A WAVELET-BASED APPROACH TO BREAKPOINT DETECTION

Milan Bašta

Abstract

Testing for the presence of breakpoints is a very important topic in the analysis of financial, economic or demographic time series. We present two wavelet-based approaches (statistical tests) to breakpoint detection available in the literature. One of the approaches is based on the maximal overlap discrete wavelet transform (MODWT), the other being based on the maximal overlap discrete wavelet packet transform (MODW*P*T). The latter approach was originally suggested as an improvement of the former one and consists of several steps. We use Monte Carlo simulations to estimate the probability of type I error for the latter approach with all the necessary steps included as part of the simulation, which mimics a real-life scenario. We show that the latter approach may not be valid under some circumstances and discuss the reasons for such a result. To our best knowledge, this (undesirable) property of the approach has not been documented in the literature.

Key words: breakpoint detection, time series, wavelets, Monte Carlo simulation

JEL Code: C12, C15, C49

Introduction

The detection of breakpoints is a very important topic in the analysis of financial, economic and demographic time series. There are several wavelet-based approaches to breakpoint detection (see e.g. Whitcher, 1998, Gabbanini et al., 2004, Cho and Fryzlewicz, 2012, Killick et al., 2013, Nason, 2013). The test by Whitcher (1998) uses the maximal overlap discrete wavelet transform (MODWT). Employing the maximal overlap discrete wavelet *packet* transform (MODW*P*T), Gabbanini et al. (2004) extended the test by Whitcher (1998) in order to handle the case where a crucial assumption of the test by Whitcher et al. (1998) is not satisfied. Gabbanini et al. (2004) provide a Monte Carlo simulation in their paper to support the argument that their extension is superior to the original test by Whitcher (1998).

However, we note that the simulation by Gabbanini et al. (2004) is performed under conditions which may often not be realistic (in real-life time series analysis). The goal of our paper is to run the simulations under realistic conditions and explore whether the test by Gabbanini et al. (2004) is indeed a valid hypothesis test under such conditions.

The paper is organized as follows. Section 1 provides an introduction to the MODWT. Section 2 presents the test by Whitcher (1998). Section 3 introduces an extension of the test suggested by Gabbanini et al. (2004). In Section 4, we perform a Monte Carlo simulation where we explore the validity of the test by Gabbanini et al. (2004) and discuss the results.

1 MODWT

A detailed introduction to the MODWT is beyond the scope of this paper and the interested reader is referred to Percival and Walden (2006, Sec. 5). We briefly introduce the notion of the MODWT wavelet filters (Section 1.1) and of the MODWT wavelet coefficients (Section 1.2). These coefficients will be used in the test by Whitcher (1998) in Section 2.

1.1 MODWT wavelet filters

There are various types of wavelets such as the Haar, D(4) wavelets, etc. (see Percival and Walden, 2006). For each type of wavelet, a special set of linear filters is available. The so-called jth level ($j = 1, 2, \dots$) wavelet filter is denoted as $\{h_{j,l}: l = 0, ..., L_j - 1\}$, where $L_j = (2^j - 1)(L_1 - 1) + 1$ is the length of the filter, L_1 being the length of the first level filter. $\{h_{j,l}\}$ is an approximate band-pass filter for the range of frequencies $[1/2^{j+1}, 1/2^j]$.

1.2 MODWT wavelet coefficients

Let us assume a stochastic process $\{X_t: t = ..., -1, 0, 1, \dots\}$. The MODWT wavelet coefficients of level j ($j = 1, 2, \dots$) for $\{X_t\}$, denoted as $\{W_{j,t}: t = \dots, -1, 0, 1, \dots\}$, are obtained by linear filtering $\{X_t\}$ with $\{h_{j,l}\}$, i.e.

$$W_{j,t} \equiv \sum_{l=0}^{L_j - 1} h_{j,l} X_{t-l}, \quad t = \dots,-1,0,1,\dots \tag{1}$$

Since $\{W_{j,t}\}$ are obtained by linear filtering $\{X_t\}$ with an approximate band-pass filter for the range of frequencies $[1/2^{j+1}, 1/2^j]$, $\{W_{j,t}\}$ capture the dynamics of $\{X_t\}$ associated with the frequency range $[1/2^{j+1}, 1/2^j]$.

