DETECTING CHANGES IN THE DEPENDENCE STRUCTURE OF A TIME SERIES

A. Dürre, R. Fried
Technische Universität Dortmund
Dortmund, GERMANY
e-mail: alexander.duerre@udo.edu

Abstract

We propose a new robust test to detect changes in the dependence structure of a time series. The test is based on empirical autocovariances of a robust transformation of the original time series. Because of the transformation we do not require any finite moments of the original time series making the test especially suitable for heavy tailed time series. We furthermore propose a lag weighting scheme which puts emphasis on changes of the autocorrelation at smaller lags. Our approach is compared to existing ones in some simulations.
Keywords: data science, dependence structure, changes detection, time series

1 Introduction

Detecting changes in the dependence structure of a time series goes back to [34] and [22]. To the best of our knowledge the first test to detect a change in the dependence structure where the possible time of change is not known a priori can be found in [33]. Since then a lot of alternatives have been proposed. In [23] estimated autocovariances of subsamples are compared to the estimation based on the whole time series. For linear models several tests have been proposed, see [3], [4], [2], [13] and [1]. CUSUM-type tests to detect changes in one or several autocovariances have been derived in [5], [24] and [14]. A test based on the auto-copula has been proposed in [7]. Tests which check stationarity of the spectrum are presented in [30], [17] and [36] and a wavelet periodogram is used in [26] and [8]. There are also proposals which compare local estimates of the spectrum with a global estimation, see [37], [28], [29] [15] and [32].
Surprisingly little attention has been paid to robustness. We want to fill the gap with a CUSUM type test based on robustified autocovariances. The testing procedure is described in Section 2 and a small simulation study in Section 3 indicates the usefulness of the proposed test.

2 Testing procedure

Denote $\mathbb{X} = X_1, \ldots, X_T$ a one dimensional time series which is stationary under the null-hypothesis. We assume in the following that $\mathbb{X}$ has a continuous marginal distribution and is strongly mixing with mixing coefficients $(a_k)_{k\in\mathbb{N}}$ fulfilling $a_k = O(k^{-1-\epsilon})$ for some $\epsilon > 0$. Strong mixing was first introduced in [35] and describes how fast the dependence between two observations decreases as the time lag between them increases,

see [6] for more details. We only want to emphasize here that a broad class of time series models is strongly mixing, like linear and GARCH processes with continuously distributed innovations, see [9] and [25].
We want to test whether the autocorrelation function of $\mathbb{X}$ stays the same, concentrating on the first p lags. We follow the approach of [16] and use bounded transformations. Before using them, the observations need to be properly standardized. Denote therefore by $\hat{\mu}$ the sample median and by $\hat{\sigma}$ the sample MAD of $\mathbb{X}$, and μ and σ their theoretical counterparts. Then we define

$$\hat{Y}_i = \psi\left(\frac{X_i - \hat{\mu}}{\hat{\sigma}}\right) \quad \text{and} \quad Y_i = \psi\left(\frac{X_i - \mu}{\sigma}\right), \quad \text{where} \quad \psi = \begin{cases} -k & x < -k \\ x & |x| \leq k \\ k & x > k \end{cases}$$

denotes the Huber-ψ function. This function was originally introduced for location estimation in [18] and basically downweights the influence of observations with large absolute values by shrinking them to more plausible values, namely $-k$ respectively k. The tuning-coefficient k determines the robustness of the test. A larger value of k is favourable under Gaussian time series whereas a smaller k is needed if the data is corrupted or heavy tailed. In [19] $k = 1.5$ is recommended as a compromise.
In the following we derive a CUSUM type test based on the Huber-transformed time series. Denote therefore $S_k^{(l)} = \sum_{t=1}^{k} \hat{Y}_i \hat{Y}_{i+l}$, then we look at

$$R_T = \max_{k=1,\ldots,\tilde{T}} \frac{1}{\tilde{T}} \begin{pmatrix} S_k^{(1)} - \frac{k}{\tilde{T}} S_{\tilde{T}}^{(1)} \\ \vdots \\ \vdots \\ S_k^{(p)} - \frac{k}{\tilde{T}} S_{\tilde{T}}^{(p)} \end{pmatrix}^T \begin{pmatrix} w_1 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & w_p \end{pmatrix} \begin{pmatrix} S_k^{(1)} - \frac{k}{\tilde{T}} S_{\tilde{T}}^{(1)} \\ \vdots \\ \vdots \\ S_k^{(p)} - \frac{k}{\tilde{T}} S_{\tilde{T}}^{(p)} \end{pmatrix}$$

where $w_1, \ldots, w_p > 0$ and $\tilde{T} = T - p$. Here are some remarks with regard to R_T :

