Robust Test for Detecting Changes in the Autocovariance Function of a Time Series

Alexander Dürre
Université libre de Bruxelles

Roland Fried
Technische Universität Dortmund

Abstract

We propose a new robust test to detect changes in the autocovariance function 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 autocovariance at smaller lags. Our approach is compared to existing ones in some simulations.

Keywords: second-order stationarity, structural break, change-point detection.

1. Introduction

Testing for second order stationarity goes back to Quenouille (1958) and Jenkins (1961). To the best of our knowledge, the first test where the possible time of change is not known a priori can be found in Priestley and Rao (1969). Since then a lot of alternatives have been proposed. In Jin, Wang, and Wang (2015) estimated autocovariances of subsamples are compared to the estimation based on the whole time series. Several tests have been proposed for linear models, see Bai (1993), Bai (1994), Andrews (1993), Davis, Huang, and Yao (1995), and Akashi, Dette, and Liu (2018). CUSUM-type tests to detect changes in one or several autocovariances have been derived in Berkes, Gombay, and Horváth (2009), Lee, Ha, Na, and Na (2003), and Dette, Wu, and Zhou (2015). A test based on the auto-copula has been proposed in Bücher, Fermanian, and Kojadinovic (2019). Tests that check stationarity of the spectrum are presented in Picard (1985), Giraitis and Leipus (1992), and Rozenholc (2001) and a wavelet periodogram is used in Nason (2013) and Cardinali and Nason (2018). There are also proposals that compare local estimates of the spectrum with a global estimation; see Von Sachs and Neumann (2000), Paparoditis (2009), Paparoditis (2010) Dette, Preuß, and Vetter (2011), and Preuß, Vetter, and Dette (2013).

Surprisingly little attention has been paid to robustness. We want to fill this gap with a CUSUM type test based on robustified autocovariances. The testing procedure is described in Section 2. A simulation study in Section 3 indicates its usefulness.

2. Testing procedure

Denote by $\mathbb{X} = X_1, \ldots, X_T$ 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 $a_k = O(k^{-3-\epsilon})$ for some $\epsilon > 0$. Strong mixing was first introduced in Rosenblatt (1956) and describes how fast the dependence between two observations decreases as the time lag between them increases. See Bradley (2005) 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 Chanda (1974) and Lindner (2009).

We want to test, whether the autocovariance function of $\mathbb{X}$ stays the same, concentrating on the first p lags. We follow the approach of Dürre and Fried (2019) and use bounded transformations. Before using them, the observations need to be properly standardized. Denote 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

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

denotes the Huber-ψ function. This function was originally introduced for location estimation in Huber (1964) and basically downweights the influence of observations with large absolute values by shrinking them to more plausible values; see Figure 1. 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 Huber (1981) $k = 1.5$ is recommended as a compromise.
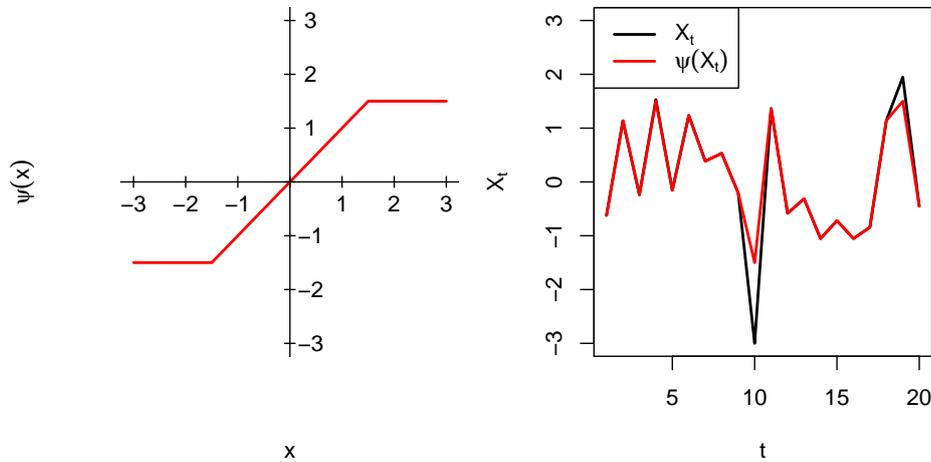

Figure 1: Huber-ψ function with $k = 1.5$ (left) and robustly transformed time series (right).