Percival and Walden (2006) show that if $\{X_t\}$ is stationary, $\{W_{j,t}\}$ is stationary with a zero mean. Moreover, if $\{W_{j,t}\}$ is *downsampled* by 2^j, it becomes approximately uncorrelated for several types of processes – this happens if the spectrum of $\{X_t\}$ is approximately flat in the frequency range $[1/2^{j+1}, 1/2^j]$ (see Percival and Walden, 2006, Sec. 9). Consequently, if $\{X_t\}$ is a stationary Gaussian process, $\{W_{j,t}\}$ *downsampled* by 2^j can be considered a Gaussian white noise.

In real-life applications, where only a portion $\{X_t: t = 0, \ldots, N-1\}$ of the process $\{X_t: t = ..., -1, 0, 1, \ldots\}$ is available, the MODWT wavelet coefficients of level j cannot be obtained for times $t = 0, \ldots, L_j - 2$ since this would require the knowledge of X_t for $t < 0$. Consequently, only $\{W_{j,t}: t = L_j - 1, \ldots, N - 1\}$ can be obtained (assuming that $N \geq L_j$).

2 Test for a variance change

Based on the properties of the *downsampled* MODWT wavelet coefficients given in Section 1, Whitcher (1998), Whitcher et al. (2000) and Percival and Walden (2006) formulate a wavelet-based hypothesis test for a *variance* change (break) in a Gaussian time series $\{X_t: t = 0, ..., N-1\}$. The idea of the test is that the MODWT wavelet coefficients of level j *downsampled* by 2^j are approximately a realization of a Gaussian white noise for many stationary Gaussian processes. Consequently, a testing approach similar to the one employed in Brown et al. (1975) and Inclan and Tiao (1994) can be adopted to test for a change in the variance of the downsampled MODWT wavelet coefficients.

More specifically, let us assume the jth level MODWT wavelet coefficients $\{W_{j,t}: t = L_j - 1, ..., N - 1\}$ for $\{X_t: t = 0, ..., N - 1\}$. Downsampling $\{W_{j,t}: t = L_j - 1, ..., N - 1\}$ by 2^j and relabeling the original coefficients, the following sequence is obtained

$$\{w_{j,1}, w_{j,2}, \ldots, w_{j,N_j}\}, \tag{2}$$

where

$$w_{j,k} \equiv W_{j,L_j - 1 + 2^j (k-1)}, \quad k = 1, \ldots, N_j, \tag{3}$$

$$N_j \equiv 1 + \lfloor (N - L_j)/2^j \rfloor \tag{4}$$

The following hypothesis test is performed to assess the significance of the change in variance

$$H_0 : \text{var}\{w_{j,1}\} = \text{var}\{w_{j,2}\} = \ldots = \text{var}\{w_{j,N_j}\}, \tag{5}$$

$$H_1 : \text{var}\{w_{j,1}\} = \ldots = \text{var}\{w_{j,z^*}\} \neq \text{var}\{w_{j,z^*+1}\} = \ldots = \text{var}\{w_{j,N_j}\}, \tag{6}$$

where z^* is an unknown index where the variance change occurred. Let the normalized cumulative sum of squares (of the downsampled MODWT wavelet coefficients) be defined as

$$p_k \equiv \frac{\sum_{z=1}^{k} w_{j,z}^2}{\sum_{z=1}^{N_j} w_{j,z}^2}, \quad k = 1, \ldots, N_j - 1. \tag{7}$$

Further, let the test statistic be defined as

$$d \equiv \max\{d^+, d^-\}, \tag{8}$$

where

$$d^+ \equiv \max_{k=1,\ldots,N_j-1}\left(\frac{k}{N_j - 1} - p_k\right), \tag{9}$$

$$d^- \equiv \max_{k=1,\ldots,N_j-1}\left(p_k - \frac{k-1}{N_j - 1}\right). \tag{10}$$

