Detecting Deviations from Stationarity

Ruprecht Puchstein

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften
an der Fakultät für Mathematik
der Ruhr-Universität Bochum

April 2014
Contents

1 Introduction 3

2 Preliminaries 7
   2.1 Stationary time series 7
      2.1.1 Definitions, properties and examples 7
      2.1.2 Estimation for stationary time series 12
   2.2 Non-stationary times series models 15
      2.2.1 Piecewise stationary linear time series 15
      2.2.2 Locally stationary linear time series 17
   2.3 Auxiliary concepts 18
      2.3.1 Cumulants 18
      2.3.2 Weak convergence of empirical processes 23

3 A new approach to the change point problem in time series analysis 27
   3.1 Testing for the presence of structural breaks 30
      3.1.1 Motivation and definitions 30
      3.1.2 An empirical process tracking discontinuity in the time direction 31
      3.1.3 Weak convergence of the empirical process 32
      3.1.4 Bootstrapping the test statistic 35
   3.2 Dividing the data set into stationary segments 41
   3.3 Practical application and finite sample properties 48
      3.3.1 Recommendations for implementation 48
      3.3.2 Simulation study 51
   3.4 Proofs 58
      3.4.1 Proof of Theorem 3.1.1 58
      3.4.2 Proof of Theorem 3.1.2 86
      3.4.3 Proof of Theorem 3.1.6 92
      3.4.4 Proof of Theorem 3.1.7 95
      3.4.5 Proof of Theorem 3.2.1 106
      3.4.6 Proof of Theorem 3.2.2 107

4 Testing for stationarity in locally stationary time series 111
   4.1 Motivation and definitions 113
   4.2 Tracking non-constancy in the time direction 115
4.3 Weak convergence of the empirical process ......................... 116
4.4 Bootstrapping the test statistic ......................................... 117
4.5 Practical application and finite sample properties .................. 119
  4.5.1 Recommendations for implementation ............................ 119
  4.5.2 Simulation study ....................................................... 120
4.6 Proofs ............................................................................. 124
  4.6.1 Proof of Theorem 4.3.2 ................................................. 124
  4.6.2 Proof of Theorem 4.4.1 ................................................. 150

Bibliography .......................................................................... 153
Chapter 1

Introduction

Time series data is the result of observations, which are made over the course of some time interval. This kind of ordered data arises in a natural way in various fields of interest, as for example in hydrology, where one observes fluctuations of tides, or in finance, where classical examples include daily exchange rate quotes or stock market returns. Fundamental to all time series data is the perception that the order in which the data was obtained contains valuable information about the phenomenon under investigation.

The field of time series analysis is concerned with the investigation of this kind of data. One of the main objectives is to develop a thorough understanding of the observed data in general and of the properties of the underlying stochastic system in particular. The reasons for this quest are manifold and might include such altruistic components as scientific curiosity. However, the driving force for most applications of time series analysis is the desire to obtain valuable information about the phenomenon at hand in order to enhance the decision making process in some related matter. In the context of hydrology, such a calculation might consider historical data of high water marks for identifying the ideal location of a new dam, which is intended to contain high tides. In the field of finance, the analysis of historic asset returns is widely used for multiple purposes as for example risk management, portfolio construction and trading signal generation.

In practice, most methods for the investigation of time series data require the practitioner to specify a model, which characterises certain properties of the data generating process. For any application, making such a choice contains a fundamental tradeoff between two contradicting features, namely complexity and tractability: On the one hand, the selected model should be able to explain a sufficient amount of complex patterns present in the data and especially must not exclude features which are fundamental to the governing stochastic process. On the other hand, the imposed assumptions have to be sufficiently strong to enable the development of statistical procedures. In this regard, the class of stationary time series models has found considerable attention in the literature and one of the most broadly employed assumptions concerning some sequence \( \{X_t\}_{t \in \mathbb{Z}} \) of (time ordered) random variables is that it belongs to the class of ARMA\((p, q)\) time series. This means that for each \( t \in \mathbb{Z} \) the
random variable $X_t$ is a stationary solution of the system

$$X_t - \sum_{j=1}^{p} \phi_j X_{t-j} = Z_t + \sum_{j=1}^{q} \theta_j Z_{t-j},$$

where $p$ denotes the AR-order, $q$ the MA-order and $\{Z_t\}_{t \in \mathbb{Z}}$ is a White Noise sequence of innovations. This type of time series models allows to track many features of serial dependence, which might be present in the data, and thus received a lot of attention in the literature. More recently and beginning with the introduction of the ARCH and GARCH models, the investigation of more flexible models, which enable the modelling of volatility clustering as it is observed in financial market return series, has gained tremendous popularity. Although the specific aspects of time interdependency, which these models are intended to explain, are very different, all models mentioned above are examples of stationary time series models. Broadly speaking, the assumption of stationarity means that the underlying stochastic process, which determines the observed realisations, is time invariant. This characteristic implies that more and more information about the underlying distributional properties becomes available as time advances and the sample size increases. This feature allows for the development of a rewarding asymptotic theory for many stationary time series models and it is for this reason that the framework of stationarity has become the dominant paradigm in the analysis of time-dependent data. For practical reasons, imposing the assumption of stationarity has the appeal that a large variety of statistical methods is available for performing inference in this kind of time series data. Out of the large literature covering the topic of stationary time series we mention Brockwell and Davis (1991), Anderson (1971), Brillinger (1981) and Kreiß and Neuhaus (2006).

Researchers as well as partitioners are often confronted with more than one time-dependent set of data and in many situations it is advisable to consider these sets as the joint realisation of one multivariate time series rather than a collection of stand-alone observations. This is due to the fact that treating the situation in the framework of multivariate time series analysis allows for the rigorous study of serial interdependence between the univariate components, which is for example of great importance in the study of return series of different asset classes or stock markets (see Tsay (2010)). A large variety of time series models considering the univariate case has been extended to the multivariate case and many of the statistical methods, which are known to yield reliable results for the analysis of univariate stationary time series, have been successfully extended to the multivariate case as well. For an overview of properties of multivariate stationary time series models and for methods of statistical inference in these models see Reinsel (1993) and Lütkepohl (2006), who explicitly consider time series with more than one component.

However, for many applications the assumption of stationarity turns out to be too restrictive and even more complicated stationary time series models seem to be unable to capture a sufficient degree of complexity in the observed data. Out of the large literature, which is concerned with the empirical analysis of financial stock market returns, we exemplarily mention Mikosch and Starica (2000), who develop a goodness of fit test for GARCH processes.
In an empirical analysis of the S&P 500, they conclude that the return series of this stock market index is likely to exhibit structural changes in the unconditional variance. This result suggests that the class of GARCH models is not able to satisfactorily explain the clustering of volatility in stock market returns, which is the purpose it was originally developed for. The general perception that even more sophisticated stationary models might not be able to explain observed data in a sufficiently accurate manner has encouraged many researchers to consider non-stationary time series models as an alternative. The general intuition of these approaches is to allow the underlying data generating system to exhibit some form of evolution over time, which by definition is excluded in the stationary framework. In this regard, the concept of locally stationary time series introduced by Dahlhaus (1996) has found considerable interest in the literature. The basic idea of this paradigm is to assume that the time series can locally be approximated by stationary models which makes an asymptotic theory feasible. While the framework of locally stationary models allows to capture a smooth evolution of several stylised facts of an observed set of data, the class of piecewise stationary time series follows another intuition for admitting features of non-stationarity. Roughly speaking, a piecewise stationary process is characterised by a certain number of break points in time, at which the underlying data generating process changes abruptly from one stationary model to another. In between of two consecutive break points, the time series can be adequately represented by some stationary model.

Methods for detecting non-stationarities in time series data and for investigating the specifics thereof are crucially important. The reason for this is twofold. Firstly, a majority of statistical procedures for analysing time-dependent data crucially depend on the assumption that the underlying data generating process is stationary. These procedures include methods for parameter estimation and forecasting techniques and have been shown to yield reliable results when applied to stationary data sets but can yield spurious conclusions if this condition does not hold. In order to justify the application of these methods, it is important to have statistical methods that are able to investigate the question whether a specific set of data can be assumed to originate from a stationary process. Secondly, if the underlying process is supposed to be non-stationary a closer understanding of the particular kind of non-stationarity present in the data can indicate how conventional methods for time series analysis can be adapted to yield solid results in the non-stationary model at hand. For example, consider the case of a time series which can be assumed to follow some piecewise stationary parametric model. In this situation, conventional methods for estimating the model parameters which are based on the whole sample are unlikely to achieve satisfying results. However, if one is able to correctly identify the stationary segments, then calculating the estimators on the individual segments becomes feasible, which in general will produce better results.

The objective of this thesis is to develop procedures that allow to determine whether an observed set of time ordered data can be assumed stationary. For this purpose, we will establish two new testing procedures which enable the practitioner to investigate the null hypothesis of a constant second order moment structure against several alternatives. Furthermore, we explain how the derived theoretical results can be exploited to obtain a deeper
understanding of the special kind of non-stationarity, which might be present in observed data in the empirical science. The structure of this discourse is as follows: In Chapter 2, we first introduce the time series models which are of relevance for this thesis. In this regard, we formally define the framework of second order stationarity in time series analysis and explain two types of non-stationary time series models, namely piecewise stationary models and locally stationary models, which allow the second order structure to evolve as time advances. The second purpose of this chapter is to give a summary of theoretical concepts that are needed for the technical proofs of this thesis. Chapter 3 is devoted to the change point problem in time series analysis. The presented discussion formally establishes a new non-parametric procedure for testing the null hypothesis of no structural breaks in a multivariate time series. Furthermore, we explain an algorithmic procedure for estimating the true number and locations of possibly multiple break points in a multivariate time series and it is shown that this method allows to consistently identify the stationary segments of the underlying process. In Chapter 4, we consider the problem of detecting smooth deviations from the stationarity assumption. For this purpose, we establish a novel non-parametric approach for testing for stationarity in the class of locally stationary time series models. The applicability of all established methods is thoroughly investigated by extensive simulation studies and the results of these surveys are given in separate sections.

At this point, I would like to express my gratitude to my supervisors Holger Dette and Philip Preuß, for whose encouragement and mentoring during the last years I am deeply grateful.
Chapter 2

Preliminaries

This chapter is devoted to the introduction of basic definitions and mathematical concepts needed for this thesis. The purpose of this presentation is twofold. On the one hand, we will present the specific time series models which we will work with. On the other hand, we intend to give a brief summary of auxiliary concept that will be mostly of technical detail and build the foundation for the formal proofs of this thesis.

2.1 Stationary time series

In the following presentations we introduce the framework of second order stationary time series, which over the last decades became a corner stone for the empirical analysis of apparently irregularly fluctuating time ordered observations [see Anderson (1971)]. We begin our review by formally presenting the concept of stationarity in time series. For this purpose, we define the mean level and the covariance function for a time discrete stochastic process and explain the properties that characterise a second order stationary time series model. We continue by giving some examples and by defining the class of linear time series. In the second part of this section, we turn our attention to a brief review of basic estimation theory for stationary time series models. In this context, we also define the periodogram, which enables a very efficient approach to statistical inference of the dependence structure of time series and will be of major importance in the following chapters. The presentations of this section are closely aligned to the demonstrations of chapters 1 and 11 in Brockwell and Davis (1991). All definitions and results presented are well established in the literature and we mention the textbooks of Anderson (1971), Kreiß and Neuhaus (2006), Brillinger (1981) and Rosenblatt (1985) where similar compositions of results can be found.

2.1.1 Definitions, properties and examples

We begin our introduction into the framework of stationary time series with the following definition.

Definition 2.1.1 (Stationary time series)

A $\mathbb{R}^d$-valued stochastic process $\{X_t\}_{t \in \mathbb{Z}}$, $X_t = (X_{t,1}, ..., X_{t,d})$, is referred to as a multivariate
stationary time series if the following properties are fulfilled:

(i) For all \( t \in \mathbb{Z} \) and for each component \( i \in \{1, \ldots, d\} \), it holds
\[
E(X_{t,i}^2) < \infty.
\]

(ii) There exists some vector \( \mu \in \mathbb{R}^d \) such that
\[
E(X_t) = \mu
\]
for all \( t \in \mathbb{Z} \).

(iii) For each \( h \in \mathbb{Z} \), there exist a positive-semidefinite matrix \( \Gamma(h) \in \mathbb{R}^{d \times d} \), \( \Gamma(h) = [\gamma_{i,j}(h)]_{i,j=1}^d \), such that
\[
E[(X_{t+h} - \mu)(X_t - \mu)^T] = \Gamma(h).
\]

A \( \mathbb{R}^d \)-valued process \( \{X_t\}_{t \in \mathbb{Z}} \) satisfying properties (1),(2) and (3) is said to be a stationary multivariate time series with mean value vector \( \mu \) and covariance matrix \( \Gamma(h) \) at lag \( h \in \mathbb{Z} \).

The following are some examples of well known stationary time series models.

**Example 2.1.1**

(i) A sequence \( \{Z_t\}_{t \in \mathbb{Z}} \) of \( d \)-variate real valued random variables is called \( d \)-variate White Noise if \( \mu = 0 \) and
\[
\Gamma(h) = \begin{cases} 
\Sigma & \text{if } h = 0 \\
0 & \text{if } h \neq 0,
\end{cases}
\]
for some covariance matrix \( \Sigma \).

(ii) A process \( \{X_t\}_{t \in \mathbb{Z}} \) is called a \( d \)-variate ARMA\((p,q)\) process, if it is a stationary solution of the difference equations
\[
X_t - \Phi_1 X_{t-1} - \ldots - \Phi_p X_{t-p} = Z_t + \Theta_1 Z_{t-1} + \ldots + \Theta_q Z_{t-q}, \tag{2.1}
\]
where \( p, q \in \mathbb{N} \), \( \{\Phi_j\}_{j=1,\ldots,p} \) and \( \{\Theta_j\}_{j=1,\ldots,q} \) are \( \mathbb{R}^{d \times d} \) valued matrices and \( \{Z_t\}_{t \in \mathbb{Z}} \) is a sequence of \( d \)-dimensional White Noise. In the literature, the characterising equations (2.1) are frequently written in the shorter form
\[
\Phi(B)X_t = \Theta(B)Z_t,
\]
where \( B \) denotes the backward shift operator satisfying \( B^j X_t = X_{t-j} \) and the polynomials \( \Phi \) and \( \Theta \) are defined by
\[
\Phi(z) : = I_d - \Phi_1 z - \ldots - \Phi_p z^p
\]
and
\[
\Theta(z) : = I_d + \Theta_1 z + \ldots + \Theta_q z^q.
\]
(iii) An ARMA\((p,0)\) process is called AR\((p)\) process and an ARMA\((0,q)\) process is denoted as a MA\((q)\) process.

We now define the class of linear time series, which will be of major importance for the considerations of this thesis.

**Definition 2.1.2 (Linear time series)**

A \(\mathbb{R}^d\)-valued stationary time series \(\{X_t\}_{t \in \mathbb{Z}}\) belongs to the class of linear time series, if there exists a sequence \(\{\Psi_l\}_{l \in \mathbb{Z}}\) of matrices \(\Psi_l := ([\Psi_l]_{i,j})_{i,j=1}^d\) which satisfies \(\sum_{l=-\infty}^{\infty} \|\Psi_l\|_\infty^2 < \infty\), and a sequence \(\{Z_t\}_{t \in \mathbb{Z}}\) of independent and identically distributed innovations \(Z_t\) with finite covariance matrix \(\Sigma \in \mathbb{R}^{d\times d}\), such that

\[
X_t = \sum_{l=-\infty}^{\infty} \Psi_l Z_{t-l}.
\]  

Throughout this thesis, we will exceptionally consider such stationary time series \(\{X_t\}_{t \in \mathbb{Z}}\) which at least locally feature a representation of the form (2.2). This restriction is motivated by the fact that it allows for a rather elegant approach to investigating stylised facts of the times series dependence structure. Before we explain these concepts in more detail, we present a few examples of stationary linear time series.

**Example 2.1.2**

(i) For a \(d\)-variate AR\((1)\) process

\[
X_t = \Phi X_{t-1} + Z_t,
\]

where \(\{Z_t\}_{t \in \mathbb{Z}}\) denotes a sequence of \(d\)-dimensional independent and identically distributed innovations with covariance matrix \(\Sigma\) and where the matrix \(\Phi\) satisfies

\[
det(I_d - z\Phi) \neq 0 \quad \text{for all} \quad z \in \mathbb{C}, \quad |z| \leq 1,
\]

we have the representation [see Example 11.3.1 in Brockwell and Davis (1991)]

\[
X_t = \sum_{j=0}^{\infty} \Phi^j Z_{t-j}.
\]

(ii) We consider a stationary ARMA\((p,q)\) time series \(\{X_t\}_{t \in \mathbb{Z}}\) satisfying the difference equations

\[
\Phi(B)X_t = \Theta(B)Z_t
\]

for some independent and identically distributed sequence \(\{Z_t\}_{t \in \mathbb{Z}}\). If the regularity condition

\[
det(\Phi(z)) \neq 0 \quad \text{for all} \quad z \in \mathbb{C}, \quad |z| \leq 1
\]

is satisfied, then
is fulfilled, then there exists exactly one stationary solution
\[ X_t = \sum_{j=0}^{\infty} \Psi_j Z_{t-j}, \]
where the sequence \( \{\Psi_j\}_{j \geq 0} \) of matrices is uniquely determined by the equations
\[ \sum_{j=0}^{\infty} \Psi_j z^j = \Phi^{-1}(z) \Theta(z), \quad |z| \leq 1 \]
[see Theorem 11.3.1 in Brockwell and Davis (1991)].

At the beginning of this section, we mentioned that the framework of stationary time series is frequently applied for the investigation of statistical properties of apparently irregularly fluctuating processes. The following result gives the covariance structure of a general linear time series of the form (2.2) and substantiates this feature.

**Theorem 2.1.3 (Covariance function of linear time series)**
Let \( \{X_t\}_{t \in \mathbb{Z}} \) denote a stationary time series featuring a linear representation of the form
\[ X_t = \sum_{l=-\infty}^{\infty} \Psi_l Z_{t-l}, \]
where the components of the sequence \( \{\Psi_l\}_{l \in \mathbb{Z}} \) are absolutely summable and where the sequence \( \{Z_t\}_{t \in \mathbb{Z}} \) is White Noise with covariance matrix \( \Sigma \). Then it is
\[ \Gamma(h) = \sum_{j=-\infty}^{\infty} \Psi_{j+h} \Sigma \Psi_j^T, \quad h \in \mathbb{Z}. \]

To continue our review of fundamental definitions and properties of stationary time series, we now turn our attention to an alternative concept for capturing the second order structure of a multivariate time series \( \{X_t\}_{t \in \mathbb{Z}} \). For this purpose, we define the spectral density matrix for a stationary time series.

**Definition 2.1.4 (Spectral density)**
For a \( \mathbb{R}^d \)-valued stationary time series \( \{X_t\}_{t \in \mathbb{Z}} \) satisfying
\[ \sum_{h=-\infty}^{\infty} \|\Gamma(h)\|_\infty < \infty, \quad (2.3) \]
we define the spectral density matrix \( f : [-\pi, \pi] \rightarrow \mathbb{C}^{d \times d} \) by
\[ f(\lambda) := \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \exp(i\lambda h) \Gamma(h). \quad (2.4) \]
The following famous theorem shows that the second order moment structure of a stationary time series is completely determined by its spectral density matrix. This result formally shows that it is sensible to consider the Fourier transform of the covariance function for an investigation of the dependence structure of a time-dependent set of data and thus builds the theoretical foundation for frequency domain methods for time series analysis.

**Theorem 2.1.5 (Spectral representation)**
Consider a stationary time series \( \{X_t\}_{t \in \mathbb{Z}} \) whose covariance function satisfies the condition (2.3). Then it is

\[
\Gamma(h) = \int_{-\pi}^{\pi} \exp(-i\lambda h) f(\lambda) d\lambda
\]
for each \( h \in \mathbb{Z} \).

While Theorem 2.1.3 gives an efficient means to calculate the covariance function \( \Gamma \) for some linear time series \( \{X_t\}_{t \in \mathbb{Z}} \), the result below shows how the spectral density matrix \( f \) of a linear time series can be explicitly specified.

**Theorem 2.1.6 (Spectral density of a linear time series)**
Let \( \{X_t\}_{t \in \mathbb{Z}} \) denote a centred stationary time series which has a linear representation of the form

\[
X_t = \sum_{l=-\infty}^{\infty} \Psi_l Z_{t-l},
\]
where the components of the sequence \( \{\Psi_l\}_{l \in \mathbb{Z}} \) are absolutely summable and where the sequence \( \{Z_t\}_{t \in \mathbb{Z}} \) is White Noise with covariance matrix \( \Sigma \). Then it is

\[
f(\lambda) = \frac{1}{2\pi} \sum_{l,m=-\infty}^{\infty} \Psi_l \Sigma \Psi_T m \exp(-i\lambda(l-m)).
\]

**(2.5)**

**Proof:** For a proof of (2.5) we employ the linear representation of the time series \( \{X_t\}_{t \in \mathbb{Z}} \) to obtain

\[
f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \exp(i\lambda h) \Gamma(h) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \exp(i\lambda h) \mathbb{E}(X_t X_{t+h}^T)
\]

\[
= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \exp(i\lambda h) \Psi_l \mathbb{E}(Z_{t-l} Z_{t+h-m}^T) \Psi_T m.
\]

Due to the fact that the sequence \( \{Z_t\}_{t \in \mathbb{Z}} \) is White Noise, we furthermore get that

\[
\mathbb{E}(Z_{t-l} Z_{t+h-m}^T) = \begin{cases} 
\Sigma & \text{if } h = m - l \\
0 & \text{else}
\end{cases}
\]

which yields the assertion. \( \square \)
2.1.2 Estimation for stationary time series

In the previous section, we have introduced the framework of stationary time series and presented fundamental properties and some examples. Furthermore, we have defined the mean value vector \( \mu \) and the covariance matrix function \( \{ \Gamma(h) \}_{h \in \mathbb{Z}} \), which are critical quantities capturing stylised facts of a multivariate stationary time series \( \{ X_t \}_{t \in \mathbb{Z}} \). The fact that for most data sets originating from real world phenomena the details pertaining to the first and second moment structure are not known demonstrates the need for statistical methods for inferring information relating to these quantities. In this section, we will shortly comment on estimating the mean level of a stationary time series and then turn our attention to the investigation of the dependence structure of a multivariate time ordered set of data. For this purpose, let us assume that \( \{ X_t \}_{t=1,...,T} \) is the realisation of a stationary time series, where the vector

\[
X_t = \mu + Y_t
\]

of observations at time \( t \) is comprised of an unknown but time-invariant mean level \( \mu \) and a centred but time-dependent and stochastic vector \( Y_t \) of fluctuations such that \( \{ Y_t \}_{t \in \mathbb{Z}} \) is a stationary time series. An intuitive estimator for the vector \( \mu \) is then defined by

\[
\bar{X}_T := \frac{1}{T} \sum_{t=1}^{T} X_t.
\]

The following theorem establishes the asymptotic normality of the estimator \( \bar{X}_T \) under the assumption that the time series \( \{ Y_t \}_{t \in \mathbb{Z}} \) of fluctuations is linear.

**Theorem 2.1.7** (Asymptotic normality of \( \bar{X}_T \))

Assume that each element of the data set \( \{ X_t \}_{t=1,...,T} \) has a representation of the form (2.6), where the time series \( \{ Y_t \}_{t \in \mathbb{Z}} \) follows the linear representation

\[
Y_t = \sum_{l=-\infty}^{\infty} \Psi_l Z_{t-l},
\]

and where \( \{ Z_t \}_{t \in \mathbb{Z}} \) is a sequence of independent identically distributed random vectors with covariance matrix \( \Sigma \) and where the sequence \( \{ \Psi_l \}_{l \in \mathbb{Z}} \) of matrices satisfies the condition

\[
\sum_{l=-\infty}^{\infty} \| \Psi_l \|_\infty < \infty.
\]

Then \( \bar{X}_T \) is asymptotically normal with mean value vector \( \mu^{\bar{X}_T} = \mu \) and covariance matrix

\[
\Sigma^{\bar{X}_T} = \frac{1}{T} \left( \sum_{l=-\infty}^{\infty} \Psi_l \right) \Sigma \left( \sum_{l=-\infty}^{\infty} \Psi_l^T \right).
\]
While the statistic $\bar{X}_T$ gives a simple estimator for the mean level of a stationary stochastic process, for many applications it is of at least the same relevance to efficiently estimate the covariance matrices $\Gamma(h)$ for various lags $h \in \mathbb{Z}$. For this purpose, we reconsider the data set $\{X_t\}_{t=1,\ldots,T}$, which was drawn from the model (2.6), where we assume without loss of generality that $\mu = 0$ (otherwise we consider the transformed data $\{X_t - \bar{X}_T\}_{t=1,\ldots,T}$). A well known and widely used estimator for the covariance matrix $\Gamma(h) = \mathbb{E}(X_{t+h}X_t^T)$ is then defined by

$$\hat{\Gamma}(h) := \begin{cases} \frac{1}{T} \sum_{t=1}^{T-h} X_{t+h}X_t^T & \text{if } 0 \leq h \leq T - 1 \\ \frac{1}{T} \sum_{t=-h+1}^{T} X_{t+h}X_t^T & \text{if } -T + 1 \leq h < 0 \end{cases}.$$ 

The following theorem shows that the estimator $\hat{\Gamma}(h)$ is consistent for $\Gamma(h)$ in the case of linear processes. For a proof of this result, see Theorem 11.2.1 in Brockwell and Davis (1991).

**Theorem 2.1.8** (Consistency of the estimator $\hat{\Gamma}(h)$)

Let $\{X_t\}_{t \in \mathbb{Z}}$ be a $\mathbb{R}^d$-valued time series having a linear representation of the form

$$X_t = \sum_{l=-\infty}^{\infty} \Psi_l Z_{t-l},$$

where $\{Z_t\}_{t \in \mathbb{Z}}$ is a sequence of independent identically distributed random vectors with covariance matrix $\Sigma$ and where the sequence $\{\Psi_t\}_{t \in \mathbb{Z}}$ of matrices fulfills

$$\sum_{l=-\infty}^{\infty} \|\Psi_l\|_\infty < \infty.$$

Then the estimator $\hat{\Gamma}(h)$ is consistent for $\Gamma(h)$ for each fixed $h \geq 0$, i.e. it holds

$$\|\hat{\Gamma}(h) - \Gamma(h)\|_\infty = o_P(1).$$

Theorem 2.1.8 assures that the estimator $\hat{\Gamma}(h)$ converges to the true covariance matrix $\Gamma(h)$ at lag $h$ for each fixed $h$. While this is a desirable property, which justifies the use of this estimator in many practical settings, we will now turn our attention to a rather different approach for statistical inference of serial dependence in time series data. To motivate this approach, we reconsider the definition of the spectral density matrix $f$ of a stationary time series, which was given in Definition 2.1.4. The fact that by Theorem 2.1.4 the second order moment structure of a time series $\{X_t\}_{t \in \mathbb{Z}}$ is completely determined by its spectral density matrix suggests the object $f$ as a starting point for further investigation. For this reason, we now focus on the study of the dependence structure of a time series in the frequency domain. In order to infer information about the spectral density and develop statistical methods for analysing its stylised facts, a means to estimate the function $f$ is required. The following definition introduces the periodogram that serves this purpose.
Definition 2.1.9 (Periodogram)
For a $\mathbb{R}^d$-valued set $\{X_t\}_{t=1,...,T}$ of data we define the periodogram $I_T(\lambda) \in \mathbb{C}^{d \times d}$ by

$$I_T(\lambda) := \frac{1}{2\pi T} \sum_{p,q=1}^{T} X_p X_q^T \exp(-i\lambda(p-q)).$$

The periodogram $I_T(\lambda)$ will be the founding block of various procedures, which will be derived throughout this thesis and are designed to examine the dependence structure of the underlying stochastic process of an observed set $\{X_t\}_{t=1,...,T}$ of time dependent data. The result below gives a first intuitive reason for considering the periodogram as a sensible object for statistical inference concerning the spectral density $f$ [see Theorem 11.7.1. in Brockwell and Davis (1991)].

Theorem 2.1.10 (Asymptotic distribution of the periodogram $I_T(\lambda)$)
Let $\{X_t\}_{t \in \mathbb{Z}}$ be a $\mathbb{R}^d$-valued time series with spectral density matrix $f$. Assume that $X_t$ has a linear representation

$$X_t = \sum_{l=-\infty}^{\infty} \Psi_l Z_{t-l},$$

where the sequence $\{\Psi_l\}_{l \in \mathbb{Z}}$ of matrices is such that

$$\sum_{l=-\infty}^{\infty} \|\Psi_l\|_\infty < \infty$$

and where $\{Z_t\}_{t \in \mathbb{Z}}$ is a sequence of independent identically distributed random vectors, which have a non-singular covariance matrix $\Sigma$ and whose components $Z_{ti}$ fulfill $E(Z_{ti}^4) < \infty$ for $i \in \{1,...,d\}$. Then the following statements hold.

(i) For each $k \in \{0,...,T\}$ and $\lambda_k = 2\pi k/T$ it holds

$$E(I_T(\lambda_k)) = f(\lambda_k) + O\left(\frac{1}{T}\right),$$

where the error term of order $O(1/T)$ is uniform in $k \in \{1,...,T\}$.

(ii) For $j,k \in \{0,...,T\}$, $\lambda_j = 2\pi j/T$, $\lambda_k = 2\pi k/T$ and $p,q,r,s \in \{1,...,d\}$ it holds

$$\text{Cov}(I_T(\lambda_j), I_T(\lambda_k)) =
\begin{cases}
[f(\lambda_j)]_{p,r} [f(\lambda_j)]_{s,q} + [f(\lambda_j)]_{p,s} [f(\lambda_j)]_{q,r} + O\left(\frac{1}{\sqrt{T}}\right) & \text{if } j = k \in \{0,T\} \\
[f(\lambda_j)]_{p,r} [f(\lambda_j)]_{s,q} + O\left(\frac{1}{\sqrt{T}}\right) & \text{if } 0 < j = k < T \\
O\left(\frac{1}{T}\right) & \text{if } j \neq k,
\end{cases}$$

where the error terms of orders $O(1/\sqrt{T})$ and $O(1/T)$ are uniform in $j$ and $k$. 

Part (i) of Theorem 2.1.10 shows that the periodogram $I_T$ is an asymptotically unbiased estimator for the spectral density $f$, which is not consistent due to part (ii). However, the fact that the periodogram estimates at different Fourier frequencies $\lambda_j \neq \lambda_k$ are asymptotically uncorrelated can be exploited to construct a consistent estimator for the spectral density at frequency $\lambda_k$. Such an estimator is defined on the basis of a local average of periodogram estimates in the frequency domain. This technique is called "smoothing" and we refer the interested reader to Brockwell and Davis (1991) for further details [see Theorem 11.7.2]. At this point, we note that this approach for constructing an estimator, which converges to the true spectral density, is widely used in various applications due to its simplicity. Nevertheless, all statistical procedures which employ some form of smoothing in the spectral domain require the statistician to specify the exact manner for taking local averages. This means that in any practical application some parameter for the bandwidth has to be specified. In most cases it is difficult to give meaningful criteria for this choice. As will become apparent in the following chapters, it is possible to circumvent this obstacle by considering specific functionals of the spectral density, which can be estimated consistently by statistics that are based on the 'raw' periodogram estimates.

2.2 Non-stationary times series models

While the framework of stationarity in time series analysis provides the practitioner with numerous methods and procedures that can be applied for statistical inference, it is frequently not easy to justify the assumption of a non-changing second order structure when considering data from a specific real world context. In some cases, this condition is obviously not fulfilled and in others it might be of necessity to somehow validate this assumption before applying any statistical method designed for stationary models. In this section, we present two classes of non-stationary models, which are able to capture stylised facts of a time-dependent second order structure and at the same time make a meaningful asymptotic theory possible. These two classes are fundamentally distinguishable by the kind of non-stationarity which they allow to be present in the data. In Section 2.2.1, we introduce the class of piecewise stationary linear time series, which broadly speaking allows the underlying stochastic process to exhibit structural breaks, but assumes that it exhibits a stationary behaviour on the various regimes. In Section 2.2.2, we then introduce the concept of local stationarity in time series. This framework allows for the second order moment structure of a time series to vary within time but makes the additional assumption that this evolution is of a smooth manner.

2.2.1 Piecewise stationary linear time series

In this section, we introduce the class of piecewise stationary linear processes, for which we have the following definition.

Definition 2.2.1 (Piecewise stationary linear time series) A sequence of centred $\mathbb{R}^d$-valued stochastic processes $\{X_{t,T}\}_{t=1,\ldots,T}$, where $X_{t,T} = (X_{t,T,1}, \ldots, X_{t,T,d})^T$ for $T \in \mathbb{N}$ $t \in \{1, \ldots, T\}$, is called piecewise stationary linear
time series, if there exists some integer $K \geq 0$, points $0 = b_0 < b_1 < \cdots < b_k < b_{K+1} = 1$ and sequences $\{\Psi_i^{(j)}\}_{i \geq 0}, j \in \{0, \ldots, K\}$, of $\mathbb{R}^{d \times d}$-valued matrices such that

\[
X_{t,T} = \begin{cases} 
\sum_{l=0}^{\infty} \Psi_l^{(0)} Z_{t-l} & \text{if } 0 = [b_0 T] < t \leq [b_1 T] \\
\sum_{l=0}^{\infty} \Psi_l^{(1)} Z_{t-l} & \text{if } [b_1 T] < t \leq [b_2 T] \\
\vdots \\
\sum_{l=0}^{\infty} \Psi_l^{(K)} Z_{t-l} & \text{if } [b_K T] < t \leq [b_{K+1} T] = T,
\end{cases}
\]

(2.8)

where $\{Z_t\}_{t \in \mathbb{Z}}$ denotes a centred White Noise process with covariance matrix $I_d$.

Note that the restriction of a unit covariance matrix of the sequence of innovations $Z_i$ can easily be relaxed and that structural breaks in the second order structure of this sequence can be included by appropriately redefining the matrices $\Psi_i^{(j)}$, $j \in \{0, \ldots, K\}$, $l \in \mathbb{Z}$. In the following, we assume that the number $K$ of structural breaks is fixed and that $K$ is minimal in the sense that for any break point $b_i$, $i \in \{1, \ldots, K\}$ it holds $\Psi_i^{(i)} \neq \Psi_i^{(i+1)}$ for at least one $l \in \mathbb{N}$. This convention implies that there is no change point in the dependency structure, if $K$ equals zero, whereas structural breaks are present whenever $K \geq 1$. For some piecewise stationary linear time series $\{X_{t,T}\}_{t=1,\ldots,T}$ of the kind (2.8), we will employ the more compact notation

\[
X_{t,T} = \sum_{l=0}^{\infty} \Psi_l(t/T) Z_{t-l} \quad t = 1, \ldots, T,
\]

(2.9)

where the functions $\Psi_l : [0, 1] \to \mathbb{R}^{d \times d}$, $l \in \mathbb{N}$, are defined by

\[
\Psi_l(u) = \sum_{j=0}^{K} \Psi_l^{(j)} \mathbb{1}_{S_j}(u)
\]

and $\mathbb{1}_{S_j}$ denotes the indicator function of the set $S_j = \{u : b_j < u \leq b_{j+1}\}$ for $j \in \{0, \ldots, K\}$.

We remark that the above definition of piecewise stationary linear time series, which relies on a double indexation of the observations $X_{t,T}$, is such that an increase of the sample size $T$ is proportionally divided between the $K + 1$ stationary segments of the underlying process. This property implies that the increase in information about the underlying distribution, which becomes available as the sample size $T$ grows, is also equally distributed over the whole time domain. As we will see, this is a key feature which makes an asymptotic theory for this kind of time series model feasible.

It is obvious that a time series model of the form (2.9) exhibits changes in its second order moment structure at the break points $b_i$, $i \in \{1, \ldots, K\}$, whenever $K > 0$. This fact can also be observed in the frequency domain: It follows from the representation (2.9) that for each $j \in \{0, \ldots, K\}$ and on the set $S_j$, the observed data $\{X_{t,T}\}_{t=1,\ldots,T}$ follows a stationary time series model with spectral density function

\[
f_j(\lambda) := \frac{1}{2\pi} \sum_{l,m=0}^{\infty} \Psi_l^{(j)}(\Psi_m^{(j)})^T \exp(-i\lambda(l-m)).
\]

(2.10)
Thus, it is straightforward to define for \( \{X_{t,T}\}_{t=1,...,T} \) the \( \mathbb{C}^{d \times d} \)-valued time-varying and piecewise constant spectral density matrix

\[
f(u, \lambda) := \frac{1}{2\pi} \sum_{l,m=0}^{\infty} \Psi_l(u)\left(\Psi_m(u)\right)^T \exp(-i\lambda(l - m)) = \sum_{j=0}^{K} f_j(\lambda)1_{S_j}(u),
\]

which locally specifies the second order moment structure of the time series model under consideration. From the definition (2.11) it follows directly that the spectral density \( f \) has discontinuities in the \( u \)-direction at the break points \( b_i \) for each \( i \in \{1,...,K\} \) whenever \( K > 0 \). In Chapter 3, we will exploit this observation in order to derive a procedure that allows to investigate the presence of structural breaks in a piecewise stationary linear time series and to consistently identify the stationary segments of the underlying process.

### 2.2.2 Locally stationary linear time series

The concept of local stationarity in time series has found considerable attention in the recent literature because many observed sets of data, which originate from diverse real world phenomena, cannot reasonably be assumed to originate from a stochastic process having a constant second order structure. Locally stationary time series models are well capable to capture stylised facts of a time-varying dependence structure and the development of procedures for statistical inference in this class of stochastic processes is thus of great importance. Following Dahlhaus and Polonik (2009) we introduce the following definition.

**Definition 2.2.2 (Locally stationary linear time series)**

A sequence of \( \mathbb{R}^d \)-valued centred stochastic processes \( \{X_{t,T}\}_{t=1,...,T} \), \( X_{t,T} = (X_{t,T,1},...,X_{t,T,d})^T \), is called a locally stationary linear time series, if for each \( T \in \mathbb{N} \) and \( t = 1,...,T \), the random vector \( X_{t,T} \) possesses a locally stationary MA(\( \infty \)) representation of the form

\[
X_{t,T} = \sum_{l=0}^{\infty} \Psi_{t,T,l}Z_{t-l},
\]

where \( \{Z_t\}_{t \in \mathbb{Z}} \) denote \( d \)-variate independent and identically distributed random variables with unit covariance matrix \( I_d \) and the coefficient matrices \( \Psi_{t,T,l}, T \in \mathbb{N}, t \in \{1,...,T\}, l \in \mathbb{N} \), are such that the following conditions hold:

(i) For all \( T \in \mathbb{N} \) and \( t \in \{1,...,T\} \), we have

\[
\sum_{l=0}^{\infty} \|\Psi_{t,T,l}\|_\infty < \infty.
\]

(ii) There exist twice continuously differentiable functions \( \Psi_l : [0,1] \to \mathbb{R}^{d \times d} \), such that

\[
\sum_{l=0}^{\infty} \sup_{t=1,...,T} \|\Psi_l(t/T) - \Psi_{t,T,l}\|_\infty = O\left(\frac{1}{T}\right)
\]

as \( T \to \infty \).
Considering the above definition, we remark that the assumption of a unit covariance matrix for the innovations \( Z_t \) is not necessary and that it is easy to incorporate the case of a time-varying covariance dependence structure. This can be achieved by appropriately redefining the matrices \( \Psi_{l,T,l} \) of the linear representation (2.12). The property (2.13), which requires that the matrices \( \Psi_{l,T,l} \) of linear factors do not change in a much too irregular fashion, assures that for sufficiently large \( T \) the time series \( \{X_{t,T}\}_{t=1,...,T} \) can be locally approximated by stationary models. For an investigation of the local second order moment structure of the time series \( \{X_{t,T}\}_{t=1,...,T} \) at some point \( u = t/T \in [0,1] \), we will focus our attention on methods which are based on the local spectral density function \( f(u, \lambda) \), which is defined by

\[
f(u, \lambda) := \frac{1}{2\pi} \sum_{l,m=0}^{\infty} \Psi_l(u) \Psi_m(u)^T \exp(-i\lambda(l-m))\]

and depends on the functions \( \Psi_l, l \in \mathbb{N} \), which approximate the coefficients of the linear representation (2.12).

\section{2.3 Auxiliary concepts}

In this section, we give a summary of fundamental mathematical concepts and methods, which we will rely on in the chapters to come. This compact overview can be divided into two parts. The first is intended to introduce the definition and basic properties of cumulants. The second part gives a brief presentation of methods for establishing weak convergence of infinitely dimensional processes.

\subsection{2.3.1 Cumulants}

In the following demonstrations, we introduce the concept of cumulants, which is a powerful tool to capture multiple stylised facts of a random distribution. Many analytical methods are founded on this concept and the following presentation of basic definitions, properties and useful applications resembles a carefully selected excerpt of the illustrations of Brillinger (1981). Throughout this section, we assume that \( Y = (Y_1, ..., Y_r) \) is a \( r \)-dimensional vector of \( \mathbb{C} \)-valued random variables, where for each component \( j \in \{1, ..., r\} \) it holds \( \mathbb{E}|Y_j|^r < \infty \).

\begin{definition} (Cumulant) \end{definition}

(i) For the \( \mathbb{C}^r \)-valued random vector \( Y \), we define the \( r \)-th cumulant \( \text{cum}(Y_1, ..., Y_r) \) by

\[
\text{cum}(Y_1, ..., Y_r) := \sum (-1)^{p-1}(p-1)!\mathbb{E}\left( \prod_{j \in \nu_1} Y_j \right) \cdots \mathbb{E}\left( \prod_{j \in \nu_p} Y_j \right),
\]

where the summation is performed over all partitions \( (\nu_1, ..., \nu_p) \), \( p = 1, ..., r \), of the set \( \{1, 2, ..., r\} \).
(ii) For a \( \mathbb{C} \)-valued random variable \( Y \) with \( \mathbb{E}|Y|^r < \infty \) and \( Y_j = Y \) for \( j \in \{1,\ldots,r\} \), we define the \( r \)-th cumulant \( \text{cum}_r(Y) \) by
\[
\text{cum}_r(Y) := \text{cum}(Y_1,\ldots,Y_r).
\]

The following theorem contains a collection of properties facilitating the treatment of cumulants of complicated statistics. We will make use of this result frequently in the technical parts of this thesis.

**Theorem 2.3.2 (Basic properties of cumulants)**

Consider a \( \mathbb{C}^r \)-valued random vector \( Y = (Y_1,\ldots,Y_r) \). The following properties for the cumulants of the vector \( Y \) hold.

1. \( \text{cum}(Y_i) = \mathbb{E}(Y_i) \) for \( i \in \{1,\ldots,r\} \).
2. \( \text{cum}(Y_i,\bar{Y}_i) = \mathbb{V}(Y_i) \) for \( i \in \{1,\ldots,r\} \).
3. \( \text{cum}(Y_i,\bar{Y}_j) = \text{Cov}(Y_i,Y_j) \) for \( i,j \in \{1,\ldots,r\} \).
4. For constants \( a_1,\ldots,a_r \in \mathbb{R} \) it holds
\[
\text{cum}(a_1Y_1,a_2Y_2,\ldots,a_rY_r) = a_1\ldots a_r \text{cum}(Y_1,\ldots,Y_r).
\]
5. \( \text{cum}(Y_1,\ldots,Y_r) \) is symmetric in all arguments.
6. If one group of the \( Y_i \)'s is independent of the other elements of the vector \( Y \), it holds
\[
\text{cum}(Y_1,\ldots,Y_r) = 0.
\]
7. Let \( Z \) denote a further \( \mathbb{C} \)-valued random element. Then it is
\[
\text{cum}(Y_1 + Z, Y_2, Y_3,\ldots,Y_r) = \text{cum}(Z,Y_2,Y_3,\ldots,Y_r) + \text{cum}(Y_1,Y_2,Y_3,\ldots,Y_r).
\]
8. Let \( c \) denote some constant. Then we have for any \( l \geq 2 \)
\[
\text{cum}(Y_1 + c,\ldots,Y_l) = \text{cum}(Y_1,\ldots,Y_l).
\]
9. Let \( X = (X_1,\ldots,X_r) \) denote a random vector that is independent from \( Y = (Y_1,\ldots,Y_r) \). Then we have
\[
\text{cum}(X_1 + Y_2,\ldots,X_r + Y_r) = \text{cum}(Y_1,\ldots,Y_r) + \text{cum}(Y_1,\ldots,Y_r).
\]

The following theorem gives an alternative means for calculating the cumulant of a collection \( Y = (Y_1,\ldots,Y_r) \) of random variables.
Theorem 2.3.3 (Alternative representation of the cumulant)
The cumulant \( \text{cum}(Y_1, \ldots, Y_r) \) is given by the coefficient of \( i^{r_1} \ldots t_r \) in the Taylor series expansion of the expression
\[
\log \left( \mathbb{E} \left( \exp \left( i \sum_{j=1}^{r} Y_j t_j \right) \right) \right)
\]
about the origin.

Next, we present a simple application of the above definitions and properties. For this purpose, we consider the cumulant structure of Gaussian random variables.

Example 2.3.1 (Cumulants of Gaussian random elements)

(i) For a Gaussian random variable \( Y \sim N(\mu, \sigma^2) \) it follows from Theorem 2.3.2 that \( \text{cum}_1(Y) = \mu \) and \( \text{cum}_2(Y) = \sigma^2 \). Furthermore, an application of Theorem 2.3.3 yields that \( \text{cum}_l(Y) = 0 \) if \( l \geq 3 \).

(ii) Let \( Y = (Y_1, \ldots, Y_r) \) denote a \( r \)-dimensional Gaussian random vector with expected value vector \( \mu = (\mu_1, \ldots, \mu_r) \) and covariance matrix \( \Sigma \). In this case Theorem 2.3.2 yields that \( \text{cum}(Y_i) = \mu_i \) for \( i \in \{1, \ldots, r\} \) and \( \text{cum}(Y_i, \bar{Y}_j) = \text{Cov}(Y_i, Y_j) \) for \( i, j \in \{1, \ldots, r\} \). Moreover, Theorem 2.3.3 and linearity arguments show that for any \( l \geq 3 \) and any subset \( \{i_1, \ldots, i_l\} \subset \{1, \ldots, r\} \) of indices it holds \( \text{cum}(Y_{i_1}, \ldots, Y_{i_l}) = 0 \).

The distinct cumulant structure of a Gaussian random vector motivates the following theorem, which constitutes a powerful tool for proving asymptotic normality for a sequence of finite dimensional random elements.

Theorem 2.3.4 (Method of cumulants)
Let \( \{Y_t\}_{t \geq 1} \), \( Y_t = (Y_{1t}, \ldots, Y_{rt}) \), be a sequence of \( \mathbb{C}^r \)-valued random elements, where for \( t \geq 1 \) \( Y_t \) has mean value vector \( \mu_t = (\mu_{1t}, \ldots, \mu_{rt}) \) and covariance matrix \( \Sigma_t = [\Sigma_{i,j}]_{i,j=1}^{r} \). Furthermore, let \( Y = (Y_1, \ldots, Y_r) \) denote a Gaussian random vector with mean value vector \( \mu = (\mu_1, \ldots, \mu_r) \) and covariance matrix \( \Sigma = [\Sigma(i,j)]_{i,j=1}^{r} \) and assume that the following properties hold:

(i) \( \lim_{t \to \infty} \text{cum}(Y_{it}) = \mu_i \) for all \( i \in \{1, \ldots, r\} \).

(ii) \( \lim_{t \to \infty} \text{cum}(Y_{it}, \bar{Y}_{jt}) = \Sigma(i,j) \) for all \( i, j \in \{1, \ldots, r\} \).

(iii) \( \lim_{t \to \infty} \text{cum}(Y_{i_1t}, Y_{i_2t}, \ldots, Y_{i_lt}) = 0 \) for all \( l \geq 3 \) and any subset \( \{i_1, \ldots, i_l\} \subset \{1, \ldots, r\} \).

Then it follows \( Y_t \Rightarrow Y \).
Theorem 2.3.4 states that weak convergence of a sequence of random vectors to a multivariate Gaussian distribution can be established by showing that all cumulants of the sequence converge to the respective cumulants of the Gaussian limiting distribution. We will employ this result frequently throughout the technical proofs of the main results in later sections. As for now, we emphasise that for an application of Theorem 2.3.4 it is necessary to establish conditions (i), (ii) and (iii), which can be quite cumbersome if the components of the sequence \( \{Y_t\}_{t \geq 1} \) are of a complicated structure. In these situations we will frequently employ the various parts of Theorem 2.3.2. To further facilitate these arguments, we now state an important result which concerns the calculation of cumulants of products of random variables. For this purpose we first define the term indecomposability for a partition of some table of indices.

**Definition 2.3.5 (Indecomposability)**

We consider a table of the form

\[
(1,1) \quad (1,2) \quad \ldots \quad (1,k_1) \\
\vdots \quad \vdots \quad \quad \quad \quad \quad \quad \vdots \\
(m,1) \quad (m,2) \quad \ldots \quad (m,k_m)
\]  

(2.14)

featuring \( m \) rows of \( k_1, \ldots, k_m \) elements respectively. For some \( 2 \leq M \) and a partition \( \nu = \{\nu_1 \cup \nu_2 \cup \ldots \cup \nu_M\} \) we define the following properties:

1. Two elements \( \nu_k \) and \( \nu_l \) of the partition \( \nu \) hook, if there exist elements \( (i_1, j_1) \in \nu_k \) and \( (i_2, j_2) \in \nu_l \) such that \( i_1 = i_2 \).

2. Two elements \( \nu_m \) and \( \nu_n \) of the partition \( \nu \) communicate, if there exists a sequence \( \nu_m = \nu_{i_1}, \ldots, \nu_{i_h} = \nu_n \) of elements of \( \nu \) such that \( \nu_{i_j} \) and \( \nu_{i_{j+1}} \) hook for \( j = 1, 2, \ldots, h - 1 \).

3. A partition \( \nu = \{\nu_1 \cup \nu_2 \cup \ldots \cup \nu_M\} \) is said to be indecomposable, if \( \nu_{i_1} \) and \( \nu_{i_j} \) communicate for all \( i, j \in \{1, \ldots, M\} \).

Having introduced the concept of indecomposability for a table of indices, we are now able to give the famous product theorem for cumulants.

**Theorem 2.3.6 (Product theorem for cumulants)**

We consider a table of random variables

\[
Y_{1,1} \quad Y_{1,2} \quad \ldots \quad Y_{1,k_1} \\
\vdots \quad \vdots \quad \quad \quad \quad \quad \quad \vdots \\
Y_{m,1} \quad Y_{m,2} \quad \ldots \quad Y_{m,k_m}
\]  

(2.15)

corresponding to the table (2.14) of indices. Furthermore, we define the column-wise product of random variables in this table by

\[
X_i := \prod_{j=1}^{k_i} Y_{i,j}
\]
for $i = 1, \ldots, 2$. Then we have the representation
\[
cum(X_1, \ldots, X_m) = \sum_{(\nu_1, \ldots, \nu_p)} \prod_{j=1}^{p} \cum(Y_{i,j}; (i,j) \in \nu_j),
\]
where the sum in (2.16) is performed over all $p \in \mathbb{N}$ and all indecomposable partitions $(\nu_1, \ldots, \nu_p)$ of the table (2.14).

The product theorem for cumulants will be very helpful for establishing the conditions (ii) and (iii) of Theorem 2.3.4 in the situations which we will encounter in Chapters 3 and 4. In order to provide a first simple illustration, we consider the following example.

**Example 2.3.2**

(i) We consider \(\mathbb{R}\)-valued random variables \(Y_{1,1}, Y_{1,2}, Y_{2,1}\) and \(Y_{2,2}\) and are interested in calculating \(\text{Cov}(Y_{1,1}Y_{1,2}, Y_{2,1}Y_{2,2})\). From Theorem 2.3.2 (iii) it follows that
\[
\text{Cov}(Y_{1,1}Y_{1,2}, Y_{2,1}Y_{2,2}) = \cum(Y_{1,1}Y_{1,2}, Y_{2,1}Y_{2,2}).
\]
From the product theorem for cumulants it can be easily seen that
\[
\text{Cov}(Y_{1,1}Y_{1,2}, Y_{2,1}Y_{2,2}) = \sum_{\nu=(\nu_1, \ldots, \nu_p)} \prod_{j=1}^{p} \cum(Y_{i,j}; (i,j) \in \nu_j),
\]
where the sum is performed over all indecomposable partitions \(\nu\) of the table
\[
(1,1) (1,2)
(2,1) (2,2).
\]

(ii) We continue the demonstrations of (i) and additionally assume that the random variables \(Y_{1,1}, Y_{1,2}, Y_{2,1}\) and \(Y_{2,2}\) are centred. Under this condition, the treatment of the term (2.17) becomes slightly less cumbersome, as the sum over \(\nu = (\nu_1, \ldots, \nu_p)\) only needs to be performed over such partitions \(\nu\), for which it holds \(|\nu_j| \geq 2\) for each \(j\). The indecomposable partitions \(\nu\) of the table (2.18) satisfying \(|\nu_j| \geq 2\) for all \(j\) are given by
\[
\{(1,1),(2,2)\} \cup \{(1,2),(2,1)\},
\{(1,1),(2,1)\} \cup \{(1,2),(2,2)\},
\{(1,1),(1,1),(2,1),(2,2)\}
\]
and we thus obtain
\[
\text{Cov}(Y_{1,1}Y_{1,2}, Y_{2,1}Y_{2,2}) = \cum(Y_{1,1}, Y_{2,2}) \cum(Y_{1,2}, Y_{2,1}) + \cum(Y_{1,1}, Y_{1,2}) \cum(Y_{2,1}, Y_{2,2})
+ \cum(Y_{1,1}, Y_{1,2}, Y_{2,1}, Y_{2,2}).
\]

(iii) We proceed with the considerations of part (ii) of this example and additionally assume that the random variables \(Y_{1,1}, Y_{1,2}, Y_{2,1}\) and \(Y_{2,2}\) are normally distributed. From Example 2.3.1 it then follows that \(\cum(Y_{1,1}, Y_{1,2}, Y_{2,1}, Y_{2,2}) = 0\) and hence we have
\[
\text{Cov}(Y_{1,1}Y_{1,2}, Y_{2,1}Y_{2,2}) = \cum(Y_{1,1}, Y_{2,2}) \cum(Y_{1,2}, Y_{2,1}) + \cum(Y_{1,1}, Y_{1,2}) \cum(Y_{1,2}, Y_{2,2}).
\]
2.3 Auxiliary concepts

2.3.2 Weak convergence of empirical processes

Throughout the course of this section, we will state a selection of results pertaining to the weak convergence of processes of infinite dimension, which will be of crucial importance for the proofs in this thesis. The subsequent presentation is closely aligned to the monograph of van der Vaart and Wellner (1996).

For the considerations below, we denote by $S$ an arbitrary set and by $l^\infty(S)$ the set of all mappings $z : S \to \mathbb{R}^{d \times d}$ such that $\|z\|_S := \sup_{s \in S} \|z(s)\|_\infty < \infty$, where $d$ denotes some positive integer and $\| \cdot \|_\infty$ the maximum norm. By introducing the distance measure $d(z_1, z_2) := \|z_1 - z_2\|_\infty$ for $z_1, z_2 \in l^\infty(S)$, we obtain that the tuple $(l^\infty(S), d)$ is a metric space. Furthermore, we consider a sequence $\{(\Omega_t, F_t, P_t)\}_{t \geq 1}$ of probability spaces and a sequence $\{X_t\}_{t \geq 1}$ of random elements of the metric space $(l^\infty(S), d)$. In order to characterise weak convergence of the sequence $\{X_t\}_{t \geq 1}$ to a Borel measure, we follow van der Vaart and Wellner (1996) and introduce the following definitions.