In the following, we derive a CUSUM type test based on the Huber-transformed time series. Denote $S_k^{(l)} = \sum_{t=1}^{k} \hat{Y}_t \hat{Y}_{t+l}$, then we look at

$$R_T = \max_{k=1,\ldots,\tilde{T}} \frac{1}{\tilde{T}} \begin{pmatrix} S_k^{(0)} - \frac{k}{\tilde{T}} S_{\tilde{T}}^{(0)} \\ \vdots \\ \vdots \\ S_k^{(p)} - \frac{k}{\tilde{T}} S_{\tilde{T}}^{(p)} \end{pmatrix}^T \begin{pmatrix} w_0 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & w_p \end{pmatrix} \begin{pmatrix} S_k^{(0)} - \frac{k}{\tilde{T}} S_{\tilde{T}}^{(0)} \\ \vdots \\ \vdots \\ S_k^{(p)} - \frac{k}{\tilde{T}} S_{\tilde{T}}^{(p)} \end{pmatrix}$$

where $w_0, \ldots, w_p > 0$ and $\tilde{T} = T - p$. Here are some remarks with regard to R_T:

- Technically, R_T tests the following hypothesis

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

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

  This is not equivalent to a test for a stationary autocovariance function up to lag p. For example, R_T will have problems to detect changes in the tail dependence since extreme values are down weighted by ψ.

- 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 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 covariance matrix of $S_{\tilde{T}}^{(0)}, \ldots, S_{\tilde{T}}^{(p)}$. R_T gets affine invariant in this case. In the time series context, this property is not desirable. The weights $w_0, \ldots, w_p$ give us more flexibility. Usually one would choose descending weights to smooth the transition between lags of the autocovariance function where one can detect a change $j = 0, \ldots, p$ to those neglected $j > p$. If there is only a change in the first autocovariance and one chooses p large, the change could be masked by the noise from the other autocovariances. Descending weights somehow counteract this problem. Without further knowledge, we suggest using $w_i = 1 - i/p$ for $i = 0, \ldots, p$.
  If one uses weights instead of Σ, the asymptotic distribution of 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 Dürre and Fried (2019). It is not explicitly stated there but effectively proved in Proposition 1 and 2 that

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

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

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

Proposition 3 in Dürre and Fried (2019) states that Σ can be consistently estimated by a kernel estimator. Denote therefore $b_{\tilde{T}} \geq 0$ a bandwidth and $k : \mathbb{R} \to [-1, 1]$ a kernel function. Then $\hat{\Sigma}$ with the elements

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

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

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

proposed in Politis and Romano (1993) works well together with $b_T = \tilde{T}^{\frac{1}{3}}$ under autoregressive processes of order 1. One can generate random variables $\tilde{R}_T^{(i)}$, $i = 1, \ldots, m$ that have asymptotically the same distribution as R_T under the null-hypothesis by the following algorithm:

- generate $(p+1) \cdot \tilde{T}$ independent standard normal random variables and store them in a $\tilde{T} \times (p+1)$ 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

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

where $W = diag(w_0, \ldots, w_p)$ and $V_{[t,]}$ is the t-th row of V.

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. We use the algorithm proposed in Schnabel and Eskow (1999) in our simulations.