Let the value of d for our actual sample be denoted as d^*. One way to obtain the p-value of the test is to employ Monte Carlo simulations. Specifically, we simulate a Gaussian white noise process of length N_j and calculate the test statistic d. We repeat this N_{sim} times, yielding N_{sim} test statistics d_1, ..., d_{Nsim} corresponding to individual runs of the simulation. The p-value of the test is approximated as a portion of instances where the test statistics from the Monte Carlo simulation were greater than or equal to d^*. If the null hypothesis of no variance change is rejected, the jth level MODWT wavelet coefficients can further be used to estimate the time of the change (see Whitcher, 1998, Percival and Walden, 2006).

3 MODWPT-based approach

Gabbanini et al. (2004) suggest that there is a potential benefit in using the maximal overlap discrete wavelet *packet* transform (MODW*P*T) instead of the MODWT to test for the variance change. Consequently, we introduce the MODW*P*T[1] in Section 3.1. The use of the MODW*P*T for the variance change detection is explained in Section 3.2 and Section 3.3.

3.1 MODWPT coefficients

[1] A detailed introduction to the MODWPT can be found in Percival and Walden (2006, Sec. 6)

The (j, n) MODW*P*T coefficients for $\{X_t: t = \dots, -1, 0, 1, \dots\}$ $(j = 1, 2, \dots; n = 1, \dots, 2^j - 1)$ are denoted as $\{W_{j,n,t}: t = \dots, -1, 0, 1, \dots\}$ and obtained by linear filtering $\{X_t\}$ with a special linear filter which can be considered an approximate band-pass filter for the frequency range $[n/2^{j+1}, (n+1)/2^{j+1}]$, the length of the filter being L_j. $\{W_{j,n,t}\}$ is, therefore, associated with the dynamics of $\{X_t\}$ in the frequency range $[n/2^{j+1}, (n+1)/2^{j+1}]$. There are two subscripts, namely j (level) and n, needed to distinguish various sets of the MODW*P*T coefficients.

$\{W_{j,n,t}\}$ $(j = 1, 2, \dots; n = 1, \dots, 2^j - 1)$ shares analogous properties to those of $\{W_{j,t}\}$ $(j = 1, 2, \dots)$ mentioned in Section 1.2. Specifically, $\{W_{j,n,t}\}$ is stationary with a zero mean if $\{X_t\}$ is stationary. $\{W_{j,n,t}\}$ is Gaussian if $\{X_t\}$ is Gaussian. Further, if downsampled by 2^j, $\{W_{j,n,t}\}$ may become approximately uncorrelated – this happens if the spectrum of $\{X_t\}$ is approximately flat in the frequency range $[n/2^{j+1}, (n+1)/2^{j+1}]$.

In real-life data analysis, the (j, n) MODW*P*T coefficients for $\{X_t: t = 0, \dots, N-1\}$ are given as $\{W_{j,n,t}: t = L_j - 1, \dots, N-1\}$.

Since the properties of $\{W_{j,n,t}\}$ $(j = 1, 2, \dots; n = 1, \dots, 2^j - 1)$ are similar to those of $\{W_{j,t}\}$ $(j = 1, 2, \dots)$, the MODW*P*T coefficients can potentially be used in the hypothesis test of Section 2 instead of the MODWT coefficients.

3.2 MODWPT coefficients for change detection

The potential benefit of the MODW*P*T for the detection of variance changes stems from the following argument. Let the largest value of j used both in the MODWT and MODW*P*T analysis be denoted as J. The actual choice of J is dictated by the requirement to have at least "a few" coefficients at level J after downsampling by 2^J.

