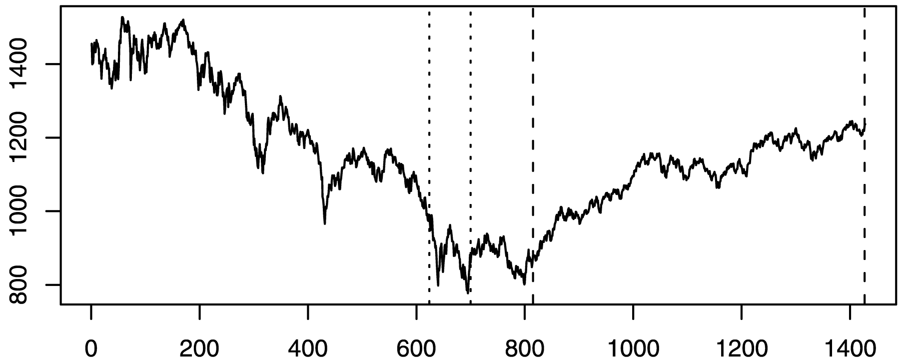

Fig. 2. S&P index.

References

1. J. Chen, A.K. Gupta, *Parametric Statistical Change Point Analysis*, Birkhäuser Verlag, Basel (2000).
2. M.I. Bhatti, J. Wang, On testing for a change-point in variance of normal distribution, *Biom. J.*, **42**, 1021–1032 (2000).
3. E. Gombay, L. Horváth, M. Hušková, Estimators and tests for change in variances, *Stat. Decis.*, **14**, 145–159 (1996).
4. A. Račkauskas, D. Zuokas, Off-line testing for a changed segment in the sample variance, *Lith. Math. J.*, **45**, 200–216 (2005).
5. G. Chen, Y.K. Choi, Y. Zhou, Nonparametric estimation of structural change points in volatility models for time series, *J. Econometrics*, **126**, 79–114 (2005).

REZIUMĖ

D. Zuokas. Pasikeitusio dispersijos segmento testavimas ir taikymas

Šiame darbe nagrinėjamas pasikeitusio dispersijos segmento testavimo ir radimo uždavinys. Šiam uždaviniui suformuluoti tolygiųjų prieauglių (UI) statistikos bei pasikeitusio segmento ilgio bei pradžios konvergavimo ir suderinamumo rezultatai. Pasiūlyta procedūra pritaikyta sudėtinio Standard & Poor's 500 akcijų indekso logaritminės grąžos dispersijos tyrimui. Rezultatai rodo, kad šis testas, priklausomai nuo parametro α, gali aptikti įvairaus ilgio pasikeitusius segmentus.