- Technically R_T tests the following hypothesis

$$H_0 : \begin{pmatrix} \mathrm{Cov}(Y_1, Y_2) \\ \vdots \\ \mathrm{Cov}(Y_1, Y_{p+1}) \end{pmatrix} = \ldots = \begin{pmatrix} \mathrm{Cov}(Y_{T-p}, Y_{T-p+1}) \\ \vdots \\ \mathrm{Cov}(Y_{T-p}, Y_T) \end{pmatrix} \quad \text{vs.}$$

$$H_1 : \ \exists k < T : \begin{pmatrix} \mathrm{Cov}(Y_{k-p}, Y_{k-p+1}) \\ \vdots \\ \mathrm{Cor}(Y_{k-p}, Y_k) \end{pmatrix} \neq \begin{pmatrix} \mathrm{Cor}(Y_k, Y_{k+1}) \\ \vdots \\ \mathrm{Cor}(Y_k, Y_{p+k}) \end{pmatrix}.$$

  This is not equivalent to a test for a stationary dependence structure. For example R_T will have problems to detect changes in the tail dependence, since extreme values are downweighted by ψ.

- We decided against calculating real robust correlations by standardizing $S_k^{(j)}$ by $S_k^{(0)}$ respectively $S_{\tilde{T}}^{(0)}$ for $j = 1, \ldots, p$. Note that we already standardize our observations by $\hat{\sigma}$. So there is no mandatory need to standardize $S_k^{(j)}$, too.

- The choice of p is crucial for the power of the test. If there is only a change in the first lag a large p only adds noise and can mask the change point. On the other hand if one chooses p to small one cannot detect changes in the higher lags. Furthermore one has to keep in mind that the estimation of Σ gets very poor if p is large compared to $\tilde{T}$. As a rule of thumb use $p < \lfloor \tilde{T}/20 \rfloor$.

- In the multivariate context it is common and beneficial to use the quadratic form with respect to Σ, the asymptotic long run variance covariance matrix of $S_{\tilde{T}}^{(1)}, \dots, S_{\tilde{T}}^{(p)}$. In this case R_T gets affine invariant. However, in the time series context this property is not desirable. The weights $w_1, \dots, w_p$ gives us more flexibility. Usually one would choose descending weights to smooth the transition between lags of the acf where one can detect a change $j = 1, \dots, p$ to those neglected $j > p$. If there is only a change in the first autocorrelation and one chooses p to large the change could be masked by the noise from the other autocorrelations. Descending weights somehow counteract this problem. Without further knowledge we suggest using $w_i = 1 - (i-1)/p$ for $i = 1, \dots, p$. A disadvantage of using weights instead of Σ is that R_T depends on the actual dependence structure of $\mathbb{X}$. Therefore one cannot use tabulated asymptotical critical values. However, one can approximate the distribution of R_T by sampling Gaussian processes with the estimated covariance structure.

Now we describe how one can approximate the distribution of R_T under the null-hypothesis. Under the above assumptions one can use Theorem 1 of [16]. It is not explicitly stated there but effectively proved in Proposition 1 and 2 that

$$\frac{1}{\sqrt{\tilde{T}}} \left[S_{\lfloor \tilde{T}x \rfloor}^{(1)} - \frac{\lfloor \tilde{T}x \rfloor}{\tilde{T}} S_{\tilde{T}}^{(1)}, \dots, S_{\lfloor \tilde{T}x \rfloor}^{(p)} - \frac{\lfloor \tilde{T}x \rfloor}{\tilde{T}} S_{\tilde{T}}^{(p)} \right]_{x\in[0,1]} \xrightarrow{w} [\boldsymbol{BB}(x)]_{x\in[0,1]}$$