**Definition 2.3.7 (Outer probability and expectation)**

Consider an arbitrary probability space $(\Omega, F, P)$.

(i) We define for some set $B \subset \Omega$ the outer probability $P^*(B)$ by

$$P^*(B) := \inf \{ P(A) | B \subset A, A \in F \}.$$ 

(ii) For some mapping $X : \Omega \to \mathbb{R}$ we define the outer expectation $E^*(X)$ of $X$ by

$$E^*(X) := \inf \{ E(U) | U \geq X, U : \Omega \to \mathbb{R} \text{ is measurable and } E(U) < \infty \}.$$ 

The following definition states the abstract concept of weak convergence for a sequence $\{X_t\}_{t \geq 1}$ of $(l^\infty(S), d)$-valued random elements.

**Definition 2.3.8 (Weak convergence)**

A sequence $\{X_t\}_{t \geq 1}$ of $(l^\infty(S), d)$-valued random elements converges weakly to a Borel measure $L$, i.e.

$$X_t \Rightarrow L$$

if for every continuous and bounded function $f : l^\infty(S) \to \mathbb{R}$ it holds

$$E^*(f(X_t)) \to \int f dL.$$
This definition of weak convergence is very similar to the definition in the case of Borel measurable random elements $X_t$. However, the latter more commonly known definition based on moments of the form $E(f(X_t))$ is not transferable to the case of random elements $X_t$ that attain values in $l^\infty(S)$. This is due to the fact that a mapping $X_t$ of the form (2.19) is in general not Borel measurable, i.e. we need a moment concept which overcomes the possibility that there exist Borel sets $L \subset l^\infty(S)$ with $X_t^{-1}(L) \notin \mathcal{F}_t$. The definitions of outer probability and outer expectation accomplish exactly this.

The following result shows how, in certain situations, weak convergence of continuous functionals of a sequence of processes follows from weak convergence of the sequence itself. We will need this theorem later for the deviation of the asymptotic distribution of certain test statistics which are derived from context-specific empirical processes.

**Theorem 2.3.9 (Continuous mapping theorem)**

Let $g : l^\infty(S) \rightarrow \mathbb{R}$ denote a function that is continuous on a subset $\bar{l} \subset l^\infty(S)$ and $\{X_t\}_{t \geq 1}$ denote a sequence of $\bar{l}$-valued random elements which converges weakly to some random element $X \in \bar{l}$, i.e. $X_t \Rightarrow X$. Then it holds

$$g(X_t) \Rightarrow g(X).$$

We now intend to present the foundation for proving weak convergence as defined in Definition 2.3.8. For this purpose, we introduce the following definitions.

**Definition 2.3.10 (Totally bounded set)**

A semi-metric space $(S,d)$ is totally bounded if for any $\varepsilon > 0$ there exists an integer $n$ and a set $\{s_1, \ldots, s_n\}$ of elements $s_i \in S$ such that the condition

$$\min_{i \in \{1, \ldots, n\}} d(s, s_i) < \varepsilon$$

holds for each element $s \in S$.

**Definition 2.3.11 (Asymptotic stochastic equicontinuity)**

Let $(S,d)$ denote a semi-metric space. A sequence $\{X_t(s)\}_{s \in S}$ of stochastic processes is called asymptotically stochastically equicontinuous with respect to the semi-metric $d$, if for every $\varepsilon, \eta > 0$ there exists some $\delta > 0$ such that

$$\limsup_{t \rightarrow \infty} P^*(\sup_{x,y,d(x,y) < \delta} |X_t(x) - X_t(y)| > \eta) < \varepsilon.$$

We are now able to present a fundamental result, which in general terms states that weak convergence of a sequence $\{X_t\}_{t \geq 1}$ of $l^\infty(S)$-valued random variables to some random element $X$ follows, if one establishes asymptotic stochastic equicontinuity for the sequence $\{X_t\}_{t \geq 1}$ and shows that the finite dimensional distributions of $X_t$ converge weakly to those of the element $X$. 
Theorem 2.3.12 (Characterisation of weak convergence)

Let \( \{X_t\}_{t \geq 1} \), \( X_t : \Omega_t \to l^\infty(S) \), denote a sequence of random processes and let \( X \) be a random element with realisations in \( l^\infty(S) \) such that the following properties hold:

(i) The finite dimensional projections of the sequence \( \{X_t\}_{t \geq 1} \) converge weakly to the finite dimensional projections of the element \( X \), i.e. for each \( k \in \mathbb{N} \) and all subsets \( \{s_1, ..., s_k\} \subset S \) it holds

\[
(X_t(s_i))_{i=1,...,k} \Rightarrow (X(s_i))_{i=1,...,k}.
\]

(ii) There exists some semi-metric \( d \) on \( S \) such that \( S \) is totally bounded with respect to \( d \) and the sequence \( \{X_t\}_{t \geq 1} \) of stochastic processes is asymptotically stochastic equicontinuous with respect to \( d \).

Then we have \( X_t \Rightarrow X \).

The theorem above and a proof thereof can be found in a similar form in van der Vaart and Wellner (1996) [see Theorems 1.5.4 and 1.5.7]. We will repeatedly use this characterisation of weak convergence of \( l^\infty(S) \)-valued random elements throughout this thesis and at this point intend to emphasise that the merit lies in its applicability. In order to formally establish weak convergence of a sequence of \( l^\infty(S) \)-valued random elements, it is sufficient to show that the sequence is asymptotically stochastic equicontinuous and that its finite dimensional projections converge to those of the postulated limiting distribution. With a view to the arguments, which will be provided in the following chapters, we already mention at this point that the convergence of the finite dimensional projections of the processes under consideration will be shown by an application of the method of cumulants [see Theorem 2.3.4]. We now state a further fundamental result pertaining to the asymptotic behaviour of a sequence of stochastic processes that is quite similar to the claim of Theorem 2.3.12 and concerns uniform convergence in probability [see. Newey (1991)].

Theorem 2.3.13 (Characterisation of uniform convergence in probability)

Let \( S \) denote some compact set and let \( \{X_t\}_{t \geq 1} \), \( X_t : \Omega_t \to l^\infty(S) \), denote a sequence of stochastic processes. Furthermore, assume that the following properties hold:

(i) The process \( \{X_t(s)\}_{s \in S} \) converges pointwise to zero in probability, i.e. for each \( s \in S \) it holds

\[
\|X_t(s)\|_\infty = o_P(1)
\]

as \( t \to \infty \).

(ii) The sequence \( \{X_t\}_{t \geq 1} \) is asymptotically stochastically equicontinuous with respect to some semi-metric \( d \) on \( S \).

Then it holds

\[
\sup_{s \in S} \|X_t(s)\|_\infty = o_P(1).
\]
In the situations, which we will encounter in later chapters, the most involved part in the application of Theorems 2.3.12 and 2.3.13 will usually lie in the proof of asymptotic stochastic equicontinuity of the sequence \( \{X_t\}_{t \geq 1} \). In order to facilitate the arguments needed for this part we now introduce the method of chaining, which will turn out to be very useful for bounding probabilities of the kind

\[
P^* \left( \sup_{x,y; d(x,y) < \delta} |X_t(x) - X_t(y)| > \eta \right)
\]

and thus constitutes an important tool for establishing asymptotic stochastic equicontinuity. Therefore, we first define some technical terms, which capture the degree of complexity of the index set \( S \) to some extent [see Pollard (1984)].

**Definition 2.3.14 (Covering numbers and covering integral)**

Let \((S, d)\) denote some semi-metric space.

(i) For some \( u > 0 \) we denote by \( N(u, d, S) \) the smallest \( u \)-net of the set \( S \) with respect to the semi-metric \( d \), i.e. \( N(u, d, S) \) equals the smallest integer \( m \) such that there exists a subset \( \{s_1, \ldots, s_m\} \subset S \) satisfying

\[
\min_{i \in \{1, \ldots, m\}} d(s, s_i) \leq u
\]

for all \( s \in S \).

(ii) For some \( \kappa > 0 \), we define the covering integral \( J(\kappa, d, S) \) of the set \( S \) with respect to the semi-metric \( d \) by

\[
J(\kappa, d, S) := \int_0^\kappa \left[ \log \left( \frac{48N(u, d, S)^2}{u} \right) \right]^2 du.
\]

Having introduced the basic concept of covering numbers and integrals, we are now able to give a slightly modified version of the chaining lemma [see Pollard (1984) section VII], which will emerge to be very useful for establishing asymptotic stochastic equicontinuity in later chapters.

**Theorem 2.3.15 (Chaining lemma)**

We consider a stochastic process \( \{Z(s)\}_{s \in S} \), where the index set \( S \) possesses a finite covering integral \( J(\kappa, d, S) \) for some semi-metric \( d \) and any \( \kappa > 0 \). Furthermore, we assume that there exists some positive constant \( D \) such that for all \( s, t \in S \) and for all \( \nu > 0 \) the inequality

\[
P\left( |Z(s) - Z(t)| > \nu d(s, t) \right) \leq 96 \exp \left( - \sqrt{\frac{\nu}{D}} \right) \tag{2.20}
\]

holds. Then there exists a countable and dense subset \( S^* \subset S \) such that

\[
P\left( \exists s, t \in S^* \text{ such that } d(s, t) < \varepsilon \text{ and } |Z(t) - Z(s)| > 26DJ(d(s, t), d, S) \right) \leq 2\varepsilon \tag{2.21}
\]

for all \( \varepsilon \in (0, 1) \). For a stochastic process \( \{Z(s)\}_{s \in S} \) featuring continuous sample paths the property (2.21) holds with set \( S^* \) being replaced by \( S \).
Chapter 3

A new approach to the change point problem in time series analysis

Due to the broad availability of methods for statistical inference in stationary processes, the assumption of stationarity has become a cornerstone for the empirical analysis of time-dependent data in the applied science. There exists a large amount of results that can be employed for the statistical analysis of stationary time series, like for example parameter estimation and forecasting techniques. These methods are known to yield reliable results when applied to sets of data generated by a stochastic process, whose stylised facts are time-invariant. However, for many real-world phenomena observed in various disciplines this assumption turns out to be too restrictive because empirical data indicate that the distributional properties of the data generating system exhibit occasional shifts in time. Therefore, the class of stationary processes is not able to capture a satisfactory level of complexity, which is present in the observed data. For this reason, the development and the investigation of alternative models allowing for more flexibility in the stylised facts of the underlying process is of high relevance for empirical research and many contributions to the scientific literature are concerned with this problem. Generally speaking, changes in the distribution of the underlying stochastic system can materialise in various manners and early research focused on shifts in the mean level of an observed set of data [see for example Sen and Srivastava (1975)]. Since these times, the investigation of changes in the mean level has attracted much attention in empirical and methodological contributions to the literature. For example, Banerjee et al. (1992) consider economic indicators for various OECD countries and investigate the hypothesis that these data sets can be adequately described by models which are stationary around broken trend lines. In James et al. (1987), the authors develop an early approach for formally testing for the presence of structural breaks in the mean level of a set of independent and homoscedastic normal random variables. Bai (1994) proposes a method for estimating the location of a single change point in the mean of a linear process and derives the consistency of the procedure. The problem of testing the presence of structural breaks in the mean level of an observed set of data and the subsequent estimation of the locations, where shifts occur, is rather well established in the literature. More recently, the investigation of switches in the regime materialising in the second order structure of the process has found much more attention. In this regard, the
analysis of changes in the variance has been of large interest due to the high importance of this parameter for various applications and especially in financial market analysis. In the context of empirical finance, an accurate estimation of the variability of asset returns is crucial for risk management purposes due to the fact that volatility measures are important input factors for the pricing of derivatives and the construction of risk and return efficient portfolios (see for example Wilmott (2006)). For many derivative contracts, the sensitivity with respect to the return variability of the underlying asset is large in comparison to other influencing factors. Thus, accurately measuring the variance of asset returns is of high importance for financial institutions. For this reason, most contributions to the empirical literature, which are concerned with the investigation of changes in the variance, consider financial market data. Fryzlewicz et al. (2006) investigate financial log returns and propose to model these data by piecewise stationary models of independent realisations having a time-dependent variance function, which is piecewise constant and exhibits an unknown number of shifts. A wavelet thresholding algorithm is proposed to estimate the parameters of the model and it is shown that this kind of processes is well suited to capture changing distributional properties of financial market returns. A similar approach is followed in Starica and Granger (2005), where a model of piecewise stationary unconditional variance is suggested for modelling stock market returns. In an empirical study covering S&P 500 returns, it is demonstrated that forecasts based on this approach outperform competing methods based on stationary long memory or Garch(1,1) models. Mikosch and Starica (2000) also investigate return patterns of stock markets and conclude that the long range dependence behaviour of these returns, which is postulated by various researchers [see for example Ding and Granger (1996), Ding et al. (1993) and Granger and Ding (1996) among many others], could merely be an artefact of changes in the unconditional variance of an GARCH($p,q$) model. Similar comparisons can be found in the paper by Granger and Hyung (2004), where an empirical analysis of the absolute returns of the S&P 500 index is provided, and in Chen et al. (2010), where a procedure for locally estimating the variance of asset returns is shown to yield better results than competing models with long memory features when applied to stock market returns. Due to this high degree of practical relevance of auto covariance instability, many analytical methods for testing the presence and estimating the location of structural breaks in the second order structure of time series data have been developed. In general, these approaches can be divided into parametric and non-parametric procedures. Parametric methods include the approaches of Carlin et al. (1992) and Page (1954), which are based on log-likelihood techniques. More recently, Davis et al. (2006) suggested a procedure which is applicable to piecewise stationary AR($p$) models. The method proposed by these authors combines a log-likelihood technique and a generic algorithm to estimate the unknown number of change points and to subsequently estimate the orders and parameters of the AR models on the respective segments. Non-parametric approaches to change point detection are preferable in many situations because the conclusions reached by their application rely on less restrictive assumptions. Inclan and Tiao (1994) explain a CUSUM-type procedure for testing the presence of possibly multiple structural breaks in the variance of a sequence of independent normally distributed random variables. Lee and Park (2001) extend this method to the case of linear processes. Lee et al. (2003) suggest to consider a test statistic which is based on a scaled sum of discrepancies of the first $m$ empirical auto covariances estimated on different
segments of the observed data in order to develop a test which detects changes in the auto covariance structure. More recently, Aue et al. (2009) proposed a testing procedure, which can be applied to a wide range of multivariate time series and allows to test the presence of a structural break in the covariance matrix of the components of the observed data.

This review of existing literature, which is by no means complete, clearly shows that the change point problem is of high relevance for practical reasons and demonstrates the diversity of approaches, which have been exploited to address this issue. A more extensive research of the scientific literature shows that the selection above is representative in the sense that there exists a general bias towards considering the univariate case. This observation is quite surprising because in many real world phenomena one observes multivariate sets of time-dependent data, which are often characterised by complex interdependencies between the univariate components. In the context of parameter estimation and forecasting, this paradigm is well established and a wide spectrum of techniques and applicable results have been designed, which allow to answer various questions in the multivariate setting as well. However, in the field of break point detection the literature is much less developed for the multivariate case and to the best of our knowledge the paper by Aue et al. (2009) is the only contribution containing a formal non-parametric testing procedure for the presence of regime shifts in the second order structure of multivariate time series. It is notable that this method is designed for detecting breaks in the covariance structure, which is a clear advancement in comparison to previous results. Nevertheless, by merely considering the covariance matrix this approach still only tracks deviations from stationarity in a small part of the second order structure of a multivariate time series and it is not able to detect structural changes, which only materialise in the auto covariances of higher orders. Thus, we conclude that the problem of testing for the presence and estimating the location of change points in the auto covariance structure of multivariate time series is a very important topic, for which until recently only very limited statistical methods have been available. The remainder of this chapter is devoted to the introduction of a new approach to the change point problem, which addresses this issue. The demonstrations are based on Preuß et al. (2013) and the organisation of the presentation is as follows. In Section 3.1, we develop a frequency domain based procedure, which allows to test the presence of possibly multiple structural breaks in the auto covariance structure of a multivariate time series. Section 3.2 considers estimation problems in the change point context. Here, we derive an algorithmic procedure which allows to consistently divide the sample into stationary segments and is able to identify the components of the spectral density matrix which exhibit discontinuities at the respective break points. We conclude this chapter by giving an overview of issues concerning the implementation of the described methods and investigate the quality of the procedures when applied to data. These demonstrations are presented in Section 3.3 and include the explanation of data driven criteria for fine tuning the regularising parameters and the evaluation of finite sample properties. All proofs and technical details are given in Section 3.4.
3.1 Testing for the presence of structural breaks

In this section, we present a new procedure for testing the presence of structural breaks in multivariate piecewise stationary time series.

3.1.1 Motivation and definitions

We assume to observe realisations of a centred $\mathbb{R}^d$-valued stochastic process $\{X_{t,T}\}_{t=1,...,T}$, where $X_{t,T} = (X_{t,T,1},...,X_{t,T,d})^T$ possesses a piecewise stationary representation of the form (2.9). As was explained in Section 2.2.1, this representation implies that the spectral density matrix $f(u,\lambda)$ exhibits points of discontinuity in $u$-direction at the break points $b_i$ ($i = 1,...,K$) whenever $K \geq 1$. In order to investigate the presence of a structural break at some point $v \in [0,1]$, we intend to compare the local averages in $u$-direction of the spectral density matrix $f$ on two consecutive intervals $[v-e,v]$ and $[v,v+e]$ of length $e$ surrounding the point $v$. Therefore, we consider the functions

$$
\lambda \mapsto \frac{1}{e} \int_{v-e}^{v} f(u,\lambda)du \quad \text{and} \quad \lambda \mapsto \frac{1}{e} \int_{v}^{v+e} f(u,\lambda)du,$$

which are identical, if $f$ is constant in the time direction $u$ on the interval $[v-e,v+e]$. In order to quantify the size of a potential structural break at time $v \in [e,1-e]$, we define the matrix

$$
E(v) := \left[ \sup_{\omega \in [0,1]} \frac{1}{e} \int_{\omega \pi}^{\omega \pi + v} f(u,\lambda)[a,b]dud\lambda - \int_{\omega \pi}^{\omega \pi + v-e} f(u,\lambda)[a,b]dud\lambda \right]_{a,b=1}^d,
$$

which gives a component wise measure for the local discontinuity of $f$ at time $v$. Note that at least one component of the matrix $E(v)$ is positive for each $v \in \{b_1,...,b_K\}$ whereas it holds

$$
\|E(v)\|_\infty = 0 \quad \text{for} \quad v \in [0,1] \setminus \{b_1,...,b_K\}
$$
as $e \to 0$. As we are interested in developing a test for the presence of structural breaks, we need a measure, which globally tracks discontinuities of the spectral density matrix $f(u,\lambda)$ in $u$-direction. For this purpose, we consider the quantity

$$
E := \sup_{v \in [0,1]} \|E(v)\|_\infty = \sup_{v,\omega \in [0,1]} \|E(v,\omega)\|_\infty,
$$

where for $v \in [e,1-e]$ and $\omega \in [0,1]$ the matrix $E(v,\omega) \in \mathbb{R}^{d \times d}$ is defined by

$$
E(v,\omega) := \frac{1}{e} \left( \int_{\omega \pi}^{\omega \pi + v} f(u,\lambda)dud\lambda - \int_{\omega \pi}^{\omega \pi + v-e} f(u,\lambda)dud\lambda \right)
$$

and for $v \notin [e,1-e]$ we set

$$
E(v,\omega) := \begin{cases} 
E(e,\omega) & \text{if } v \leq e \\
E(1-e,\omega) & \text{if } v \geq 1 - e
\end{cases}.
$$
Under the null hypothesis

\[ H_0 : K = 0 \]  \hspace{1cm} (3.3)

of no structural breaks, we have \( E = 0 \), while the quantity \( E \) is strictly positive if structural breaks are present. In order to obtain a test for the null hypothesis (3.3) against the alternative

\[ H_1 : K \geq 1, \]  \hspace{1cm} (3.4)

it is natural to determine an estimator \( \hat{E}_T \) for the deterministic quantity \( E \) and to reject the null hypothesis whenever this estimator attains 'large' values. For developing a statistical test which allows to prove the existence of structural breaks at a controlled type I error, we proceed as follows: In Section 3.1.2, we introduce for \((v, \omega) \in [0, 1]^2\) a statistic \( \hat{E}_T(v, \omega) \), which is an asymptotically unbiased estimator for the deterministic matrix \( E(v, \omega) \). We consider the supremum of the empirical process \( \{ \hat{E}_T(v, \omega) \}_{(v,\omega) \in [0,1]^2} \), which we will denote by \( \hat{E}_T \), as a test statistic for the hypothesis (3.3). In Section 3.1.3, we continue our discussion with an in-depth investigation of the asymptotic behaviour of the empirical process \( \{ \hat{E}_T(v, \omega) \}_{(v,\omega) \in [0,1]^2} \) and derive results pertaining to the weak convergence of an appropriately scaled version of this empirical process under the null hypothesis (3.3). As we will see, the continuous mapping theorem immediately provides us with the asymptotic distribution of the test statistic \( \hat{E}_T \). However, it will become apparent that this distribution depends on the unknown spectral density matrix \( f \) of the original time series data \( \{X_{t,T}\}_{t=1,...,T} \), which implies that its quantiles are not readily available. Therefore, this result does not yield a straightforward means of constructing a formal testing procedure. In Section 3.1.4, we develop an algorithmic procedure that solves this problem in an efficient way. In this regard, we explain how resampling methods can be used to accurately estimate the unknown quantiles of the limiting distribution of the test statistic \( \hat{E}_T \) and we formally establish the asymptotic accuracy of a bootstrap based test for the null hypothesis (3.3).

### 3.1.2 An empirical process tracking discontinuity in the time direction

In order to obtain an estimator of the quantity \( E \) defined in (3.1), we first introduce an estimator for the matrix \( E(v, \omega) \), \((v, \omega) \in [0, 1]^2\) defined in (3.2). Therefore, we consider some sequence \( N = N_T \) of even integers satisfying \( N_T/T \to \epsilon \) and consider for \( u \in [0, 1] \) the local periodogram

\[ I_N(u, \lambda) := \frac{1}{2\pi N} \sum_{r,s=0}^{N-1} X_{\lfloor uT \rfloor - N/2+1+s,T}X_{\lfloor uT \rfloor - N/2+1+r,T}^T \exp(-i\lambda(s-r)). \hspace{1cm} (3.5) \]

Note that the periodogram (3.5) is computed from the \( N \) observations \( \{X_{\lfloor uT \rfloor - N/2+1+1,T}, \ldots, X_{\lfloor uT \rfloor + N/2,T}\} \) surrounding the observation \( X_{\lfloor uT \rfloor} \) at time \( \lfloor uT \rfloor \). An estimator for the matrix \( E(v, \omega) \) is then defined by

\[ \hat{E}_T(v, \omega) := \frac{1}{N} \sum_{k=-\lfloor \omega(N-1)/2 \rfloor}^{\lfloor \omega N/2 \rfloor} \left( I_N(v + N/(2T), \lambda_k) - I_N(v - N/(2T), \lambda_k) \right). \hspace{1cm} (3.6) \]
if \( v \in [N/T, 1 - N/T] \), where \( \lambda_k = 2\pi k/N, \) \( k \in \mathbb{N} \), denote the Fourier frequencies to the base \( N \). For \( v \notin [N/T, 1 - N/T] \) we define

\[
\hat{E}_T(v, \omega) := \begin{cases} \hat{E}_T(\frac{N}{T}, \omega) & \text{if } v \in [0, \frac{N}{T}) \\ \hat{E}_T(1 - \frac{N}{T}, \omega) & \text{if } v \in (1 - \frac{N}{T}, 1] \end{cases}
\]

Considering the definition of the quantity \( E(v, \omega) \) given in (3.2), we note that the basic idea underlying the construction of the estimator \( \hat{E}_T(v, \omega) \) is to replace the integral in \( \lambda \)-direction by a Riemann sum, where the averaged time varying spectral density matrices

\[
\frac{1}{c} \int_{v}^{v+e} f(u, \lambda) du \quad \text{and} \quad \frac{1}{c} \int_{v-e}^{v} f(u, \lambda) du
\]

on the intervals \([v, v + e]\) and \([v - e, v]\) are replaced by the local periodograms

\[
I_N(v + N/(2T), \lambda) \quad \text{and} \quad I_N(v - N/(2T), \lambda),
\]

which are computed using the \( N \) observations surrounding the observations \( X_{[vT] + N/2} \) and \( X_{[vT] - N/2} \) respectively. Now, it is natural to estimate the quantity \( E \), which globally measures the presence of structural breaks and was defined in (3.1), with the statistic

\[
\hat{E}_T := \sup_{(v, \omega) \in [0,1]^2} ||\hat{E}_T(v, \omega)||_{\infty} = \max_{v \in [N/T,1-N/T]} \sup_{\omega \in [0,1]} ||\hat{E}_T(v, \omega)||_{\infty}. \tag{3.7}
\]

Intuitively, under the null hypothesis (3.3) we would expect the estimator \( \hat{E}_T \) to attain small values, while large realisations of \( \hat{E}_T \) indicate some kind of discontinuity of the spectral density \( f \).

The following section is devoted to an investigation of the asymptotic properties of the process \( \{\hat{E}_T(v, \omega)\}_{(v, \omega) \in [0,1]^2} \) and the statistic \( \hat{E}_T \) under the null hypothesis (3.3) of no structural breaks and the alternative (3.4) of at least one structural break.

### 3.1.3 Weak convergence of the empirical process

In this section, we present fundamental results concerning the asymptotic distribution of the empirical process \( \{\hat{E}_T(v, \omega)\}_{(v, \omega) \in [0,1]^2} \). For this purpose we distinguish two scenarios for the asymptotic size of the block length \( N = N_T \), on which the local spectral estimates of the matrix \( f \) are based. More precisely, we consider the following settings:

1. There exists a constant \( c \geq 2/ \min_{i=1,\ldots,K+1} |b_i - b_{i+1}| \) such that

\[
\frac{N_T}{T} \to \frac{1}{c}, \tag{3.8}
\]

as \( T \to \infty \).

2. There exists some \( \epsilon > 0 \) such that

\[
\frac{N_T}{T} \to 0 \quad \text{and} \quad \frac{T^\epsilon}{N_T} \to 0 \tag{3.9}
\]

as \( T \to \infty \).
The first setting (3.8) assumes that the sequence $N$ is chosen such that $N$ is of the same order as the sample size $T$, while assuring that the block size is asymptotically small enough to fit twice between each pair of consecutive break points $[b_lT]$ and $[b_{l+1}T]$ in the time domain. The second scenario (3.9) is constructed such that the relative block size $N/T$ vanishes, where the second part of (3.9) assures that the block length does not shrink too quickly. At this point, we merely emphasise that the asymptotic distributional properties of the process $\{\hat{E}_T(v, \omega)\}_{(v, \omega) \in [0, 1]^2}$ under setting (1) and (2) differ in a fundamental way and that this fact will be made precise by Theorems 3.1.1 and 3.1.2: Under setting (2), the projections of $\hat{E}_T$ at time points $v_1, v_2 \in [0, 1]$, $v_1 \neq v_2$, are asymptotically uncorrelated due to the fact that for $T$ large enough it holds $2N/T < |v_1 - v_2|$ and the estimates $\hat{E}_T(v_1, \omega_1)$ and $\hat{E}_T(v_2, \omega_2)$ for $\omega_1, \omega_2 \in [0, 1]$ are based on two blocks of data, which are divided by an ever increasing set of observations in between. In contrast, the setting (1) allows for a non-degenerate covariance structure between projections at points $(v_1, \omega_1)$ and $(v_2, \omega_2)$ as long as $|v_1 - v_2| \leq 2/c$. These observations are made rigorous by Theorems 3.1.1 and 3.1.2, which derive the asymptotic distribution of the empirical process $\{\hat{E}_T(v, \omega)\}_{(v, \omega) \in [0, 1]^2}$ under conditions (3.8) and (3.9) respectively.

**Theorem 3.1.1** (Weak convergence of the empirical process $\{\sqrt{N}\hat{E}_T(v, \omega)\}_{(v, \omega) \in [0, 1]^2}$) Suppose that the coefficients in the representation (2.9) satisfy

$$
\sum_{l=0}^{\infty} \sup_{u \in [0, 1]} \|\Psi_l(u)\|_\infty |l| < \infty \tag{3.10}
$$

and that the sequence $\{N_T\}_{T \in \mathbb{N}}$ complies with (3.8). Then the following statements hold:

a) Under the null hypothesis (3.3) of no structural breaks, the process $\{\sqrt{N}\hat{E}_T(v, \omega)\}_{(v, \omega) \in [0, 1]^2}$ converges weakly to a centred Gaussian process $\{B(v, \omega)\}_{(v, \omega) \in [0, 1]^2}$, i.e.

$$
\{\sqrt{N}\hat{E}_T(v, \omega)\}_{(v, \omega) \in [0, 1]^2} \Rightarrow \{B(v, \omega)\}_{(v, \omega) \in [0, 1]^2}. \tag{3.11}
$$

Here, for $v_i \in [\frac{1}{c}, 1 - \frac{1}{c}]$ ($i = 1, 2$) the covariance kernel of $\{B(v, \omega)\}_{(v, \omega) \in [0, 1]^2}$ is given by

$$
\text{Cov}([B(v_1, \omega_1)]_{a_1, b_1}, [B(v_2, \omega_2)]_{a_2, b_2}) = \begin{cases} 
0 & \text{if } \frac{2}{c} \leq |v_2 - v_1| \\
-2 - |v_2 - v_1| \frac{1}{2\pi} \int_{-\min(\omega_1, \omega_2)^\pi}^{\min(\omega_1, \omega_2)^\pi} \rho_{a_1, a_2, b_1, b_2}(\lambda) d\lambda & \text{if } \frac{1}{c} \leq |v_2 - v_1| \leq \frac{2}{c}, \\
[2 - 3|v_2 - v_1|] \frac{1}{2\pi} \int_{-\min(\omega_1, \omega_2)^\pi}^{\min(\omega_1, \omega_2)^\pi} \rho_{a_1, a_2, b_1, b_2}(\lambda) d\lambda & \text{if } 0 \leq |v_2 - v_1| \leq \frac{1}{c}
\end{cases} \tag{3.12}
$$

where the function $\rho_{a_1, a_2, b_1, b_2}$ is defined by

$$
\rho_{a_1, a_2, b_1, b_2}(\lambda) := f_{a_1, a_2}(\lambda) f_{b_1, b_2}(-\lambda) + f_{a_1, b_2}(\lambda) f_{b_1, a_2}(-\lambda).
$$
If \( v_i \notin \left[ \frac{1}{c}, 1 - \frac{1}{c} \right] \) for at least one \( i \in \{1, 2\} \), the covariance kernel is given by
\[
\text{Cov}(\{B(v_1, \omega_1)\}_{i_1, b_1}, \{B(v_2, \omega_2)\}_{i_2, b_2}) = \text{Cov}(\{B(a_c(v_1), \omega_1)\}_{i_1, b_1}, \{B(a_c(v_2), \omega_2)\}_{i_2, b_2}),
\]
where we defined \( a_c(v) := \min(\max(v, \frac{1}{c}), 1 - \frac{1}{c}) \).

b) Under the alternative (3.4), there exists a constant \( C > 0 \) such that
\[
\lim_{T \to \infty} P\left( \sup_{\omega \in [0, 1]} \left| \hat{E}_T(b, \omega) \right| > C \right) = 0
\]
for all \( (r, a, b) \in \{1, \ldots, K\} \times \{1, \ldots, d\}^2 \) with \( \sup_{\omega \in [0, 1]} \left| E(b, \omega) \right| > 0 \).

**Theorem 3.1.2** (Weak convergence of the empirical process \( \{N^{-1}\hat{E}_T(v, \omega)\}_{(v, \omega) \in [0, 1]^2} \))

Suppose that the coefficients in the representation (2.9) satisfy (3.10) and that the sequence \( \{N_T\}_{T \in \mathbb{N}} \) complies with (3.9). Then the following statements hold:

a) Under the null hypothesis (3.3) of no structural breaks, it is
\[
\sup_{(v, \omega) \in [0, 1]^2} \left| \hat{E}_T(v, \omega) \right| = o_P(N^{-\gamma})
\]
for any \( \gamma \in (0, 1/2) \).

b) Under the alternative (3.4), there exists a constant \( C > 0 \) such that
\[
\lim_{T \to \infty} P\left( \sup_{\omega \in [0, 1]} \left| \hat{E}_T(b, \omega) \right| > C \right) = 0
\]
for all \( (r, a, b) \in \{1, \ldots, K\} \times \{1, \ldots, d\}^2 \) with \( \sup_{\omega \in [0, 1]} \left| E(b, \omega) \right| > 0 \).

Theorem 3.1.1 implies that under the absence of structural breaks the empirical process \( \{\sqrt{N}\hat{E}_T(v, \omega)\}_{(v, \omega) \in [0, 1]^2} \) has a Gaussian limit if the condition (3.8) is satisfied, whereas Theorem 3.1.2 implies that \( \{N^{-1}\hat{E}_T(v, \omega)\}_{(v, \omega) \in [0, 1]^2} \) converges to zero uniformly in probability for any \( \gamma \in (0, 1/2) \) if (3.9) holds. Both results reveal a lot of information about the asymptotic properties of the test statistic (3.7). In fact, the continuous mapping theorem [see Theorem 2.3.9] implies that from Theorem 3.1.1 it follows that \( \sqrt{N}\hat{E}_T \) converges to the supremum of a Gaussian process, which is uniquely defined by the covariance kernel (3.12). However, since the second order structure of this process crucially depends on the unknown spectral density \( f \) of the data generating process \( \{X_{t,T}\}_{t=1,\ldots,T} \), we are not provided with a straightforward means for obtaining quantiles of the limiting distribution. Theorems 3.1.1 and 3.1.2, providing a lot of insight into the theoretical properties of the test statistic \( \hat{E}_T \), thus do not single-handedly allow for the construction of a test for the null hypothesis (3.3) [see Dahlhaus (2009) or Preuß et al. (2012) for a similar situation]. However, as will be explained in the following section, they build the theoretical justification for a bootstrap based procedure, which allows to accurately estimate the quantiles of the statistic \( \hat{E}_T \) under the null hypothesis. To summarise, in this section we have obtained the following insight into the asymptotic behaviour of the test statistic \( \hat{E}_T \) under the null hypothesis (3.3) of no structural breaks:
3.1 Testing for the presence of structural breaks

(1) In the case (3.8) of an asymptotically constant and non-vanishing relative block size \( N_T/T \), it holds

\[
\sqrt{N} \hat{E}_T \Rightarrow \sup_{(v, \omega) \in [0,1]^2} ||B(v, \omega)||_{\infty},
\]

where \( \{B(v, \omega)\}_{(v, \omega) \in [0,1]^2} \) is a Gaussian process with covariance kernel (3.12).

(2) In the case (3.9) of an asymptotically vanishing relative block size \( N_T/T \), it holds

\[
N^\gamma \hat{E}_T = o_p(1)
\]

for any \( \gamma \in (0, 1/2) \).

3.1.4 Bootstrapping the test statistic

As was shortly pointed out at the closing of the previous section, Theorems 3.1.1 and 3.1.2 by themselves do not provide a means for the construction of an asymptotic test with a controlled type I error, as the limiting distribution of the test statistic \( \hat{E}_T \) either depends on unknown quantities pertaining to the unknown spectral density matrix \( f \) (Theorem 3.1.1) or is degenerated (Theorem 3.1.2). This implies that under either configuration (3.8) or (3.9) we do not have a reliable source for critical values for a meaningful definition of a rejection region for the statistic \( \hat{E}_T \) yet. In order to resolve this obstacle, we intend to apply resampling methods to obtain estimates for the desired quantiles. More precisely, we proceed by explaining how a bootstrap procedure, which is closely related to the one dimensional AR(\( \infty \)) bootstrap introduced by Kreiß (1988), can be employed to accurately estimate the quantiles of \( \hat{E}_T \) under the null hypothesis. Before we formally explain the procedure and establish the accuracy of the resulting testing procedure, we state the following central assumptions:

Assumption 3.1.3

The stationary \( \mathbb{R}^d \)-valued process \( \{X_t\}_{t \in \mathbb{Z}} \) with spectral density function \( g(\lambda) = \int_0^1 f(u, \lambda)du \) has an AR(\( \infty \)) representation of the form

\[
X_t = \sum_{j=1}^{\infty} a_j X_{t-j} + \Sigma^{1/2} Z_t,
\]

where \( \{Z_t\}_{t \in \mathbb{Z}} \) denotes a sequence of independent \( d \)-dimensional \( N(0, I_d) \) distributed random variables, \( \Sigma \in \mathbb{R}^{d \times d} \) is positive definite and \( \{a_j\}_{j \in \mathbb{N}} \) is a sequence of \( d \times d \) matrices satisfying

\[
\det \left( I_d - \sum_{j=1}^{\infty} z^j a_j \right) \neq 0 \text{ for } |z| \leq 1 \quad \text{and} \quad \sum_{j=0}^{\infty} |j|^m \|a_j\|_{\infty} < \infty
\]

for some \( m \geq 1 \).

The main motivation of the proposed resampling procedure is founded on the idea that the underlying data generating process, which determines the distribution of the sample
\{X_{t,T}\}_{t=1,...,T}, can be approximated by an AR(p) model if the order \( p \) of the autoregressive model is chosen sufficiently large. Therefore, we consider an increasing sequence \( p = p(T) \to \infty \) as \( T \to \infty \) and approximate the stationary process \( \{X_t\}_{t \in \mathbb{Z}} \) defined in (3.17) by an AR(p) model with coefficients

\[
\begin{align*}
(a_{1,p}, \ldots, a_{p,p}) &:= \arg \min_{b_1, \ldots, b_{p}} \text{tr} \left( \mathbb{E}[\left( X_t - \sum_{j=1}^{p} b_{j,p} X_{t-j} \right) \left( X_t - \sum_{j=1}^{p} b_{j,p} X_{t-j} \right)^T] \right) \\
\end{align*}
\]

(3.19)

and innovations with covariance matrix

\[
\Sigma_p = \mathbb{E}\left( (X_t - \sum_{j=1}^{p} a_{j,p} X_{t-j})(X_t - \sum_{j=1}^{p} a_{j,p} X_{t-j})^T \right).
\]

For increasing order \( p = p(T) \), the coefficients \( (a_{1,p}, \ldots, a_{p,p}) \) in (3.19) resemble the sequence \( \{a_j\}_{j \in \mathbb{N}} \) in (3.17) more closely and it is thus reasonable to create replicates \( \hat{E}_T^\ast \) of the test statistic \( \hat{E}_T \) by employing the following algorithm.

**Algorithm 3.1.4 (Autoregressive bootstrap)**

1) Determine a consistent estimator \( (\hat{a}_{1,p}, \ldots, \hat{a}_{p,p}) \) of the minimiser in (3.19).

2) Simulate data from the model

\[
X^\ast_{t,T} = \sum_{j=1}^{p} \hat{a}_{j,p} X^\ast_{t-j,T} + \hat{\Sigma}_p^{1/2} Z^\ast_j,
\]

(3.20)

where for \( T \in \mathbb{N} \) the random variables \( Z^\ast_j \) are independent \( \mathcal{N}(0, I_d) \) distributed,

\[
\hat{\Sigma}_p = \frac{1}{T - p} \sum_{j=p+1}^{T} (\hat{z}_j - \bar{\hat{z}}_T)(\hat{z}_j - \bar{\hat{z}}_T)^T,
\]

\[
\hat{z}_j := X_{j,T} - \sum_{i=1}^{p} \hat{a}_{i,p} X_{j-i,T} \quad (j = p + 1, \ldots, T) \quad \text{and} \quad \bar{\hat{z}}_T := \frac{1}{T - p} \sum_{j=p+1}^{T} \hat{z}_j.
\]

3) Define \( \hat{E}_T^\ast \) in the same way as the statistic \( \hat{E}_T \) was defined in (3.7), where the observations \( \{X_{t,T}\}_{t=1,\ldots,T} \) are replaced by the bootstrap replicates \( \{X^\ast_{t,T}\}_{t=1,\ldots,T} \).

With the autoregressive bootstrap method of Algorithm 3.1.4 at hand, we are finally able to compute estimates of critical values for \( \hat{E}_T \). For a formal testing procedure for the null hypothesis (3.3) of no structural breaks we hence suggest to apply the following algorithm.

**Algorithm 3.1.5 (Test for structural breaks)**

1) Choose some level \( \alpha \in (0,1) \) of significance and an integer \( B \) for the amount of replicates employed in the estimation of the \((1 - \alpha)\)-quantile \( q_{1-\alpha}^{E_T} \) of the test statistic \( \hat{E}_T \) under the null hypothesis.
2) Calculate the test statistic $\hat{E}_T$ according to (3.7).

3) Choose the order $p$ for the fitted autoregressive model and determine a consistent estimator $(\hat{a}_{1,p}, ..., \hat{a}_{p,p})$ of the minimizer in (3.19).

4) Apply Algorithm 3.1.4 to generate $B$ replicates $\{\hat{E}_{T,1}^*, ..., \hat{E}_{T,B}^*\}$ of the test statistic and estimate the quantile $q_{1-\alpha}^{\hat{E}_T}$ by

$$q_{1-\alpha}^{\hat{E}_T} := \hat{E}_{T,(\lceil(1-\alpha)B\rceil)},$$

where $\hat{E}_{T,1}, ..., \hat{E}_{T,B}$ denotes the ordered bootstrap sample.

5) Reject the null hypothesis (3.3) of no break points, if

$$\hat{E}_T > q_{1-\alpha}^{\hat{E}_T}.$$  (3.22)

To obtain an intuition for the reasoning behind this procedure, note that under the null hypothesis the sample $\{X_{t,T}\}_{t=1,...,T}$ originates from a stationary process that has the same spectral density matrix as the process $\{X_t\}_{t \in \mathbb{Z}}$ defined in (3.17). By computing estimators for the minimiser in (3.19), we approximate this process by a finite order autoregressive model. The consistency of the estimators $\hat{a}_{j,p}$ implies that the distribution of the bootstrap samples $\{X_{t,T}^\ast\}_{t=1,...,T}$ generated according to (3.20) is similar to the distribution of the original sample $\{X_{t,T}\}_{t=1,...,T}$.

In the following demonstrations, we will formally show that Algorithm 3.1.5 and the rejection rule (3.22) in fact lead to an asymptotic level $\alpha$ test for the null hypothesis (3.3). In order to achieve this, we again distinguish between the cases (3.8) and (3.9) for the sequence $\{N_T\}_{T \in \mathbb{N}}$. In Theorem 3.1.6, we consider the cases of the block size converging to a fixed fraction of the sample size $T$ and in Theorem 3.1.7 we investigate the asymptotic behaviour of the bootstrap process $\{\hat{E}_T^\ast(v, \omega)\}_{(v, \omega) \in [0,1]^2}$ in the cases $N_T/T \to 0$ as $T \to \infty$.

**Theorem 3.1.6** (Weak convergence of the bootstrap process $\{\sqrt{N}_T \hat{E}_T^\ast(v, \omega)\}_{(v, \omega) \in [0,1]^2}$)

Let the assumptions of Theorem 3.1.1 and Assumption 3.1.3 be fulfilled. Furthermore, suppose that the following conditions on the growth rate of the sequence $p = p(T)$, the estimates $\hat{a}_{j,p}$ and the true AR parameters $a_j$ and $a_{j,p}$ defined in (3.17) and (3.19) are satisfied:

i) There exist sequences $p_{\min}(T)$ and $p_{\max}(T)$ such that the order $p$ of the fitted autoregressive process satisfies $p = p(T) \in [p_{\min}(T), p_{\max}(T)]$ with $p_{\max}(T) \geq p_{\min}(T) \to \infty$ and

$$p_{\max}^3(T) \sqrt{\frac{\log(T)}{T}} = O(1).$$  (3.23)

ii) The estimators $\hat{a}_{j,p}$ of the AR parameters $a_{j,p}$ satisfy

$$\max_{1 \leq j \leq p(T)} ||\hat{a}_{j,p} - a_{j,p}||_\infty = O_P\left(\sqrt{\frac{\log(T)}{T}}\right)$$  (3.24)

uniformly with respect to $p \in [p_{\min}(T), p_{\max}(T)]$. 


iii) The matrices \( \hat{\Sigma}_p \) and \( \Sigma \) satisfy

\[
\| \hat{\Sigma}_p - \Sigma \|_\infty \xrightarrow{P} 0.
\]

Then, as \( T \to \infty \), we have that conditionally on the sample \( X_{1,T}, \ldots, X_{T,T} \) it holds

\[
\left\{ \sqrt{N} \hat{E}_T^* (v, \omega) \right\}_{(v, \omega) \in [0,1]^2} \Rightarrow \left\{ B_{H_0} (v, \omega) \right\}_{(v, \omega) \in [0,1]^2}, \tag{3.25}
\]

almost surely, where \( B_{H_0} (v, \omega) \) denotes a Gaussian process, with covariance kernel

\[
\text{Cov}(B_{H_0}(v_1, \omega_1), B_{H_0}(v_2, \omega_2)) = \begin{cases} 0 & \text{if } \frac{2}{c} \leq |v_2 - v_1| \\ -[2 - |v_2 - v_1|c] \frac{1}{2\pi} \int_{-\min(\omega_1, \omega_2)\pi}^{\min(\omega_1, \omega_2)\pi} \rho_{H_0}^{a_1, a_2, b_1, b_2} (\lambda) d\lambda & \text{if } \frac{1}{c} \leq |v_2 - v_1| \leq \frac{2}{c} \\ [2 - 3|v_2 - v_1|c] \frac{1}{2\pi} \int_{-\min(\omega_1, \omega_2)\pi}^{\min(\omega_1, \omega_2)\pi} \rho_{H_0}^{a_1, a_2, b_1, b_2} (\lambda) d\lambda & \text{if } 0 \leq |v_2 - v_1| \leq \frac{1}{c} \end{cases} \tag{3.26}
\]

for \( v_i \in [\frac{1}{c}, 1 - \frac{1}{c}] \), where the function \( \rho_{H_0}^{a_1, a_2, b_1, b_2} \) is defined by

\[
\rho_{H_0}^{a_1, a_2, b_1, b_2} (\lambda) := \int_0^1 f_{a_1, a_2} (u, \lambda) f_{b_1, b_2} (u, -\lambda) du + \int_0^1 f_{a_1, b_2} (u, \lambda) f_{b_1, a_2} (u, -\lambda) du.
\]

If \( v_i \notin [\frac{1}{c}, 1 - \frac{1}{c}] \) for at least one \( i \in \{1, 2\} \), the covariance kernel is given by

\[
\text{Cov}(B_{H_0}(v_1, \omega_1), B_{H_0}(v_2, \omega_2)) = \text{Cov}(B_{H_0}(a_c(v_1), \omega_1), B_{H_0}(a_c(v_2), \omega_2)).
\]

where \( a_c(v) = \min(\max(v, \frac{1}{c}), 1 - \frac{1}{c}) \).

Theorem 3.1.6 implies that the suggested procedure for estimating the quantiles of the distribution of \( \hat{E}_T \), which is summarised in Algorithm 3.1.5, yields accurate estimates of the critical values for the test statistic \( \hat{E}_T \), in the case of \( N_T/T \to 1/c \). In fact, under the null hypothesis the covariance kernels (3.12) and (3.26) of the processes \( \{B(v, \omega)\}_{(v, \omega) \in [0,1]^2} \) and \( \{B_{H_0}(v, \omega)\}_{(v, \omega) \in [0,1]^2} \) are the same and hence the continuous mapping theorem assures that from (3.25) it follows that

\[
\sqrt{N} \hat{E}_T^* \Rightarrow \sup_{(v, \omega) \in [0,1]^2} \| B(v, \omega) \|_\infty.
\]

conditionally on the sample \( X_{1,T}, \ldots, X_{T,T} \) almost surely. This property together with (3.15) implies that in the case of no structural breaks and conditionally on the observed data the asymptotic distributions of the bootstrap replicates \( \hat{E}_T^* \) and the test statistic \( \hat{E}_T \) are the same in the scenario \( N_T/T \to 1/c \). Furthermore, consistency of the test (3.22) follows from the fact that under the presence of structural breaks the test statistic \( \hat{E}_T \) becomes larger than some positive constant [see Theorem 3.1.1 part b)], while the bootstrap samples \( \hat{E}_T^* \) converge to zero by Theorem 3.1.6.

For the scenario (3.9) of a shrinking relative block length \( N_T/T \to 0 \), the theoretical justification for the proposed bootstrap method is founded on the following theorem:
3.1 Testing for the presence of structural breaks

**Theorem 3.1.7**

Let the assumptions of Theorem 3.1.2 and Theorem 3.1.6 be fulfilled, where the condition (3.23) is replaced by

\[ \frac{p_3(T \sqrt{\log(T)} \sqrt{N}}{\sqrt{T}} = o(1). \]  

(3.27)

Furthermore, assume that the property (3.18) holds with \( m \geq 2 \). Then there exists a sequence of \( \mathbb{R}^{d \times d} \)-valued random processes \( \{ \hat{E}_{T,a}(v, \omega) \}_{(v, \omega) \in [0,1]} \) such that the following statements hold:

a) Under the null hypothesis (3.3) of no structural breaks, it holds

\[ \sup_{(v, \omega) \in [0,1]^2} \| \hat{E}_T(v, \omega) \|_\infty \overset{D}{=} \sup_{(v, \omega) \in [0,1]^2} \| \hat{E}_{T,a}(v, \omega) \|_\infty \]  

for any \( T \in \mathbb{N} \).

b) If \( K \geq 0 \), we have

\[ \frac{\sup_{(v, \omega) \in [0,1]^2} \| \hat{E}_T(v, \omega) \|_\infty - \sup_{(v, \omega) \in [0,1]^2} \| \hat{E}_{T,a}(v, \omega) \|_\infty}{\sqrt{E \sup_{(v, \omega) \in [0,1]^2} \| \hat{E}_{T,a}(v, \omega)^2 \|_\infty}} = o_P(1) \]  

(3.29)

as \( T \to \infty \).

c) For any \( \gamma \in (0, 1/2) \), it holds

\[ N^\gamma \sup_{(v, \omega) \in [0,1]^2} \| \hat{E}_{T,a}(v, \omega) \|_\infty = o_P(1). \]

Theorem 3.1.7 indicates that the distributions of the random variables

\[ \sup_{(v, \omega) \in [0,1]^2} \| \hat{E}_T(v, \omega) \|_\infty \sqrt{E \sup_{(v, \omega) \in [0,1]^2} \| \hat{E}_T(v, \omega)^2 \|_\infty} \]

and

\[ \sup_{(v, \omega) \in [0,1]^2} \| \hat{E}_{T,a}(v, \omega) \|_\infty \sqrt{E \sup_{(v, \omega) \in [0,1]^2} \| \hat{E}_{T,a}(v, \omega)^2 \|_\infty} \]

are asymptotically the same. This observation yields that the \((1-\alpha)\)-quantile of the distributions of \( \sqrt{N}\hat{E}_T \) and \( \sqrt{N}\hat{E}_{T,a} \) are close for large values of \( T \), which implies that the bootstrap assisted procedure for the estimation of the critical values yields accurate estimates and that the test (3.22) has asymptotic level \( \alpha \) in the setting (3.9) [see Paparoditis (2010) for a similar argument]. Consistency of the test (3.22) in the case \( N_T/T \to 0 \) follows from Theorem 3.1.2 and Theorem 3.1.7: While \( \hat{E}_T \) becomes larger than some positive constant under the alternative (3.4) due to Theorem 3.1.2 b) it is assured by Theorem 3.1.7b) and c) that each of the bootstrap replicates \( \hat{E}_{T,a} \) converges to zero.
Remark 3.1.8
As one of many possible modifications of the test (3.22) it can also be of interest to consider the statistic
\[ \hat{E}^{(2)}_T := \sup_{v \in [0,1]} ||\hat{E}_T(v,1)||_\infty. \]
Here, similar arguments can be used to obtain a bootstrap assisted test for the null hypothesis that the covariance matrix of the time series is constant over time. Such a test has already been developed by Aue et al. (2009). However, this observation shows that the test statistic \( \hat{E}_T \) is by no means the universally most appropriate quantity to consider for the identification of sudden changes of stylised facts in the second order structure of an observed set \( \{X_{1,T}, ..., X_{T,T}\} \) of data. In fact, the procedure, which is summarised in Algorithm 3.1.5, can easily be adapted to several other interesting hypotheses. Technically speaking, this can be achieved by choosing some appropriate function \( \phi(u, \lambda) : [0,1] \times [-\pi, \pi] \to \mathbb{C}^{d \times d} \) and investigating functionals of the form
\[ \frac{1}{\epsilon} \left( \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \phi(u, \lambda) f(u, \lambda) dud\lambda - \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \phi(u, \lambda) f(u, \lambda) dud\lambda \right) \]
instead of the matrices \( E(v, \omega), (v, \omega) \in [0,1]^2 \). For example, the choice \( \phi(u, \lambda) = I_d \exp(-i\lambda k) \) yields a test for constancy of the covariance function at a specific lag \( k \in \mathbb{N} \). A hypothesis of this type can be of interest if the statistician knows in advance that only certain lags have an impact on the dependency structure of the underlying process.

Remark 3.1.9
In the academic literature, statistics with a similar structure as \( \hat{E}_T(v, \omega) \) have been considered in various contexts and under different conditions and some references can be found in Giraitis and Leipus (1990). These authors conduct an investigation of the asymptotic distributional properties of the quantity
\[ \sup_{k \in \{1, ..., T\}} \sup_{\omega \in [0,1]} |\hat{F}_k(\omega) - \hat{F}_{T-k}^*(\omega)|, \quad (3.30) \]
where \( \hat{F}_k(\omega) = \int_0^{\omega} I_{(1, ..., k)}(\lambda) d\lambda \) and \( \hat{F}_{T-k}^*(\omega) = \int_0^{\pi} I_{(k+1, ..., T)}(\lambda) d\lambda \) and \( I_{(k_1, ..., k_2)}(\lambda) \) denotes the periodogram based on the data \( X_{k_1}, X_{k_1+1}, ..., X_{k_2} \) for \( k_1, k_2 \in \{1, ..., T\} \). By considering the statistic (3.30), the authors obtain a measure for the presence of structural breaks. However, this approach can only be employed to detect at most one break point. Furthermore, the authors show that an appropriately standardised version of the test statistic converges to some non-degenerate limit whose quantiles are not readily available. Therefore, the construction of a computable level \( \alpha \) test based on statistics of the type (3.30) is still a challenging problem. We will come back to the problem of estimating quantiles of the asymptotic distribution of such estimators in Chapter 4. At this point, we merely remark that the difficulty of this problem mainly results from the fact that statistics of the form (3.30) compare spectral density estimates that are based on two segments of different length. In our current setting, in which both blocks have size \( N \), this is not the case and we therefore postpone this discussion to a later time.
Remark 3.1.10
The assumption of Gaussianity of the innovation process \( \{Z_t\}_{t \in \mathbb{Z}} \) is imposed to simplify technical arguments in the proofs. An extension of the test (3.22) and an adoption of the proposed resampling method to the cases of general linear processes is straightforward but comes at the cost of a considerable increase of technicality and notation in the proofs. In fact, the proofs of Theorems 3.1.1, 3.1.2, 3.1.6 and 3.1.7 can be modified to address for non-Gaussian innovations and the only modification, which has to be made in an practical application of the bootstrap procedure, concerns the innovations \( \{Z_t^*\}_{t=1,...,T} \), which are generated for the calculation of the bootstrap replicates \( \hat{E}_{T,(1)}^*, ..., \hat{E}_{T,(B)}^* \). In the case of non-Gaussian innovations of the data generating process, the appropriate choice of the bootstrap replicates \( Z_t^* \) in Algorithm 3.1.5 has to mimic the fourth cumulant of the true underlying innovation process \( \{Z_t\}_{t \in \mathbb{Z}} \). This issue is discussed in more detail in Kreiß and Paparoditis (2012).