3. Simulations

We assess our approach in a simulation study. We compare our method with tests for second order stationarity which are available in R (R Core Team 2019). Two wavelet based tests proposed in Nason (2013) and Cho (2016) are implemented in the packages `locits` (Nason 2016) respectively `unsystation` (Cho 2018). A revised version of the ANOVA test originally proposed in Priestley and Rao (1969) is implemented in the package `factal` (Constantine and Percival 2017). We abbreviate these tests by *Wav*, *Rpar* and *Anova*. These methods can cope with multiple break points, so we expect them to have inferior power in the one change-point setting. A copula based method (Bücher *et al.* 2019) is implemented in the package `npcp` (Kojadinovic 2019). This test can detect quite general changes in the dependence structure up to lag p but no changes in the marginal variance. We choose $p \in \{3, 5\}$ and abbreviate the tests as *Copula3* and *Copula5*. Furthermore, we compare with tests that can detect changes in autoregressive processes: a likelihood based (Davis *et al.* 1995) and a quantile based (Qu 2008). Both methods require the order p of the AR process. We choose $p \in \{1, 3, 5\}$ and denote the likelihood based tests by *AR1*, *AR3*, *AR5* and the quantile based tests by *Quan1*, *Quan3*, *Quan5*. A test for a change in the spectrum (Picard 1985) is abbreviated as *spec*.
Our proposed methods depend on the maximal considered lag $p \in \{3, 5\}$ and the robustness parameter $k \in \{1.5, 1000\}$. The former choice of k is robust and the test is abbreviated as *HCov*. The later choice is non-robust and similar to a covariance based test proposed in Berkes *et al.* (2009). The test is denoted by *Cov*. Furthermore, we compare different weights: linear decreasing weights $w_i = 1 - i/p$, $i = 0, \ldots, p$ are indicated by d, constants weights $w_i = 1$, $i = 0, \ldots, p$ by e and weights generated by the estimated asymptotic long run covariance matrix $\hat{\Sigma}$ by s. In this notation *Cov5d* denotes the non robust test with $k = 1000$, $p = 5$ and decreasing weights.

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
Cov3d	2.5	3.1	3.6	3.9	4.1	4.4	1.7	1.8	2.1	3.5	3.6	3.5
Cov5d	1.7	2.4	3.1	5.1	5.6	5.7	1.3	1.7	2.0	4.6	4.7	4.7
Cov3e	2.1	2.5	3.3	2.7	3.0	3.5	1.6	1.7	2.1	2.2	2.5	2.5
Cov5e	1.5	1.8	2.7	2.6	3.3	3.5	1.3	1.6	1.9	2.1	2.4	2.6
Cov3s	5.7	2.8	2.4	29.1	21.0	11.3	16.9	11.7	7.2	35.9	31.8	25.1
Cov5s	13.1	3.4	2.1	35.7	36.1	23.1	21.8	15.3	10.5	37.0	42.2	38.4
HCov3d	1.6	1.8	2.5	4.1	4.4	4.5	1.6	2.2	2.4	5.1	4.8	5.5
HCov5d	0.8	1.3	1.9	4.5	5.0	4.9	0.8	1.5	1.9	5.6	5.4	5.9
HCov3e	2.0	2.5	3.2	3.7	4.2	4.5	1.9	2.9	3.5	4.2	4.3	5.0
HCov5e	1.4	2.0	2.9	3.2	4.3	4.5	1.3	2.0	2.9	3.7	4.3	4.8
HCov3s	4.3	2.2	2.4	21.3	10.1	5.1	4.2	2.0	2.6	29.0	14.9	7.0
HCov5s	9.8	2.3	2.1	32.4	26.6	10.8	9.8	2.4	2.2	36.5	34.1	14.3
Wav	0.8	3.6	3.7	1.3	4.0	4.0	7.7	29.1	33.1	5.3	22.6	27.2
Rpar	5.3	6.5	6.4	4.7	5.4	6.3	46.0	62.1	76.4	44.4	61.0	76.9
Anova	68.4	6.1	6.6	51.9	13.8	11.3	52.5	3.4	3.6	53.4	11.6	6.8
Copula3	4.0	5.1	7.0	1.8	2.1	3.3	3.5	5.2	6.8	1.6	1.9	2.5
Copula5	0.6	1.4	3.4	1.0	1.5	2.6	0.4	1.3	3.3	0.9	1.4	2.1
AR1	3.5	3.7	4.0	9.6	8.3	7.7	10.9	11.4	11.5	21.7	20.7	21.1
AR3	8.2	7.9	7.9	14.8	12.9	11.1	20.1	19.5	20.0	29.1	28.1	27.0
AR5	38.8	32.1	30.3	50.0	41.3	35.6	45.9	42.3	39.7	55.9	49.7	47.0
Quan1	3.4	4.0	4.4	2.4	3.3	3.8	3.2	3.5	4.2	2.5	3.2	4.1
Quan3	2.7	3.4	4.2	2.3	3.0	3.5	2.3	3.1	3.6	2.2	2.7	3.8
Quan5	2.5	3.3	4.2	2.5	3.3	3.9	2.0	2.8	3.5	2.6	3.0	3.8
Spec	4.8	4.6	5.5	6.0	10.3	15.9	56.4	69.5	80.0	21.1	35.4	49.2

