MSc Thesis Applied Mathematics

Change Point Detection in the Autocorrelation Function of Stationary Time Series: A Comparative Analysis of Different Methods

Ronan Timon

Supervisor: Annika Betken, Giorgio Micali

March, 2025

Department of Applied Mathematics
Faculty of Electrical Engineering,
Mathematics and Computer Science

UNIVERSITY OF TWENTE.

Contents

Abstract

In this thesis the issue of detecting structural changes in the autocovariance function of a stationary time series is explored. In particular a modified CUSUM-test that detects changes is the main subject. All necessary background is introduced in as much detail as possible. Asymptotic properties that are known in the literature but not proven explicitly will be proven here. Simulations supplement this work to show the size and power of the test in comparison to two tests in the literature. Two Cusum-tests proposed by Berkes et al[5], and another based on ordinal-patterns by Betken et al[6]. ARMA-models will be used to demonstrate the power, runtime, strengths and weaknesses of each test.

Introduction

Time series data is any kind of data with timestamps attached to it. From revenue and stock prices, to EEG data and weather patterns, time series data is prolific and occurs in almost every industry. Modelling techniques are well-researched and the state of the art is advanced in this area. If a structural change were to occur at some point in time these techniques become less effective, and such changes in regime need to be accounted for. Structural changes can happen for any number of reasons, some can be seen, predicted, and readily prepared for, and some are invisible, subtle, and might only be noticed after the fact. A large franchise opening a branch in a new country can affect revenue for a particular competing retailer, this can be anticipated and prepared for. Or perhaps a change in tax policy or some unknown phenomenon reduces revenue, this could be unexpected and only noticed after the fact. This has been a pressing issue for as long as Time Series Statistics has been a branch of study, and in this thesis a pivotal question is addressed: "How can you detect a structural break in your data?"
Such changes could be found manually by data analysis, but for large or constantly updating datasets this is infeasible. A robust method to do this quickly and efficiently has been an area of study since as early as 1955 with Page[30] proposing a method to detect changes in mean. More pertinent to this thesis are the methods of detecting autocovariance changes. At the time of writing the state of the art can be found in Dürre & Fried [39].
CUSUM tests are the main subject of this work, say one wishes to test the following hypothesis, let $(X_i)_{i\in[1,T]}$ be a time series:

$$\begin{aligned}
\mathcal{H}_0 :&\quad \mathbb{E}[X_1] = \mathbb{E}[X_2] = \cdots = \mathbb{E}[X_T] \\
\mathcal{H}_1 :&\quad \exists k \in \mathbb{N} \text{ s.t. } \mathbb{E}[X_1] = \cdots = \mathbb{E}[X_{k-1}] \neq \mathbb{E}[X_k] = \mathbb{E}[X_{k+1}] = \cdots = \mathbb{E}[X_T]
\end{aligned}$$

Under the null hypothesis the sample means $\hat{\mu}_k = \frac{1}{k}\sum_{i=1}^{k} X_i$ and $\hat{\mu}_T = \frac{1}{T}\sum_{i=1}^{T} X_i$ should be close to each other. After rearranging and taking the difference the standard CUSUM-test for a change in mean at k is

$$\mathrm{CS}(k) = \left| \sum_{i=1}^{k} X_i - \frac{k}{T} \sum_{i=1}^{T} X_i \right| \tag{1.1}$$

so under the null hypothesis this difference should be small, to detect a change in mean one may take the maximum for $k \in T$ for this. This is a mean change-point-detection test.
This is to be extended to test for a change in autocovariance at k using the following

problem

$$\mathcal{H}_0) \qquad \begin{pmatrix} \mathrm{Cov}(Y_1, Y_1) \\ \vdots \\ \mathrm{Cov}(Y_1, Y_{1+p}) \end{pmatrix} = \begin{pmatrix} \mathrm{Cov}(Y_{T-p}, Y_{T-p}) \\ \vdots \\ \mathrm{Cov}(Y_{T-p}, Y_T) \end{pmatrix} \tag{1.2}$$

$$\mathcal{H}_1) \quad \exists k < T: \qquad \begin{pmatrix} \mathrm{Cov}(Y_{k-p}, Y_{k-p}) \\ \vdots \\ \mathrm{Cov}(Y_{k-p}, Y_k) \end{pmatrix} \neq \begin{pmatrix} \mathrm{Cov}(Y_k, Y_k) \\ \vdots \\ \mathrm{Cov}(Y_k, Y_{k+p}) \end{pmatrix} \tag{1.3}$$

This will require some refinement of the mean change-point-detection test. Namely a quadratic form of the CUSUM test to detect changes in autocovariance of multiple lags will be introduced for a stationary time series. Some known results will be shown here, and proven in more detail than exists in the literature, particularly showing the power of the test under the null-hypothesis. The results will be verified and tested against some other contemporary tests for some ARMA models and eventually some datasets taken from a variety of industries.

Theoretical Background

2.1 Related Works

Page[30] is the earliest known source to propose methods for finding structural breaks via changes in the mean of a time series model. Robustness was not a concern at this point. It wasn't until works like Quenouille[34] and Jenkins[23] that second-order changes were looked into. Quenouille via the correlation structure of two lengths of time series, and Jenkins through spectral analysis of the time series. The subject, of course, was highly motivated by its promising application to financial data.

Many studies developed and tested techniques in this context (Wichern[41], Picard[32], Tsay[40], Tang & MacNeil[38], Kim[26], Lee & Park[27].) Wichern use a moving block procedure to detect variance change in AR(1) processes, likelihood arguments are used to develop estimates of the changepoint. Davis et al[12] also look at AR processes, and propose a genetic algorithm to find optimal windows for piecewise modelling using AR models. Chen & Gupta[10] use a Schwarz information criteria for detecting parameter changes. For results on AR(p) models see also Huskoba et al[19], [18] and Gombay[16], where an efficient score vector is used.

Picard[32] compare non-parametric models to parametric using a likelihood ratio test, and derive the asymptotic distribution of their estimator. Tang & Macneil also showed convergence of their test for ARMA processes. Kim et al[26] look at ARMA processes too, they use a Monte Carlo method to show size and power of their test and derive asymptotic properties.

Csorgo & Horvat[11] show results only when the underlying distribution is known. Baufays[4] improve Wichern's results using Baysian/Machine learning, for a similar approach to estimate the posterior distribution of the shift point see Abraham & Wei[1]. Nyblom[29] instead test for changes in regression coefficients.

Banjeera & Urga[3] and Perron[31] both provide a good resource of structural stability of various time series. For further resources see Jandhyala et al[22], who examine various tests on a broad selection of data.

Galeano & Pena[15] propose a non-parametric test for structural breaks in the variance matrix of multivariate time series. Recent papers on second order structural breaks see Killick et al[25], who develop a likelihood-based hypothesis test for a locally stationary wavelet model (LSW model), and Preuss et al[33], where the estimated spectral distribution of different segments is compared.

Inclan & Tiao[20] use an iterated cumulative sum of squares (ICSS), which is very similar to a CUSUM-based test, to detect changepoints in the variance of a multi-dimensional time series. Gombay et al[16] do the same, only for a CUSUM-based test and test their results on financial data. Kim[26] took a broad approach; they showed their CUSUM-test had good size and power using a Monte Carlo method and also derived the asymptotic properties. Au & Horvath[2] develop a non-parametric CUSUM-type test using the vech$(\cdot)$ operation on the outer product of a multidimensional time series. vech$(\cdot)$ sums all of the values below the diagonals in the column of a matrix. They derived asymptotic properties under both the null and the alternative, and they examined financial data from four companies in different industries to test their results.

2.2 Notation

Let $(X_i)_{i\in[1,T]} = \{X_1, \dots, X_T\}$, $X_i \in \mathbb{R}\ \forall i \in [1, T]$ be a time series, which is stationary under the null-hypothesis. Denote the median by μ and the Median Absolute Deviation (MAD) by $\sigma = \text{median}|X_i - \mu|$, the sample median and MAD shall be denoted as $\hat{\mu} = \underset{i\in[1,T]}{\text{median}}\ X_i$ and $\hat{\sigma} = \underset{i\in[1,T]}{\text{median}}|X_i - \hat{\mu}|$. Additionally denote the autocovariance function of a stationary time series by $\gamma(l) = \mathbb{E}\left((X_i - \mu)(X_{i+l} - \mu)\right)$. for two random vectors X, Y belonging to a probability space $(\mathbb{R}^p, \mathcal{A}, \mathcal{P})$ the Covariance matrix is given by

$$\text{Cov}\left(X, Y\right)_{i,j} = \text{Cov}(X_i, Y_j) = \mathbb{E}[X_i Y_j] - \mathbb{E}[X_i]\mathbb{E}[Y_j] \tag{2.1}$$

A matrix A is said to be positive definite if it has positive eigenvalues only, and will be denoted by $A \succ 0$ The Huber-ϕ function is defined as

$$\phi(x) = \begin{cases} x & \text{if } |x| < k \\ k & \text{if } |x| \geq k \end{cases} \tag{2.2}$$

will be written as $\phi(x)$ often in this paper. $L^\infty(\mathbb{R}^p, \mathbb{R}^s)$ is the space of functions $f : \mathbb{R}^p \to \mathbb{R}^s$ such that $\exists C \in \mathbb{R}$ such that $\|f\|_\infty < C\ \forall x \in \mathbb{R}$. where $\|f\|_\infty = \sup_{x\in\mathbb{R}} |f(x)|$. Furthermore, let $\|X\|_v = \sqrt{v^\top X v}$ be the matrix norm.
A sequence a_n belonging to a topological space $(S, \boldsymbol{S})$ that converges in to a shall be denoted $a_n \overset{S}{\to} a$. Convergence in probability shall be denoted instead by $a_n \overset{S}{\Rightarrow} a$.

2.3 Background

Here the class of time series that are the subject of this paper are introduced, that is to say strong-mixing, stationary time series. After this Convergence results needed, and the space that CUSUM-tests are in, the Skorohod Space, will be motivated and defined. Finally additional results on mixing, and on weak convergence that are used throughout the paper will be stated in this section.
First standardize the series $(X_i)_{i\in[1,T]}$ in the following manner

$$Y_{i,T} = \phi\left(\frac{X_i - \hat{\mu}}{\hat{\sigma}}\right) \quad Y_i = \phi\left(\frac{X_i - \mu}{\sigma}\right) \tag{2.3}$$

where Y_i is the result of the true standardization, i.e. as $T \to \infty$. The Huber function ϕ down-weights the effects of outliers, it will be shown later that CUSUM-statistics are quite sensitive to extreme values. Median and MAD have been shown to be consistent under strong mixing conditions[44]. Note that under the null hypothesis the standardized series $(Y_{i,T})_{i\in[1,T]}$ has median close to 0 and MAD close to 1.

Definition 1 (strong mixing)**.** *(Rosenblatt 1956[35])*
Let $(X_t)_{t\in[1.T]}$ be a time series. For $-\infty \leq m \leq p \leq \infty$, let $\mathcal{F}_m^p$ denote the σ-field of events generated by the random variables $\{X_j, m \leq j \leq p\}$. For any two fields $\mathcal{A}$ and $\mathcal{B} \subset \mathcal{F}$, consider the following measure of dependence [35]:

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup |P(A \cap B) - P(A)P(B)|, \quad A \in \mathcal{A}, B \in \mathcal{B}, \tag{2.4}$$

and the mixing coefficients:

$$\alpha_0 = \frac{1}{4} \quad \textit{and} \quad \alpha_n = \alpha(\mathcal{F}_{-\infty}^0, \mathcal{F}_n^\infty) \qquad \textit{for all} \quad n \in \mathbb{N}^*. \tag{2.5}$$

If $\alpha_n \to 0$ as $n \to \infty$ the sequence $\mathbb{X}$ is strongly mixing.

Strong mixing implies that future values of a time series are "almost independent" of the initial values. A general overview of mixing can be found in Bradley's paper[8], whose main results on strong α-mixing are from Chande[9]. Strong mixing conditions on ARMA schemes can be found in Withers[42].
This paper shall follow the same setting proposed by Dürre & Fried, where the series $(X_i)_{i\in[1,T]}$ is assumed to be strongly mixing with coefficients $(\alpha_k)_{k\in\mathbb{N}}$ satisfying $\alpha_k = O(k^{-3-\epsilon})$ for some $\epsilon > 0$; that is to say the mixing coefficients diminish in a manner at least cubic. Strong mixing is a useful property for the study of asymptotic properties of a time series. One might ask what this means for the mixing of the process $(Y_{i,T})_{i\in[1,T]}$. This is addressed in the following result

Lemma 1. *For a strongly mixing sequence $(X_i)_{i\in\mathbb{Z}}$ with mixing coefficients α_n and Borel-measurable function $f : (X, \mathcal{F}_X) \to (Y, \mathcal{F}_y)$, $(f(X_i))_{i\in\mathbb{Z}}$ is also strongly mixing with coefficients $\alpha(f(X_i)) \leq \alpha(X_i)$.*

This is proven by the author, but it should be noted that this is actually a narrowing of a result from Bradley(*theorem 5.2, pg. 20*[8]). $(Y_i)_{i\in[1,T]} = \phi(\frac{X_i - \mu}{\sigma})_{i\in[1,T]}$ meets the requirements for this lemma.
It can be broken up into a composition of the affine function $f(x) = \dfrac{x-\mu}{\sigma}$, followed by the function $\phi(x)$.
$f(x)$ is Borel-measurable in the standard topology as all affine functions are.
The preimage of any open set $V \subset \mathbb{R}$ in ϕ has several options:

1. $k \in V$ or $-k \in V$: The pre-image is $\{x \in (-\infty, -k]\} \cup (V \cap [-k, k])$ or $\{x \in [k, \infty)\} \cup (V \cap [-k, k])$ respectively, Both are the union of Borel-sets, and therefore Borel-sets

2. $V \cap [-k, k] \subseteq [-k, k]$: The preimage is an open subset of $[-k, k]$ and therefore in $\mathcal{B}$

So $\phi(\cdot)$ is Borel-measurable too, so their composition makes for a α-mixing process with mixing constants satisfying $\alpha'_k \leq \alpha_k$.
A second property of the time series $(Y_{i,T})_{i\in[1,T]}$ is its long run variance

$$u = \lim_{T\to\infty} \operatorname{Var}\left(\frac{1}{\sqrt{T}} \sum_{i=1}^{T} Y_{i,T}\right) \tag{2.6}$$

For the needs of this text it is better to look at the long-run covariance matrix, for a p-dimensional time series $(\boldsymbol{Y}_{i,T})_{i\in[1,T]}$ let

$$U = \lim_{T\to\infty} \mathrm{Var}(\frac{1}{\sqrt{T}} \sum_{i=1}^{T} \boldsymbol{Y}_{i,T}) \tag{2.7}$$

be the long-run covariance matrix. It is not guaranteed that this limit is bounded. In principle, a strongly mixing sequence becomes eventually independent. Meaning that the covariance terms are always decreasing.
This can be shown with the use of a covariance inequality which is derived in Davydov[13]. The inequality requires that two moments p and q be bounded for a series such that $\frac{1}{p} + \frac{1}{q} < 1$, and that the mixing coefficients satisfy $\alpha_k \leq Ca^k$ for some $a \leq 1$ and $C \in \mathbb{R}$.

$$|\mathrm{Cov}(Y_1, Y_{1+j})| \leq C\alpha(|j|)^{1-\frac{1}{p}-\frac{1}{q}} \left([\mathbb{E}[Y_1]^p]^{\frac{1}{p}} [\mathbb{E}[Y_{1+j}]^q]^{\frac{1}{q}}\right)$$

Thanks to the mixing conditions $C = \frac{1}{4}$ is a sufficient upper bound, also notice $\alpha(|j|)^{1-\frac{1}{q}-\frac{1}{p}} \to 0$ regardless of what moments p, and q are chosen. The norms are bounded too and well defined thanks to the function ϕ. This gives a nice bound, leading to bounded Long run covariance

$$\sum_{j=1}^{T} |\mathrm{Cov}(Y_1, Y_{1+j})| \leq \sum_{j=1}^{T} \left| C\alpha(|j|)^{1-\frac{1}{p}-\frac{1}{q}} \left([\mathbb{E}[Y_1]^p]^{\frac{1}{p}} [\mathbb{E}[Y_{1+j}]^q]^{\frac{1}{q}}\right)\right|$$

All terms inside the summation are bounded, so the entire summation is bounded too, therefore

$$\sum_{j=1}^{T} |\mathrm{Cov}(Y_1, Y_{1+j})| < \infty$$

as $T \to \infty$.
Often long run variance is instead written as follows:

$$\lim_{T\to\infty} \mathrm{Var}(\frac{1}{\sqrt{T}} \sum_{i=1}^{T} \boldsymbol{Y}_i) \tag{2.8}$$

$$= \lim_{T\to\infty} \mathrm{Cov}(\frac{1}{\sqrt{T}} \sum_{i=1}^{T} \boldsymbol{Y}_i, \frac{1}{\sqrt{T}} \sum_{i=1}^{T} \boldsymbol{Y}_i) \tag{2.9}$$

$$= \lim_{T\to\infty} \frac{1}{T} \sum_{i=1}^{T} \sum_{j=1}^{T} \mathrm{Cov}(\boldsymbol{Y}_i^{(j)}, \boldsymbol{Y}_i^{(j)}) \tag{2.10}$$

$$= \lim_{T\to\infty} \sum_{k=-(T-1)}^{T-1} (\frac{T-k}{T}) \mathrm{Cov}(\boldsymbol{Y}_i, \boldsymbol{Y}_{i+k}) \tag{2.11}$$

$$= \lim_{T\to\infty} \mathrm{Cov}(\boldsymbol{Y}_1, \boldsymbol{Y}_1) + 2\sum_{k=1}^{T-1} (\frac{T-k}{N}) \mathrm{Cov}(\boldsymbol{Y}_i, \boldsymbol{Y}_{i+k}) \tag{2.12}$$

$$= \mathrm{Cov}(\boldsymbol{Y}_1, \boldsymbol{Y}_1) + 2\sum_{k=1}^{\infty} \mathrm{Cov}(\boldsymbol{Y}_i, \boldsymbol{Y}_{i+k}) \tag{2.13}$$

Long run covariance is incredibly useful here, as it can be shown that the covariance matrix of a partial sum $\sum_{i=1}^{k} X_i$ of random variables is close to the long run covariance matrix,

lets call this U. This will come up later.
Convergence & The Skorohod Space:
As asymptotic properties of CUSUM-tests are the main subject of this paper the space in which this convergence occurs is a critical detail. A CUSUM-test is a function of random variables, so it would be natural to induce a topology with $|\cdot|_\infty$. This leads to problems in the set-up in the thesis. To see this recall the test(1.1)

$$\mathrm{CS}_T = \max_{1\le k\le T-1}\left|\sum_{i=1}^{k} X_i - \frac{k}{T}\sum_{i=1}^{T} X_i\right| \tag{1.1}$$

For starters, functions composed partial sums of random variables are not continuous in $\mathbb{R}$. The space of these functions must be defined

Definition 2 (Skorohod Space)**.** *Let (M, d) be a metric space and let $E \subseteq \mathbb{R}$. A function $f : E \to M$ is called a* ***càdlàg function*** *if, for every $t \in E$,*

- *the left limit $f(t-) := \lim_{s\to t^-} f(s)$ exists; and*

- *the right limit $f(t+) := \lim_{s\to t^+} f(s)$ exists and equals $f(t)$.*

That is, f is right-continuous with left limits. Let the ***Skorohod Space*** *$D[0,1]$ be the collection of càdlàg functions on the interval $[0,1]$*

The term càdlàg is derived from the French "continue à droit, limite à gauche", and is the home of CUSUM-tests. To see this denote partial sums as $S_{\lfloor Tx\rfloor} = \sum_{i=1}^{\lfloor Tx\rfloor} X_i$ parametrized by $x \in [0,1]$. Notice that $f(x)$ is a collection of disjoint constants:

when $x \in [0, \frac{1}{T})$: $S_{\lfloor Tx\rfloor}(\omega) = X_1(\omega)$

when $x \in [\frac{1}{T}, \frac{2}{T})$: $S_{\lfloor Tx\rfloor}(\omega) = X_1(\omega) + X_2(\omega)$

$\vdots$

If the max norm $|\cdot|_\infty$ were equipped the space would not be separable. That is to say, there does not exist a sequence $(x_i)_{i\in\mathbb{N}}$ of functions such that every non-empty open subset of the space contains at least one element of the sequence. This can be constructed by counter example, consider the following uncountable family of functions

$$\{\zeta_t(x) = \mathbb{1}_{[0,t)}(x), t \in [0,1]\} \tag{2.14}$$

notice that for any two elements of this family, $d_{|\cdot|_\infty}(\zeta_t, \zeta_s) = \sup_{x\in[0,1]} |\zeta_t(x) - \zeta_s(x)| = \sup_{x\in[0,1]} \mathbb{1}_{t,s} = 1$. i.e. regardless of the choice of t, s $(t \neq s)$ these functions are within 1 from each other in the $|\cdot|_\infty$ topology. So these elements of $D[0,1]$ are not dense countable. This resolve this consider the following topology.