3.2 Dividing the data set into stationary segments

While the problem of testing for structural breaks in an observed set \( \{X_{t,T}\}_{t=1,...,T} \) of data is a very important topic [see the discussion at the beginning of this chapter], from a practical point of view, the mere knowledge of the presence of shifts in the distributional properties of the time series at various points in time is seldom satisfying by itself. The awareness that the data set at hand is likely to exhibit more than one regime frequently leads to the canonical question how to identify the points in time where one regime ends and another one begins. Whereas in the previous sections we have constructed a procedure which allows to test the null hypothesis (3.3) of no structural breaks at a controlled type I error, we now intend to develop a method which answers the below follow-up questions arising whenever the null hypothesis (3.3) is rejected:

1. How many breaks are present?
2. Where are the breaks located?
3. What are the characteristics of each break point, i.e. which components of the spectral density matrix \( f \) exhibit discontinuities?

In the following discussion, we will develop a procedure which consists of three steps and detects simultaneously the number, location and corresponding components of multiple structural breaks. Therefore, in the first step we estimate sets containing ‘potential break points’. Roughly speaking, these sets contain all points \( v \in \{N/T, ..., (T - N)/T\} \), for which the various components of the estimate \( \hat{E}_T(v, \omega) \) indicate that a structural break is ‘likely’ to be present. In the second step, these sets are used to extract a final set of estimators for the number and location of the break points. In the third step, we identify for each detected change point the components of the spectral density matrix \( f \) exhibiting discontinuities. In order to formally describe the method, we first recall the definition of the empirical process \( \{E_T(v, \omega)\}_{(v, \omega) \in [0,1]^2} \) given in (3.6) and choose some constant \( \gamma \in (0, 1/2) \). We proceed...
by identifying potential break points according to step I:

**Step I** (Identification of sets containing break points)

We consider a point \( v \in \{N/T, (N+1)/T, ..., (T-N)/T\} \) as a candidate for a structural break in the component \((a, b)\) if the inequality

\[
N^\gamma \sup_{\omega \in [0, 1]} \| \hat{E}_T(v, \omega) \|_{a,b} > \varepsilon_{T,a,b}(v)
\]

holds, where \( \varepsilon_{T,a,b}(v) \) is a threshold satisfying \( \lim \inf_{T \to \infty} \varepsilon_{T,a,b}(v) \geq C \) for some constant \( C > 0 \) and \( \varepsilon_{T,a,b}(v) = o(N^\gamma) \) uniformly in \( v \in [0, 1] \).

In practical applications, employing the decision rule (3.31) for each \((a, b) \in \{1, ..., d\}^2\) will identify \( K_T \geq 0 \) subsets \( R_1, ..., R_{K_T} \subset \{N/T, ..., 1 - N/T\} \) containing all points, where changes in the components of the spectral density matrix are 'likely' to be present. Each point contained in the conjunction of these sets undergoes a refined analysis in a second step of the procedure. Before we explain how, a set of final break point estimates is filtered out in this second step, we present an example illustrating the construction of the sets \( R_1, ..., R_{K_T} \).

**Example 3.2.1** (Illustration of Step I)

We consider the bivariate model

\[
X_{t,T} = \sum_{j=1}^{4} 1_{(t-1)T,4T}(t) \Theta_j Z_t,
\]

where the matrices \( \Theta_1, \Theta_2, \Theta_3, \Theta_4 \) are defined by

\[
\Theta_1 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \Theta_2 := \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \quad \Theta_3 := \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \quad \Theta_4 := \begin{pmatrix} \sqrt{2} & \sqrt{2} \\ 0 & 2 \end{pmatrix}
\]

and \( \{Z_t\}_{t \in \mathbb{Z}} \) is a two dimensional Gaussian White Noise process. The spectral density matrix \( \mathbf{f} \) of a bivariate time series following model (3.32) exhibits 3 break points, where the first change only involves the first component \( [\mathbf{f}]_{1,1} \), the second only concerns the component \( [\mathbf{f}]_{2,2} \) and the third break point leaves the components \( [\mathbf{f}]_{1,1} \) and \( [\mathbf{f}]_{2,2} \) unchanged but appears in the cross spectrum \( [\mathbf{f}]_{1,2} \) and \( [\mathbf{f}]_{2,1} \), which determines the dependence structure of the two components. Figure 3.1 contains a plot of a typical set of data of length \( T = 2048 \) generated by model (3.32), where the dashed vertical lines indicate the true break points in the univariate time series. Note that the third break point only corresponds to a change in the dependence structure of the two univariate data sets and cannot easily be seen from the time series plot. Figure 3.2 shows the four plots of the component wise functions

\[
v \mapsto N^\gamma \sup_{\omega \in [0, 1]} \| \hat{E}_T(v, \omega) \|_{a,b},
\]
3.2 Dividing the data set into stationary segments

Figure 3.1: Simulated data from the model (3.32) for the sample size $T = 2048$. The vertical dashed lines denote the true break points in the univariate time series.

$a, b \in \{1, 2\}$ (solid lines), where $N = 256$ and $\gamma = 0.4$. In each component we added a plot of the threshold level

$$v \mapsto \varepsilon_{T,a,b}(v),$$

$a, b \in \{1, 2\}$ (red lines), where we used a data driven choice for this quantity, which will be formally defined in Section 3.3. It can easily be seen that for each component $(a, b)$ the test statistic exceeds the level $\varepsilon_{T,a,b}(v)$ in a neighbourhood of the break point. For the simulated set of data we obtain the sets

$$R_1 = \left\{ \frac{311}{2048}, \ldots, \frac{592}{2048} \right\}, \quad R_2 = \left\{ \frac{809}{2048}, \ldots, \frac{1132}{2048} \right\} \quad \text{and} \quad R_3 = \left\{ \frac{1453}{2048}, \ldots, \frac{1699}{2048} \right\}$$

of potential break points. At this point, we already emphasise that the local maxima of the functions $v \mapsto N^7 \sup_{\omega \in [0,1]} |\hat{E}(v, \omega)|_{a,b}$ are close to the true change points. This observation will be exploited in the the second step of the procedure, which is designed to reduce the sets $R_1$, $R_2$ and $R_3$.

The following result shows that for an increasing sample size the subsets $R_1, \ldots, R_{K_T}$ are contained in neighbourhoods of radius $N/T$ of the ’true’ break points, i.e. it holds

$$\bigcup_{j=1}^{K_T} R_j \subset \bigcup_{a,b \in \{1, \ldots, d\}} \mathcal{I}_{T,a,b}(b_1, \ldots, b_K),$$

where

$$\mathcal{I}_{T,a,b}(b_1, \ldots, b_K) := \bigcup_{j=1}^{K} \left\{ \left\lfloor \frac{b_j T}{T} \right\rfloor - N, \ldots, \left\lceil \frac{b_j T}{T} \right\rceil + N \right\},$$

(3.33)
A new approach to the change point problem in time series analysis

Figure 3.2: Plots of the functions \( v \mapsto N^\gamma \sup_{\omega \in [0,1]} |[\hat{E}_T(v,\omega)]_{a,b}| \) (solid lines) and \( v \mapsto \varepsilon_{T,a,b}(v) \) \((a,b = 1,2)\) (red lines) with vertical dashed lines at the true break points.

Theorem 3.2.1

Assume that the condition (3.10) is satisfied and that the sequence \( \{N_T\} \subseteq \mathbb{N} \) satisfies either (3.8) or (3.9). Furthermore, for all \( a,b \in \{1,...,d\}, v \in [0,1] \) and \( \gamma \in (0,\frac{1}{2}) \) let \( \{\varepsilon_{T,a,b}(v)\} \subseteq \mathbb{N} \) denote a sequence satisfying

\[ \varepsilon_{T,a,b}(v) = o(N^\gamma) \quad \text{and} \quad \lim_{T \to \infty} \inf_{v \in [0,1]} \varepsilon_{T,a,b}(v) \geq C \]

for some constant \( C > 0 \). Then the detection rule (3.31) is accurate in the following sense:

a) The probability that the decision rule (3.31) of Step I indicates a potential structural break at a rescaled time point, which has a distance of at least \( N/T \) from each of the break points \( b_1,...,b_K \), vanishes asymptotically, i.e.

\[ P\left( \bigcup_{a,b \in \{1,...,d\}} \bigcup_{v \in \mathcal{I}_{T,a,b}(b_1,...,b_K)} \left\{ N^\gamma \sup_{\omega \in [0,1]} |[\hat{E}_T(v,\omega)]_{a,b}| > \varepsilon_{T,a,b}(v) \right\} \right) \xrightarrow{T \to \infty} 0, \quad (3.34) \]

where \( \mathcal{I}_{T,a,b}(b_1,...,b_K) = \{N/T, (N+1)/T, ..., (T-N)/T\} \setminus \mathcal{I}_{T,a,b}(b_1,...,b_K) \), and where the set \( \mathcal{I}_{T,a,b}(b_1,...,b_K) \) was defined in (3.33).

b) The probability that the procedure detects all structural breaks converges to one, i.e.

\[ P\left( \bigcap_{v \in \{b_1,...,b_K\}} \bigcap_{(a,b) \in B(v)} \left\{ N^\gamma \sup_{\omega \in [0,1]} |[\hat{E}_T(v,\omega)]_{a,b}| > \varepsilon_{T,a,b}(v) \right\} \right) \xrightarrow{T \to \infty} 1, \quad (3.35) \]

where the set \( B(v) \) is defined by

\[ B(v) := \{(a,b) \in \{1,...,d\}^2 | \sup_{\omega \in [0,1]} |[E(v,\omega)]_{a,b}| > 0\}. \quad (3.36) \]
Recall that in Step I we use the decision rule (3.31) to identify sets $R_j$, $j \in \{1, \ldots, K T\}$ of possible break points. Theorem 3.2.1 shows that these sets have the following important characteristics:

1. With an asymptotic probability of one the sets $R_j$ are contained in the neighbourhoods
   \[
   \{(\lfloor b_j T \rfloor - N)/T, \ldots, (\lfloor b_j T \rfloor + N)/T\}
   \]
   of those points $b_j$, for which there exists a change in at least one of the components of the spectral density matrix $f$.

2. With asymptotic probability of one each break point $b_i$, $i \in \{1, \ldots, K\}$ is contained in the set
   \[\bigcup_{j=1}^{K T} R_j.\]

Property (1) assures that the proposed procedure asymptotically does not identify possible break points at locations which are far away from any discontinuities in the spectral density $f$. Meanwhile, the characteristic (2) implies that each true break point $b_j$ is identified as a potential break point with probability converging to one. However, as has already been demonstrated in Example 3.2.1, the inequality (3.31) is usually satisfied in a neighbourhood of each break point. Consequently, for a finite sample size, the number of possible change points detected by the rule (3.31) is usually much larger than the true number $K \in \mathbb{N}$.

In order to obtain an accurate estimate $\hat{K}$ for the amount of structural breaks, which are present in the data, we need a procedure to extract a final set $(\hat{b}_1, \ldots, \hat{b}_K)$ of break point estimates from the set $\bigcup_{j=1}^{K T} R_j$. For this purpose, we employ the following fundamental idea: For a certain set $R_j$ of points in
\[\{N/T, \ldots, 1 - N/T\}\]
we identify the point $\tilde{b} \in R_j$ for which the local deviation from stationarity is maximal and then remove all points of an sufficiently large interval surrounding the point $\tilde{b}$ from the set $R_j$. The number and locations of the break points $(b_1, \ldots, b_K)$ are then estimated by the following algorithmic procedure, which embodies this general idea:

**Step II (Localisation of structural breaks)**

1. Let $\hat{K}$ denote the number of elements $v \in \{N/T, (N+1)/T, \ldots, (T-N)/T\}$, for which the inequality (3.31) holds for at least one component $(a, b) \in \{1, \ldots, d\}^2$. We denote the corresponding elements by $\hat{b}_1, \ldots, \hat{b}_K$ and define the sets
   \[\hat{B}_P := \{\hat{b}_1, \ldots, \hat{b}_K\} \quad \text{and} \quad \hat{B}_D = \emptyset\]
   of potential and detected break points.

2. If the set $\hat{B}_P$ is not empty, we add the element $\tilde{b} \in \hat{B}_P$ to the set $\hat{B}_D$ for which the quantity
   \[\sup_{(a,b)\in\{1,\ldots,d\}^2} \left( \sup_{\omega \in [0,1]} N^\gamma |[\hat{E}_T(\tilde{b}, \omega)]_{a,b}| \right)\]
   is maximal and reduce the set $\hat{B}_P$ by all elements which are contained in the neighbourhood $[\tilde{b} - (1 + \tau) N/T, \tilde{b} + (1 + \tau) N/T]$ of the detected break point $\tilde{b}$ for some $\tau > 0$, i.e. we replace $\hat{B}_P$ by
   \[\hat{B}_P \setminus [\tilde{b} - (1 + \tau) N/T, \tilde{b} + (1 + \tau) N/T].\] (3.37)
(3) Step (2) is repeated until \( \hat{B}_P = \emptyset \).

(4) We redefine \( \hat{K} = |\hat{B}_D| \) and choose \((\hat{b}_1, \ldots, \hat{b}_\hat{K})\) as the elements of \( \hat{B}_D \) such that \( \hat{b}_i < \hat{b}_{i+1} \) for \( i = 1, \ldots, \hat{K} - 1 \).

We reconsider the situation introduced in Example 3.2.1 in order to demonstrate how the algorithm of Step II can be employed to reach a final set of break point estimators.

**Example 3.2.2** (Illustration of Step II)

We continue the discussion of Example 3.2.1, where for \( T = 2048 \) we identified for a realisation \( \{X_t, T\}_{t=1, \ldots, T} \) of the model (3.32) three sets \( R_1, R_2 \) and \( R_3 \) of potential break points. At the beginning of Step II the sets of possible and determined break points are thus given by

\[
\hat{B}_P = R_1 \cup R_2 \cup R_3 \quad \text{and} \quad \hat{B}_D = \emptyset
\]

respectively. For this illustration we choose \( \tau = 0.1 \) and emphasise that in the situation at hand, the choice of \( \tau \) has no influence on the composition of the final break point estimators as long as \( \tau \) is chosen sufficiently small. In the first iteration of the algorithm of Step II, we add the element \( \tilde{b} = \frac{511}{2048} \) to \( \hat{B}_D \) and remove all elements, which do not have a distance of at least \( (1 + \tau)N/T = (1 + \tau)\frac{256}{2048} \) from \( \tilde{b} = \frac{511}{2048} \), from the set \( \hat{B}_P \). This leaves the set \( \hat{B}_P = R_2 \cup R_3 \). In the next iteration, we add the element \( \tilde{b} = \frac{1537}{2048} \) to \( \hat{B}_D \) and reduce the set \( \hat{B}_P \) to \( \hat{B}_P = R_2 \). In the last step, we add the point \( \tilde{b} = \frac{1026}{2048} \) and obtain

\[
\hat{B}_D = \left\{ \frac{511}{2048}, \frac{1026}{2048}, \frac{1537}{2048} \right\} \quad \text{and} \quad \hat{B}_P = \emptyset.
\]

At this point, the procedure terminates because \( \hat{B}_P = \emptyset \) and yields \( \hat{K} = 3 \) as an estimator for the number of break points. The locations of the break points are estimated by

\[
(\hat{b}_1, \hat{b}_2, \hat{b}_3) = \left( \frac{511}{2048}, \frac{1026}{2048}, \frac{1537}{2048} \right).
\]

Thus, in this example all break points are detected and as can be seen in Figure 3.2, the respective components of the spectral density matrix \( f \), which are responsible for these changes, are identified as well.

The procedure summarised in Step II yields an estimator \( \hat{K} \) for the amount of regime shifts present and a vector \((\hat{b}_1, \ldots, \hat{b}_\hat{K})\) of estimators for the locations of these break points. Step II thus allows to answer questions (1) and (2) raised at the beginning of this section. In order to investigate the characteristic properties of the detected break points we intend to identify the components of the spectral density matrix \( f \) exhibiting discontinuities at the respective locations. For this purpose, we suggest the following refined analysis:
Step III (Determination of the characteristics of structural breaks)

For each detected break point \( \hat{b}_i \) and each component \((a, b) \in \{1, \ldots, d\}^2\), we say that the change materialises in the component \((a, b) \in \{1, \ldots, d\}^2\) if the inequality
\[
N^\gamma \sup_{\omega \in [0, 1]} |[\hat{E}_T(\hat{b}_i, \omega)]_{a,b}| > \varepsilon_{T,a,b}(\hat{b}_i)
\] (3.38)
holds.

The below theorem formally establishes that the detection and characterisation procedure comprised of Steps I, II and III consistently estimates the true number \(K\) and locations \((b_1, \ldots, b_K)\) of the break points and accurately identifies the components of the spectral density matrix \(f\), which exhibit a change at the respective locations with asymptotic probability one.

**Theorem 3.2.2** (Consistency of the detection procedure)

Assume that condition (3.10) holds and that the sequence \(\{N_T\}_{T \in \mathbb{N}}\) satisfies either (3.8) or (3.9). Furthermore, for all \(a, b \in \{1, \ldots, d\}\), \(v \in [0, 1]\) and \(\gamma \in (0, 1/2)\) let \(\varepsilon_{T,a,b}(v)\) denote a sequence satisfying
\[
\varepsilon_{T,a,b}(v) = o(N^\gamma) \quad \text{and} \quad \liminf_{T \to \infty} \inf_{v \in [0,1]} \varepsilon_{T,a,b}(v) \geq C
\]
for some constant \(C > 0\). Then the following properties hold:

a) The estimator \(\hat{K}\) for the number \(K\) of break points is consistent, i.e.
\[
\hat{K} \xrightarrow{P} K
\] (3.39)
as \(T \to \infty\).

b) For each \(i \in \{1, \ldots, K\}\) the estimator \(\hat{b}_i\) consistently estimates break point \(b_i\), i.e.
\[
\hat{b}_i \xrightarrow{P} b_i
\] (3.40)
as \(T \to \infty\).

c) For each break point \(b_i, i \in \{1, \ldots, K\}\), the identification of the affected components by the inequality (3.38) is consistent, i.e.
\[
P\left( \bigcap_{i \in \{1, \ldots, K\}} \left( S_{1,T}(\hat{b}_i) \cap S_{2,T}(\hat{b}_i) \right) \right) \xrightarrow{T \to \infty} 1,
\]
where for \(\hat{b}_i\) the events \(S_{1,T}(\hat{b}_i)\) and \(S_{2,T}(\hat{b}_i)\) are defined by
\[
S_{1,T}(\hat{b}_i) := \bigcap_{(a,b) \in B(\hat{b}_i)} \left\{ N^\gamma \sup_{\omega \in [0,1]} |[\hat{E}_T(\hat{b}_i, \omega)]_{a,b}| > \varepsilon_{T,a,b}(\hat{b}_i) \right\}
\]
and
\[
S_{2,T}(\hat{b}_i) := \bigcap_{(a,b) \in \{1, \ldots, d\}^2 \setminus B(\hat{b}_i)} \left\{ N^\gamma \sup_{\omega \in [0,1]} |[\hat{E}_T(\hat{b}_i, \omega)]_{a,b}| \leq \varepsilon_{T,a,b}(\hat{b}_i) \right\}
\]
and for \(v \in [0,1]\) the set \(B(v)\) is defined in (3.36).
3.3 Practical application and finite sample properties

In this section, we demonstrate the applicability of the proposed method for testing for the presence of structural breaks and the procedure for estimating the amount and locations of the break points. For this purpose, we first define some data driven choices for the regularising parameters, which have to be specified for the implementation, and explain their underlying intuition. We briefly comment on the topic of run time efficient implementation and subsequently turn to an investigation of the finite sample properties of the procedures by means of a simulation study. These empirical results can be found in Section 3.3.2.

3.3.1 Recommendations for implementation

The proposed test for structural breaks and the corresponding localisation method depend on the choice of several regularisation parameters. For the application of the test (3.22), these parameters are the block length $N$ and the order $p$ of the autoregressive model for the bootstrap procedure. To estimate the amount and location of the break points, it is additionally required to choose the parameters $\gamma$ and $\tau$. In this section, we give recommendations for choosing these parameters that allow to optimally adapt the testing and estimation methods to various real world settings. On the one hand, the proposed rules for fine tuning will be in accordance with the asymptotic theory presented in the previous sections. On the other hand, they are designed with a focus on good performance in finite sample situations.

We first consider the choice of $\gamma$. Generally speaking, increasing the value of $\gamma \in (0, 1/2)$ tends to increase the number of potential break points $\hat{K}$, which are identified in Step I of the localisation procedure. This naturally implies that the amount of final break points filtered out in Step II tends to increase in lockstep with the value of $\gamma$. This property is of practical importance for example if it is desirable to avoid an underestimation of the true number of break points. For the implementation, which is used to obtain the empirical results in the next Section, we chose the value $\gamma = 0.4$, which gives a good performance in every model under consideration.

The performance of the proposed algorithm for estimating the break points is rather insensitive with respect to the configuration of the the parameter $\tau$ and for the empirical results presented below we choose $\tau = 0.01$.

For the choice of the threshold sequence $\varepsilon_{T,a,b}(v)$, which is used in (3.31) to identify potential break points, we note that it follows by similar calculations, as are provided in the proof of Theorem 3.1.1, that the variance of $[\bar{E}_T(v, \omega)]_{a,b}$ satisfies

$$\lim_{T \to \infty} \frac{1}{N/T} \text{Var}(\sqrt{N}[\bar{E}_T(v, \omega)]_{a,b}) = \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{1}{2\epsilon} \int_{v-\epsilon}^{v+\epsilon} \left( |f(u, \lambda)|_{aa} |f(u, \lambda)|_{bb} + |f(u, \lambda)|_{ab}|^2 \right) du d\lambda.$$

For a data driven choice of the sequence $\varepsilon_{T,a,b}(v)$, we consider for fixed $(v, \omega) \in [0, 1]^2$, the
estimator
\[ M_{T,a,b}(v, \omega) = \frac{1}{N} \sum_{k=-[\omega N]}^{[\omega N]} \mathbb{I}_{2N}(v, \lambda_{k,2N})_{aa} \mathbb{I}_{2N}(v, \lambda_{k,2N})_{bb} \]

for the asymptotic local variance, which is based on a block of length $2N$ surrounding the observation $X_{[vT]}$. Arguments similar to those presented in the proofs of Theorems 3.1.1 and 3.1.2 yield that the random variables $N^\gamma \mathbb{E}_T(v, \omega)_{a,b}$ and $N^\gamma \mathbb{E}_T(v', \omega')_{a,b}$ are asymptotically independent whenever $v \neq v'$. Thus, we recommend to specify the threshold sequence by

\[ \varepsilon_{T,a,b}(v) = \sqrt{2M_{T,a,b}(v, 1) \log \left( \frac{d(d+1)T}{2N} \right)}, \] (3.41)

where $d$ denotes the dimension of the time series under consideration. We note that under the non-restrictive condition

\[ \inf_{u \in [0,1]} \min_{i \in \{1, \ldots, d\}} [f(u, \lambda)]_{ii} > 0 \]

it follows that the inequality

\[ \liminf_{T \to \infty} \varepsilon_{T,a,b}(v) \geq C > 0 \]

holds for all $v \in [0,1]$, $a, b \in \{1, \ldots, d\}$ with probability converging to 1. This shows that the sequence (3.41) complies with the conditions of Theorem 3.2.1 and thus leads to the consistency of the estimators $\hat{K}$ and $(\hat{b}_1, \ldots, \hat{b}_k)$ according to Theorem 3.2.2.

For the choice of the window length $N$, there exists a similar tradeoff as in the case of the parameter $\gamma$. When confronted with a data set that exhibits only a small amount of switches in regime and features long segments, on which the underlying process is stationary, a rather large choice for the parameter $N$ is beneficial, as it increases the precision of the local spectral density estimates and thus leads to higher precision of the localisation of break points according to Step II. However, in the case of many break points with small distances in between it is crucial to choose $N$ sufficiently small to assure that the procedure detects discrepancies between short consecutive segments of different regimes. Nevertheless, the problem of break point detection implies by its nature that we do not have this kind of information about the amount and length of the regimes in the underlying process in advance. The following data-dependent algorithmic procedure tries to resolve this problem.

**Algorithm 3.3.1 (Choice of the window length $N$)**

1. We consider a set $V_T := \{N_1, \ldots, N_n\}$ of even integers $N_i$, which satisfy

\[ \sqrt{T} \leq N_1 < N_2 < \ldots < N_n \leq T^{5/6}. \]

2. We determine for each $N_i \in V_T$ the number $\hat{K}_T(N_i)$ of break points detected by the algorithm of Steps I and II.
We define
\[ i^* := \sup \{ i \in \{2, ..., n(T)\} | \hat{K}_T(N_{i-1}) \leq \hat{K}_T(N_i) \}, \]
where \( \sup \emptyset = -\infty \).

We set
\[ N^* = \begin{cases} 
N_i & \text{if } i^* \leq n(T) \\
N_{n(T)} & \text{if } i^* = -\infty
\end{cases} \]
and use \( N = 2N^* \) for the test of structural breaks and \( N = N^* \) for the estimation of \( K \) and the localisation of the break points.

The idea of the above algorithm is to apply the detection method for every \( N \in V_T \) and to select \( N^* \) as the largest \( N \in V_T \) for which there is no additional break point detected for the next smaller \( N \in V_T \). The intention of this procedure is to choose \( N \) as large as possible, while at the same time choosing \( N \) sufficiently small to identify short stationary segments. Note also that the recommended choice of \( N \) is different for the testing and the detection procedure. This recommendation is motivated by the following observation. When the data generating process contains two consecutive regimes, where the smaller of the two segments has length \( m \), the test (3.22) works best for the choice \( N = m \), as the discrepancy between the spectral densities on these two segments is estimated with the highest possible accuracy for this configuration. Meanwhile, for the detection method the best configuration is \( N = m/2 \) due to the asymptotic theory, which assumes that the distance between two consecutive change points is at least \( 2/e \).

For the implementation of the AR(\( \infty \)) bootstrap, we select the order \( p \) for the fitted AR model as the minimiser of the AIC criterion, i.e. we choose
\[ \hat{p} = \arg\min_p \frac{2\pi}{T} \sum_{k=1}^{T/2} \left( \log(\det[f_{\hat{\theta}(p)}(\lambda_{k,T})]) + \text{tr}\left[(f_{\hat{\theta}(p)}(\lambda_{k,T}))^{-1}I_T(\lambda_{k,T})\right]\right) + \frac{p}{T} \] (3.42)
[see Whittle (1951)]. Here, \( f_{\hat{\theta}(p)} \) denotes the spectral density of the fitted stationary AR(\( p \)) process and \( I_T \) is the usual periodogram calculated under the assumption of stationarity with the corresponding Fourier frequencies \( \lambda_{k,T} = 2\pi k/T \). Finally, for our implementation of the AR(\( \infty \)) bootstrap, we use the famous Yule-Walker estimators for \( \hat{a}_{j,p} \). In this context, we remark that in Hannan and Kavalieris (1986) it was shown that the Yule-Walker and the least squares estimators satisfy the condition (3.24) and that our choice of estimators thus is in accordance with the asymptotic theory of Section 3.1.

For the implementation of the testing and detection procedures, we additionally restrict ourselves to configurations of the parameter \( N \) which equal a power of two. More precisely, we consider \( n = \left[ \log_2(T^{5/6}) \right] - \left[ \log_2(\sqrt{T}) \right] + 1 \) and \( N_i = 2^{[\log_2(\sqrt{T})]+1+i} \) for \( i = 1, ..., n \). This choice allows for an application of the fft algorithm in the calculation of the local periodogram yielding a significant reduction in computational time.
3.3.2 Simulation study

In this section, we present the results of an extensive simulation study that investigates the finite sample performance of the test (3.22) and the estimation procedure comprised of Steps I, II and III. In this empirical analysis, we also provide a comparison with competing methods. For all results presented below, we chose the regularising parameters according to the criteria, which were explained in the previous section. Furthermore, we emphasise that for all time series models, which we consider in the following, the sequence \( \{Z_t\}_{t \in \mathbb{Z}} \) is assumed to be Gaussian White Noise with covariance matrix given by the identity whenever no other second order moment structure is specified.

We now begin our empirical study of the new methods and demonstrate the applicability of the bootstrap assisted test (3.22).

Size of the test (3.22)

We begin with a study of the nominal level of the bootstrap test. For the results which are reported below, we used 1000 simulation runs for estimating the rejection probabilities and employed 300 bootstrap replications for estimating the critical values. For our size study we consider the bivariate MA(1) and AR(1) models

\[
X_{t,T} = Z_t + \begin{pmatrix} \theta & 0.2 \\ 0.2 & \theta \end{pmatrix} Z_{t-1}
\]  

(3.43)

\[
X_{t,T} = \begin{pmatrix} \phi & 0.2 \\ 0.2 & \phi \end{pmatrix} X_{t-1} + Z_t.
\]  

(3.44)

Table 3.1 shows the rejection frequencies obtained for different values of the parameters \( \theta \) and \( \phi \) and various sample sizes \( T \). It can be seen that the nominal level is underestimated for small sample sizes, but it is evident that the approximation improves with increasing sample size.

<table>
<thead>
<tr>
<th>( H_0: ) Model (3.43)</th>
<th>( H_0: ) Model (3.44)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta = 0.5 )</td>
<td>( \phi = 0.5 )</td>
</tr>
<tr>
<td>( \theta = -0.5 )</td>
<td>( \phi = -0.5 )</td>
</tr>
<tr>
<td>( T )</td>
<td>( 5% )</td>
</tr>
<tr>
<td></td>
<td>( 10% )</td>
</tr>
<tr>
<td>128</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>0.057</td>
</tr>
<tr>
<td>256</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>0.065</td>
</tr>
<tr>
<td>512</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>0.096</td>
</tr>
</tbody>
</table>

Table 3.1: Simulated nominal level of the test (3.22) for the models (3.43) and (3.44) with different choices of \( \theta \), \( \phi \) and for various sample sizes \( T \).

Power of the test (3.22)

We now present the results of an extensive simulation study that compares the new test (3.22) with the CUSUM type procedure developed in Aue et al. (2009). We note that this proce-
procedure is specifically designed to test for constancy of the covariance matrix $\text{Cov}(X_{t,T}, X_{t,T})$ of a multivariate time series. The reported rejection frequencies are obtained from 1000 simulation runs in the case of the CUSUM test of Aue et al. (2009) and 500 runs in the case of the new test (3.22). For a comparison, we first consider the bivariate models

$$X_{t,T} = \sum_{l=0}^{K} 1_{[l,b_l+1],[b_{l+1},T]}(t) \left( \begin{array}{cc} \theta_l & 0.2 \\ 0.2 & \theta_l \end{array} \right) Z_{l-1} + Z_t$$  \hfill (3.45)

$$X_{t,T} = \sum_{l=0}^{K} 1_{[l,b_l+1],[b_{l+1},T]}(t) \left( \begin{array}{cc} \phi_l & 0.2 \\ 0.2 & \phi_l \end{array} \right) X_{t-1,T} + Z_t$$  \hfill (3.46)

$$X_{t,T} = \sum_{l=0}^{K} 1_{[l,b_l+1],[b_{l+1},T]}(t) \left( \begin{array}{cc} \sigma_l & 0.2 \\ 0.2 & \sigma_l \end{array} \right) Z_l$$  \hfill (3.47)

for different choices of the number $K$ and location $b = (b_1,\ldots,b_K)$ of the break points and parameters $\Sigma := (\sigma_0,\ldots,\sigma_K)$, $\Phi := (\phi_0,\ldots,\phi_K)$ and $\Theta := (\theta_0,\ldots,\theta_K)$. The simulated rejection probabilities in models (3.45), (3.46) and (3.47) are displayed in Table 3.2.

<table>
<thead>
<tr>
<th>model</th>
<th>$b$</th>
<th>parameter</th>
<th>$T = 128$</th>
<th>$T = 256$</th>
<th>$T = 512$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3.22) Aue (3.22) Aue (3.22) Aue (3.22) Aue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3.45)</td>
<td>$(\frac{1}{4}, \frac{5}{4}, \frac{1}{4})$</td>
<td>$\Theta = (1, -1.5, 1, -1.5)$</td>
<td>0.294</td>
<td>0.066</td>
<td>0.602</td>
</tr>
<tr>
<td></td>
<td>$(\frac{1}{2})$</td>
<td>$\Theta = (1, -1.5)$</td>
<td>0.276</td>
<td>0.257</td>
<td>0.516</td>
</tr>
<tr>
<td>(3.46)</td>
<td>$(\frac{1}{4}, \frac{5}{4}, \frac{1}{4})$</td>
<td>$\Phi = (0.5, -0.5, 0.5, -0.5)$</td>
<td>0.000</td>
<td>0.046</td>
<td>0.288</td>
</tr>
<tr>
<td></td>
<td>$(\frac{1}{2})$</td>
<td>$\Phi = (0.5, -0.5)$</td>
<td>0.080</td>
<td>0.047</td>
<td>0.254</td>
</tr>
<tr>
<td>(3.47)</td>
<td>$(\frac{1}{4}, \frac{5}{4}, \frac{1}{4})$</td>
<td>$\Sigma = (1, 2, 1, 0.5)$</td>
<td>1.000</td>
<td>0.213</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>$(\frac{1}{2})$</td>
<td>$\Sigma = (1, 2)$</td>
<td>0.844</td>
<td>0.818</td>
<td>0.992</td>
</tr>
<tr>
<td></td>
<td>$\emptyset$</td>
<td>$\Sigma = (1)$</td>
<td>0.030</td>
<td>0.033</td>
<td>0.038</td>
</tr>
</tbody>
</table>

Table 3.2: Empirical rejection frequencies of the test (3.22) and the CUSUM type procedure of Aue et al. (2009). The nominal level is $\alpha = 0.05$ and the models are defined in (3.45), (3.46) and (3.47), where different choices of break points $b = (b_1,\ldots,b_K)$ and AR parameters $\Phi = (\phi_0,\ldots,\phi_K)$, MA-parameters $\Theta = (\theta_0,\ldots,\theta_K)$ and standard deviations $\Sigma = (\sigma_0,\ldots,\sigma_K)$ are considered.

The first part of Table 3.2 shows that in the MA-model (3.45) the new test yields a substantially larger power than the test of of Aue et al. (2009) in the case of multiple break points, while the power is slightly smaller for $T \in \{256, 512\}$ if there exists only one break point. For the piecewise stationary AR model (3.46) we observe that the new test (3.22) significantly outperforms the CUSUM type test of Aue et al. (2009) regardless of the specific model configuration. The lower part of Table 3.2 contains the simulated power of the test (3.22) and the procedure of Aue et al. (2009) for a selection of specifications of the White Noise model (3.47). Here, we observe that for small sample sizes the new method significantly outperforms the competing test of Aue et al. (2009) for models featuring multiple break points. In the case of one break point, the new method performs similarly for large sample sizes and
yields slightly better results for small sample sizes. Note that these findings are remarkable since the breaks in the model (3.47) only materialise in the covariance structure, and the test of Aue et al. (2009) is particularly designed for detecting break points of this structure. Since our approach is able to detect break points in a much wider class of alternatives one might expect a loss in power compared to a procedure which is specifically constructed to test for such changes.

We continue our comparison by considering a selection of more complex piecewise stationary times series models. Therefore, we consider the five dimensional piecewise stationary MA(1) model

\[ X_{t,T} = \sum_{l=0}^{K} 1[[b_lT+1, [b_{l+1}]T]](t) \Theta_l Z_{t-1} + Z_t, \] (3.48)

where \( K = 2, (b_0, b_1, b_2, b_3) = (0, 1/2, 3/4, 1) \) and the matrices \( \Theta_0, \Theta_1 \) and \( \Theta_2 \) are defined by

\[
\Theta_0 := \begin{pmatrix}
1 & -1 & -1 & -1 & -1 \\
1 & 1 & -1 & -1 & -1 \\
1 & 1 & 1 & -1 & -1 \\
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1
\end{pmatrix},
\Theta_1 := \begin{pmatrix}
1 & -1 & -1 & -1 & -1 \\
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1
\end{pmatrix},
\Theta_2 := \begin{pmatrix}
1 & -1 & -1 & -1 & -1 \\
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1
\end{pmatrix},
\]

and the two-dimensional AR(2) process defined by

\[
X_{t,T} = \begin{cases}
\begin{pmatrix}
0.9 & 0 \\
0 & 1.32
\end{pmatrix} X_{t-1,T} + \begin{pmatrix}
0 & 0 \\
0 & -0.81
\end{pmatrix} X_{t-2,T} & \text{if } \frac{1}{7} \leq \frac{t}{T} \leq \frac{1}{2} \\
\begin{pmatrix}
1.68 & 0 \\
0 & 1.68
\end{pmatrix} X_{t-1,T} + \begin{pmatrix}
-0.81 & 0 \\
0 & -0.81
\end{pmatrix} X_{t-2,T} & \text{if } \frac{1}{2} < \frac{t}{T} \leq \frac{3}{4} \\
\begin{pmatrix}
1.32 & 0 \\
0 & 0.90
\end{pmatrix} X_{t-1,T} + \begin{pmatrix}
-0.81 & 0 \\
0 & 0
\end{pmatrix} X_{t-2,T} & \text{if } \frac{3}{4} < \frac{t}{T} \leq 1
\end{cases}
+ Z_t, \quad (3.49)
\]

where \( u \in [0, 1] \) and \( \{ Z_t \}_{t \in \mathbb{Z}} \) is a two dimensional centred Gaussian sequence with \( \text{Var}(Z_{t,1}) = \text{Var}(Z_{t,2}) = 1 \); \( \text{Cov}(Z_{t,1}, Z_{t,2}) = 0.5 \). Furthermore, we investigate how the new procedure performs if the assumptions of the asymptotic theory are not satisfied. Therefore, we include the bivariate smoothly changing AR(1) model

\[
X_{t,T} := \Phi(\frac{t}{T}) X_{t-1,T} + Z_t, \quad (3.50)
\]

where the matrix \( \Phi(u) \) is defined by

\[
\Phi(u) := I_{[0, 1/2]}(u) \begin{pmatrix}
0.8 & 0.3 \\
-0.6 & 0.1
\end{pmatrix} + I_{(1/2, 9/16)}(u) \begin{pmatrix}
13.6 - 25.6u & 0.3 \\
-0.6 & 0.1
\end{pmatrix} + I_{(9/16, 1]}(u) \begin{pmatrix}
-0.8 & 0.3 \\
-0.6 & 0.1
\end{pmatrix},
\]
and the piecewise locally stationary AR(1) and MA(1) models

\[ X_{t,T} = \Phi \left( \frac{t}{T} \right) X_{t-1,T} + Z_t, \quad (3.51) \]
\[ X_{t,T} = \Theta \left( \frac{t}{T} \right) Z_{t-1} + Z_t, \quad (3.52) \]

where the time-varying matrices \( \Phi(u) \) and \( \Theta(u) \) are defined by

\[ \Phi(u) = \Theta(u) := \begin{pmatrix} 0.8 - 3.2u & 0.3 \\ -0.6 & 0.1 \end{pmatrix} I_{[0,1/2]}(u) + \begin{pmatrix} 2.4 - 3.2u & 0.3 \\ -0.6 & 0.1 \end{pmatrix} I_{(1/2,1]}(u). \]

Note that the models (3.51) and (3.52) describe locally stationary AR(1)- and MA(1)-models, which have one structural break at time \( b_1 = 1/2 \). We also consider the piecewise stationary and piecewise locally stationary models

\[ X_{t,T} := \sum_{j=0}^{6} 1_{[b_jT+1, b_{j+1}T]}(t) \sigma_j Z_t, \quad (3.53) \]
\[ X_{t,T} := \sum_{j=0}^{6} 1_{[b_jT+1, b_{j+1}T]}(t) \left( 1 + \sin \left( \frac{t}{T} \right) \right) \sigma_j Z_t, \quad (3.54) \]

where \( \{ Z_t \}_{t \in \mathbb{Z}} \) is a two dimensional centred Gaussian process with \( \text{Var}(Z_{t,1}) = \text{Var}(Z_{t,2}) = 1; \text{Cov}(Z_{t,1}, Z_{t,2}) = 0.5 \) and \( (b_1, \ldots, b_6) = (2/16, 4/16, 5/16, 9/16, 11/16, 14/16) \) and \( (\sigma_0, \ldots, \sigma_6) = (1, 2, 1, 2, 1, 2, 1) \). We remark that models (3.51), (3.52) and (3.54) do not posses a representation of the form (2.8) and thus do not formally fulfil the assumptions imposed to derive the asymptotic theory. However, the results needed to show the consistency for the estimation procedure can be extended to the setting of piecewise locally stationary time series models and we therefore include these models into our empirical analysis.

Table 3.3 summarises the simulated rejection probabilities of the new test (3.22) and the procedure of Aue et al. (2009) for these more sophisticated models. It can be seen that for the MA(1) model (3.48) the new test performs similarly as the procedure of Aue et al. (2009) for large sample sizes, while the power is smaller in the case \( T = 128 \). In the AR(2)-model (3.49) the test (3.22) yields larger power. From the results corresponding to the model (3.50), it can be observed that both procedures are able to detect smooth changes in the coefficients of an AR(1)-model and that the new test shows slightly better performance in small sample sizes. For the piecewise locally stationary AR- and MA-models (3.51) and (3.52), we recognise that the test of Aue et al. (2009) yields higher power values than the new test. Finally, in the models (3.53) and (3.54), which are characterised by piecewise constant and smoothly varying variance functions with multiple break points respectively, the new procedure has clear advantages.

**Finite sample properties of the localisation method**

In this paragraph, we illustrate the performance of the procedure for estimating the amount and the locations of the break points. We first consider the models (3.48), (3.50), (3.53)
and (3.54). For each of these models, we generated a histogram that shows the empirical distribution of the estimated break points $\widehat{b} = (\widehat{b}_1, \ldots, \widehat{b}_K)$ in 100 simulation runs. These results are graphically depicted in Figure 3.3. From the upper left part of Figure 3.3, we observe that in the five dimensional MA(1) model (3.48) both break points are correctly localised and that the precision grows in lockstep with the sample size. Furthermore, it can be seen that the first break point $b_1 = 1/2$ is less frequently detected, which can be explained by the fact that this structural break is less pronounced. The right upper part of Figure 3.3 contains the results for the non-stationary AR(1)-model (3.50), and we conclude from this histogram that the new method is also able to detect a smooth transition from one stationary segment to another. In the lower part of Figure 3.3, the empirical distributions of the break point estimators in models (3.53) (left panel) and (3.54) (right panel) are shown. From these results, we conclude that the proposed detection method is well suited for estimating multiple break points, which separate stationary and locally stationary segments of different length.

We continue our investigation of the finite sample properties of the segmentation procedure by providing a comparison with the procedure of Davis et al. (2006). We emphasise that the method proposed by these authors assumes piecewise stationary AR-models. We applied the new method and the procedure of Davis et al. (2006) to estimate the amount and locations of the change points in the models (3.49), (3.51) and (3.52). The results of these experiments are graphically presented in Figure 3.4, which contains the empirical distribution of the estimated break points under both procedures for various sample sizes. It can be observed that in the case of the AR(2) model (3.49) the new non-parametric as well as the parametric method of Davis et al. (2006) work well and detect both structural changes. Moreover, the histograms are centred at the true break points and the empirical distributions of the break point estimators exhibit decreasing variation with increasing sample size. For small sample sizes, it is apparent that the parametric method of Davis et al. (2006) outperforms the non-parametric detection rule substantially. In the next step, we compare the relative performance of the detection rules in situations, which do not fulfil the restrictive assumption that the time series possesses locally an AR($p$)-representation. Therefore, we simulated data from the models (3.51) and (3.52) and generated the corresponding histograms of

<table>
<thead>
<tr>
<th>model</th>
<th>$T = 128$</th>
<th>$T = 256$</th>
<th>$T = 512$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3.48)</td>
<td>0.696</td>
<td>0.901</td>
<td>0.992</td>
</tr>
<tr>
<td>(3.49)</td>
<td>0.604</td>
<td>0.120</td>
<td>0.926</td>
</tr>
<tr>
<td>(3.50)</td>
<td>0.834</td>
<td>0.752</td>
<td>0.996</td>
</tr>
<tr>
<td>(3.51)</td>
<td>0.072</td>
<td>0.092</td>
<td>0.088</td>
</tr>
<tr>
<td>(3.52)</td>
<td>0.038</td>
<td>0.108</td>
<td>0.070</td>
</tr>
<tr>
<td>(3.53)</td>
<td>0.182</td>
<td>0.042</td>
<td>0.568</td>
</tr>
<tr>
<td>(3.54)</td>
<td>0.400</td>
<td>0.056</td>
<td>0.850</td>
</tr>
</tbody>
</table>

Table 3.3: Empirical rejection frequencies of the test (3.22) and the CUSUM type procedure of Aue et al. (2009). The nominal level is $\alpha = 0.05$. 

3.3 Practical application and finite sample properties

55
the break point estimators, which are displayed in the middle and lower part of Figure 3.4. In these situations, the assumptions of the Auto-PARM procedure of Davis et al. (2006) are not satisfied and it can be observed that the parametric approach does not yield reliable results. For example, in the piecewise locally stationary AR-model (3.51) it seems to deliver satisfactory results for sample sizes $T \in \{512, 1025\}$. However, for the sample size $T = 2048$ it detects too many structural changes, which indicates some inconsistency. A similar conclusion is reached for the case of a locally stationary MA(1) model of the form (3.52), which exhibits one structural break at time $b_1 = 1/2$. In comparison to these findings, the non-parametric procedure developed in this chapter constitutes a significant improvement. It seems to be more flexible with respect to the model assumptions and yields much better results.
Figure 3.4: Histograms of the empirical distribution of $\hat{b} = (\hat{b}_1, \ldots, \hat{b}_K)$ based on 100 simulation runs in models (3.49), (3.51) and (3.52) for sample sizes $T \in \{512, 1024, 2048\}$. Left panels: the new segmentation procedure. Right panels: Auto-PARM procedure of Davis et al. (2006).
### 3.4 Proofs

In this section, we present the formal proofs for the theoretical results of Chapter 3.

#### 3.4.1 Proof of Theorem 3.1.1

For notational convenience, we restrict ourselves to the case $d = 1$, since the more general scenario is treated completely analogously by using linearity arguments and the independence of the components of the random vectors $Z_t, t \in \mathbb{Z}$. Throughout this chapter, $C$ denotes a universal constant, which does not depend on the sample size and can vary from line to line in the calculations. Furthermore, throughout the following arguments we denote by $\varepsilon' > 0$ a constant that can be arbitrarily small but must be positive.

**Proof of Theorem 3.1.1 a):** From Theorem 2.3.12, it follows that the assertion of Theorem 3.1.1a) is a consequence of the following two results:

- **Lemma 3.4.1 (Convergence of the finite dimensional projections)**

  The finite dimensional projections of the process $\{\sqrt{N} \hat{E}_T(v, \omega)\}_{(v, \omega) \in [0, 1]^2}$ converge to the finite dimensional projections of the process $\{B(v, \omega)\}_{(v, \omega) \in [0, 1]^2}$, i.e. for every $k \geq 1$ and any $(v_1, \omega_1), \ldots, (v_k, \omega_k) \in [0, 1]^2$ it holds

  \[
  \sqrt{N}(\hat{E}_T(v_1, \omega_1), \ldots, \hat{E}_T(v_k, \omega_k)) \Rightarrow (B(v_1, \omega_1), \ldots, B(v_k, \omega_k)). 
  \] (3.55)

- **Lemma 3.4.2 (Asymptotic stochastic equicontinuity)**

  For every $\beta \in (0, 1/3)$, the process $\{\sqrt{N} \hat{E}_T(v, \omega)\}_{(v, \omega) \in [0, 1]^2}$ is asymptotically stochastically equicontinuous with respect to the semimetric

  \[
  d_\beta((v_1, \omega_1), (v_2, \omega_2)) := (|v_1 - v_2| + |\omega_1 - \omega_2|)^{\beta/2}, \quad (3.56)
  \]

  i.e. for every $\eta, \varepsilon > 0$, there exists a $\delta > 0$ such that

  \[
  \lim_{T \to \infty} P \left( \sup_{y_1, y_2 \in [0, 1]^2, d_\beta(y_1, y_2) < \delta} \sqrt{N} |\hat{E}_T(y_1) - \hat{E}_T(y_2)| > \eta \right) < \varepsilon, \quad (3.57)
  \]

  where $y_1 = (v_1, \omega_1)$ and $y_2 = (v_2, \omega_2)$.

We continue by formally establishing Lemmas 3.4.1 and 3.4.2.

**Proof of Lemma 3.4.1**

The assertion of Lemma 3.4.1 follows, if we show that for each $k \in \mathbb{N}$ and all $(v_1, \omega_1), \ldots, (v_k, \omega_k) \in [0, 1]^2$, all cumulants of the random vector $(\sqrt{N} \hat{E}_T(v, \omega_i))_{i=1,\ldots,k}$ converge to the corresponding cumulants of the vector $(B(v, \omega_i))_{i=1,\ldots,k}$. As the random vector...
(\(B(v_i, \omega_i)\))_{i=1, \ldots, k} is multivariate normally distributed, we obtain by Example 2.3.1 that we have to show the following claims:

1. The process \(\{\sqrt{N} \hat{E}_T(v, \omega)\}_{(v, \omega) \in [0,1]^2}\) is asymptotically centred. More specifically, for all \((v, \omega) \in [0,1]^2\) it holds
   \[
   \mathbb{E}(\sqrt{N} \hat{E}_T(v, \omega)) = O(\frac{1}{\sqrt{N}}). \tag{3.58}
   \]

2. The asymptotic covariance kernel of the empirical process \(\{\sqrt{N} \hat{E}_T(v, \omega)\}_{(v, \omega) \in [0,1]^2}\) is given by (3.12), i.e. for all \((v_1, \omega_1), (v_2, \omega_2) \in [0,1]^2\) it holds
   \[
   \text{Cov}(\sqrt{N} \hat{E}_T(v_1, \omega_1), \sqrt{N} \hat{E}_T(v_2, \omega_2)) = \text{Cov}(B(v_1, \omega_1), B(v_2, \omega_2)) + R(T), \tag{3.59}
   \]
   where the error term \(R(T)\) is uniform in \((v, \omega) \in [0,1]^2\) and of order \(O(1/N^\alpha)\) for any \(\alpha \in (0, 1/2)\).