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 $\rho \in \{0, 0.8\}$, different distributions for $(\epsilon_t)_{t=1,\ldots,T}$, namely the standard normal and a t-distribution with 3 degrees of freedom, and different $T \in \{128, 256, 512\}$. Results based on 10000 repetitions are summarized in Table 1. Only *Cov*, *HCov* with decreasing and equal weights, *Copula*, and *Quan* give reliable results in all cases. *Wav* and *Rpar* are strongly oversized under t_3 innovations, *spec* is additionally anti-conservative for large ρ. *Anova* needs very large sample sizes. Furthermore one can observe that *Cov* and *HCov* are conservative for small T in case of equal or decreasing weights. The weights based on $\hat{\Sigma}$ lead to strongly anti-conservative tests if T is small and ρ is large.

To asses power under H_1, we first look at autoregressive models of order one:

$$X_t = \begin{cases} \epsilon_t, & t = 1, \ldots, 128 \\ \rho X_{t-1} + \epsilon_t, & t = 129, \ldots, 256 \end{cases} \quad \text{for} \quad \rho \in \{0, 0.05, \ldots, 0.55, 0.6\}$$

and $\epsilon_t \sim N(0,1)$ for $t \in \mathbb{Z}$. The autocovariance function changes from $\rho(k) = 0$ to $\rho(k) = \rho^k/(1-\rho^2)$ for $k \in \mathbb{N}$. Results based on 10000 repetitions can be seen in Figure 2. In the first line on the left, we see different versions of our proposed method. The tests with decreasing weights clearly outperform the others. Surprisingly, the robust tests have higher power than the non robust ones. Furthermore, we notice that the test considering 5 lags does not lose much against the one that only considers 3 lags. On the right, we see different versions of *AR*, *Copula*, and *Quan* and notice that *Copula3* clearly outperforms *Copula5*. Furthermore,

Quan1 has considerably larger power than *Quan3* and *Quan5* whereas *AR3* and especially *AR5* have large size distortions under H_0. We choose the versions that behave best and compare them in the second row of Figure 2. There we see that *Copula3* has the largest power for small ρ and *AR1* for larger ρ. The former is quite remarkable since copula based tests have power against arbitrary changes of the dependence structure. Our test *HCov3d* lags some power compared to *Copula3* and *Quan1* which are also robust to some degree.
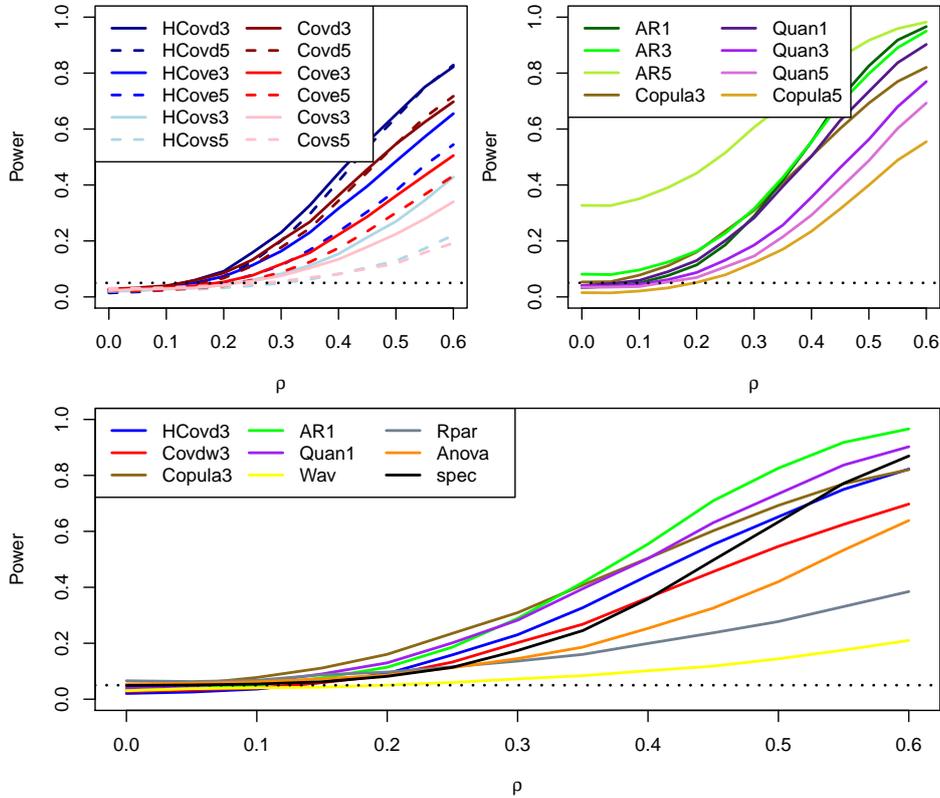
Figure 2: Simulated power under a change from independent random variables $\epsilon_t \sim N(0,1)$, $t = 1, \ldots, 128$ to an AR(1) model: $X_t = \rho X_{t-1} + \epsilon_t$ with $\epsilon_t \sim N(0,1)$ for $t = 129, \ldots, 256$. In the first line on the left are different versions of *HCov* and *Cov*, on the right are different versions of *AR*, *Copula*, and *Quan* and in the second line the best versions are compared.