where $[\boldsymbol{B}(x)]_{x\in[0,1]}$ is a Gaussian process with mean function $g(x) = \mathbf{0}$ and covariance function $\gamma(x, y) = x(1-y)\Sigma$ for $0 \le x \le y \le 1$. Here, Σ is the asymptotic long run covariance matrix defined by

$$\Sigma = \sum_{h=-\infty}^{\infty} \mathrm{Cov} \left(\begin{bmatrix} Y_1 Y_2 \\ \vdots \\ Y_1 Y_{1+p} \end{bmatrix}, \begin{bmatrix} Y_{1+h} Y_{2+h} \\ \vdots \\ Y_{1+h} Y_{1+p+h} \end{bmatrix} \right).$$

Proposition 3 in [16] states that Σ can be consistently estimated by a kernel estimator. Denote therefore $b_{\tilde{T}} \ge 0$ a bandwidth and $k : \mathbb{R} \to [-1, 1]$ a kernel function. Then $\hat{\Sigma}$ with the elements

$$\hat{\Sigma}_{[i,j]} = \frac{1}{T} \sum_{t=1}^{\tilde{T}} \sum_{s=1}^{\tilde{T}} (\hat{Y}_s \hat{Y}_{s+i} - S_{\tilde{T}}^{(i)})(\hat{Y}_t \hat{Y}_{t+j} - S_{\tilde{T}}^{(j)}) k\left(\frac{|s-t|}{b_{\tilde{T}}} \right)$$

is the related kernel estimator. Simulations indicate that the flat-top kernel

$$k(x) = \begin{cases} 1 & 0 \le |x| \le 0.5 \\ 2 - 2|x| & 0.5 < |x| \le 1 \\ 0 & |x| > 1 \end{cases}$$

proposed [31] in works well together with $b_T = \tilde{T}^{\frac{1}{3}}$ under autoregressive processes of order 1. One can generate random variables $\tilde{R}_T^{(i)},\ i = 1, \ldots, m$, which have asymptotically the same distribution as R_T under the null-hypothesis by the following algorithm:

- generate $p \cdot \tilde{T}$ independent standard normal random variables and store them in a $\tilde{T} \times p$ matrix Z
- reproduce the cross sectional dependence by multiplying Z with L of the Cholesky decomposition $\hat{\Sigma} = LL^T$: Set $V = Z \cdot L$
- calculate the weighted test statistic

$$\tilde{R}_T = \frac{1}{T} \max_{k=1,\ldots,\tilde{T}} \left(\sum_{t=1}^{k} V_{[t,]} - \frac{k}{\tilde{T}} \sum_{t=1}^{T} V_{[t,]} \right) W \left(\sum_{t=1}^{k} V_{[t,]} - \frac{k}{\tilde{T}} \sum_{t=1}^{T} V_{[t,]} \right)^T$$

By this algorithm one can generate random variables to calculate approximate p-values very fast. We recommend using a modified Cholesky decomposition to safeguard against numeric instabilities which could arise especially if T is small compared to p. In our simulations we used the algorithm proposed in [38].

3 Simulations

We want to assess our approach in a small simulation study. We compare our method with tests for second order stationarity which are available in R, namely two wavelet based tests [26] and [10], which are implemented in the packages [27] respectively [11], and a revised version of the ANOVA test originally proposed in [33], which is implemented in the package [12]. We abbreviate these tests by *Wav*, *Rpar* and *Anova*. Note that all these methods are constructed with multiple break points in mind, so we expect our method to perform comparatively well in the one change-point setting. Usually we set $p = 3$ and use the abbreviation *HCov* if we use $k = 1.5$ and *Cov* if we use $k = 1000$, which is effectively a covariance based test and not robust.