Consequently, there is a total of J sets of the MODWT wavelet coefficients (associated with frequency ranges $[1/2^{j+1}, 1/2^j]$, for $j = 1, \dots, J$) and a total of $(2^1 - 1) + \dots + (2^J - 1) \geq J$ sets of the MODW*P*T coefficients (associated with frequency ranges $[n/2^{j+1}, (n+1)/2^{j+1}]$, for $j = 1, \dots, J$; $n = 1, \dots, 2^j - 1$). When working with the MODWT sets only, it can happen that none of the MODWT sets is uncorrelated after downsampling by 2^j because the spectrum of $\{X_t\}$ is not approximately flat in any of the frequency ranges $[1/2^{j+1}, 1/2^j]$, for $j = 1, \dots, J$. However, the chance of having a flat spectrum in *none* of the ranges $[n/2^{j+1}, (n+1)/2^{j+1}]$, $j = 1, \dots, J$; $n = 1, \dots, 2^j - 1$, is lower. Consequently, it is more likely to find such a combination of j and n $(j = 1, \dots, J; n = 1, \dots, 2^j - 1)$ for which the corresponding (j, n) MODW*P*T coefficients will approximately be uncorrelated after downsampling by 2^j. Note, that the assumption of no correlation is an inherent assumption of the test for variance change.

3.3 The procedure

Assuming an input time series $\{X_t = 0, \ldots, N - 1\}$, we may follow the procedure proposed by Gabbanini et al. (2004) to test for a variance break in the time series:

1.) We start with level $j = 1$.

2.) For a given level j, we perform the Ljung-Box test (Ljung and Box, 1978) of no autocorrelation for the (j, n) MODW*P*T coefficients in turn for $n = 1, \ldots, 2^j - 1$. If the null hypothesis of no autocorrelation is not rejected for at least one of the values of n, we select the optimal (j, n) MODW*P*T coefficients as those corresponding to the value of n associated with the lowest Ljung-Box test statistic and move to step **3.)** below. On the other hand, if the null hypothesis of no autocorrelation is rejected for all $n = 1, \ldots, 2^j - 1$, and if $j < J$, we increase the level j by one, i.e. $j \rightarrow j + 1 \leq J$, and repeat step **2.)** with the new level; whereas if $j = J$, we select the optimal (j, n) MODW*P*T coefficients as those corresponding to the value of n associated with the lowest Ljung-Box test statistic and move to step **3.)** below.

3.) Employing the (j, n) MODW*P*T coefficients selected in the previous step, the hypothesis test for the variance change (break) is performed (Section 2).

4 Monte Carlo simulation

Gabbanini et al. (2004) perform a Monte Carlo simulation to estimate the probability of type I error of the above described MODW*P*T-based test for the variance change (break). However, the simulation run by Gabbanini et al. (2004) does not correspond to a real-life scenario since (in their simulation) Gabbanini et al. (2004) assume that the “optimal” doublet (j, n) is not selected by the procedure outlined in Section 3.3, but is known *in advance* (specifically, they assume the (2, 3) doublet for the ARMA process considered in their simulation).

Consequently, employing the R software (R Core Team, 2014) and the wmtsa contributed R package (Constantine and Percival, 2013), we perform a simulation different from that given in Gabbanini et al. (2004) in the sense that we assume that the doublet (j, n) is *not* known in advance and has to be selected by the procedure outlined in Section 3.3. This is a more realistic scenario compared to the one considered by Gabbanini et al. (2004) in their simulation. Since the extent of the paper is limited, we will be interested only in the validity of the test (i.e. in the fact whether the probability of type I error matches the nominal significance level) and will illustrate the results for one specific ARMA process only.

4.1 Setting of the simulation

We assume the following ARMA(1, 1) process:

$$X_t = 0.9X_{t-1} + a_t + 0.5a_{t-1}, \tag{11}$$

where $\{a_t\}$ is a Gaussian white noise with unit variance. Further, we use D(4) wavelets ($L_1 = 4$ for D(4) wavelets). The Ljung-Box test is performed at a nominal significance level of 0.05 with the lag parameter set to 3. Further, a nominal significance level of 0.05 is assumed in the test for the variance change outlined in Section 2.