Definition 3 (The Skorohod Topology)**.** *(Billingsley, [7], page 123) Let Λ denote the set of strictly increasing, continuous mappings of $[0,1]$ onto itself. If $\lambda \in \Lambda$, then $\lambda(0) = 0$ and $\lambda(1) = 1$. For x and y in $D[0,1]$ define $d(x,y)$ to be the infimum of those positive ϵ for which there exists $\lambda \in \Lambda$ satisfying:*

$$\sup_t |\lambda(t) - t| = \sup_t |t - \lambda^{-1}(t)| < \epsilon \tag{2.15}$$

and

$$\sup_t |x(t) - y(\lambda(t))| = \sup_t |x(\lambda^{-1}(t)) - y(t)| < \epsilon \tag{2.16}$$

This can be expressed more compactly by

$$d(x, y) = \underset{\lambda \in \Lambda}{inf}\{\|\lambda - I\| + \|x - y(\lambda))\|\} \tag{2.17}$$

This metric induces the Skorohod Topology D[0,1]

The functions λ essentially "tie together" all of the jumps and discontinuities. And the smallest λ is chosen such that the error between the gap between the functions x and y and their parameterized $\lambda(x)$ and $\lambda(y)$ is no more than ϵ of a jump.

Convergence in this topology shall be denoted $\overset{D[0,1]}{\to}$. Now that the topological space has been clarified the necessary definitions for treating convergence of random variables can be introduced.

Definition 4 (weak convergence)**.** *(Pollet,[21], Definition 18.1, pg. 151) Let P_n and P be probability measures on $\mathbb{R}^d$ $(d \geq 1)$. The sequence P_n converges weakly to P if $\int f(x)P_n(dx)$ converges to $\int f(x)P(dx)$ for each f which is real-valued, continuous and bounded on $\mathbb{R}^d$*

let $\overset{D[0,1]}{\Rightarrow}$ denote weak convergence in the Skorohod space. Notice that the definitions does not vary over the function f but instead over a sequence of probability measures P_n. Since the induced measure $P_X(A) = P(X \in A)$ entirely characterizes the distribution of X, the following extension can be made

Definition 5 (Convergence in Distribution)**.** *(Pollet,[21], Definition 18.2, pg. 151) Let $(X_i)_{i\in\mathbb{N}}$, X in $\mathbb{R}^p$-valued random variables. X_i converges in distribution to X if the distribution measures P_{X_i} converges weakly to P_X. We write $X_i \overset{\mathcal{D}}{\to} X$.*

These results are essential for studying the asymptotic properties of CUSUM-tests, and will be used in key results in this paper. A process that is often discussed alongside CUSUM-tests is Brownian Motion. This is a mathematical object based on the motion of a particle suspended in fluid, and is generalized as a stochastic process $(B_t)_{t\in\mathbb{R}}$ under four conditions:[24]

1. $B_0 = 0$
2. the sample trajectories $t \mapsto B_t$ are continuous, with probability one
3. for a finite sequence of time $t_0 < t_1 < \cdots < t_n$, the increments

   $$B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}}$$

   are mutually independent random variables
4. For any given times $0 \leq s \leq t$, $B_t - B_s$ has Gaussian distribution $N(0, t - s)$ with 0 mean and variance $t - s$

In short a Brownian motion is fixed at $t = 0$, a.s. continuous in t, increments are independent Gaussian distributed.
Additional Results:
The earlier definitions and lemmas are the backbone of the results that will follow in this paper, but to make use of these tools and properties a slew of additional theorems and lemmas will be required.

Theorem 1. *(Continuous Mapping Theorem) (Billingsley[7], page 20)*
Suppose h maps a metric space S, with Borel σ-field $\boldsymbol{S}$ onto another metric space S' with metric ρ and Borel σ-field $\boldsymbol{S}'$. If h is measurable $\boldsymbol{S}/\boldsymbol{S}'$, then each probability P on $(S', \boldsymbol{S}')$ induces on $(S, \boldsymbol{S})$ a probability Ph^{-1} defined by $Ph^{-1}(A) = P(h^{-1}A)$. Let P_n be a sequence of probability measures, and let $D_h \subset S'$ be the collection of points where h^{-1} is not continuous.
If $P_n \to P$ and $PD_h = 0$, then $P_n h^{-1} \to Ph^{-1}$.

The continuous mapping theorem allows for a function with known convergence behaviour to be studied in place of a trickier function, provided there is a Borel measurable map h^{-1} between the two functions.

Theorem 2. *(Slutsky's Theorem[37]) Let X_n,Y_n be sequences of random elements. if X_n converges in distribution to a random element X and Y_n converges in probability to a constant $c \neq 0$ then*

- $X_n + Y_n \stackrel{\mathcal{D}}{\to} X + c$
- $X_n Y_n \stackrel{\mathcal{D}}{\to} Xc$
- $X_n / Y_n \stackrel{\mathcal{D}}{\to} X/c$

where $\stackrel{\mathcal{D}}{\to}$, again, is convergence in distribution. Another tool in this class of results is the Cramer-Wold theorem.

Theorem 3 (Cramer-Wold Theorem)**.** *(Wooldridge White [43], Proposition 4.1, pg. 221) Let $\{W_n\}_{n\in\mathbb{N}}$ be a sequence of random elements of $D^l[0,1]$, and let W be a random element of $D^l[0,1]$, an l-dimensional vector composed of elements of the Skorohod space $D[0,1]$ (not necessarily Brownian motion.) Then $W_n \stackrel{D^l[0,1]}{\to} W$ if an only if*

$$\sum_{i=1}^{l} \lambda_i W_{n,i} \stackrel{D[0,1]}{\Rightarrow} \sum_{i=1}^{l} \lambda_i W_i$$

for each linear combination $\boldsymbol{\lambda}$ with $\boldsymbol{\lambda}^T \boldsymbol{\lambda} = 1$

The Cramer-Wold theorem considers the linear combination of elements of a vector, and if all such combinations converge then so too does the vector. This is undoubtedly much easier than showing element-wise convergence of a vector. Two additional results will be used later in this paper that are more specific. Both of these will be listed here

Theorem 4 (Weak Law of Large Numbers[17])**.** *Let $(X_i)_{i\in[1,T]}$ be a strongly mixing sequence, with sample mean $\bar{X}_T = \frac{1}{T}\sum_{t=1}^{T} X_t$. if*

$$\lim_{B\to\infty} \sup_{T} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[X_t \mathbb{1}_{|X_t|>B}] = 0$$

then

$$\mathbb{E}\Big|\bar{X}_T - \mathbb{E}[\bar{X}_T]\Big| \xrightarrow{S} 0$$

and

$$\bar{X}_T - \mathbb{E}[\bar{X}_T] \overset{S}{\Rightarrow} 0$$

as $T \to \infty$, for a topological space S.

This is needed as the weak law of a large numbers is strictly defined for independent, identically distributed random variables. Note that the original definition requires only weak mixing, where $\lim_{T\to\infty} \frac{1}{T}\sum_{k=1}^{T} X_t = 0$. This has been replaced since strong mixing implies weak mixing. If $\lim_{k\to\infty} \alpha_k = 0$ then the sample averages will also tend to 0, meaning that $\lim_{T\to\infty} \frac{1}{T}\sum_{k=1}^{T} \alpha_k = 0$. The second result is the following.

Theorem 5. *(Merlevede & Peligrad[28] pg. 8) Suppose that $(X_i)_{i\in[\mathbb{Z}]}$ is a strictly stationary, centered, strong mixing sequence with finite second moment. Additionally for $S_n = \sum_{j=1}^{n} X_j$, quartile function*

$$Q_w(u) = inf\{t \geq 0 : P(W > t) \leq u\} \tag{2.18}$$

and mixing constants α_n assume:

- $\underset{n\to\infty}{lim\ inf} \frac{\mathbb{E}[S_n^2]}{n} > 0$
- $\int_0^{\alpha_n} Q^2_{|X_0|}(u)du = o(\frac{1}{n})$ *as* $n \to \infty$

Then $W_n(t) = \dfrac{\sum_{i=1}^{\lfloor nt \rfloor} X_i}{\sqrt{(\pi/2)}\mathbb{E}|S_n|}$ *converges in distribution to a standard Brownian Motion in* $D[0, 1]$.

The conditions for this result are either assumed, or may be shown with a some effort. With this work done the main body of the paper may begin.

The modified CUSUM-test: R_T

In this section it will be shown how a mean change-point detection CUSUM-test like (1.1) can be adapted to detect changes in autocovariance at lag l. The asymptotic properties of this test will be shown, and only then will it be extended to the quadratic CUSUM-test R_T. In order to prove the asymptotic properties of R_T a detour will be taken to the related test $W_T^2(x)$, this is an earlier test made by Dürre & Fried[39] and the progenitor to R_T. The limiting distribution of $W_T^2(x)$ will first be shown so that R_T's distribution may be stated as a corollary.

3.1 Autocovariance Changepoint-Detection

Recall that by standardization the series $(Y_{i,T})_{i\in[1,T]}$ has median close to 0 and MAD close to 1. The closeness here is largely determined by the size of the sample T, since consistency has been shown[44] this means

$$\lim_{T\to\infty}\left(\underset{i\in[1,T]}{\text{median}}\; Y_{i,T}\right) = 0 \quad \& \quad \lim_{T\to\infty}\left(\underset{i\in[1,T]}{\text{median}}|Y_{i,T} - \hat{\sigma}|\right) = 1 \tag{3.1}$$

Define a new series by $Z_i^{(l)} = Y_i Y_{i+l}$ for $i \in [1, T-l]$, and some lag $l \in [0, \dots, p]$, $p \in \mathbb{N}$. Since $(Y_i)_{i\in[1,T]}$ is stationary under the null-hypothesis note that $\mathbb{E}[Z_i^{(l)}] = \mathbb{E}[Y_i Y_{i+l}] = \gamma(l)$, the autocovariance function at l. In fact, under $\mathcal{H}_0$

$$\mathbb{E}[Z_1^{(l)}] = \mathbb{E}[Z_2^{(l)}] = \cdots = \mathbb{E}[Z_{T-l}^{(l)}] = \gamma(l)$$

This can be leveraged to detect autocovariance changes: for a centred, stationary process the test

$$\begin{aligned} \mathcal{H}_0^{(l)})\quad & \text{cov}(Y_1, Y_{1+l}) = \text{cov}(Y_2, Y_{2+l}) = \cdots = \text{cov}(Y_{T-l}, Y_T) \\ \mathcal{H}_1^{(l)})\quad & \exists k \in [1, T-l], \text{cov}(Y_1, Y_{1+l}) = \cdots = \text{cov}(Y_k, Y_{k+l}) \neq \text{cov}(Y_{k+1}, Y_{k+1+l}) = \cdots = \text{cov}(Y_{T-l}, Y_T) \end{aligned}$$

can be rewritten as a mean changepoint-test

$$\mathcal{H}_0^{(l)})\quad \mathbb{E}[Z_1^l] = \mathbb{E}[Z_2^l] = \cdots = \mathbb{E}[Z_{T-l}^l] \tag{3.2}$$

$$\mathcal{H}_1^{(l)})\quad \exists k \in [1, T-l], \mathbb{E}[Z_1^l] = \cdots = \mathbb{E}[Z_k^l] \neq \mathbb{E}[Z_{k+1}^l] = \cdots = \mathbb{E}[Z_{T-l}^l] \tag{3.3}$$

A standard choice of test-statistic for the test (3.2) would be the CUSUM-test(1.1), for $\tilde{T} = T - l$ note that a sample autocovariance taken over size k and size $\tilde{T}$ should be roughly close to each other under the null hypothesis (3.2), i.e.

$$\begin{aligned}
\frac{1}{k}\sum_{i=1}^{k} Z_i^{(l)} &\simeq \frac{1}{\tilde{T}}\sum_{i=1}^{\tilde{T}} Z_i^{(l)} \\
\frac{1}{k}\sum_{i=1}^{k} Z_i^{(l)} - \frac{1}{\tilde{T}}\sum_{i=1}^{\tilde{T}} Z_i^{(l)} &\simeq 0 \\
\sum_{i=1}^{k} Z_i^{(l)} - \frac{k}{\tilde{T}}\sum_{i=1}^{\tilde{T}} Z_i^{(l)} &\simeq 0 \qquad (\text{Under } \mathcal{H}_0^{(l)})
\end{aligned}$$

Once more, large values of this are unlikely under the null-hypothesis. So by letting $S_k^{(l)} = \sum_{i=1}^{k} Z_i^{(l)}$ and taking the maximum over $k \in [1, T-l]$ the test becomes

$$\mathrm{CS}_{\tilde{T}}^{(l)} := \max_{1\le k\le \tilde{T}-1} \frac{1}{\sqrt{T u}} \left| S_k^{(l)} - \frac{k}{\tilde{T}} S_{\tilde{T}}^{(l)} \right| \tag{3.4}$$

where u is the long-run covariance. In order to study the asymptotic distribution of equation-3.4 it must instead be written as

$$CS_{\tilde{T}}^{(l)} = \sup_{x\in[0,1]} \frac{1}{\sqrt{T u}} \left| S_{\lfloor \tilde{T}x\rfloor}^{(l)} - x S_{\lfloor \tilde{T}x\rfloor}^{(l)} \right| \tag{3.5}$$

The asymptotic distribution of $CS_{\tilde{T}}^{(l)}$(3.5) is well-known, and it's distribution can be shown a number of ways. These depend very much on the assumptions made in the paper. This paper shall prove these results under the assumptions of strong-mixing, to this end mixing conditions given by Merleveed & Peligrad[28] provide sufficient conditions to find the asymptotic distribution.

Lemma 2. *Suppose $(Z_i)_{i\in[1,T]}$ is a strictly stationary, centred, strongly mixing sequence with finite second moment. Furthermore, let*

$$\begin{aligned}
&\liminf_{\tilde{T}\to\infty} \frac{\mathbb{E}[(S_{\tilde{T}}^{(l)})^2]}{\tilde{T}} > 0, \\
&\int_0^{\alpha_{\tilde{T}}} Q_{|X_0|}^2(u)\, du = o\left(\frac{1}{n}\right).
\end{aligned}$$

Where $Q(x)$ is the quartile function defined in (2.18). Then, $CS_{\tilde{T}}^{(l)} \xrightarrow{\mathcal{D}} BB(x)$, where $BB(x)$ is a standard Brownian Bridge

This is the first instance in this paper of Long-run covariance affecting the limiting distribution of a CUSUM-test, which will come up many times over the course of this paper. However u contains a limit and must be estimated. Durre & Fried use a kernel estimator and show that this is consistent for u, Berkes [5] use a Bartlett estimator as their test depends on the long-run variance of a fixed lag $\gamma(l)$. Betken et al[6] instead choose to use a sample estimator for u before, and after the change-point. This can be used to either estimate the critical values of the test per dataset[39][5], or to normalize the test and create an uncorrelated limiting distribution[14][6].

3.2 Testing autocovariance changes at multiple lags: R_T

The aim of this section is to extend the CUSUM-test (3.4) such that it can test multiple lags simultaneously and to make a start at showing the asymptotic distribution of this test. If a sufficiently large change occurs in at least one lag then the test will reject the new hypothesis

$$\mathcal{H}_0) \qquad \begin{pmatrix} \mathrm{Cov}(Y_1, Y_1) \\ \vdots \\ \mathrm{Cov}(Y_1, Y_{1+p}) \end{pmatrix} = \cdots = \begin{pmatrix} \mathrm{Cov}(Y_{T-p}, Y_{T-p}) \\ \vdots \\ \mathrm{Cov}(Y_{T-p}, Y_T) \end{pmatrix} \tag{3.6}$$

$$\mathcal{H}_1) \quad \exists k < T: \qquad \begin{pmatrix} \mathrm{Cov}(Y_{k-p}, Y_{k-p}) \\ \vdots \\ \mathrm{Cov}(Y_{k-p}, Y_k) \end{pmatrix} \neq \begin{pmatrix} \mathrm{Cov}(Y_k, Y_k) \\ \vdots \\ \mathrm{Cov}(Y_k, Y_{k+p}) \end{pmatrix} \tag{3.7}$$

The same process can be applied here. Due to centring the hypothesis can be restated.

$$\mathcal{H}_0) \qquad \begin{pmatrix} \mathbb{E}[Z_1^0] \\ \vdots \\ \mathbb{E}[Z_1^p] \end{pmatrix} = \cdots = \begin{pmatrix} \mathbb{E}[Z_{T-p}^0] \\ \vdots \\ \mathbb{E}[Z_{T-p}^p] \end{pmatrix} \tag{3.8}$$

$$\mathcal{H}_{1)} \quad \exists k < T: \qquad \begin{pmatrix} \mathbb{E}[Z_k^0] \\ \vdots \\ \mathbb{E}[Z_k^p] \end{pmatrix} \neq \begin{pmatrix} \mathbb{E}[Z_{k+1}^0] \\ \vdots \\ \mathbb{E}[Z_{k+1}^p] \end{pmatrix} \tag{3.9}$$

To begin to make a multi-lag test define the vector

$$\boldsymbol{S}_T^p = \begin{pmatrix} S_T^0 \\ S_T^1 \\ \vdots \\ S_T^p \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^T Y_i Y_i \\ \sum_{i=1}^{T-1} Y_i Y_{i+1} \\ \vdots \\ \sum_{i=1}^{T-p} Y_i Y_{i+p} \end{pmatrix}$$

Then the multi-dimensional CUSUM-test can be defined

$$R_T := \max_{1<k<\tilde{T}} \frac{1}{\tilde{T}} \left\| \boldsymbol{S}_T^p - \frac{k}{T} \boldsymbol{S}_{\lfloor k\tilde{T} \rfloor}^p \right\|_W^2 \tag{3.10}$$

$$= \max_{1<k<\tilde{T}} \frac{1}{\tilde{T}} \begin{bmatrix} S_k^{(0)} - \frac{k}{\tilde{T}} S_{\tilde{T}}^{(0)} \\ \vdots \\ \vdots \\ S_k^{(p)} - \frac{k}{\tilde{T}} S_{\tilde{T}}^{(p)} \end{bmatrix}^\top \begin{bmatrix} w_0 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & w_p \end{bmatrix} \begin{bmatrix} S_k^{(0)} - \frac{k}{\tilde{T}} S_{\tilde{T}}^{(0)} \\ \vdots \\ \vdots \\ S_k^{(p)} - \frac{k}{\tilde{T}} S_{\tilde{T}}^{(p)} \end{bmatrix} \tag{3.11}$$

Where W is a diagonal matrix with values $w_0, \dots, w_p$. If a sufficiently large change were to occur at a lag $l \in [0, p]$ for some point then a large value would occur at the l^{th} entry. There is little precedent to choose a matrix norm using the diagonal matrix W, to the authors knowledge no other test of it's kind in the literature does this. In fact, choosing the inverse long-run covariance matrix U^{-1}(2.7) arguably gives a better limiting distribution. W provides flexibility in the test in several ways:

1. Methods for estimating U are computationally expensive to calculate. Using a kernel method gives each component complexity $O(\tilde{T}^2)$, total complexity of estimating U is $O\left((\frac{p+1}{2})^2 \tilde{T}^2\right)$. This can become problematic when testing higher lags, which is the entire benefit of the test R_T.