3. The cumulants of orders \(l \geq 3\) vanish. More precisely, for all integers \(l \geq 3\) and \((v_1, \omega_1), \ldots, (v_l, \omega_l) \in [0,1]^2\) we have
   \[
   \text{cum}(\sqrt{N} \hat{E}_T(v_1, \omega_1), \ldots, \sqrt{N} \hat{E}_T(v_l, \omega_l)) \leq 2^l (2l)! C \frac{\log(N)^{l-1}}{N^{l/2-1}} = o(1). \tag{3.60}
   \]

For the proof of assertions (1), (2) and (3), we define for \(y = (v, \omega) \in [0,1]^2\) the function
\[
\phi_{y,T}(j, \lambda) := 1_{\left[ -\frac{2\pi |(N-1)/2|}{N} , \frac{2\pi |N/2|}{N} \right]}(\lambda) \left[ 1_{\{u(v,T)T+N/2\}+N/2}(j) - 1_{\{u(v,T)T-N/2\}-N/2}(j) \right], \tag{3.61}
\]
where \(u(v, T)\) is defined by
\[
u(v, T) := \begin{cases} \frac{N}{T} & \text{if } v < \frac{N}{T} \\ v & \text{if } \frac{N}{T} \leq v \leq 1 - \frac{N}{T} \\ 1 - \frac{N}{T} & \text{if } v > 1 - \frac{N}{T} \end{cases}
\]
This notation implies that the representation
\[
\hat{E}_T(y) = \frac{1}{N} \sum_{j=1}^{T} \sum_{k=-[(N-1)/2]}^{N/2} \phi_{y,T}(j, \lambda_k) I_N(\frac{j}{T}, \lambda_k) \tag{3.62}
\]
holds for all \(y \in [0,1]^2\). Furthermore, we introduce for \(y \in [0,1]^2\) the quantity
\[
E_{N,T}(y) := \frac{T}{N} \left( \int_{-\omega \pi}^{\omega \pi} \int_{-\omega N/T}^{\omega N/T} f(u, \lambda) du d\lambda - \int_{-\omega \pi}^{\omega \pi} \int_{-\omega N/T}^{\omega N/T} f(u, \lambda) du d\lambda \right) \tag{3.63}
\]

**Proof of (1):** Without loss of generality, we assume \(v \in (1/c, 1 - 1/c)\), where \(c\) denotes the
constant satisfying \( c = \lim_{T \to \infty} T/N \). This assumption implies \( u(v, T) = v \) and it is simple to see that the cases \( v \leq 1/c \) and \( v \geq 1 - 1/c \) follow by the same arguments. For a proof of (3.58), we show that the more general statement

\[
\mathbb{E}\left( \sqrt{N} \left( \hat{E}_T(y) - E_{N,T}(y) \right) \right) = O\left( \frac{1}{\sqrt{N}} \right)
\]

(3.64)

holds for all \( y \in [0, 1]^2 \) under the null hypothesis as well as under the alternative. Due to the fact that \( E_{N,T}(y) \) vanishes for all \( y \in [0, 1]^2 \) under the null hypothesis, part (1) immediately follows. [The more general statement (3.64) will be useful in the proof of part b) of Theorem 3.1.1]. For notational convenience, we define \( \psi_l(t/T) := \Psi_l(t/T) \) and obtain for any \( \omega \in [0, 1] \)

\[
\mathbb{E}(\hat{E}_T(y)) = \frac{1}{2\pi N^2} \sum_{j=1}^{N/2} \sum_{k=-(N-1)/2}^{N/2} \phi_{y,T}(j, \lambda_k) \sum_{p,q=0}^{N-1} \exp(-i\lambda_k(p - q))
\]

\[
\times \mathbb{E}(X_{j-\frac{N}{2}+1+p,T}X_{j-\frac{N}{2}+1+q,T})
\]

\[
= \frac{1}{2\pi N^2} \sum_{j=1}^{T} \sum_{k=-(N-1)/2}^{N/2} \phi_{y,T}(j, \lambda_k) \sum_{l,m=0}^{\infty} \psi_l(j - \frac{N}{2} + 1 + p) \psi_m(j - \frac{N}{2} + 1 + q)
\]

\[
\times \sum_{p,q=0}^{N-1} \exp(-i\lambda_k(p - q)) \mathbb{E}(Z_{j-\frac{N}{2}+1+p-l}Z_{j-\frac{N}{2}+1+q-m}),
\]

(3.65)

where we employed the piecewise stationary linear representation (2.9) for the time series \( \{X_t,T\}_{t=1,...,T} \). The independence of the innovation process \( \{Z_t\}_{t \in \mathbb{Z}} \) implies the identity

\[
\mathbb{E}(Z_i Z_j) = \delta_{ij} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}
\]

(3.66)

which yields that the condition \( q = p - l + m \) has to hold such that the respective summands in (3.65) do not vanish. By employing the notation

\[
A_{T,1}(v) := \left\{ [vT] - \frac{N}{2}, \left[ vT \right] + \frac{N}{2} \right\}
\]

(3.67)
for \( v \in (1/c, 1 - 1/c) \), we obtain that the expression (3.65) is equal to

\[
\frac{1}{2\pi N^2} \sum_{j \in A_{T,1}(v)} \sum_{k=-(N-1)/2}^{N/2} \phi_{y,T}(j, \lambda_k) \sum_{l,m=0}^{\infty} \sum_{p=0}^{N-1} \frac{1}{0 \leq p - l + m \leq N-1} \psi_l\left(\frac{j - \frac{N}{2} + 1 + p}{T}\right) \psi_m\left(\frac{j - \frac{N}{2} + 1 + p - l + m}{T}\right) \exp\left(-i\lambda_k (m - l)\right)
\]

\[
= \frac{1}{2\pi N^2} \sum_{j \in A_{T,1}(v)} \sum_{k=-(N-1)/2}^{N/2} \phi_{y,T}(j, \lambda_k) \sum_{l,m=0}^{\infty} \sum_{p=0}^{N-1} \frac{1}{0 \leq p - l + m \leq N-1} \psi_l\left(\frac{j - \frac{N}{2} + 1 + p}{T}\right)
\]

\[
\times \psi_m\left(\frac{j - \frac{N}{2} + 1 + p}{T}\right) + \psi_l\left(\frac{j - \frac{N}{2} + 1 + p}{T}\right) \exp\left(-i\lambda_k (m - l)\right)
\]

\[
\times \left(\psi_m\left(\frac{j - \frac{N}{2} + 1 + p - l + m}{T}\right) - \psi_m\left(\frac{j - \frac{N}{2} + 1 + p}{T}\right)\right) \exp\left(-i\lambda_k (m - l)\right)
\]

\[
=: I_{1,T}(y) + I_{2,T}(y), \quad (3.68)
\]

where the quantities \( I_{1,T} \) and \( I_{2,T} \) are defined by

\[
I_{1,T}(y) := \frac{1}{2\pi N^2} \sum_{j \in A_{T,1}(v)} \sum_{k=-(N-1)/2}^{N/2} \phi_{y,T}(j, \lambda_k) \sum_{l,m=0}^{\infty} \sum_{p=0}^{N-1} \psi_l\left(\frac{j - \frac{N}{2} + 1 + p}{T}\right)
\]

\[
\times \psi_m\left(\frac{j - \frac{N}{2} + 1 + p}{T}\right) \exp\left(-i\lambda_k (m - l)\right)
\]

\[
I_{1,T}(y) := \frac{1}{2\pi N^2} \sum_{j \in A_{T,1}(v)} \sum_{k=-(N-1)/2}^{N/2} \phi_{y,T}(j, \lambda_k) \sum_{l,m=0}^{\infty} \sum_{p=0}^{N-1} \frac{1}{0 \leq p - l + m \leq N-1} \psi_l\left(\frac{j - \frac{N}{2} + 1 + p}{T}\right)
\]

\[
\times \left(\psi_m\left(\frac{j - \frac{N}{2} + 1 + p - l + m}{T}\right) - \psi_m\left(\frac{j - \frac{N}{2} + 1 + p}{T}\right)\right) \exp\left(-i\lambda_k (m - l)\right)
\]

The assertion (3.64) is a consequence of the following two claims, the proof of which is our next objective.

\[
I_{1,T}(y) = E_{N,T}(y) + O\left(\frac{1}{N}\right), \quad (3.69)
\]

\[
I_{2,T}(y) = O\left(\frac{1}{N}\right). \quad (3.70)
\]

Regarding (3.69), we note that we can drop the restriction \( 0 \leq p - l + m \leq N - 1 \) in the summation with respect to \( p \) by making an error which is bounded by the term

\[
C \frac{1}{2\pi N^2} \sum_{k=-(N-1)/2}^{N/2} \sum_{l,m=0}^{\infty} \sup_{u \in [0,1]} |m - l| \sup_{u \in [0,1]} |\psi_l(u)| \sup_{u \in [0,1]} |\psi_m(u)|
\]
and therefore is of order $O(1/N)$ due to condition (3.10). Thus, we obtain that $I_{1,T}$ is equal to

$$\frac{1}{2\pi N^2} \sum_{j \in A_{T,1(v)} \cap (-N^1/2)}^{N^1/2} \sum_{k=-N^1/2}^{N^1/2} \phi_{y,T}(j, \lambda_k) \sum_{l,m=0}^{\infty} \sum_{p=1}^{N-1} \psi_l(j - N^1/2 + 1 + p) \psi_m(j - N^1/2 + 1 + p) \times \exp(-i\lambda_k(m - l)) + O(\frac{1}{N})$$

$$= \frac{1}{N} \sum_{j \in A_{T,1(v)} \cap (-N^1/2)}^{N^1/2} \phi_{y,T}(j, \lambda_k) \sum_{l,m=0}^{\infty} \sum_{p=1}^{N-1} \frac{T}{N} \int_{j/T - \pi N^1/2}^{j/T + \pi N^1/2} \psi_l(u) \psi_m(u) \exp(-i\lambda_k(m - l)) du + O(\frac{1}{N})$$

$$= E_{N,T}(y) + O(\frac{1}{N}),$$  \hspace{1cm} (3.71)

where the second equality follows from the piecewise constancy of the functions $\psi_l(u)$. With regards to the claim (3.70), we first note that, because of $\psi$, where the second equality follows from the piecewise constancy of the functions $\psi_l(u)$. Under the null hypothesis of no structural breaks, the linear coefficients $\psi_l(t/T)$ of the time series $\{X_{t,T}\}_{t=1,\ldots,T}$ do not depend on the rescaled time $u = t/T$. Therefore, $|\psi_l(t/T)|$ is bounded by $O(1/N)$. For the proof of part (2), we assume without loss of generality that $\omega_1 \leq \omega_2$ and note that, under the null hypothesis of no structural breaks, the linear coefficients $\psi_l(t/T)$ of the time series $\{X_{t,T}\}_{t=1,\ldots,T}$ do not depend on the rescaled time $u = t/T$. Therefore, $|\psi_l(t/T)|$ is bounded by $O(1/N)$. \hspace{1cm} (3.72)
fore, we define $ψ_l := ψ_l(u)$ and employ the representation

$$X_{l,T} = \sum_{l=0}^{∞} ψ_l Z_{l-l}$$

in the following demonstrations. Without loss of generality, we assume that $v_1, v_2 \in (1/c, 1-1/c)$ and for notational convenience we define the set

$$A_{T,2}(v_1, v_2) := A_{T,1}(v_1) \times A_{T,1}(v_2),$$

(3.73)

where the set $A_{T,1}(v)$ for $v \in (1/c, 1-1/c)$ was defined in (3.67). Furthermore, we introduce the distance

$$Δ(j_1, j_2) := |j_1 - j_2|$$

for $(j_1, j_2) \in A_{T,2}(v_1, v_2)$. The definition of the functions $φ_{y,T} \psi$ given in (3.61) implies that

$$φ_{y,T}(j_1, λ_{k_1})φ_{y,T}(j_2, λ_{k_2}) = 0 \text{ for } (j_1, j_2) \notin A_{T,2}(v_1, v_2)$$

and we thus obtain

$$\text{Cov}(\sqrt{N} \hat{E}_T(y_1), \sqrt{N} \hat{E}_T(y_2))$$

$$= \frac{1}{N} \sum_{(j_1, j_2) \in A_{T,2}(v_1, v_2)} \sum_{k_1, k_2 = -\lfloor (N-1)/2 \rfloor}^{N/2} φ_{y,T}(j_1, λ_{k_1})φ_{y,T}(j_2, λ_{k_2}) \text{cum}(I_N(j_1 \frac{T}{T}, λ_{k_1}), I_N(j_2 \frac{T}{T}, λ_{k_2}))$$

$$= \frac{1}{(2\pi)^2 N^3} \sum_{(j_1, j_2) \in A_{T,2}(v_1, v_2)} \sum_{k_1, k_2 = -\lfloor (N-1)/2 \rfloor}^{N/2} φ_{y,T}(j_1, λ_{k_1})φ_{y,T}(j_2, λ_{k_2}) \sum_{p_1, p_2 = 0}^{N-1} \sum_{q_1, q_2 = 0}^{N-1} \sum_{l, m, n, o = 0}^{∞}$$

$$\times ψ_l ψ_{m} ψ_{n} ψ_{o} \exp(-iλ_{k_1}(p_1 - q_1)) \exp(-iλ_{k_2}(p_2 - q_2))$$

$$\times \text{cum}(Z_{j_1 + p_1 + 1 - l} Z_{j_1 + q_1 + 1 - m}, Z_{j_2 + p_2 + 1 - l} Z_{j_2 + q_2 + 1 - o})$$

$$= \sum_{(j_1, j_2) \in A_{T,2}(v_1, v_2)} B_T(j_1, j_2),$$

(3.74)

where for $(j_1, j_2) \in A_{T,2}(v_1, v_2)$ the term $B_T(j_1, j_2)$ is defined by

$$B_T(j_1, j_2) := \frac{1}{(2\pi)^2 N^3} \sum_{k_1, k_2 = -\lfloor (N-1)/2 \rfloor}^{N/2} φ_{y,T}(j_1, λ_{k_1})φ_{y,T}(j_2, λ_{k_2}) \sum_{p_1, p_2 = 0}^{N-1} \sum_{q_1, q_2 = 0}^{N-1} \sum_{l, m, n, o = 0}^{∞}$$

$$\times ψ_l ψ_{m} ψ_{n} ψ_{o} \exp(-iλ_{k_1}(p_1 - q_1)) \exp(-iλ_{k_2}(p_2 - q_2))$$

$$\times \text{cum}(Z_{j_1 + p_1 + 1 - l} Z_{j_1 + q_1 + 1 - m}, Z_{j_2 + p_2 + 1 - l} Z_{j_2 + q_2 + 1 - o}).$$

The product theorem for cumulants [see Theorem 2.3.6] and the Gaussianity of the $Z_t$ imply

$$\text{cum}(Z_a Z_b, Z_c Z_d) = \text{cum}(Z_a, Z_d) \text{cum}(Z_b, Z_c) + \text{cum}(Z_a, Z_c) \text{cum}(Z_b, Z_d)$$
A new approach to the change point problem in time series analysis

[see Example 2.3.1]. Hence, the term \( B_T(j_1, j_2) \) can be split up into

\[
\frac{1}{(2\pi)^2 N^2} \sum_{k_1, k_2 = -[(N-1)/2]}^{N/2} \phi_{y_1, T}(j_1, \lambda_{k_1}) \phi_{y_2, T}(j_2, \lambda_{k_2}) \sum_{p_1, p_2 = 0}^{N-1} \sum_{q_1, q_2 = 0}^{N-1} \sum_{l, m, n, o = 0}^{\infty} \psi_l \psi_m \psi_n \psi_o \\
x \exp(-i\lambda_{k_1}(p_1 - q_1)) \exp(-i\lambda_{k_2}(p_2 - q_2)) \\
\times \left[ \text{cum}(Z_{j_1 + p_1 + 1 - l}, Z_{j_2 + q_2 + 1 - o}) \text{cum}(Z_{j_1 + q_1 + 1 - m}, Z_{j_2 + p_2 + 1 - n}) \\
+ \text{cum}(Z_{j_1 + p_1 + 1 - l}, Z_{j_2 + p_2 + 1 - n}) \text{cum}(Z_{j_1 + q_1 + 1 - m}, Z_{j_2 + q_2 + 1 - o}) \right] \\
= V_{1, T}(j_1, j_2) + V_{2, T}(j_1, j_2),
\]

(3.75)

where \( V_{1, T}(j_1, j_2) \) and \( V_{2, T}(j_1, j_2) \) are defined implicitly. For the following arguments, let \( a_T \) denote some sequence satisfying

\[
a_T \equiv 0 \quad \text{and} \quad \frac{a_T}{N^\alpha} \to 0
\]

(3.76)

for some \( \alpha > 0 \) as \( T \to \infty \). We proceed by demonstrating that the claims

\[
V_{1, T}(j_1, j_2) = (1 - \Delta(j_1, j_2)) \frac{1}{N} \sum_{k = -[(N-1)/2]}^{N/2} \phi_{y_1, T}(j_1, \lambda_k) \phi_{y_2, T}(j_2, \lambda_k) f^2(\lambda_k) + R(T),
\]

(3.77)

\[
V_{2, T}(j_1, j_2) = (1 - \Delta(j_1, j_2)) \frac{1}{N} \sum_{k = -[(N-1)/2]}^{N/2} \phi_{y_1, T}(j_1, \lambda_k) \phi_{y_2, T}(j_2, \lambda_k) f^2(\lambda_k) + R(T)
\]

(3.78)

hold uniformly in \( y_1, y_2 \in [0, 1]^2 \), where the error term \( R(T) \) is of order \( O(1/a_T^{1-\varepsilon} + a_T/N^{1-\varepsilon}) \) uniformly in \( y_1, y_2 \in [0, 1]^2 \).

For a proof of (3.77), we note that it follows from the property

\[
\text{cum}(Z_i, Z_j) = \begin{cases} 
1 & \text{if } i = j \\
0 & \text{else} 
\end{cases}
\]

(3.79)

that the conditions

\[
q_2 = p_1 - l + o + j_1 - j_2,
\]

(3.80)

\[
q_1 = p_2 - n + m + j_2 - j_1
\]

(3.81)
have to hold. This implies that $V_{1,T}(j_1, j_2)$ is equal to

$$\frac{1}{(2\pi)^2 N^3} \sum_{k_1, k_2 = -(N-1)/2}^{N/2} \phi_{y_1,T}(j_1, \lambda_{k_1}) \phi_{y_2,T}(j_2, \lambda_{k_2}) \sum_{l,m,n,o=0}^\infty \psi_l \psi_m \psi_n \psi_o$$

$$\times \sum_{p_1,p_2=0}^{N-1} \exp(-i\lambda_{k_1}(p_1 - p_2 + n - m - j_2 + j_1)) \exp(-i\lambda_{k_2}(p_2 - p_1 + l - o - j_1 + j_2))$$

$$- \frac{1}{(2\pi)^2 N^3} \sum_{k_1, k_2 = -(N-1)/2}^{N/2} \phi_{y_1,T}(j_1, \lambda_{k_1}) \phi_{y_2,T}(j_2, \lambda_{k_2}) \sum_{l,m,n,o=0}^\infty \psi_l \psi_m \psi_n \psi_o \exp(-i\lambda_{k_1}(n - m))$$

$$\times \exp(-i\lambda_{k_2}(l - o)) \sum_{p_1,p_2=0}^{N-1} \exp(-i(\lambda_{k_1} - \lambda_{k_2})(p_1 - p_2 - j_2 + j_1))$$

$$=: V_{1,k_1=k_2}(j_1, j_2) + V_{1,k_1\neq k_2}(j_1, j_2), \quad (3.82)$$

where $V_{1,k_1=k_2}(j_1, j_2)$ and $V_{1,k_1\neq k_2}(j_1, j_2)$ denote the summations over all $(k_1, k_2)$ with $k_1 = k_2$ and $k_1 \neq k_2$ respectively. Here, the summation over $p_1$ and $p_2$ is performed with respect to the restrictions

$$0 \leq p_1 - l + o + j_1 - j_2 \leq N - 1,$$

$$0 \leq p_2 - n + m + j_2 - j_1 \leq N - 1,$$

which follow from the equalities (3.80) and (3.81) because of $q_1, q_2 \in \{0, ..., N - 1\}$. We first consider the quantity $V_{1,k_1=k_2}(j_1, j_2)$, for which we obtain by simple calculations

$$V_{1,k_1=k_2}(j_1, j_2) = \frac{1}{(2\pi)^2 N^3} \sum_{k_1, k_2 = -(N-1)/2}^{N/2} \phi_{y_1,T}(j_1, \lambda_{k_1}) \phi_{y_2,T}(j_2, \lambda_{k_2}) \sum_{l,m,n,o=0}^\infty \psi_l \psi_m \psi_n \psi_o$$

$$\times \exp(-i\lambda_{k_2}(n - m)) \exp(-i\lambda_{k_2}(l - o))$$

$$\times \max(N - 1 - |j_1 - j_2 - l + o|, 0) \max(N - 1 - |j_2 - j_1 + m - n|, 0). \quad (3.85)$$

The summability condition (3.10) implies that for $(j_1, j_2) \in A_{T,2}(v_1, v_2)$ satisfying $\Delta(j_1, j_2) > N$ the term $|V_{1,k_1=k_2}(j_1, j_2)|$ is bounded by

$$\frac{C}{N^3} \sum_{k, j = -(N-1)/2}^N \phi_{y_1,T}(j_1, \lambda_{k_1}) \phi_{y_2,T}(j_2, \lambda_{k_2}) \sum_{l,m,n,o=0}^\infty |\psi_l \psi_m \psi_n \psi_o| |o - l||m - n| = O\left(\frac{1}{N^3}\right).$$
For \((j_1, j_2) \in A_{T,2}(v_1, v_2)\) satisfying \(\Delta(j_1, j_2) < N\), the bound

\[
\left| \frac{1}{(2\pi)^2 N^3} \sum_{k=-[(N-1)/2]}^{N/2} \overline{O_{m,n}} O_{l,m,n,o} \sum_{l,m,n,o=0}^\infty \psi_l \psi_m \psi_n \psi_o \times \left[ \max(N - 1 - |j_1 - j_2 - l + o|, 0) \max(N - 1 - |j_2 - j_1 + m - n|, 0) 
- \max(N - 1 - |j_1 - j_2|, 0) \max(N - 1 - |j_2 - j_1|, 0) \right] \right| \\
\leq \frac{C}{N^3} \sum_{k=-[(N-1)/2]}^{N/2} \sum_{l,m,n,o=0}^\infty \|\psi_l \psi_m \psi_n \psi_o\| \left( N |o - l| + N |m - n| \right) = O\left( \frac{1}{N} \right),
\]

which is due to the summability condition (3.10), warrants that we can drop the \(l, m, n\) and \(o\) terms in the arguments of the \(O(\cdot)\) functions by including an error term of order \(O(1/N)\). Therefore, it follows from (3.85) that

\[
V_{1,k_1=k_2}(j_1, j_2) = \frac{1}{(2\pi)^2 N^3} \sum_{k=-[(N-1)/2]}^{N/2} \phi_{y_1, T}(j_1, \lambda_k) \phi_{y_2, T}(j_2, \lambda_k) \sum_{l,m,n,o=0}^\infty \psi_l \psi_m \psi_n \psi_o \\
\times \exp(-i\lambda_k (n - m)) \exp(-i\lambda_k (l - o)) \\
\times \max(N - 1 - |j_1 - j_2|, 0) \max(N - 1 - |j_2 - j_1|, 0) + O\left( \frac{1}{N} \right) \\
= \frac{1}{N} (1 - \Delta(j_1, j_2))^2 \sum_{k=-[(N-1)/2]}^{N/2} \phi_{y_1, T}(j_1, \lambda_k) \phi_{y_2, T}(j_2, \lambda_k) f^2(\lambda_k) + O\left( \frac{1}{N} \right).
\]

Now, we consider the quantity \(V_{1,k_1\neq k_2}(j_1, j_2)\) arising from the application of the product theorem in (3.75). In a first step, we show that the error that is made by changing the restrictions in the summation over \(p_1\) and \(p_2\) from (3.83) and (3.84) to

\[
0 \leq p_1 + j_1 - j_2 \leq N - 1, \\
0 \leq p_2 + j_2 - j_1 \leq N - 1,
\]

is of order \(O(1/N^1-c')\) for any \(c' > 0\). For this purpose and later application, we note that the geometric series formula and the identity \(|1 - \exp(ix)| = 2|\sin(x/2)|\) imply that the equality

\[
\left| \sum_{p_1=0}^{N-1} \exp(-i(\lambda_{k_1} - \lambda_{k_2})(p_1 - j_2 + j_1)) \right| = \left| \sum_{p=0}^{N-1-\Delta(j_1,j_2)} \exp(-i(\lambda_{k_1} - \lambda_{k_2})p) \right| \\
= \left| 1 - \exp\left( \frac{2\pi i (k_1 - k_2) \Delta(j_1,j_2)}{N} \right) \right| = \left| \frac{\sin(\frac{\pi (k_1 - k_2) \Delta(j_1,j_2)}{N})}{\sin(\frac{\pi (k_1 - k_2)}{N})} \right|
\]

(3.88)
holds for $k_1 \neq k_2$ and $(j_1, j_2) \in A_{T,2}(v_1, v_2)$ with $\Delta(j_1, j_2) < N$. This identity shows that in the definition of $V_{1,k_1 \neq k_2}(j_1, j_2)$ given in (3.82) the condition in the summation over $p_2$ can be changed to $0 \leq p_2 + j_2 - j_1 \leq N - 1$ by committing an error which is not larger than

$$\sum_{k_1, k_2 = -[(N-1)/2]}^{N/2} \sum_{l, m, n, o=0}^{\infty} |\psi_l \psi_m \psi_n \psi_o| |m - n|$$

Due to (3.10), the above term is bounded by

$$\frac{C}{N^3} \sum_{k_1, k_2 = -[(N-1)/2]}^{N/2} \left| \frac{1}{\sin(\frac{\pi (k_1 - k_2)}{N})} \right| + \frac{C}{N^3} \sum_{k_1, k_2 = -[(N-1)/2]}^{N/2} \left| \frac{1}{\sin(\frac{\pi (k_1 - k_2)}{N})} \right|.$$

Next, we make use of the fact that there exists a constant $C > 0$ such that the inequality

$$|\sin(\pi x)| > \begin{cases} C|x| & \text{if } |x| \in [0, 1/2] \\ C(1 - |x|) & \text{if } |x| \in (1/2, 1] \end{cases}$$

holds. This bound implies that (3.89) is at most of order

$$\frac{C}{N^2} \sum_{k_1, k_2 = -[(N-1)/2]}^{N/2} \frac{1}{|k_1 - k_2|} + \frac{C}{N^2} \sum_{k_1, k_2 = -[(N-1)/2]}^{N/2} \frac{1}{|N - |k_1 - k_2||}$$

where the last equality holds for any $\varepsilon' > 0$. Thus, we have shown that the error which is made by changing the condition (3.84) to (3.87) is of order $O(1/N^{1-\varepsilon'})$. The same arguments yield that the restriction in the summation over $p_1$ can be changed from (3.83) to (3.86) by
including an error term of order $O(1/N^{1-\varepsilon'})$. Hence, we get

$$V_{1,k_1 \neq k_2}(j_1, j_2) = \frac{1}{(2\pi)^2 N^3} \sum_{k_1, k_2 = -[(N-1)/2]}^{N/2} \phi_{y_1,T}(j_1, \lambda_{k_1}) \phi_{y_2,T}(j_2, \lambda_{k_2}) \sum_{l,m,n,o=0}^{\infty} \psi_l \psi_m \psi_n \psi_o$$

$$\times \exp(-i\lambda_{k_1}(n - m)) \exp(-i\lambda_{k_2}(l - o)) \sum_{(p_1, p_2)=0}^{N-1} \phi_{y_2,T}(j_2, \lambda_{k_2}) \sum_{l,m,n,o=0}^{\infty} \psi_l \psi_m \psi_n \psi_o$$

$$\times \exp(-i(\lambda_{k_1} - \lambda_{k_2})(p_1 - p_2 - j_2 + j_1)) + O\left(\frac{1}{N^{1-\varepsilon'}}\right). \quad (3.91)$$

The identity (3.88) and the bound (3.90) for the sin function imply that in (3.91) the sum over all $(k_1, k_2)$ satisfying

$$a_T < |k_1 - k_2| < N - a_T$$

is bounded by

$$\leq \frac{C}{N} \sum_{k_1, k_2 = -[(N-1)/2]}^{N/2} \frac{1}{N^2 \sin^2\left(\frac{\pi (k_1 - k_2)}{N}\right)} + \frac{C}{N} \sum_{k_1, k_2 = -[(N-1)/2]}^{N/2} \frac{1}{N^2 \sin^2\left(\frac{\pi (k_1 - k_2)}{N}\right)}$$

$$\leq \frac{C}{N} \sum_{k_1 = -[(N-1)/2]}^{N} \frac{1}{a_T^{2-\varepsilon'}} \sum_{k_2 = 0}^{N/2} \frac{1}{k_2^{2+\varepsilon'}} = O\left(\frac{1}{a_T^{-\varepsilon'}}\right)$$

for any $\varepsilon' > 0$ and we obtain

$$V_{1,k_1 \neq k_2}(j_1, j_2) = \frac{1}{(2\pi)^2 N^3} \sum_{k_1, k_2 = -[(N-1)/2]}^{N/2} \phi_{y_1,T}(j_1, \lambda_{k_1}) \phi_{y_2,T}(j_2, \lambda_{k_2}) \sum_{l,m,n,o=0}^{\infty} \psi_l \psi_m \psi_n \psi_o$$

$$\times \exp(-i\lambda_{k_1}(n - m)) \exp(-i\lambda_{k_2}(l - o))$$

$$\times \sum_{(p_1, p_2)=0}^{N-1} \exp(-i(\lambda_{k_1} - \lambda_{k_2})(p_1 - p_2 - j_2 + j_1)) + O\left(\frac{1}{N^{1-\varepsilon'}} + \frac{1}{a_T^{-\varepsilon'}}\right), \quad (3.92)$$

where the summation over $(k_1, k_2)$ is performed with respect to the restriction

$$|k_1 - k_2| \leq a_T \quad \text{or} \quad |k_1 - k_2| \geq N - a_T. \quad (3.93)$$
3.4 Proofs

An application of the identity (3.88) shows that (3.92) equals

\[
\frac{1}{N^3} \sum_{\substack{k_1, k_2=-(N-1)/2 \\ k_1 \neq k_2, (3.93)}}^{N/2} \phi_{y_1, T}(j_1, \lambda_{k_1}) \phi_{y_2, T}(j_2, \lambda_{k_2}) f(\lambda_{k_1}) f(\lambda_{k_2}) \left| \frac{\sin(\pi(k_1-k_2) \Delta)}{\sin(\pi(k_2-k_1) N)} \right|^2 \\
+ O\left( \frac{1}{N^{1-\varepsilon}} + \frac{1}{a_T^{1-\varepsilon}} \right),
\]

(3.94)

Our objective for the further treatment of the quantity in (3.94) is to change \(\lambda_{k_2}\) to \(\lambda_{k_1}\) in the argument of the function \(f\). For this purpose, we first establish an upper bound for the error involved in this transition. By means of a Taylor expansion, we get that this error is at most of size

\[
\leq \frac{C}{N^3} \sum_{\substack{k_1, k_2=-(N-1)/2 \\ k_1 \neq k_2, (3.95)}}^{N/2} \frac{|k_1 - k_2|}{N} \left| \frac{\sin(\pi(k_1-k_2) \Delta)}{\sin(\pi(k_1-k_2) N)} \right|^2 \\
\leq \frac{C}{N^2} \sum_{k_1, k_2=-(N-1)/2}^{N/2} \frac{|k_1 - k_2|}{N} \left| \frac{1}{\sin^2(\pi(k_1-k_2) N)} \right| + \frac{C}{N^3} \sum_{k_1, k_2=-(N-1)/2}^{N/2} \frac{|k_1 - k_2|}{N} \left| \frac{1}{\sin^2(\pi(k_1-k_2) N)} \right|
\]

\(:=S_{(3.95)} + S_{(3.96)}\),

where \(S_{(3.95)}\) and \(S_{(3.96)}\) denote the sums over \((k_1, k_2)\) with respect to the conditions

\[
1 \leq |k_1 - k_2| \leq a_T, \quad \text{(3.95)} \\
N - a_T \leq |k_1 - k_2| \leq N - 1, \quad \text{(3.96)}
\]

respectively. Now, we assume that \(T\) is sufficiently large, such that \(a_T/N \leq 1/2\) and employ the inequality (3.90) for the \(\sin\) function to get

\[
S_{(3.95)} \leq \frac{C}{N^2} \sum_{k_1, k_2=-(N-1)/2}^{N/2} \frac{1}{|k_1 - k_2|} \leq \frac{C}{N^2} \sum_{k_1=-(N-1)/2}^{N} \sum_{k_2=1}^{a_T} \frac{1}{k_2} \\
= O\left( \frac{\log(a_T)}{N} \right) = O\left( \frac{1}{N^{1-\varepsilon}} \right),
\]

(3.97)
for any $\varepsilon' > 0$. Moreover, by an application of (3.90) we get the inequality

$$S_{(3.96)} \leq C \frac{N^2}{N^2} \sum_{k_1, k_2 = -[(N-1)/2]}^{N/2} \frac{|k_1 - k_2|}{[N - (k_1 - k_2)]^2}$$

which is due to the fact that for each fixed $k_2$ the sum over $k_1$ is of order $O(1/N)$. In the next step, we substitute the sum over $k_2$ by the respective integral to obtain that the above expression is bounded by

$$\leq C \frac{N^2}{N^2} \left( \sum_{k_1 = -[(N-1)/2]}^{N/2} \int_{N/2 - a_T + [(N-1)/2] + k_1 + 1}^{N/2 - 1} \frac{(k_2 - k_1)}{[N - k_1 + k_2]^2} dk_2 \right) + O(1/N)$$

$$\leq C \frac{N^2}{N^2} \left( \sum_{k_1 = -[(N-1)/2]}^{N/2} \left[ \frac{N}{N - k_1 + t} + \log([N - k_1 + t]) \right]^{N/2 - 1}_{N/2 - a_T + [(N-1)/2] + k_1 + 1} \right)$$

$$+ O\left( \frac{1}{N} \right) = O\left( \frac{a_T \log(N)}{N} \right) = O\left( \frac{a_T}{N^{1 - \varepsilon'}} \right) \quad (3.98)$$

The bounds (3.97) and (3.98) imply that (3.94) is equal to

$$\frac{1}{N^3} \sum_{k_1, k_2 = -[(N-1)/2]}^{N/2} \phi_{y_1, t}(j_1, \lambda_{k_1})\phi_{y_2, t}(j_2, \lambda_{k_2})f^2(\lambda_{k_1}) \left| \frac{\sin(\frac{\pi(k_1 - k_2)}{N})}{\sin(\frac{\pi(k_1 - k_2)}{N})} \right|^2 + O\left( \frac{1}{a_T} + \frac{a_T}{N^{1 - \varepsilon'}} \right) \quad (3.99)$$

for any $\varepsilon' > 0$. We continue by showing that in (3.99) the sum over all $(k_1, k_2)$ with $|k_1 - k_2| \geq N - a_T$ is of vanishing order. For this purpose, we use the fact that $f \leq C$
and the inequality (3.90) to get
\[
\frac{1}{N^3} \sum_{k_1, k_2 = -\lceil (N-1)/2 \rceil}^{N/2} \phi_{y_1, T}(j_1, \lambda_{k_1}) \phi_{y_2, T}(j_2, \lambda_{k_2}) f^2(\lambda_{k_1}) \left| \frac{\sin \left( \frac{\pi(k_1-k_2)}{N} \Delta \right)}{\sin \left( \pi \frac{(k_1-k_2)}{N} \right)} \right|^2 \\
\leq C \frac{1}{N^3} \sum_{k_1, k_2 = -\lceil (N-1)/2 \rceil}^{N/2} \sin^2 \left( \frac{\pi(k_1-k_2)}{N} \right) \leq C \frac{1}{N} \sum_{k_1, k_2 = -\lceil (N-1)/2 \rceil}^{N/2} \frac{1}{[N - (k_1 - k_2)]^2} \\
= C \left( \sum_{k_1 = -\lceil (N-1)/2 \rceil}^{N/2-2} \sum_{k_2 = -\lceil (N-1)/2 \rceil + k_1 + 2}^{N/2} \frac{1}{[N - k_1 + k_2]^2} + O \left( \frac{1}{N} \right) \right)
\]

where in the last step we exploited that the sum over \( k_1 \) is of order \( O(1/N) \) for each fixed \( k_2 \). Next, we replace the sum over \( k_2 \) by the corresponding integral and obtain that the above expression is not larger than
\[
\frac{C}{N} \left( \sum_{k_1 = -\lceil (N-1)/2 \rceil}^{N/2-2} \int_{-\lceil (N-1)/2 \rceil + k_1 + 1}^{N/2} \frac{1}{[N - k_1 + k_2]^2} dk_2 \right) \\
+ \sum_{k_1 = N/2 - a_T + 1}^{N/2} \int_{-\lceil (N-1)/2 \rceil + k_1 + 1}^{N/2} \frac{1}{[N - k_1 + k_2]^2} dk_2 \\
\leq C \left( \sum_{k_1 = -\lceil (N-1)/2 \rceil}^{N/2-2} \left[ \frac{1}{[N - k_1 + k_2]^2} \right] \right) \\
+ \sum_{k_1 = N/2 - a_T + 1}^{N/2} \left[ \frac{1}{[N - k_1 + k_2]^2} \right] = O \left( \frac{a_T}{N} \right).
\]

We have just shown that (3.99) is up to an error term of order \( O(1/a_T^{1-\varepsilon'} + a_T/N^{1-\varepsilon'}) \) equal to
\[
\frac{1}{N^3} \sum_{k_1, k_2 = -\lceil (N-1)/2 \rceil}^{N/2} \phi_{y_1, T}(j_1, \lambda_{k_1}) \phi_{y_2, T}(j_2, \lambda_{k_2}) f^2(\lambda_{k_1}) \left| \frac{\sin \left( \frac{\pi(k_1-k_2)}{N} \Delta \right)}{\sin \left( \pi \frac{(k_1-k_2)}{N} \right)} \right|^2.
\]

By applying the approximation \( |\sin(x)| = |x| + E(x) \), where the error term satisfies \( E(x) \leq C x^3 \) uniformly in \( x \in [0, \pi/2] \), we obtain that for sufficiently large \( T \) the expression (3.100)
is up to an error of order $O(1/N)$ equal to

$$\frac{1}{N^3} \sum_{k_1, k_2 = -\lceil (N-1)/2 \rceil}^{N/2} \phi_{y_1, T}(j_1, \lambda_{k_1}) \phi_{y_2, T}(j_2, \lambda_{k_2}) \frac{f^2(\lambda_{k_1})}{\Delta(\lambda_{k_1})} \sin\left(\frac{\pi(k_1 - k_2)}{N} \Delta\left(j_1, j_2\right) \right)^2, \quad (3.101)$$

where we used the fact that because of $|k_1 - k_2|/N \leq a_T/N \to 0$ the error for each summand $(k_1, k_2)$ in the transition from (3.100) to (3.101) is at most of order

$$-\frac{1}{N^3} \left| \sin\left(\frac{\pi(k_1 - k_2)}{N}\right) \right|^2 \left( \frac{1}{\Delta(k_1 - k_2)} \right)^2 \left( \frac{1}{\pi(k_1 - k_2)} \right)^2 \leq \frac{1}{N^3} \left| \sin\left(\frac{\pi(k_1 - k_2)}{N}\right) \right|^2 \left( \frac{1}{\Delta(k_1 - k_2)} \right)^2 \left( \frac{1}{\pi(k_1 - k_2)} \right)^2 \leq \frac{1}{N^3} \left( 1 - C\left( \frac{k_1 - k_2}{N} \right)^2 - C\left( \frac{k_1 - k_2}{N} \right)^2 \right) = O\left( \frac{1}{N^3} \right).$$

Hence, it follows that

$$V_{1,k_1 \neq k_2}(j_1, j_2) = \frac{1}{N} \sum_{k_1 = -\lceil (N-1)/2 \rceil}^{N/2} \phi_{y_1, T}(j_1, \lambda_{k_1}) f^2(\lambda_{k_1}) \sum_{k_2 = -\lceil (N-1)/2 \rceil}^{N/2} \phi_{y_2, T}(j_2, \lambda_{k_2}) \sin\left(\frac{\pi(k_1 - k_2)}{N} \Delta(j_1, j_2) \right)^2 + O\left( \frac{1}{N^1 - \epsilon'} + \frac{a_T}{N^1 - \epsilon'} \right)$$

$$= \frac{1}{N} \sum_{k_1 = -\lceil (N-1)/2 \rceil}^{N/2} \phi_{y_1, T}(j_1, \lambda_{k_1}) f^2(\lambda_{k_1}) L_{T,y_2}(r_{k_1}, j_1, j_2) + O\left( \frac{1}{N^1 - \epsilon'} + \frac{a_T}{N^1 - \epsilon'} \right),$$

where for $y_2 = (v_2, \omega_2)$ the function $L_{T,y_2}$ is defined by

$$L_{T,y_2}(\lambda_{k_1}, j_1, j_2) := \sum_{k_2 = -\lceil (N-1)/2 \rceil}^{N/2} \phi_{y_2, T}(j_2, \lambda_{k_2}) \sin\left(\frac{\pi(k_1 - k_2)}{N} \Delta(j_1, j_2) \right)^2. \quad (3.102)$$

We continue our discussion of the quantity $V_{1,k_1 \neq k_2}(j_1, j_2)$ by investigating the function $L_{T,y_2}$. Therefore, we note that the definition of the function $\phi_{y_2, T}$ implies that $\phi_{y_2, T}(j_1, \lambda_{k_2}) = 0$ for $j_2 \notin A_{T,1}(v_2)$ and $L_{T,y_2}$ thus vanishes for $j_2 \notin A_{T,1}(v_2)$. For $j_2 \in A_{T,1}(v_2)$ we distinguish the
cases

\[ k_1 \in \{-\lfloor \omega_2 N/2 \rfloor + a_T + 1, \ldots, \lfloor \omega_2 N/2 \rfloor - a_T\}, \]
\[ k_1 \in \{\lfloor \omega_2 N/2 \rfloor + a_T + 1, \ldots, N/2\} \cup \{-\lfloor (N - 1)/2 \rfloor, \ldots, -\lfloor \omega_2 N/2 \rfloor - a_T - 1\} \]

for \( \lambda_{k_1} \) and proceed by establishing the identity

\[
L_{T,y_2}(\lambda_{k_1}, j_1, j_2) = \begin{cases} 
\frac{\Delta(j_1, j_2)(N - \Delta(j_1, j_2))}{N^2} + O\left(\frac{1}{a_T^{-\varepsilon'}}\right) & \text{if (3.102)} \\
0 & \text{if (3.103)} 
\end{cases} 
\]

for \( j_2 \in A_{T,1}(v_2) \) and any \( \varepsilon' > 0 \). For this purpose, we recall that we assumed without loss of generality that \( \omega_1 \leq \omega_2 \). Considering the case (3.102), the definition of the function \( \phi_{y_2,T} \) given in (3.61) implies that

\[
L_{T,y_2}(\lambda_{k_1}, j_1, j_2) = \sum_{k_2=-\lfloor \omega_2(N-1)/2 \rfloor}^{\lfloor \omega_2 N/2 \rfloor} \left| \frac{\sin\left(\frac{\pi(k_1-k_2)}{N}N\right)}{\pi(k_1-k_2)} \right|^2 = 2 \sum_{l=1}^{a_T} \frac{\sin^2\left(\frac{\pi\Delta(j_1, j_2)}{N}\right)}{(\pi l)^2},
\]

where we used the symmetry of \( \sin^2(x) \) and \( x^2 \) and the fact that for \( k_1 \) satisfying (3.102) the restriction \( |k_1 - k_2| \leq a_T \) is binding. The inequality

\[
\sum_{l=a_T+1}^{\infty} \frac{\sin^2\left(\frac{\pi\Delta(j_1, j_2)}{N}\right)}{(\pi l)^2} \leq C \sum_{l=a_T+1}^{\infty} \frac{1}{l^2} \leq C \frac{1}{a_T^{1-\varepsilon'}} \sum_{l=a_T+1}^{\infty} \frac{1}{l^{1+\varepsilon'}} = O\left(\frac{1}{a_T^{1-\varepsilon'}}\right),
\]

which holds for any \( \varepsilon' > 0 \), yields that (3.105) is equal to

\[
2 \sum_{l=1}^{\infty} \frac{\sin^2\left(\frac{\pi\Delta(j_1, j_2)}{N}\right)}{(\pi l)^2} + O\left(\frac{1}{a_T^{1-\varepsilon'}}\right).
\]

An application of the identities \( \sin^2(x) = \frac{1}{2}(1 - \cos(2x)) \) and \( \sum_{k=1}^{\infty} \frac{\cos(kx)}{k^2} = \left(\frac{x}{2}\right)^2 - \frac{x^2}{12} \) [see Jolley (1961)] finally shows that, for \( j_2 \in A_{T,1}(v_2) \) and \( k_1 \) satisfying (3.102), we have

\[
L_{T,y_2}(\lambda_{k_1}, j_1, j_2) = \sum_{l=1}^{\infty} \frac{1 - \cos\left(\frac{2\pi\Delta(j_1, j_2)}{N}\right)}{(\pi l)^2} + O\left(\frac{1}{a_T^{-\varepsilon'}}\right) = \frac{\pi^2}{6} - \left(\frac{\pi\Delta(j_1, j_2)}{N} - \frac{\pi}{2}\right)^2 + O\left(\frac{1}{a_T^{1-\varepsilon'}}\right) = \frac{\Delta(j_1, j_2)^2}{N^2} = \frac{\Delta(j_1, j_2)(N - \Delta(j_1, j_2))}{N^2} + O\left(\frac{1}{a_T^{-\varepsilon'}}\right).
\]
In order to complete the proof of (3.104), we now consider the case (3.103) for $k_1$ and again assume that $j_2 \in A_{T,1}(v_2)$. In this scenario, for each fixed $k_1$ the restriction $|k_1 - k_2| \leq a_T$ cannot hold for any $k_2 \in \{-[\omega_2(N-1)/2], \ldots, [\omega_2N/2]\}$ [note that we assumed $\omega_1 \leq \omega_2$], and the sum over $k_2$ thus vanishes.

The property (3.104) now enables us to conclude that

$$V_{1,k_1\neq k_2}(j_1, j_2) = \frac{1}{N} \sum_{k_1=-(N-1)/2}^{N/2} \phi_{y_1,T}(j_1, \lambda_{k_1})\phi_{y_2,T}(j_2, \lambda_{k_1}) f^2(\lambda_{k_1}) \frac{(N - \Delta(j_1, j_2))\Delta(j_1, j_2)}{N^2} + O(\frac{1}{a_T^{1-\varepsilon'}} + \frac{a_T}{N^{1-\varepsilon'}}).$$

By combining the above results for $V_{1,k_1=k_2}(j_1, j_2)$ and $V_{1,k_1\neq k_2}(j_1, j_2)$ we finally get

$$V_{1,T}(j_1, j_2) = V_{1,k_1=k_2}(j_1, j_2) + V_{1,k_1\neq k_2}(j_1, j_2) = (1 - \frac{\Delta(j_1, j_2)}{N}) \frac{1}{N} \sum_{k=-(N-1)/2}^{N/2} \phi_{y_1,T}(j_1, \lambda_k)\phi_{y_2,T}(j_2, \lambda_k) f^2(\lambda_k) + O(\frac{1}{a_T^{1-\varepsilon'}} + \frac{a_T}{N^{1-\varepsilon'}})$$

for any $\varepsilon' > 0$.

For a proof of the claim (3.78), we consider the second term $V_{2,T}(j_1, j_2)$ resulting from the application of the product theorem for cumulants in (3.75). For this expression, the property (3.79) implies that the conditions

$$p_1 = p_2 + l - n - j_1 + j_2,$$
$$q_1 = q_2 + m - o - j_1 + j_2$$

and

$$0 \leq p_2 + l - n - j_1 + j_2 \leq N - 1,$$
$$0 \leq q_2 + m - o - j_1 + j_2 \leq N - 1$$

(3.106) (3.107)
have to hold, which implicates that

\[
V_{2,T}(j_1, j_2) = \frac{1}{(2\pi)^2 N^3} \sum_{k_1,k_2=-(N-1)/2}^{N/2} \phi_{y_1,T(j_1, \lambda_k)} \phi_{y_2,T(j_2, \lambda_k)} \sum_{l,m,n,o=0}^{\infty} \psi_l \psi_m \psi_n \psi_o \times \sum_{p_2,q_2=0}^{p_2 + q_2 = 0} \exp(-i \lambda_{k_1} (p_2 - q_2 + l - n - m + o)) \exp(-i \lambda_{k_2} (p_2 - q_2))
\]

\[
= \frac{1}{(2\pi)^2 N^3} \sum_{k_1,k_2=-(N-1)/2}^{N/2} \phi_{y_1,T(j_1, \lambda_k)} \phi_{y_2,T(j_2, \lambda_k)} \sum_{l,m,n,o=0}^{\infty} \psi_l \psi_m \psi_n \psi_o \times \sum_{p_2,q_2=0}^{p_2 + q_2 = 0} \exp(-i \lambda_{k_1} (l - n - m + o)) \sum_{p_2,q_2=0}^{p_2 + q_2 = 0} \exp(-i (\lambda_{k_1} + \lambda_{k_2}) (p_2 - q_2))
\]

\[
= : V_{2,k_1=-k_2}(j_1, j_2) + V_{2,k_1\neq-k_2}(j_1, j_2),
\]

(3.108)

where \( V_{2,k_1=-k_2}(j_1, j_2) \) and \( V_{2,k_1\neq-k_2}(j_1, j_2) \) denote the sums over all \((k_1, k_2)\) satisfying \( k_1 = -k_2 \) and \( k_1 \neq -k_2 \) respectively. Using the same arguments as were provided in the treatment of the term \( V_{1,k_1=k_2}(j_1, j_2) \), it can be seen that \( V_{2,k_1=-k_2}(j_1, j_2) \) is of order \( O(1/N^2) \), if \( \Delta(j_1, j_2) \geq N \), and in the case \( \Delta(j_1, j_2) < N \), is equal to

\[
\frac{1}{N} (1 - \frac{\Delta(j_1, j_2)}{N})^2 \sum_{k=-(N-1)/2}^{N/2} \phi_{y_1,T(j_1, \lambda_k)} \phi_{y_2,T(j_2, \lambda_k)} f^2(\lambda_k) + O(\frac{1}{N^{1-\varepsilon}})
\]

(3.109)

for any \( \varepsilon' > 0 \). Regarding the term \( V_{2,k_1\neq-k_2}(j_1, j_2) \), similar arguments as were provided in the treatment of the term \( V_{1,T}(j_1, j_2) \) can be performed to show that

\[
V_{2,k_1\neq-k_2}(j_1, j_2) = \frac{1}{N} \sum_{k_1=-(N-1)/2}^{N/2} \phi_{y_1,T(j_1, \lambda_k)} \phi_{y_2,T(j_2, \lambda_k)} f^2(\lambda_k) \frac{(N - \Delta(j_1, j_2)) \Delta(j_1, j_2)}{N^2} + O\left(\frac{1}{a_T^{1-\varepsilon'}} + \frac{a_T}{N^{1-\varepsilon'}}\right).
\]

(3.110)

Properties (3.109) and (3.110) imply the claim (3.78).

For the terms \( B_T(j_1, j_2) = V_{1,T}(j_1, j_2) + V_{2,T}(j_1, j_2) \) for \( (j_1, j_2) \in A_{T,2}(v_1, v_2) \), it subsequently follows

\[
B_T(j_1, j_2) = (1 - \frac{\Delta(j_1, j_2)}{N}) \frac{2}{N} \sum_{k=-(N-1)/2}^{N/2} \phi_{y_1,T(j_1, \lambda_k)} \phi_{y_2,T(j_2, \lambda_k)} f^2(\lambda_k)
\]

\[
+ O\left(\frac{1}{a_T^{1-\varepsilon'}} + \frac{a_T}{N^{1-\varepsilon'}}\right).
\]
uniformly with respect to \((j_1, j_2) \in A_{T,2}(v_1, v_2)\) satisfying \(\Delta(j_1, j_2) < N\) and \(\varepsilon' > 0\). For the case \(v_i \in (1/c, 1 - 1/c)\) for \(i = 1, 2\), the definition of the functions \(\phi_{y_1,T}\) and \(\phi_{y_2,T}\) and the set \(A_{T,2}(v_1, v_2)\) yields

\[
\text{Cov}(\sqrt{N} \hat{E}_T(v_1, \omega_1), \sqrt{N} \hat{E}_T(v_2, \omega_2)) = \sum_{(j_1, j_2) \in A_{T,2}(v_1, v_2)} B_T(j_1, j_2) + O\left( \frac{1}{a_T^{1/2}} + \frac{a_T}{N^{1-\varepsilon'}} \right)
\]

\[
= \begin{cases} 
0 & \text{if } \frac{2}{c} \leq v_2 - v_1 \\
-2(v_2 - v_1)[1 - (v_2 - v_1)] \int_{-\min(\omega_1, \omega_2)/\pi}^{\min(\omega_1, \omega_2)/\pi} f^2(\lambda) d\lambda & \text{if } \frac{1}{c} < v_2 - v_1 \leq \frac{2}{c} \\
2 - 3(v_2 - v_1)[1 - (v_2 - v_1)] \int_{-\min(\omega_1, \omega_2)/\pi}^{\min(\omega_1, \omega_2)/\pi} f^2(\lambda) d\lambda & \text{if } 0 \leq v_2 - v_1 \leq \frac{1}{c} \\
+ O\left( \frac{1}{a_T^{1/2}} + \frac{a_T}{N^{1-\varepsilon'}} \right), & \text{otherwise}
\end{cases}
\]

where \(\varepsilon' > 0\) is an arbitrary positive constant and \(a_T\) is any sequence, which satisfies the conditions (3.76). [The assertion for the cases \(v_i < 1/c\) and \(v_i > 1 - 1/c\) for at least one \(i \in \{1, 2\}\) follows by similar arguments.] Choosing the sequence \(a_T\) such that \(a_T = CN^{1/2}\) for some constant \(C\) we finally obtain

\[
\text{Cov}(\sqrt{N} \hat{E}_T(v_1, \omega_1), \sqrt{N} \hat{E}_T(v_2, \omega_2)) = \text{Cov}(B_T(v_1, \omega_1), B_T(v_2, \omega_2)) + O\left( \frac{1}{N^{1/2(1-\varepsilon')}} + \frac{1}{N^{1/2-\varepsilon'}} \right)
\]

for any \(\varepsilon' > 0\), which completes the proof of (2).