We estimate the probability of type I error for the following three scenarios:

1. Doublet (j, n) = (1, 1) is chosen *in advance* to be used in the test, the selection procedure outlined in steps **1.)** and **2.)** in Section 3.3 *not* being performed. We have verified that (the stochastic process of) the (1, 1) MODW*P*T coefficients downsampled by 2^1 is rather strongly autocorrelated at lag 1 for the above given ARMA process. Consequently, Scenario 1 is expected to provide a hypothesis test which is not valid (i.e. a test whose probability of type I error does not match the nominal significance level).
2. Doublet (j, n) = (2, 2) is chosen *in advance* to be used in the test, the selection procedure outlined in steps **1.)** and **2.)** in Section 3.3 *not* being performed. We have verified that (the stochastic process of) the (2, 2) MODW*P*T coefficients downsampled by 2^2 is only weakly autocorrelated for the above given ARMA process. Scenario 2 is thus expected to lead to an approximately valid hypothesis test.
3. The test is performed as outlined in Section 3.3, i.e. it also includes the “selection steps”. We use $J = 2$ (for all the lengths of the input time series explored, see below).

For each of the three scenarios, we run the simulation for the following lengths (N) of the time series: 32, 64, 128, 256, 512 and 1024. For each length, 10,000 simulations are performed. Subsequently, since the process is simulated under H_0 (i.e. without any variance change), the probability of type I error can readily be estimated. The results are provided in Table 1, the estimated probabilities of type I error for the three scenarios being denoted as P(rej. H_0, Scenario 1), P(rej. H_0, Scenario 2) and P(rej. H_0, Scenario 3).

Tab. 1: Estimated probabilities of type I error (nominal significance level = 0.05)

	N = 32	N = 64	N = 128	N = 256	N = 512	N = 1024

P(rej. H_0, Scenario 1)	0.071	0.088	0.100	0.114	0.119	0.121
P(rej. H_0, Scenario 2)	0.053	0.048	0.052	0.059	0.063	0.061
P(rej. H_0, Scenario 3)	0.068	0.062	0.054	0.057	0.063	0.060

Source: Own construction.

As has been expected, the estimated probability of type I error deviates (largely) from the nominal significance level (0.05) for Scenario 1. On the other hand, the estimated probability of type I error is not far from the nominal significance level for Scenario 2.

Interesting results (different from those for Scenario 2) have been obtained for Scenario 3. To elaborate on the results for Scenario 3, we can note that

$$P(rej.\, H_0, \text{Scenario 3}) = \sum_{(j,n)} P(rej.\, H_0 \mid j, n) P(j, n), \qquad (12)$$

where $P(rej.\, H_0 \,|\, j,\, n)$ is the estimated probability of type I error (for Scenario 3) *conditional* on doublet (j, n) being selected as the optimal doublet in the selection procedure, and $P(j, n)$ denotes the estimated probability that doublet (j, n) will be selected as the optimal doublet in the selection procedure. Summation in Equation 12 runs over the following four doublets: (1, 1), (2, 1), (2, 2) and (2, 3).

Table 2 provides the estimated probabilities $P(rej.\, H_0 \,|\, j,\, n)$ and $P(j, n)$ for $N = 32$. It is interesting to note that (for $N = 32$) the conditional probabilities $P(rej.\, H_0 \,|\, 1,\, 1)$ and $P(rej.\, H_0 \,|\, 2,\, 2)$ differ from $P(rej.\, H_0$, Scenario 1) and $P(rej.\, H_0$, Scenario 2). Specifically, $P(rej.\, H_0 \,|\, 1,\, 1)$ is closer to the nominal significance level than $P(rej.\, H_0 \,|\, 2,\, 2)$ despite the fact that $P(rej.\, H_0$, Scenario 1) is further from the nominal significance level than $P(rej.\, H_0$, Scenario 2). Paradoxically, this suggests that (for $N = 32$) the (2, 2) doublet is "no more optimal" for performing the test *conditional on the fact that* the doublet was selected as the optimal one by the selection procedure. The conditional probabilities $P(rej.\, H_0 \,|\, j,\, n)$ are further weighted by $P(j, n)$ in Equation 12. Since the Ljung-Box test has a low power when $N = 32$, it often fails to reject the null hypothesis of no autocorrelation for the (1, 1) doublet and the selection procedure often terminates with the (1, 1) doublet being selected as the optimal one.