2. The diagonals of U^{-1} scale the components of the vectors based on their variance. This is incredibly useful as it reduces the likelihood of false positives in the test statistic due to unusually large values caused by higher variance. One can set W to be the diagonals of U^{-1} to do this much more cheaply than computing the entire matrix.

3. The covariance function are often "tight" i.e. $\gamma(l)$ diminishes rapidly as l increases, this is always the case when a sequence is strongly mixing thanks to Davydov's inequality[13]. Choosing W such that the diagonals descend from 1 to 0 to prevent proportionally large changes at higher lags which may not indicate a structural change.

The long-run variance of interest for this test is that of the vector of values $[Z_i^0, \ldots, Z_i^p]^\top$:

$$U = \lim_{h\to\infty} \operatorname{Cov}\left(\begin{pmatrix} Y_1Y_1 \\ Y_1Y_2 \\ \vdots \\ Y_1Y_{1+p} \end{pmatrix}, \begin{pmatrix} Y_{1+h}Y_{1+h} \\ Y_{1+h}Y_{2+h} \\ \vdots \\ Y_{1+h}Y_{1+h+p} \end{pmatrix}\right) \tag{3.12}$$

This can be estimated using a kernel-based method:

$$\hat{U}_{i,j} = \frac{1}{T}\sum_{t=1}^{\tilde{T}}\sum_{s=1}^{\tilde{T}}\left(\hat{Y}_s\hat{Y}_{s+i} - \frac{1}{\tilde{T}}S_{\tilde{T}}^{(i)}\right)\left(\hat{Y}_t\hat{Y}_{t+j} - \frac{1}{\tilde{T}}S_{\tilde{T}}^{(j)}\right)k\left(\frac{|s-t|}{b_{\tilde{T}}}\right) \tag{3.13}$$

where $k : \mathbb{R} \to [-1, 1]$ is a kernel function and $b_{\tilde{T}}$ is the bandwidth. The choice of kernel function seems highly data dependent, and the choice should be based on the dimension of the test[14], which relates to the number of lags p taken in this setting. The flat-top kernel is chosen here

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

in the same fashion as in Dürre and Fried(2020)[39], where $b_{\tilde{T}} = \tilde{T}^{\frac{1}{3}}$. This kernel-based estimator is shown to be weakly consistent for U(3.12), but it has not been shown for a series standardized in the same manner as Y.

Lemma 3. *The Long-run covariance estimator for the series $(Y_{i,T})_{i\in[1,T]}$, $\hat{U}$, gotten from the estimated standardization $Y_{i,T} = \phi\left(\frac{X_i - \hat{\mu}}{\hat{\sigma}}\right)$, is a consistent estimator for U*

This matrix, and its estimator $\hat{U}$, will be essential in characterizing the limiting distribution of R_T. The proof of this distribution is known, and loosely proven in the literature. To prove this, in more detail than currently exists, the following steps will be taken

1. Show the limiting distribution of the column $\frac{1}{\sqrt{T}}(\boldsymbol{S}_{\tilde{T}}^l - \tau\boldsymbol{S}_{\lfloor\tilde{T}x\rfloor}^l)_{x\in[0,1]}$

2. Use this result to show the limiting distribution of the test

$$\left(W_T^2(x)\right)_{x\in[0,1]} = \left(\left\|\boldsymbol{S}_T^l - x\boldsymbol{S}_{\lfloor Tx\rfloor}^l\right\|_{\hat{U}^{-1}}^2\right)_{x\in[0,1]}$$

3. Finally, show the limiting distribution of R_T as a Corollary

Step-1 will be shown here. Step-2 will require a detour, $W_T^2(x)$ is an earlier iteration of the test R_T for which the asymptotic distribution has been shown in some capacity. The next section will detail this step, and from there step-3 comes quite naturally.
First, for step-1:

Proposition 1. *$(X_i)_{i\in\mathbb{N}}$ be a 1-dimensional stationary, and strongly mixing sequence with mixing coefficients $(\alpha_k)_{k\in\mathbb{N}}$ satisfying $\alpha_k = O(k^{-3-\epsilon})$ for some $\epsilon > 0$. Let $\phi : \mathbb{R} \to \mathbb{R}$ be a bounded, non-zero, function such that $Y_i = \phi(\frac{X_i-\mu}{\sigma_i})$ where μ is the median and σ_i is the MAD.*
for some p and $\tilde{T} = T - p$, and assume that matrix

$$U = \sum_{h=-\infty}^{\infty} \mathrm{Cov}\begin{pmatrix} Y_1Y_1 \\ \vdots \\ Y_1Y_{1+p} \end{pmatrix}\begin{pmatrix} Y_{1+h}Y_{1+h} \\ \vdots \\ Y_{1+h}Y_{1+h+p} \end{pmatrix}$$

is positive definite. Then

$$\frac{1}{\sqrt{\tilde{T}}}(\boldsymbol{S}^p_{\lfloor Tx\rfloor} - \frac{\lfloor Tx\rfloor}{T}\boldsymbol{S}^p_T) \xrightarrow{\mathcal{D}} BB(x), \tag{3.14}$$

where $BB(t)$ is an p-dimensional Brownian-Bridge with covariance matrix $\mathrm{Var}\,(BB(t)) = t(1-t)U$.

Remark:
Most of these requirements are assumed in the setting of this paper, the only one that has not been addressed yet in this paper is that U is positive definite. This implies that there is no linear-dependence between any of the columns of U. A column represents the long-run covariance of one particular lag with every other lag being tested, and so any two points $Z^l_{i_1}, Z^{(l)}_{i_2}$ should not correlate with a particular column $[Z^{(1)}_j, \ldots, Z^{(p)}_j]^\top$ in the same way. This is not a strong requirement in a strongly mixing series, nor is it common that a series does not fulfil this, but the sample covariance $\hat{U}$ is almost never positive definite. This will be discussed more in the Simulations section, and a work-around shown.

3.3 $W_T^2(x)$ and the limiting distribution of R_T

In this section the related test $W_T^2(x)$ is introduced, it's asymptotic properties shown, and finally the asymptotic size of R_T is shown.

$$W_T^2(x) = \frac{1}{T}\left\|\boldsymbol{S}^{(p)}_{\lfloor Tx\rfloor} - x\boldsymbol{S}^{(p)}_T\right\|^2_{\hat{U}^{-1}} \tag{3.15}$$

$$= \frac{1}{T}\left(\boldsymbol{S}^{(p)}_{\lfloor Tx\rfloor} - x\boldsymbol{S}^{(p)}_T\right)^\top \hat{U}^{-1}\left(\boldsymbol{S}^{(p)}_{\lfloor Tx\rfloor} - x\boldsymbol{S}^{(p)}_T\right) \tag{3.16}$$

Dürre & Fried introduced this test in their 2019 paper[14]. It is a more generic test but has only been stated here for such that it can detect changes in cross-sectional dependence. This was done since it is only used for comparison, and for results concerning R_T. After applying a functional like $\sup_{x\in[0,1]}$ its only surface-level difference is its use of $\hat{U}^{-1}$ in place of W. Additional assumptions are made for this test, most of which are covered by our more constrained system.

Assumption 1 Let $(X_i)_{i\in\mathbb{N}}$ be strongly-mixing with mixing constants $(a_k)_{k\in\mathbb{N}}$ fulfilling $a_k = O(k^{-2-\epsilon})$
Remark 1 This is a weaker mixing assumption than what is already made in this paper on the series $(X)_{\in\mathbb{N}}$, which is $a_k = O(k^{-3-\epsilon})$. Any results that hold for this mixing must also hold for a stronger mixing series.
Assumption 2. Let $\psi : \mathbb{R}^p \to \mathbb{R}^s$ be a function fulfilling:

- $\psi \in L^\infty(\mathbb{R}^s, \mathbb{R}^p)$
- Every component of ϕ is two times continuous differentiable in $\mathbb{R}\backslash D$ and there exists $C_1, C_2 > 0$ and $|\phi^{(i)}('x)^T x| \le C_1$ and $x^T \phi''^{(i)}(x)x \le C_2, \forall x \in \mathbb{R}^C$ and $i = 1, \dots, s$

Remark 2. The function ϕ is not the only function that may be used to pre-process in this setting, standardizing is still done using μ and MAD. These assumptions cover a variety of requirements, such as non-degenerate limits, robustness, etc. This is important as a variety of functions can be used to preprocess the time series $\mathbb{X}$. As we only focus on the function ϕ these assumptions are true, and not really in need of stating. The ϕ function is continuous at all points. $\phi(x)$ is undefined at $\{x \in \mathbb{R} : |x| = k\}$, this requires a little care but is addressed in lemma-3.
Assumption 3. $det(U) > 0$ where

$$U = \sum_{h=-\infty}^{\infty} \mathrm{Cov}\left(\begin{pmatrix} Y_1Y_1 \\ \vdots \\ Y_1Y_{1+p} \end{pmatrix} \begin{pmatrix} Y_{1+h}Y_{1+h} \\ \vdots \\ Y_{1+h}Y_{1+p+h} \end{pmatrix} \right) \tag{3.17}$$

Remark 3. This is another way of requiring $U \succ 0$, which is already assumed in prop-1.
Assumption 4. Let $(\mu_T)_{T\in\mathbb{N}}$ and $(\sigma_T)_{T\in\mathbb{N}}$ be real valued stochastic processes fulfilling

$$\mu_T - \mu = O_{\mathbb{P}}(T^{-\frac{1}{2}}) \text{ and } \sigma_T - \sigma = O_{\mathbb{P}}(T^{-\frac{1}{2}})$$

Remark 4. Dürre and Fried note that this is likely not necessary, as it has been shown in Yoshihara(1995)[44] that median and MAD are consistent under strong mixing and continuous density of innovations.
These results, and the remainder of the setting already assumed in this paper, are necessary for the following result:

Theorem 6. *Let X_i be a stationary $1-$dimensional series. Let $\psi : \mathbb{R} \to \mathbb{R}$ be a bounded function such that $Y_i = \psi\left(\frac{X_i - \mu}{\sigma}\right)$ where μ is the median and σ is the MAD. Furthermore, let assumptions 1-4 hold. Then*

$$(W_T(x)^2) \xrightarrow{\mathcal{D}} \sum_{i=0}^{p} \bar{B}B_i(x)^2 \,, \tag{3.18}$$

where $(\bar{B}B_i(x))_{i\in[0,p]}$ are mutually independent standard Brownian Bridges.

This is an incredibly powerful result, because the components of the vector are decorrelated. Meaning that under the null-hypothesis the test statistic, and the critical values used, are not dependent on the underlying data This is the benefit of normalizing the test with the matrix $\hat{U}$. It is quite natural now to state the following

Corollary 1. $(X_i)_{i\in\mathbb{N}}$ *be a* 1*-dimensional, centred, stationary, and strongly mixing sequence with mixing coefficients* $(\alpha_k)_{k\in\mathbb{N}}$ *satisfying* $\alpha_k = O(k^{-3-\epsilon})$ *for some* $\epsilon > 0$. *Let* $\phi : \mathbb{R} \to \mathbb{R}$ *be a bounded, non-zero function such that* $Y_i = \phi(\frac{X_i-\mu}{\sigma_i})$ *where* μ *is the median and* σ_i *is the MAD. Let* $U \succ 0$ *where* U *is the long-run covariance* (3.17). *Then*

$$R_T(x) \xrightarrow{\mathcal{D}} (\boldsymbol{BB}(x))^\top (\boldsymbol{BB}(x)) \tag{3.19}$$

Where $\boldsymbol{BB}(x)$ *is a* $p+1$ *dimensional Brownian Bridge with covariance matrix*

$$Cov(\boldsymbol{BB}(x), \boldsymbol{BB}(y)) = x(1-y)U \ .$$

This is already a slightly weaker result than that of $W_T^2(x)$ since the asymptotic distribution is dependent on the underlying series, and its long-run covariance U. However p-values can be estimated quicker, and it might be that other methods of weighting the test give similar with less runtime. The effectiveness of this sacrifice will be the topic of the next section.

Simulations

The test will be examined to see if it has sufficient power in a number of examples, and some features of the test will be verified. Three different choices of weight values $w_0, \ldots, w_p$ of (3.11) will be tested. Constant weights $w_i = 1$ may be a good choice if little is known about the underlying data, and there is no obvious choice of weighting matrix W that can improve the test, this shall be denoted R^e_T.
Descending weights, modelled by $w_i = 1 - \frac{i}{p}$ can down-weight smaller changes at later lags. This may be a good choice when strong-mixing is assumed, as the covariance decreases at later lags. A small shift could cause an unusually large value, so this weighting may reduce type 1 errors at higher lags. At $p = 1$ this creates a 0 entry in the second diagonal of W making for a test for change in variance only. It will therefore be omitted until testing $p \geq 2$. This will be denoted as R^d_T
Finally, weights taken from the diagonals of $\hat{U}$, where $w_i = \hat{U}^{-1}_{i,i}$ could prevent errors by suppressing large variance in the series $(Z_i)_{i \in [1,\tilde{T}]}$ caused by the underlying data. This choice will be denoted by R^s_T
w_i are defined in the range $i \in \{0, \ldots, p\}$. For comparison two tests detecting changes in the autocovariance structure of the time series are selected.

Test 1

Berkes et. al. [5] suggest a similar CUSUM-based test that can be used to detect changes in autocovariance. Consider the stationary series $(X)_{i \in [T]}$ and, for

$$M^{(p)}_n(t) = \left\{ \frac{1}{n^2} \sum_{i=1}^{\lfloor (n+1)t \rfloor} \left((X_i - \bar{X}_n)(X_{i-p} - \bar{X}_n) - \bar{X}^{(p)}_n \right) \right. \quad ,$$

define the following two CUSUM test statistics:

$$B^{(p)}_T = \frac{1}{\hat{\gamma}(p)} \sup_{0<t<1} M^{(p)}_n(t) \tag{4.1}$$

$$B'^{(p)}_T = \frac{1}{\hat{\gamma}(p)} \sup_{0<t<1} \left| M^{(p)}_n(t) \right| \tag{4.2}$$

Where $\hat{\gamma}_n(p)$ is an estimator for

$$\gamma^2(p) = \lim_{n \to \infty} \frac{1}{n} \operatorname{Var} \left(\sum_{i=1}^{n} (X_i - \mu)(X_{i-p} - \mu) \right)$$

at some lag p.

Berkes et. al. do not outline how $\gamma^2(p)$ is estimated. Since they use a Bartlett estimator to estimate the long run variance the same method shall be used here.

$$\hat{\gamma}^2(p) = \frac{1}{n}\sum_{i=1}^{n}\left((X_i-\mu)(X_{i-r}-\mu) - \frac{1}{n}\sum_{j=1}^{n}(X_j-\mu)(X_{j-r}-\mu)\right) \tag{4.3}$$

$$+ 2\sum_{j=1}^{q}\left(1-\frac{j}{q+1}\right)\left(\frac{1}{n-j}\right)\sum_{i=i}^{n-j}\left((X_i-\mu)(X_{i-r}-\mu) - \frac{1}{i}\sum_{l=1}^{i}(X_{l+j}-\mu)(X_{l-r}-\mu)\right) \tag{4.4}$$

$$\cdot\left((X_{i+j}-\mu)(X_{i+j-r}-\mu) - \frac{1}{i+j}\sum_{l=1}^{i+j}(X_l-\mu)(X_{l+j-r}-\mu)\right) \tag{4.5}$$

Berkes' et al do not suggest any form of pre-processing before applying $B_T^{(p)}$ and $B_T^{'(p)}$, and both tests only detect changes at a chosen lag p. Furthermore $B_T^{'(p)}$ is a two-tailed test which is uncommon for CUSUM-tests. These differences make for a good choice for comparison to R_T, as they could show if the additionally assumptions, preprocessing, and computation time necessarily improve results.

Test 2

Betken and Micali et. al. [6] proposed a test to detect structural changes using ordinal patterns. For time series data $(\xi_i)_{i\in[0,r]}$ let $(X_i)_{i\in[1,r]}$ be the series of increments where $X_t := \xi_t - \xi_{t-1}$. Additionally let $(\pi_0,\dots,\pi_r)$ be its permutation, where π_j is the rank of X_j based on its magnitude in the series $(X_i)_{i\in[1,r]}$. Let the function Π perform this ordering, where $\mathcal{S}_r$ is the set of $r+1$ permutations:

$$\Pi : \mathbb{R}^{r+1} \to \mathcal{S}_r, \quad (X_0 \dots X_r) \to (\pi_0,\dots,\pi_r) \tag{4.6}$$

The test is based on estimating the turning rate of the series across a sliding window, over length 3 the set of $\tau = \{(0,2,1),(2,0,1),(1,2,0),(1,0,2)\}$ is the collection of permutations where a change occurs. In the differenced series this is a zero-crossing. For a differenced time series of size T and a block of size m the collection of $T_b = \lfloor\frac{T+2}{m+2}\rfloor$ random variables $\hat{q}_{1,m}\dots,\hat{q}_{T_b,m}$ each represent the turning rate over a block of size m and are defined by

$$\hat{q}_{j,m} = \frac{1}{m}\sum_{i=0}^{m-1}\sum_{\gamma\in\tau}\mathbb{1}(\Pi(X_{(j-1)(m+2)}, X_{(j-1)(m+2)+i+1}, X_{(j-1)(m+2)+i+2}) = \gamma) \qquad j=1,\dots,T_b \tag{4.7}$$

Sinn and Keller [36] show that a change in this zero-crossing estimator series indicates a change in the autocovariance structure of a time series. To detect a change in this zero-crossing estimator Betken and Micali propose a CUSUM test

$$G_T = \max_{k=1,\dots,T_b}\left|\sum_{j=1}^{k}\hat{q}_{j,m} - \frac{k}{T_b}\sum_{j=1}^{n_n}\hat{q}_{j,m}\right| \tag{4.8}$$

Power of this test is known when $(X_i)_{i\in[1,T]}$ is stationary, admits a linear representation $X_i = \sum_{j=0}^{\infty} a_j Z_{t-j}$ with $\sum_{j=0}^{\infty}|a_j| < \infty$, with $(Z_j)_{j\in\mathbb{N}} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,\sigma_Z^2)$ admitting continuous bounded density and finite second moment $\mathbb{E}[|Z_j|^2] < \infty$.

For $(Z_j)_{j\in\mathbb{Z}}$ forming a stationary Gaussian time series, it can be shown that a change in the mean of the turning rate series is associated with a change at lag 1 of the autocovariance structure of $(X_t)_t$. However, since ordinal patterns distributions do not vary under the assumption of stationarity (null hypothesis), a corresponding change in the expectation of the turning rate must be reflected in some structural change of the underlying time series. Thus, the test do not restrict to detecting changes in autocovariance structure only. On top of this flexibility it has a fast computation time compared to R_T: calculating permutations for small lengths has negligible computation time, $\hat{q}_{j,m}$ has complexity $O(|\tau| \cdot m)$, and G_T itself is linear in T . This makes for a good test of comparison as it promises the same flexibility as R_T in a smaller package, and requires almost no additional assumptions to work.