**Proof of (3):** For the proof of part (3), we use the notation

\[
Y_{i,1} := Z_{j_i - N/2 + 1 + p_i - m_i}, \quad Y_{i,2} := Z_{j_i - N/2 + 1 + q_i - n_i}
\]

for \(i = 1, \ldots, l\) and, for \(v_1, \ldots, v_l \in [0, 1]\), introduce the set

\[
A_{T,2}(v_1, \ldots, v_l) := \{[u(v_1, T)T] - N/2, [u(v_1, T)T] + N/2] \times \ldots \times [u(v_l, T)T] - N/2, [u(v_l, T)T] + N/2\}.
\]

By employing the definition (3.61) of the function \(\phi_{y,T}\), \(y \in [0, 1]^2\), the linear representation (2.9) of the time series \(\{X_{t,T}\}_{t=1,...,T}\) and the product theorem for cumulants [see Theorem...
2.3.6], we get

\[
\text{cum}(\sqrt{N} \hat{E}_T(y_1), ..., \sqrt{N} \hat{E}_T(y_l)) = \frac{1}{N^{3l/2}(2\pi)^l} \sum_{(j_1, ..., j_l) \in A_{T,l}(v_1, ..., v_l) \ k_1, ..., k_l = \lfloor (N-1)/2 \rfloor} \sum_{m_1=0}^{N/2} \sum_{n_1=0}^{N/2} \sum_{p_1=0}^{N/2} \sum_{q_1=0}^{N/2} \prod_{s=1}^{l} [\phi_{y_s,T}(j_s, \lambda_{k_s}) \psi_{m_s} \psi_{n_s} \exp(-\lambda_{k_s}(p_s - q_s))] \times \text{cum}(Y_{1,1}Y_{1,2}, Y_{2,1}Y_{2,2}, ..., Y_{l,1}Y_{l,2}),
\]

\[
= : \sum_{\nu} V(\nu).
\]

Here, the term \( V(\nu) \) is defined by

\[
V(\nu) := \frac{1}{N^{3l/2}(2\pi)^l} \sum_{(j_1, ..., j_l) \in A_{T,l}(v_1, ..., v_l) \ k_1, ..., k_l = \lfloor (N-1)/2 \rfloor} \sum_{m_1=0}^{N/2} \sum_{n_1=0}^{N/2} \sum_{p_1=0}^{N/2} \sum_{q_1=0}^{N/2} \prod_{s=1}^{l} \phi_{y_s,T}(j_s, \lambda_{k_s}) \psi_{m_s} \psi_{n_s} \exp(-\lambda_{k_s}(p_s - q_s)) \times \text{cum}(Y_{a,b}; (a, b) \in \nu_1) \cdots \text{cum}(Y_{a,b}; (a, b) \in \nu_l)
\]

and the summation is carried out over all indecomposable partitions \( \nu = (\nu_1, ..., \nu_k) \) of the table

\[
\begin{array}{ccc}
Y_{1,1} & Y_{1,2} \\
\vdots & \vdots \\
Y_{l,1} & Y_{l,2}
\end{array}
\]

(3.113)

Due to the Gaussianity of the random variables \( Y_{i,j} \), we only have to consider partitions \( \nu = (\nu_1, ..., \nu_l) \) with \( l \) elements [see Example 2.3.1]. As the number of indecomposable partitions of the table (3.113) is bounded by \( 2^l(2l)! \) [see Dahlhaus (1988)], it is furthermore sufficient to establish the inequality

\[
V(\nu) \leq C_l \frac{\log(N)^{l-1}}{N^{l/2-1}}
\]

for all indecomposable partitions \( \nu \) of table (3.113). Without loss of generality, we restrict ourselves to the indecomposable partition

\[
\bar{\nu} := \bigcup_{i=1}^{l-1} (Y_{i,1}, Y_{i+1,2}) \cup (Y_{l,1}, Y_{1,2}),
\]

(3.114)
which by simple calculations corresponds to the dominating term in the sum (3.112). The independence of the innovations $Z_t$, $t \in \mathbb{Z}$, implies that the equations

$$q_{i+1} = p_i - m_i + n_{i+1} + j_i - j_{i+1} \quad \text{for} \quad i \in \{1, \ldots, l - 1\},$$

$$q_1 = p_1 - m_1 + n_1 - j_1 + j_l,$$

have to hold for the respective summands not to vanish. Therefore, we obtain

$$V(\bar{\nu}) = \frac{1}{N^{2l}} \sum_{(j_1, \ldots, j_l) \in A_{T,I}(v_1, \ldots, v_l)} N/2 \sum_{k_1, \ldots, k_l = -\lfloor (N-1)/2 \rfloor}^{N/2} \prod_{s=1}^{l} \phi_{y_s,T}(j_s, \lambda_k) \sum_{m_1, \ldots, m_l, n_1, \ldots, n_l = 0}^{\infty} \sum_{p_1, \ldots, p_l = 0}^{N-1} \sum_{s=1}^{l} \prod_{s=1}^{l} \exp(-i\lambda_k (p_s - p_{s-1} + m_{s-1} - n_s - j_{s-1} + j_s)),$$

where summation over $p_1, \ldots, p_l$ is performed with respect to the restrictions

$$0 \leq p_i - m_i + n_{i+1} + j_i - j_{i+1} \leq N - 1 \quad \text{for} \quad i \in \{1, \ldots, l - 1\},$$

$$0 \leq p_l - m_l + n_1 - j_1 + j_l \leq N - 1.$$

Because of $p_i \in \{0, \ldots, N - 1\}$, the inequalities (3.117) and (3.118) can only hold, if

$$|n_{i+1} - m_i + j_i - j_{i+1}| \leq N \quad \text{for} \quad i \in \{1, \ldots, l - 1\},$$

$$|n_1 - m_l - j_1 + j_l| \leq N$$

for each summand in (3.116). If we combine (3.119) and (3.120) with the inequality

$$\left| \frac{1}{N} \sum_{k = -\lfloor (N-1)/2 \rfloor}^{N/2} \phi_{y,T}(j, \lambda_k) e^{-i\lambda_k r} \right| \leq \frac{C}{|r \mod N/2|},$$

which holds uniformly with respect to $y \in [0, 1]^2$ and for all $r \in \mathbb{N}$ with $r \mod N/2 \neq 0$ [see (A2) in Eichler (2008)], we obtain that (3.116) is bounded by

$$\frac{C l}{N^{2l}} \sum_{(j_1, \ldots, j_l) \in A_{T,I}(v_1, \ldots, v_l)} \sum_{m_1, \ldots, m_l = 0}^{\infty} \prod_{s=1}^{l} \psi_{m_s} \psi_{n_s} \sum_{p_1, \ldots, p_l = 0}^{N-1} \sum_{s=1}^{l} \prod_{s=1}^{l} \frac{1}{|p_s - p_{s-1} + m_{s-1} - n_s - j_{s-1} + j_s|} \prod_{s=1}^{l} 1(p_s \notin \{z_{s1}, z_{s2}\}),$$

(3.121)
where we identified the indices 0 with $l$ and $l + 1$ with 1, defined
\[
z_{s1} := p_{s-1} - m_{s-1} + n_s + j_{s-1} - j_s,
\]
\[
z_{s2} := p_{s+1} + m_s - n_{s+1} - j_s + j_{s+1}
\]
and used that the cases with $p_s = z_{s1}$, $p_s = z_{s2}$ or $|p_s - p_{s-1} + m_{s-1} - n_s - j_{s-1} + j_s| \geq N/2$
for some $s \in \{1, \ldots, l\}$ are of the same or smaller order than (3.121). For the following arguments, we define the set
\[
A_i := [0, N - 1] \setminus \{[z_{i1} - 1, z_{i1} + 1] \cup [z_{i2} - 1, z_{i2} + 1]\}
\]
for $i = 1, \ldots, l$. This notation implies that the expression (3.121) is bounded by
\[
\frac{C^l}{N^{l/2}} \sum_{(j_1, \ldots, j_l) \in A_{l,i}(v_1, \ldots, v_t)} \prod_{s=1}^{\infty} \prod_{m_{s-1} - n_s - j_{s-1} + j_s \leq N}^{m_s - n_{s+1} - j_s + j_{s+1} \leq N} \prod_{s=1}^{l} \psi_{m_i, \psi_{n_s}}
\]
\[
\times \prod_{s=1}^{N-1} \prod_{p_1 = 0}^{N} \prod_{1 \leq i \leq l} \frac{1}{|p_s - p_{s-1} + m_{s-1} - n_s - j_{s-1} + j_s|} d(p_1, \ldots, p_l). \quad (3.122)
\]
For a treatment of the above term, we first consider the integral over the set $A_1 \times \ldots \times A_{l-1}$
for, which we obtain the upper bound
\[
\int_{A_1 \times \ldots \times A_{l-1}} \frac{1}{|p_1 - p_i + m_i - n_1 - j_1 + j_i|} \prod_{s=2}^{l} \frac{1}{|p_s - p_{s-1} + m_{s-1} - n_s - j_{s-1} + j_s|} d(p_1, \ldots, p_l)
\]
\[
\leq \int_{A_2 \times \ldots \times A_{l-1}} \left[ \log \left( \frac{1}{|p_1 - p_i + m_i - n_1 - j_1 + j_i|} \right) + \log \left( \frac{1}{|p_2 - t + m_2 - n_2 - j_2 + j_2|} \right) \right] d(p_2, \ldots, p_l)
\]
\[
\times \prod_{s=3}^{l} \frac{1}{|p_s - p_{s-1} + m_{s-1} - n_s - j_{s-1} + j_s|} d(p_2, \ldots, p_l), \quad (3.123)
\]
where the expression $[\ldots]_{\partial A_i}$ means that we evaluate the function inside the brackets at
all points of the boundary of $A_i$ and subsequently compute the sum of these values. The restrictions on the $p_i$, $m_i$ and $n_i$ in the sum (3.121) and the definition of the set $A_1$ imply
that the log-terms in this expression have arguments that are at most of size $2N < 2lN$.
Therefore, (3.123) is not larger than
\[
\int_{A_2 \times \ldots \times A_{l-1}} \frac{\log(2N)}{|p_1 + m_1 - n_1 - j_1 + j_1|} \prod_{s=3}^{l} \frac{1}{|p_s - p_{s-1} + m_{s-1} - n_s - j_{s-1} + j_s|} d(p_1, \ldots, p_l).
\]
By using the same argument in the integration over \( p_2, ..., p_{l-1} \), it becomes apparent that the \( p_i \) and \( j_i \) terms in the denominator vanish in a telescoping sum and we thus obtain that (3.123) is bounded by

\[
\log(2lN)^{l-1} \left| \frac{m_1 - n_1 + m_2 - n_2 + \ldots + m_l - n_l}{m_1 - n_1 + m_2 - n_2 + \ldots + m_l - n_l} \right|.
\]

Together with the summability condition (3.10), this implies that the quantity (3.121) is not larger than

\[
C^l \frac{\log(N)^{l-1}}{N^{l/2}},
\]

where we additionally used that the sum over all terms with \( m_1 - n_1 + \ldots + m_l - n_l = 0 \) is of smaller order and in the last step employed the fact that \( \text{card}(A_{T,l}(v_1, \ldots, v_l)) \leq 2^l \). \( \square \)

**Proof of Lemma 3.4.2**

For a proof of the inequality (3.57), we consider the representation

\[
\hat{E}_T(\phi_{y,T}) := \hat{E}_T(y) = \frac{1}{N} \sum_{j=1}^{T} \sum_{k=-\lfloor (N-1)/2 \rfloor}^{\lfloor N/2 \rfloor} \phi_{y,T}(j, \lambda_k) I_N(\frac{j}{T}, \lambda_k),
\]

with the function \( \phi_{y,T} \) for \( y \in [0,1]^2 \) defined in (3.61). This identification implies

\[
\{ \sqrt{N} \hat{E}_T(y) \}_{y \in [0,1]^2} = \{ \sqrt{N} \hat{E}_T(\phi_{y,T}) \}_{\phi_{y,T} \in \mathcal{F}_T},
\]

where the symbol \( \mathcal{F}_T \) denotes the class of functions

\[
\mathcal{F}_T := \{ \phi_{y,T} | y \in \mathcal{P}_T \}
\]

with the set \( \mathcal{P}_T \) being defined by

\[
\mathcal{P}_T := \left\{ 0, \frac{1}{T}, \frac{2}{T}, \ldots, 1 - \frac{1}{T}, 1 \right\} \times \left\{ 0, \frac{1}{N}, \ldots, 1 - \frac{1}{N}, 1 \right\}.
\]
This representation implies the equality
\[
P\left( \sup_{y_1, y_2 \in [0,1]^2, d_\beta(y_1, y_2) < \delta} \sqrt{N} |\hat{E}_T(y_1) - \hat{E}_T(y_2)| > \eta \right) = P\left( \sup_{y_1, y_2 \in \mathcal{P}_T, d_\beta(y_1, y_2) < \delta} \sqrt{N} |\hat{E}_T(y_1) - \hat{E}_T(y_2)| > \eta \right)
\]
and for a proof of Lemma 3.4.2, we thus have to show that
\[
\forall \eta, \varepsilon > 0 : \exists \delta > 0 \text{ such that } P\left( \sup_{y_1, y_2 \in \mathcal{P}_T, d_\beta(y_1, y_2) < \delta} \sqrt{N} |\hat{E}_T(y_1) - \hat{E}_T(y_2)| > \eta \right) \leq \varepsilon \quad (3.125)
\]
for \( T \) sufficiently large. In order to achieve this goal, we need the following technical results, whose proof will be postponed.

**Lemma 3.4.3**
For every \( \beta \in (0, 1/3) \), there exists a constant \( C > 0 \) such that for sufficiently large \( T \in \mathbb{N} \) the inequality
\[
P(\sqrt{N} |\hat{E}_T(y_1) - \hat{E}_T(y_2)| > \eta d_\beta(y_1, y_2)) \leq 96 \exp\left(-\sqrt{\eta/4C}\right) \quad (3.126)
\]
holds uniformly in \( y_1, y_2 \in \mathcal{P}_T \) and for all \( \eta > 0 \), where \( d_\beta \) denotes the semi-metric
\[
d_\beta((v_1, \omega_1), (v_2, \omega_2)) = (|v_1 - v_2| + |\omega_1 - \omega_2|)^{\beta/2},
\]
introduced in Lemma 3.4.2.

**Lemma 3.4.4**
For the covering integral
\[
J(\kappa, d_\beta, \mathcal{P}_T) = \int_0^\kappa \left[ \log \left( \frac{48N(u, d_\beta, \mathcal{P}_T)}{u} \right) \right]^2 du
\]
of the set \( \mathcal{P}_T \) with respect to the semi-metric
\[
d_\beta((v_1, \omega_1), (v_2, \omega_2)) = (|v_1 - v_2| + |\omega_1 - \omega_2|)^{\beta/2},
\]
it holds
\[
\lim_{\kappa \to 0} \lim_{T \to \infty} J(\kappa, d_\beta, \mathcal{P}_T) = 0. \quad (3.127)
\]
Before we present the proofs of Lemma 3.4.3 and Lemma 3.4.4, we show how the property (3.125) is implied by these results: For this purpose, let \( \varepsilon > 0 \) and \( \eta > 0 \) be arbitrarily small and fixed. Lemma 3.4.3 allows us to apply the Chaining Lemma [see Theorem 2.3.15], which yields the probability bound
\[
P\left( \exists y_1, y_2 \in \mathcal{P}_T \text{ with } d_\beta(y_1, y_2) \leq \delta \text{ and } \sqrt{N} |\hat{E}_T(y_1) - \hat{E}_T(y_2)| > 26CJ_T(d_\beta(y_1, y_2)) \right) \leq 2\delta. \quad (3.128)
\]
This inequality gives the following upper bound for the probability in (3.125) [see Dahlhaus (1988) for a similar argument]

\[
P\left( \sup_{y_1, y_2 \in \mathcal{P}_T} \sqrt{N} |\hat{E}_T(y_1) - \hat{E}_T(y_2)| > \eta \right)
= P\left( \sup_{y_1, y_2 \in \mathcal{P}_T} \sqrt{N} |\hat{E}_T(y_1) - \hat{E}_T(y_2)| > \eta, \eta \geq 26CJ(\delta, d_\beta, \mathcal{P}_T) \right) + P\left( \eta < 26CJ(\delta, d_\beta, \mathcal{P}_T) \right)
\leq 2\delta + P\left( \eta < 26CJ(\delta, d_\beta, \mathcal{P}_T) \right),
\]

where in the last step the inequality (3.128) was used. Note that the event \{\eta < 26CJ(\delta, d_\beta, \mathcal{P}_T)\} is deterministic and that Lemma 3.4.4 implies that for \delta sufficiently small and \(T\) large enough the opposite inequality \(\eta > 26CJ(\delta, d_\beta, \mathcal{P}_T)\) is valid. This shows that

\[
P\left( \sup_{y_1, y_2 \in \mathcal{P}_T} \sqrt{N} |\hat{E}_T(y_1) - \hat{E}_T(y_2)| > \eta \right) \leq 2\delta
\]

for \delta small enough and \(T\) sufficiently large. Thus, the statement (3.125) follows by choosing \delta sufficiently small.

For the proof of Lemma 3.4.2 it remains to establish Lemma 3.4.3 and Lemma 3.4.4.

**Proof of Lemma 3.4.3**

In order to establish Lemma 3.4.3, we use the fact that there exists a constant \(C > 0\) such that, for all \(y_1, y_2 \in [0, 1]^2\) and for all even integers \(k \in \mathbb{N}\), we have the bound

\[
\mathbb{E}\left( N^{k/2} (\hat{E}_T(y_1) - \hat{E}_T(y_2))^k \right) \leq (2k)!C^k d_\beta(y_1, y_2)^k.
\]

Before we demonstrate how the moment inequality (3.130) can be established, we show how the bound (3.126) follows from this property. For this purpose, we define \(\varepsilon := \eta d_\beta(y_1, y_2)\) and consider the function \(h(x) := \exp(tx^{1/2})\) for some fixed value \(t > 0\). From the Markov
inequality, it follows that

\[
P(\sqrt{N}|\hat{E}_T(y_1) - \hat{E}_T(y_2)| > \varepsilon) \leq \frac{\mathbb{E}\left[\exp\left(t(\sqrt{N}|\hat{E}_T(y_1) - \hat{E}_T(y_2)|)^{1/2}\right)\right]}{\exp(t\varepsilon^{1/2})} = \exp(-t\varepsilon^{1/2}) \sum_{m=0}^{\infty} \frac{t^m}{m!} \mathbb{E}[\left(\sqrt{N}|\hat{E}_T(y_1) - \hat{E}_T(y_2)|\right)^m].
\]  

(3.131)

Furthermore, for each \(j \in \{1, 2, 3\}\) it holds by Jensens inequality

\[
\mathbb{E}\left[(\sqrt{N}|\hat{E}_T(y_1) - \hat{E}_T(y_2)|)^{4l+j}\right] \leq \left(\mathbb{E}[\left(\sqrt{N}|\hat{E}_T(y_1) - \hat{E}_T(y_2)|\right)^{2l+2}]\right)^{\frac{4l+j}{2l+4}}.
\]

(3.132)

By an application of the moment inequality (3.130), it can easily be seen that (3.132) is bounded by

\[
(4l + 4)!Cd_\beta(y_1,y_2)^{4l+j}
\]

uniformly in \(y_1, y_2\) and for all \(l \geq 0\) and \(j \in \{1, 2, 3\}\). Thus, we obtain that (3.131) is not larger than

\[
\exp(-t\varepsilon^{1/2}) \sum_{m=0}^{\infty} t^m [Cd_\beta(y_1,y_2)]^{\frac{m}{2}} (m+1)(m+2)(m+3),
\]

which for the choice

\[
t = \frac{1}{\sqrt{4Cd_\beta(y_1,y_2)}}
\]

and by using straightforward calculations can be shown to yield the upper bound (3.126). In order to complete the proof of Lemma 3.4.3, it remains to establish the moment inequality (3.130).

**Proof of (3.130):** For a proof of the inequality (3.130), we show that the bound

\[
|\text{cum}_l[\sqrt{N}(\hat{E}_T(y_1) - \hat{E}_T(y_2))]| \leq (2l)!Cd_\beta(y_1,y_2)^l
\]

(3.134)

holds uniformly in \(y_1, y_2 \in [0, 1]^2\) and for all integers \(l \in \mathbb{N}\). Then, the statement (3.130) follows by an application of the product theorem for cumulants [see Theorem 2.3.6] along
the same line of arguments as provided in Dahlhaus (1988):

\[ \mathbb{E}\left(N^{k/2} | \hat{E}_T(y_1) - \hat{E}_T(y_2)|^k \right) = \left| \sum_{\{P_1, \ldots, P_m\} \text{ is partition of } \{1, \ldots, k\}} \left( \prod_{j=1}^{m} \text{cum}_1|P_j| (\sqrt{N}(\hat{E}_T(y_1) - \hat{E}_T(y_2))) \right) \right| \]

\[ \leq \left| \sum_{\{P_1, \ldots, P_m\} \text{ is partition of } \{1, \ldots, k\}} \left( \prod_{j=1}^{m} (2|P_j|)! C^{2|P_j|} \beta(y_1, y_2)^{|P_j|} \right) \right| \leq (2k)! 2^k C^k \beta(y_1, y_2)^k. \quad (3.135) \]

It remains to illustrate how the bound (3.134) for the cumulants of the increments of the process \( \{\sqrt{N}\hat{E}_T(y)\}_{y \in [0,1]^2} \) ensues.

**Proof of (3.134):** Without loss of generality, we assume that \( y_1 = (v_1, \omega_1), y_2 = (v_2, \omega_2) \in \mathcal{P}_T \) are arbitrary but fixed such that \( v_1 \neq v_2 \) and \( \omega_1 \leq \omega_2 \) and consider the cases \( l = 1, l = 2 \) and \( l \geq 3 \) separately.

For the case \( l = 1 \), we obtain by the same arguments as were exploited in the proof of Lemma 3.4.1 part (1) that

\[ \text{cum}_1(\sqrt{N}(\hat{E}_T(y_1) - \hat{E}_T(y_2))) = \sqrt{N} \mathbb{E}(\hat{E}_T(y_1) - \hat{E}_T(y_2)) = O\left(\frac{1}{\sqrt{N}}\right) \]

\[ = O\left(\frac{1}{\sqrt{T}}\right) \leq C(|v_1 - v_2| + |\omega_1 - \omega_2|)^{1/2} \leq C(|v_1 - v_2| + |\omega_1 - \omega_2|)^{\beta/2} \leq 2! 2^1 C^1 \beta(y_1, y_2), \]

where we used that for \( y_1, y_2 \in \mathcal{P}_T, y_1 \neq y_2 \) it holds

\[ \frac{1}{T} \leq C(|v_1 - v_2| + |\omega_1 - \omega_2|) \]

and that there exists a constant \( C \) such that \( x^{1/2} \leq C x^{\beta/2} \) for all \( x \in [0,2] \).

For the case \( l = 2 \), we make use of results and arguments provided in the proof of Lemma 3.4.1 part (2). These yield

\[ \text{cum}_2(\sqrt{N}(\hat{E}_T(y_1) - \hat{E}_T(y_2))) \]

\[ = N \left( \text{Var}(\hat{E}_T(y_1)) + \text{Var}(\hat{E}_T(y_2)) - 2\text{Cov}(\hat{E}_T(y_1), \hat{E}_T(y_2)) \right) \]

\[ = \text{Var}(B(y_1)) + \text{Var}(B(y_2)) - 2\text{Cov}(B(y_1), B(y_2)) + O\left(\frac{1}{N^{\beta}}\right) \]

\[ (3.137) \]

for any \( \beta \in (0, 1/3) \). For the further treatment of the quantity (3.137), we provide a case-by-case analysis and investigate the cases \( 0 \leq |v_1 - v_2| \leq 1/c, 1/c < |v_1 - v_2| \leq 2/c \) and \( 2/c < |v_1 - v_2| \) separately [note that \( c \) denotes the constant satisfying \( T/N \to c \)]. For all cases, we apply the covariance structure (3.12) of the limiting process \( \{B(v, \omega)\}_{(v,\omega) \in [0,1]^2} \) in our calculations. For \( v_1, v_2 \in [0,1] \) satisfying \( 0 \leq |v_1 - v_2| \leq 1/c \), we obtain that (3.137) is
bounded by [note that we assumed \( \omega_1 \leq \omega_2 \)]

\[
\frac{1}{\pi} \left[ 2 \int_{-\omega_1}^{\omega_1} f^2(\lambda) d\lambda + 2 \int_{-\omega_2}^{\omega_2} f^2(\lambda) d\lambda - 2(2 - 3|v_1 - v_2|c) \int_{\min(\omega_1, \omega_2)}^{\min(\omega_1, \omega_2)} f^2(\lambda) d\lambda \right] + O\left( \frac{1}{N^{\beta}} \right)
\]

\[
= \frac{1}{\pi} \left[ 2 \int_{-\omega_1}^{\omega_1} f^2(\lambda) d\lambda + 2 \int_{-\omega_2}^{\omega_2} f^2(\lambda) d\lambda + 6|v_1 - v_2|c \int_{\min(\omega_1, \omega_2)}^{\min(\omega_1, \omega_2)} f^2(\lambda) d\lambda \right] + O\left( \frac{1}{N^{\beta}} \right)
\]

\[
\leq C(|v_1 - v_2| + |\omega_1 - \omega_2|) + (|v_1 - v_2| + |\omega_1 - \omega_2|)^{\beta}
\]

\[
\leq C(|v_1 - v_2| + |\omega_1 - \omega_2|)^{\beta} \leq 4!^2 C^2 d_\beta(y_1, y_2)^2,
\]

where in the second step (3.136) was used. For \( 1/c < |v_1 - v_2| \leq 2/c \), the term (3.137) is not larger than

\[
\frac{1}{\pi} \left[ 2 \int_{-\omega_1}^{\omega_1} f^2(\lambda) d\lambda + 2 \int_{-\omega_2}^{\omega_2} f^2(\lambda) d\lambda - 2(2 - |v_1 - v_2|c) \int_{\min(\omega_1, \omega_2)}^{\min(\omega_1, \omega_2)} f^2(\lambda) d\lambda \right] + O\left( \frac{1}{N^{\beta}} \right)
\]

\[
= \frac{1}{\pi} \left[ 2 \int_{-\omega_1}^{\omega_1} f^2(\lambda) d\lambda + 2 \int_{-\omega_2}^{\omega_2} f^2(\lambda) d\lambda + 2|v_1 - v_2|c \int_{\min(\omega_1, \omega_2)}^{\min(\omega_1, \omega_2)} f^2(\lambda) d\lambda \right] + O\left( \frac{1}{N^{\beta}} \right)
\]

\[
\leq C(|v_1 - v_2| + |\omega_1 - \omega_2|) + (|v_1 - v_2| + |\omega_1 - \omega_2|)^{\beta} \leq 4!^2 C^2 d_\beta(y_1, y_2)^2.
\]

In the case \( 2/c < |v_1 - v_2| \), the covariance between the random variables \( B(y_1) \) and \( B(y_2) \) vanishes and we thus obtain that (3.137) does not exceed the threshold

\[
\frac{1}{\pi} \left[ 2 \int_{-\omega_1}^{\omega_1} f^2(\lambda) d\lambda + 2 \int_{-\omega_2}^{\omega_2} f^2(\lambda) d\lambda \right] \leq C \leq C(|v_1 - v_2| + |\omega_1 - \omega_2|) \leq 4!^2 C^2 d_\beta(y_1, y_2)^2,
\]

which completes the proof of (3.134) in the case \( l = 2 \).

Next, we consider some integer \( l \geq 3 \) and define the function \( \phi := \phi_{y_1, T} - \phi_{y_2, T} \), where \( \phi_{y, T} \) for \( y \in [0, 1]^2 \) was introduced in (3.61). This notation enables the convenient identification

\[
\hat{E}_T(y_1) - \hat{E}_T(y_2) = \frac{1}{N} \sum_{j=1}^{T} \sum_{k=-\lfloor (N-1)/2 \rfloor}^{N/2} \phi(j, \lambda_k) I_N\left( \frac{j}{T}, \lambda_k \right)
\]

and allows to perform the same calculations as were provided in the treatment of the cumulants of higher orders in the proof of Lemma 3.4.1 [see the proof of part (3)], which yield

\[
\text{cum}_l\left( \sqrt{N}(\hat{E}_T(y_1) - \hat{E}_T(y_2)) \right) \leq (2l)!^2 C^l O\left( \frac{\log(N)^{l-1}}{N^{l/2-1}} \right) = (2l)!^2 C^l O\left( \frac{1}{N^{l(1/2-1/3-\varepsilon')}} \right)
\]

for any \( \varepsilon' > 0 \) and \( l \geq 3 \). By choosing \( \varepsilon' = 1/6 - \beta/2 \) and applying the inequality (3.136),
we finally obtain
\[
\text{cum}_t \left( \sqrt{N} (\hat{E}_T(y_1) - \hat{E}_T(y_2)) \right) \leq (2l)!2^l C^l O(\frac{1}{\sqrt{N}^{\beta/2}}) \leq (2l)!2^l C (|v_1 - v_2| + |w_1 - w_2|)^{\beta/2}
\]
\[= (2l)!2^l C^l d_{\beta}(y_1, y_2)^l,
\]
which completes the proof of (3.134).

Proof of Lemma 3.4.4

The definition of the semi-metric \(d_{\beta}\) implies the bound
\[
N(u, d_{\beta}, P_T) \leq \frac{C}{u^{4/\beta}}
\]
for the covering numbers \(N(u, d_{\beta}, P_T)\), which yields
\[
J(\kappa, d_{\beta}, P_T) = \int_0^{\kappa} \left[ \log \left( \frac{48N(u, d_{\beta}, P_T)^2}{u^2} \right) \right]^2 du \leq C \int_0^{\kappa} \left[ \log(48u^{8/\beta - 1}) \right]^2 du
\]
\[= C \left[ \int_0^{\kappa} \log(48)^2 du - 2 \int_0^{\kappa} \log(48) \log(u^{8/\beta + 1}) du + \int_0^{\kappa} (\log(u^{8/\beta + 1}))^2 du \right] \to 0
\]
as \(\kappa \to 0\).

Proof of Theorem 3.1.1 b): Under the alternative \(K \geq 1\), there exists for each break point \(b_r, r \in \{1, ..., K\}\) some value \(\omega_r\) such that the deterministic quantity \(|E_{N,T}(b_r, \omega_r)|\), which was defined in (3.63), is positive. This together with the property (3.64) and the fact
\[
N^{1/2} |\hat{E}_T(y) - E_{N,T}(y)| = O_P(1),
\]
which can be shown by similar arguments as were provided in the proof of Lemma 3.4.1, yields the assertion.

3.4.2 Proof of Theorem 3.1.2

As in the proof of Theorem 3.1.1, we consider the case \(d = 1\) and emphasise that \(C\) denotes a constant, which does not depend on the sample size and can vary from line to line in the calculations.

Proof of Theorem 3.1.2 a): Theorem 2.3.13 implies that the assertion is a consequence of the following results:
Lemma 3.4.5 (Point wise convergence to zero in probability)
For every \( y = (v, \omega) \in [0, 1]^2 \) and \( \gamma \in (0, 1/2) \), it holds
\[
N^\gamma \hat{E}_T(y) = o_P(1) .
\] (3.139)

Lemma 3.4.6 (Asymptotic stochastic equicontinuity)
For every \( \gamma \in (0, 1/2) \) and \( \eta, \varepsilon > 0 \), there exists some \( \delta > 0 \) such that
\[
\lim_{T \to \infty} P \left( \sup_{(y_1, y_2) \in [0, 1]^2: d_2(y_1, y_2) < \delta} N^\gamma |\hat{E}_T(y_1) - \hat{E}_T(y_2)| > \eta \right) < \varepsilon ,
\] (3.140)
where \( d_2(y_1, y_2) \) denotes the euclidean distance between \( y_1 = (v_1, \omega_1) \) and \( y_2 = (v_2, \omega_2) \).

Proof of Lemma 3.4.5

The claim (3.139) follows by similar arguments as given in the proof of Theorem 4.3.2 a), where we showed that \( N^{1/2} \hat{E}_T(y) \) is asymptotically normally distributed.

Proof of Lemma 3.4.6

For a proof of Lemma 3.4.6, we proceed by a similar course of argument as in the proof of Lemma 3.4.2 and employ the representation
\[
\hat{E}_T(y) = \frac{1}{N} \sum_{j=1}^{T} \sum_{k=-\lfloor (N-1)/2 \rfloor}^{N/2} \phi_{y,T}(j, \lambda_k) I_N(j, \lambda_k),
\]
which was introduced in (3.62), where the function \( \phi_{y,T} \) for \( y = (v, \omega) \in [0, 1]^2 \) was defined in (3.61). Moreover, we define by
\[
\rho_{2,T,\gamma}(y_1, y_2) := \left( \frac{1}{N^{2(1-\gamma)}} \sum_{j=1}^{T} \sum_{k=-\lfloor (N-1)/2 \rfloor}^{N/2} (\phi_{y_1,T}(j, \lambda_k) - \phi_{y_2,T}(j, \lambda_k))^2 \right)^{1/2} (3.141)
\]
a semi-metric on the set
\[
\mathcal{P}_T = \left\{ \frac{N}{T}, \frac{N+1}{T}, ..., 1 - \frac{N}{T} \right\} \times \left\{ 0, \frac{1}{N}, ..., 1 - \frac{1}{N}, 1 \right\} .
\] (3.142)

Simple calculations show that there exists a constant \( C > 0 \) such that the inequality
\[
\rho_{2,T,\gamma}(y_1, y_2) \leq d_{T,\gamma}(y_1, y_2)
\]
\[
: = C/2 \frac{N^\gamma}{N^{2-\gamma}} \left( \sqrt{|\omega_2 - \omega_1|} + \sqrt{|v_2 - v_1|} \right) I_{\{v_1 \neq v_2\}} + C \frac{N^{1-\gamma}}{N^{2-\gamma}} I_{\{v_1 = v_2\}}
\] (3.143)
holds. Note that \( d_{T,\gamma} \) defines a further semi-metric on \( \mathcal{P}_T \) and that, due to the fact that \( d_{T,\gamma}(y_1, y_2) \to 0 \) as \( T \to \infty \) for all \( y_1, y_2 \in [0,1]^2 \), we have

\[
\{(y_1, y_2)|d_2(y_1, y_2) \leq \delta\} \subset \{(y_1, y_2)|d_{T,\gamma}(y_1, y_2) \leq \delta\}
\]

for \( T \) sufficiently large and any \( \delta > 0 \). This property implies that

\[
P\left( \sup_{y_i \in [0,1]^2; \atop d_2(y_i,y_j) < \delta} N^\gamma|\hat{E}_T(y_1) - \hat{E}_T(y_2)| > \eta \right) \leq P\left( \sup_{y_i \in \mathcal{P}_T; \atop d_{T,\gamma}(y_1,y_2) < \delta} N^\gamma|\hat{E}_T(y_1) - \hat{E}_T(y_2)| > \eta \right) \quad (3.144)
\]

for every \( \gamma \in (0,1/2), \delta > 0 \) and sufficiently large \( T \in \mathbb{N} \). The bound (3.144) shows that for a proof of (3.140) it is sufficient to demonstrate that the probability on the right-hand side of (3.144) becomes arbitrarily small for \( T \) sufficiently large. The arguments of the proof of Lemma 3.4.2 reveal that we achieve this goal by establishing the following results:

**Lemma 3.4.7**

There exists a constant \( C > 0 \) such that the inequality

\[
|\text{cum}_l(N^\gamma(\hat{E}_T(y_1) - \hat{E}_T(y_2)))| \leq (2l)!C^ld_{T,\gamma}(y_1, y_2)^l \quad (3.145)
\]

holds for all \( y_1, y_2 \in [0,1]^2 \) and integers \( l \in \mathbb{N} \).

**Lemma 3.4.8**

For the covering integral

\[
J(\kappa, d_{T,\gamma}, \mathcal{P}_T) = \int_0^\kappa \left[ \log \left( \frac{48N(u,d_{T,\gamma},\mathcal{P}_T)}{u} \right) \right]^2 du
\]

of the set \( \mathcal{P}_T \) with respect to the semi-metric

\[
d_{T,\gamma}(y_1, y_2) := \frac{C/2}{N^{\frac{1}{2}-\gamma}}(\sqrt{\omega_2 - \omega_1} + \sqrt{\nu_2 - \nu_1})1_{\{v_1T=v_2T\}} + \frac{C}{N^{\frac{1}{2}-\gamma}}1_{\{v_1T\neq v_2T\}},
\]

which was introduced in (3.143), it holds

\[
\lim_{\kappa \to 0} \lim_{T \to \infty} J(\kappa, d_{T,\gamma}, \mathcal{P}_T) = 0. \quad (3.146)
\]

**Proof of Lemma 3.4.7**

Without loss of generality, we assume that \( l \) is even and, for \( y_1, y_2 \in \mathcal{P}_T \), define the function \( \phi := \phi_{y_1,T} - \phi_{y_2,T} \), where \( \phi_{y,T} \) for \( y \in \mathcal{P}_T \) was introduced in (3.61). This notation implies
that for \( y_1, y_2 \in \mathcal{P}_T \) the increment \( \hat{E}_T(y_1) - \hat{E}_T(y_2) \) of the process \( \{\hat{E}_T(y)\}_{y \in [0,1]^2} \) can be described by the expression

\[
\hat{E}_T(y_1) - \hat{E}_T(y_2) = \frac{1}{N} \sum_{j=1}^{T} \sum_{k=-\lfloor (N-1)/2 \rfloor}^{N/2} \phi(j, \lambda_k) I_N(\frac{j}{T}, \lambda_k).
\]

This representation allows for the application of the product theorem for cummulants [Theorem 2.3.6] in the same way as was explained in part (3) of the proof of Lemma 3.4.1 to obtain

\[
\text{cum}_1\left(N^\gamma(\hat{E}_T(y_1) - \hat{E}_T(y_2))\right) = \sum_\nu V_\gamma(\nu),
\]

where the term \( V_\gamma(\nu) \) is defined by

\[
V_\gamma(\nu) := \frac{1}{N^{l(2-2\gamma)}(2\pi)^l} \sum_{j_1,..,j_l=1}^{T} \sum_{k_1,..,k_l=-\lfloor (N-1)/2 \rfloor}^{N/2} \sum_{m_1=0}^{\infty} \sum_{n_1=0}^{\infty} \sum_{p_1=0}^{N-1} \prod_{s=1}^{l} \phi(j_s, \lambda_{k_s})
\]

\[
\times \prod_{s=1}^{l} \left[ \psi_{m_s} \psi_{n_s} \exp(-\lambda_{k_s}(p_s - q_s)) \right] \text{cum}(Y_{a,b}; (a,b) \in \nu_1) \cdots \text{cum}(Y_{a,b}; (a,b) \in \nu_l),
\]

and the summation is carried out over all indecomposable partitions \( \nu = (\nu_1, ..., \nu_k) \) of the table (3.113). As the number of indecomposable partitions of the table (3.113) is bounded by \( 2^l(2l)! \) [see Dahlhaus (1988)], in order to establish the claim (3.145), it is sufficient to show that the inequality

\[
V_\gamma(\nu) \leq C^l d_{\gamma}(y_1, y_2)^l
\]

holds uniformly for all indecomposable partitions \( \nu \). Without loss of generality, we restrict ourselves to the indecomposable partition \( \bar{\nu} \) defined in (3.114) [Terms corresponding to other partitions can be treated in a similar manner.]. For this partition, the equations (3.115), which follow from the independence of the innovation process \( \{Z_t\}_{t \in \mathbb{Z}} \), and the definition of the function \( \phi \) imply

\[
V(\bar{\nu}) = \frac{1}{N^{l(2-2\gamma)}(2\pi)^l} \sum_{(j_1,..,j_l) \in A_{T,\gamma}(v_1,v_2)^l} \sum_{k_1,..,k_l=-\lfloor (N-1)/2 \rfloor}^{N/2} \sum_{m_1=0}^{\infty} \sum_{n_1=0}^{\infty} \sum_{p_1=0}^{N-1} \prod_{s=1}^{l} \phi(j_s, \lambda_{k_s})
\]

\[
\times \prod_{s=1}^{l} \left[ \psi_{m_s} \psi_{n_s} \exp(-i\lambda_{k_1}(p_1 - p_l + m_l - n_1 + j_1 - j_l)) \right]
\]

\[
\times \prod_{s=2}^{l} \exp(-\lambda_{k_s}(p_s - p_{s-1} + m_{s-1} - n_s - j_{s-1} + j_s)),
\]

The theorem given in [Theorem 3.4.6] implies that

\[
\text{cum}_1\left(N^\gamma(\hat{E}_T(y_1) - \hat{E}_T(y_2))\right) = \sum_\nu V_\gamma(\nu),
\]

where the term \( V_\gamma(\nu) \) is defined by

\[
V_\gamma(\nu) := \frac{1}{N^{l(2-2\gamma)}(2\pi)^l} \sum_{(j_1,..,j_l) \in A_{T,\gamma}(v_1,v_2)^l} \sum_{k_1,..,k_l=-\lfloor (N-1)/2 \rfloor}^{N/2} \sum_{m_1=0}^{\infty} \sum_{n_1=0}^{\infty} \sum_{p_1=0}^{N-1} \prod_{s=1}^{l} \phi(j_s, \lambda_{k_s})
\]

\[
\times \prod_{s=1}^{l} \left[ \psi_{m_s} \psi_{n_s} \exp(-i\lambda_{k_1}(p_1 - p_l + m_l - n_1 + j_1 - j_l)) \right]
\]

\[
\times \prod_{s=2}^{l} \exp(-\lambda_{k_s}(p_s - p_{s-1} + m_{s-1} - n_s - j_{s-1} + j_s)),
\]

The theorem given in [Theorem 3.4.6] implies that

\[
\text{cum}_1\left(N^\gamma(\hat{E}_T(y_1) - \hat{E}_T(y_2))\right) = \sum_\nu V_\gamma(\nu),
\]

where the term \( V_\gamma(\nu) \) is defined by

\[
V_\gamma(\nu) := \frac{1}{N^{l(2-2\gamma)}(2\pi)^l} \sum_{(j_1,..,j_l) \in A_{T,\gamma}(v_1,v_2)^l} \sum_{k_1,..,k_l=-\lfloor (N-1)/2 \rfloor}^{N/2} \sum_{m_1=0}^{\infty} \sum_{n_1=0}^{\infty} \sum_{p_1=0}^{N-1} \prod_{s=1}^{l} \phi(j_s, \lambda_{k_s})
\]

\[
\times \prod_{s=1}^{l} \left[ \psi_{m_s} \psi_{n_s} \exp(-i\lambda_{k_1}(p_1 - p_l + m_l - n_1 + j_1 - j_l)) \right]
\]

\[
\times \prod_{s=2}^{l} \exp(-\lambda_{k_s}(p_s - p_{s-1} + m_{s-1} - n_s - j_{s-1} + j_s)),
\]
where the restrictions (3.117) in the summation over \( p_1, \ldots, p_l \) follow from the fact that \( q_1, \ldots, q_l \in \{0, \ldots, N - 1\} \) and the set \( A_{T,3}(v_1, v_2) \) is defined by

\[
A_{T,3}(v_1, v_2) := \{[v_1 T] - N/2, [v_1 T] + N/2, [v_2 T] - N/2, [v_2 T] + N/2\}.
\]  

(3.149)

Now, we rename the \( m_i \) and \( n_i \) [we denote \( m_i \) by \( n_i \) and replace \( n_i \) by \( m_{i-1} \)], which yields the conditions

\[
0 \leq q_{i+1} = p_i + m_i - n_i + j_i - j_{i+1} \leq N - 1 \text{ for } i \in \{1, 3, \ldots, l - 1\} \tag{3.150}
\]

\[
0 \leq q_{i+1} = p_i + m_i - n_i + j_i - j_{i+1} \leq N - 1 \text{ for } i \in \{2, 4, \ldots, l - 2\} \tag{3.151}
\]

\[
0 \leq q_1 = p_1 + m_1 - n_1 + j_1 - j_1 \leq N - 1 \tag{3.152}
\]

for the summation over \( p_1, \ldots, p_l \). In the next step, we divide the sum over \( p_i, m_i \) and \( n_i \) into the product of two sums, namely the sums over \( p_i, m_i \) and \( n_i \) for even values \( i \in \{1, \ldots, l\} \) and the sum over \( p_i, m_i \) and \( n_i \) for odd values \( i \in \{1, \ldots, l\} \). By an application of the Cauchy-Schwarz inequality, we obtain the bound

\[
V_\gamma(\bar{\nu}) \leq \sqrt{V_{1,\gamma}(\bar{\nu}) V_{2,\gamma}(\bar{\nu})},
\]

where

\[
V_{1,\gamma}(\bar{\nu}) := \frac{1}{N^{l(2-\gamma)}(2\pi)^l} \sum_{(j_1, \ldots, j_l) \in A_{T,3}(v_1, v_2)^l} N/2 \prod_{k \in \{1, 3, \ldots, l-1\}} \phi^2(j_{s_2}, \lambda_{s_1}) \times \left| \sum_{p_1, p_2, \ldots, p_{l-1} = 0}^{N-1} m_1, m_3, m_5, \ldots \prod_{s \in \{1, 3, \ldots, l-1\}} \psi_{m_s} \psi_{n_s} \exp(-i(\lambda_{k_s} - \lambda_{k_{s+1}})s) \right|^2 \tag{3.153}
\]

and the quantity \( V_{2,\gamma}(\bar{\nu}) \) is defined in an analogous manner with the respective summations carried out over \( p_i, m_i, n_i \) for even \( i \in \{1, \ldots, l\} \) and with respect to conditions (3.151) and (3.152). Simple calculations reveal that the term \( V_{1,\gamma}(\bar{\nu}) \) is bounded by

\[
\frac{C}{N^{l(1-\gamma)}} \sum_{(j_1, \ldots, j_l) \in A_{T,3}(v_1, v_2)^l} \prod_{k \in \{1, 3, \ldots, l-1\}} \phi^2(j_{s_2}, \lambda_{s_1}) |H_{T,1}(j_1, j_2, \ldots, j_l, \lambda_1, \lambda_3, \ldots, \lambda_{l-1})|,
\]
where the function $H_{T,1}$ is defined by

$$H_{T,1}(j_1, j_2, \ldots, j_t, \lambda_{k_1}, \lambda_{k_3}, \ldots, \lambda_{k_{T-1}}) := \frac{1}{N^l} \sum_{p_1, \ldots, p_{T-1}=0}^{N-1} \sum_{m_1, m_2, \ldots, m_{T-1}=0}^{\infty} \prod_{s \in \{1, \ldots, l-1\}} \exp(-i\lambda_{k_s}(p_s - \bar{p}_s)) \prod_{s \in \{1, \ldots, l-1\}} (\psi_{m_s, \psi_{n_s}}) \times \psi_{\bar{m}_s, \psi_{\bar{n}_s}}$$

and the expression (3.150) denotes the condition (3.150) with $p_i, m_i, n_i$ replaced by $\bar{p}_i, \bar{m}_i, \bar{n}_i$. The identity

$$\frac{1}{N} \sum_{k=-\left\lfloor \frac{N-1}{2} \right\rfloor}^{N/2} \exp(\lambda_k p) = \begin{cases} 1 & \text{if } p = lN \text{ for some } l \in \mathbb{N} \\ 0 & \text{else} \end{cases}$$

together with the restriction $0 \leq \bar{p}_i \leq N-1$ implies that, for any fixed choice of $p_i, m_i, n_i, \bar{m}_i, \bar{n}_i$, there is only one choice for $\bar{p}_i$ such that the corresponding summand in the term (3.154) does not vanish. Therefore, we get

$$\sum_{(j_2, j_3, \ldots, j_l) \in A_{T,3}(v_1, v_2)^{l/2}} |H_{T,1}(j_1, j_2, \ldots, j_l, \lambda_{k_1}, \lambda_{k_3}, \ldots, \lambda_{k_{T-1}})| \leq \sum_{(j_2, j_3, \ldots, j_l) \in A_{T,3}(v_1, v_2)^{l/2}} \prod_{s \in \{1, \ldots, l-1\}} |\psi_{m_s, \psi_{n_s}}| \psi_{\bar{m}_s, \psi_{\bar{n}_s}} = 4^{l/2} \left( \sum_{n=0}^{\infty} |\psi_n| \right)^{2l}, \quad (3.155)$$

where we employed the fact that $\text{card}(A_{T,3}(v_1, v_2)) = 4$. The summability condition (3.10) shows that (3.155) is equal to $C^l$ for some constant $C > 0$ and we thus obtain that $V_{1,\gamma}(\bar{v})$ is bounded by

$$\frac{C^l}{N^{l(1-\gamma)}} \sum_{j_1, j_3, \ldots, j_{T-1}=1}^{T} \sum_{k_1, k_3, \ldots, k_{T-1}=-[(N-1)/2]}^{N/2} \left( \prod_{s \in \{1, \ldots, l-1\}} \phi^2(j_s, \lambda_{k_s}) \right),$$

which is equal to $C^l \rho_{2, T, \gamma}(y_1, y_2)^l$ for some constant $C > 0$. Since similar calculations yield the same upper bound for the quantity $V_2(\bar{v})$, the property (3.148) follows by

$$V_\gamma(\bar{v}) \leq \sqrt{V_{1,\gamma}(\bar{v})V_{2,\gamma}(\bar{v})} \leq C^l \rho_{2, T, \gamma}(y_1, y_2)^l \leq C^l d_{T,\gamma}(y_1, y_2)^l,$$

where the bound (3.143) for the semi-metric $\rho_{2, T, \gamma}$ was used in the last step. \hfill $\Box$
Proof of Lemma 3.4.8

For the proof of Lemma 3.4.8, note that the definition of the semi-metric \( d_{T,\gamma} \) implies that there exists some constant \( C > 0 \) such that

\[
N(u, d_{T,\gamma}, P_T) \leq \begin{cases} 
TN & \text{for } u < C/N^{1/2-\gamma} \\
1 & \text{for } u \geq C/N^{1/2-\gamma}.
\end{cases}
\]

This bound yields that, for any \( \kappa > 0 \), the covering integral \( J(\kappa, d_{T,\gamma}, P_T) \) of the set \( P_T \) with respect to the semi-metric \( d_{T,\gamma} \) is bounded by

\[
J\left(\frac{C}{N^{1/2-\gamma}}, d_{T,\gamma}, P_T\right) + 1_{\{\kappa > C/N^{1/2-\gamma}\}} \int_{C/N^{1/2-\gamma}}^{\kappa} \left[ \log\left(\frac{48N(u, d_{T,\gamma}, P_T)^2}{u}\right)\right]^2 du
\leq \int_{0}^{C/N^{1/2-\gamma}} \left[ \log(48T^2N^2)^2 - 2 \log(48T^2N^2) \log(u) + \log(u)^2\right] du + \left| \int_{C/N^{1/2-\gamma}}^{\kappa} \left[ \log\left(\frac{48}{u}\right)\right]^2 du \right|.
\]

The fact \( T^e/N \to 0 \) implies that the right hand side of the above expression vanishes as \( T \to \infty \) and \( \kappa \to 0 \).

Proof of Theorem 3.1.2 b): Assertion b) of Theorem 3.1.2 can be established using the same arguments as were explained in the proof of part b) of Theorem 3.1.1.

3.4.3 Proof of Theorem 3.1.6

As in the previous proof, we restrict ourselves without loss of generality to the case \( d = 1 \). Furthermore, we suppress the argument \( T \) when referring to the sequence \( p = p(T) \).

For a proof of Theorem 3.1.6, we define the process

\[
X_{t}^{AR}(p) := \sum_{j=1}^{p} a_{j,p} X_{t-j}^{AR}(p) + Z_{t}^{AR}(p)
\]

for \( p \geq 1 \), where \((a_{1,p},...,a_{p,p})\) denotes the minimiser in (3.19) and \( \{Z_{t}^{AR}(p)\}_{t \in \mathbb{Z}} \) denotes a sequence of independent centred Gaussian random variables with variance

\[
\sigma_{p} = \mathbb{E}\left[\left( X_{t} - \sum_{j=1}^{p} a_{j,p} X_{t-j} \right)^2 \right].
\]

This definition implies that \( \{X_{t}^{AR}(p)\}_{t \in \mathbb{Z}} \) corresponds to the best approximation of the observed data \( \{X_{t,T}\}_{t=1,...,T} \) by an AR(\( p \)) model. Assumption 3.1.3 allows an application of Lemma 2.3 in Kreiß et al. (2011), which shows that the time series model \( \{X_{t}^{AR}(p)\}_{t \in \mathbb{Z}} \) has
a MA(\infty) representation of the form

$$X^A_R(p) = \sum_{j=0}^{\infty} \psi^A_R(p) Z_{t-j}(p)$$

for \( p \) sufficiently large. Moreover, Assumption 3.1.3 and the condition (3.24), which implies the consistency of the estimators \( \hat{a}_{j,p}, j \in \{1,\ldots,p\} \) for the coefficients \( a_{j,p}, j \in \{1,\ldots,p\} \), together with Lemma 2.3 of Kreiß et al. (2011) yield that the bootstrap time series \( \{X^*_t\}_{t=1,\ldots,T} \) has a MA(\infty) representation of the form

$$X^*_t = \sum_{l=0}^{\infty} \hat{\psi}^A_R(p) Z_{t-l}^*$$

for \( p \) sufficiently large. Note that (3.157) corresponds to a stationary time series model. For a proof of Theorem 3.1.6, we therefore intend to apply the same argumentation as in the proof of Theorem 3.1.1 a). For this purpose, we recall that in the proof of Theorem 3.1.1 a) we relied on the fact that the error terms arising in various calculations could all be bounded by applying the property

$$\left( \sum_{m=0}^{\infty} |\psi_m| \right)^{q_1} \left( \sum_{l=0}^{\infty} l |\psi_l| \right)^{q_2} = O\left( \frac{1}{N} \right)$$

for \( q_1, q_2 \in \mathbb{N} \). This bound is an obvious consequence of the summability condition (3.10), which we imposed on the sequence \( \{\psi_l\}_{l \in \mathbb{N}} \) of linear coefficients. The proof of Theorem 3.1.6 can be conducted in the same way as the proof of Theorem 3.1.1 a), if we show that the error terms

$$\left( \sum_{m=0}^{\infty} |\hat{\psi}_m^A(p)| \right)^{q_1} \left( \sum_{l=0}^{\infty} l |\hat{\psi}^A_R(p)| \right)^{q_2} \frac{N}{N}$$

which are random, are of order \( O_P(1/N) \). Hence, the assertion of Theorem 3.1.6 follows if we show that

$$\sum_{l=0}^{\infty} |l| |\hat{\psi}^A_R(p)| = O_P(1).$$

(3.158)

The inequality

$$\sum_{l=0}^{\infty} |l| |\hat{\psi}^A_R(p)| \leq \sum_{l=0}^{\infty} |l| |\hat{\psi}^A_R(p) - \psi^A_R(p)| + \sum_{l=0}^{\infty} |l| |\psi^A_R(p) - \psi_l| + \sum_{l=0}^{\infty} |l| |\psi_l|$$
and the summability condition (3.10) imply that, for establishing the claim (3.158), it is sufficient to show the following claims:

\[
\sum_{l=0}^{\infty} ||l|| \hat{\psi}^AR_{l}(p) - \psi^AR_{l}(p) || = O_P(1), \tag{3.159}
\]

\[
\sum_{l=0}^{\infty} ||l|| \psi^AR_{l}(p) - \psi_{l} || = O(1). \tag{3.160}
\]

For a proof of (3.159) and (3.160), we will make use of results of Kreiß et al. (2011). Lemma 2.3 of Kreiß et al. (2011) implies that the complex polynomials

\[ A_p(z) := 1 - \sum_{k=1}^{p} a_{j,p} z^k \quad \text{and} \quad \hat{A}_p(z) := 1 - \sum_{k=1}^{p} \hat{a}_{j,p} z^k \]

have no roots within the closed unit disc \( \{ z | |z| \leq 1 + 1/p \} \) if \( p \) is sufficiently large. An application of Cauchy’s inequality for holomorphic functions to the difference

\[ \hat{A}_p^{-1}(z) - A_p^{-1}(z) := \sum_{k=1}^{\infty} [\hat{\psi}^AR_k(p) - \psi^AR_k(p)] z^k, \]

as in the proof of Lemma 2.5 in Kreiß et al. (2011), yields the bound

\[
|l||\hat{\psi}^AR_{l}(p) - \psi^AR_{l}(p)|| \leq |l| \left( \frac{1}{1 + 1/p} \right)^l \max_{|z|=1+1/p} \left| \frac{A_p(z) - \hat{A}_p(z)}{A_p(z) \hat{A}_p(z)} \right| \leq |l| \left( \frac{1}{1 + 1/p} \right)^l \max_{|z|=1+1/p} \frac{\sum_{k=1}^{p} |\hat{a}_{k,p} - a_{k,p}|(1 + \frac{1}{p})^k}{|A_p(z) \hat{A}_p(z)|} = p|l| \left( 1 + \frac{1}{p} \right)^{-l} O_P \left( \sqrt{\frac{\log(T)}{T}} \right), \tag{3.161}
\]

which is uniform in \( l \) and \( p \). An application of the geometric series formula \( \sum_{k=0}^{\infty} k q^k = \frac{q}{(1-q)^2} \) for \( |q| < 1 \) shows that

\[
\sum_{l=0}^{\infty} |l||\hat{\psi}^AR_{l}(p) - \psi^AR_{l}(p)|| \leq O_P \left( p(T) \sqrt{\frac{\log(T)}{T}} \right) \sum_{l=0}^{\infty} l \left( \frac{p}{1 + p} \right)^l = O_P \left( p^3 \max(T) \sqrt{\frac{\log(T)}{T}} \right) = O_P(1)
\]
due to the condition (3.23). This completes the proof of (3.159). In order to establish the claim (3.160), we apply Lemma 2.4 of Kreiß et al. (2011), which implies the bound

$$\sum_{l=0}^{\infty} |l||\psi_l^{AR}(p) - \psi_l| \leq \sum_{l=0}^{\infty} |1 + l||\psi_l^{AR}(p) - \psi_l| \leq C \sum_{j=p+1}^{\infty} (1 + j)|a_j| = O(1),$$

where the last identity follows from Assumption 3.1.3.

\[\square\]

### 3.4.4 Proof of Theorem 3.1.7

In the following demonstration, we restrict ourselves to the case \(d = 1\) and suppress the argument \(T\) when referring to the sequence \(p = p(T)\).

For a proof of Theorem 3.1.7, note that, due to Assumption 3.1.3, the stationary process \(\{X_t\}_{t\in\mathbb{Z}}\) with spectral density \(g(\lambda) = \int_{0}^{1} f(u, \lambda)du\) possesses a MA(\(\infty\)) representation

\[X_t = \sum_{l=0}^{\infty} \psi_l Z_{t-l}.\] (3.162)

Moreover, under the null hypothesis of no structural breaks, it is \(g = f\) and the functions \(\psi_l(u) (l \in \mathbb{N})\) in the linear representation (2.9) do not depend on the rescaled time \(t/T = u \in [0,1]\). Therefore, we denote the constant value of the function \(\psi_l(u)\) by \(\psi_l\), \((l \in \mathbb{N})\).

Furthermore, we define \(\{X_t^*\}_{t\in\mathbb{Z}}\) as the process which is obtained by replacing the innovations \(Z_t\) in the representation (3.162) by the bootstrap replicates \(Z_t^*\), which are generated according to step 2) of Algorithm 3.1.4, i.e. we set

\[X_t^* := \sum_{l=0}^{\infty} \psi_l Z_{t-l}^*.\] (3.163)

for \(t \in \{1, ..., T\}\). Now, we define the process \(\{\hat{E}_{T,a}^*(y)\}_{y\in[0,1]^2}\) in the same way as the empirical process \(\{\hat{E}_T(y)\}_{y\in[0,1]^2}\) was defined in (3.6), where the random variables \(\{X_{t,T}\}_{t=1,...,T}\) are replaced by \(\{X_t^*\}_{t=1,...,T}\).

In the following discussion, we show that the empirical process \(\{\hat{E}_{T,a}^*(y)\}_{y\in[0,1]^2}\) fulfils parts a), b) and c) of Theorem 3.1.7.

**Proof of part a):** The definition of the random variables \(X_t^*, \ t \in \{1,...,T\}\), implies that, under the null hypothesis (3.3) of no structural breaks, the sequences \(\{X_{t,T}\}_{t=1,...,T}\) and \(\{X_t^*\}_{t=1,...,T}\) have the same distribution. The distributional equivalence (3.28) obviously follows.