Now we evaluate the behaviour under heavy tails. We look at

$$X_t = \begin{cases} \epsilon_t, & t = 1, \ldots, 128 \\ 0.35 X_{t-1} + \epsilon_t, & t = 129, \ldots, 256 \end{cases}$$

where the innovations ϵ_t are t-distributed with different degrees of freedom $df \in \{1, \ldots, 10\}$. Results are based on 10000 repetitions. In the first line of Figure 3 on the left, one can see that *HCov3d* has the largest power. Usually *HCov* gains power under more heavy tailed innovations which corresponds to small df, and *Cov* loses power. The exemption is *Covs*, which is anti-conservative under heavy tails. In Figure 3 on the right, we see that *Copula3* considerably outperforms *Copula5*, whereas *Quan1* has larger power than *Quan3* and *Quan5*. We decided against adding the results of *AR3* and *AR5* since they are strongly anti-conservative under heavy tails. We compare the best tests at the bottom of Figure 3 and notice that *Quan1* has the largest power apart from $df = 1$. In this case *HCov3d* is most powerful.

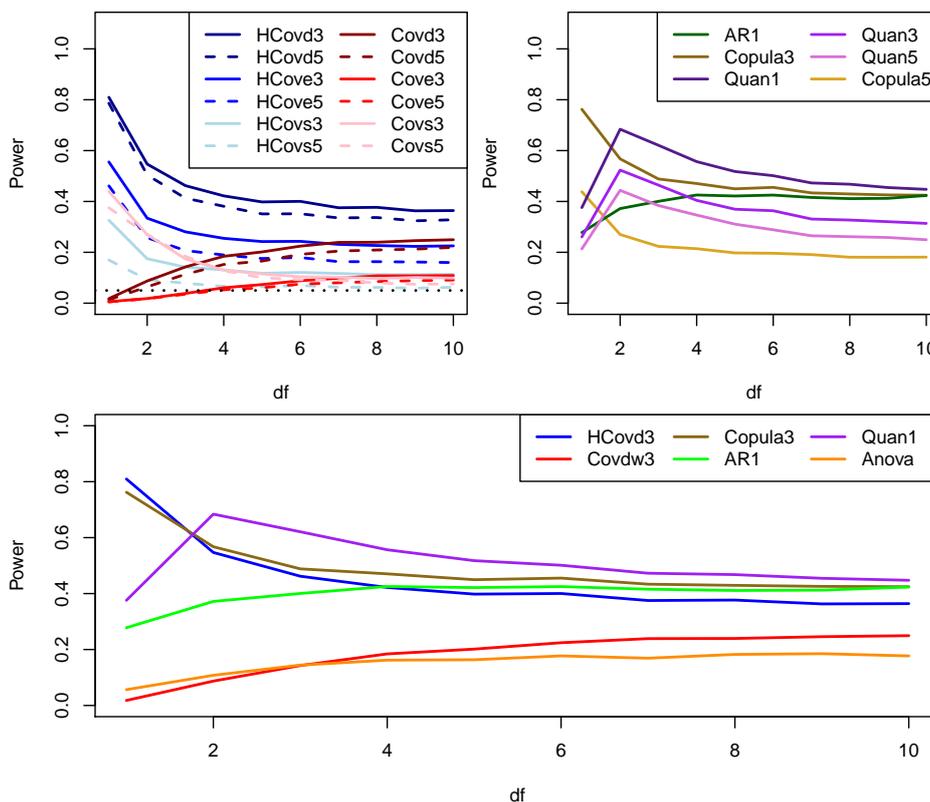Finally we want to evaluate the influence of the time-lag where the autocovariance function