4.1 Estimating Critical Values

Before experiments can be conducted it is necessary to estimate p-values. Since the test involves a norm, and a supremum is being applied it is a 1-tailed test. A critical value must be estimated for each choice of lag p, and the weighting $w_0, \ldots, w_p$ too.
The limiting distribution of the test includes the long-run covariance U, and so a critical value must also be estimated per dataset being tested. The following procedure is suggested in Dürre & Fried[39]:

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

2. reproduce the cross sectional dependence by multiplying Z with L of the Cholesky decomposition $\hat{U} = LL^T$ of the long-run covariance of the dataset of interest. Set $V = ZL$.

3. calculate the weighted test statistic

$$\tilde{R}_T = \frac{1}{T} \max_{k=1,\ldots,\tilde{T}} (\sum_{t=1}^{k} V_{[t,]} - \frac{k}{\tilde{T}} \sum_{t=1}^{\tilde{T}} V_{[t,]}) W (\sum_{t=1}^{k} V_{[t,]} - \frac{k}{\tilde{T}} \sum_{t=1}^{\tilde{T}} V_{[t,]}) \tag{4.9}$$

Step 1 and 2 together generate correlated data under the null-hypothesis. Step 3 generates a sample statistic $\tilde{R}_{\tilde{T}}$, which can be repeated many times to estimate the distribution for this particular choice of weighting, lag, and dataset. Critical values at any desired significance level can be taken from the quantiles. In this paper the significance will be taken to be 5%.
This is an application of the Monte Carlo method, and is useful for estimating critical values when an analytical solution is not readily available. The drawback is that this could be computationally intense, fortunately most of the time complexity of this process comes from estimating the matrix U - which need only be done once per dataset & weighting - and the number of sample statistics $\tilde{R}_{\tilde{T}}$ that are generated.
It is worth noting that estimating U using the kernel method in line (3.13) could be affected by the new regime if the change-point is early in the dataset, and perhaps should only be estimated for a portion of the dataset to prevent this. This could make for a poor estimation of U due to a small sample, and deciding where to cut off the estimating may be tricky when the exact change-point is unknown (indeed, in a practical application of

the test R_T it likely is not.)
Critical values for several datasets, lags, and weightings can be found in table(4.1). Something that can be seen already from this table is that critical values are sensitive to the type of data, and the maximum lag being tested. AR(1) Heavytail, which here denotes an AR(1) model with white noise ϵ_t t-distributed with 5 degrees of freedom, experiences a large jump from $T = 100$ to $T = 500$. This hints at the fact that the test might have poor power for heavytail models. Additionally, the AR(1) experiences a large jump in critical value for $p = 3$ over the two sample sizes, this may affect the size of the test over higher lags, and increase the amount of false positives.

		95th Percentile					
W	**p**	**AR(1)**		**AR(1) Heavytail**		**MA(1)**	
		T=100	**T=500**	**T=100**	**T=500**	**T=100**	**T=500**
	0	6.1423	6.6068	4.107	10.825	3.4844	4.7256
e	1	3.67024	4.741151	5.073585	13.1973	1.633085	3.722304
	3	6.3277	16.3715	8.2018	60.480096	5.94137	6.039725
	0	5.3154	6.4292	4.4615	10.546	3.4968	4.8262
d	1	1.9786	7.4186	8.1899	6.7745	3.5809	10.517
	3	3.7106	13.8782	5.97125	53.39764	5.21845	4.551697
	0	0.62855	2.2094	0.42038	0.7665	2.7439	1.0645
s	1	1.397702	1.654273	1.014046	2.347253	0.663165	1.940664
	3	3.9838	11.31809	0.81145	9.3283	2.49872	4.95891

TABLE 4.1: 95th percentile values for different models and sample sizes, where e,d, and s refer to the choice of weight matrix W.

4.2 Run time and complexity

Something that will be apparent throughout this part of the paper is that R_T will not have results for sample sizes larger than $T = 500$, and lags $p > 3$ will rarely appear. This is due to the limited computing power of the author, paired with the greater complexity of the algorithm. A small experiment to determine the average run time required to use each in a variety of scenarios.
All scripts are run in Matlab, on an Asus Vivobook with Intel Core-i7 CPU, and 8 gigabytes of RAM. and results can be found in tables (4.2) to (4.4).At lag 1 R_T^e,R_T^d, B_T, and B_T' all take less than one millisecond per iteration per test statistic. G_T takes a little bit longer, but does not seem to suffer from running higher samples.
R_T requires the computation of $\tilde{T}$ test statistics, the $(p+1)\tilde{T}$ cumulative sums S_k^p can be precomputed, each one with $O(k)$, and the matrix multiplication can be done in only $p + 1$ multiplications since W is diagonal. R_T therefore has total time complexity is $O\left((p+1)\tilde{T}\right)$, and scales relatively well in T and p.

R_T^s is the slowest of all tests examined, and increases dramatically in runtime for higher samples T. This can be entirely attributed to the estimation of the diagonals of the long-run-covariance matrix $\hat{U}$. Indeed the kernel method used involves a double sum over $\tilde{T}$, and each involves a computation of a cumulative sum. If the Cumulative Sums $S_{\tilde{T}}^p$ are pre-computed the total complexity if $\hat{U}_{i,j} = O(\tilde{T}^2)$, which must be computed $p+1$ times. The bottle-neck in computing power for all R_T tests is in fact the matrix $\hat{U}$ that must be estimated, and factorized to estimate p-values. For sample sizes higher than 3 and samples larger than 500 the matrix $\hat{U}$ can not be computed in acceptable time on the computer being used.

	$T = 100$	$T = 500$	$T = 2000$
G_T	0.01413	0.0038655	0.011185
R_T^e	0.0061709	0.00010816	0.00013246
R_T^d	0.00040296	5.9085e-05	9.675e-05
R_T^s	0.049672	2.4555	128.72
B_T'	0.0038594	0.00046854	0.0010246
B_T	0.0012601	0.00049515	0.0009838

Table 4.2: Runtime Results for Lag = 1

	$T = 100$	$T = 500$	$T = 2000$
R_T^e	6.6775e-05	0.00011123	0.00016282
R_T^d	5.0615e-05	6.869e-05	0.00013906
R_T^s	0.070947	3.6889	193.29
B_T'	0.00037804	0.00046277	0.0010125
B_T	0.00035018	0.00045823	0.00094086

Table 4.3: Runtime Results for Lag = 2

	$r = 3$			$r = 5$		
	$T = 100$	$T = 500$	$T = 2000$	$T = 100$	$T = 500$	$T = 2000$
R_T^e	0.00021916	8.972e-05	0.00023064	0.00032436	0.00010988	0.0006473
R_T^d	0.00017937	7.037e-05	0.00017176	0.00023621	9.2465e-05	0.00060691
R_T^s	132.12	5.0013	265.23	0.40667	7.3578	420.67
B_T'	0.0011603	0.0004481	0.0011003	0.0010972	0.00043094	0.00098113
B_T	0.00080713	0.00041662	0.0010448	0.0010799	0.00042909	0.00092548

Table 4.4: Runtime Results for Lag = 3 and 5

4.3 Sensitivity to changepoint, and the Huber ϕ function

In this section a some general properties of R_T that were highlighted earlier in the paper shall be addressed. Namely sensitivity to location of change-point, and outliers. Firstly, note that CUSUM-type tests are more effective when the change occurs in the middle of a sample. This is because it has an equal amount of data on both regimes, meaning sufficient data to accumulate deviations under both regimes in equal amount. To test this, three models were considered.

$$X_t = \begin{cases} \phi_0 X_{t-1} + \epsilon_t & t \in [1, \lfloor Tx \rfloor] \\ \phi_a X_t + \epsilon_t & t \in [\lfloor Tx \rfloor + 1, T] \end{cases} \quad , \qquad X_t = \begin{cases} \phi_0 \epsilon_{t-1} + \epsilon_t & t \in [1, \lfloor Tx \rfloor] \\ \phi_a \epsilon_t + \epsilon_t & t \in [\lfloor Tx \rfloor + 1, T] \end{cases} \tag{4.10}$$

Where the first is an AR(1) process, and the second is an MA(1) process, both with $\epsilon \sim \mathcal{N}(0, 1)$. And finally

$$X_t = \begin{cases} \phi_0 X_{t-1} + \epsilon_t & t \in [1, \lfloor Tx \rfloor] \\ \phi_a X_t + \epsilon_t & t \in [\lfloor Tx \rfloor + 1, T] \end{cases} \tag{4.11}$$

Where instead ϵ is t-distributed with 5 degrees of freedom, this process will be written as AR1ht (meaning AR(1), heavytail) in the graphs below. The power was examined for R_T with weighting e ($w_i = 1$), and s ($w_i = \hat{U}_{i,i}$), coefficients $\phi_0, \theta_0 = 0.3$, $\phi_a, \theta_a = 0.8$ and changepoint $x \in \{0.1, 0.2, \ldots, 0.9\}$.
Results can be found in figures-4.1. Both tests perform worse when the change-point is close to the start or end of the series, regardless of the model. The middle graph, where the series with t-distributed ϵ, has the strongest performance regardless of location.
R_T^s is least effective here, this seems to imply that this instance of the test performs worse when the variance if larger and will be addressed in later sections.
The MA(1) model is both most sensitive to the change-point, and and performs worst for R_T^e, this hints that processes that are 'more random' can benefit more from the weighting used in R_T^s.

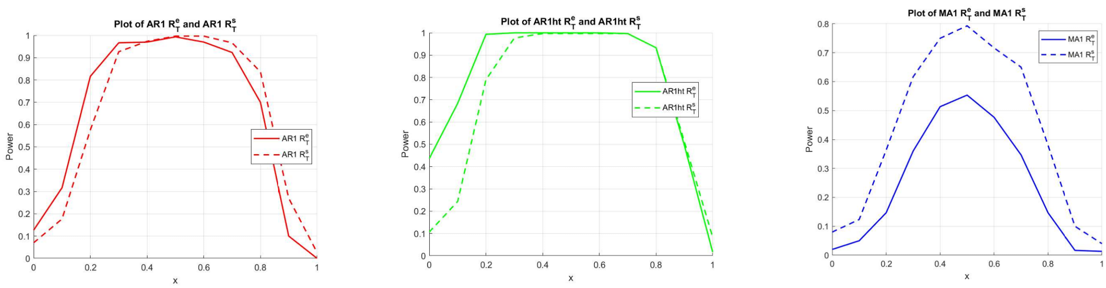


FIGURE 4.1: Power at changepoint for $\phi_0 = 0.3$ to $\phi_a = 0.8$ for different changepoints $x \in [0, 1]$

Secondly, CUSUM-type tests are sensitive to outliers, functionals like sup & max easily give false positives if the data contains a sharp change in estimate. The function ϕ downweights the affect of these outliers, to test this power will be examined for R_T when the cap of the huber-phi function is $k = \{1.5, 1000\}$ for a dataset with frequent outliers. The following experiment is proposed.

1. Generate an MA(1), with $\epsilon_t \sim \mathcal{N}(0, \sigma)$, time series of length T under the null-hypothesis
2. at a random point in the series add a constant $c \in [1, 4]$

The results over the set c can be found in figure-4.2. The graph shows that the size of R_T^e is quite poor when a large boundary is chosen for ϕ, regardless of how large the outlier actually is.

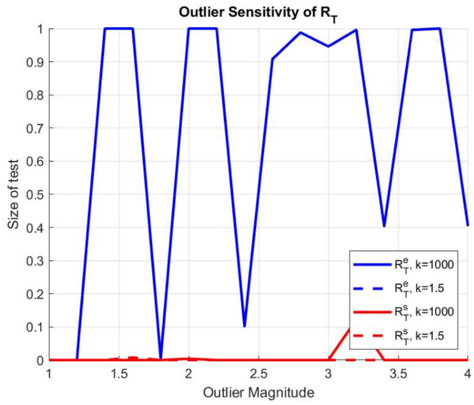


FIGURE 4.2: Size of R_T for outliers of different magnitudes

What is peculiar is that R_T^s has good size regardless of the choice of k inside the function ϕ. The minor spike after magnitude 3 seems more like an anomaly as the size is close to 0 afterwards. This seems to indicate that robust standardizing, both of the time-series X_i and the test R_T, is effective at suppressing the effect of outliers. However R_T^e, $k = 1.5$ suffers from no anomalies, so it seems to be slightly more effective to apply ϕ.

4.4 Effectiveness of testing $p = 0$

R_T does not compare directly to G_T or B_T and B'_T even when changes in lag 1 are examined only. This is because even at lag 1 R_T considers changes in variance. This makes for unfair comparison, since it is difficult to create a change in autocovariance at one location only without affecting the variance. To test if R_T truly benefits from this the following two modified tests will be examined.

$$R_T^{1,e} = \max_{k \in [1, T-1]} \left(S_k^1 - \frac{k}{T-1} S_T^1 \right)^2 \quad \text{and} \quad R_T^{1,s} = \max_{k \in [1, T-1]} \hat{U}_{2,2}^{-1} \left(S_k^1 - \frac{k}{T-1} S_T^1 \right)^2 \tag{4.12}$$

Where $\hat{U}_{2,2}^{-1}$ is the inverse of the second diagonal of the matrix $\hat{U}$. These two tests will be compared to R_T^e and R_T^s for AR(1) both with white noise ϵ_t normally distributed with $\mathcal{N}(0, 1)$ and t-distributed with 5 degrees of freedom, and MA(1) for sample sizes $T \in \{100, 500\}$ and changepoints $x \in \{0.1, 0.25, 0.5\}$, with a parameter change $\phi_0 = 0.3$ to $\phi_a = 0.8$.
Results can be seen in tables 4.5 to 4.7. These results contradict the assumption that testing for variance too creates an advantage or should improve results. Across the board R_T^1 outperforms or matches R_T. These results, together with observations made on the critical value, imply that any advantage gotten from testing multiple changes simultaneously likely makes for a slightly unstable critical region for the test. For the remainder of this section R_T alone will be tested.

4.5 Power when $p = 1$

In this section the power under the alternative hypothesis will be tested under several ARMA models at lag $p = 1$. This will be compared to the other tests introduced at the

	$T = 100$			$T = 500$		
$\tau =$	0.10	0.25	0.50	0.10	0.25	0.50
R_T^e	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
R_T^s	0.9210	0.9990	0.9840	1.0000	1.0000	1.0000
$R_T^{1,e}$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$R_T^{1,s}$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

TABLE 4.5: Power of both R_T and R_T^1 tests for AR(1) with change at different change-points, for $T = \{100, 500\}$.

	$T = 100$			$T = 500$		
$\tau =$	0.10	0.25	0.50	0.10	0.25	0.50
R_T^e	0.865	0.923	0.904	0.91	0.949	0.957
R_T^s	0.268	0.465	0.282	0.372	0.513	0.917
$R_T^{1,e}$	0.978	0.983	0.988	0.988	0.2999	1.0000
$R_T^{1,s}$	0.548	0.754	0.862	0.633	0.849	0.981

TABLE 4.6: Power of both R_T and R_T^1 tests for MA(1) with change at different change-points, for $T = \{100, 500\}$.

start of this section.

AR(1) The first example generated by synthetic data is an AR(1) process

$$X_t = \begin{cases} \phi_0 X_{t-1} + \epsilon_t & t \in [1, \lfloor Tx \rfloor] \\ \phi_a X_{t-1} + \epsilon_t & t \in [\lfloor Tx \rfloor + 1, T] \end{cases}$$

where $|\phi_0|, |\phi_a| < 1$ and the change-point $x \in (0, 1)$, the autocovariance of this process has three cases.

- **Case 1.** $t, t + k \leq \lfloor Tx \rfloor$

$$Cov(X_t, X_{t+k}) = \begin{cases} \dfrac{1}{1 - \phi_0^2} & k = 0 \\ \dfrac{\phi_0^k}{1 - \phi_0^2} & k > 0 \end{cases}$$

- **Case 2.** $t \leq \lfloor Tx \rfloor$, and $t + k \geq \lfloor Tx \rfloor + 1$

$$\begin{aligned} Cov(X_t, X_{t+k}) &= Cov(X_t, \phi_a X_{t+k-1} + \epsilon_{t+k}) \\ &= Cov(X_t, \phi_a(\phi_a X_{t+k-2} + \epsilon_{t+k-1}) + \epsilon_{t+k}) \\ &\vdots \\ &= Cov(X_t, \phi_a^m X_{t+k-m} + \sum_{i=0}^{m} \phi_a^i \epsilon_{t+k-i}) \end{aligned}$$

	$T = 100$			$T = 500$		
$\tau =$	0.10	0.25	0.50	0.10	0.25	0.50
R_T^e	0.905	0.917	0.88	1.0000	1.0000	1.0000
R_T^s	0.997	0.999	0.999	0.998	1.0000	1.0000
$R_T^{1,e}$	1.0000	1.0000	0.995	1.0000	1.0000	1.0000
$R_T^{1,s}$	0.994	0.984	0.924	1.0000	1.0000	1.0000

TABLE 4.7: Power of the tests for AR(1), with white noise ϵ_t t-distributed, with change at different change-points, for $T = \{100, 500\}$.

Where $m = \lfloor Tx \rfloor - k$, and the second term in the covariance is uncorrelated with X_t, so we disregard it going forward.

$$\begin{aligned} Cov(X_t, X_{t+k}) &= \phi_a^m \operatorname{Cov}(X_t, X_{t+k-m}) \\ &= \phi_a^m \operatorname{Cov}(X_t, \phi_0 X_{t+k-m-1} + \epsilon_{t+k-m-2}) \\ &\vdots \\ &= \phi_a^m \operatorname{Cov}(X_t, \phi_0^{k-m} X_t + \sum_{i=0}^{k-m} \phi_0^i \epsilon_{t+k-m-i}) \end{aligned}$$

Giving the final covariance function

$$\operatorname{Cov}(X_t, X_{t+k}) = \begin{cases} \dfrac{1}{1-\phi_0^2} & k = 0 \\ \phi_a^m \phi_0^{k-m} & k > 0 \end{cases}$$

- **Case 3.** $t, t+k \geq \lfloor Tx \rfloor + 1$

$$Cov(X_t, X_{t+k}) = \begin{cases} \dfrac{1}{1-\phi_a^2} & k = 0 \\ \dfrac{\phi_a^k}{1-\phi_a} & k > 0 \end{cases}$$

A brief examination of the power of all three tests for this AR(1) process can be found in figure 4.3. Here ϕ_0 is fixed at 0.3 and the power is instead measured over the change $h = \phi_a - \phi_0$ with three choice of change-point $x \in (0, 1)$. Additionally table-4.10 shows power for different choices of T for the same change-points.
On a surface level R_T^e has better size under the null-hypothesis, but worse power across most of the graph than R_T^s. G_T is much more consistent, with a size very close to 0.05 under the null-hypothesis, and has shallower increase as h goes up.
B_T fails to detect the change at all, regardless of location and sample size. B_T' performs much better, but still has quite poor power, This discrepancy can only be because of the use of a supremum. B_T likely contains many "large" results that were negative, and in cases where a change in covariance creates a large negative value in the test it goes undetected. Indeed a change from $\phi_0 = 0.3$ to $\phi_a = 0.8$ would create a large negative value within the CUSUM.

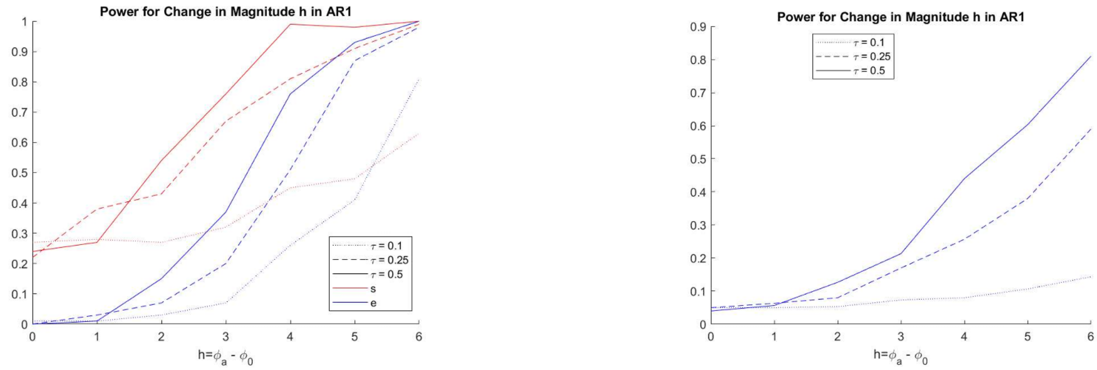


FIGURE 4.3: Power of R_T(left) and G_T(right) for different changes h

Interestingly the power becomes worse the closer to $x = 0.5$ the changepoint is. This is likely because of the bartlett estimator, when it has an equal amount of data on both regimes it does not fit very well, and so standardizes the test poorly. It manages to overcome this when the sample size is sufficiently large, however.