First we evaluate the behaviour under the null hypotheses. We look at AR(1) models $X_t = \rho X_{t-1} + \epsilon_t$ for $t = 1, \ldots, T$ with parameters $\rho \in \{0, 0.8\}$, different distributions for the innovations $(\epsilon_t)_{t=1,\ldots,T}$, namely the standard normal and a t-distribution with 3 degrees of freedom, and different length $T \in \{128, 256, 512\}$. Results are based on 10000 repetitions and summarized in Table 1. We can see that *HCov* holds its size very well under serial dependence and heavy tails, whereas *Wave*, *Rpar* and to some degree also *Cov* have problems in the later case. Surprisingly *Anova* needs at least $T = 256$ to work well.

To asses power under H_1 we look at $X_t = \begin{cases} \epsilon_t, & t = 1, \ldots, 128 \\ 0.3X_{t-1} + \epsilon_t, & t = 129, \ldots, 256 \end{cases}$. Here the autocorrelation function changes from $\rho(k) = 0$ to $\rho(k) = 0.3^k$ for $k \in \mathbb{N}$. We use t-distributions with different degrees of freedom to investigate the influence of heavy tails. Results based on 16000 repetitions can be seen in Figure 1 on the left. The

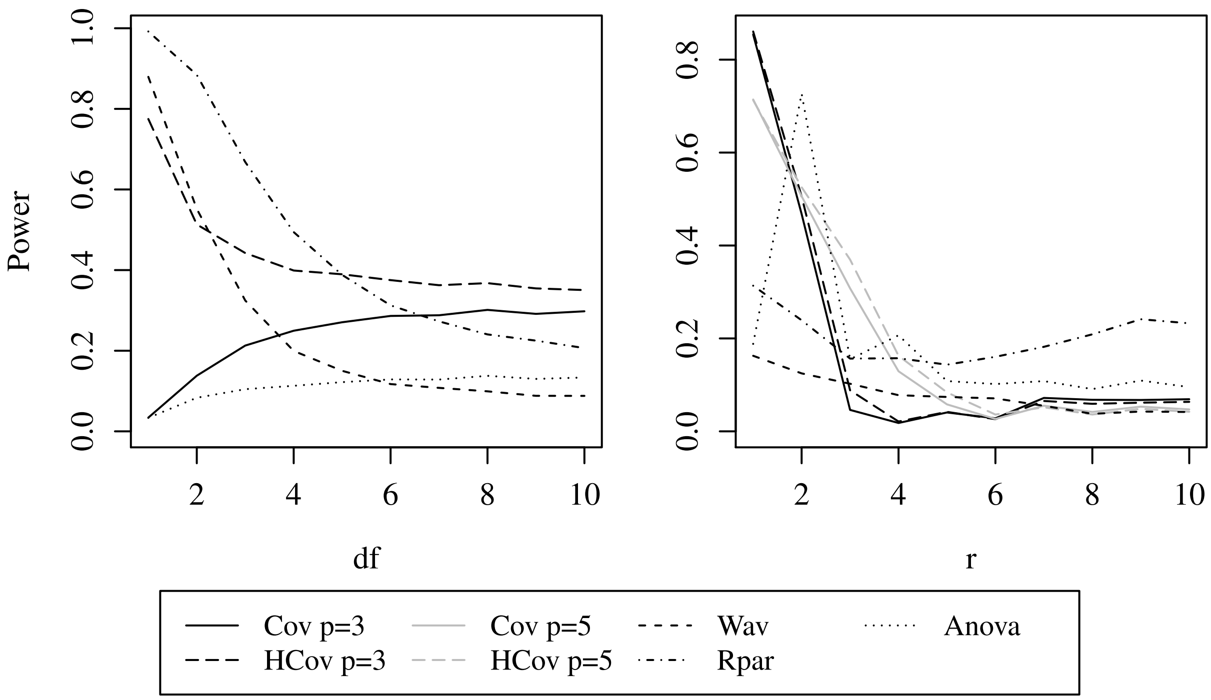

Figure 1: Empirical power under a change from independent observations to an AR(1) model with t distributed innovations with various degrees of freedom (left) and a change to an MA(r) model with normal innovations (right).

Table 1: Empirical size in percent under AR(1) models with different ρ, normal and t_3 distributed innovations and various time series lengths T at a nominal level of 0.05.