Figure 3: Simulated power under a change from independent random variables $\epsilon_t \sim t_{df}$, $t = 1, \ldots, 128$ to an AR(1) model: $X_t = 0.35 X_{t-1} + \epsilon_t$ with $\epsilon_t \sim t_{df}$ for $t = 129, \ldots, 256$. In the first line on the left are different versions of *HCov* and *Cov*, on the right are different versions of *AR*, *Copula*, and *Quan* and in the second line the best versions are compared.

changes and look at

$$
X_t = \begin{cases} \epsilon_t & t = 1, \ldots, 128 \\ \epsilon_t + 0.8\epsilon_{t-r} & t = 129, \ldots, 256 \end{cases}
$$

where $\epsilon_t \sim N(0,1)$. The autocovariance function changes from $\gamma(k) = I_{\{k=0\}}$ to $\gamma(k) = 1.8 I_{\{k=0\}} + 0.8 I_{\{k=r, r \neq 0\}}$ for $k \in \mathbb{N}_0$ and $r = 0, \ldots, 10$. Note that $r = 0$ corresponds to a change of the marginal variance. Results based on 10000 repetitions can be seen in Figure 4. In the first line on the left, we see different versions of *HCov* and *Cov*. It is not clear which one behaves best overall. All tests have the largest power if there is only a shift in the variance ($r = 0$). Furthermore, they have power > 0.05 for all r since there is always a change in the marginal variance. The power of tests with decreasing weights decrease as expected, whereas the power of the tests with equal weights increase slightly from $k = 1$ to $k = p$. Usually *Cov* has larger power than *HCov* for $k = 0$ and $k = p+1, \ldots, 10$.

On the right, we compare different versions of *Copula*, *AR*, and *Quan*. *Copula3* has uniformly larger power than *Copula5*, *AR3* outperforms *AR1*, and *Quan5* dominates *Quan3* and *Quan1*. At the bottom of Figure 4, we compare the best tests and notice that *spec* behaves best overall. *Anova*, *Copula*, *Quan*, and *AR* have problems to detect changes in the marginal variance ($r = 0$) and *Quan*, *Anova*, *Copula*, and *Wav* have problems if r is large.