	$T = 100$			$T = 500$			$T = 2000$		
$x =$	0.1	0.25	0.5	0.1	0.25	0.5	0.1	0.25	0.5
R_T^e	0.954	0.942	0.889	1.0	1.0	1.0			
R_T^s	0.468	0.648	0.881	0.587	0.997	1.0			
G_T	0.183	0.221	0.2590	0.073	0.246	0.39	0.156	0.721	0.916
$B_T^{(1)}$	0	0	0	0	0	0	0	0	0
$B_T^{'(1)}$	0.393	0.205	0.044	0.441	0.009	0.109	0.560	0.243	1.00

TABLE 4.8: Power under AR(1) with change $\phi_0 = 0.3$ to $\phi_a = 0.8$ at different change-points, for $T = \{100, 500, 2000\}$

MA(1) Next the power of the test under a moving average process is examined

$$X_t = \begin{cases} \theta_0 \epsilon_t + \epsilon_{t-1} \\ \theta_a \epsilon_t + \epsilon_{t-1} \end{cases} \tag{4.13}$$

Again $\epsilon_t \sim \mathcal{N}(0, 1)$ normally distributed at all points $t \in [0, T]$, there are three cases to consider here

- **Case 1.** $t, t + k \leq \lfloor Tx \rfloor$

$$\text{Cov}(X_t, X_{t+k}) = \begin{cases} 1 + \theta_0^2 & k = 0 \\ \theta_0 & k = 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.14}$$

- **Case 2.** $t \leq 150$, and $t + k \geq \lfloor Tx \rfloor + 1$

$$\text{Cov}(X_t, X_{t+k}) = \text{Cov}(\theta_0 \epsilon_t + \epsilon_{t-1}, \theta_a \epsilon_{t+k})$$

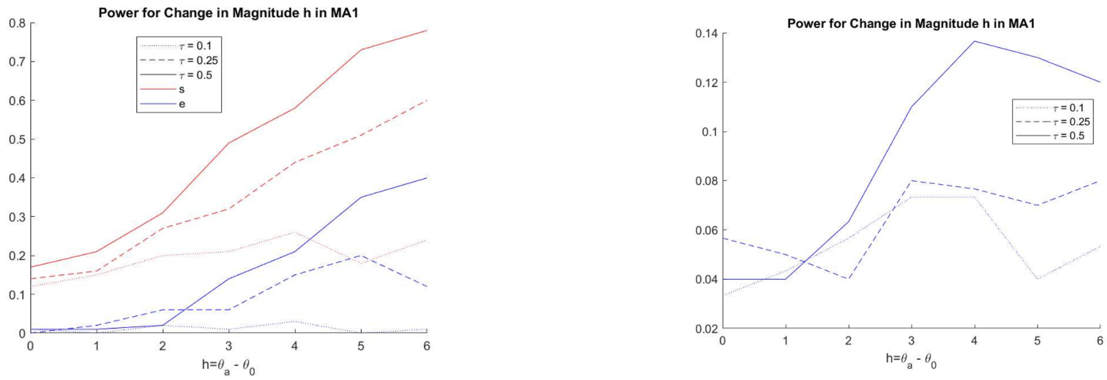


FIGURE 4.4: Power of R_T(left) and G_T(right) for different changes h, for $T = 300$

this is 0 in all cases, as k must be at least 1 for this example to happen, and this creates no overlap in the ϵ terms.

- **Case 3.** $t, t+k \geq \lfloor Tx \rfloor + 1$

$$\text{Cov}(X_t, X_{t+k}) = \begin{cases} 1 + \theta_a^2 & k = 0 \\ \theta_a & k = 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.15}$$

	$T = 100$			$T = 500$			$T = 2000$		
x	0.1	0.25	0.5	0.1	0.25	0.5	0.1	0.25	0.5
R_T^e	0.391	0.433	0.456	0.587	0.869	0.94			
R_T^s	0.666	0.781	0.835	0.752	0.966	0.997			
G_T	0.194	0.196	0.185	0.061	0.145	0.215	0.099	0.451	0.706
$B_T^{(1)}$	0	0	0	0	0	0	0	0	0
$B_T^{\prime(1)}$	0.456	0.267	0.084	0.5	0.047	0.018	0.154	0.006	0.935

TABLE 4.9: Power under MA1 with change $\theta_0 = 0.3$ to $\theta_a = 0.8$ at different change-points, for $T = \{100, 500, 2000\}$

Here R_T^s performs far better than R_T^e, this was hinted at by the better consistency of the p-values for R_T^s across different sample sizes for this model. G_T struggled greatly with this model, a slight change in the coefficient likely did not induce a large enough change in the turning rate for significant power.

Heavy-tail AR(1) An AR model with a ϵ_t t-distributed is looked at next.

$$X_t = \begin{cases} \phi_0 X_{t-1} + \epsilon_t & t \in [1, \lfloor Tx \rfloor] \\ \phi_a X_{t-1} + \epsilon_t & t \in [\lfloor Tx \rfloor + 1, T] \end{cases} \tag{4.16}$$

Where ϵ_t is $t-$distributed with degrees of freedom $df = 5$. The covariance structure of this time series also has three cases.

- **Case 1.** $t, t+k \leq \lfloor Tx \rfloor$

$$\begin{aligned}\text{Cov}(X_t, X_{t+k} =& \text{Cov}(\phi_0 X_{t+k-1} + \epsilon_{t+k-1})\\ &\vdots\\ =& \text{Cov}(X_t, \phi_0^k X_t + \sum_{i=0}^{k} \phi_0^i \epsilon_{t+k-i})\end{aligned}$$

there is no choice of i in the sum where ϵ_{t+k-i} coincides with something which gives a non-zero covariance, so the sum is 0.

$$\begin{aligned}\text{Cov}(X_t, X_{t+k}) =& \phi_0^k \operatorname{Var} X_t\\ =& \begin{cases}\infty \quad 1 < df \leq 2\\ \frac{\frac{df}{df-2}}{1-\phi_0^2} \quad k=0,\ df > 2\\ \frac{\phi_0^k(\frac{df}{df-2})}{1-\phi_0^2} \quad k \geq 1,\ df > 2\end{cases}\end{aligned}$$

- **Case 2.** $t \leq \lfloor Tx \rfloor$, and $t+k \geq \lfloor Tx \rfloor + 1$

$$\begin{aligned}Cov(X_t, X_{t+k}) &= Cov(X_t, \phi_a X_{t+k-1} + \epsilon_{t+k})\\ &= Cov(X_t, \phi_a(\phi_a X_{t+k-2} + \epsilon_{t+k-1}) + \epsilon_{t+k})\\ &\vdots\\ &= Cov(X_t, \phi_a^m X_{t+k-m} + \sum_{i=0}^{m} \phi_a^i \epsilon_{t+k-i})\end{aligned}$$

Where $m = \lfloor Tx \rfloor - k$, and there is no choice of i in the sum that gives non-zero covariance with X_t so it is 0 and can be disregarded.

$$\begin{aligned}Cov(X_t, X_{t+k}) &= \phi_a^m \operatorname{Cov}(X_t, X_{t+k-m})\\ &= \phi_a^m \operatorname{Cov}(X_t, \phi_0 X_{t+k-m-1} + \epsilon_{t+k-m-2})\\ &\vdots\\ &= \phi_a^m \operatorname{Cov}(X_t, \phi_0^{k-m} X_t + \sum_{i=0}^{k-m} \phi_0^i \epsilon_{t+k-m-i})\end{aligned}$$

Giving the final covariance function

$$\text{Cov}(X_t, X_{t+k}) = \begin{cases}\infty \quad 1|df \leq 2\\ \frac{\frac{df}{df-1}}{1-\phi_0^2} \quad k=0,\ df > 2\\ \phi_a^m \phi_0^{k-m} \quad k > 0,\ df > 2\end{cases}$$

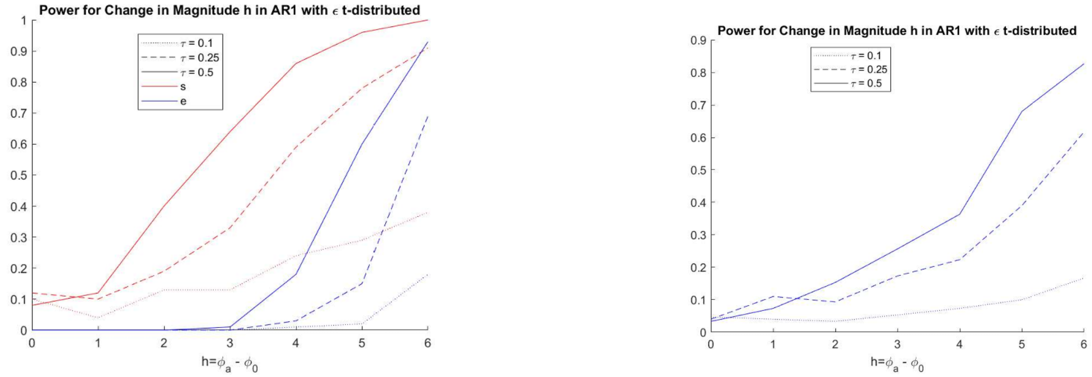


FIGURE 4.5: Power of R_T(left) and G_T(right) for different changes h

- **Case 3.** $t, t+k \geq \lfloor Tx \rfloor + 1$
  In the same manner as case 1

$$
\begin{aligned}
\mathrm{Cov}(X_t, X_{t+k} =& \mathrm{Cov}(\phi_a X_{t+k-1} + \epsilon_{t+k-1}) \\
&\vdots \\
=& \mathrm{Cov}(X_t, \phi_a^k X_t + \sum_{i=0}^{k} \phi_a^i \epsilon_{t+k-i}) \\
=& \phi_a^k \operatorname{Var} X_t \\
=& \begin{cases} \infty & 1 < df \leq 2 \\ \dfrac{\frac{df}{df-2}}{1-\phi_a^2} & k=0,\ df > 2 \\ \dfrac{\phi_a^k(\frac{df}{df-2})}{1-\phi_a^2} & k \geq 1,\ df > 2 \end{cases}
\end{aligned}
$$

	$T = 100$			$T = 500$			$T = 2000$		
x	0.1	0.25	0.5	0.1	0.25	0.5	0.1	0.25	0.5
R_T^e	0.913	0.91	0.861	1.0	1.0	1.0			
R_T^s	0.994	1.0	0.999	1.0	1.0	1.0			
G_T	0.179	0.22	0.246	0.067	0.229	0.364	0.143	0.707	0.927
$B_T^{(1)}$	0	0	0	0	0	0	0	0	0
$B_T^{\prime(1)}$	0.289	0.168	0.034	0.37	0.04	0.07	0.063	0.044	0.985

TABLE 4.10: Power under AR(1), with ϵ_i t-distributed with 5 degrees of freedom, with change $phi_0 = 0.3$ to $\phi_a = 0.8$ at different changepoints, for $T = \{100, 500, 2000\}$

Both G_T and R_T^s performed remarkably well for this model. R_T^s benefitted greatly from the normalisation done by the diagonals of $\hat{U}_{i,i}$, R_T^e failed to detect changes until $h > 0.3$, this is likely due to the extremely large critical values estimated for this case.
A broader experiment is conducted in tables 4.11 to 4.14. All possible changes in parameter in $\phi_0, \theta_0 \in \{-0.3, \ldots, 0.3\}$ and $\phi_a, \theta_a \in \{-0.5 \ldots, 0.5\}$ are examined for the three ARMA

schemes for sample sizes $T = 500$ and again for $T = 2000$. Tables for R_T were not made here due to the time constraints of the model.
These tables clarify some things about G_T and both Berkes tests B_T, B'_T. Firstly, any poor performance by G_T earlier can be attributed to small sample size. G_T demonstrates good power in table-4.11 and table-4.12 when $T = 2000$. Size close to 0.05 under the null-hypothesis can be seen in the diagonals of both tables too. G_T can also be run extremely quickly compared to R_T, making it feasible for datasets with large size.
Results for B_T and B'_T fail to match those in the Berkes paper[5], and continue to perform poorly in tables-4.13 and 4.14. B'_t shows good size when there is a change in sign in the coefficients, and almost never detects a change otherwise. B_T has good power in some cases when the coefficient changes from negative to positive, and when a large, positive θ_0 changes to a θ_a close to 0. The biggest indicator of the issue is the fact that power is almost 0 at every point where $\phi_0 = 0$. A poor estimation of $\gamma(r)$ would explain this discrepancy, and without knowing how it was estimated in the original paper this can not be improved.

$\phi_0 \backslash \phi_a$	-0.5	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5	
-0.3	10.3	6	5.6	7	14.5	27.7	45.4	66.1	79	86.5	92.5	
-0.2	21.1	12.1	6.3	4	6.9	15.5	30.8	50.7	62.3	77.7	86.3	
-0.1	34.8	24.8	16.1	7.2	3.9	7.2	13.4	32.5	45.2	62.6	70.9	
0.0	51.7	39.8	26.6	15	8.1	5.5	5.7	15	31.1	43.9	53.2	$T = 500$
0.1	72.5	58.9	45.9	30	14.9	6.9	4.1	6.7	15.8	25.8	34.7	
0.2	85.2	77	64.5	47.5	30.4	16.1	8.1	5.2	6.4	12.2	19.9	
0.3	93.4	88.1	80.1	63.2	47.6	29.3	13.6	6.5	4.3	7.3	10.1	
-0.3	31.1	14.2	3.5	16.4	48	80.1	96	99.6	100	100	100	
-0.2	66.9	40.7	15	4.8	17.2	54.4	83.9	97.3	99.3	99.8	100	
-0.1	90.8	75	48.8	16.8	5.4	19.6	52	85.7	95.4	99.1	99.9	
0.0	97.5	93.9	79	54.3	20.1	3.6	18.3	55.4	80.7	94	98.3	$T = 2000$
0.1	99.5	99.5	96.9	85.4	55.6	17	5	17	48	74.8	89.2	
0.2	100	99.9	98.8	96.7	84.4	52.9	16.5	4.9	14.7	42.1	65.9	
0.3	100	100	100	99	96.2	79.2	48.4	13.4	3.9	12.6	35.4	

TABLE 4.11: G_T for a MA(1) process with change at $x = 0.5$ and parameter change ϕ_0 to ϕ_a when $T = 500$ and $T = 2000$

4.6 Power when $p > 1$

In this section, a change in autocorrelation at lags greater than 1 is examined. Given that strong-mixing is assumed in this paper it is of particular interest to see if the location of the change within the autocovariance function affects power. Furthermore, it is of interest to see if R_T performs better if a change is distributed across multiple lags, say by changing the magnitude of the variance of ϵ in an ARMA scheme.
MA(r) models
The following moving average process will be tested here.

$$X_t = \begin{cases} \epsilon_t & t \in [1, \lfloor Tx \rfloor] \\ \theta\epsilon_{t-r} + \epsilon_t & t \in [\lfloor Tx \rfloor + 1, T] \end{cases} \tag{4.17}$$

$\phi_0 \backslash \phi_a$	-0.5	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5	
-0.3	20.6	8.3	4.1	7.9	16.6	36.1	58.7	75	87.6	94.6	98.2	
-0.2	36.8	20.1	8.9	5.4	8.7	19.6	36	56.9	77.2	87.8	94.7	
-0.1	57.7	35.3	17.9	7.9	4.5	8.9	19.5	38.9	56.7	76.2	88.8	
0.0	75.5	56.2	37.6	17.7	9.4	4	8.1	19.4	36.5	57	75.3	T=500
0.1	87.5	76.6	60	37.7	21.3	8.3	4.9	7	19.8	36	60.9	
0.2	95.3	88	73.6	59.3	39.5	18.4	8	4.3	7.3	18	36	
0.3	97	94.8	85.7	76.8	57.3	38	18.7	7.6	4.9	7.3	19.3	
-0.3	63.1	23.8	5.3	22.6	61.7	91	98.6	100	100	100	100	
-0.2	92	63	20.6	4.6	23.6	66	92.5	98.8	99.8	99.9	100	
-0.1	98.7	91.1	62.3	22.4	3.2	18.5	64.1	89.6	99.2	99.9	100	
0.0	99.9	98.6	90.9	61.7	21.7	4.4	25.8	64.6	93.2	98.5	100	T=2000
0.1	100	99.9	98.5	91.3	64	22.8	4.5	20	60.4	89	98.4	
0.2	100	100	99.6	99.1	91.6	61.9	22	4.2	20.2	62.9	89.4	
0.3	100	100	100	100	98.9	90.8	60.2	22.7	4.1	19.8	67.1	

TABLE 4.12: G_T for a AR(1) process, with ϵ_t t-distributed with 5 degrees of freedom, change at $x = 0.5$ and parameter change ϕ_0 to ϕ_a when $T = 500$ and $T = 2000$

Where $\epsilon \sim \mathcal{N}(0,1)$, and $r \in \mathbb{N}$. There are several cases for the covariance function $\gamma(l)$ here:

Case 1. $t, t+k < \lfloor Tx \rfloor$

$$\text{Cov}\left(X_t, X_{t+k}\right) = \text{Cov}\left(\epsilon_t, \epsilon_{t+k}\right) = \begin{cases} 1 & k = 0 \\ 0 & \text{otherwise} \end{cases}$$

Case 2. $t < \lfloor Tx \rfloor$, $t+k > \lfloor Tx \rfloor + 1$

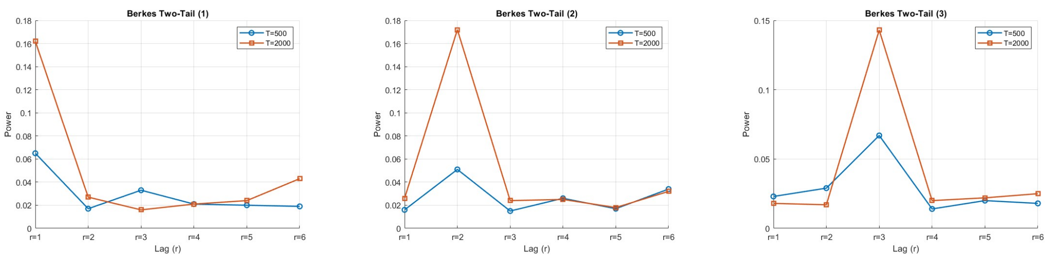
$$\begin{aligned} \text{Cov}\left(X_t, X_{t+k}\right) &= \text{Cov}\left(\epsilon_t, \epsilon_{t+k} + \theta\epsilon_{t+k-r}\right) \\ &= \text{Cov}\left(\epsilon_t, \epsilon_{t+k}\right) + \text{Cov}\left(\epsilon_t, \theta\epsilon_{t+k-r}\right) \\ &= \begin{cases} 1 & k = 0, r \neq 0 \\ \theta & |k| = r \neq 0 \\ 1+\theta & k = r = 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Case 3. $t, t+k > \lfloor Tx \rfloor + 1$

$$\begin{aligned} \text{Cov}\left(X_t, X_{t+k}\right) &= \text{Cov}\left(\epsilon_t + \theta\epsilon_{t-r}, \epsilon_{t+k} + \theta\epsilon_{t+k-r}\right) \\ &= \text{Cov}\left(\epsilon_t, \epsilon_{t+k}\right) + \text{Cov}\left(\epsilon_t, \theta\epsilon_{t+k-r}\right) + \text{Cov}\left(\theta\epsilon_{t-r}, \epsilon_{t+k}\right) + \text{Cov}\left(\theta\epsilon_{t-r}, \theta\epsilon_{t-r+k}\right) \\ &= \begin{cases} 1+\theta^2 & k = 0, r \neq 0 \\ \theta & |k| = r \neq 0 \\ 1+2\theta+\theta^2 & k = r = 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

$\phi_0 \backslash \phi_a$	-0.5	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5	
-0.3	7.6	9.6	5.3	10.4	51.7	92.5	99.6	100	100	100	100	
-0.2	3.2	6.1	5.1	5.4	17.5	70.2	98.6	100	100	100	100	
-0.1	0.1	2.3	3.9	4.2	6.0	25.6	78.5	97.2	99.9	100	100	
0.0	0	0	0.1	0.5	3.0	4.3	3.6	0.7	0	0	0	B'_T
0.1	100	100	99.7	97.6	77.8	26.9	4.3	3.8	4.1	1.2	0.3	
0.2	100	100	100	100	97.2	78.5	21.2	6.1	4.9	6.1	3.0	
0.3	100	100	100	100	99.7	93.7	57.5	10.5	5.5	8.6	7.5	
-0.3	0.5	3.4	8.5	11.8	22.8	31.0	35.2	48.9	59.5	70.1	75.7	
-0.2	0.1	0.2	2.0	4.1	10.8	21.1	33.9	53.8	77.5	85.0	89.9	
-0.1	0	0	0.3	0.6	4.8	20.9	52.3	83.8	94.6	96.9	97.5	
0.0	0	0	0	0.9	2.8	4.5	1.6	0.1	0	0	0	B_T
0.1	0	0	2.1	13.9	34.9	29.6	3.8	0.5	0	0	0	
0.2	0.3	2.6	17.7	48.0	83.1	85.1	39.7	4.6	0.1	0	0	
0.3	4.4	18.4	47.2	81.2	97.7	97.3	72.9	26.4	3.3	0.2	0	

TABLE 4.13: B_T and B'_T for a MA(1) process, change at $x = 0.5$ and parameter change ϕ_0 to ϕ_a.