**Proof of part b):** For a proof of Theorem 3.1.7 part b), we reconsider for \(p \geq 1\) the time series \(\{X_t^{AR}(p)\}_{t\in\mathbb{Z}}\), which has already been of relevance in the proof of Theorem 3.1.6 and
is defined by

\[ X_t^{AR}(p) = \sum_{j=1}^{p} a_{j,p} X_{t-j}^{AR}(p) + Z_t^{AR}(p), \]  

(3.164)

where the coefficients \( a_{1,p}, \ldots, a_{p,p} \) were introduced in (3.19) and the \( Z_t^{AR}(p) \) are independent centred Gaussian random variables with variance

\[ \sigma_p = E\left[ (X_t - \sum_{j=1}^{p} a_{j,p} X_{t-j})^2 \right]. \]

Since \( \{X_t\} \) is the stationary process with spectral density \( g(\lambda) = \int_0^1 f(u, \lambda) du \), the process \( \{X_t^{AR}(p)\} \) corresponds to the ‘best’ approximation of the process (3.17) by an AR\((p)\) model. Lemma 2.3 in Kreiß et al. (2011) implies that for sufficiently large \( T \in \mathbb{N} \) the approximating process \( \{X_t^{AR}(p)\} \) and the bootstrap analog \( \{X_t^*, T\} \) in (3.20) have MA\((\infty)\) representations of the form

\[ X_t^{AR}(p) = \sum_{j=0}^{\infty} \psi_j^{AR}(p) Z_{t-j}^{AR}(p) \]  

(3.165)

\[ X_t^*, T = \sum_{j=0}^{\infty} \hat{\psi}_j^{AR}(p) Z_{t-j}^*. \]  

(3.166)

In order to enhance further arguments, we introduce the set \( W_T \) as the event on which the inequality

\[ \sqrt{N} \sum_{m,n=0}^{\infty} |n| |\hat{\psi}_m^{AR}(p)\hat{\psi}_n^{AR}(p) - \psi_m\psi_n| \leq 1 \]  

(3.167)

holds. From the arguments presented in the proof of Theorem 3.1.1, it is easily seen that there exists \( \tilde{y} \in [0, 1]^2 \) such that \( |\tilde{E}_{T,a}(\tilde{y})|^2 \geq C/N \) for some constant \( C > 0 \). This implies that (3.29) can be proven by showing that

\[ \sqrt{N} \sup_{y \in [0,1]^2} |\tilde{E}_{T,a}^*(y) - \tilde{E}_{T,a}^*(y)| = o_P(1). \]  

(3.168)

Now, we define a further empirical process \( \{\tilde{E}_T(y)\} \) by

\[ \tilde{E}_T(y) := \sqrt{N}(\tilde{E}_T^*(y) - \tilde{E}_{T,a}^*(y)) \times 1_{W_T}. \]  

(3.169)

This definition implies that, for a proof of the identity (3.168), it is sufficient to show that the process \( \{\tilde{E}_T(y)\} \) converges to zero uniformly in probability and that the probability of the event \( W_T \) converges to one as \( T \to \infty \). Theorem 2.3.13 implies that we achieve this by establishing the three results below.
Lemma 3.4.9 (Asymptotic stochastic equicontinuity)
For every \( \beta \in (0, \varepsilon) \), where \( \varepsilon \) denotes the constant introduced in (3.9), the empirical process \( \{ \tilde{E}_T(y) \}_{y \in [0,1]^2} \) is asymptotically stochastically equicontinuous with respect to the semi-metric
\[
d_\beta((v_1, \omega_1), (v_2, \omega_2)) = (|v_1 - v_2| + |\omega_1 - \omega_2|)^{\beta/2},
\]
i.e. for every \( \eta, \varepsilon > 0 \) there exists some \( \delta > 0 \) such that
\[
\lim_{T \to \infty} P\left( \sup_{y_1, y_2 \in [0,1]^2 : d_\beta(y_1, y_2) < \delta} |\tilde{E}_T(y_1) - \tilde{E}_T(y_2)| > \eta \right) < \varepsilon,
\]
where \( y_1 = (v_1, \omega_1) \) and \( y_2 = (v_2, \omega_2) \).

Lemma 3.4.10 (Pointwise convergence to zero in probability)
The process \( \{ \tilde{E}_T(y) \}_{(v, \omega) \in [0,1]^2} \) converges pointwise to zero in probability, i.e. for every \( y = (v, \omega) \in [0,1]^2 \) it holds
\[
\tilde{E}_T(y) = o_P(1).
\]

Lemma 3.4.11
For the set \( W_T \) defined in (3.167) we have
\[
P(W_T) \to 1
\]
as \( T \to \infty \).

Proof of Lemma 3.4.9
We recall the definition of the set \( \mathcal{P}_T \) given in (3.142) and the definition of the function \( \phi_y, T \), \( y \in \mathcal{P}_T \) introduced in (3.61). This notation implies that the above defined process has the representation
\[
\{ \tilde{E}_T(y) \}_{y \in [0,1]^2} = \{ \tilde{E}_T(y) \}_{y \in \mathcal{P}_T},
\]
with \( \tilde{E}_T(y) = \left( \frac{1}{\sqrt{N}} \sum_{j=1}^{T} \sum_{k=-\lfloor (N-1)/2 \rfloor}^{N/2} \phi_y(j, \lambda_k) \left( I_N^{(X_t^*, T)}(\frac{j}{T}, \lambda_k) - I_N^{(X_t)}(\frac{j}{T}, \lambda_k) \right) \right) \times 1_{W_T},
\]
where \( I_N^{(X_t^*, T)} \) and \( I_N^{(X_t)} \) denote the periodogram estimates based on the random variables \( \{X_t^*, T\}_{t=1,...,T} \) and \( \{X_t\}_{t=1,...,T} \) respectively. As in the proof of (3.140), the assertion of Lemma 3.4.9 follows, if we show that there exists a constant \( C > 0 \) such that the inequality
\[
|\text{cum}_l(\tilde{E}_T(y_1) - \tilde{E}_T(y_2))| \leq (6l)!C^l d_\beta(y_1, y_2)^l,
\]
holds uniformly in \( y_1, y_2 \in \mathcal{P}_T \) and for all \( l \in \mathbb{N} \). Therefore, we assume without loss of generality that \( y_1, y_2 \in \mathcal{P}_T \) are fixed such that \( y_1 \neq y_2 \), define the function \( \phi := \phi_{y_1, T} - \phi_{y_2, T} \)
and treat the cases \( l = 1, l = 2 \) and \( l = 3 \) separately.

In the case \( l = 1 \), we obtain by the independence of the innovations \( \{Z_t^*\}_{t \in \mathbb{Z}} \) and the estimates \( \{\hat{\psi}_m^{AR}(p)\}_{m \in \mathbb{Z}} \), that

\[
\mathbb{E}(\hat{E}_T(y_1) - \hat{E}_T(y_2)) = \frac{1}{2\pi N^{3/2}} \sum_{j=1}^{T} \sum_{k=-(N-1)/2}^{N/2} \phi(j, \lambda_k) \sum_{p=0}^{N-1} \exp(-i\lambda_k(p - q)) \times \mathbb{E}[1_{W_T}(X_{j-N/2+1+p,T}^*X_{j-N/2+q,T}^* - X_{j-N/2+1+p}^*X_{j-N/2+q})]
\]

\[
= \frac{1}{2\pi N^{3/2}} \sum_{j=1}^{T} \sum_{k=-(N-1)/2}^{N/2} \phi(j, \lambda_k) \sum_{l,m=0}^{\infty} \mathbb{E}[1_{W_T}(\hat{\psi}_l^{AR}(p)\hat{\psi}_m^{AR}(p) - \psi_l\psi_m)] \times \sum_{p,q=0}^{N-1} \exp(-i\lambda_k(p - q)) \mathbb{E}(Z_{j-N/2+1+p}^*Z_{j-N/2+1+q-m}^*) \tag{3.175}
\]

where we employed the MA(\( \infty \)) representations (3.166) and (3.163) for \( X_{l,T}^* \) and \( X_t^* \) respectively. The independence of the bootstrap innovations \( \{Z_t^*\}_{t \in \mathbb{Z}} \), which implies

\[
\mathbb{E}(Z_t^*Z_j^*) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}, \tag{3.176}
\]

yields that the condition \( q = p - l + m \) has to hold for the respective summands in (3.175) not to vanish. Therefore, we obtain that (3.175) is equal to

\[
\frac{1}{2\pi N^{3/2}} \sum_{j=1}^{T} \sum_{k=-(N-1)/2}^{N/2} \phi(j, \lambda_k) \sum_{l,m=0}^{\infty} \exp(-i\lambda_k(m - l)) \mathbb{E}[1_{W_T}(\hat{\psi}_l^{AR}(p)\hat{\psi}_m^{AR}(p) - \psi_l\psi_m)] \times \sum_{p=0}^{N-1} 1 \tag{3.177}
\]

The bound (3.167), which holds on the event \( W_T \), and the arguments provided in the demonstrations regarding the identity (3.69) yield that the error made by dropping the restriction \( 0 \leq p-l+m \leq N-1 \) in the summation over \( p \) is at most of order \( O(1/\sqrt{N}) \). Hence, we obtain that (3.177) is equal to

\[
\mathbb{E}\left(1_{W_T}\frac{1}{2\pi N^{3/2}} \sum_{j=1}^{T} \sum_{k=-(N-1)/2}^{N/2} \phi(j, \lambda_k) \sum_{l,m=0}^{\infty} \exp(-i\lambda_k(m - l)) (\hat{\psi}_l^{AR}(p)\hat{\psi}_m^{AR}(p) - \psi_l\psi_m) \right) + O\left(\frac{1}{\sqrt{N}}\right) = O\left(\frac{1}{\sqrt{N}}\right), \tag{3.178}
\]
where we used the definition of the function $\phi$ in the last step. The asymptotics (3.9) of $N$ and $T$ and the bound

$$\frac{1}{T} \leq (|v_1 - v_2| + |\omega_1 - \omega_2|),$$

(3.179)

which holds for all $(v_1, \omega_1) \neq (v_2, \omega_2) \in \mathcal{P}_T$, furthermore imply that (3.178) is not larger than

$$C \frac{1}{T^{\varepsilon/2}} \leq C(\{v_1 - v_2| + |\omega_1 - \omega_2|\}^{\varepsilon/2} \leq C d_9(y_1, y_2),$$

which completes the proof of (3.174) in the case $l = 1$. For the case $l = 2$, simple calculations show that

$$\sum_{\nu} V(\nu),$$

(3.180)

where the set $\mathcal{A}_{T,2}(v_1, v_2)$ for $v_1, v_2 \in [0, 1]$ was introduced in (3.73), the quantity $V(\nu)$ is defined by

$$V(\nu) := \frac{1}{(2\pi)^2 N^3} \sum_{(j_1, j_2) \in \mathcal{A}_{T,2}(v_1, v_2)} \sum_{k_1, k_2 = -[(N-1)/2]}^{N/2} \phi(j_1, k_1) \phi(j_2, k_2) \prod_{p_1, p_2 = 0}^{N-1} \prod_{q_1, q_2 = 0}^{N-1} \sum_{l, m, n, \omega = 0}^{\infty} \exp(-i\lambda_{k_1}(p_1 - q_1)) \exp(-i\lambda_{k_2}(p_2 - q_2))$$

and the summation is performed over all indecomposable partitions $\nu = (\nu_1, ..., \nu_p)$ of the table

$$\begin{align*}
Y_{1,1} &:= 1_{\mathcal{W}_T} \times (\hat{\psi}_1^{AR}(p)\hat{\psi}_m^{AR}(p) - \psi_1\psi_m) & Y_{1,2} &:= Z_{j_1 + p_1 + 1 - l}^{*} \quad Y_{1,3} := Z_{j_1 + q_1 + 1 - m}^{*} \\
Y_{2,1} &:= 1_{\mathcal{W}_T} \times (\hat{\psi}_n^{AR}(p)\hat{\psi}_o^{AR}(p) - \psi_n\psi_o) & Y_{2,2} := Z_{j_2 + p_2 + 1 - n}^{*} \quad Y_{2,3} := Z_{j_2 + q_2 + 1 - o}^{*} \quad (3.181)
\end{align*}$$

Due to the independence of the sequences $\{Z_t^*\}_{t \in \mathbb{Z}}$ and $\{\hat{\psi}_m^{AR}(p)\}_{m \in \mathbb{N}}$, we only need to consider partitions $\nu$, that are solely comprised of elements of either the first column of (3.181) or of the second and third column of (3.181) [see Theorem 2.3.2]. Furthermore, the sub par-
A new approach to the change point problem in time series analysis

and we will show that

\[ |V(\tilde{\nu})| \leq C^2 d_\beta(y_1, y_2)^2. \]

The inequality (3.174) for the case \( l = 2 \) then follows due to the fact that the term \( V(\nu) \) can be bounded in the same way for all indecomposable partitions of (3.181) and that \( 12!3^2 \) is an upper bound for the number of partitions of this scheme. The property \( \text{cum}(Z_i^*, Z_j^*) = \delta_{ij} \) implies that the relations

\begin{align*}
0 &\leq q_2 = p_1 - l + o + j_1 - j_2 \leq N - 1, \\
0 &\leq q_1 = p_2 - n + m + j_2 - j_1 \leq N - 1
\end{align*}

have to hold for each non-vanishing summand in the definition of \( V(\tilde{\nu}) \) and we obtain

\[
V(\tilde{\nu}) = \frac{1}{(2\pi)^2 N^3} \sum_{(j_1,j_2) \in A_{T/2}(v_1,v_2)} \sum_{k_1,k_2 = 0}^{N/2} \phi(j_1, \lambda_{k_1}) \phi(j_2, \lambda_{k_2}) \sum_{l,m,n,o=0}^{N-1} \sum_{(p_1,p_2=0)}^{\infty} \exp(-i\lambda_{k_1}(p_1 - p_2 - m + n - j_2 + j_1)) \exp(-i\lambda_{k_2}(p_2 - p_1 + l - o - j_1 + j_2)) \times \mathbb{E}\left[ 1_{W_T} \times (\hat{\psi}_i^{AR}(p)\hat{\psi}_m^{AR}(p) - \psi_i\psi_m) \right] \mathbb{E}\left[ 1_{W_T} \times (\hat{\psi}_n^{AR}(p)\hat{\psi}_o^{AR}(p) - \psi_n\psi_o) \right].
\]

The calculations presented in the proof of Lemma 3.4.1 to reach (3.91) together with the bound (3.167), which holds on the set \( W_T \), yield that we can change the conditions (3.182) and (3.183) to

\begin{align*}
0 &\leq q_2 = p_1 + j_1 - j_2 \leq N - 1, \\
0 &\leq q_1 = p_2 + j_2 - j_1 \leq N - 1
\end{align*}

respectively by making an error that is at most of order \( O(1/N^{1-\varepsilon'}) \) for any \( \varepsilon' > 0 \). This implies that the absolute value of \( V(\tilde{\nu}) \) is at most of size

\[
\frac{C^2}{N^4} \sum_{k_1,k_2 = 0}^{N/2} \left| \sum_{l,m=0}^{\infty} \left[ \mathbb{E}\left( 1_{W_T} \sqrt{N}[\hat{\psi}_i^{AR}(p)\hat{\psi}_m^{AR}(p) - \psi_i\psi_m] \right) \exp(i\lambda_{k_1}m - i\lambda_{k_2}l) \right] \right| \times \left| \sum_{n,o=0}^{\infty} \left[ \mathbb{E}\left( 1_{W_T} \sqrt{N}[\hat{\psi}_n^{AR}(p)\hat{\psi}_o^{AR}(p) - \psi_n\psi_o] \right) \exp(-i\lambda_{k_1}n + i\lambda_{k_2}o) \right] \right| \times \left| \sum_{p_1,p_2=0}^{N-1} \exp(-i(\lambda_{k_1} - \lambda_{k_2})(p_1 - p_2)) \right| + O\left( \frac{1}{N^{1-\varepsilon'}} \right).
\]
which, due to (3.167), is not larger than

\[
\frac{C^2}{N^4} \sum_{k_1, k_2=-\lfloor (N-1)/2 \rfloor}^{N/2} \left| \sum_{p_1, p_2=0}^{N-1} \exp(-i(\lambda_{k_1} - \lambda_{k_2})(p_1 - p_2)) \right| + O\left(\frac{1}{N^{1-\varepsilon}}\right)
\]

\[= V'_{k_1=k_2,T} + V'_{k_1 \neq k_2,T} + O\left(\frac{1}{N^{1-\varepsilon}}\right),\]

where the symbols \(V'_{k_1=k_2,T}\) and \(V'_{k_1 \neq k_2,T}\) denote the sums over all \((k_1, k_2)\) satisfying \(k_1 = k_2\) and \(k_1 \neq k_2\) respectively. It is obvious to see that the quantity \(V'_{k_1=k_2,T}\) is of order \(O(1/N)\) and we continue by considering \(V'_{k_1 \neq k_2,T}\). Therefore, we employ (3.88) and the bound (3.90) for the sin function to get

\[
V'_{k_1 \neq k_2} \leq \frac{C^2}{N^4} \sum_{k_1, k_2=-\lfloor (N-1)/2 \rfloor}^{N/2} \frac{1}{\sin^2\left(\frac{k_1-k_2}{N}\right)} \leq \frac{C^2}{N^2} \sum_{k_1, k_2=-\lfloor (N-1)/2 \rfloor}^{N/2} \frac{1}{|k_1-k_2|^2} + \frac{C^2}{N^2} \sum_{k_1, k_2=-\lfloor (N-1)/2 \rfloor}^{N/2} \frac{1}{(N - |k_1-k_2|)^2} \leq \frac{C^2}{N^2} \sum_{k_1=-\lfloor (N-1)/2 \rfloor}^{N/2} \sum_{k=1}^{N} \frac{1}{k^2} \leq \frac{C^2}{N}.
\]

Thus, we have shown that for any \(\varepsilon' > 0\) it holds

\[
\text{cum}_2\left(\bar{E}_T(y_1) - \bar{E}_T(y_2)\right) \leq \frac{C^2}{N^{1-\varepsilon'}} \leq \frac{C^2}{2^{1-\varepsilon}(1-\varepsilon')} \leq C^2 2! (|v_1 - v_2| + |\omega_1 - \omega_2|)^{\varepsilon(1-\varepsilon')},
\]

where we used \(T^e/N \to 0\) and employed the bound (3.179). By taking \(\varepsilon'\) small enough, we obtain the value \(C^2 [d_\beta(y_1, y_2)]^2\) as an upper bound for this expression, which completes the proof of (3.174) in the case \(l = 2\).

Finally, we consider the case \(l = 3\) and obtain by an application of the product theorem for cumulants [see Theorem 2.3.6] that for each \(y_1 = (v_1, \omega_1), y_2 = (v_2, \omega_2)\) we have

\[
\text{cum}_l\left(\bar{E}_T(y_1) - \bar{E}_T(y_2)\right) = \frac{1}{N^{2l}(2\pi)^l} \sum_{(j_1, \ldots, j_l) \in \Lambda_{T,3}(v_1, v_2)} \sum_{k_1, \ldots, k_l=-\lfloor (N-1)/2 \rfloor}^{N/2} \sum_{m_1, \ldots, m_l=0}^{\infty} \sum_{p_1, \ldots, p_l, q_1, \ldots, q_l=0}^{\infty} \sum_{s=1}^{N-1} \prod_{s=1}^{l} [\phi(j_s, \lambda_{k_s}) \exp(-\lambda_{k_s}(p_s - q_s))] \times \text{cum}(Y_1, Y_2, Y_3, Y_{2,1} Y_{2,2} Y_{2,3}, \ldots, Y_{1,1} Y_{1,2} Y_{1,3}),
\]

\[= \sum_{\nu} V(\nu), \quad (3.186)\]
where for \( v_1, v_2 \in [0, 1] \) the set \( A_{T,3}(v_1, v_2) \) was introduced in (3.149), \( \phi \) denotes the function 
\( \phi := \phi_{y_1,T} - \phi_{y_2,T} \), the quantity \( V(\nu) \) is defined by 
\[
V(\nu) := \frac{1}{N^{N/2} (2\pi)^l} \sum_{(j_1, \ldots, j_l) \in A_{T,3}(v_1, v_2)} \sum_{k_1, \ldots, k_l = -[(N-1)/2]}^{N/2} \sum_{m_1, \ldots, m_l = 0}^{\infty} \sum_{n_1, \ldots, n_l = 0}^{\infty} \sum_{p_1, \ldots, p_l = 0}^{\infty} \sum_{q_1, \ldots, q_l = 0}^{\infty} \prod_{s=1}^{l} \phi(j_s, \lambda_{k_s}) \exp(-\lambda_{k_s}(p_s - q_s)) \cum(Y_{a,b}; (a, b) \in \nu_1) \cdots \cum(Y_{a,b}; (a, b) \in \nu_l),
\]
and the summation is performed with respect to all indecomposable partitions \( \nu = (\nu_1, \ldots, \nu_p) \) of the table 
\[
\begin{array}{ccc}
Y_{1,1} & Y_{1,2} & Y_{1,3} \\
Y_{2,1} & Y_{2,2} & Y_{2,3} \\
\vdots & & \vdots \\
Y_{l,1} & Y_{l,2} & Y_{l,3}
\end{array}
\]  
(3.187)
where we introduced the notation 
\[
Y_{i,1} := 1_{W_T} \times (\hat{\psi}_{m_i}^{AR}(p)\hat{\psi}_{n_i}^{AR}(p) - \psi_{m_i}\psi_{n_i}) \quad Y_{i,2} := Z_{j_i+p_i-1=m_i} \quad Y_{i,3} := Z_{j_i+q_i+1-n_i},
\]
for \( i \in \{1, 2, \ldots, l\} \). Furthermore, we note that, due to the independence of the sequence \( \{Z_t^*\}_{t \in \mathbb{Z}} \) of the innovations and the sequence \( \{\hat{\psi}_{m}^{AR}(p)\}_{m \in \mathbb{N}} \) of random linear coefficients, in the summation in (3.186) it is only necessary to consider those partitions \( \nu \) of table (3.187), which can be represented as the conjunction \( \nu = \nu_1 \cup \nu_2 \) of two partitions \( \nu^1 \) and \( \nu^2 \), where \( \nu^1 \) is a partition of the first column of (3.187) and \( \nu^2 \) is a partition of the second and third column of (3.187). Due to Theorem 2.3.2, all other partitions do not have a contribution. Moreover, the Gaussianity of the innovations \( \{Z_t^*\}_{t \in \mathbb{Z}} \) implies that only those partitions \( \nu^2 \) of columns two and three need to be considered which are comprised of \( l \) elements [see Example 2.3.1]. Thus, it is easily seen that (3.186) is bounded by 
\[
|\sum_{\nu^2} \sum_{\nu^1 \in (\nu^2)} V(\nu^1 \cup \nu^2)|,
\]
where the first summation is performed with respect to all partitions \( \nu^2 \) of columns two and three and the second summation is carried out with respect to all partitions \( \nu^1 \) of column one satisfying the condition \( \star(\nu^2) \) defined by 
\[
\star(\nu^2) :\iff \nu = \nu^1 \cup \nu^2 \text{ is an indecomposable partition of the table (3.187)}.
\]
In the following demonstrations, we will show that, for each partition \( \nu^2 \), it holds 
\[
|\sum_{\nu^1 \in (\nu^2)} V(\nu^1 \cup \nu^2)| \leq C^l(3l)! [d_\beta(y_1, y_2)]^l.
\]  
(3.189)
The assertion (3.174) is then implied by the fact that there exist at most \((2!)^l 2^j\) partitions \(\nu^2\) of the second and third column of (3.187). For a proof of (3.189), we consider, without loss of generality, the partition

\[
\nu^2 := \bigcup_{i=1}^{l-1} (Y_{i,2}, Y_{i+1,3}) \cup (Y_{1,2}, Y_{1,3}),
\]

(3.190)

for which we obtain the conditions

\[
0 \leq q_{i+1} = p_i - m_i + n_{i+1} + j_i - j_{i+1} \leq N_1 \quad \text{for} \quad i \in \{1, ..., l - 1\},
\]

(3.191)

\[
0 \leq q_1 = p_1 - m_1 + n_1 - j_1 + j_1 \leq N_1,
\]

(3.192)

which result from the independence of the innovations \(\{Z_t^k\}_{t \in \mathbb{Z}}\). Because of \(p_i \in \{0, ..., N - 1\}\), the restrictions (3.191) and (3.192) imply the bounds

\[
|n_{i+1} - m_i + j_i - j_{i+1}| \leq N \quad \text{for} \quad i \in \{1, ..., l - 1\},
\]

(3.193)

\[
|n_1 - m_1 - j_1 + j_1| \leq N.
\]

(3.194)

Hence, we have that \(\left| \sum_{s=2}^{N/2} V(\nu^1 \cup \nu^2) \right|\) is equal to

\[
\left| \sum_{s=1}^{N/2} \frac{1}{N^{2l}(2\pi)^l} \sum_{(j_1, ..., j_l) \in A_{r,3}(v_1, v_2)^l} \prod_{k=1}^{l-1} \phi(j_k, \lambda_{k_i}) \exp(-i\lambda_{k_1} (p_1 - p_l + m_l - n_1 + j_1 - j_l)) \right|
\]

\[
\times \prod_{s=2}^{l} \exp(-i\lambda_{k_1} (p_s - p_{s-1} + m_{s-1} - n_s - j_{s-1} + j_s)) \sum_{s=1}^{N/2} \text{cum}(Y_{a,t}; (a, t) \in \nu^1).
\]

Simple calculations show that the above expression is the same as

\[
\left| \frac{1}{N^{2l}(2\pi)^l} \sum_{(j_1, ..., j_l) \in A_{r,3}(v_1, v_2)^l} \prod_{k=1}^{l-1} \phi(j_k, \lambda_{k_i}) \exp(-i\lambda_{k_1} (p_1 - p_l + m_l - n_1 + j_1 - j_l)) \right|
\]

\[
\times \prod_{s=2}^{l} \exp(-i\lambda_{k_1} (p_s - p_{s-1} + m_{s-1} - n_s - j_{s-1} + j_s)) \sum_{s=1}^{N/2} \text{cum} \left( \sqrt{N} Y_{a,t}; (a, t) \in \nu^1 \right).
\]
By applying the same arguments as in the transition from the term (3.116) to (3.124) in part (3) of the proof of Lemma 3.4.1, we get that the above quantity is not larger than

\[
\frac{C}{N} \sum_{(j_1, \ldots, j_l) \in A_{\nu, \bar{a}}(0, 0)} \sum_{m_1, n_1 = 0}^{\infty} \sum_{p_1 = 0}^{N - 1} \log(2N)^{-l - 1} |m_1 - n_1 + m_2 - n_2 + \ldots + m_l - n_l| \times |\sum_{\nu^1} \text{cum}(\sqrt{N} Y_{a,1}; (a, 1) \in \nu^1)\|
\]

\[
\leq \frac{C}{N} N \log(2N)^{-l - 1} \sum_{m_1 \ldots m_l = 0}^{\infty} \sum_{n_1 \ldots n_l = 0}^{\infty} |\sum_{\nu^1} \text{cum}(\sqrt{N} Y_{a,1}; (a, 1) \in \nu^1)\| \sum_{\nu_p} |\sum_{\nu^1} \text{cum}(\sqrt{N} Y_{a,1}; (a, 1) \in \nu_p)\|
\]

(3.195)

for some constant \(C > 0\) and any \(\varepsilon' > 0\). Note that the definition of the cumulant [see Definition 2.3.1] implies the inequality

\[
\sum_{m_1, n_1 = 0}^{\infty} |\sum_{(i, 1) \in \nu^1} \text{cum}(\sqrt{N} Y_{i,1}; (i, 1) \in \nu^1)|
\]

\[
\leq \sum_{m_1, n_1 = 0}^{\infty} \sum_{\nu^1 \text{partition of } \nu^1} q! \mathbb{E}\left( \prod_{(i, 1) \in P_1} \sqrt{N} |Y_{i,1}| \right) \ldots \mathbb{E}\left( \prod_{(i, 1) \in P_q} \sqrt{N} |Y_{i,1}| \right)
\]

\[
= \sum_{\nu^1 \text{partition of } \nu^1} q! \mathbb{E}\left( \prod_{(i, 1) \in P_1} \sqrt{N} \sum_{m_1, n_1 = 0}^{\infty} |Y_{i,1}| \right) \ldots \mathbb{E}\left( \prod_{(i, 1) \in P_q} \sqrt{N} \sum_{m_1, n_1 = 0}^{\infty} |Y_{i,1}| \right)
\]

for each element \(\nu^1_j, j \in \{1, \ldots, p\}\), of the partition \(\nu^1\). We now employ the fact that there exist at most \([\text{card}(\nu^1_j)]!2^{\text{card}(\nu^1_j)}\) partitions \(P = (P_1, \ldots, P_q)\) of the elements contained in \(\nu^1_j\), the definition of the random variable \(Y_{i,1}\) and the property (3.167), which holds on the set \(W_T\), to obtain the inequality

\[
\sum_{m_1, n_1} |\sum_{(i, 1) \in \nu^1_j} \text{cum}(\sqrt{N} Y_{a,1}; (a, 1) \in \nu^1_j)| \leq [\text{card}(\nu^1_j)]!2^{\text{card}(\nu^1_j)} \leq [2\text{card}(\nu^1_j)]!2^{\text{card}(\nu^1_j)}
\]
for all \( j = 1, ..., p \). From these inequalities, it easily follows that (3.195) is bounded by

\[
\frac{C^l}{N^{(l-1)(1-\varepsilon')}} \sum_{\nu^1 \ast (\nu^2)} (2l)!2^l \leq \frac{C^l}{N^{(l-1)(1-\varepsilon')}} (2l)!2^l \leq \frac{C^l}{N^{(l-1)(1-\varepsilon')}} (3l)!,
\]

where the first inequality is due to the fact that the number of partitions \( \nu^1 \) of column one is bounded by \( 2^l \). Moreover, the assumption (3.9) regarding the growth rate of the block size \( N \) and the property (3.179) shows that it holds

\[
| \sum_{\nu^1 \ast (\nu^2)} V(\nu^1 \cup \bar{\nu}^2) | \leq \frac{C^l}{N^{l/2(1-\varepsilon')}} (3l) \leq C^l (3l)!d_\beta(y_1, y_2)
\]

for all \( l \geq 3 \), where the last inequality is valid for sufficiently small \( \varepsilon' \) and any \( \beta \in (0, \varepsilon) \). This completes the proof of (3.174).

Proof of Lemma 3.4.10

In order to show that the random variable \( \bar{E}_T(y) \) is of order \( o_P(1) \) for each \( y \in [0, 1]^2 \), we assume that \( y \in [0, 1]^2 \) is fixed and arbitrary and note that it is sufficient to show

\[
E(\bar{E}_T(y)) = o(1),
\]

\[
V(\bar{E}_T(y)) = o(1).
\]

These claims follow by the same arguments as were employed in the demonstrations of (3.174) for the cases \( l = 1 \) and \( l = 2 \).

Proof of Lemma 3.4.11

For a proof of (3.173), note that we have the bound

\[
\sum_{m,n=0}^{\infty} |n| |\psi_m \psi_n - \hat{\psi}^{AR}_m(p) \hat{\psi}^{AR}_n(p)|
\]

\[
\leq \sum_{m=0}^{\infty} |\psi_m| \sum_{n=0}^{\infty} |n| |\psi_n - \hat{\psi}^{AR}_n(p)| + \sum_{m=0}^{\infty} |\hat{\psi}^{AR}_m(p)| \sum_{n=0}^{\infty} |n| |\psi_n - \hat{\psi}^{AR}_n(p)|.
\]

Thus, for establishing the property

\[
\sqrt{N} \sum_{m,n=0}^{\infty} |n| |\psi_m \psi_n - \hat{\psi}^{AR}_m(p) \hat{\psi}^{AR}_n(p)| = o_P(1)
\] (3.196)
it is sufficient to show the following equalities:

\[
\sqrt{N} \sum_{n=0}^{\infty} |n| |\hat{\psi}_n - \hat{\psi}^{AR}_n(p)| = o_P(1), \tag{3.197}
\]

\[
\sum_{n=0}^{\infty} |\hat{\psi}^{AR}_n(p)| = O_P(1). \tag{3.198}
\]

In order to establish the claims (3.197) and (3.198), we employ several arguments, which were presented in Kreiß et al. (2011). As in the proof of Theorem 3.1.6, where (3.159) and (3.160) were shown, an application of Lemma 2.3 of Kreiß et al. (2011) yields the bound

\[
\sum_{n=0}^{\infty} \frac{|\hat{\psi}^{AR}_n(p)|}{1 + \frac{p}{(p+1)^2}} = O_P(1), \tag{3.199}
\]

which holds for sufficiently large \(T\) and \(p\) and is uniform in \(l \in \mathbb{N}\) and \(p\). This bound together with the geometric series formula implies

\[
\sum_{n=0}^{\infty} \frac{|n| |\hat{\psi}_n^{AR} - \hat{\psi}_n(p)|}{1 + \frac{p}{(p+1)^2}} = O_P(1).
\]

Properties (3.199) and (3.200) imply (3.197) and the property (3.198) follows from (3.197) together with the summability condition (3.10).

\[\square\]

**Proof of part c):** Part c) is the bootstrap analog of Theorem 3.1.2 a) and is therefore shown by combining the arguments given in the proof of part b) with the reasoning in the proof of Theorem 3.1.2 a).

\[\square\]

### 3.4.5 Proof of Theorem 3.2.1

For a proof of Theorem 3.2.1, we define the deterministic matrix

\[
E_{N,T}(v, \omega) := \frac{T}{N} \left( \int_{-\omega \pi}^{\omega \pi} \int_{v-N/T}^{v} f(u, \lambda) du d\lambda - \int_{-\omega \pi}^{\omega \pi} \int_{v-N/T}^{v} f(u, \lambda) du d\lambda \right)
\]
for \((v, \omega) \in [0,1]^2\), \(N, T \in \mathbb{N}\) and emphasise that similar arguments as were provided in the proofs of Theorems 3.1.1 and 3.1.2 can be employed to show that

\[
N^\gamma \sup_{v, \omega \in [0,1]^2} \| \hat{E}_T(v, \omega) - E_{N,T}(v, \omega) \|_\infty = o_P(1) \tag{3.201}
\]

for any \(\gamma \in (0, 1/2)\) under the conditions (3.8) or (3.9).

**Proof of part a):** The equality

\[
\sup_{\omega \in [0,1]} |[E_{N,T}(v, \omega)]_{a,b}| = 0 \quad \text{for} \quad v \in \overline{I}_{T,a,b}(b_1, ..., b_K)
\]

and the inequality \(\varepsilon_{T,a,b}(v) > C\), which holds for some strictly positive constant \(C\), implies that for all \((a, b) \in \{1, ..., d\}^2\) and sufficiently large \(N\) and \(T\) we have

\[
P\left( \bigcup_{v \in \overline{I}_{T,a,b}(b_1, ..., b_K)} \left\{ N^\gamma \sup_{\omega \in [0,1]} |[\hat{E}_T(v, \omega)]_{a,b}| > \varepsilon_{T,a,b}(v) \right\} \right)
\leq P\left( N^\gamma \sup_{v \in \overline{I}_{T,a,b}(b_1, ..., b_K)} \sup_{\omega \in [0,1]} |[\hat{E}_T(v, \omega)]_{a,b}| > C \right)
\leq P\left( N^\gamma \sup_{v \in \overline{I}_{T,a,b}(b_1, ..., b_K)} \sup_{\omega \in [0,1]} |[\hat{E}_T(v, \omega)]_{a,b} - [E_{N,T}(v, \omega)]_{a,b}| > C \right)
\leq P\left( N^\gamma \sup_{v \in [0,1]} \sup_{\omega \in [0,1]} |[\hat{E}_T(v, \omega) - E_{N,T}(v, \omega)]|_\infty > C \right) \xrightarrow{T \to \infty} 0,
\]

where we used (3.201) in the last step. The property (3.34) follows due to the fact that the dimension \(d\) is finite and does not depend on \(T\).

**Proof of part b):** The claim (3.35) is a direct consequence of Theorems 3.1.1 part b) and 3.1.2 part b), which imply that, for each break point \(b_i\) and each component \((a, b) \in B(b_i)\), it holds

\[
P\left( \sup_{\omega \in [0,1]} N^\gamma |[\hat{E}_T(b_r, \omega)]_{a,b}| > \varepsilon_{T,a,b}(b_r) \right) \to 1
\]

under either the condition (3.8) or (3.9).

\[\square\]

### 3.4.6 Proof of Theorem 3.2.2

In order to prove parts a), b) and c) of Theorem 3.2.2, we define for \(v \in [0,1]\) the deterministic sequence

\[
\bar{E}_{N,T}(v) := \sup_{\omega \in [0,1]} |E_{N,T}(v, \omega)|_\infty \tag{3.202}
\]
and note that the quantity $\bar{E}_{N,T}(v)$ attains local maxima at the true break points $b_j$, $j \in \{1, ..., K\}$. Additionally, we define its empirical counterpart by

$$\hat{E}_T(v) := \sup_{\omega \in [0,1]} \| \bar{E}_T(v,\omega) \|_\infty.$$  \hfill (3.203)

The property (3.201) yields the identity

$$N^\gamma \sup_{v \in [0,1]} |\hat{E}_T(v) - \bar{E}_{N,T}(v)| = o_P(1),$$  \hfill (3.204)

which will be used in the following arguments.

**Proof of part a):** For a proof of (3.39), we define for $i \in \{1, ..., K\}$ the event

$$A_T(b_i) := \left\{ \left| \arg\max_{v \in [b_i-\tau,b_i+\tau]} \bar{E}_{N,T}(v) - \arg\max_{v \in [b_i-\tau,b_i+\tau]} \hat{E}_T(v) \right| \leq \frac{\tau}{2} \right\},$$

where $\tau$ denotes the constant used in the second step in the localisation procedure (see Step II). In order to show consistency of the estimator $\hat{K}$, we consider the division

$$P(\hat{K} = K) = P(\hat{K} = K \mid \bigcap_{i \in \{1, ..., K\}} A_T(b_i)) \times P\left( \bigcap_{i \in \{1, ..., K\}} A_T(b_i) \right)
+ P(\hat{K} = K \mid \left( \bigcap_{i \in \{1, ..., K\}} A_T(b_i)^C \right)) \times P\left( \left( \bigcap_{i \in \{1, ..., K\}} A_T(b_i)^C \right) \right).$$

Thus, for a proof of (3.39) it is sufficient to show that

$$P(\hat{K} = K \mid \bigcap_{i \in \{1, ..., K\}} A_T(b_i)) \to 1,$$  \hfill (3.205)

$$P\left( \bigcap_{i \in \{1, ..., K\}} A_T(b_i) \right) \to 1$$  \hfill (3.206)

as $T \to \infty$. For this purpose, we note that the proposed method (3.37) for reducing the set $\hat{B}_P$ of potential break points and claims a) and b) of Theorem 3.2.1 imply (3.205). Moreover, (3.204) assures that the probability of the set $A_T(b_i)$ converges to one for each $i \in \{1, ..., K\}$. This implies (3.206).

**Proof of part b):** In the scenario (3.9) of a shrinking relative block size, the consistency of the estimators $\hat{b}_i$ is implied by the fact that the sets $R_j$ of potential break points are included in neighbourhoods $\{[b_i T - N)/T, ..., (b_i T + N)/T\}$ of radius $N/T$ of one of the true break points $b_i$, $i \in \{1, ..., K\}$. Therefore, (3.40) follows from $N/T \to 0$.

In the case (3.8), which assumes that the relative block length $N/T$ converges to some fixed fraction $1/c$, where $c \geq 2/\min_{i=1, ..., K+1} |b_i - b_{i+1}|$, consistency of the break point estimates follows from the property (3.204) and the fact that the deterministic sequence $\bar{E}_{N,T}(v)$ at-
Proof of part c): The accuracy of the proposed method for the characterisation of the detected break points follows from part b) and (3.201).
Chapter 4

Testing for stationarity in locally stationary time series

A large amount of methods for statistical inference in stationary time series models has been developed during the course of the last decades. The consequence of this effort is a well-established asymptotic theory for various stationary time series models and a vast number and diversity of procedures for investigating statistical properties of observed datasets originating from a second order stationary stochastic process. The resulting availability of a well-equipped toolbox for conducting empirical analysis in stationary models caused the assumption of second order stationarity to emerge as the dominant paradigm in modern time series analysis. In empirical studies, this assumption is thus frequently imposed in order to justify the application of various methods, as for example parameter estimation and forecasting techniques. However, for the results and conclusions, which are reached by employing these methods, to be reliable, it is of major importance to validate the assumption of a non-changing second order structure of the underlying process. For this reason, the availability of procedures for goodness-of-fit testing is of crucial relevance from a practical point of view. This chapter is devoted to a novel approach, which was introduced in Puchstein and Preuß (2013).

Broadly speaking, methods for validating the assumption of second order stationarity can be developed in either parametric or non-parametric settings. Procedures designed for a specific parametric context include the method of Sakiyama and Taniguchi (2003), which allows to test composite hypothesis, and the approach of Changli et al. (2009), which intends to detect non-constancies of the parameters of a vector autoregressive model. All parametric approaches for validating the assumption of stationarity have in common that they critically depend on the correct model choice. However, this fact represents a significant drawback from a practical point of view since the validation of the stationarity assumption is usually carried out at an early stage of any empirical analysis, which implies that a well-founded specification of the correct parametric time series model has not yet taken place. For this reason, the class of non-parametric approaches is preferable from a fundamental point of view and a large amount of scientific literature has contributed to this area of research by proposing various techniques for validating the assumption of stationarity in the class of locally stationary models.
One of the first to suggest approaching the question of stationarity in the frequency domain were Priestley and Subba Rao (1969), who propose a method that is based on estimates of evolutionary spectra at different points in time. Another early approach, which adopts the framework of locally stationary processes, can be found in Neumann and von Sachs (1997). These authors suggest an adaptive wavelet based estimator for the time-variant spectral density. In von Sachs and Neumann (2000), a formal testing procedure for the null hypothesis of stationarity is developed, which is based on a Haar wavelet series expansion of the local periodogram estimates. For an application of this method, it is necessary to choose the correct amount of wavelet coefficients and a data driven choice for this tuning parameter is not provided. The method proposed in Dwivedi and Subba Rao (2011) follows a Portmanteau-type approach and its test statistic is based on empirical covariances, which are summed up to a specific order $m$. For the definition of a formal test, it is shown that the test statistic is approximately chi-square distributed under the null hypothesis. In Jentsch and Subba Rao (2012), this approach is extended to the multivariate setting, which is generally to be considered an advancement for many applications characterised by the presence of several time-dependent processes, which are best treated as one multivariate sample with a distinctive dependence structure. However, the application of this method requires the practitioner to specify several bandwidth and tuning parameters, for whose selection there do not exist data driven criteria. In Paparoditis (2010), a rather different approach for testing the time-dependence of the spectral density is suggested. In order to obtain an empirical measure for the deviation from stationarity, an estimator for the $L_2$ distance between the local spectral density and its global best approximation by a stationary spectral density is defined and the null hypothesis is rejected whenever this estimate surpasses a certain threshold. For estimating the local spectral density, this approach relies on periodogram estimates, whose computation require the specification of a block length $N$ and a taper function $\tau(\cdot)$. Critical values for the test statistic are obtained by applying a bootstrap method. Paparoditis (2009) follows a similar idea. In this article, a related test statistic, which also tracks the $L_2$ distance between local spectral density estimates and a global stationary approximation, is shown to have a Gaussian limit. This theoretical result allows for the construction of an asymptotic level $\alpha$ test, which does not rely on resampling methods. However, the practical implementation of this approach also requires the choice of several smoothing kernels and tuning parameters, but data driven criteria for their specifications are only partly available. The approach of Dette et al. (2011) constitutes a major advancement in this regard. In this contribution, the authors consider the deterministic quantity

$$D^2 = \min_g \int_{-\pi}^{\pi} \int_{0}^{1} (f(u, \lambda) - g(\lambda))^2 dud\lambda,$$

where the minimum is calculated over the set of all spectral densities $g$ corresponding to a stationary model. It is obvious that this quantity vanishes, if $f$ does not depend on $u$, and that it is strictly positive if the local spectral density $f(u, \lambda)$ is non-constant on a set of positive Lebesgue measure in the time direction. For the construction of a formal test for the null hypothesis of stationarity, the authors define an estimator $\hat{D}_T^2$ for the unknown quantity $D^2$, which is composed of local periodogram estimates of the functions $f(u, \lambda)$ and $f^2(u, \lambda)$,
and subsequently derive the asymptotic normality of an appropriately scaled version of this statistic. The practical as well as theoretical merit of this idea lies in the simplicity of the test statistic, which does not depend on any smoothing kernels or bandwidth for calculating the periodogram estimates. For the implementation of the proposed test, it is only required to choose a block length \( N \) specifying the range of data employed for the local estimation of the spectral density \( f \). We conclude this review of existing approaches, which is by no means exhaustive, by mentioning the approach of Preuß et al. (2012) and Preuß and Vetter (2012). The underlying idea for gauging the presence of non-stationarities in the data generating process exploited by these authors and the algorithmic nature of the resulting test closely resembles the approach, which will be presented in this chapter. For the derivation of a testing procedure which allows to answer the question of stationarity in an observed set of data originating from a locally stationary time series, Preuß et al. (2012) consider the Kolmogorov-Smirnov type distance measure

\[
\frac{1}{2\pi} \sup_{(v, \omega) \in [0,1]^2} \left| \int_0^{\omega \pi} \int_0^{v} f(u, \lambda) d\lambda d\lambda - v \int_0^{\omega \pi} \int_0^{1} f(u, \lambda) d\lambda d\lambda \right|.
\]

This quantity vanishes in the case of stationarity and is strictly positive if the second order structure of the time series changes with time. For obtaining a formal test, an empirical version of (4.1) is defined and it is shown that an appropriately scaled version of this statistic converges to the supremum of a Gaussian random process. A bootstrap procedure is shown to yield accurate estimates of critical values, which cannot easily be obtained from the limiting distribution, since it depends on the unknown spectral density of the data generating process. In order to define the test statistic mimicking the quantity (4.1), the authors divide the whole data set of length \( T \) into \( M \) blocks of length \( N \). Subsequently, an estimator for expressions of the form \( \int_0^{\omega \pi} \int_0^{v} f(u, \lambda) d\lambda d\lambda \) is computed by approximating both integrals by Riemann sums over local periodogram estimates. This implies that the practical application of the proposed method requires to choose some value for the block length \( N \), which is used for the local spectral estimation.

The remainder of this chapter is concerned with the presentation of a novel approach that contributes to the area of non-parametric testing of the stationarity assumption in the spectral domain. The new method complements the existing literature well and in fact constitutes a major advancement in comparison to existing procedures due to the fact that it solves many of the above mentioned shortcomings.

4.1 Motivation and definitions

We assume to observe a sequence of centred \( \mathbb{R}^d \)-valued stochastic processes \( \{X_{t,T}\}_{t=1,...,T} \), where, for each \( T \in \mathbb{N} \) and \( t = 1, ..., T \) the random vector \( X_{t,T} = (X_{t,T,1}, ..., X_{t,T,d})^T \) possesses a locally stationary representation of the form (2.12). As indicated in Section 2.2.2, the local
Testing for stationarity in locally stationary time series

Spectral density matrix \( f \) for this locally stationary time series model is given by

\[
f(u, \lambda) = \frac{1}{2\pi} \sum_{l,m=0}^{\infty} \Psi_l(u)\Psi_m(u)^T \exp(-i\lambda(l-m)).
\]

In this chapter, we intend to develop a testing procedure, which allows to determine whether \( f \) depends on the time parameter \( u \), or whether, in fact, it is time-invariant. In order to approach this question, we will draw a comparison between local averages of the local spectral density \( f(u, \lambda) \) and its global best approximation by a stationary function \( f(\lambda) \).

More precisely, we consider for some \( v \in [0,1] \) the functions

\[
\lambda \mapsto \int_0^v f(u, \lambda)du \quad (4.2)
\]

\[
\lambda \mapsto v \int_0^1 f(u, \lambda)du. \quad (4.3)
\]

and note that in the case of no time-dependence of the spectral density, the functions (4.2) and (4.3) are identical for each \( v \in [0,1] \). In order to obtain a component wise measure for the discrepancy of the average spectral density on the interval \([0,v]\) from its global best constant approximation, we define for \( \omega \in [0,1] \) the matrix

\[
D(v, \omega) := \frac{v}{2\pi} \left( \int_0^\omega \int_0^v f(u, \lambda)du d\lambda - v \int_0^1 \int_0^1 f(u, \lambda)du d\lambda \right) \quad (4.4)
\]

and consider the quantity

\[
D(v) := \sup_{\omega \in [0,1]} \|D(v, \omega)\|_{\infty},
\]

which is strictly positive, if the functions (4.2) and (4.3) differ in at least one component on a set of positive Lebesgue measure. For a global measure for the time-dependence of the spectral density matrix \( f \) it is therefore intuitive to consider the supremum of the function \( D(v), v \in [0,1] \). For this reason, we turn our attention to a further investigation of the quantity

\[
D := \sup_{(v, \omega) \in [0,1]^2} \|D(v, \omega)\|_{\infty} \quad (4.5)
\]

for the development of a testing procedure. It follows from the definition of the matrices \( D(v, \omega), (v, \omega) \in [0,1]^2 \) that \( D \) vanishes if the null hypothesis

\[
H_0 : \ f(u, \lambda) \text{ is independent of } u \quad (4.6)
\]

holds, while \( D \) is strictly positive in the case of the alternative

\[
H_1 : \ f(u, \lambda) \text{ is non-constant in } u\text{-direction on a set of positive Lebesgue measure} \quad (4.7)
\]

In order to derive a formal level \( \alpha \) test of the null hypothesis (4.6), we develop an estimator for the quantity \( D \) and reject (4.6), if this estimate surpasses an appropriately chosen threshold.
Hence, we proceed as follows: In the next section, we introduce an estimator $\hat{D}_T(v, \omega)$ for the matrix $D(v, \omega)$, $(v, \omega) \in [0,1]^2$, define an empirical process $\{\hat{D}_T(v, \omega)\}_{(v, \omega) \in [0,1]^2}$, which tracks the behaviour of the deterministic collection $\{D(v, \omega)\}_{(v, \omega) \in [0,1]^2}$ of matrices, and choose the supremum $\hat{D}_T$ of this process as the test statistic. Subsequently, we show that an appropriately scaled version of $\hat{D}_T$ converges weakly to the supremum of a centred Gaussian process, if the null hypothesis holds. Furthermore, we explain how critical values for $\hat{D}_T$ can be estimated by applying the AR($\infty$) bootstrap, which has already been introduced in Section 3.1.4.

4.2 Tracking non-constancy in the time direction

We introduce an empirical counterpart for the matrix $D(v, \omega)$, $(v, \omega) \in [0,1]^2$ defined in (4.4) by constructing an estimator for integrals of the form

$$\int_0^v f(u, \lambda) du$$

for some $v \in [0,1]$.

In order to estimate expressions of this kind, we take the number $2 \lfloor v T / 2 \rfloor$ (i.e. the largest even integer not exceeding the value $v T$) and calculate the periodogram

$$I_{2 \lfloor v T / 2 \rfloor}(\lambda) := \frac{1}{4\pi \lfloor v T / 2 \rfloor} \sum_{r,s=0}^{2 \lfloor v T / 2 \rfloor-1} X_1 s T X_1^T s T \exp(-i \lambda (s - r))$$

on the basis of the first $2 \lfloor v T / 2 \rfloor$ data points. As will be shown in the proof of Theorem 4.3.2 below, this estimator is asymptotically unbiased for the integral $\frac{1}{v} \int_0^v f(u, \lambda) du$ and we thus obtain a reasonable estimator for $D(v, \omega)$, $(v, \omega) \in [0,1]^2$, by

$$\hat{D}_T(v, \omega) := v \left( \frac{1}{T} \sum_{k=1}^{\lfloor \omega [v T / 2] \rfloor} I_{2 \lfloor v T / 2 \rfloor}(\lambda_{k,2[v T / 2]}) - \frac{v}{T} \sum_{k=1}^{\lfloor \omega / T \rfloor} I_{2 [T / 2]}(\lambda_{k,T}) \right),$$

(4.8)

where for $n \in \mathbb{N}$, $\lambda_{k,n} = 2\pi k / n$ ($k = 1,...,n$) denote the Fourier frequencies to the base $n$. Note that the estimator for $D(v, \omega)$ is constructed by replacing the integral in $\lambda$-direction by a Riemann sum over the respective Fourier frequencies and substituting the estimators

$$\frac{2 \lfloor v T / 2 \rfloor}{T} I_{2 \lfloor v T / 2 \rfloor}(\lambda) \quad \text{and} \quad I_{2 [T / 2]}(\lambda)$$

for the integrals

$$\int_0^v f(u, \lambda) du \quad \text{and} \quad \int_0^1 f(u, \lambda) du$$
respectively. Having found a sensible estimator for matrices \( \{ \mathbf{D}(v, \omega) \}_{(v, \omega) \in [0,1]^2} \), an empirical analog to the deterministic quantity \( D \) introduced in (4.5) is defined by

\[
\hat{D}_T := \sup_{(v, \omega) \in [0,1]^2} \| \hat{D}_T(v, \omega) \|_\infty. \tag{4.9}
\]

In the following section, we will derive crucial properties of the asymptotic distribution of the process \( \{ \hat{D}_T(v, \omega) \}_{(v, \omega) \in [0,1]^2} \) and the test statistic \( \hat{D}_T \) under the null hypothesis of a constant spectral density matrix and the alternative.

### 4.3 Weak convergence of the empirical process

We now turn to the investigation of the asymptotic distribution of the test statistic \( \hat{D}_T \) defined in (4.9). In order to formally establish weak convergence of the empirical process \( \{ \hat{D}_T(v, \omega) \}_{(v, \omega) \in [0,1]^2} \) to a centred Gaussian process, we first state the following critical assumptions concerning the sequence \( \{ \mathbf{X}_{t,T} \}_{t=1}^{1, \ldots, T} \):

**Assumption 4.3.1**

The sequence \( \{ \mathbf{X}_{t,T} \}_{t=1}^{1, \ldots, T} \) is a locally stationary time series in the sense of Definition 2.2.2 and the following conditions hold:

i) In the representation (2.12) of the vector \( \mathbf{X}_{t,T}, \ t \in \{1, \ldots, T\}, \ \{ \mathbf{Z}_t \}_{t \in \mathbb{Z}} \) is a sequence of independent \( \mathcal{N}(0, \mathbf{I}_d) \) distributed random vectors.

ii) The approximating functions \( \Psi_l : [0, 1] \rightarrow \mathbb{R}^{d \times d}, \ l \in \mathbb{N} \) satisfying (2.13), fulfill the conditions

\[
\sum_{l=0}^{\infty} \sup_{u \in [0,1]} \| \Psi_l(u) \|_\infty |l| < \infty, \tag{4.10}
\]

\[
\sum_{l=0}^{\infty} \sup_{u \in [0,1]} \| \Psi_l'(u) \|_\infty |l| < \infty, \tag{4.11}
\]

\[
\sum_{l=0}^{\infty} \sup_{u \in [0,1]} \| \Psi_l''(u) \|_\infty < \infty. \tag{4.12}
\]

The above assumptions are rather standard in the framework of locally stationary time series models [see Dahlhaus (2000), Dahlhaus and Polonik (2009) or Paparoditis (2009)].