We conclude that no test dominates the others in all considered situations. Our proposal *HCov* behaves well in all of them. It is the only test which can cope with heavy tailed distributions and can detect changes in the marginal variance simultaneously. We notice that decreasing weights yield good results, especially under autoregressive processes. The choice of p influences the power of the test. It seems that choosing p too large is better than choosing it too small.
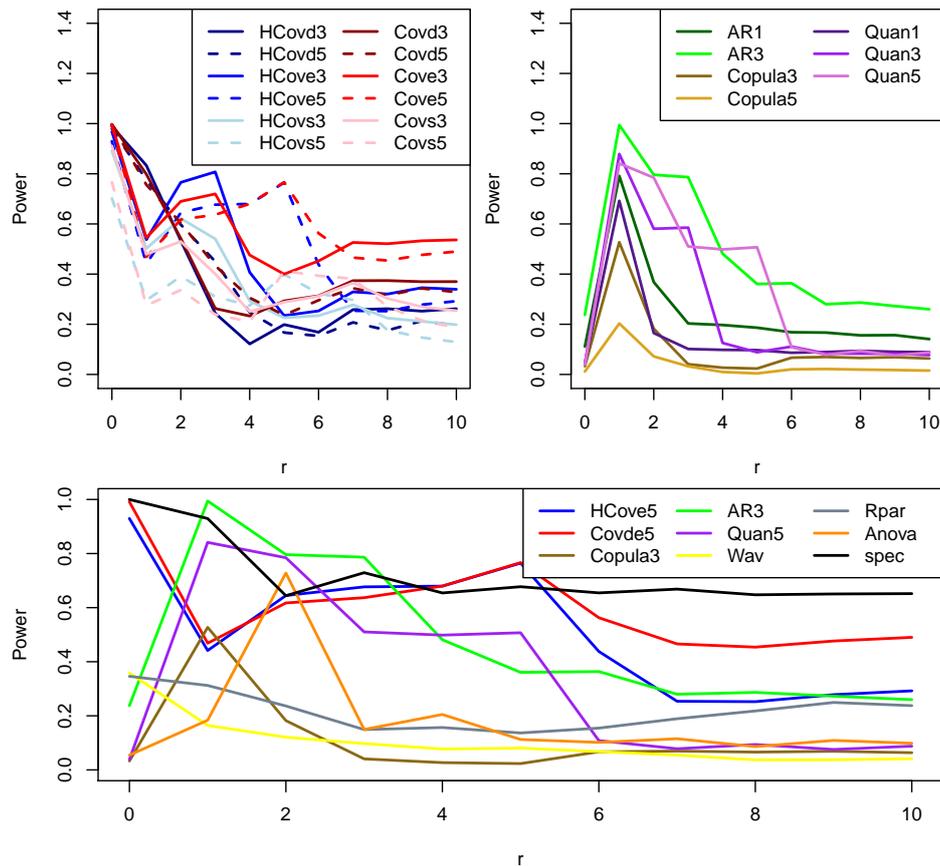
Figure 4: Simulated power under a change from independent random variables $\epsilon_t \sim N(0,1)$, $t = 1, \ldots, 128$ to a moving average model: $X_t = \epsilon_t + 0.8\epsilon_{t-r}$ with $\epsilon_t \sim N(0,1)$ for $t = 129, \ldots, 256$. In the first line on the left are different versions of *HCov* and *Cov*, on the right are different versions of *AR*, *Copula* and *Quan* and in the second line the best versions are compared.

In all simulations, a misspecification of p is not as fatal as a misspecification of similar tuning parameters of the competing methods *AR*, *Quan*, and *Copula*.

Acknowledgements

References

Akashi F, Dette H, Liu Y (2018). "Change-Point Detection in Autoregressive Models with no Moment Assumptions." *Journal of Time Series Analysis*, **39**(5), 763–786.

Andrews DWK (1993). "Tests for Parameter Instability and Structural Change with Unknown Change Point." *Econometrica*, **61**(4), 821–856.

Bai J (1993). "On the Partial Sums of Residuals in Autoregressive and Moving Average Models." *Journal of Time Series Analysis*, **14**(3), 247–260.

Bai J (1994). "Weak Convergence of the Sequential Empirical Processes of Residuals in ARMA Models." *The Annals of Statistics*, **22**(4), 2051–2061.

Berkes I, Gombay E, Horváth L (2009). "Testing for Changes in the Covariance Structure of Linear Processes." *Journal of Statistical Planning and Inference*, **139**(6), 2044–2063.

Bradley RC (2005). "Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions." *Probability surveys*, **2**, 107–144.

Bücher A, Fermanian JD, Kojadinovic I (2019). "Combining Cumulative Sum Change-Point Detection Tests for Assessing the Stationarity of Univariate Time Series." *Journal of Time Series Analysis*, **40**(1), 124–150.

Cardinali A, Nason GP (2018). "Practical Powerful Wavelet Packet Tests for Second-order Stationarity." *Applied and Computational Harmonic Analysis*, **44**(3), 558–583.

Chanda KC (1974). "Strong Mixing Properties of Linear Stochastic Processes." *Journal of Applied Probability*, **11**(2), 401–408.

Cho H (2016). "A Test for Second-order Stationarity of Time Series Based on Unsystematic Sub-samples." *Stat*, **5**(1), 262–277.

Cho H (2018). *unsystation: Stationarity Test Based on Unsystematic Sub-Sampling*. R package version 0.2.0, URL https://CRAN.R-project.org/package=unsystation.