	N(0,1)						t_3					
ρ	0			0.8			0			0.8		
T	128	256	512	128	256	512	128	256	512	128	256	512
Cov	3	4	5	2	3	4	4	6	7	2	2	3
HCov	3	4	5	3	4	4	3	4	5	3	3	4
Wav	1	3	4	2	4	4	8	29	34	6	23	27
Rpar	5	6	7	5	6	6	46	62	76	44	61	76
Anova	67	6	6	52	14	10	53	3	3	54	11	7

robust test *HCov* dominates its competitors for $df = 10$ and even gains power as df decreases. This could be a result of good leverage points. We see that the wavelet based tests have a higher power at some point, though this is completely driven by their anticonservatism under the null-hypothesis.
Finally we want to evaluate the influence of the time-lag where the autocorrelation function changes and look at $X_t = \begin{cases} \epsilon_t & t = 1, \ldots, 128 \\ \epsilon_t + 0.8\epsilon_{t-r} & t = 129, \ldots, 256 \end{cases}$. Here the acf changes from $\rho(k) = 0$ to $\rho(k) = 0.8/(1 + 0.8^2)I_{\{k=r\}}$ for $k \in \mathbb{N}$. Results under normal innovations and 16000 repetitions can be seen in Figure 1 at the right. Our tests with the choice $p = 3$ can only detect changes of the acf at lag 1 and 2. We also run our tests with $p = 6$ and noticed that the power for smaller lags deteriorates a little while we can now detect changes up to lag 4. Apart from *Rpar* all tests loose power as the lag of change r increases.
In summary our robust change-point test behaves well under linear models with and without heavy tails. The choice of p is crucial and determines up to which lag changes in the acf can be detected. In case of doubt one should choose it rather larger than smaller to be able to detect changes in the acf at higher lags.

References

[1] Akashi F., Dette H., Liu Y. (2018). Change point detection in autoregressive models with no moment assumptions. *Journal of Time Series Analysis*. Vol. **39**, pp. 763-786.

[2] Andrews D.W.K. (1993). Tests for Parameter Instability and Structural Change with unknown Change Point. *Econometrica*. Vol. **61**, pp. 821-856.

[3] Bai J. (1993). On the partial sums of residuals in autoregressive and moving average models. *Journal of Time Series Analysis*. Vol. **14**, pp. 247-260.

[4] Bai J. (1994). Weak convergence of the sequential empirical processes of residuals in ARMA models. *The Annals of Statistics*. Vol. **22**, pp. 2051-2061.

[5] Berkes I., Gombay E., Horvath, L. (2009). Testing for changes in the covariance structure of linear processes. *Journal of Statistical Planning and Inference*. Vol. **139**, pp. 2044-2063.

[6] Bradley R.C. (2005). Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys*. Vol. **2**, pp. 107-144.

[7] Bucher A., Fermanian J.D., Kojadinovic I. (2019). Combining Cumulative Sum Change-Point Detection Tests for Assessing the Stationarity of Univariate Time Series. *Journal of Time Series Analysis*. Vol. 40, pp. 124-150.

[8] Cardinali A., Nason G. (2018). Practical powerful wavelet packet tests for second-order stationarity. *Applied and Computational Harmonic Analysis*. Vol. **44**, pp. 558-583.

[9] Chanda K.C. (1974). Strong mixing properties of linear stochastic processes. *Journal of Applied Probability*. Vol. **11**, pp. 401-408.

[10] Cho H. (2016). A test for second-order stationarity of time series based on unsystematic sub-samples. *Stat*. Vol. **5**, pp. 262-277.

[11] Cho H. (2018). unsystation - Stationarity Test Based on Unsystematic Sub-Sampling. *R-Package*, Version 0.2.0.

[12] Constantine W., Percival D. (2017). fractal - A Fractal Time Series Modeling and Analysis Package. *R-Package*, Version 2.0.4.

[13] Davis R.A., Huang D., Yao, Y. (1995). Testing for a Change in the Parameter Values and Order of an Autoregressive Model. *The Annals of Statistics*. Vol. **23**, pp. 282-304.

[14] Dette H., Wu W., Zhou, Z. (2015). Change point analysis of second order characteristics in non-stationary time series. *arXiv e-prints*. arXiv:1503.08610.