Tab. 2: $N = 32$, P(rej. $H_0 \,|\, j, n$) and $P(j, n)$

	$n = 1$	$n = 2$	$n = 3$
$j = 1$	0.0652, 0.7872	X	X
$j = 2$	0.0779, 0.0719	0.0840, 0.0726	0.0747, 0.0683

Source: Own construction.

Different results are obtained for N = 1024 (Table 3). Here, the selection procedure "never" selects the (1, 1) doublet as the optimal one and mostly chooses the (2, 2) doublet. This is not surprising since the Ljung-Box test has a high power when N = 1024 and is capable of selecting those coefficients that are truly the least correlated after downsampling. Moreover, for N = 1024, $P(rej.\,H_0\,|\,2,\ 2)$ = 0.0600, which is very close to P(rej. H_0, Scenario 2) = 0.061. As a result, P(rej. H_0, Scenario 3) is close to P(rej. H_0, Scenario 2) for N = 1024.

Tab. 3: N = 1024, P(rej. H_0 | j, n) and P(j, n)

	n = 1	n = 2	n = 3
j = 1	NA , 0.0000	X	X
j = 2	0.0698, 0.0086	0.0600, 0.9900	0.0000, 0.0014

Source: Own construction.

Conclusion

We have shown that the wavelet-based test for variance change proposed by Gabbanini et al. (2004) as an improvement of the test by Whitcher (1998) does not always guarantee a valid hypothesis test, the reason for this being the fact that the test by Gabbanini et al. (2004) is a complex multistep procedure and its performance has to be evaluated with all the steps included. If all the steps are included, the probability of type I error may differ from the nominal significance level and the test may become invalid. Consequently, we recommend selecting the optimal doublet in advance based on a priori information if possible rather than by the means of the Ljung-Box test. If the Ljung-Box test is indeed employed to select the optimal doublet, bootstrap methods can be recommended to explore the validity of the test.

Acknowledgment

This paper was written with the support of the Czech Science Foundation (GA CR), grant no. P402/12/P507, Modelling of Financial and Economic Time Series - Application and Comparison of Wavelet and Traditional Methods.

References

Brown, R. L., Durbin, J., & Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)*, *37*, 149-192.

Cho, H., & Fryzlewicz, P. (2012). Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *Statistica Sinica*, *22*, 207-229.

Constantine, W., & Percival, D. (2013). wmtsa: Wavelet Methods for Time Series Analysis. R package version 2.0-0. URL: http://CRAN.R-project.org/package=wmtsa.

Gabbanini, F., Vannucci, M., Bartoli, G., & Moro, A. (2004). Wavelet packet methods for the analysis of variance of time series with application to crack widths on the Brunelleschi dome. *Journal of Computational and Graphical Statistics*, *13*(3), 639-658.

Inclan, C., & Tiao, G. C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association*, *89*(427), 913-923.

Killick, R., Eckley, I. A., & Jonathan, P. (2013). A wavelet-based approach for detecting changes in second order structure within nonstationary time series. *Electronic Journal of Statistics*, *7*, 1167-1183.

Ljung, G. M., & Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297-303.

Nason, G. (2013). A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *75*(5), 879-904.

Percival, D. B., & Walden, A. T. (2006). *Wavelet methods for time series analysis* (Vol. 4). Cambridge university press.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

Whitcher, B., Guttorp, P., & Percival, D. B. (2000). Multiscale detection and location of multiple variance changes in the presence of long memory. *Journal of Statistical Computation and Simulation*, *68*(1), 65-87.

Whitcher, B. J. (1998). *Assessing nonstationary time series using wavelets* (Doctoral dissertation, University of Washington).

Contact

Milan Bašta

Department of Statistics and Probability, University of Economics in Prague

Katedra statistiky a pravděpodobnosti, Fakulta informatiky a statistiky, Vysoká škola ekonomická v Praze, nám. W. Churchilla 1938/4, 130 67 Praha 3
milan.basta@vse.cz