Constantine W, Percival D (2017). *fractal: A Fractal Time Series Modeling and Analysis Package*. R package version 2.0-4, URL https://CRAN.R-project.org/package=fractal.

Davis RA, Huang D, Yao YC (1995). "Testing for a Change in the Parameter Values and Order of an Autoregressive Model." *The Annals of Statistics*, **23**(1), 282–304.

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*, **106**(495), 1113–1124.

Dette H, Wu W, Zhou Z (2015). "Change Point Analysis of Second Order Characteristics in Non-stationary Time Series." *arXiv preprint arXiv:1503.08610*.

Dürre A, Fried R (2019). "Robust Change Point Tests by Bounded Transformations." *arXiv preprint arXiv:1905.06201*.

Giraitis L, Leipus R (1992). "Testing and Estimating in the Change-point Problem of the Spectral Function." *Lithuanian Mathematical Journal*, **32**(1), 15–29.

Huber PJ (1964). "Robust Estimation of a Location Parameter." *The Annals of Mathematical Statistics*, **35**(1), 73–101.

Huber PJ (1981). *Robust Statistics*. Wiley series in probability and mathematical statistics. Wiley, New York. ISBN 0471418056.

Jenkins GM (1961). "General Considerations in the Analysis of Spectra." *Technometrics*, **3**(2), 133–166.

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 (Statistical Methodology)*, **77**(5), 893–922.

Kojadinovic I (2019). *npcp: Some Nonparametric CUSUM Tests for Change-Point Detection in Possibly Multivariate Observations*. R package version 0.1-11, URL https://CRAN.R-project.org/package=npcp.

Lee S, Ha J, Na O, Na S (2003). "The Cusum Test for Parameter Change in Time Series Models." *Scandinavian Journal of Statistics*, **30**(4), 781–796.

Lindner AM (2009). "Stationarity, Mixing, Distributional Properties and Moments of GARCH (p, q)–processes." In *Handbook of financial time series*, pp. 43–69. Springer.

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.

Nason G (2016). *locits: Tests of Stationarity and Localized Autocovariance*. R package version 1.7.3.

Paparoditis E (2009). "Testing Temporal Constancy of the Spectral Structure of a Time Series." *Bernoulli*, **15**(4), 1190–1221.

Paparoditis E (2010). "Validating Stationarity Assumptions in Time Series Analysis by Rolling Local Periodograms." *Journal of the American Statistical Association*, **105**(490), 839–851.

Picard D (1985). "Testing and Estimating Change-points in Time Series." *Advances in Applied Probability*, **17**(4), 841–867.

Politis DN, Romano JP (1993). "On a Family of Smoothing Kernels of Infinite Order." *Computing science and statistics*, pp. 141–141.

Preuß P, Vetter M, Dette H (2013). "A Test for Stationarity Based on Empirical Processes." *Bernoulli*, **19**(5B), 2715–2749.

Priestley MB, Rao TS (1969). "A Test for Non-stationarity of Time-series." *Journal of the Royal Statistical Society: Series B (Methodological)*, **31**(1), 140–149.

Qu Z (2008). "Testing for Structural Change in Regression Quantiles." *Journal of Econometrics*, **146**(1), 170–184.

Quenouille MH (1958). "The Comparison of Correlations in Time-series." *Journal of the Royal Statistical Society: Series B (Methodological)*, **20**(1), 158–164.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

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*, **42**(1), 43.

Rozenholc Y (2001). "Nonparametric Tests of Change-Points with Tapered Data." *Journal of Time Series Analysis*, **22**(1), 13–43.

Schnabel RB, Eskow E (1999). "A Revised Modified Cholesky Factorization Algorithm." *SIAM Journal on optimization*, **9**(4), 1135–1148.

Von Sachs R, Neumann MH (2000). "A Wavelet-based Test for Stationarity." *Journal of Time Series Analysis*, **21**(5), 597–613.

Affiliation:

Alexander Dürre
European Center for Advanced Research in Economics and Statistics
Université libre de Bruxelles
BE-1050 Brussels, Belgium
E-mail: alexander.durre@ulb.be
URL: https://www.statistik.tu-dortmund.de/duerre-en.html