[15] Dette H., Preuß P., Vetter M. (2011). A measure of stationarity in locally stationary processes with applications to testing. *Journal of the American Statistical Association*. Vol. **106**, pp. 1113-1124.

[16] Dürre A., Fried R. (2019). Robust change point tests by bounded transformations. *arXiv e-prints*. arXiv:1905.06201.

[17] Giraitis L., Leipus R. (1992). Testing and estimating in the change-point problem of the spectral function. *Lithuanian Mathematical Journal*. Vol. **32**, pp. 15-29.

[18] Huber P.J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*. Vol. **35**, pp. .73-101.

[19] Huber P.J., (2011). Robust Statistics. In *International Encyclopedia of Statistical Science*. Springer, pp. 1248-1251.

[20] Huskova M., Kirch C., Praskova Z., Steinebach, J. (2007). On the detection of changes in autoregressive time series I. Asymptotics. *Journal of Statistical Planning and Inference*. Vol. **137**, pp. 1243-1259.

[21] Huskova M., Kirch C., Praskova Z., Steinebach, J. (2007). On the detection of changes in autoregressive time series II. Resampling Procedures. *Journal of Statistical Planning and Inference*. Vol. **138**, pp. 1697-1721.

[22] Jenkins G.M. (1961). The comparison of correlations in time series. *Journal of the Royal Statistical Society: Series B (Methodological)*. Vol. **20**, pp. 158-164.

[23] Jin L., Wang S. Wang H. (2015). A new non-parametric stationarity test of time series in the time domain. *Journal of the Royal Statistical Society: Series B (Methodological)*. Vol. **77**, pp. 893-922.

[24] Lee S., Ha J. Na O., Na, S. (2015). The Cusum Test for Parameter Change in Time Series Models. *Scandinavian Journal of Statistics*. Vol. **30**, pp. 781-796.

[25] Lindner A.M. (2009). Stationarity, mixing, distributional properties and moments of GARCH (p, q)–processes. In *Handbook of financial time series*. Springer, pp. 43-69.

[26] 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)*. Vol. **75**, pp. 879-904.

[27] Nason G. (2016). Locits - Test of Stationarity and Localized Autocovariance. *R-Package*, Version 1.7.3.

[28] Paparoditis E. (2009). Testing temporal constancy of the spectral structure of a time series. *Bernoulli*. Vol. **15**, pp. 1190-1221.

[29] Paparoditis E. (2010). Validating stationarity assumptions in time series analysis by rolling local periodograms. *Journal of the American Statistical Association*. Vol. **105**, pp. 839-851.

[30] Picard D. (1985). Testing and estimating change-points in time series. *Advances in applied probability*. Vol. **17**, pp. 841-867.

[31] Politis D.N, Romano J.P. (1993). On a family of smoothing kernels of infinite order. {Computing Science and Statistics, Proceedings of the 25th Symposium on the Interface}, pp.141-145.

[32] Preuš P., Vetter M., Dette H. (2013). A test for stationarity based on empirical processes. *Bernoulli*. Vol. **19**, pp. 2715-2749.

[33] Priestley M.B., Rao T.S. (1969). A test for non-stationarity of time-series. *Journal of the Royal Statistical Society: Series B (Methodological)*. Vol. **31**, pp. 140-149.

[34] Quenouille M.H. (1958). The comparison of correlations in time series. *Journal of the Royal Statistical Society: Series B (Methodological)*. Vol. **20**, pp. 158-164.

[35] Rosenblatt M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*. Vol. **42**, pp. 43-47.

[36] Rosenholc Y. (2001). Nonparametric Tests of Change-Points with Tapered Data. *Journal of Time Series Analysis*. Vol. **22**, pp. 13-43.

[37] Von Sachs R., Neumann M.H. (2000). A wavelet-based test for stationarity. *Journal of Time Series Analysis*. Vol. **21**, pp. 597-613.

[38] Schnabel R.B., Eskow E. (1999). A revised modified Cholesky factorization algorithm. *SIAM Journal on Optimization*. Vol. **9**, pp. 1135-1148.