The following theorem specifies the asymptotic behaviour of the empirical process \( \{ \sqrt{T} \hat{D}_T(v, \omega) \}_{(v, \omega) \in [0,1]^2} \) both under the null hypothesis and the alternative.
4.4 Bootstrapping the test statistic

Theorem 4.3.2 (Weak convergence of the empirical process \( \{\sqrt{T} \hat{D}_T(v, \omega)\}_{(v, \omega) \in [0,1]^2} \))

Suppose Assumption 4.3.1 is fulfilled. Then the following statements hold:

a) Under the null hypothesis (4.6), the empirical process \( \{\sqrt{T} \hat{D}_T(v, \omega)\}_{(v, \omega) \in [0,1]^2} \) converges weakly to a centred Gaussian process \( \{G(v, \omega)\}_{(v, \omega) \in [0,1]^2}, i.e.

\[
\{\sqrt{T} \hat{D}_T(v, \omega)\}_{(v, \omega) \in [0,1]^2} \Rightarrow \{G(v, \omega)\}_{(v, \omega) \in [0,1]^2},
\]

where the covariance kernel of \( \{G(v, \omega)\}_{(v, \omega) \in [0,1]^2} \) is given by

\[
\text{Cov}(G(v_1, \omega_1)|a_1, b_1, G(v_2, \omega_2)|a_2, b_2) = \frac{1}{2\pi} v_1 v_2 (\min(v_1, v_2) - v_1 v_2)
\times \int_0^{\min(\omega_1, \omega_2)} f_{a_1, b_2}(\lambda) f_{b_1, a_2}(-\lambda) d\lambda \quad (4.13)
\]

for \((v_1, \omega_1), (v_2, \omega_2) \in [0, 1]^2\) and \((a_1, b_1), (a_2, b_2) \in \{1, \ldots, d\}^2\).

b) Under the alternative (4.7), there exists a constant \(C > 0\), such that for all \((a, b) \in \{1, \ldots, d\}^2\)

\[
\lim_{T \to \infty} P\left( \sup_{(v, \omega) \in [0,1]^2} ||\hat{D}_T(v, \omega)|a,b| > C \right) = 1 \quad \text{if} \quad \sup_{(v, \omega) \in [0,1]^2} ||D(v, \omega)|a,b| > 0
\]

\[
\sup_{(v, \omega) \in [0,1]^2} ||\hat{D}_T(v, \omega)|a,b| = O_p(T^{-1/2}) \quad \text{if} \quad \sup_{(v, \omega) \in [0,1]^2} ||D(v, \omega)|a,b| = 0.
\]

The above theorem shows that under the null hypothesis, the empirical process \( \{\sqrt{T} \hat{D}_T(v, \omega)\}_{(v, \omega) \in [0,1]^2} \) converges to a Gaussian process \( \{G(v, \omega)\}_{(v, \omega) \in [0,1]^2} \). By the continuous mapping theorem, it immediately follows that

\[
\sqrt{T} \hat{D}_T \Rightarrow \sup_{(v, \omega) \in [0,1]^2} ||G(v, \omega)||_{\infty}. \quad (4.14)
\]

The property (4.14) determines the asymptotic distribution of the rather complicated estimator \( \hat{D}_T \). Nevertheless, the fact that the covariance kernel of the process \( \{G(v, \omega)\}_{(v, \omega) \in [0,1]^2} \) given in (4.13) crucially depends on the unknown spectral density matrix \( f \) of the underlying data generating process means that we cannot derive a formal testing procedure from the property (4.14) [see Dahlhaus (2009) or Preuß et al. (2012) for similar situations].

4.4 Bootstrapping the test statistic

In this section, we explain how the AR(\(\infty\)) bootstrap, which has already been employed in Section 3.1.4 to derive an asymptotic test for the presence of structural breaks, can be applied in the present context as well. For this purpose, we assume that the \(\mathbb{R}^d\)-valued process \( \{X_t\}_{t \in \mathbb{Z}} \) with spectral density function \( g(\lambda) = \int_0^1 f(u, \lambda) du \) satisfies Assumption 3.1.3 and generate bootstrap statistics \( \hat{D}_T \) in an analog manner as replications \( \hat{E}_T^* \) of the statistic \( E_T \) were generated in Algorithm 3.1.4.
The following result formally shows that the replicates $\hat{D}_T^*$ have similar distributional properties as the test statistic $\hat{D}_T$ under the null hypothesis.

**Theorem 4.4.1** *(Weak convergence of the empirical process)*  
Suppose that Assumptions 4.3.1 and 3.1.3 holds. Furthermore, assume that conditions i), ii) and iii) of Theorem 3.1.6 are fulfilled. Then it follows that conditionally on the sample $\{X_{t,T}\}_{t=1,...,T}$ 
$$\sqrt{T} \hat{D}_T^* \Rightarrow \sup_{(v,\omega)\in[0,1]^2} \|G(v,\omega)\|_{\infty}$$  
(4.16)

as $T \to \infty$ almost surely, where $G_H(v,\omega)$ denotes a centred Gaussian process with covariance kernel 
$$\text{Cov}(G_H(v_1,\omega_1), G_H(v_2,\omega_2)) = \frac{1}{2\pi} v_1 v_2 (\min(v_1, v_2) - v_1 v_2) \times \int_0^{\min(\omega_1,\omega_2)} \left( \int_0^1 f_{a_1 b_1}(u,\lambda) du \right) \left( \int_0^1 f_{a_2 b_2}(u,-\lambda) du \right) d\lambda. \quad (4.15)$$

We note that the covariance kernel of the process $\{G_H(v,\omega)\}_{(v,\omega)\in[0,1]^2}$ is similar to the one which determines the limiting process in Theorem 4.3.2 a). The only difference is that the spectral density $f$ is replaced by the time averaged version $\int_0^1 f(u,\lambda) du$ corresponding to the best approximation of $f(u,\lambda)$ by a stationary spectral density. In fact, this implies that under the null hypothesis (4.6) of a non-changing spectral density matrix the covariance kernels (4.13) and (4.15) are identical. Summarising, we have:

1. The continuous mapping theorem [see Theorem 2.3.9] implies that under the null hypothesis (4.6), the bootstrap statistic $\hat{D}_T^*$ conditionally on the data $\{X_{t,T}\}_{t=1,...,T}$ has the same limiting distribution as the test statistic $\hat{D}_T$, i.e.
   $$\sqrt{T} \hat{D}_T^* \Rightarrow \sup_{(v,\omega)\in[0,1]^2} \|G(v,\omega)\|_{\infty}$$  
   (4.16)

   conditionally on $\{X_{t,T}\}_{t=1,...,T}$ almost surely, where the process $\{G(v,\omega)\}_{(v,\omega)\in[0,1]^2}$ was introduced in Theorem 4.3.2.

2. Under the alternative (4.7), the bootstrap replicates $\sqrt{T} \hat{D}_T^*$ conditionally on the data $\{X_{t,T}\}_{t=1,...,T}$ converge weakly to the supremum of a Gaussian process with covariance kernel (4.15), i.e.
   $$\sqrt{T} \hat{D}_T^* \Rightarrow \sup_{(v,\omega)\in[0,1]^2} \|G_H(v,\omega)\|_{\infty}$$  
   (4.17)

   conditionally on $\{X_{t,T}\}_{t=1,...,T}$ almost surely.

The above insight shows that it is sensible to choose the following procedure for testing the hypothesis (4.6):
Algorithm 4.4.2 (Test for stationarity)

1) Choose some level $\alpha \in (0, 1)$ of significance and an integer $B$ for the amount of replicates, which are employed in the estimation of the $(1-\alpha)$-quantile $q_{1-\alpha}^{\hat{D}_T}$ of the test statistic $\hat{D}_T$ under the null hypothesis.

2) Calculate the test statistic $\hat{D}_T$.

3) Choose the order $p$ for the fitted autoregressive model and determine a consistent estimator $(\hat{a}_{1,p}, \ldots, \hat{a}_{p,p})$ of the minimiser in (3.19).

4) Generate $B$ replicates $\{\hat{D}_{T,1}^*, \ldots, \hat{D}_{T,B}^*\}$ and estimate the quantile $q_{1-\alpha}^{\hat{D}_T}$ by

$$\hat{q}_{1-\alpha}^{\hat{D}_T} := \hat{D}_{T,(\lfloor(1-\alpha)\rfloor B)},$$

where $\hat{D}_{T,(1)}, \ldots, \hat{D}_{T,(B)}$ denotes the ordered bootstrap sample.

5) Reject the null hypothesis (4.6) of stationarity, if

$$\hat{D}_T > \hat{q}_{1-\alpha}^{\hat{D}_T}.$$ 

Note that the property (4.16) implies that the estimator (4.18) is consistent for the $(1-\alpha)$-quantile of the test statistic $\hat{D}_T$ under the null hypothesis. Algorithm 4.4.2 therefore defines a testing procedure with asymptotic level $\alpha$. Furthermore, it is easy to see that this procedure is consistent: Theorem 4.3.2 b) implies that, under the alternative (4.7), the test statistic $\hat{D}_T$ surpasses some positive constant $C > 0$, whereas (4.17) implies that the estimate $\hat{q}_{1-\alpha}^{\hat{D}_T}$ for the critical value vanishes with probability converging to one.

4.5 Practical application and finite sample properties

In this section, we give an overview of matters relating to the applicability of the described method for testing for stationarity in locally stationary models. Therefore, we first comment on issues relating to the implementation of the bootstrap test and then proceed by presenting data on its finite sample properties.

4.5.1 Recommendations for implementation

As was mentioned above, the main advantage of the test (4.19) in comparison to all other non-parametric procedures proposed so far is founded in the simplicity of the test statistic $\hat{D}_T$, which does not depend on any regularising parameters. In fact, the implementation of Algorithm 4.4.2 into well-working code is straightforward and for the obtained results presented in the next paragraph we choose the order of the $AR$ model, which is fitted for estimating the quantile $q_{1-\alpha}^{\hat{D}_T}$ according to (3.42). For the Bootstrap procedure, we fit an autoregressive model by calculating the Yule-Walker estimators for the minimiser in (3.19).
Note that due to Hannan and Kavalieris (1986), where it is shown that the Yule-Walker estimators and the least squares estimators fulfil the condition (3.24), this implementation is in accordance with the asymptotic theory of the previous sections. Finally, we remark that for the calculation of the test statistic $\hat{D}_T$ and the replicates $\hat{D}_T^*$ we only consider data sets of size $T = 2^i$ for some $i \in \mathbb{N}$ and values $v \in \left\{ \frac{T}{2^i} | 1 \leq i \leq \lfloor \log_2(T/2) \rfloor \right\}$. This implies that for the computation of the test statistic $\hat{D}_T$ and its bootstrap replicates $\hat{D}_T^*$ the periodogram estimates $I_{2^i v T/2}$ and $I_{2^i T/2}$ are based on blocks of length equal to a power of 2, which allows for an efficient application of the fft-algorithm and thus results in a significant reduction of computational runtime.

### 4.5.2 Simulation study

In this section, we investigate the finite sample properties of the test (4.19) by means of a simulation study. First, we analyse how well the proposed procedure approximates the nominal level under various models having a constant second order structure. We continue by considering several non-stationary models and estimate the obtained power of the new test. Furthermore, we compare the new approach to existing procedures and demonstrate that it performs well in comparison to other approaches. In the following presentation, all reported empirical quantiles are based on 1000 simulation runs with each run containing 200 bootstrap replications.

**Size of the test**

We first examine how well the nominal level is approximated under the null hypothesis. For this reason, we start by exploring the performance of the test (4.19) in the univariate MA(1) and AR(1) models

\[
X_{t,T} = \theta Z_{t-1} + Z_t, \quad (4.20)
\]

\[
X_{t,T} = \phi X_{t-1,T} + Z_t, \quad (4.21)
\]

where the innovations $Z_t, t \in \mathbb{Z}$, denote independent standard normally distributed random variables. We proceed with a size investigation for the bivariate MA(1) and AR(1) models

\[
X_{t,T} = \begin{pmatrix} \theta & 0.2 \\ 0.2 & \theta \end{pmatrix} Z_{t-1} + Z_t, \quad (4.22)
\]

\[
X_{t,T} = \begin{pmatrix} \phi & 0.2 \\ 0.2 & \phi \end{pmatrix} X_{t-1,T} + Z_t, \quad (4.23)
\]

where the $Z_t, t \in \mathbb{Z}$, denote independent bivariate normally distributed random vectors with unite covariance matrix. The estimated rejection probabilities of the test (4.19) for the models (4.20) – (4.23) under different configurations of the parameters $\theta$ and $\phi$ and for various sample sizes $T$ are presented in Tables 4.1 and 4.2. For the univariate, as well as for the multivariate models, the new test seems to be conservative for small sample sizes $T$, ...
while the approximation improves with increasing sample size.

<table>
<thead>
<tr>
<th></th>
<th>$\text{H}_0$: Model (4.20)</th>
<th>$\text{H}_0$: Model (4.21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta = 0.5$</td>
<td>0.035</td>
<td>0.041</td>
</tr>
<tr>
<td>$\theta = 0$</td>
<td>0.019</td>
<td>0.009</td>
</tr>
<tr>
<td>$\theta = -0.5$</td>
<td>0.003</td>
<td>0.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\text{H}_0$: Model (4.22)</th>
<th>$\text{H}_0$: Model (4.23)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta = 0.5$</td>
<td>0.047</td>
<td>0.043</td>
</tr>
<tr>
<td>$\theta = -0.5$</td>
<td>0.038</td>
<td>0.031</td>
</tr>
</tbody>
</table>

Table 4.1: Empirical rejection frequencies of the test (4.19) for the models (4.20) and (4.21) for different choices for the parameters $\theta$, $\phi$ and the sample size $T$ at nominal levels $\alpha \in \{0.05, 0.1\}$.

<table>
<thead>
<tr>
<th></th>
<th>$\text{H}_0$: Model (4.22)</th>
<th>$\text{H}_0$: Model (4.23)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta = 0.5$</td>
<td>0.047</td>
<td>0.043</td>
</tr>
<tr>
<td>$\theta = -0.5$</td>
<td>0.038</td>
<td>0.031</td>
</tr>
</tbody>
</table>

Table 4.2: Empirical rejection frequencies of the test (4.19) for the models (4.22) and (4.23) for different choices for the parameters $\theta$, $\phi$ and the sample size $T$ at nominal levels $\alpha \in \{0.05, 0.1\}$.

### Power of the test

We proceed with our investigation of finite sample properties by assessing the power of the test (4.19) in various non-stationary time series models. For this reason, we consider the univariate and multivariate models

$$X_{t,T} = (1 + \frac{t}{T})Z_t,$$

$$X_{t,T} = -0.9\sqrt{\frac{T}{T}}X_{t-1,T} + Z_t,$$

$$X_{t,T} = \begin{cases} 0.5X_{t-1,T} + Z_t & \text{if } 1 \leq t \leq \frac{T}{2} \\ -0.5X_{t-1,T} + Z_t & \text{if } \frac{T}{2} + 1 \leq t \leq T, \end{cases}$$

$$X_{t,T} = \begin{pmatrix} \sigma_t \\ 0.1 \\ 0.1 \sigma_t \end{pmatrix} Z_t, \quad \sigma_t = \frac{1}{10} + \frac{1}{2} \cos(\frac{\pi t}{2T}),$$

$$X_{t,T} = \begin{pmatrix} \phi_t \\ 0.1 \\ 0.1 \phi_t \end{pmatrix} X_{t-1,T} + Z_t, \quad \phi_t = \frac{9}{10} \cos(\frac{t}{T} \pi),$$
where in each model \( \{ Z_t \}_{t=1,\ldots,T} \) and \( \{ Z_t^{(2)} \}_{t=1,\ldots,T} \) denote sequences of univariate and bivariate standard normally distributed random variables. Note that, except for the model (4.26), the above processes fit into the class of locally stationary processes. The obtained estimates for the power of the test (4.19) in models (4.24)–(4.28) are contained in Table 4.3. The results show that for all considered alternatives, the rejection frequencies are significantly higher than the nominal level and increase with the sample size, which could be expected on the basis of the theoretical results. For a comparison of the relative merit of the new approach with existing procedures, we summarise the results of the extensive simulation study, which was presented in Preuß et al. (2012) and compares these authors’ new procedure with those proposed in Paparoditis (2010), Dwivedi and Subba Rao (2011) and Dette et al. (2011). Table 4.4 contains the essence of this empirical analysis. Each cell of Table 4.4 contains the lowest and highest power estimate of each of the procedures under consideration, where each approach was evaluated for several choices of the respective regularising parameters. The intervals reveal the lowest and the highest rejection frequencies obtained in models (4.24)–(4.26) at nominal level \( \alpha = 0.1 \) for any configuration [note that the wide range of the intervals is due to several tests included in the investigation which all depend on several regularisation parameters]. These results show that in model (4.24), the new method outperforms all procedures contained in the comparison group. This observation is independent of the specific choices of tuning parameters for the competing methods. In model (4.25), the power of the new test (4.19) falls only slightly short of the best attainable power in any of the competing procedures for any choice of regularising parameters. For model (4.26), some of the methods, which were compared in Preuß et al. (2012), outperform the test (4.19) but only if the regularising parameters are optimally chosen. As a conclusion, we emphasise that the wide range of the intervals in table (4.4) highlight one important advantage of the new procedure, namely that it is independent of any tuning parameters. By using the new approach, the practitioner is not in danger of obtaining artificial results by varying the regularisation parameters, and the test (4.19) therefore presents a significant improvement in the context of testing for stationarity.
### Table 4.3: Empirical rejection frequencies of the test (4.19) for the models (4.24) – (4.28), sample sizes $T \in \{64, 128, 256\}$ and nominal levels $\alpha \in \{0.05, 0.1\}$.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>0.371</td>
<td>0.512</td>
<td>0.093</td>
<td>0.096</td>
<td>0.207</td>
</tr>
<tr>
<td>128</td>
<td>0.667</td>
<td>0.776</td>
<td>0.301</td>
<td>0.182</td>
<td>0.356</td>
</tr>
<tr>
<td>256</td>
<td>0.951</td>
<td>0.981</td>
<td>0.751</td>
<td>0.416</td>
<td>0.738</td>
</tr>
</tbody>
</table>

### Table 4.4: Empirical rejection frequencies for the tests of Paparoditis (2010), Dwivedi and Subba Rao (2011), Dette et al. (2011) and Preuß et al. (2012) for the models (4.24) – (4.28), sample sizes $T \in \{64, 128, 256\}$ and nominal level $\alpha = 0.1$.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>[0.126, 0.444]</td>
<td>[0.100, 0.328]</td>
<td>[0.056, 0.344]</td>
</tr>
<tr>
<td>128</td>
<td>[0.16, 0.772]</td>
<td>[0.114, 0.578]</td>
<td>[0.116, 0.566]</td>
</tr>
<tr>
<td>256</td>
<td>[0.226, 0.978]</td>
<td>[0.210, 0.868]</td>
<td>[0.176, 0.922]</td>
</tr>
</tbody>
</table>
4.6 Proofs

In this section, we formally derive the theoretical results of this chapter by presenting the technical proofs of Theorems 4.3.2 and 4.4.1.

4.6.1 Proof of Theorem 4.3.2

Without loss of generality, we restrict ourselves to the case \( d = 1 \) and note that the more general case is treated analogously by employing linearity arguments and using the independence of the components of the sequence \( \{ Z_t \}_{t \in \mathbb{Z}} \) of random vectors. Throughout this section, \( C \) denotes some universal positive constant that may vary from line to line. Furthermore, \( \varepsilon' > 0 \) denotes a positive constant, which can be arbitrarily small.

Proof of part a): We employ Theorem 2.3.12, which implies that the assertion is a consequence of the following two lemmas:

**Lemma 4.6.1 (Convergence of the finite dimensional projections)**

The finite dimensional projections of the empirical process \( \{ \sqrt{T} \hat{D}_T(v, \omega) \}_{(v, \omega) \in [0,1]^2} \) converge weakly to the finite dimensional projections of the process \( \{ G(v, \omega) \}_{(v, \omega) \in [0,1]^2} \), i.e. for every \( k \geq 1 \), and any \( (v_1, \omega_1), \ldots, (v_k, \omega_k) \in [0,1]^2 \) it holds

\[
\sqrt{T} (\hat{D}_T(v_1, \omega_1), \ldots, \hat{D}_T(v_k, \omega_k)) \Rightarrow (G(v_1, \omega_1), \ldots, G(v_k, \omega_k)).
\] (4.29)

**Lemma 4.6.2 (Asymptotic stochastic equicontinuity)**

For any fixed constant \( \beta \in (0,1/4) \), the empirical process \( \{ \sqrt{T} \hat{D}_T(v, \omega) \}_{(v, \omega) \in [0,1]^2} \) is asymptotically stochastically equicontinuous with respect to the semi metric

\[
d_\beta ((v_1, \omega_1), (v_2, \omega_2)) = (|v_1 - v_2| + |\omega_1 - \omega_2|)^{\beta/2},
\] (4.30)
i.e. for every \( \eta, \varepsilon > 0 \) there exists some \( \delta > 0 \) such that

\[
limit_{T \to \infty} P \left( \sup_{(v_1, \omega_1), (v_2, \omega_2) \in [0,1]^2: d_\beta ((v_1, \omega_1), (v_2, \omega_2)) < \delta} \sqrt{T} |\hat{D}_T(v_1, \omega_1) - \hat{D}_T(v_2, \omega_2)| > \eta \right) < \varepsilon.
\] (4.31)

Proof of Lemma 4.6.1:

For a proof of Lemma 4.6.1, let \( k \geq 1 \) and \( (v_1, \omega_1), \ldots, (v_k, \omega_k) \in [0,1]^2 \) be arbitrary but fixed. In order to prove (4.29), we employ Theorem 2.3.4, i.e. we will show that all cumulants of the random vector \( (\sqrt{T} \hat{D}_T(v_i, \omega_i))_{i=1, \ldots, k} \) converge to the respective cumulants of the vector \( (G(v_i, \omega_i))_{i=1, \ldots, k} \). The demonstrations of Example 2.3.1 and the Gaussianity of the vector \( (G(v_i, \omega_i))_{i=1, \ldots, k} \) imply that we achieve this goal by showing that the following three claims hold under the null hypothesis:
(1) The process \( \{ \sqrt{T} \hat{D}_T(v, \omega) \} \) is asymptotically centred. More precisely, for all \((v, \omega) \in [0, 1]^2\) it holds
\[
E(\sqrt{T} \hat{D}_T(v, \omega)) = O\left( \frac{1}{\sqrt{T}} \right). \tag{4.32}
\]

(2) For all \((v_1, \omega_1), (v_2, \omega_2) \in [0, 1]^2\), the asymptotic covariance of \(\sqrt{T} \hat{D}_T(v_1, \omega_1)\) and \(\sqrt{T} \hat{D}_T(v_2, \omega_2)\) is given by (4.13). More precisely, for all \(\alpha \in (0, 1/4)\) we have
\[
\text{Cov}(\sqrt{T} \hat{D}_T(v_1, \omega_1), \sqrt{T} \hat{D}_T(v_2, \omega_2)) = \frac{1}{2\pi} v_1 v_2 (\min(v_1, v_2) - v_1 v_2) \times \int_0^{\min(\omega_1, \omega_2)} f^2(\lambda) d\lambda + O\left( \frac{1}{T^\alpha} \right). \tag{4.33}
\]

(3) The cumulants of orders \(l \geq 3\) vanish. More precisely, for all \(l \geq 3\) and \((v_i, \omega_i) \in [0, 1]^2\), \(i \in \{1, ..., l\}\) we have
\[
\text{cum}(\sqrt{T} \hat{D}_T(v_1, \omega_1), ..., \sqrt{T} \hat{D}_T(v_l, \omega_l)) \leq 2^l (2l)! C^l \frac{\log(T)^l}{T^{l/2-1}} = o(1).
\]

Proof of part (1): In order to facilitate the arguments needed in the proof of part b) of Theorem 4.3.2, we show that the slightly more general statement
\[
E\left( \sqrt{T} (\hat{D}_T(v, \omega) - D(v, \omega)) \right) = O\left( \frac{1}{\sqrt{T}} \right) \tag{4.34}
\]
holds for all \((v, \omega) \in [0, 1]^2\) both under the null hypothesis and the alternative, where the deterministic quantity \(D(v, \omega)\) was defined in (4.4) and vanishes under the null hypothesis. Obviously, (4.34) follows from the property
\[
E\left( \sqrt{T} \sum_{k=1}^{[vT/2]} I_{2|vT/2|}(\lambda_k, 2|vT/2|) - \frac{v}{2\pi} \int_0^{\infty} \int_0^v f(u, \lambda) dud\lambda \right) = O\left( \frac{1}{\sqrt{T}} \right), \tag{4.35}
\]
and we thus focus on a proof of (4.35). For this purpose, we define
\[
\psi_{l,T,t} := \Psi_{l,T,t} \quad \text{and} \quad \psi_l(t) := \Psi_l(t) \quad \text{for} \quad l \leq 2.
\]
and employ the locally stationary representation (2.12) of the time series \( \{ X_{t,T} \}_{t=1,...,T} \) to obtain

\[
E \left( \frac{v}{T} \sum_{k=1}^{[\omega vT/2]} I_{2[vT/2]}(\lambda_{k,2[vT/2]}) \right) 
= E \left( \frac{v}{T} \sum_{k=1}^{[\omega vT/2]} \frac{1}{4\pi vT/2} \sum_{p,q=0}^{2[vT/2]-1} \psi_{1+p,T}\psi_{1+q,T} \exp(-i\lambda_{k,2[vT/2]}(p-q)) \right) 
= v \sum_{k=1}^{[\omega vT/2]} \frac{1}{4\pi vT/2} \sum_{l,m=0}^{\infty} \sum_{p,q=0}^{2[vT/2]-1} \psi_{1+p,T,l}\psi_{1+q,T,m} E(Z_{1+p-T}Z_{1+q-m}) \exp(-i\lambda_{k,2[vT/2]}(p-q)).
\]

Now, we make use of (2.13) to approximate the above term by

\[
\frac{v}{T} \sum_{k=1}^{[\omega vT/2]} \frac{1}{4\pi vT/2} \sum_{l,m=0}^{\infty} \sum_{p,q=0}^{2[vT/2]-1} \psi_{l}(1+pT)\psi_{m}(1+qT) E(Z_{1+p-T}Z_{1+q-m}) \times \exp(-i\lambda_{k,2[vT/2]}(p-q)) + O(\frac{1}{T}),
\]

where we used that the time-varying coefficients \( \psi_{t,T,l} \) can be replaced by the approximating functions \( \psi_{l}(t/T) \) by making an error that is at most of order \( O(1/T) \). The condition

\[
E(Z_{i}Z_{j}) = \delta_{ij} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}
\]

yields that the equality

\[ q = p - l + m \]

has to hold for all non-vanishing summands in (4.36). Furthermore, from \( q \in \{1, ..., 2\lfloor vT/2 \rfloor - 1\} \) it follows that the inequality

\[ 0 \leq p - l + m \leq 2\lfloor vT/2 \rfloor - 1 \]

has to hold for all summands which have a contribution. Thus, we get that (4.36) equals

\[
\frac{v}{T} \sum_{k=1}^{[\omega vT/2]} \frac{1}{2\pi} \sum_{l,m=0}^{\infty} \exp(-i\lambda_{k,2[vT/2]}(l-m)) \frac{1}{2[vT/2]} \sum_{p=0}^{2[vT/2]-1} \psi_{l}(1+pT) 
\times \psi_{m}(\frac{1+p-l+m}{T}) + O(\frac{1}{T}).
\]"
The summability condition (4.10) for the linear coefficients \( \psi_l(u), l \in \mathbb{N}, u \in [0, 1] \), yields

\[
\sum_{l,m=0}^{\infty} \left| \sum_{p=0}^{2[vT/2]-1} \psi_l\left(\frac{1+p}{T}\right)\psi_m\left(\frac{1+p-l+m}{T}\right) - \sum_{p=0}^{2[vT/2]-1} \psi_l\left(\frac{1+p}{T}\right)\psi_m\left(\frac{1+p-l+m}{T}\right) \right|
\]

\[
\leq \sum_{l,m=0}^{\infty} |l-m| \sup_{u \in [0,1]} |\psi_l(u)| \sup_{u \in [0,1]} |\psi_m(u)| \leq C, \tag{4.40}
\]

where we used the fact that dropping the condition (4.38) adds at most \(|l-m|\) terms in the summation over \( p \), each of which is at most of size \( \sup_{u \in [0,1]} |\psi_l(u)| \sup_{u \in [0,1]} |\psi_m(u)| \). The bound (4.40) implies that the restriction in the summation over \( p \) in (4.39) can be dropped by including an error term of order \( O(1/T) \). Hence, we obtain that (4.39) is equal to

\[
\frac{v}{T} \sum_{k=1}^{[\omega vT/2]} \frac{1}{2\pi} \sum_{l,m=0}^{\infty} \exp(-i\lambda_k 2[vT/2](l-m)) \frac{1}{2[vT/2]} \sum_{p=0}^{2[vT/2]-1} \left[ \psi_l\left(\frac{1+p}{T}\right)\psi_m\left(\frac{1+p}{T}\right) \right.
\]
\[
+ \left. \psi_l\left(\frac{1+p}{T}\right)\left(\psi_m\left(\frac{1+p-l+m}{T}\right) - \psi_m\left(\frac{1+p}{T}\right)\right) \right] + O\left(\frac{1}{T}\right)
\]

\[
= E_{T,1} + E_{T,2} + O\left(\frac{1}{T}\right),
\]

where we define the quantities \( E_{T,1} \) and \( E_{T,2} \) by

\[
E_{T,1} := \frac{v}{T} \sum_{k=1}^{[\omega vT/2]} \frac{1}{2\pi} \sum_{l,m=0}^{\infty} \exp(-i\lambda_k 2[vT/2](l-m)) \frac{1}{2[vT/2]} \sum_{p=0}^{2[vT/2]-1} \psi_l\left(\frac{1+p}{T}\right)\psi_m\left(\frac{1+p}{T}\right)
\]

\[
E_{T,2} := \frac{v}{T} \sum_{k=1}^{[\omega vT/2]} \frac{1}{2\pi} \sum_{l,m=0}^{\infty} \exp(-i\lambda_k 2[vT/2](l-m)) \frac{1}{2[vT/2]} \sum_{p=0}^{2[vT/2]-1} \psi_l\left(\frac{1+p}{T}\right)
\]

\[
\times \left(\psi_m\left(\frac{1+p-l+m}{T}\right) - \psi_m\left(\frac{1+p}{T}\right)\right).
\]

In order to establish the claim (4.35), it is sufficient to show

\[
E_{T,1} = \frac{v}{2\pi} \int_0^\omega \int_0^v f(u, \lambda) du d\lambda + O\left(\frac{1}{T}\right), \tag{4.41}
\]

\[
E_{T,2} = O\left(\frac{1}{T}\right). \tag{4.42}
\]
Regarding the statement (4.41), we note that simple calculations yield

\[
E_{T,1} = v \sum_{k=1}^{[vT/2]} \frac{1}{2\pi} \sum_{l,m=0}^{\infty} \exp(-i\lambda_{k,2}[vT/2](l - m)) \frac{1}{v} \int_0^v \psi_l(u) \psi_m(u) du + O(\frac{1}{T})
\]

\[
= \left[ \frac{vT/2}{T} \right]^{[vT/2]} \sum_{k=1}^{[vT/2]} \int_0^v f(u, \lambda_{k,2}[vT/2]) du + O(\frac{1}{T})
\]

\[
= \frac{v}{2\pi} \int_0^{\omega^2} \int_0^v f(u, \lambda) d\lambda du + O(\frac{1}{T}).
\]

With respect to the claim (4.42), a Taylor expansion implies that the absolute value of \(E_{T,2}\) is bounded by

\[
\frac{C}{T} \sum_{k=1}^{[vT/2]} \sum_{l,m=0}^{\infty} \frac{1}{2[vT/2]} \sum_{p=0}^{2[vT/2]-1} \left| \psi_l \left( \frac{1+p}{T} \right) \left( \psi_m \left( \frac{1+p - l - m}{T} \right) - \psi_m \left( \frac{1+p}{T} \right) \right) \right|
\]

\[
\leq C \sum_{l,m=0}^{\infty} \sup_{u \in [0,1]} \left| \psi_l(u) \right| \left| \frac{m - l}{T} \right| \sup_{u \in [0,1]} \left| \psi'_m(u) \right| = O(\frac{1}{T}),
\]

where in the last step we employed the conditions (4.10) and (4.11). This completes the proof of (4.35).

**Proof of part (2):** For the calculation of the covariances of the empirical process \(\{\sqrt{T}D_T(v, \omega)\}_{(v, \omega) \in [0,1]^2}\), we first consider the process \(\{\sqrt{T}D_{T,1}(v, \omega)\}_{(v, \omega) \in [0,1]^2}\), where

\[
\hat{D}_{T,1}(v, \omega) := \frac{v}{T} \sum_{k=1}^{[vT/2]} I_{2[vT/2]}(\lambda_{k,2}[vT/2])
\]

for \((v, \omega) \in [0,1]^2\). In the following calculations, we will demonstrate that the asymptotic covariance process \(\{\sqrt{T}D_{T,1}(v, \omega)\}_{(v, \omega) \in [0,1]^2}\) satisfies

\[
\text{Cov}(\sqrt{T}D_{T,1}(v_1, \omega_1), \sqrt{T}D_{T,1}(v_2, \omega_2)) = \frac{1}{2\pi} \min(v_1, v_2)v_1v_2
\]

\[
\times \int_0^{\min(\omega_1, \omega_2)\pi} f^2(\lambda) d\lambda + O(\frac{1}{T^3}) \quad (4.43)
\]

for \((v_1, \omega_1), (v_2, \omega_2) \in [0,1]^2\) and any \(\alpha \in (0,1/4)\). Because of the relation

\[
\sqrt{T}D_T(v, \omega) = \sqrt{T}\left( \hat{D}_{T,1}(v, \omega) - v^2 \hat{D}_{T,1}(1, \omega) \right),
\]
it follows by linearity arguments that
\[
\text{Cov} \left( \sqrt{T} \hat{D}_T(v_1, \omega_1), \sqrt{T} \hat{D}_T(v_2, \omega_2) \right)
= T \text{Cov} \left( \hat{D}_{T,1}(v_1, \omega_1) - v_1^2 \hat{D}_{T,1}(1, \omega_1), \hat{D}_{T,1}(v_2, \omega_2) - v_2^2 \hat{D}_{T,1}(1, \omega_2) \right)
= T \left( \text{Cov} \left( \hat{D}_{T,1}(v_1, \omega_1), \hat{D}_{T,1}(v_2, \omega_2) \right) + v_1^2 v_2^2 \text{Cov} \left( \hat{D}_{T,1}(1, \omega_1), \hat{D}_{T,1}(1, \omega_2) \right) \right)
- v_1^2 \text{Cov} \left( \hat{D}_{T,1}(1, \omega_1), \hat{D}_{T,1}(v_2, \omega_2) \right) - v_2^2 \text{Cov} \left( \hat{D}_{T,1}(v_1, \omega_1), \hat{D}_{T,1}(1, \omega_2) \right)
\]
\[
= \frac{1}{2\pi} \int_{\min(\omega_1, \omega_2) \pi}^{\infty} f^2(\lambda) d\lambda \left( \min(v_1, v_2) v_1 v_2 + v_1^2 v_2^2 - v_1^2 v_2^2 - i \frac{1}{T} \right)
= \frac{1}{2\pi} v_1 v_2 (\min(v_1, v_2) - v_1 v_2) \int_{\min(\omega_1, \omega_2) \pi}^{\infty} f^2(\lambda) d\lambda + O \left( \frac{1}{T^\alpha} \right),
\]
which completes the proof of (4.33). In order to show the claim (4.43), we assume without loss of generality that \(v_1 \leq v_2\) and obtain by the linearity of the cumulant [see Theorem 2.3.2] that
\[
\text{Cov} \left( \sqrt{T} \hat{D}_{T,1}(v_1, \omega_1), \sqrt{T} \hat{D}_{T,1}(v_2, \omega_2) \right)
= \frac{v_1 v_2}{T} \sum_{k_1=1}^{[\omega_1 v_1 T/2]} \sum_{k_2=1}^{[\omega_2 v_2 T/2]} \text{cum} \left( I_{2[v_1 T/2]}(\lambda_{k_1,2[v_1 T/2]}), I_{2[v_2 T/2]}(\lambda_{k_2,2[v_2 T/2]}) \right)
= \frac{v_1 v_2}{T} \sum_{k_1=1}^{[\omega_1 v_1 T/2]} \sum_{k_2=1}^{[\omega_2 v_2 T/2]} \frac{1}{(2\pi)^2 \sqrt{v_1 T/2} \sqrt{v_2 T/2}} \sum_{p_1,q_1=0}^{2[v_1 T/2] - 1 \omega_1 \omega_2} \sum_{p_2,q_2=0}^{2[v_2 T/2] - 1 \omega_1 \omega_2}
\times \exp \left( -i \lambda_{k_1,2[v_1 T/2]}(p_1 - q_1) \right) \exp \left( -i \lambda_{k_2,2[v_2 T/2]}(p_2 - q_2) \right)
\times \text{cum} \left( X_{1+p_1,T} X_{1+q_1,T}, X_{1+p_2,T} X_{1+q_2,T} \right),
\]
where we employed the definition of the periodogram in the second step. As we assume that the time series \(\{X_{l,T}\}_{l=1,\ldots,T}\) is stationary, the linear coefficients \(\psi_l(u), \ l \in \mathbb{N}, \ u \in [0,1]\), in the representation (2.12) do not depend on the rescaled time \(u = t/T\). Therefore, we define \(\psi_1 := \psi_1(u)\) for notational convenience. By employing this notation, we see that the above term is equal to
\[
\frac{v_1 v_2}{T} \sum_{k_1=1}^{[\omega_1 v_1 T/2]} \sum_{k_2=1}^{[\omega_2 v_2 T/2]} \frac{1}{(2\pi)^2 \sqrt{v_1 T/2} \sqrt{v_2 T/2}} \sum_{p_1,q_1=0}^{2[v_1 T/2] - 1 \omega_1 \omega_2} \sum_{p_2,q_2=0}^{2[v_2 T/2] - 1 \omega_1 \omega_2}
\times \exp \left( -i \lambda_{k_1,2[v_1 T/2]}(p_1 - q_1) \right) \exp \left( -i \lambda_{k_2,2[v_2 T/2]}(p_2 - q_2) \right)
\times \text{cum} \left( Z_{1+p_1-l} Z_{1+q_1-m}, Z_{1+p_2-n} Z_{1+q_2-o} \right).
\]
The product theorem for cumulants [see Theorem 2.3.6] and the Gaussianity of the innovation sequence \( \{Z_t\}_{t \in \mathbb{Z}} \) imply

\[
\text{cum}(Z_a Z_b, Z_c Z_d) = \text{cum}(Z_a, Z_d)\text{cum}(Z_b, Z_c) + \text{cum}(Z_a, Z_c)\text{cum}(Z_b, Z_d)
\]

[see Example 2.3.1], which allows to split the term (4.44) into the sum of two summands, which we denote by \( V_{T,1} \) and \( V_{T,2} \) and which are defined by

\[
V_{T,1} := \frac{v_1 v_2}{T} \sum_{k_1=1}^{\lfloor v_1 T/2 \rfloor} \sum_{k_2=1}^{\lfloor v_2 T/2 \rfloor} \frac{1}{(2\pi)^2 [v_1 T/2] [v_2 T/2]} \sum_{p_1,q_1=0}^{2[v_1 T/2]-1} \sum_{p_2,q_2=0}^{2[v_2 T/2]-1} \exp(-i\lambda_{k_1,2 [v_1 T/2]} (p_1 - q_1)) \exp(-i\lambda_{k_2,2 [v_2 T/2]} (p_2 - q_2)) \sum_{l,m,n,o=0}^\infty \psi_l \psi_m \psi_n \psi_o
\]

\[
V_{T,2} = \frac{v_1 v_2}{T} \sum_{k_1=1}^{\lfloor v_1 T/2 \rfloor} \sum_{k_2=1}^{\lfloor v_2 T/2 \rfloor} \frac{1}{(2\pi)^2 [v_1 T/2] [v_2 T/2]} \sum_{p_1,q_1=0}^{2[v_1 T/2]-1} \sum_{p_2,q_2=0}^{2[v_2 T/2]-1} \exp(-i\lambda_{k_1,2 [v_1 T/2]} (p_1 - q_1)) \exp(-i\lambda_{k_2,2 [v_2 T/2]} (p_2 - q_2)) \sum_{l,m,n,o=0}^\infty \psi_l \psi_m \psi_n \psi_o
\]

\[
\text{cum}(Z_{1+p_1-1}, Z_{1+q_2-2} \circ \text{cum}(Z_{1+q_1-m}, Z_{1+p_2-n})
\]

We note that the claim (4.43) follows from the statements

\[
V_{T,1} = \frac{1}{2\pi} v_1 v_2 \min(v_1, v_2) \int_0^{\min(\omega_1, \omega_2)\pi} f^2(\lambda) d\lambda + O\left(\frac{1}{T^\alpha}\right),
\]

\[
V_{T,2} = O\left(\frac{1}{T^{1-\varepsilon'}}\right)
\]

for some \( \alpha \in (0, 1/4) \) and \( \varepsilon' > 0 \). In the following discussion, we will establish the claims (4.47) and (4.48).

**Proof of (4.47):** Note that the property (4.37) implies that the relations

\[
q_2 = p_1 - l + o \quad \text{and} \quad q_1 = p_2 - n + m
\]

have to hold for all non-vanishing summands in the definition of \( V_{T,1} \). Because of \( q_1 \in \{0, \ldots, 2[v_1 T/2] - 1\} \) and \( q_2 \in \{0, \ldots, 2[v_2 T/2] - 1\} \), (4.49) additionally yields the restrictions

\[
0 \leq p_1 - l + o \leq 2[v_2 T/2] - 1,
\]

\[
0 \leq p_2 - n + m \leq 2[v_1 T/2] - 1
\]
for $p_1$ and $p_2$. Thus, we obtain

$$V_{T,1} = \frac{v_1v_2}{T} \sum_{k_1=1}^{\lfloor v_1T/2 \rfloor} \sum_{k_2=1}^{\lfloor v_2T/2 \rfloor} \frac{1}{(2\pi)^2} \sum_{l,m,n,o=0}^{\infty} \psi_l\psi_m\psi_n\psi_o \exp(-i\lambda_{k_1,2[v_1T/2]}(m-n))$$

\[
\times \exp(-i\lambda_{k_2,2[v_2T/2]}(l-o)) \frac{1}{4[v_1T/2][v_2T/2]} \sum_{p_1=0}^{2[v_1T/2]-1} \sum_{p_2=0}^{2[v_2T/2]-1}
\]

\[
\times \exp(-i(\lambda_{k_1,2[v_1T/2]} - \lambda_{k_2,2[v_2T/2]})(p_1 - p_2)).
\]

(4.52)

In the next step, we will show that the conditions (4.50) and (4.51) in the summations with respect to $p_1$ and $p_2$ can be replaced by

$$0 \leq p_1 \leq 2[v_2T/2] - 1,$$

(4.53)

$$0 \leq p_2 \leq 2[v_1T/2] - 1$$

(4.54)

respectively and that the error involved in this transition is of order $O(1/T^{1-\varepsilon'})$ for any $\varepsilon' > 0$. We only illustrate this fact for the case (4.50) regarding $p_1$ and the analog claim regarding the summation over $p_2$ follows by the same arguments. The error made by changing the restriction in the summation over $p_1$ from (4.50) to (4.53) is not larger than

$$\frac{Cv_1v_2}{T} \sum_{k_1=1}^{\lfloor v_1T/2 \rfloor} \sum_{k_2=1}^{\lfloor v_2T/2 \rfloor} \sum_{l,m,n,o=0}^{\infty} \frac{|o - l||\psi_l\psi_m\psi_n\psi_o|}{4[v_1T/2][v_2T/2]} \sum_{p_2=0}^{2[v_2T/2]-1}
\]

\[
\times \exp(-i(\lambda_{k_1,2[v_1T/2]} - \lambda_{k_2,2[v_2T/2]})(-p_2)),
\]

(4.55)

where we used the fact that the cardinality of the set over which the summation with respect to $p_1$ is performed is altered by at most $|o - l|$ elements by dropping the $l$ and $o$ terms in the condition (4.50). Now, we divide the quantity (4.55) into two sums, which we identify with the symbols $E_{(4.56)}$ and $E_{(4.57)}$ reflecting the summations over tuples $(k_1, k_2)$ satisfying the conditions

$$|2k_2[v_2T/2] - 2k_1[v_1T/2]| \leq 2[v_1T/2],$$

(4.56)

$$2[v_1T/2] < |2k_1[v_2T/2] - 2k_2[v_1T/2]| \leq 2[v_1T/2][v_2T/2]$$

(4.57)

respectively. We show that for every $\varepsilon' > 0$ it holds

$$E_{(4.56)} = O\left(\frac{1}{T}\right),$$

(4.58)

$$E_{(4.57)} = O\left(\frac{1}{T^{1-\varepsilon'}}\right).$$

(4.59)

The claim (4.58) is an obvious consequence of the summability condition (4.10) and the fact that for fixed $k_1$ there exist at most three values for $k_2$ such that the restriction (4.56)
is fulfilled. For a proof of \((4.59)\) and later applications, we note that the geometric series formula and the identity \(|1 - \exp(ix)| = 2|\sin(x/2)|\) imply that the inequality

\[
\left| \sum_{p=0}^{n} \exp(-i(\lambda_{k_1,2,v_1 T/2} - \lambda_{k_2,2,v_2 T/2})) \right| = \left| \frac{1 - \exp(i2\pi(2k_1\lfloor v_2 T/2 \rfloor - 2k_2\lfloor v_1 T/2 \rfloor))}{1 - \exp(-i2\pi(2k_1\lfloor v_2 T/2 \rfloor - 2k_2\lfloor v_1 T/2 \rfloor))} \right| \]

\[
= \left| \frac{\sin(\pi(2k_1\lfloor v_2 T/2 \rfloor - 2k_2\lfloor v_1 T/2 \rfloor)n)}{\sin(\pi(2k_1\lfloor v_2 T/2 \rfloor - 2k_2\lfloor v_1 T/2 \rfloor))} \right| \leq \left| \sin\left(\frac{\pi(2k_1\lfloor v_2 T/2 \rfloor - 2k_2\lfloor v_1 T/2 \rfloor)}{4\lfloor v_1 T/2 \rfloor\lfloor v_2 T/2 \rfloor}\right) \right|^{-1}
\]

holds for \(n \geq 1\) and \((k_1, k_2)\) such that \(\lambda_{k_1,2,v_1 T/2} - \lambda_{k_2,2,v_2 T/2} \neq 0\). It is easily seen how this bound can be exploited to show that \(E_{(4.57)}\) is not larger than

\[
\frac{Cv_1 v_2}{T} \sum_{k_1=1}^{\lfloor v_1 T/2 \rfloor} \left| \sum_{k_2=1}^{\lfloor v_2 T/2 \rfloor} \sum_{l,m,n,o=0}^{\infty} |o - l| |\psi_l \psi_m \psi_n \psi_o| \frac{1}{4\lfloor v_1 T/2 \rfloor\lfloor v_2 T/2 \rfloor} \times \left| \sin\left(\frac{\pi(2k_1\lfloor v_2 T/2 \rfloor - 2k_2\lfloor v_1 T/2 \rfloor)}{4\lfloor v_1 T/2 \rfloor\lfloor v_2 T/2 \rfloor}\right) \right|^{-1}.
\]

At this point, we make use of the property

\[
|\sin(\pi x)| > \begin{cases} 
C|x| & \text{if } |x| \in [0, 1/2] \\
C(1 - |x|) & \text{if } |x| \in (1/2, 1]
\end{cases}
\]

(4.61)
of the sin function to obtain that the above term is not larger than

\[
\frac{Cv_1 v_2}{T} \sum_{k_1=1}^{\lfloor v_1 T/2 \rfloor} \left| \sum_{k_2=1}^{\lfloor v_2 T/2 \rfloor} \sum_{l,m,n,o=0}^{\infty} |o - l| |\psi_l \psi_m \psi_n \psi_o| \frac{1}{|2k_1\lfloor v_2 T/2 \rfloor - 2k_2\lfloor v_1 T/2 \rfloor|} \times \left| \sin\left(\frac{\pi(2k_1\lfloor v_2 T/2 \rfloor - 2k_2\lfloor v_1 T/2 \rfloor)}{4\lfloor v_1 T/2 \rfloor\lfloor v_2 T/2 \rfloor}\right) \right|^{-1},
\]

(4.62)

where in the second step the summability condition \((4.10)\) was used. The condition \((4.57)\), which assures that \(|k_1\lfloor v_2 T/2 \rfloor/\lfloor v_1 T/2 \rfloor - k_2| \geq 1\), yields that \((4.62)\) is at most of size

\[
\frac{Cv_1 v_2}{T} \sum_{k_1=1}^{\lfloor v_1 T/2 \rfloor} \left| \sum_{k_2=1}^{\lfloor v_2 T/2 \rfloor} \frac{1}{|k_1\lfloor v_2 T/2 \rfloor/\lfloor v_1 T/2 \rfloor - k_2|} \times \left| \sin\left(\frac{\pi(2k_1\lfloor v_2 T/2 \rfloor - 2k_2\lfloor v_1 T/2 \rfloor)}{4\lfloor v_1 T/2 \rfloor\lfloor v_2 T/2 \rfloor}\right) \right|^{-1},
\]

(4.63)

where in the second step the summability condition \((4.10)\) was used. The condition \((4.57)\), which assures that \(|k_1\lfloor v_2 T/2 \rfloor/\lfloor v_1 T/2 \rfloor - k_2| \geq 1\), yields that \((4.62)\) is at most of size

\[
\frac{Cv_1 v_2}{T} \sum_{k_1=1}^{\lfloor v_1 T/2 \rfloor} \left| \sum_{k_2=1}^{\lfloor v_2 T/2 \rfloor} \frac{1}{|k_1\lfloor v_2 T/2 \rfloor/\lfloor v_1 T/2 \rfloor - k_2|} \times \left| \sin\left(\frac{\pi(2k_1\lfloor v_2 T/2 \rfloor - 2k_2\lfloor v_1 T/2 \rfloor)}{4\lfloor v_1 T/2 \rfloor\lfloor v_2 T/2 \rfloor}\right) \right|^{-1},
\]

(4.63)
for any $\varepsilon' > 0$. This completes the proof of (4.59).