(A) B'_T at lag 1 for 4.17

(B) B'_T at lag 2 for 4.17

(C) B'_T at lag 3 for 4.17

FIGURE 4.6: B'_T at lags$\in \{1, 2, 3\}$ over model-4.17

This time series will be examined for R_T with several choices of maximum lag p to be tested. $\theta = 0.8$ will be chosen first, and power will be estimated $r \in \{1, \ldots, 6\}$.
This will be the only tests done for R_T, due to it's hefty runtime once $p > 1$. But θ will be varied for $B_T^{(r)}$, $B_T^{(r)'}$, and G_T since they can all be run in adequate time. Results for B'_T are found in figure 4.7, B_T are found in figures, and G_T is in

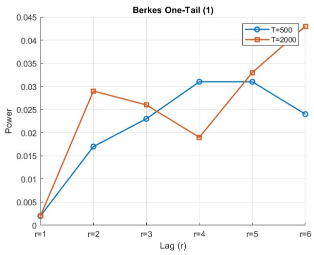
AR(r) **models** A similar auto-regressive model will be studied

$$X_t = \begin{cases} \epsilon_t & t \in [1, \lfloor Tx \rfloor] \\ \phi X_{t-r} + \epsilon_t & t \in [\lfloor Tx \rfloor + 1, T] \end{cases} \tag{4.18}$$

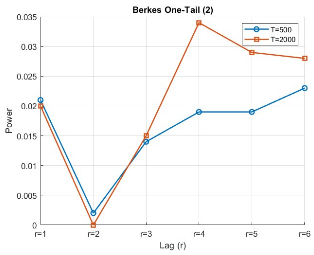
The auto-covariance function is determined by 3 cases:

Case 1. $t, t+k \leq \lfloor Tx \rfloor$

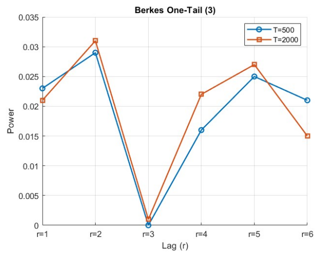
$$\text{Cov}\left(X_t, X_{t+k}\right) = \text{Cov}\left(\epsilon_t, \epsilon_{t+k}\right) = \begin{cases} 1, & k = 0 \\ 0, & k > 0 \end{cases}$$



(A) B_T at lag 1 for 4.17



(B) B_T at lag 2 for 4.17



(C) B_T at lag 3 for 4.17

FIGURE 4.7: B_T at lags$\in \{1, 2, 3\}$ over model-4.17

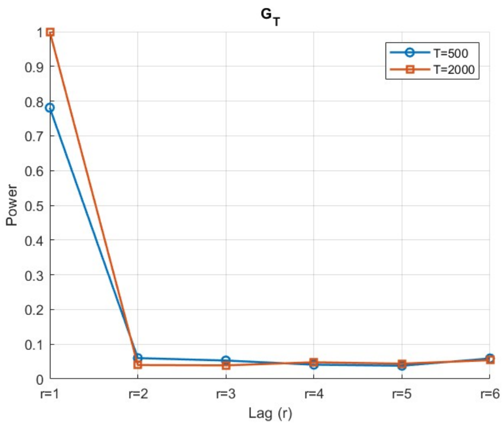


FIGURE 4.8: G_T over model-4.17

$\phi_0 \backslash \phi_a$	-0.5	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5	
-0.3	7.2	4.8	4.0	23.5	70.9	96.7	100	100	100	100	100	
-0.2	1.3	3.5	5.6	5.2	24.8	70.9	97.4	99.6	100	100	100	
-0.1	0.9	0.6	2.9	5.0	6.2	22.9	77.0	97.6	99.9	100	100	
0.0	0	0	0.1	0.9	3.5	3.4	4.1	0.8	0.1	0	0	$B_T^{'}$
0.1	100	100	100	96.2	74.2	21.3	4.9	3.5	2.4	0.7	0.2	
0.2	100	100	100	99.9	98.5	79.5	19.9	4.3	4.5	3.9	1.0	
0.3	100	100	100	100	99.8	96.6	66.5	28.1	6.1	5.4	6.5	
-0.3	0.1	1.4	3.4	11.0	21.0	32.2	41.6	50.6	64.0	71.5	86.0	
-0.2	0	0	0.6	7.0	12.5	24.3	39.0	58.5	79.4	90.6	96.8	
-0.1	0	0	0.2	0.4	6.6	18.1	51.8	84.7	93.6	97.0	99.6	
0.0	0	0	0	0.8	4.7	4.5	2.3	0.3	0	0	0	B_T
0.1	0	0	1.3	13.0	36.3	26.2	4.9	0.2	0	0	0	
0.2	0	2.4	14.1	46.0	85.4	88.3	42.4	5.4	0.3	0.1	0	
0.3	1.4	16.4	45.3	81.2	98.1	99.1	83.2	41.8	8.7	0.4	0	

TABLE 4.14: B_T and $B_T^{'}$ for a AR(1) process, with ϵ_t t-distributed, change at $x = 0.5$ and parameter change ϕ_0 to ϕ_a .

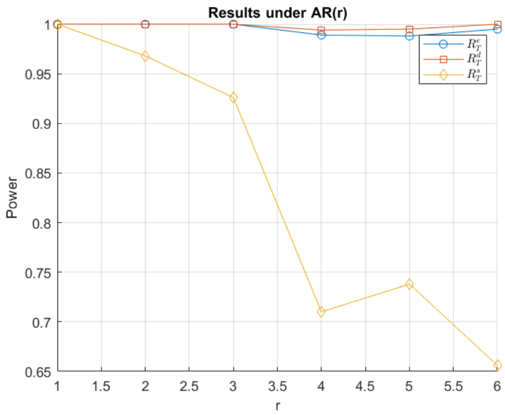


FIGURE 4.9: R_T for AR(r) model, with change at $x = 0.5$

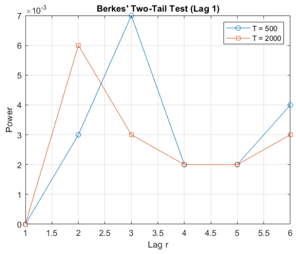
Case 2. $t \leq \lfloor Tx \rfloor$, $t + k \geq \lfloor Tx \rfloor + 1$

$$\begin{aligned} \text{Cov}\left(X_t, X_{t+k}\right) &= \phi^m \, \text{Cov}\left(\epsilon_t, \epsilon_{t+k-mr}\right) \quad \text{for } m = \lfloor \frac{k}{r} \rfloor \\ &= \begin{cases} \phi^m, & t + k - mr = t \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

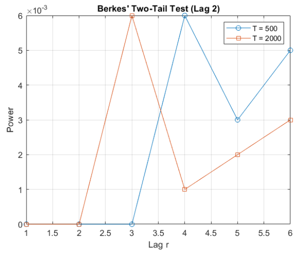
Case 3. $t, t + k \geq \lfloor Tx \rfloor + 1$

$$\text{Cov}\left(X_t, X_{t+k}\right) = \phi^{\lfloor k/r \rfloor}$$

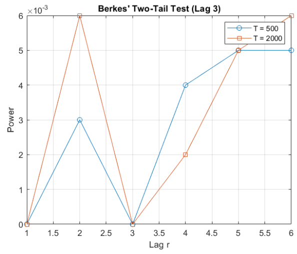
Results for power are found in tables -



(A) B'_T at lag 1 for 4.18

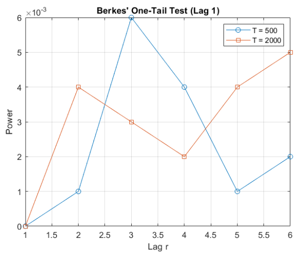


(B) B'_T at lag 2 for 4.18

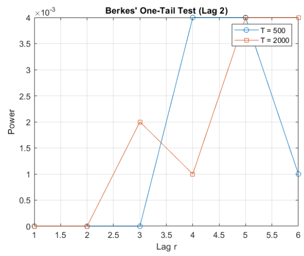


(C) B'_T at lag 3 for 4.18

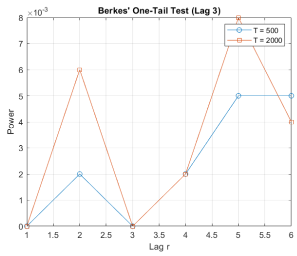
FIGURE 4.10: B'_T at lags$\in \{1, 2, 3\}$ over model-4.17



(A) B_T at lag 1 for 4.18



(B) B_T at lag 2 for 4.18



(C) B_T at lag 3 for 4.18

FIGURE 4.11: B_T at lags$\in \{1, 2, 3\}$ over model-4.18

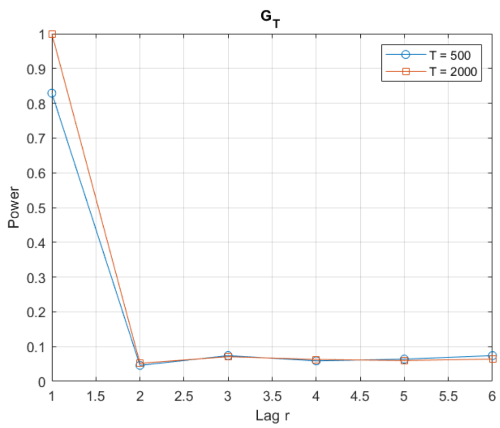


FIGURE 4.12: G_T over model-4.18

Discussion

R_T has been shown to perform quite well over a number of models. Furthermore, it performs better than any of the other tests for small sample sizes ($T \simeq 500$.) The two biggest issues that R_T faces are its poor runtime, and poor size under the null-hypothesis. There may be room to improve this runtime, a better PC would be the most obvious. The bottleneck is mostly due to the sheer number of partial sums that are taken in this test, so improving the runtime significantly is unlikely.
The size under the null-hypothesis is only improved when the diagonals of the matrix W were taken to be the diagonals of $\hat{U}^{-1}$. Results could be further improved if $\hat{U}$ were estimated using a sample up to only the point k, i.e. giving the test the following form

$$R_T = \max_{k\in[1,\tilde{T}]} \left\| \boldsymbol{S}_k^p - \frac{k}{\tilde{T}} \boldsymbol{S}_{\tilde{T}}^p \right\|_{W_k}$$

$$\text{Where, } W_k = \begin{pmatrix} w_{k,o} & 0 & \dots & 0 \\ 0 & w_{k,1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & w_{k,p} \end{pmatrix}, \quad w_k, i = \hat{U}_{i,i}^k$$

$$\text{and } \hat{U}_{i,i}^k = \sum_{t=1}^{k} \sum_{s=1}^{k} \left(\hat{Y}_i \hat{Y}_{i+s} - \frac{1}{k} S_k^i \right) \left(\hat{Y}_i \hat{Y}_{i+t} - \frac{1}{k} S_k^i \right) k \left(\frac{|s-t|}{b_{\tilde{T}}} \right).$$

This type of standardization has precedent, Berkes et al [5] did not outline it explicitly but stated that they estimated $\gamma^2(r)$ up until a known changepoint. G_T likely performs so well because it does this too, but would likely be infeasible until the runtime is addressed.
The power of R_T under the null-hypothesis was shown in this thesis in more detail than exists in the literature. What was a 6 line proof in Dürre & Fried[14] has been extended to a proposition and a theorem showing all details needed.
The Power of R_T under the alternative was beyond the scope of this thesis, and would be a good point to continue this work. Dürre & Fried themselves note that this would require a lot of care because of the use of a non-linear function ϕ, and the standardization used with it too.[14].
Another important detail of R_T is its strong-mixing requirement. It is difficult to verify if data is, indeed, strong mixing, and most authors do not demonstrate the strong-mixing property of their data. A more thorough method for examining if a series is strong-mixing, and just how strong this is, would be a good addition to the work done here.
B_T failed to match its performance in the literature, and as stated this is likely due to the estimator $\hat{\gamma}^2(r)$. Improving this with a better estimator was not possible in the time remaining in this thesis.
G_T outperformed every other test for sufficiently high samples. For higher samples it

seems like the best choice, and there is a lot of room to continue work on this test. It may be possible to detect changes beyond the first lag using G_T, though these changes in turning rate are very fine and might need a different treatment before applying the test.

Conclusion

In this work the modified CUSUM-test R_T is introduced, some results proven, and demonstrated through implementation and simulations. The power of the test under the null-hypothesis with strong-mixing conditions is proven in greater detail than exists in the literature currently. This result is verified through simulations, comparing the test the similar CUSUM-tests B_T and B'_T, and the ordinal-pattern based test G_T.
The Power of R_T was shown to be good under certain choices of weight-matrix W, and with smaller sample sizes than the other tests. It's runtime, especially when testing changes in the autocovariance greater than 1 is a huge deficit to the test and has little room for improvement outside of a stronger computer. The size of the test was shown to be poor for constant weights R^e_T, and seemed prone to giving false positives. R^s_T was the best version of the test under most circumstances, especially moving average processes. The test additionally suffers from its flexibility, as critical values appear to grow unstable when attempting to test many lags.
B_T and B'_T were not implemented in a way that matched performance in the paper they were taken from, and failed to give good results across all examples tested. This can be attributed to poor standardization, and only B'_T gave results that could compete with the other tests in this paper.
G_T outperformed R_T, and demonstrated better size under the null-hypothesis, all with a much faster runtime. Until this test is modified to possibly detect changes in autocovariance greater than 1 R_T seems to be the best choice for detecting structural changes in autocovariance at high lags.

Acknowledgments

Firstly, and most importantly, I must thank my supervisors Annika and Giorgio. Annika for never failing to put me at ease and on the right path to progress, and Giorgio for our frequent (and often unscheduled) conversations to resolve issues. I don't think anybody else would have been so okay with their master students moving country during their thesis, but they have both been incredibly understanding at every step in this journey.
I would like to thank my parents for pretending to understand what I was doing, and for their unwavering support. My partner, Charlene, deserves special thanks for keeping me fed and watered, and for her unlimited kindness during this thesis.
I thank all of my friends in Enschede, in particular Jorg, Lavinia, and Eva who gave me a home away from home.
I must finally thank Paul for proofreading my thesis so many times, my housemate Caróg for the walks home from the library, and Sam and Stewart for all of the coffee breaks.

Bibliography

[1] Bovas Abraham and William Wei. Inferences about the parameters of a time series model with changing variance. *Metrika*, 31:183–194, 02 1984.

[2] Alexander Aue, Siegfried Hörmann, Lajos Horváth, and Matthew Reimherr. Break detection in the covariance structure of multivariate time series models. 2009.

[3] Anindya Banerjee and Giovanni Urga. Modelling structural breaks, long memory and stock market volatility: an overview. *Journal of Econometrics*, 129(1-2):1–34, 2005.

[4] P Baufays and JP Rasson. Variance changes in autoregressive models. *Time series analysis: theory and practice*, 7:119–127, 1985.

[5] István Berkes, Edit Gombay, and Lajos Horváth. Testing for changes in the covariance structure of linear processes. *Journal of Statistical Planning and Inference*, 139(6):2044–2063, 2009.

[6] Annika Betken, Giorgio Micali, and Johannes Schmidt-Hieber. Ordinal patterns based change points detection. *arXiv preprint arXiv:2502.03099*, 2025.

[7] Patrick Billingsley. *Convergence of Probability Measures.* Wiley, 7 1999.

[8] Richard C. Bradley. Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys*, 2(none), January 2005.

[9] K. C. Chanda. Strong mixing properties of linear stochastic processes. *Journal of Applied Probability*, 11(2):401–408, 1974.

[10] Jie Chen and Arjun K Gupta. Testing and locating variance changepoints with application to stock prices. *Journal of the American Statistical association*, 92(438):739–747, 1997.

[11] Miklós Csörgö and Lajos Horváth. Limit theorems in change-point analysis. *(No Title)*, 1997.

[12] Richard A Davis, Thomas C M Lee, and Gabriel A Rodriguez-Yam. Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association*, 101(473):223–239, 2006.

[13] Yu A Davydov. Convergence of distributions generated by stationary stochastic processes. *Theory of Probability & Its Applications*, 13(4):691–696, 1968.

[14] Alexander Dürre and Roland Fried. Robust change point tests by bounded transformations. 5 2019.

[15] Pedro Galeano and Daniel Peña. Covariance changes detection in multivariate time series. *Journal of Statistical Planning and Inference*, 137(1):194–211, 2007.

[16] Edit Gombay. Change detection in autoregressive time series. *Journal of Multivariate Analysis*, 99(3):451–464, 2008.

[17] Bruce Hansen. A weak law of large numbers under weak mixing. 2019.

[18] Marie Hušková, Claudia Kirch, Zuzana Prášková, and Josef Steinebach. On the detection of changes in autoregressive time series, ii. resampling procedures. *Journal of Statistical Planning and Inference*, 138(6):1697–1721, 2008.

[19] Marie Hušková, Zuzana Prášková, and Josef Steinebach. On the detection of changes in autoregressive time series i. asymptotics. *Journal of Statistical Planning and Inference*, 137(4):1243–1259, 2007.

[20] Carla Inclán and George C. Tiao. Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association*, 89(427):913–923, 1994.

[21] J. Jacod and P. Protter. *Probability Essentials*. Universitext. Springer Berlin Heidelberg, 2004.

[22] Venkata Jandhyala, Stergios Fotopoulos, Ian MacNeill, and Pengyu Liu. Inference for single and multiple change-points in time series. *Journal of Time Series Analysis*, 34(4):423–446, 2013.

[23] G. M. Jenkins. General considerations in the analysis of spectra. *Technometrics*, 3(2):133–166, 1961.

[24] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113. Springer New York, 1998.