Similar arguments yield that the restriction in the summation over $p_2$ can be changed from (4.51) to (4.54) by inclusion of an error term of order $O(1/T^{1-\varepsilon'})$ and we thus obtain that

$$V_{1,T} = \frac{v_1 v_2}{T} \sum_{k_1=1}^{[v_1 T/2]} \sum_{k_2=1}^{[v_2 T/2]} f(\lambda_{k_1,2[v_1 T/2]} f(\lambda_{k_2,2[v_2 T/2]}) \frac{1}{4[v_1 T/2][v_2 T/2]}$$

$$\times \sum_{p_1,p_2=0}^{2 \min([v_1 T/2],[v_2 T/2])-1} \exp(-i(\lambda_{k_1,2[v_1 T/2]} - \lambda_{k_2,2[v_2 T/2]})(p_1 - p_2)) + O(\frac{1}{T^{1-\varepsilon'}}).$$

In the next step, we intend to replace $\lambda_{k_2,2[v_2 T/2]}$ by $\lambda_{k_1,2[v_1 T/2]}$ in the argument of the function $f$. By employing a Taylor expansion, it is straightforward to see that the error involved in this transition is bounded by [note that we assumed $v_1 \leq v_2$]

$$\frac{C v_1 v_2}{T} \sum_{k_1=1}^{[v_1 T/2]} \sum_{k_2=1}^{[v_2 T/2]} \frac{|\lambda_{k_1,2[v_1 T/2]} - \lambda_{k_2,2[v_2 T/2]}|}{4[v_1 T/2][v_2 T/2]}$$

$$\times \sum_{p_1,p_2=0}^{2[v_1 T/2]-1} \exp(-i(\lambda_{k_1,2[v_1 T/2]} - \lambda_{k_2,2[v_2 T/2]})(p_1 - p_2)) =: S_{(4.56)} + S_{(4.57)},$$

where $S_{(4.56)}$ and $S_{(4.57)}$ denote the sums over all $(k_1, k_2)$ satisfying conditions (4.56) and (4.57) respectively. We will show that for every $\varepsilon' > 0$ it holds

$$S_{(4.56)} = O(\frac{1}{T}),$$

$$S_{(4.57)} = O(\frac{1}{T^{1-\varepsilon'}}).$$

We first consider the sum $S_{(4.56)}$ and observe that for fixed $k_1$, there exist at most two values for $k_2$ such that the condition (4.56) holds and the respective summand is non-vanishing. Furthermore, for each of these values we have $|\lambda_{k_1,2[v_1 T/2]} - \lambda_{k_2,2[v_2 T/2]}| \leq 2\pi(2[v_2 T/2])^{-1}$ and we hence obtain

$$S_{(4.56)} \leq \frac{v_1 v_2}{T} \sum_{k_1=1}^{[v_1 T/2]} \frac{2\pi}{[v_2 T/2]} = O(\frac{1}{T}).$$
We continue by establishing the statement (4.65). For this purpose, we note that the same arguments as employed in (4.60) yield the inequality

\[
\sum_{p_1,p_2=0}^{2[v_1T/2]-1} \exp(-i(\lambda_{k_1,2v_1T} - \lambda_{k_2,2v_2T})(p_1 - p_2)) \leq \left| (1 - \exp(-i(\lambda_{k_1,2v_1T} - \lambda_{k_2,2v_2T}))(1 - \exp(i(\lambda_{k_1,2v_1T} - \lambda_{k_2,2v_2T}))\right|^2 \leq \sin \left( \frac{(\lambda_{k_1,2v_1T} - \lambda_{k_2,2v_2T})}{2} \right)^2, \tag{4.66}
\]

for each \((k_1, k_2)\) satisfying (4.57). By using this bound, it can easily be seen that \(S_{(4.57)}\) is at most of the size

\[
\frac{Cv_1v_2}{T} \sum_{k_1=1}^{v_1T/2} \sum_{k_2=1}^{v_2T/2} \frac{\left| 2k_1 [v_2T/2] - 2k_2 [v_1T/2] \right|}{\left(4[v_1T/2][v_2T/2]\right)^2} \frac{1}{\sin^2 \left( \frac{\pi(2k_1 [v_2T/2] - 2k_2 [v_1T/2])}{4[v_1T/2][v_2T/2]} \right)},
\]

and an application of (4.61) yields

\[
\frac{Cv_1v_2}{T} \sum_{k_1=1}^{v_1T/2} \sum_{k_2=1}^{v_2T/2} \frac{\left| 2k_1 [v_2T/2] - 2k_2 [v_1T/2] \right|}{\left(4[v_1T/2][v_2T/2]\right)^2} \frac{1}{k_1 [v_2T/2] - k_2} = \frac{Cv_1v_2}{T} \sum_{k_1=1}^{v_1T/2} \sum_{k_2=1}^{v_2T/2} \frac{1}{k_1 [v_2T/2] - k_2}
\]

as an upper bound. Employing the fact \(\left| k_1 [v_2T/2] / [v_1T/2] - k_2 \right| \geq 1\), which is due to the condition (4.57), it is easy to see that the above term is of order

\[
\frac{C}{T[v_1T/2]} \sum_{k_1=1}^{v_1T/2} \log(\left| v_2T/2 \right|) = O(\frac{\log(T)}{T}) = O(\frac{1}{T^{1-\varepsilon}}),
\]

for any \(\varepsilon' > 0\). This completes the proof of (4.65). Now, the properties (4.64) and (4.65) imply that

\[
V_{1,T} = \frac{v_1v_2}{T} \sum_{k_1=1}^{[\omega_1[v_1T/2]]} f^2(\lambda_{k_1,2v_1T/2}) \frac{1}{4[v_1T/2][v_2T/2]} \sum_{k_2=1}^{[\omega_2[v_2T/2]]} \exp(-i \frac{2\pi(k_1[v_2T/2] - k_2[v_1T/2])}{4[v_1T/2][v_2T/2]}(p_1 - p_2)) + O(\frac{1}{T^{1-\varepsilon}}). \tag{4.67}
\]

For the following arguments, let \(a_T\) denote some sequence satisfying

\[
\frac{a_T}{T^2} \to 0 \quad \text{and} \quad \frac{T^{1+\alpha}}{a_T} \to 0 \tag{4.68}
\]
for some $\alpha > 0$ as $T \to \infty$. We split the sum over $(k_1, k_2)$ in (4.67) in three sums with respect to the conditions

$$0 = |k_12v_2T/2| - k_22v_1T/2|, \quad (4.69)$$
$$1 \leq |k_12v_2T/2| - k_22v_1T/2| \leq v_1v_2\alpha_T, \quad (4.70)$$
$$v_1v_2\alpha_T \leq |k_12v_2T/2| - k_22v_1T/2|. \quad (4.71)$$

In the first step, we will show that the sum over all $(k_1, k_2)$, which satisfy the restriction (4.71), is of vanishing order. To achieve this goal, we employ (4.66) and the inequality (4.61) to obtain

$$\frac{v_1v_2}{T} \sum_{k_1=1}^{\omega_1[v_1T/2]} f^2(\lambda_{k_12[v_1T/2]}) \frac{1}{4[v_1T/2][v_2T/2]} \sum_{k_2=1}^{\omega_2[v_2T/2]} \left| \sum_{p_1, p_2=0}^{2[v_1T/2]-1} \left| \omega_2[v_2T/2] \right| \right| \left| \omega_2[v_2T/2] \right| \sum_{k_2=1}^{\omega_2[v_2T/2]} \frac{1}{(k_12v_2T/2) - k_22v_1T/2)]^2}$$

$$\leq \frac{Cv_1v_2}{T} \sum_{k_1=1}^{\omega_1[v_1T/2]} \frac{v_2T/2}{v_1T/2} \sum_{k_2=1}^{\omega_2[v_2T/2]} \frac{1}{(k_12v_2T/2) - k_22v_1T/2)}^2.$$

Bearing the conditions $1 \leq |k_12v_2T/2|/v_1T/2| - k_2|$ and $v_1v_22\alpha_T/2v_1T/2 \geq v_22\alpha$ in mind, we see that the above quantity is of order

$$\frac{Cv_2^2}{T} \sum_{k_1=1}^{\omega_1[v_1T/2]} \sum_{k_2=2v_2T/2}^{\infty} \frac{1}{k_2^2} = \frac{Cv_2^2}{T} \sum_{k_1=1}^{\omega_1[v_1T/2]} O(1/(v_22\alpha_{1-\epsilon'})) = O(\frac{1}{T\alpha(1-\epsilon')}) \quad (4.72)$$

for each $\epsilon' > 0$. Hence, we obtain from (4.67) that

$$V_{1,T} = \frac{1}{T} \sum_{k_1=1}^{\omega_1[v_1T/2]} f^2(\lambda_{k_12[v_1T/2]}) A_T(v_1, v_2, k_1, \omega_2) + O(\frac{1}{T\alpha(1-\epsilon')}), \quad (4.73)$$
where the term $A_T$ is comprised of the summations with respect to conditions (4.69) and (4.70), i.e.

$$A_T(v_1, v_2, k_1, \omega_2) := \frac{v_1v_2}{4[v_1T/2][v_2T/2]} \sum_{k_2=1}^{[\omega_2[v_2T/2]]} \frac{2[v_1T/2]^{-1}}{4[v_1T/2][v_2T/2]} \exp(-i \frac{2\pi(k_12[v_2T/2] - k_22[v_1T/2])}{4[v_1T/2][v_2T/2]}(p_1 - p_2)).$$

For the arguments employed below let $b_T$ denote an arbitrary sequence that satisfies

$$\frac{T^{1/2+\alpha}}{b_T} \to 0 \quad \text{and} \quad \frac{b_T}{T^{1-\alpha}} \to 0,$$

where $\alpha > 0$ denotes the constant which was introduced for the definition of the sequence $a_T$ in (4.68). In the next step, we conduct a detailed investigation of the function $A_T$. Therefore, we distinguish the cases

$$k_1 \in \{b_T, ..., [\omega_2[v_1T/2]] - b_T\}$$

$$k_1 \in \{[\omega_2[v_1T/2]] + b_T, ..., [v_1T/2]\}$$

and show

$$A_T(v_1, v_2, k_1, \omega_2) = \begin{cases} v_1v_2 + O(\frac{1}{T}) & \text{if } (4.75) \\ O(\frac{1}{T}) & \text{if } (4.76) \end{cases}$$

uniformly in $k_1$. We consider the case (4.75) first and define the indicator function

$$\mathbb{I}(k_1) := \begin{cases} 1 & \text{if } k_12[v_2T/2] = \frac{t}{v_1T/2} \text{ for some } t \in \{1, ..., [\omega_2[v_2T/2]]\} \\ 0 & \text{else} \end{cases}.$$ 

This notation, an application of the geometric series formula as in (4.66) and the identity $|1 - \exp(ix)|^2 = 4\sin^2(x/2)$ imply that in the case (4.75) the function $A_T(v_1, v_2, k_1, \omega_2)$ is equal to

$$v_1v_2 \left( \mathbb{I}(k_1) \frac{v_1T/2}{[v_2T/2]} + \frac{1}{4[v_1T/2][v_2T/2]} \sum_{k_2=1}^{[\omega_2[v_2T/2]]} \frac{\sin^2(\frac{\pi(k_12[v_2T/2] - k_22[v_1T/2])}{2[v_2T/2]})}{\sin^2(\frac{\pi(k_12[v_2T/2] - k_22[v_1T/2])}{4[v_1T/2][v_2T/2]})} \right),$$

where we used that in the case $\mathbb{I}(k_1) = 1$ the summand with $k_2 = k_1[v_2T/2]/[v_1T/2]$ in the definition of $A_T(v_1, v_2, k_1, \omega_2)$ corresponds to $[k_12[v_2T/2] - k_22[v_1T/2]] = 0$ and has the
value $v_1v_2[v_1T/2]/[v_2T/2]$. Now, we introduce the more compact notation

$$x(k_1, k_2) := \pi (k_1[2v_1T/2] - k_2[2v_1T/2]) / (4[v_1T/2][v_2T/2])$$

and employ the approximation $\sin(x) = x + E(x)$ with $|E(x)| \leq Cx^3$ uniformly in $x \in [0, 1/2]$ to obtain that the error, which is made by exchanging $[x(k_1, k_2)]^2$ for $\sin^2(x(k_1, k_2))$ in the denominator of the second summand in (4.78), is bounded by

$$\left| \frac{v_1v_2}{4[v_1T/2][v_2T/2]} \sum_{k_2=1}^{[v_2T/2]} \frac{1}{[x(k_1, k_2)]^2} - \frac{1}{x(k_1, k_2) + E(x(k_1, k_2))]^2} \right| \leq \frac{v_1v_2}{4[v_1T/2][v_2T/2]} \sum_{k_2=1}^{[v_2T/2]} \frac{2|x(k_1, k_2)||E(x(k_1, k_2))| + E^2(x(k_1, k_2))}{[x(k_1, k_2)]^4 + 2|[x(k_1, k_2)]^2||E(x(k_1, k_2))| + [x(k_1, k_2)]^2E^2(x(k_1, k_2))}$$

$$\leq \frac{v_1v_2}{4[v_1T/2][v_2T/2]} \sum_{k_2=1}^{[v_2T/2]} \frac{x(k_1, k_2)^4(2C' + C'[x(k_1, k_2)]^2)}{x(k_1, k_2)]^4(1 - 2C'[x(k_1, k_2)]^2 - C'[x(k_1, k_2)]^4)}$$

$$= \frac{v_1v_2}{4[v_1T/2][v_2T/2]} \sum_{k_2=1}^{[v_2T/2]} \frac{2C' + C'[x(k_1, k_2)]^2}{1 - 2C'[x(k_1, k_2)]^2 - C'[x(k_1, k_2)]^4}. \quad (4.79)$$

The conditions (4.70) and $a_T/T^2 \to 0$ imply that for sufficiently large $T$ the terms $x(k_1, k_2)$ are smaller than any arbitrary positive constant $C''$. Taking $C''$ small enough shows that the denominator of each summand in (4.79) is larger than some positive constant for $T$ sufficiently large. This implies that (4.79) is of the order

$$\frac{Cv_1v_2}{4[v_1T/2][v_2T/2]} \sum_{k_2=1}^{[v_2T/2]} 1 = O\left(\frac{1}{T}\right)$$

and yields

$$A_T(v_1, v_2, k_1, \omega_1) = v_1v_2 \left( \mathcal{H}(k_1) \frac{v_1T/2}{[v_2T/2]} + 4[v_1T/2][v_2T/2] \sum_{k_2=1}^{[v_2T/2]} \frac{\omega_2[v_2T/2]}{2[v_1T/2]} \right) \times \frac{\sin^2\left(\pi(k_1[2v_1T/2] - k_2[2v_1T/2])\right)}{\left(\pi(k_1[2v_1T/2] - k_2[2v_1T/2])\right)^2} + O\left(\frac{1}{T}\right). \quad (4.80)$$
For the arguments below, we introduce the definitions

\[ x_1(k_1) := \inf \{ k_1 [v_2 T/2] - n [v_1 T/2], \quad n \in \mathbb{N} \text{ such that } n [v_1 T/2] < k_1 [v_2 T/2] \} \]

\[ x_2(k_1) := \inf \{ n [v_1 T/2] - k_1 [v_2 T/2], \quad n \in \mathbb{N} \text{ such that } n [v_1 T/2] > k_1 [v_2 T/2] \} \]

\[ t(k_1) := k_1 [v_2 T/2] / [v_1 T/2]. \]

Note that this notation implies \( x_1(k_1) = x_2(k_1) = 2[v_1 T/2] \) in the case \( I(k_1) = 1 \) and \( x_1(k_1) + x_2(k_1) = 2[v_1 T/2] \) in the case \( I(k_1) = 0 \). By splitting the sum over \( k_2 \) up into the summands which satisfy

\[ k_1 [v_2 T/2] - k_2 [v_1 T/2] > 0, \]

\[ k_1 [v_2 T/2] - k_2 [v_1 T/2] < 0, \]

respectively, we obtain from (4.80) that

\[ A_T(v_1, v_2, k_1, \omega_2) = v_1 v_2 \left\{ I(k_1) \frac{[v_1 T/2]}{[v_2 T/2]} + 4[v_1 T/2] [v_2 T/2] \left( \sum_{l=0}^{[t(k_1)]} \frac{\sin^2 \left( \frac{\pi(x_1(k_1) + l [v_1 T/2])}{2[v_2 T/2]} \right)}{\left( \pi(x_1(k_1) + l [v_1 T/2]) \right)^2} \right) + \frac{\sin^2 \left( \frac{\pi(x_2(k_1) + l [v_1 T/2])}{2[v_2 T/2]} \right)}{\left( \pi(x_2(k_1) + l [v_1 T/2]) \right)^2} \right\} + O(1) \]

\[ = v_1 v_2 \left\{ I(k_1) \frac{[v_1 T/2]}{[v_2 T/2]} + \frac{[v_2 T/2]}{[v_1 T/2]} \left( \sum_{l=0}^{\min([t(k_1)] - [x_1(k_1)]}, \frac{\sin^2 \left( \frac{\pi(x_1(k_1) + l [v_1 T/2])}{2[v_2 T/2]} \right)}{\left( \pi(x_1(k_1) + l [v_1 T/2]) \right)^2} \right) + \frac{\sin^2 \left( \frac{\pi(x_2(k_1) + l [v_1 T/2])}{2[v_2 T/2]} \right)}{\left( \pi(x_2(k_1) + l [v_1 T/2]) \right)^2} \right\} + O(1). \]  

For the following treatment of the above term, we intend to expand the summations over \( l \) from zero to \( \infty \). For this purpose, we will first show that the alteration made in the quantity (4.81) by expanding the sums over \( l \) to \( \infty \) is of order \( O(1/T^{\alpha(1-\varepsilon')}) \) for any \( \varepsilon' > 0 \), where \( \alpha \) denotes a constant such that (4.68) and (4.74) hold for the sequences \( a_T \) and \( b_T \) as \( T \to \infty \). In order to achieve this, we note that the conditions (4.68) and (4.74) imply for the upper
boundaries of the sums in (4.81) that
\[
\min\left\{ t(k_1), \frac{v_1v_2a_T - x_1(k_1)}{2\lfloor \frac{v_1T}{2} \rfloor} \right\} \geq C \min\left( \frac{bTv_2}{v_1}, \frac{av_2}{T} - 1 \right) \\
\geq CT^\alpha \min\left( \frac{v_2}{v_1}, v_2 \right) = C v_2 T^\alpha
\]
\[
\min\left( \frac{v_1v_2a_T - x_2(k_1)}{2\lfloor v_1T/2 \rfloor}, \lfloor \frac{v_2T}{2} \rfloor - \lfloor t(k_1) \rfloor \right) \geq C \min\left( \frac{v_1v_2a_T}{v_1T} - 1, v_2 T\left(\frac{v_2}{2} - \frac{\omega_2/2v_1T - b_T}{v_1T}\right) \right) \\
\geq CT^\alpha \min\left( v_2, \frac{v_2}{v_1}, v_2 \right) = C v_2 T^\alpha
\]
for \( T \) large enough. Together with the bound
\[
\sum_{k=v_2T^\alpha}^\infty \frac{1}{k^2} \leq \frac{1}{v_2^{1-\epsilon} T^{\alpha(1-\epsilon)}} \sum_{k=v_2T^\alpha}^\infty \frac{1}{k^{1+\epsilon}} = O\left( \frac{1}{v_2^{1-\epsilon} T^{\alpha(1-\epsilon)}} \right),
\]
it follows from (4.81) that
\[
A_T(v_1, v_2, k_1, \omega_2) = v_1v_2 \left\{ \mathbb{I}(k_1) \frac{v_1T/2}{v_2T/2} + \frac{v_1T/2}{v_1T/2} \left( \sum_{l=0}^\infty \frac{\sin^2\left( \frac{\pi(x_1(k_1)+2l)}{2v_1T/2} \right)}{\pi(x_1(k_1))/(2\lfloor v_1T/2 \rfloor) + l} \right)^2 \\
+ \sum_{l=0}^\infty \frac{\sin^2\left( \frac{\pi(x_2(k_1)+2l)}{2v_1T/2} \right)}{\pi(x_2(k_1))/(2\lfloor v_1T/2 \rfloor) + l} \right) \right\} + O\left( \frac{1}{T^{\alpha(1-\epsilon)}} \right) \quad (4.82)
\]
for any \( \epsilon' > 0 \). From this term, the identity (4.77) can be derived by simple calculations and we only demonstrate this for the case \( \mathbb{I}(k_1) = 1 \), which implies \( x_1(k_1) = x_2(k_1) = 2\lfloor v_1T/2 \rfloor \). By using the identities
\[
\sin^2(x) = \frac{1}{2}(1 - \cos(2x)) \quad \text{and} \quad \sum_{k=1}^\infty \frac{\cos(kx)}{k^2} = \left( \frac{x - \pi}{2} \right)^2 - \frac{\pi^2}{12}
\]
[see Jolley (1961)] it is easy to see that (4.82) is, up to a term of order \( O(1/T^{\alpha(1-\epsilon)}) \), equal to
\[
v_1v_2 \left\{ \frac{v_1T/2}{v_2T/2} + \frac{v_2T/2}{v_1T/2} \left( \sum_{l=1}^\infty \frac{2}{\pi l^2} \sin^2\left( \frac{\pi l}{v_1T/2} \right) \right) \right\} \\
v_1v_2 \left\{ \frac{v_1T/2}{v_2T/2} + \frac{v_2T/2}{v_1T/2} \left( \sum_{l=1}^\infty \frac{1}{\pi l^2} - \frac{\pi^2}{12} \sum_{l=1}^\infty \frac{\cos(2\pi l/v_1T/2)}{l^2} \right) \right\} \\
v_1v_2 \left\{ \frac{v_1T/2}{v_2T/2} + \frac{v_2T/2}{v_1T/2} \left( \frac{\pi^2}{6} - \frac{\pi^2}{4} \right) + \frac{\pi^2}{12} \right\} \\
v_1v_2 \left\{ \frac{v_1T/2}{v_2T/2} + \frac{v_2T/2}{v_1T/2} \left( - \frac{\pi^2}{4} \right) + \frac{\pi^2}{12} \right\} = v_1v_2.
\]
which completes the proof of (4.77) in the case (4.75). To prove of the identity (4.77) in the
case (4.76), we apply the bound (4.66) and (4.61) to obtain that the quantity $A_T(v_1, v_2, k_1, \omega_2)$
is of the order
\[
v_1v_2 \left( 4\left\lfloor \frac{v_1T}{2} \right\rfloor \left\lfloor \frac{v_2T}{2} \right\rfloor \frac{1}{\pi(k_12\left\lfloor \frac{v_2T}{2} \right\rfloor - k_22\left\lfloor \frac{v_1T}{2} \right\rfloor)^2} \right) \sum_{k_2=1}^{\lfloor \omega_2/v_2T/2 \rfloor} \frac{1}{(\pi(2\left\lfloor \omega_2/v_1T/2 \right\rfloor + 2b_T\left\lfloor v_2T/2 \right\rfloor - k_22\left\lfloor v_1T/2 \right\rfloor)^2)} \sum_{k_3=1}^{\lfloor \omega_2/v_2T/2 \rfloor} \frac{1}{(\pi b_T)^2} = O\left( \frac{T}{b_T^2} \right) = O\left( \frac{1}{T^\alpha} \right),
\]
where in the second step we used that $k_2$ is not larger than $\lfloor \omega_2/v_2T/2 \rfloor$. This completes
the proof of (4.77) and we now show how (4.47) follows. Due to $A_T(v_1, v_2, k_1, \omega_2) = O(1)$
for all $k_1$, it is straightforward to see that (4.73) implies
\[
V_{1,T} = \frac{1}{T} \sum_{k_1=b_T}^{\lfloor \omega_1/v_1T/2 \rfloor} f^2(\lambda_{k_1,2\left\lfloor v_1T/2 \right\rfloor})A_T(v_1, v_2, k_1, \omega_2) + O\left( \frac{b_T}{T} \right) + O\left( \frac{1}{T^{\alpha(1-\epsilon)}} \right),
\]
Now, we employ the property (4.77) concerning the function $A_T$ and the condition $b_T/T = O(1/T^\alpha)$
to obtain from (4.83) [note that we assumed $v_1 \leq v_2$]
\[
V_{T,1} = v_1v_2 \frac{\left\lfloor \frac{v_1T}{2} \right\rfloor}{T} \frac{1}{\left\lfloor \frac{v_1T}{2} \right\rfloor} \sum_{k_1=1}^{\min(\omega_1,\omega_2)/v_1T/2} f^2(\lambda_{k_1,2\left\lfloor v_1T/2 \right\rfloor}) + O\left( \frac{1}{T^{\alpha(1-\epsilon)}} \right)
\]
\[
= \frac{\min(v_1, v_2)v_1v_2}{2\pi} \int_0^{\min(\omega_1,\omega_2)\pi} f^2(\lambda)d\lambda + O\left( \frac{1}{T^{\alpha(1-\epsilon)}} \right),
\]
which completes the proof of the claim (4.47).

**Proof of (4.48):** Note that the independence of the innovations implies that the equations
\[
p_1 = p_2 + l - n,
q_1 = q_2 - o + m,
\]
have to hold for the non-vanishing terms in (4.46). Therefore, we obtain the restrictions
\[
0 \leq p_2 + l - n \leq 2\left\lfloor \frac{v_1T}{2} \right\rfloor - 1, \quad (4.85)
0 \leq q_2 - o + m \leq 2\left\lfloor \frac{v_1T}{2} \right\rfloor - 1, \quad (4.86)
\]
for the summation over $p_2$ and $q_2$. By simple calculations, it follows that

$$
V_{T,2} = \frac{v_1 v_2}{T} \sum_{k_1=1}^{[\omega_1 v_1 T/2]} \sum_{k_2=1}^{[\omega_2 v_2 T/2]} \frac{1}{(2\pi)^2} \sum_{l,m,n,o=0}^{\infty} \psi_l \psi_m \psi_n \psi_o \exp(-i \lambda_{k_1,2[v_1 T/2]} (l - n - m + o))
$$

$$
\times \frac{1}{4[v_1 T/2][v_2 T/2]} \sum_{p_2=0}^{2[v_2 T/2]-1} \sum_{q_2=0}^{2[v_2 T/2]-1} \exp(-i(\lambda_{k_1,2[v_1 T/2]} + \lambda_{k_2,2[v_2 T/2]})(p_2 - q_2)). \quad (4.87)
$$

By using analogous arguments in the demonstrations concerning the properties (4.58) and (4.59), we can change the conditions (4.85) and (4.86) to

$$
0 \leq p_2 \leq 2[v_1 T/2] - 1, \quad (4.88)
$$

$$
0 \leq q_2 \leq 2[v_1 T/2] - 1, \quad (4.89)
$$

with an additional error term of order $O(1/T^{1-\varepsilon'})$ (for any $\varepsilon' > 0$). Then, we employ the summability condition (4.10) and obtain that (4.87) is bounded by

$$
C v_1 v_2 \sum_{k_1=1}^{[\omega_1 v_1 T/2]} \sum_{k_2=1}^{[\omega_2 v_2 T/2]} \frac{1}{4[v_1 T/2][v_2 T/2]} \sum_{p_2=0}^{2[v_2 T/2]-1} \sum_{q_2=0}^{2[v_2 T/2]-1} \exp(-i(\lambda_{k_1,2[v_1 T/2]} + \lambda_{k_2,2[v_2 T/2]})(p_2 - q_2)) + O(\frac{1}{T^{1-\varepsilon'}}). \quad (4.90)
$$

For the next step of the proof, we introduce the condition

$$
(k_1, k_2) \neq ([v_1 T/2], [v_2 T/2]), \quad (4.91)
$$

which intends to exclude the case $\lambda_{k_1,2[v_1 T/2]} + \lambda_{k_2,2[v_2 T/2]} = 2\pi$ from the following considerations. Note that this case is only of relevance, if $\omega_1 = \omega_2 = 1$ and that in this situation the summand in (4.90) which corresponds to $(k_1, k_2)$ satisfying (4.91) is of order $O(1/T)$. Thus, the inequality (4.66) implies that (4.90) is bounded by

$$
\frac{C v_1 v_2}{T} \sum_{k_1=1}^{[\omega_1 v_1 T/2]} \sum_{k_2=1}^{[\omega_2 v_2 T/2]} \frac{1}{4[v_1 T/2][v_2 T/2]} \sin^2\left(\frac{\pi(k_1[v_2 T/2] + k_2[v_1 T/2])}{4[v_1 T/2][v_2 T/2]}\right)^{-2} + O(\frac{1}{T^{1-\varepsilon'}}). \quad (4.92)
$$

For the further treatment of the above quantity, we specify the conditions

$$
|k_1[v_2 T/2] + k_2[v_1 T/2]| \leq 2[v_1 T/2][v_2 T/2] \quad (4.93)
$$

$$
2[v_1 T/2][v_2 T/2] < |k_1[v_2 T/2] + k_2[v_1 T/2]| \leq 4[v_1 T/2][v_2 T/2] - 2[v_1 T/2] \quad (4.94)
$$
for \((k_1, k_2)\) corresponding to the cases
\[
\begin{align*}
\lambda_{k_1,2[v_1T/2]} + \lambda_{k_2,2[v_2T/2]} &\in (0, \pi/2], \\
\lambda_{k_1,2[v_1T/2]} + \lambda_{k_2,2[v_2T/2]} &\in (\pi/2, \pi),
\end{align*}
\]
respectively. In order to show the statement (4.48), we split the term (4.92) up into the sums over \((k_1, k_2)\) satisfying (4.93) and (4.94), which we denote by \(S_{(4.93)}\) and \(S_{(4.94)}\), and establish the properties
\[
S_{(4.93)} = O\left(\frac{1}{T^{1-\varepsilon'}}\right), \tag{4.95}
\]
\[
S_{(4.94)} = O\left(\frac{1}{T^{1-\varepsilon'}}\right). \tag{4.96}
\]
Regarding the statement (4.95), we observe that inequalities (4.66) and (4.61) imply
\[
S_{(4.93)} \leq \frac{Cv_1v_2}{T} \sum_{k_1=1}^{[\omega_1[v_1T/2]]} \frac{1}{4[v_1T/2][v_2T/2]} \sum_{k_2=1}^{[\omega_2[v_2T/2]]} \frac{1}{(k_1[v_2T/2] + k_2)^2 (4[v_1T/2][v_2T/2])^2} \leq \frac{Cv_1v_2}{T} \sum_{k_1=1}^{[v_1T/2]} \sum_{k_2=1}^{[v_2T/2]} \frac{1}{k_1[v_2T/2]/[v_1T/2]} = O\left(\frac{\log(T)}{T}\right) = O\left(\frac{1}{T^{1-\varepsilon'}}\right)
\]
for any \(\varepsilon' > 0\). With reference to the claim (4.96), we obtain by simple calculations using (4.66) and (4.61) that
\[
S_{(4.94)} \leq \frac{Cv_1v_2}{T} \sum_{k_1=1}^{[\omega_1[v_1T/2]]} \frac{1}{4[v_1T/2][v_2T/2]} \sum_{k_3=1}^{[\omega_2[v_2T/2]]} \frac{1}{(1 - \frac{k_1[v_2T/2] + k_2[v_1T/2]}{4[v_1T/2][v_2T/2])^2} \leq \frac{Cv_1v_2}{T} [v_2T/2] [v_1T/2] \sum_{k_1=1}^{[v_1T/2]} \sum_{k_2=1}^{[v_2T/2]} \frac{1}{2[v_2T/2] - (k_1[v_2T/2]/[v_1T/2] + k_2)^2}.
\]
By splitting the sum over \((k_1, k_2)\) up into the summands with \(k_1 = \lfloor v_1 T/2 \rfloor\) and \(k_1 \neq \lfloor v_1 T/2 \rfloor\), we get

\[
S_{(4.94)} \leq C v_1 v_2 \frac{v_2 T/2}{v_1 T/2} \frac{\lfloor v_1 T/2 \rfloor}{v_1 T/2} \sum_{k_1=1}^{\lfloor v_1 T/2 \rfloor} \sum_{k_2=1}^{v_2 T/2} \frac{1}{(2 \lfloor v_2 T/2 \rfloor - (k_1 \lfloor v_2 T/2 \rfloor + k_2))^2}
\]

\[
+ C v_1 v_2 \frac{v_2 T/2}{v_1 T/2} \frac{\lfloor v_2 T/2 \rfloor}{v_1 T/2} \sum_{k_2=1}^{\lfloor v_2 T/2 \rfloor} \frac{1}{(\lfloor v_2 T/2 \rfloor - k_2)^2}
\]

Now, we use the fact that in both of the above terms for any fixed \(k_1\) the sum over \(k_2\) is of order \(O(1/T)\), which implies

\[
S_{(4.94)} \leq C v_1 v_2 \frac{v_2 T/2}{v_1 T/2} \frac{\lfloor v_1 T/2 \rfloor}{v_1 T/2} \sum_{k_1=2}^{\lfloor v_1 T/2 \rfloor} \sum_{k_2=1}^{v_2 T/2} \frac{1}{(2 \lfloor v_2 T/2 \rfloor - (k_1 \lfloor v_2 T/2 \rfloor + k_2))^2}
\]

\[
+ C v_1 v_2 \frac{v_2 T/2}{v_1 T/2} \frac{\lfloor v_2 T/2 \rfloor}{v_1 T/2} \sum_{k_2=2}^{\lfloor v_2 T/2 \rfloor} \frac{1}{(\lfloor v_2 T/2 \rfloor - k_2)^2} + O \left( \frac{1}{T} \right)
\]

(4.97)

For the second term in the above expression, we obtain the bound

\[
C v_1 v_2 \frac{v_2 T/2}{v_1 T/2} \frac{\lfloor v_2 T/2 \rfloor}{v_1 T/2} \sum_{k_2=2}^{\lfloor v_2 T/2 \rfloor} \frac{1}{(\lfloor v_2 T/2 \rfloor - k_2)^2} \leq C v_1^2 \frac{\lfloor v_2 T/2 \rfloor}{v_1 T/2} \sum_{k_2=2}^{\lfloor v_2 T/2 \rfloor} \frac{1}{(\lfloor v_2 T/2 \rfloor - k_2)^2} dk_2
\]

\[
\leq C v_2^2 \left( \frac{1}{v_2 T/2 - k_2} \right)^{\lfloor v_2 T/2 \rfloor - 1} = O \left( \frac{1}{T} \right).
\]

For the first term in (4.97), simple calculations yield that it is bounded by

\[
C v_2^2 \sum_{k_2=1}^{\lfloor v_2 T/2 \rfloor} \left( \frac{1}{2 \lfloor v_2 T/2 \rfloor - k_1 \lfloor v_2 T/2 \rfloor} - t \right)^{\lfloor v_2 T/2 \rfloor + 1} = O \left( \frac{1}{T} \right)
\]

\[
= C v_2^2 \sum_{k_2=1}^{\lfloor v_2 T/2 \rfloor} \frac{1}{v_2 T/2 - k_1 \lfloor v_2 T/2 \rfloor / \lfloor v_1 T/2 \rfloor - 1} + O \left( \frac{1}{T} \right)
\]

\[
\leq C v_2^2 \int_{1}^{\lfloor v_1 T/2 \rfloor} \frac{1}{v_2 T/2 - k_1 \lfloor v_2 T/2 \rfloor / \lfloor v_1 T/2 \rfloor - 1} dk_1 + O \left( \frac{1}{T} \right)
\]

\[
\leq C v_2^2 \left[ \log(|v_2 T/2 - 1| + |v_2 T/2|/|v_1 T/2| t)) \right]^{\lfloor v_1 T/2 \rfloor - 1} + O \left( \frac{1}{T} \right) = O \left( \frac{\log(T)}{T} \right),
\]

which complete the proof of (4.96). Properties (4.95) and (4.96) imply that \(V_{T,2}\) is of order \(O(1/T^{1-\varepsilon})\) for any \(\varepsilon > 0\).

Finally, the claim (4.43), which relates to the covariance structure of the process
\( \{ \sqrt{T} \hat{D}_T(v, \omega) \}_{(v, \omega) \in [0,1]^2} \) follows from (4.47) and (4.48) due to the fact that for each value \( \alpha \in (0, 1/4) \) sequences \( a_T \) and \( b_T \) can be found such that (4.68) and (4.74) are fulfilled.

**Proof of part (3):** For the treatment of the cumulants of orders \( l \geq 3 \), we employ the representation

\[
\hat{D}_T(v, \omega) = \frac{1}{T} \sum_{j=1}^{j/2} \sum_{k=1}^{j/2} \phi_{v,\omega,T}(j, \lambda_{k,j}) I_j(\lambda_{k,j}),
\]

where the functions \( \phi_{v,\omega,T} \) for \( (v, \omega) \in [0, 1]^2 \) are defined by

\[
\phi_{v,\omega,T}(j, \lambda) := v \{ 2[vT/2] \} (j) \{ 2[\min(vT, T/2)] \} (\lambda) - v^2 \{ 2[vT/2] \} (j) \{ 2[\min(vT, T/2)] \} (\lambda).
\]

For notational convenience, we define

\[
Y_{i,1} := Z_{1+p_i-m_i}, \quad Y_{i,2} := Z_{1+q_i-n_i}
\]

and introduce for \( v_1, ..., v_l \in [0, 1] \) the set

\[
A_T(v_1, ..., v_l) := \{ 2[v_1T/2], 2[T/2] \} \times \ldots \times \{ 2[v_lT/2], 2[T/2] \}.
\]

Using this notation and applying the product theorem for cumulants [see Theorem 2.3.6], the Gaussianity and independence of the innovation sequence \( \{ Z_t \}_{t \in \mathbb{Z}} \) and the definition of the functions \( \phi_{v,\omega,T} \), we obtain

\[
cum(\sqrt{T} \hat{D}_T(v_1, \omega_1), ..., \sqrt{T} \hat{D}_T(v_l, \omega_l)) = \sum_\nu V(\nu),
\]

where the quantity \( V(\nu) \) is defined by

\[
V(\nu) := \frac{1}{T^{l/2}} \sum_{(j_1, ..., j_l) \in A_T(v_1, ..., v_l)} \prod_{k=1}^{j_1/2} \sum_{m_1, ..., m_l = 0}^{j_1/2} \sum_{n_1, ..., n_l = 0}^{j_1/2} \prod_{s=1}^{l} \frac{1}{j_s} \sum_{p_1, q_1 = 0}^{j_1-1} \ldots \sum_{p_l, q_l = 0}^{j_l-1}
\]

\[
\times \prod_{s=1}^{l} [\phi_{v,\omega,T}(j_s, \lambda_{k_s,j_s}) \psi_{m_s} \psi_{n_s} \exp(-i\lambda_{k_s,j_s}(p_s - q_s))]
\]

\[
\times \prod_{s=1}^{l} \sum_{(a,b) \in \nu} \prod_{s=1}^{l} \sum_{(a,b) \in \nu} \text{cum}(Y_{a,b}; (a, b) \in \nu_l).
\]

Here, the summation is performed over all indecomposable partitions \( (\nu_1, ..., \nu_l) \) of the scheme

\[
Y_{1,1} \quad Y_{1,2} \\
\vdots \quad \vdots \\
Y_{l,1} \quad Y_{l,2}
\]

that are comprised of exactly \( l \) elements, each of which features two components [see Example 2.3.1]. It is easy to see that the dominating term in the sum above corresponds to the
seen that the absolute value of (4.98) is bounded by

$$j$$

which holds uniformly in partition

$$\tilde{\nu} := \bigcup_{i=1}^{l-1} (Y_{i,1}, Y_{i+1,2}) \cup (Y_{l,1}, Y_{l,2})$$

and that all other summands are of smaller order. By simple calculations, we obtain

$$V(\tilde{\nu}) = \frac{1}{T^{l/2}} \sum_{(j_1, \ldots, j_l) \in A_T(v_1, \ldots, v_l)} \sum_{k_1=1}^{j_1/2} \cdots \sum_{k_l=1}^{j_l/2} \prod_{s=1}^{l} \phi_{v_s, \omega_s, r}(j_s, \lambda_{k_s}, j_s) \sum_{m_1, \ldots, m_l=0}^{\infty} \sum_{n_1, \ldots, n_l=0}^{\infty} \prod_{s=1}^{l} \left[ \psi_{m_s, \psi_n} \right] \times \exp(-i\lambda_{k_1,j_1} (p_1 - p_l + m_l - n_l))$$

$$\times \prod_{s=2}^{l} \exp(-i\lambda_{k_s,j_s} (p_s - p_{s-1} + m_{s-1} - n_s)), \quad (4.98)$$

where the conditions

$$0 \leq q_1 = p_l - m_l + n_l \leq j_1, \quad (4.99)$$

$$0 \leq q_{i+1} = p_l - m_l + n_{i+1} \leq j_{i+1} \quad i = 1, \ldots, l - 1 \quad (4.100)$$

follow from the independence of the innovations \{\(Z_t\)\} and the specific form of the partition \(\tilde{\nu}\). The restrictions (4.99) and (4.100) imply that |\(n_i - m_i| \leq \max(j_i, j_{i+1})|\) for i = 1, ..., l - 1 have to hold. By additionally applying the bound

$$\left| \frac{1}{j_s} \sum_{k_s=1}^{j_s/2} \phi_{v_s, \omega_s, r}(j_s, \lambda_{k_s}, j_s) \exp(-i\lambda_{k_s,j_s} (r)) \right| \leq \frac{C}{r \mod j_s/2},$$

which holds uniformly in \(j_s, v, \omega\) and \(r \mod j_s/2 \neq 0\) [see (A.2) in Eichler (2008)], it can be seen that the absolute value of (4.98) is bounded by

$$\frac{C}{T^{l/2}} \sum_{(j_1, \ldots, j_l) \in A_T(v_1, \ldots, v_l)} \sum_{m_1, \ldots, m_l=1}^{\infty} \sum_{n_1, \ldots, n_l=1}^{\infty} \prod_{s=1}^{l} \left[ \psi_{m_s, \psi_n} \right] \sum_{p_1=0}^{j_1-1} \sum_{p_2=0}^{j_2-1} \sum_{p_3=0}^{j_3-1} \prod_{s=1}^{l} \left[p_s - p_{s-1} + m_{s-1} - n_s \right] \times \prod_{s=1}^{l} 1(\{p_s \notin \{z_{s1}, z_{s2}\}\}), \quad (4.101)$$

where we identify the indices 0 with l and l + 1 with 1, define \(z_{s1} := p_{s-1} - m_{s-1} + n_s\) and \(z_{s2} := p_{s+1} + m_s - n_{s+1}\) and exploit the fact that the sums over |\(p_s - p_{s-1} + m_{s-1} - n_s| \geq j_s/2
and \( p_s \in \{z_{s1}, z_{s2}\} \) are of smaller or same order. Now, we define the sets
\[
A_i := [0, j_i - 1) \setminus \{z_{i1} - 1, z_{i1} + 1\} \cup \{z_{i2} - 1, z_{i2} + 1\}
\]
for \( i = 1, ..., l \) and by substituting the summation over \( p_i \) for \( i \in \{2, ..., l\} \) by the integrals over the sets \( A_i \) for \( i \in \{2, ..., l\} \), we obtain that (4.101) is bounded by
\[
\frac{C^l}{T^{l/2}} \sum_{(j_1, ..., j_l) \in \{m_1, ..., m_l\}} \sum_{A_T(v_1, ..., v_l)} \sum_{n_1, ..., n_l=1}^{\infty} \prod_{s=1}^{l} \left| \psi_{m_s} \psi_{n_s} \right|
\times \prod_{p_s=0}^{j_1-1} A_2 \times \cdots \times A_l \left| \frac{1}{p_s - p_{s-1} + m_{s-1} - n_{s}} \right| \prod_{s=1}^{l} 1(p_s \notin \{z_{s1}, z_{s2}\}) d(p_2, ..., p_l).
\]
The arguments employed in the proof of the statement (3.60) and the fact that \( \text{card}(A_T(v_1, ..., v_l)) = 2^l \) yield that the above quantity is not larger than
\[
\frac{C^l}{T^{l/2}} \sum_{(j_1, ..., j_l) \in \{m_1, ..., m_l\}} \sum_{A_T(v_1, ..., v_l)} \sum_{n_1, ..., n_l=1}^{\infty} \prod_{s=1}^{l} \left| \psi_{m_s} \psi_{n_s} \right| \prod_{m_1 - n_1 + m_2 - n_2 + \ldots + m_l - n_l \neq 0} 1
\times \sum_{p_s=0}^{j_1-1} \log(2lT)^{-1} = C^l O\left(\frac{\log(T)^l}{T^{l/2-1}}\right) = o(1)
\]
for some constant \( C > 0 \), which completes the proof of part (3). \( \square \)

**Proof of Lemma 4.6.2:**

In order to establish asymptotic stochastic equicontinuity for the sequence \( \{\sqrt{T} \tilde{D}_T(v, \omega)\}_{(v, \omega) \in [0, 1]^2} \), we first note that we have the decomposition
\[
\tilde{D}_T(v, \omega) = \tilde{D}_{T, 1}(v, \omega) - \tilde{D}_{T, 2}(v, \omega),
\]
where \( \tilde{D}_{T, 1}(v, \omega) \) and \( \tilde{D}_{T, 2}(v, \omega) \) for \((v, \omega) \in [0, 1]^2\) are defined by
\[
\tilde{D}_{T, 1}(v, \omega) := \frac{v}{T} \sum_{k=1}^{\lfloor \omega / 2^{T/2} \rfloor} I_{2^{T/2}}(\lambda_{k, 2^{T/2}}),
\]
\[
\tilde{D}_{T, 2}(v, \omega) := \frac{v^2}{T} \sum_{k=1}^{\lfloor \omega / 2^{T/2} \rfloor} I_{T}(\lambda_{k, 2^{T/2}}).
\]
From (4.102), it follows that it is sufficient to show asymptotic stochastic equicontinuity for the processes \( \{\sqrt{T} \tilde{D}_{T, 1}(v, \omega)\}_{(v, \omega) \in [0, 1]^2} \) and \( \{\sqrt{T} \tilde{D}_{T, 2}(v, \omega)\}_{(v, \omega) \in [0, 1]^2} \) separately. For the sake
of brevity, we restrict ourselves to the first summand in (4.102) and note that the equivalent statement for the second summand can be obtained by employing the fact $D_{T,2}(v, \omega) = v^2 D_{T,1}(1, \omega)$. For this proof, we use the representation

$$
\hat{D}_{T,1}(v, \omega) = \frac{1}{T} \sum_{j=1}^{T} \sum_{k=1}^{[j/2]} \phi_{v,\omega,T}^{(1)}(j, \lambda_{k,j}) I_j(\lambda_{k,j}),
$$

where the function $\phi_{v,\omega,T}^{(1)}$ for $(v, \omega) \in [0, 1]^2$ is defined by

$$
\phi_{v,\omega,T}^{(1)}(j, \lambda_{k,j}) := v 1_{(2[jvT/2])}(j) 1_{[0, 2\pi/2]}(\lambda_{k,j}).
$$

(4.103)

Furthermore, the distance $d_\beta$ defined in (4.30) is a semi-metric on the set

$$
P_T := \left\{ (v, \omega) \middle| v \in \{2/T, 4/T, ..., 1\}, \omega \in \{0, 1/(2\sqrt{vT/2})\}, ..., 1 - 1/(2\sqrt{vT/2})\} \right\}
$$

and we have the equality

$$
P\left( \sup_{(v_1, \omega_1), (v_2, \omega_2) \in [0, 1]^2 : d_\beta((v_1, \omega_1), (v_2, \omega_2)) < \delta} \sqrt{T} |\hat{D}_{T,1}(v_1, \omega_1) - \hat{D}_{T,1}(v_2, \omega_2)| > \eta \right)
= P\left( \sup_{(v_1, \omega_1), (v_2, \omega_2) \in P_T : d_\beta((v_1, \omega_1), (v_2, \omega_2)) < \delta} \sqrt{T} |\hat{D}_{T,1}(v_1, \omega_1) - \hat{D}_{T,1}(v_2, \omega_2)| > \eta \right).
$$

(4.104)

In order to show the equivalent of the statement (4.31) for the process $\{\sqrt{T} \hat{D}_{T,1}(v, \omega)\}_{(v, \omega) \in [0, 1]^2}$, it is therefore sufficient to show that for each $\eta > 0$ and $\epsilon > 0$ there exists some $\delta > 0$ such that the inequality

$$
P\left( \sup_{(v_1, \omega_1), (v_2, \omega_2) \in P_T : d_\beta((v_1, \omega_1), (v_2, \omega_2)) < \delta} \sqrt{N} |\hat{D}_{T,1}(v_1, \omega_1) - \hat{D}_{T,1}(v_2, \omega_2)| > \eta \right) \leq \epsilon
$$

(4.105)

holds for $T$ sufficiently large. As in the proof of Lemma 3.4.1, this assertion is implied by the following two results:

**Lemma 4.6.3**

For every $\beta \in (0, 1/4)$, there exists some constant $C > 0$ such that for sufficiently large $T \in \mathbb{N}$ the inequality

$$
P(\sqrt{T}|\hat{D}_{T,1}(v_1, \omega_1) - \hat{D}_{T,1}(v_2, \omega_2)| > \eta d_\beta((v_1, \omega_1), (v_2, \omega_2))) \leq 96 \exp(-\sqrt{\eta/C})
$$

(4.106)

holds uniformly in $(v_1, \omega_1), (v_2, \omega_2) \in P_T$ and for all $\eta > 0$, where $d_\beta$ denotes the semi-metric

$$
d_\beta((v_1, \omega_1), (v_2, \omega_2)) = (|v_1 - v_2| + |\omega_1 - \omega_2|)^{\beta/2}.
$$
Lemma 4.6.4
For the covering integral
\[
J(\kappa, d_\beta, \mathcal{P}_T) = \int_0^\infty \left[ \log \left( \frac{48 \mathcal{N}(u, d_\beta, \mathcal{P}_T)}{u} \right) \right]^2 du
\]
of the set \( \mathcal{P}_T \) with respect to the semi-metric
\[
d_\beta((v_1, \omega_1), (v_2, \omega_2)) = (|v_1 - v_2| + |\omega_1 - \omega_2|)^{\beta/2},
\]
it holds
\[
\lim_{\kappa \to 0} \lim_{T \to \infty} J(\kappa, d_\beta, \mathcal{P}_T) = 0.
\]

Proof of Lemma 4.6.3

By similar arguments, as are presented in the proof of Lemma 3.4.3, we restrict ourselves to a proof of the inequality
\[
\mathbb{E}\left(T^{k/2}(\hat{D}_{T,1}(v_1, \omega_1) - \hat{D}_{T,1}(v_2, \omega_2))^k\right) \leq (2k)!C^k d_\beta((v_1, \omega_1), (v_2, \omega_2))^k,
\]
which holds uniformly in \((v_1, \omega_1), (v_2, \omega_2) \in \mathcal{P}_T\) and all even integers \(k \in \mathbb{N}\). Then, the arguments provided in the proof of Lemma 3.4.3 yield the assertion (4.106). As was shown in (3.135), the statement (4.107) follows if we show that the inequality
\[
\text{cum}_l(\sqrt{T}(\hat{D}_{T,1}(v_1, \omega_1) - \hat{D}_{T,1}(v_2, \omega_2))) \leq (2l)!C^l d_\beta((v_1, \omega_1), (v_2, \omega_2))^l
\]
holds uniformly in \((v_1, \omega_1), (v_2, \omega_2) \in \mathcal{P}_T\) and \(l \in \mathbb{N}\).

Proof of (4.108): In order to show (4.108), we consider the cases \(l = 1, l = 2\) and \(l \geq 3\) separately. We take \((v_1, \omega_1) \neq (v_2, \omega_2) \in \mathcal{P}_T\) and assume without loss of generality that \(v_1 \leq v_2, \omega_1 \leq \omega_2\). In the case \(l = 1\) (4.32) yields that for \((v_1, \omega_1) \neq (v_2, \omega_2) \in \mathcal{P}_T\)
\[
\text{cum}_1(\sqrt{T}(\hat{D}_{T,1}(v_1, \omega_1) - \hat{D}_{T,1}(v_2, \omega_2))) = O\left(\frac{1}{\sqrt{T}}\right)
\]
\[
\leq C(|v_1 - v_2| + |\omega_1 - \omega_2|)^{1/2} \leq C2^l l d_\beta((v_1, \omega_1), (v_2, \omega_2))
\]
where in the second step we used the fact that
\[
\frac{1}{T} \leq C(|v_1 - v_2| + |\omega_1 - \omega_2|).
\]
For the case \(l = 2\), we obtain by following the argumentation presented in the calculations of the covariances and by applying the linearity of the covariance that for any \(\beta \in (0, 1/4)\)
it holds
\[
\text{cum}_2(\sqrt{T}(\hat{D}_{T,1}(v_1, \omega_1) - \hat{D}_{T,1}(v_2, \omega_2))) \\
\leq \frac{1}{2\pi} \left( \int_{\omega_1}^{\omega_2} f^2(\lambda) d\lambda (v_1^3 + v_2^3 - 2v_1^2 v_2) + v_2^3 \int_{\omega_1}^{\omega_2} f^2(\lambda) d\lambda \right) + O\left( \frac{1}{T^\beta} \right) \\
\leq \frac{1}{2\pi} \left( \int_{\omega_1}^{\omega_2} f^2(\lambda) d\lambda v_2 (v_2^2 - v_1^2) + \int_{\omega_1}^{\omega_2} f^2(\lambda) d\lambda \right) + O\left( \frac{1}{T^\beta} \right) \\
\leq C(|v_1 - v_2| + |\omega_1 - \omega_2|) + (|v_1 - v_2| + |\omega_1 - \omega_2|)^\beta \leq 2^2! C^2 (|v_1 - v_2| + |\omega_1 - \omega_2|)^\beta \\
= 2^2! C^2 d_\beta((v_1, \omega_1), (v_2, \omega_2))^2,
\]
where we employed (4.109) and the fact that there exists a constant \( C > 0 \) such that the bound \( x \leq C x^\beta \) holds uniformly in \( x \in [0, 2] \).
For the treatment of the case \( l \geq 3 \), we define
\[
\phi(j, \lambda) := \phi^{(1)}_{\nu_1, \omega_1, T}(j, \lambda) - \phi^{(1)}_{\nu_2, \omega_2, T}(j, \lambda),
\]
where the function \( \phi^{(1)}_{\nu, \omega, T} \) for \( (v, \omega) \in [0, 1]^2 \) was defined in (4.103). This definition yields the representation
\[
\sqrt{T}(\hat{D}_{T,1}(v_1, \omega_1) - \hat{D}_{T,1}(v_2, \omega_2)) = \frac{1}{\sqrt{T}} \sum_{j=1}^{T} \sum_{k=1}^{j/2} \phi(j, \lambda_{k,j}) I_j(\lambda_{k,j})
\]
for the increments of the process \( \{ \sqrt{T} \hat{D}_{T,1}(v, \omega) \}_{(v, \omega) \in [0, 1]^2} \). By following the various steps in the calculation of the higher order cumulants in the proof of (4.29), we get that
\[
\text{cum}_l(\sqrt{T}(\hat{D}_{T,1}(v_1, \omega_1) - \hat{D}_{T,1}(v_2, \omega_2))) = 2^l (2l)! C^l O\left( \frac{\log(T)^{l-1}}{T^{l/2 - 1}} \right)
\]
uniformly in \((v_1, \omega_1), (v_2, \omega_2) \in \mathcal{P}_T\) and for any \( \varepsilon > 0 \). By taking \( \varepsilon = 1/6 - \beta/2 \) and employing (4.109) we thus obtain
\[
\text{cum}_l(\sqrt{T}(\hat{D}_{T,1}(v_1, \omega_1) - \hat{D}_{T,1}(v_2, \omega_2))) \leq 2^l (2l)! C^l d_\beta((v_1, \omega_1), (v_2, \omega_2))^l,
\]
which completes the proof of Lemma 4.6.3. \( \square \)

**Proof of Lemma 4.6.4**

The definition of the semi-metric \( d_\beta \) implies the bound
\[
N(u, d_\beta, \mathcal{P}_T) \leq \frac{C}{u^{4/\beta}}
\]
for the covering numbers $N(u, d, \mathcal{P}_T)$. From this bound, it follows that

$$J(\kappa, d, \mathcal{P}_T) = \int_0^\kappa \left[ \log \left(48N(u, d, \mathcal{P}_T)^2u^{-1}\right) \right]^2 du \leq C \int_0^\kappa [\log(48u^{-8/\beta-1})]^2 du$$

$$= C \left[ \int_0^\kappa \log(48)^2 du - 2 \int_0^\kappa \log(48) \log(u^{8/\beta+1}) du + \int_0^\kappa (\log(u^{8/\beta+1}))^2 du \right] \to 0$$
as $\kappa \to 0$.

**Proof of part b) of Theorem 4.3.2:** By similar arguments, as were employed in the proof of part a), we obtain that under the Assumptions 4.3.1 it holds

$$\sup_{(v, \omega) \in [0, 1]^2} \sqrt{T} |\hat{D}_T(v, \omega) - \mathbb{E}(D_T(v, \omega))| = O_P(1).$$

Therefore, the claim follows from the statement (4.34).

### 4.6.2 Proof of Theorem 4.4.1

For the proof of Theorem 4.4.1, we employ the same course of argument as in the proof of Theorem 3.1.6 and consider without loss of generality the case $d = 1$. Furthermore, we suppress the argument $T$ when referring to the sequence $p = p(T)$.

We note that the process

$$X_t^{AR}(p) := \sum_{j=1}^p a_{j,p} X_{t-j}^{AR}(p) + Z_t^{AR}(p),$$

where the coefficients $(a_{1,p}, \ldots, a_{p,p})$ were defined in (3.19) and $\{Z_t^{AR}(p)\}_{t \in \mathbb{Z}}$ denotes a sequence of independent centred Gaussian random variables with variance

$$\sigma_p = \mathbb{E} \left[ (X_t - \sum_{j=1}^p a_{j,p} X_{t-j})^2 \right],$$
corresponds to the best approximation of the observed data $\{X_{t,T}\}_{t=1, \ldots, T}$ by an AR($p$) model. As was explained in the proof of Theorem 3.1.6, it follows from Assumption 3.1.3 that the time series model $\{X_t^{AR}(p)\}_{t \in \mathbb{Z}}$ has a MA($\infty$) representation of the form

$$X_t^{AR}(p) = \sum_{j=0}^\infty \psi_j^{AR}(p) Z_{t-j}^{AR}(p)$$
for \( p \) sufficiently large. Moreover, Assumption 3.1.3 and the condition (3.24) yield that the bootstrap sample \( \{X_{t,T}^*\}_{t=1,\ldots,T} \) has a \( MA(\infty) \) representation

\[
X_{t,T}^* = \sum_{l=0}^{\infty} \hat{\psi}^{AR}_l(p)Z^*_{t-l}
\]

(4.110)

for \( p \) sufficiently large. Note that the representation (4.110) corresponds to a stationary time series model following a linear representation, whose coefficients are random. For a proof of Theorem 4.4.1, we intend to modify the arguments, which were made in the proof of Theorem 4.3.2, where we established weak convergence of the empirical process \( \{\sqrt{T}D_T(v,\omega)\}_{(v,\omega)\in[0,1]^2} \). Therefore, we note that for the proof of Theorem 4.3.2 we used the property

\[
\frac{(\sum_{m=0}^{\infty} |\psi_m|)^{q_1}(\sum_{l=0}^{\infty} |\psi_l|)^{q_2}}{T} = O\left(\frac{1}{T}\right)
\]

(4.111)

for \( q_1, q_2 \in \mathbb{N} \) for bounding various error terms arising in the calculations. In the situation of Theorem 4.3.2, the property (4.111) is a consequence of the summability condition (4.10), which we imposed on the deterministic sequence \( \{\psi_l\}_{l\in\mathbb{N}} \) of linear coefficients of the \( MA(\infty) \) representation of the time series \( \{X_{t,T}\}_{t=1,\ldots,T} \). In order to make sure that the arguments provided in the proof of Theorem 4.3.2 carry over to the situation of Theorem 4.4.1, where the sequence \( \{\hat{\psi}^{AR}_l\}_{l\in\mathbb{N}} \) is stochastic, we have to show that the error terms

\[
\frac{(\sum_{m=0}^{\infty} |\hat{\psi}^{AR}_m(p)|)^{q_1}(\sum_{l=0}^{\infty} l|\hat{\psi}^{AR}_l(p)|)^{q_2}}{T}
\]

are of order \( O_P(1/T) \). Thus, the assertion of Theorem 4.4.1 is a consequence of the property

\[
\sum_{l=0}^{\infty} |l||\hat{\psi}^{AR}_l(p)| = O_P(1).
\]

(4.112)

However, this identity follows from the statements (3.159) and (3.160), which have been established in the proof of Theorem 3.1.6.


Selbstständigkeitserklärung

Hiermit versichere ich, dass ich diese Promotionsarbeit mit dem Thema ”Detecting Deviations from Stationarity” selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt wurden, sowie Zitate kenntlich gemacht habe.
Persönliche Daten
Geburtsdatum: 12/06/1987
Geburtsort: Münster

Ausbildung

seit 08/2011 Promotion im Bereich Statistik, Ruhr-Universität, Bochum.


Vertiefung ökonomischer und finanzmarktheoretischer Konzepte zur Ergänzung analytischer Fähigkeiten, insbesondere Volkswirtschaftslehre, Finanzmark- und Wettbewerbstheorie und Bilanzierungsrechnung.

10/2010-06/2011 Auslandsjahr, Durham University, Durham, Großbritannien.
Kurse in angewandter Mathematik, insbesondere Finanzmathematik und Decision Theory.

Note: 1.4 (von 1.0-4.0). Thematische Schwerpunkte: Theorie der Firma, Mikroökonomie und Ökonometrie. Bachelor Arbeit im Bereich Ressourcenökonomik: "Intertemporale Allokation nicht erneuerbarer natürlicher Ressourcen".

06/1997-06/2005 Abitur (Gymnasium), Schiller-Schule, Bochum.
Note: 2.0 (von 1.0-4.0). Besondere Leistungen: Ich habe die 10. Klasse übersprungen und im Rahmen des Projektes "Schüleruniversität" an Veranstaltungen verschiedener Fakultäten der Ruhr-Universität Bochum teilgenommen.

Berufserfahrung

Mitwirkung an Projekten im Bereich Risikomanagement mit Schwerpunkt Bankenwesen.

seit 04/2012 Übungsgruppenleiter und Administrator für die Veranstaltung "Statistische Methodenlehre für Psychologen", Ruhr-Universität, Bochum.

seit 08/2011 Statistische Beratung, Ruhr-Universität, Bochum.