[25] R. Killick, I. A. Eckley, and P. Jonathan. A wavelet-based approach for detecting changes in second order structure within nonstationary time series. *Electronic Journal of Statistics*, 7(none):1167 – 1183, 2013.

[26] Jae-Young Kim. Detection of change in persistence of a linear time series. *Journal of Econometrics*, 95(1):97–116, 2000.

[27] Sangyeol Lee and Siyun Park. The cusum of squares test for scale changes in infinite order moving average processes. *Scandinavian Journal of Statistics*, 28:625–644, 12 2001.

[28] Florence Merlevede and Magda Peligrad. The functional central limit theorem under the strong mixing condition. *The Annals of Probability*, 28(3):1336–1352, 2000.

[29] Jukka Nyblom. Testing for the constancy of parameters over time. *Journal of the American Statistical Association*, 84(405):223–230, 1989.

[30] E. S. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.

[31] Pierre Perron et al. Dealing with structural breaks. *Palgrave handbook of econometrics*, 1(2):278–352, 2006.

[32] Dominique Picard. Testing and estimating change-points in time series. *Advances in Applied Probability*, 17(4):841–867, 1985.

[33] Philip Preuß, Ruprecht Puchstein, and Holger Dette. Detection of multiple structural breaks in multivariate time series, 2013.

[34] M. H. Quenouille. The comparison of correlations in time-series. *Journal of the Royal Statistical Society. Series B (Methodological)*, 20(1):158–164, 1958.

[35] Murray Rosenblatt. Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics*, 27(3):832 – 837, 1956.

[36] Mathieu Sinn and Karsten Keller. Estimation of ordinal pattern probabilities in gaussian processes with stationary increments. *Computational Statistics Data Analysis*, 55(4):1781–1790, 2011.

[37] E. Slutsky. Über stochastische Asymptoten und Grenzwerte. 1925.

[38] S. M. Tang and I. B. MacNeill. The effect of serial correlation on tests for parameter change at unknown time. *The Annals of Statistics*, 21(1):552–575, 1993.

[39] Dürre Fried. Robust test for detecting changes in the autocovariance function of a time series. *Austrian Journal of Statistics*, 49(4):35–45, 2020.

[40] Ruey S Tsay. Outliers, level shifts, and variance changes in time series. *Journal of forecasting*, 7(1):1–20, 1988.

[41] Dean W. Wichern, Robert B. Miller, and Der-Ann Hsu. Changes of variance in first-order autoregressive time series models-with an application. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 25(3):248–256, 1976.

[42] C. S. Withers. Conditions for linear processes to be strong-mixing. *Zeitschrift f r Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57:477–480, 12 1981.

[43] Jeffrey M. Wooldridge and Halbert White. Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory*, 4(2):210–230, 1988.

[44] Ken-ichi Yoshihara. The bahadur representation of sample quantiles for sequences of strongly mixing random variables. *Statistics & probability letters*, 24(4):299–304, 1995.

Appendix

Lemma 1 *For a strongly mixing sequence $(X_i)_{i\in\mathbb{Z}}$ with mixing coefficients α_n and Borel-measurable function $f : (X, \mathcal{F}_X) \to (Y, \mathcal{F}_y)$, $(f(X_i))_{i\in\mathbb{Z}}$ is also strongly mixing with coefficients $\alpha(f(X_i)) \leq \alpha(X_i)$.*

Proof. $f : (X, \mathcal{F}_X) \to (Y, \mathcal{F}_Y)$ is Borel-measurable if the preimage $f^{-1}(V)$ of any open set $V \subset Y$ is measurable in $\mathcal{F}_Y$, where $\mathcal{F}_Y$ is the smallest sigma-field generated by all open sets in the target space Y.
This ensures that applying f to a process $(X_i)_{i\in\mathbb{Z}}$ preserves measurability with respect to the original sigma-field. Specifically, for every $I \subset \mathbb{Z}$, we have:

$$\sigma(f(X_i), i \in I) \subset \sigma(X_i, i \in I) \tag{5.1}$$

In other words, the sigma-field generated by $(f(X_i))_{i\in I}$ is a sub-sigma-field of that generated by $(X_i)_{i\in I}$.
To show why this implies strong-mixing of $(f(X_i))_{i\in I}$ note that by applying f to any sets $A \in \mathcal{F}^0_{-\infty}$ and $B \in \mathcal{F}^\infty_n$ and taking the supremum gives the following inequality

$$\sup_{A' \in f(\mathcal{F}^0_{-\infty}),\, B' \in f(\mathcal{F}^\infty_n)} \|P(A' \cap B') - P(A')P(B')\| \leq \sup_{A \in \mathcal{F}^0_{-\infty},\, B \in \mathcal{F}^\infty_n} \|P(A \cap B) - P(A)P(B)\| = \alpha(X_n) \tag{5.2}$$

As the sets $f(\mathcal{F}^0_{-\infty})$ and $f(\mathcal{F}^\infty_n)$ at most retain the element of $\mathcal{F}^0_{-\infty}$ and $\mathcal{F}^\infty_n$ that results in the supremum on the right, and otherwise create a smaller expression.
Since $\alpha(f(X_i))$ is therefore bounded above by $\alpha(X_i)$ and below by 0 the squeeze theorem gives the mixing coefficients $\alpha(f(X_i)) \to 0$ as $i \to \infty$ □

Lemma 4. *$\hat{U}$ with estimated standardization $\hat{Y}_i = \phi\left(\frac{X_i - \hat{\mu}}{\sigma}\right)$ is consistent for U*

$$\hat{U} = \frac{1}{\tilde{T}} \sum_{s=1}^{\tilde{T}} \sum_{t=1}^{\tilde{T}} \left(\hat{Y}_s \hat{Y}_{s+i} - \frac{1}{\tilde{T}} S^{(i)}_{\tilde{T}}\right) \left(\hat{Y}_t \hat{Y}_{t+j} - \frac{1}{\tilde{T}} S^{(j)}_{\tilde{T}}\right) k\left(\frac{s-t}{b_T}\right)$$

with kernel function $k(x)$, $\tilde{T} = T - p$ and $S^{(p)}_{\tilde{T}} = \sum_{i=1}^{\tilde{T}} \hat{Y}_i \hat{Y}_{i+p}$

Proof. Let $S'^{(i)}_{\tilde{T}} = \sum_{k=1}^{\tilde{T}} Y_k Y_{k+i}$ be the series gotten from the standardized series $Y_i = \phi\left(\frac{X_i - \mu}{\sigma}\right)$ with true μ and σ. Without loss of generality, take the difference

$$\frac{1}{\tilde{T}} \sum_{s=1}^{\tilde{T}} \sum_{t=1}^{\tilde{T}} \left(\hat{Y}_s \hat{Y}_{s+i} - \frac{1}{\tilde{T}} S^{(i)}_{\tilde{T}}\right) \left(\hat{Y}_t \hat{Y}_{t+j} - \frac{1}{\tilde{T}} S^{(j)}_{\tilde{T}}\right) k\left(\frac{s-t}{b_T}\right) - \frac{1}{\tilde{T}} \sum_{s=1}^{\tilde{T}} \sum_{t=1}^{\tilde{T}} (Y_s Y_{s+i} - \gamma(i)) (Y_t Y_{t+j} - \gamma(j)) k\left(\frac{s-t}{b_T}\right) \tag{5.3}$$

and perform the following decomposition

$$
\begin{aligned}
(5.3) =& \frac{1}{\tilde{T}} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\hat{Y}_s \hat{Y}_{s+i} - \frac{1}{\tilde{T}} S_{\tilde{T}}^{(j)} \right) \left(\hat{Y}_t \hat{Y}_{t+j} - Y_t Y_{t+j} - \frac{1}{\tilde{T}} \left(S_{\tilde{T}}^{(j)} - S_{\tilde{T}}^{'(i)} \right) \right) \\
&+ \frac{1}{\tilde{T}} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\hat{Y}_s \hat{Y}_{s+i} - \frac{1}{\tilde{T}} S_{\tilde{T}}^{(j)} \right) \left(\hat{Y}_s Y_{s+i} - Y_s Y_{s+i} - \frac{1}{\tilde{T}} \left(S_{\tilde{T}}^{(i)} - S_{\tilde{T}}^{'(i)} \right) \right) \\
&+ \frac{1}{\tilde{T}} \sum_{t=1}^{T} \sum_{s=1}^{\tilde{T}} \left(Y_s Y_{s+i} - \gamma(i) \right) \left(\gamma(j) - \frac{1}{\tilde{T}} S_{\tilde{T}}^{'(j)} \right) \\
&+ \frac{1}{\tilde{T}} \sum_{t=1}^{\tilde{T}} \sum_{s=1}^{\tilde{T}} \left(Y_t Y_{t+j} - \gamma(j) \right) \left(\gamma(i) - \frac{1}{\tilde{T}} S_{\tilde{T}}^{'(j)} \right) \\
=& K_1 + K_2 + K_3 + K_4
\end{aligned}
$$

Denote these four summands by K_1, K_2, K_3, and K_4 respectively. Before going any further note the second order Taylor decomposition:

$$
\begin{aligned}
\phi\left(\frac{X-\mu}{\sigma}\right) =& \phi\left(X + \left[\frac{1}{\sigma} - 1\right] X - \frac{\mu}{\sigma}\right) \\
=& \phi(X) + \phi'(X)\left(\left[\frac{1}{\sigma} - 1\right] X - \frac{\mu}{\sigma}\right) + \frac{1}{2}\phi''(X)\left(\left[\frac{1}{\sigma} - 1\right] X - \frac{\mu}{\sigma}\right)^2 + \mathrm{R}(X)
\end{aligned}
\tag{5.4}
$$

Where the remainder $\mathrm{R}(T)$ depends on all higher order derived of $\phi(x)$. Care is required here, as ϕ is not differentiable everywhere. If taken piecewise the derivative becomes:

$$
\phi'(x) = \begin{cases} 1 \quad \text{if } |x| < k \\ \text{undefined if } |x| = k \\ 0 \quad \text{otherwise} \end{cases}
$$

$\phi'(x)$ has a discontinuity at $x = \{-k, k\}$, but the random variable X such that $\frac{X-\mu}{\sigma} = k$ occurs with probability 0 as it two single points. So this is piecewise, almost-surely continuous . Similarly the second derivative is the a.s. continuous function

$$
\phi''(X) = \begin{cases} \text{undefined if } |x| = k \\ 0 \text{ otherwise} \end{cases}
$$

At any point $|X| \geq k$ gives $\phi'(X) = 0$, so (5.3) is 0 at these points and these terms disappear. Furthermore, since the remainder $\mathrm{R}(X)$ depends on higher order derivates of $\phi(x)$ it has $\mathrm{R}(X) \overset{\text{a.s}}{=} 0$ for all X. So convergence need only be checked for the set $\{X \in \mathbb{X} : \quad |X| < k\}$. Furthermore $\phi''(X)$ is 0 everywhere, and the final component of (5.4) can be disregarded.
Apply this taylor expansion, starting with K_1. Looking first at product of the first two

terms of the second bracket gives

$$\begin{aligned}\hat{Y}_t\hat{Y}_{t+j} - Y_tY_{t+j} = &\left(\phi(X_t) - \left[\frac{1}{\hat{\sigma}} - 1\right] X_t - \frac{\hat{\mu}}{\hat{\sigma}}\right) \left(\phi(X_{t+j}) + \left[\frac{1}{\hat{\sigma}} - 1\right] X_{t+j} - \frac{\hat{\mu}}{\hat{\sigma}}\right) \\
&- \left(\phi(X_t) - \left[\frac{1}{\sigma} - 1\right] X_t - \frac{\mu}{\sigma}\right) \left(\phi(X_{t+j}) + \left[\frac{1}{\sigma} - 1\right] X_{t+j} - \frac{\mu}{\sigma}\right) \\
=&\phi(X_t)\phi(X_{t+j}) - \phi(X_t)\phi(X_{t+j}) \\
&+\phi(X_t)\left(\left(\frac{1}{\hat{\sigma}} - \frac{1}{\sigma}\right) X_t - \frac{\hat{\mu}}{\hat{\sigma}} + \frac{\mu}{\sigma}\right) \\
&+\phi(X_{t+j})\left(\left(\frac{1}{\hat{\sigma}} - \frac{1}{\sigma}\right) X_{t+j} - \frac{\hat{\mu}}{\hat{\sigma}} + \frac{\mu}{\sigma}\right) \\
&+X_tX_{t+j}\left(\left(\frac{1}{\hat{\sigma}} - 1\right)^2 - \left(\frac{1}{\sigma} - 1\right)^2\right) \\
&+ (X_t + X_{t+j})\left(\left(\frac{1}{\sigma} - 1\right)\frac{\mu}{\sigma} - \left(\frac{1}{\hat{\sigma}} - 1\right)\frac{\hat{\mu}}{\hat{\sigma}}\right) \\
&+ \left(\frac{\hat{\mu}}{\hat{\sigma}}\right)^2 - \left(\frac{\mu}{\sigma}\right)^2\end{aligned}$$

Thanks to consistency of $\hat{\mu}$ and $\hat{\sigma}$, and an application of the continuous mapping theorem, every difference involving an estimated $\hat{\mu}$ and $\hat{\sigma}$ and their corresponding μ and σ converge in probability to 0. Looking at the Taylor-expansion applied to the remaining terms in K_1's second bracket:

$$\begin{aligned}\frac{1}{\tilde{T}}\left(S_{\tilde{T}}^{(j)} - S_{\tilde{T}}^{\prime(j)}\right) =&\frac{1}{\tilde{T}}\left(\sum_{l=1}^{\tilde{T}} \hat{Y}_l\widehat{Y_{l+j}} - \sum_{l=1}^{\tilde{T}} Y_lY_{l+j}\right) \\
=&\frac{1}{\tilde{T}}\left(\sum_{l=1}^{\tilde{T}}\left(\hat{Y}_l\hat{Y}_{l+j} - Y_lY_{l+j}\right)\right)\end{aligned}$$

Applying the same Taylor expansion to the values inside the sum gives a similar expression

$$\begin{aligned}\frac{1}{\tilde{T}}\sum_{l=1}^{\tilde{T}}\left(\hat{Y}_l\hat{Y}_{l+j} - Y_lY_{l+j}\right) =&\frac{1}{\tilde{T}}\sum_{l=1}^{\tilde{T}}\left[\left(\phi(X_l) - \left(\frac{1}{\hat{\sigma}} - 1\right) X_l - \frac{\hat{\mu}}{\hat{\sigma}}\right)\left(\phi(X_{l+j}) + \left(\frac{1}{\hat{\sigma}} - 1\right) X_{l+j} - \frac{\hat{\mu}}{\hat{\sigma}}\right)\right. \\
&\left. - \left(\phi(X_l) - \left(\frac{1}{\sigma} - 1\right) X_l - \frac{\mu}{\sigma}\right)\left(\phi(X_{l+j}) + \left(\frac{1}{\sigma} - 1\right) X_{l+j} - \frac{\mu}{\sigma}\right)\right] \\
=&\frac{1}{\tilde{T}}\sum_{l=1}^{\tilde{T}}\left[\phi(X_l)\phi(X_{l+j}) - \phi(X_l)\phi(X_{l+j})\right. \\
&+ \phi(X_l)\left(\left(\frac{1}{\hat{\sigma}} - \frac{1}{\sigma}\right) X_l - \frac{\hat{\mu}}{\hat{\sigma}} + \frac{\mu}{\sigma}\right) \\
&+ \phi(X_{l+j})\left(\left(\frac{1}{\hat{\sigma}} - \frac{1}{\sigma}\right) X_{l+j} - \frac{\hat{\mu}}{\hat{\sigma}} + \frac{\mu}{\sigma}\right) \\
&+ X_lX_{l+j}\left(\left(\frac{1}{\hat{\sigma}} - 1\right)^2 - \left(\frac{1}{\sigma} - 1\right)^2\right) \\
&+ (X_l + X_{l+j})\left(\left(\frac{1}{\sigma} - 1\right)\frac{\mu}{\sigma} - \left(\frac{1}{\hat{\sigma}} - 1\right)\frac{\hat{\mu}}{\hat{\sigma}}\right) \\
&\left.+ \left(\frac{\hat{\mu}}{\hat{\sigma}}\right)^2 - \left(\frac{\mu}{\sigma}\right)^2\right]. \qquad (5.5)\end{aligned}$$

where again, by the assumption that μ and σ are consistent, every difference that appears here converges to 0 in probability. So the expression K_1 converges in probability, K_2 does so in the same way. Again breaking it's second bracket into the first two and last two terms.

$$\begin{aligned}
\hat{Y}_s\hat{Y}_{s+i} - Y_tY_{t+j} = &\left(\phi(X_s) - \left[\frac{1}{\hat{\sigma}} - 1\right] X_s - \frac{\hat{\mu}}{\hat{\sigma}}\right)\left(\phi(X_{s+i}) + \left[\frac{1}{\hat{\sigma}} - 1\right] X_{s+i} - \frac{\hat{\mu}}{\hat{\sigma}}\right) \\
&- \left(\phi(X_s) - \left[\frac{1}{\sigma} - 1\right] X_s - \frac{\mu}{\sigma}\right)\left(\phi(X_{s+i}) + \left[\frac{1}{\sigma} - 1\right] X_{s+i} - \frac{\mu}{\sigma}\right) \\
=&\phi(X_s)\phi(X_{s+i}) - \phi(X_s)\phi(X_{s+i}) \\
&+\phi(X_s)\left(\left(\frac{1}{\hat{\sigma}} - \frac{1}{\sigma}\right) X_s - \frac{\hat{\mu}}{\hat{\sigma}} + \frac{\mu}{\sigma}\right) \\
&+\phi(X_{s+i})\left(\left(\frac{1}{\hat{\sigma}} - \frac{1}{\sigma}\right) X_{s+i} - \frac{\hat{\mu}}{\hat{\sigma}} + \frac{\mu}{\sigma}\right) \\
&+X_sX_{s+i}\left(\left(\frac{1}{\hat{\sigma}} - 1\right)^2 - \left(\frac{1}{\sigma} - 1\right)^2\right) \\
&+ (X_s + X_{s+i})\left(\left(\frac{1}{\sigma} - 1\right)\frac{\mu}{\sigma} - \left(\frac{1}{\hat{\sigma}} - 1\right)\frac{\hat{\mu}}{\hat{\sigma}}\right) \\
&+ \left(\frac{\hat{\mu}}{\hat{\sigma}}\right)^2 - \left(\frac{\mu}{\sigma}\right)^2
\end{aligned}$$

Which converges to 0 in probability, the last two terms are also the same

$$\begin{aligned}
\frac{1}{\tilde{T}}\left(S^i_{\tilde{T}} - S'^{(i)}_{\tilde{T}}\right) =&\frac{1}{\tilde{T}}\left(\sum_{s=1}^{\tilde{T}} \hat{Y}_s\hat{Y}_{s+i} - \sum_{s=1}^{\tilde{T}} Y_sY_{s+i}\right) \\
=&\frac{1}{\tilde{T}}\sum_{s=1}^{\tilde{T}}\left(\hat{Y}_s\hat{Y}_{s+i} - Y_sY_{s+i}\right) \\
=&\frac{1}{\tilde{T}}\sum_{s=1}^{\tilde{T}}\Bigg[\left(\phi(X_s) - \left(\frac{1}{\hat{\sigma}} - 1\right) X_s - \frac{\hat{\mu}}{\hat{\sigma}}\right)\left(\phi(X_{s+i}) + \left(\frac{1}{\hat{\sigma}} - 1\right) X_{s+i} - \frac{\hat{\mu}}{\hat{\sigma}}\right) \\
&\quad - \left(\phi(X_s) - \left(\frac{1}{\sigma} - 1\right) X_s - \frac{\mu}{\sigma}\right)\left(\phi(X_{s+i}) + \left(\frac{1}{\sigma} - 1\right) X_{s+i} - \frac{\mu}{\sigma}\right)\Bigg] \\
=&\frac{1}{\tilde{T}}\sum_{s=1}^{\tilde{T}}\Bigg[\phi(X_s)\phi(X_{s+i}) - \phi(X_s)\phi(X_{s+i}) \\
&+ \phi(X_s)\left(\left(\frac{1}{\hat{\sigma}} - \frac{1}{\sigma}\right) X_s - \frac{\hat{\mu}}{\hat{\sigma}} + \frac{\mu}{\sigma}\right) \\
&+ \phi(X_{s+i})\left(\left(\frac{1}{\hat{\sigma}} - \frac{1}{\sigma}\right) X_{s+i} - \frac{\hat{\mu}}{\hat{\sigma}} + \frac{\mu}{\sigma}\right) \\
&+ X_sX_{s+i}\left(\left(\frac{1}{\hat{\sigma}} - 1\right)^2 - \left(\frac{1}{\sigma} - 1\right)^2\right) \\
&+ (X_s + X_{s+i})\left(\left(\frac{1}{\sigma} - 1\right)\frac{\mu}{\sigma} - \left(\frac{1}{\hat{\sigma}} - 1\right)\frac{\hat{\mu}}{\hat{\sigma}}\right) \\
&+ \left(\frac{\hat{\mu}}{\hat{\sigma}}\right)^2 - \left(\frac{\mu}{\sigma}\right)^2\Bigg].
\end{aligned} \tag{5.6}$$

Each difference converges to 0 in probability, so K_2 also converges to 0 in probability. K_3 and K_4 require a different approach, and come about more easily. Taking a look at the second bracket in K_3

$$\gamma(j) - \frac{1}{\tilde{T}} S_{\tilde{T}}^{(j)}$$

where $\gamma(j)$ is the expected value of the series $(Y_i Y_{i+j})_{i\in\mathbb{N}}$, notice that $\frac{1}{\tilde{T}} S_{\tilde{T}}^{'(j)} = \frac{1}{\tilde{T}} \sum_{i=1}^{\tilde{T}} Y_i Y_{i+j}$ is an estimator for the mean of the series $(Y_i Y_{i+j})_{i\in\mathbb{N}}$. The Law of Large Numbers can be applied, and $\frac{1}{\tilde{T}} S_{\tilde{T}}^{'(j)} \xrightarrow{P} \gamma(j)$- meaning K_3 converges to 0 in probability. The same argument applies to K_4, looking at the second bracket, and applying the central limit theorem as shown above gives $\frac{1}{T} S_{\tilde{T}}^{(i)} \xrightarrow{P} \gamma(i)$ and the whole expression converges in probability to 0. □

Proposition 1. *$(X_i)_{i\in\mathbb{N}}$ be a 1-dimensional stationary, and strongly mixing sequence with mixing coefficients $(\alpha_k)_{k\in\mathbb{N}}$ satisfying $\alpha_k = O(k^{-3-\epsilon})$ for some $\epsilon > 0$. Let $\phi : \mathbb{R} \to \mathbb{R}$ be a bounded, non-zero, function such that $Y_i = \phi(\frac{X_i - \mu}{\sigma_i})$ where μ is the median and σ_i is the MAD.*
for some p and $\tilde{T} = T - p$, and assume that matrix

$$U = \sum_{h=-\infty}^{\infty} \operatorname{Cov} \begin{pmatrix} Y_1 Y_1 \\ \vdots \\ Y_1 Y_{1+p} \end{pmatrix} \begin{pmatrix} Y_{1+h} Y_{1+h} \\ \vdots \\ Y_{1+h} Y_{1+h+p} \end{pmatrix}$$

is positive definite. Then

$$\frac{1}{\sqrt{\tilde{T}}} (\boldsymbol{S}_{\lfloor Tx \rfloor}^p - \frac{\lfloor Tx \rfloor}{T} \boldsymbol{S}_T^p) \xrightarrow{\mathcal{D}} BB(x) \tag{5.7}$$

Where $BB(t)$ is an p-dimensional Brownian-Bridge with covariance matrix $\operatorname{Var}(BB(t)) = t(1-t))U$.

Proof. Begin by applying a Cramer-wold device, i.e. $\frac{1}{\sqrt{T}} \left(\boldsymbol{S}_{\lfloor Tx \rfloor}^p - \frac{\lfloor Tx \rfloor}{\tilde{T}} \boldsymbol{S}_T^p \right) \xrightarrow{\mathcal{D}} BB(x)$ if and only if:

$$(\boldsymbol{\lambda})^\top \frac{1}{\sqrt{T}} \left(\boldsymbol{S}_{\lfloor Tx \rfloor}^p - \frac{\lfloor Tx \rfloor}{\tilde{T}} \boldsymbol{S}_T^p \right) \xrightarrow{\mathcal{D}} (\boldsymbol{\lambda})^\top (BB(x)) \tag{5.8}$$

for all $\lambda \in \mathbb{R}^{p+1}$ such that $\boldsymbol{\lambda}^\top \boldsymbol{\lambda} = 1$. The left side of (5.8) becomes

$$\frac{1}{\sqrt{T}} \sum_{i=0}^{p} \lambda_i S_{\lfloor Tx \rfloor}^i - \frac{\lfloor Tx \rfloor}{\tilde{T}} \left(\frac{1}{\sqrt{T}} \sum_{i=0}^{p} \lambda_i S_{\lfloor Tx \rfloor}^i \right) \tag{5.9}$$

To show convergence apply the continuous mapping theorem, define h^{-1} as follows:

$$h^{-1} : D[0,1] \to D[0,1] \tag{5.10}$$
$$s.t \quad f(x) \mapsto f(x) - x f(1) \tag{5.11}$$

This function is continuous in the Skorohod topology, so it need only be shown that

$$f(x) = \frac{1}{\sqrt{\tilde{T}}} \sum_{i=0}^{p} \lambda_i S_{\lfloor Tx \rfloor}^i \xrightarrow{\mathcal{D}} \sigma(\lambda_0, \ldots, \lambda_{p+1}) \mathbf{B}(x) \tag{5.12}$$

Where $B(x)$ is a $p+1$-dimension Brownian-Motion in the Skorohod space $D[0,1]$, and σ depends on $\lambda_0, \dots, \lambda_p$. To retrieve this first the convergence property of the series $Z_i^l = \frac{\lambda_i}{\sqrt{\tilde{T}}}(Y_i Y_{i+l} - \gamma(l))$ must be shown. After that another appropriate map g^{-1} will allow an application of the continuous mapping theorem and give the final convergence. Applying theorem 1.3 of Merleveed & Peligrad[28] to the series $(Z_i^l)_{i \in [1,\tilde{T}]}$ gives

$$k(x) = \frac{\sum_1^{\lfloor Tx \rfloor} Z_i^l}{\sqrt{\frac{\pi}{2}}\mathbb{E}[\left|\sum_{i=1}^T Z_i^l\right|]} \xrightarrow{\mathcal{D}} B(x) \tag{5.13}$$

Where $B(x)$ is a Brownian Motion. All requirements for this theorem to be used are met

- The addition of a constant $\gamma(l)$ does not affect the stationarity of $Y_i Y_{i+l}$, so Z_i^l is also stationary.
- $\mathbb{E}[Z_i^l] = \mathbb{E}[Y_i Y_{i+l}] - \gamma(l) = \gamma(l) - \gamma(l) = 0$, so the series is centred.
- By lemma 1 the series Z_i^l is at least as strongly mixing as the series $(Y_i Y_{i+l})_{i \in [1,\tilde{T}]}$, and in turn $(Y_i)_{i \in [1,T]}$.
- Since the series $(Y_i)_{i \in [1,T]}$ is stationary it's second moment is fixed, therefore the second moment of $(Y_i Y_{i+l})_{i \in [1,\tilde{T}]}$ is bounded.
- To show $\liminf \frac{\mathbb{E}[\left(\sum_{i=1}^T Z_i^l\right)^2]}{T} > 0$ see that

$$\begin{aligned}
\frac{1}{T}\mathbb{E}[\left(\sum_{i=1}^T Z_i^l\right)^2] &= \frac{1}{T}\mathbb{E}[\sum_{i=1}^T \sum_{j=1}^T Z_i^l Z_j^l] \\
&= \frac{1}{T}\sum_{i=1}^T \sum_{j=1}^T \mathbb{E}[Z_i^l Z_j^l] \\
&= \frac{1}{T}\sum_{i=1}^T \sum_{j=1}^T \mathbb{E}[(Y_i Y_{i+l} - \gamma(l))(Y_j Y_{j+l} - \gamma(l))] \\
&= \frac{1}{T}\sum_{i=1}^T \sum_{j=1}^T \operatorname{Cov} Y_i Y_{i+l} Y_j Y_{j+l} \\
&= \sum_{k=-(T-l}^{T-l} \frac{T-k}{T} \operatorname{Cov} Y_i Y_{i+l} Y_j Y_{j+l} \\
&\xrightarrow{T \to \infty} U_{l,k}
\end{aligned}$$

  Where $U_{l,k} > 0$ since $U \succ 0$.

- $Q(u)$(2.18) is bounded above by 1. Therefore $\int_0^{\alpha_T'} Q^2_{|Z_0^l|}(u)du$ has $[1, \alpha_T'] \subseteq [1, T^{-3-\epsilon}]$ as the bounds being integrating over, for some $\epsilon > 0$, where α_k' are the mixing coefficients of the series $(Z_i^l)_{i \in [1,\tilde{T}]}$. Therefore the entire integral is bounded above the magnitude of this set times the maximum height of $Q^2(u)$ which is 1. So

$$\int_0^{\alpha_T'} Q^2_{|Z_0^l|}(u)du \leq T^{-3-\epsilon} = o(T^{-3-\epsilon})$$

  And the final requirement is met.

So the conditions for (5.13) are met. An application of the continuous mapping theorem gives (5.12), by defining function g^{-1}

$$g^{-1} : D[0,1] \to D[0,1] \tag{5.14}$$

$$s.t. \quad k(x) \mapsto \sqrt{\frac{\pi}{2}} \left(k(x) + \gamma(l)\right) \mathbb{E}[|k(1)|] \tag{5.15}$$

is affine, and retrieves the function $f(x)$ in line (5.12). So since $f(x) \overset{D[0,1]}{\Rightarrow} B(x)$ then through the continuous mapping theorem and the map h^{-1}

$$\frac{1}{\sqrt{T}} \sum_{i=0}^{p} \lambda_i S^i_{\lfloor Tx \rfloor} - \frac{\lfloor Tx \rfloor}{\tilde{T}} \left(\frac{1}{\sqrt{T}} \sum_{i=0}^{p} \lambda_i S^i_{\lfloor Tx \rfloor} \right) \overset{\mathcal{D}}{\to} \sum_{i=0}^{p} \lambda_i (B_i(x) - \frac{\lfloor Tx \rfloor}{\tilde{T}} B_i(x)) = \sum_{i=0}^{p} \lambda_i BB_i(x)$$

Where $BB_i(x)$ are Brownian Bridges. Which satisfies the requirements for the Cramer-Wold device (5.8), so

$$\frac{1}{\sqrt{\tilde{T}}} \left(\boldsymbol{S}_{\lfloor Tx \rfloor^p} - \frac{\lfloor Tx \rfloor}{\lfloor Tx \rfloor} \boldsymbol{S^p_T} \right) \overset{\mathcal{D}}{\to} BB(x)$$

And we are done, to find the covariance $\mathrm{Cov}(BB(x), BB(y))$ perform the following.

$$\mathrm{Var}\, S^p_{\lfloor Tx \rfloor} = \mathrm{Var} \begin{pmatrix} \sum_{i=1}^{\lfloor Tx \rfloor} Y_i Y_{i+1} \\ \vdots \\ \sum_{i=1}^{\lfloor Tx \rfloor} Y_i Y_{i+p} \end{pmatrix}$$

This is a $p \times p$ matrix where each entry is written as

$$\begin{aligned} \Sigma_{l,m} &= \mathrm{Cov}(\sum_{i=1}^{\lfloor Tx \rfloor} Y_i Y_{i+l}, \sum_{j=1}^{\lfloor Tx \rfloor} Y_j Y_{i+m}) \\ &= \sum_{i=1}^{\lfloor Tx \rfloor} \sum_{i=1}^{\lfloor Tx \rfloor} \mathrm{Cov}(Y_i Y_{i+l}, Y_j Y_{j+m}) \end{aligned}$$

An application of kernel estimator to estimate this inner sum gives

$$\begin{aligned} &= \sum_{i=1}^{\lfloor Tx \rfloor} \hat{U}_{l,m} + O(1) \\ &= \lfloor Tx \rfloor \hat{U}_{l,m} + O(1) \\ &\approx \lfloor Tx \rfloor \hat{U}_{l,m} \end{aligned}$$

So $\mathrm{Var}\, \boldsymbol{S}^p_{\lfloor Tx \rfloor} = k\hat{U}$. Note that this argument involves a heuristic in places. The variance of the total expression is therefore

$$\begin{aligned} \frac{1}{\tilde{T}} \left(\mathrm{Var} \left(\boldsymbol{S}^p_{\lfloor Tx \rfloor} - \frac{\lfloor Tx \rfloor}{T} \boldsymbol{S}^p_T \right) \right) &= \frac{1}{\tilde{T}} \left(\mathrm{Var}\, \boldsymbol{S}^p_{\lfloor Tx \rfloor} + \left(\frac{\lfloor Tx \rfloor}{\tilde{T}} \right)^2 \mathrm{Var}\, \boldsymbol{S}^p_{\tilde{T}} - 2 \left(\frac{\lfloor Tx \rfloor}{\tilde{T}} \right) \mathrm{Cov}\, \boldsymbol{S}^p_{\lfloor Tx \rfloor} \boldsymbol{S}^p_T \right) \\ &= \frac{1}{\tilde{T}} \left(\lfloor Tx \rfloor U + \left(\frac{\lfloor Tx \rfloor}{\tilde{T}} \right)^2 TU - 2 \frac{\lfloor Tx \rfloor}{\tilde{T}} \lfloor Tx \rfloor \right) \\ &= U \left(\frac{\lfloor Tx \rfloor}{\tilde{T}} + \left(\frac{\lfloor Tx \rfloor}{T} \right)^2 - 2 \left(\frac{\lfloor Tx \rfloor}{T} \right)^2 \right) \\ &= x(1-x)U \end{aligned}$$

□

Theorem 6. *Let X_i be a $1-$dimensional, stationary and strongly mixing sequence with mixing coefficients satisfying $\alpha_k = O(k^{-3-\epsilon})$ for some $\epsilon > 0$. Let $\phi : \mathbb{R} \to \mathbb{R}$ be a bounded function such that $Y_i = \phi\left(\frac{X_i - \mu}{\sigma}\right)$ where μ is the median and σ is the MAD. let $\det U > 0$ where U is the long run covariance. Then*

$$(W_T(x)^2)_{x\in[0,1]} \xrightarrow{\mathcal{D}} (\sum_{i=0}^{p} \bar{B}B_i(x)^2)_{x\in[0,1]} \tag{5.16}$$

Where $\bar{B}B_i(x))_{i\in[0,p]}$ are mutually independent standard Brownian Bridges.

Proof. From proposition 1 that $\frac{1}{\sqrt{\tilde{T}}}\left(\boldsymbol{S}^p_{\lfloor Tx\rfloor} - \frac{\lfloor Tx\rfloor}{T}\boldsymbol{S}^p_T\right) \xrightarrow{\mathcal{D}} \mathbf{BB}(x)$, a p-dimensional Brownian Bridge.
$\hat{U}$ is positive semi-definite and so allows a Cholesky Decomposition $\hat{U} = (\hat{U}^{\frac{1}{2}})^T(\hat{U}^{\frac{1}{2}})$, using this with Slutsky's Theorem gives

$$\hat{U}^{-\frac{1}{2}}\left(\frac{1}{\sqrt{\tilde{T}}}\left(\boldsymbol{S}^p_{\lfloor Tx\rfloor} - \frac{\lfloor Tx\rfloor}{T}\boldsymbol{S}^p_T\right)\right) \xrightarrow{\mathcal{D}} \hat{U}^{-\frac{1}{2}}\mathbf{BB}(x) \tag{5.17}$$

Let $g^{-1} : \mathrm{D}[0,1]^p \to \mathrm{D}[0,1]$ be a map $X \mapsto X^T X$, note that with this map (5.17) becomes $W_T^2(x)$ since

$$\begin{aligned}
&g^{-1}\left(\hat{U}^{-\frac{1}{2}}\left(\frac{1}{\sqrt{\tilde{T}}}\left(\boldsymbol{S}^p_{\lfloor Tx\rfloor} - \frac{\lfloor Tx\rfloor}{T}\boldsymbol{S}^p_T\right)\right)\right)\\
&= \left(\hat{U}^{-\frac{1}{2}}\left(\frac{1}{\sqrt{\tilde{T}}}\left(\boldsymbol{S}^p_{\lfloor Tx\rfloor} - \frac{\lfloor Tx\rfloor}{T}\boldsymbol{S}^p_T\right)\right)\right)^T\left(\hat{U}^{-\frac{1}{2}}\left(\frac{1}{\sqrt{\tilde{T}}}\left(\boldsymbol{S}^p_{\lfloor Tx\rfloor} - \frac{\lfloor Tx\rfloor}{T}\boldsymbol{S}^p_T\right)\right)\right)\\
&= \frac{1}{\tilde{T}}\left(\boldsymbol{S}^p_{\lfloor Tx\rfloor} - \frac{\lfloor Tx\rfloor}{T}\boldsymbol{S}^p_T\right)^T(\hat{U}^{-\frac{1}{2}})^T\hat{U}^{-\frac{1}{2}}\left(\boldsymbol{S}^p_{\lfloor Tx\rfloor} - \frac{\lfloor Tx\rfloor}{T}\boldsymbol{S}^p_T\right)\\
&= \frac{1}{\tilde{T}}\left(\boldsymbol{S}^p_{\lfloor Tx\rfloor} - \frac{\lfloor Tx\rfloor}{T}\boldsymbol{S}^p_T\right)^T\hat{U}^{-1}\left(\boldsymbol{S}^p_{\lfloor Tx\rfloor} - \frac{\lfloor Tx\rfloor}{T}\boldsymbol{S}^p_T\right)\\
&= W_T^2(x)
\end{aligned}$$

So by the Continuous Mapping Theorem $W_T^2(x) \xrightarrow{\mathcal{D}} (\hat{U}^{-\frac{1}{2}}\mathbf{BB}(x))^T(\hat{U}^{-\frac{1}{2}}\mathbf{BB}(x))$. All that is left to show now is that this results in a sum of squares of mutually independent Brownian Bridges $\mathrm{BB}_i(x)$ $i = 0, \ldots, p$

$$\begin{aligned}
\mathrm{Var}\left(\hat{U}^{-\frac{1}{2}}\mathbf{BB}(x)\right) &= \hat{U}^{-\frac{1}{2}}\,\mathrm{Var}\left(\mathbf{BB}(x)\right)(\hat{U}^{-\frac{1}{2}})^T\\
&= (\hat{U}^{-\frac{1}{2}})\left(x(1-x)\hat{U}\right)(\hat{U}^{-\frac{1}{2}})^T\\
&= x(1-x)I_p
\end{aligned}$$

Where I_p is the $(p+1)\times(p+1)$ identity matrix. Off-diagonals in this matrix represent the covariance between components of $\hat{U}^{-\frac{1}{2}}\mathbf{BB}(x)$, since these are 0 the vectors components are mutually independent, call this new vector $\bar{\mathbf{B}}\mathbf{B}(x)$. So the final expression is $W_T^2(x) \xrightarrow{\mathcal{D}} \sum_{i=0}^{p}\bar{\mathrm{B}}\mathrm{B}_i(x)$ where $\bar{\mathrm{B}}\mathrm{B}_i(x)$ are mutually independent Brownian Bridges. □